The HLA-B landscape of Africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection

Sanchez-Mazas, Alicia; Černý, Viktor; Di, Da; Buhler, Stéphane; Podgorna, Eliska; Chevallier, Elodie; Brunet, Lydie; Weber, Stephan Ole Daniel; Kervaire, Barbara Nelly; Testi, Manuela; Andreani, Marco; Tiercy, Jean-Marie; Villard, Jean; Nunes, Jose Manuel

How to cite

SANCHEZ-MAZAS, Alicia et al. The HLA-B landscape of Africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. In: Molecular Ecology, 2017, vol. 26, n° 22, p. 6238–6252. doi: 10.1111/mec.14366

This publication URL: https://archive-ouverte.unige.ch/unige:102913

Publication DOI: 10.1111/mec.14366

Received: 10 June 2016 | Accepted: 11 September 2017

DOI: 10.1111/mec.14366

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

The HLA-B landscape of Africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection

Alicia Sanchez-Mazas[1,2*] | Viktor Černý[3] | Da Di[1] | Stéphane Buhler[1,4] | Eliška Podgorná[5] | Elodie Chevallier[1] | Lydie Brunet[1,4] | Stephan Weber[1] | Barbara Kervaire[4] | Manuela Testi[6] | Marco Andreani[6] | Jean-Marie Tiercy[2,4] | Jean Villard[2,4] | José Manuel Nunes[1,2*]

[1]Department of Genetics and Evolution – Anthropology Unit, Laboratory of Anthropology, Genetics and Peopling History (AGP), University of Geneva, Geneva, Switzerland

[2]Institute of Genetics and Genomics in Geneva (IGE3), Geneva, Switzerland

[3]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

[4]Department of Genetic and Laboratory Medicine, Transplantation Immunology Unit and National Reference Laboratory for Histocompatibility (UIT/LNRH), Geneva University Hospitals, Geneva, Switzerland

[5]Department of the Archaeology of Landscape and Archaeobiology, Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, Czech Republic

[6]Laboratory of Immunogenetics and Transplant Biology, IME Foundation, Policlinic of the University of Tor Vergata, Rome, Italy

Correspondence
Alicia Sanchez-Mazas, Department of Genetics and Evolution – Anthropology Unit, University of Geneva Sciences II, Geneva, Switzerland.
Email: alicia.sanchez-mazas@unige.ch

Funding information
Swiss National Science Foundation, Grant/Award Number: 31003A_144180; Grant Agency of the Czech Republic, Grant/Award Number: 13-37998S-P505

Abstract

Human leukocyte antigen (HLA) genes play a key role in the immune response to infectious diseases, some of which are highly prevalent in specific environments, like malaria in sub-Saharan Africa. Former case–control studies showed that one particular HLA-B allele, B*53, was associated with malaria protection in Gambia, but this hypothesis was not tested so far within a population genetics framework. In this study, our objective was to assess whether pathogen-driven selection associated with malaria contributed to shape the HLA-B genetic landscape of Africa. To that aim, we first typed the HLA-A and -B loci in 484 individuals from 11 populations living in different environments across the Sahel, and we analysed these data together with those available for 29 other populations using several approaches including linear modelling on various genetic, geographic and environmental parameters. In addition to relevant signatures of populations' demography and migrations history in the genetic differentiation patterns of both HLA-A and -B loci, we found that the frequencies of three HLA alleles, B*53, B*78 and A*74, were significantly associated with *Plasmodium falciparum* malaria prevalence, suggesting their increase through pathogen-driven selection in malaria-endemic environments. The two HLA-B alleles were further identified, by high-throughput sequencing, as B*53:01:01 (in putative linkage disequilibrium with one HLA-C allele, C*04:01:01:01) and B*78:01 in all but one individuals tested, making them appropriate candidates to malaria protection. These results highlight the role of environmental factors in the evolution of the HLA polymorphism and open key perspectives for functional studies focusing on HLA peptide-binding properties.

KEYWORDS
African populations, geographic patterns, HLA polymorphism and disease associations, human population genetics, malaria protection, pathogen-driven selection

*Both authors contributed equally to the work.

1 | INTRODUCTION

The genetic history of Africa is one of the most fascinating in evolutionary genetics. First, this continent is assumed to be the homeland of modern humans according to both fossil records going back between 195,000 and 300,00 years (Aubert et al., 2012; Hublin et al., 2017; McDougall, Brown, & Fleagle, 2005; Richter et al., 2017) and genetic evidence (Campbell, Hirbo, Townsend, & Tishkoff, 2014; for a review); second, it is currently inhabited by 1.2 billion people speaking more than 2,000 languages (Lewis, Simons, & Fennig, 2016) and living in highly diverse environments, from arid deserts to savannahs, tropical rain forests and extended highlands. The genetic structure of African populations is thus expected to reveal highly heterogeneous genetic signals due both to a complex demographic history characterized by thousands of years of populations' migrations, assimilations and/or replacements (Excoffier, Pellegrini, Sanchez-Mazas, Simon, & Langaney, 1987; Poloni et al., 2009; Tishkoff et al., 2009), and to various kinds and/or intensities of natural selection operating in response to peculiar pathogenic environments and resources conditioning human survival and lifestyles (Gomez, Hirbo, & Tishkoff, 2014; Triska et al., 2015).

In this context, the analysis of the major histocompatibility complex (MHC)-encoded molecules, named human leukocyte antigens (HLA) in our species, is particularly stimulating as they assume vital functions in the organism by presenting pathogen-derived peptides to T cells as the main initial step of the adaptive immune response (Parham & Janeway, 2015). This is the reason why the huge level of polymorphism observed at HLA genes is generally regarded as a consequence of balancing selection favouring heterozygous individuals in pathogen-rich environments, an early hypothesis (Doherty & Zinkernagel, 1975; Hughes & Nei, 1988) that was further sustained, at the population level, by neutrality tests (Buhler & Sanchez-Mazas, 2011; Solberg et al., 2008) and correlation studies (Prugnolle et al., 2005; Qutob et al., 2012; Sanchez-Mazas, Lemaitre, & Currat, 2012). However, single HLA (more generally, MHC) loci may also display reduced levels of diversity that would not necessarily reduce the fitness of individuals thanks to the high polymorphism existing at other loci (Buhler, Nunes, & Sanchez-Mazas, 2016; de Groot & Bontrop, 2013). In addition, specific HLA alleles are known or supposed to confer resistance or susceptibility to various auto-immune (e.g., IDDM, coeliac disease) and infectious (e.g., HIV, tuberculosis, malaria) diseases (reviewed by Naranbhai & Carrington, 2017; Sollid, 2017; and Trowsdale & Knight, 2013; among others), indicating that directional selection may also drive the evolution of HLA genes. As an example, the HLA-DRB1*12:02:01 allele has been suggested to have increased in frequency in a Mongolian population that migrated from the north to the southwest of China about 700 years ago as a consequence of its settlement in a new environment where this allele would have been protective (Sun et al., 2015).

In Africa, a study conducted some 25 years ago in Gambia showed that one particular allele of the HLA-B locus, HLA-B*53, exhibited higher frequencies in a healthy control sample than in patients suffering from malaria and concluded that this allele was protective against *Plasmodium falciparum* (*P. falciparum* hereafter) infection (Hill et al., 1991). However, probably because HLA data for African populations remained very scarce since then, it is not known whether this putative protection had significant effects on HLA-B frequencies in this continent. Actually, the patterns of HLA genetic variation observed either worldwide or within continents are generally well predicted by geography (Buhler & Sanchez-Mazas, 2011; Cao et al., 2004; Currat, Poloni, & Sanchez-Mazas, 2010; Di & Sanchez-Mazas, 2011, 2014; Di, Sanchez-Mazas, & Currat, 2015; Mack et al., 2000; Sanchez-Mazas, 2001; Sanchez-Mazas, Buhler, & Nunes, 2013), suggesting that natural selection did not blur the signatures of human migration history. However, because pathogens are also geographically distributed, demographic history and natural selection may generate confounding effects that may be difficult to disentangle on simple patterns of inter-population differentiation (Meyer, Single, Mack, Erlich, & Thomson, 2006). This motivates the use of more powerful approaches where various kinds of genetic, geographic and environmental parameters are simultaneously taken into account.

In this study, our objective was to assess whether directional selection driven by *P. falciparum* infection could have shaped the HLA genetic landscape of Africa, particularly at the HLA-B locus where one allele, HLA-B*53, was suggested to be protective to malaria. To address this question, we first typed the HLA-A and -B loci in almost 500 individuals from 11 populations located in distinct areas of the Sahel (transitional ecoregion lying between the Sahara Desert to the north and the wooded savannas to the south) where HLA variation had seldom been investigated—thus largely improving the data set suitable for this study—and where malaria exhibits contrasted levels of endemicity. We then analysed these new data together with those published for 29 other populations from different regions of Africa through several approaches including linear modelling, which allowed to take into account the prevalence of *P. falciparum* malaria across the whole continent in addition to other genetic, geographic and environmental parameters. We analysed HLA-A simultaneously to HLA-B in the same population samples because the former locus is assumed to be less intensely submitted to selection than the latter (Buhler & Sanchez-Mazas, 2011; Di et al., 2015; Sanchez-Mazas et al., 2013; Solberg et al., 2008) and is thus expected to reveal contrasted results in case of substantial selection acting on HLA-B. Finally, because our results uncovered several HLA alleles as putative targets of *P. falciparum*-driven selection, we determined the precise nucleotide sequence of these alleles using a high-throughput sequencing technology.

2 | MATERIAL AND METHODS

2.1 | Population samples

The data used in this study represent 484 individuals from 11 populations (Sudanese Arabs (SUD), Rashaayda Bedouins (RAS), Beja Hadendowa (BEJ), Nubians (NUB), Daza (DAZ), Baggara Arabs (BAG),

Maba (MAB), Dangaléat (DAN), Mossi (MOS), Gurunsi Kassena (GRS) and Gourmantché (GOU)) following distinct lifestyles as sedentary farmers, semi-nomadic or nomadic pastoralists across the Sahel (Sudan, Chad and Burkina Faso) and specifically HLA-typed for this study (*Sahel set* hereafter, Table 1), plus 29 population samples from published sources (Assane et al., 2010; Cao et al., 2004; Galgani et al., 2013; Gomez-Casado et al., 2000; Gourraud et al., 2014; Hajjej et al., 2006; Mack & Erlich, 2007; Mahfoudh et al., 2013; Modiano et al., 2001; Norman et al., 2013; Nunes et al., 2010; Piancatelli et al., 2004; Riccio et al., 2013; Sanchez-Mazas, 2007; Sanchez-Mazas & Tiercy, 2007; Spinola, Bruges-Armas, Middleton, & Brehm, 2005; Testi et al., 2015; Torimiro et al., 2006). These 40 populations (Table S1) are distributed among 19 countries of Eastern (ESAFR), Central (CSAFR), Western (WSAFR) and Southern (SSAFR) sub-Saharan Africa as well as of North Africa (NAFR), and speak languages belonging to three of the four main African linguistic families, that is, Niger-Congo (NC), Afro-Asiatic (AA) and Nilo-Saharan (NS) (Lewis et al., 2016).

2.2 | DNA extraction and HLA typing (Sahel set)

All individuals from the *Sahel set* were healthy unrelated volunteers who gave their informed verbal consent for the study and whose sampling was authorized by the appropriate authorities of their respective country, as described elsewhere (Podgorná et al., 2015). DNA was extracted from saliva samples using *Oragene*^{TM} DNA (OG 500) collection kits as described in the protocols provided by the supplier of the technology used (DNA Genotek, Ottawa, ON, Canada) and typed for HLA-A and -B exons 2 and 3 (peptide-binding region) by reverse PCR-SSO typing on microbeads arrays (LABType, One Lambda Inc., Canoga Park, CA, USA). Whenever necessary, ambiguities were resolved by PCR-SSP (Olerup, Stockholm, Sweden). In a further step, B*53-, B*78- and A*74-positive individuals (except some B*78 and A*74 carriers for which initial typings already revealed the precise allele) were retyped by a high-throughput sequencing technology (Holotype HLA X2, Omixon Biocomputing Ltd, Budapest, Hungary and NXType NGS, One Lambda Inc., Canoga Park, CA, USA) to identify the exact molecular variant.

2.3 | Environmental parameters

A detailed map of *P. falciparum* malaria prevalence in Africa provided by the Malaria Atlas Project (http://www.map.ox.ac.uk Gething et al., 2011) was used in this study. This map was drawn from a database of *P. falciparum* parasite rates (*PfPR*) corresponding to the proportions of individuals sampled in each location for which the pathogen was detected in the blood, standardized for an age range comprised between 2 and 10 years old. The data were then interpolated to provide these values in a grid of 5 × 5 km illustrated by a continuous coloured map for the whole continent. To analyse the relationship between HLA allele frequencies and *P. falciparum* malaria prevalence in Africa, we assigned *PfPR* values to the locations of the 40 population samples used in this study (Table S1). This assignment step was made using the QGIS PLUGIN *Point sampling tool* [https://github.com/borysiasty/pointsamplingtool, version: 0.4 in QGIS Official Plugin Repository], with the precise geographical coordinates of

TABLE 1 New population samples analysed in this study (Sahel set)

Acronym	Population name	Linguistic family	Linguistic branch	Geographic region	Country	Location	LAT	LONG	Lifestyle	A-N	B-N
SUD	Sudanese Arabs	AA	Semitic	ESAFR	Sudan	Karima to Kerma	18.41	30.75	Sedentary	46	46
RAS	Rashaayda Bedouins	AA	Semitic	ESAFR	Sudan	Abu Talha	15.35	36.30	Semi-nomadic	52	52
BEJ	Beja Hadendowa	AA	Cushitic	ESAFR	Sudan	North of Kassala	15.55	36.34	Nomadic	48	48
NUB	Nubians	NS	Eastern-Sudanic	ESAFR	Sudan	Kerma to Wadi Halfa	20.79	30.48	Sedentary	54	54
DAZ	Daza	NS	Saharan	CSAFR	Chad	Between Faya, Ounianga and Fada	18.18	20.55	Semi-nomadic	41	41
BAG	Baggara Arabs	AA	Semitic	CSAFR	Chad	Batha, west of Ati	13.18	17.95	Nomadic	50	50
MAB	Maba	NS	Maban	CSAFR	Chad	Abeche	13.86	20.83	Sedentary	41	41
DAN	Dangaléat	AA	Chadic	CSAFR	Chad	West of Mongo	12.16	18.52	Sedentary	49	49
MOS	Mossi	NC	Atlantic-Congo	WSAFR	Burkina-Faso	East of Ziniare	12.59	−1.34	Sedentary	35	35
GRS	Gurunsi Kassena	NC	Atlantic-Congo	WSAFR	Burkina-Faso	Po	11.17	−1.14	Sedentary	32	32
GOU	Gourmantché	NC	Atlantic-Congo	WSAFR	Burkina-Faso	Near Pama	11.24	0.72	Sedentary	36	36

Linguistic families: AA, Afro-Asiatic; NS, Nilo-Saharan; NC, Niger-Congo.
Geographic regions: ESAFR, Eastern sub-Saharan Africa; CSAFR, Central sub-Saharan Africa; WSAFR, Western sub-Saharan Africa.
LAT: latitude (in decimals); LONG: longitude (in decimals); A-N and B-N: number of individuals successfully typed for HLA-A and HLA-B, respectively.

the samples, except for four of them for which no precise geographical coordinates were available (SUD2, KEN, DOG, GUI) and to which we assigned country average values (for Sudan, Kenya, Mali and Guinea-Bissau, respectively). We used the *PfPR* values from year 2000 (i.e., before malaria control efforts were undertaken, Bhatt et al., 2015) to account for the fact that the samples analysed in this study were composed (to few exceptions) of individuals born before year 2000. The map is shown in Figure 1, where the 40 HLA-typed populations are also located based on their geographic coordinates.

To assess a possible relationship between the HLA diversity of each population and its potential exposure to pathogens, we also used the number of infectious diseases (NID) observed in each African country (as provided by the GIDEON online database, http://www.gideononline.com, accessed on March 2016) as a measure of pathogen richness and we assigned these values to the populations tested in this study (Table S1).

2.4 | Statistical analyses

Due to the scarcity of African population samples typed for HLA at resolution levels higher than the first-field (which corresponds to allele groups (or lineages) according to the official HLA nomenclature (Robinson et al., 2015)), we used that same resolution level for the newly typed populations, thus allowing to compare them with as many other data as possible. As the new population samples are represented by an average of 44 individuals per population (484 individuals of 11 populations), this level of resolution also allows to avoid ambiguous allele assignments, to reduce sampling bias while estimating HLA allele and haplotype frequencies and hence to use reliable statistics for the analyses.

Intrapopulation analyses were first carried out on the 11 population samples of the *Sahel set*. We used the GENE[RATE] programs (Nunes, 2015; Nunes, Buhler, Roessli, & Sanchez-Mazas, 2014) to estimate allele and haplotype frequencies with an expectation–maximization algorithm; to test Hardy–Weinberg equilibrium (HWE) with a likelihood-ratio test; to test selective neutrality with a bootstrapped Ewens–Watterson–Slatkin (BEWS) using false discovery rate (FDR) correction for multiple testing; and to test linkage disequilibrium (LD) for each HLA-A~B haplotype (the HLA-A and HLA-B loci being at about 1.4 Mb and 0.8 cM from each other, Litwin, 1989) using standardized residuals.

Interpopulation analyses were then performed on the total set of 40 populations. We first plotted nonmetrical multidimensional scaling (NMDS, Kruskal, 1964) analyses based on Prevosti, Ocaña, and

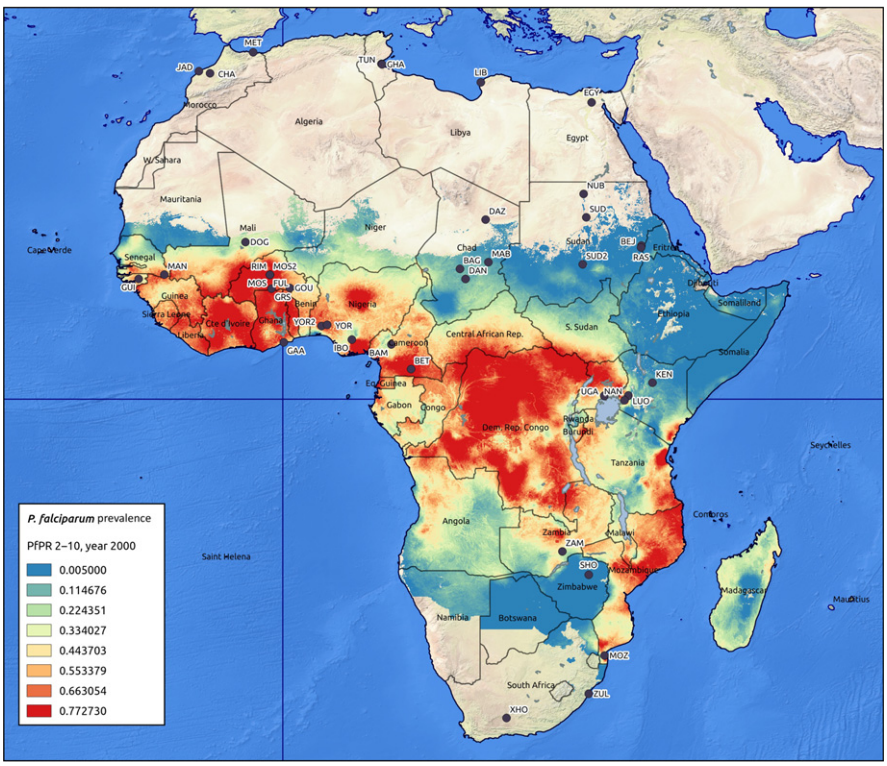

FIGURE 1 Map of Africa showing the prevalence of *Plasmodium falciparum* malaria (estimated by *P. falciparum* parasite rates *PfPr*, coloured areas) according to the Malaria Atlas Project (http://www.map.ox.ac.uk) and the location of the populations analysed in this study. Population names (see Table S1 for details): CHA, Moroccans Chaouya; JAD, Moroccans El Jadida; MET, Moroccans Metalsa; TUN, Tunisians South; GHA, Tunisians Ghannouch; LIB, Libyans Benghazi; EGY, Egyptians; SUD2, Sudanese; SUD, Sudanese Arabs; RAS, Rashaayda Bedouins; BEJ, Beja Hadendowa; NUB, Nubians; LUO, Luo; NAN, Nandi; KEN, Kenyans; UGA, Ugandans; DAZ, Daza; BAG, Baggara Arabs; MAB, Maba; DAN, Dangaléat; BAM, Bamileke; BET, Beti; IBO, Ibo; YOR, Yoruba; YOR2, Yoruba; GAA, Ga-Adangbe; MOS, Mossi; MOS2, Mossi; RIM, Rimaibé; FUL, Fulani; GRS, Gurunsi Kassena; GOU, Gourmantche; DOG, Dogons; GUI, Guiné-Bissauans; MAN, Mandenka; MOZ, Mozambicans; ZAM, Zambians; SHO, Shona; ZUL, Zulu; XHO, Xhosa

Alonso's (1975) genetic distances to assess the overall genetic relationships among populations at each locus. Spatial autocorrelation was checked by analysing Moran's I coefficients of each allele as a function of geographic distance (estimated by Vincenty's ellipsoid method) classes (Barbujani, 2000), and possible population structures were assessed by performing analyses of genetic variance (ANOVA, Excoffier, Smouse, & Quattro, 1992) on predefined geographic and/or linguistic groups. Finally, significant associations between different genetic (heterozygosity, number of alleles and allele frequencies), geographic (latitude, longitude) and environmental (*P. falciparum* malaria prevalence and pathogen richness) variables were assessed by means of linear modelling. For each association sought, an initial model containing all independent variables having a Pearson or Spearman correlation coefficient with the response variable larger than 0.5 was reduced to a final (retained) model including the smallest number of independent variables as possible and being nonsignificantly different from the initial model. To avoid multicollinearity problems, we selected among the highly correlated (i.e., value of the coefficient larger than 0.7) independent variables (in our case, allele frequencies) the one that was most widely distributed among the population samples (i.e., different from zero in a larger number of population samples). When the number of independent variables in the initial model was larger than 4, we used backwards stepwise regression (Venables & Ripley, 2002) and we checked again whether relevant variables not retained in the final model (e.g., latitude, longitude or pathogen richness) were really not significant. Finally, all retained models were validated using diagnostic plots and formal Shapiro and Levene tests for normality and homoscedasticity of the residuals.

All statistical analyses were performed with R (R Core Team 2013) (also described by Dalgaard, 2008) environment packages RESHAPE2 (Wickham, 2007), GGPLOT2 (Wickham, 2009), MASS (Venables & Ripley, 2002), VEGAN (Oksanen et al., 2016), NCF (Bjornstad, 2015) and GEOSPHERE (Hijmans, 2015), except the population substructure (ANOVA) analyses for which we used the ARLEQUIN software (Excoffier & Lischer, 2010).

3 | RESULTS

3.1 | HLA-A and -B genetic profiles of populations across the Sahel

HLA-A and -B allele frequencies and the results of HWE and BEWS tests estimated for the 11 populations newly typed in this study (*Sahel set*) are reported in Table 2. We do not observe any deviation from HWE (all $p > .05$ at both loci). A significant departure from selective neutrality towards an excess of heterozygotes is found in a few cases at both locus A (in GOU and DAN) and B (in BAG, DAZ, GRS, NUB and DAN), in line with what is observed at the continent level (Table S2). Interestingly, the HLA-A and -B genetic profiles of the Sahelian populations differ markedly depending either on their lifestyle, that is, as (semi-)nomadic pastoralists or sedentary farmers, or on their geographic region, that is, Western or Eastern/Central/Northern. Regarding the (semi-)nomadic pastoralist populations, the RAS and BEJ from Sudan exhibit particularly high frequencies of several HLA alleles (A*02: above 45%; B*50 and B*51: 15-26%; and B*07: 17% in RAS), and of two HLA haplotypes in significant positive linkage disequilibrium (A*02~B*07: 17% in RAS; A*02~B*50: 21% in BEJ) (Table S3). These two populations therefore show a lower heterozygosity (*He*) at both loci (A: below 80%; B: below 90%, compared to other African populations, Table S2). Another HLA-B allele, B*39, also reaches a relatively high frequency (13%–17%) in the (semi-)nomadic DAZ and BAG from Chad. By contrast, the sedentary farmers exhibit a higher frequency of B*15 (10%–17%) than the pastoralist populations. Regarding regional variation, higher frequencies of several alleles (A*23, B*42 and B*53, often above 15%–20%) are observed in Western (MOS, GRS and GOU from Burkina Faso) Africans than in the other populations.

3.2 | Overall HLA-A and -B genetic landscape of Africa

The HLA frequencies of the total set of 40 populations, that is, the 11 typed in this study and the 29 taken from published studies (reported in Table S2) were used to compute genetic distances and plot the two-dimensional (2D) NMDS shown in Figure 2a,b for loci HLA-A and HLA-B, respectively (additional three-dimensional (3D) NMDS are shown in Fig. S1). The NMDS of the two loci resemble globally to each other regarding the outlier positions of some (semi-)nomadic populations (RAS, BEJ and FUL, the latter being located in Burkina Faso) as well as the differentiation of three geographic regions, namely SSAFR, WSAFR and NAFR (plotted at the bottom left, bottom right and top right of the 2D graphs, respectively). In both NMDS, ESAFR and CSAFR also overlap extensively with each other and with NAFR, whereas SSAFR is only discriminated from the other regions at the HLA-A locus. These observations agree with the results of the ANOVA analyses (Table S4), which indicate that ESAFR and CSAFR are undifferentiated to each other according to both loci ($F_{CT} \cong 0$ with $p = .428$ at locus A and $p = .294$ at locus B) and that the geographic regions are more often discriminated ($F_{CT} > F_{SC}$) according to HLA-A, which slightly increases the overall geographic structure at this locus.

3.3 | Geographic variation of HLA-A and -B alleles

We performed additional analyses to assess whether particular HLA alleles were more significantly associated than others to this overall genetic pattern. Spatial autocorrelation was found to be significant for many alleles at both loci (Table S5). However, the patterns revealed by the individual autocorrelograms are heterogeneous (Fig. S2), as only a few alleles show typical *clines* (i.e., continuous increase or decrease of allele frequencies with geographic distance) or *depressions* (i.e., clinal variation encompassing only a part of the studied area) (as defined by Barbujani, 2000), for example, A*33 and B*35 (which reveal a *depression* when all African populations are included, and a *cline* when NAFR populations are removed from the

TABLE 2 HLA-A and HLA-B allele frequencies and summary statistics in the 11 populations typed in this study (Sahel set)

	Sudan				Chad				Burkina Faso		
	SUD	***RAS***	***BEJ***	**NUB**	***DAZ***	***BAG***	**MAB**	**DAN**	**MOS**	**GRS**	**GOU**
HLA-A											
N	46	52	48	54	41	50	41	49	35	32	36
A*01	0.065	0.048	0.073	**0.111**	**0.110**	0.080	0.098	0.020	0.046	0.047	0.014
A*02	**0.228**	**0.462**	**0.479**	**0.343**	**0.293**	**0.320**	**0.354**	**0.194**	**0.227**	**0.188**	**0.157**
A*03	0.098	**0.125**	0.042	0.056	0.049	0.070	0	0.071	0.030	0.031	0.086
A*11	0	0.010	0.010	0.009	0	0.020	0	0	0	0	0
A*23	0.098	0.029	0.021	0.009	0.037	0.020	0.024	0.082	**0.152**	**0.234**	**0.200**
A*24	**0.130**	0.029	0.073	0.056	0.073	0.010	0.037	0.041	0	0	0
A*25	0	0	0	0	0	0	0	0	0	0	0
A*26	0.011	0	0	0	0	0.030	0	0.031	0	0	0
A*29	0.044	0	0.010	0.009	**0.122**	0.010	0.049	0	0.015	0	0.014
A*30	0.098	0.058	**0.125**	**0.148**	**0.159**	**0.110**	**0.232**	**0.174**	**0.167**	**0.141**	**0.129**
A*31	0.022	0.077	0	0.009	0	0.070	0	0.079	0	0	0
A*32	0.044	0.048	0.063	0.065	0	0.040	0.037	0.061	0	0	0
A*33	0.022	0	0.042	0.056	0.037	0.050	**0.110**	0.034	0.061	**0.172**	**0.143**
A*34	0.011	0	0	0.028	0.073	0.020	0.012	0.051	0.030	0.031	0
A*36	0	0	0	0	0	0	0	0.031	**0.106**	0.016	0.071
A*43	0	0	0	0	0	0	0	0	0	0	0
A*66	0.033	0.019	0.010	0.009	0	0	0	0	0.030	0.031	0.057
A*68	0.087	0.019	0.021	0.083	0.012	**0.120**	0.049	**0.122**	0.091	0.047	0.071
A*69	0	0.019	0	0	0	0	0	0	0	0	0
A*74	0.011	0.058	0.031	0.009	0.024	0.030	0	0.010	0.030	0.063	0.057
A*80	0	0	0	0	0.012	0	0	0	0.015	0	0
A*blank	0	0	0	0	0	0	0	0	0	0	0
k	15	13	13	15	12	15	10	14	14	11	12
He	0.884	0.751	0.735	0.827	0.846	0.848	0.791	0.888	0.868	0.849	0.874
HWE	1.000	1.000	1.000	1.000	0.876	0.072	1.000	1.000	1.000	1.000	1.000
BEWS-min	0.042	0.565	0.705	0.394	0.094	0.179	0.336	**0.005**	0.106	0.152	**0.011**
BEWS-max	0.569	0.905	0.943	0.818	0.722	0.804	0.736	0.397	0.775	0.746	0.433
HLA-B											
N	46	52	48	54	41	50	41	49	35	32	36
B*07	0.076	**0.173**	0.042	0.037	0.026	0.020	0.039	0.031	0.014	0.078	**0.104**
B*08	0	0	0.021	0	0.039	0.060	0.064	0.020	0.014	0.016	0.028
B*13	0.011	0.019	**0.125**	0.046	0.000	0.020	0.090	0.020	0.014	0.016	0
B*14	0.011	0	0.010	0.065	0.026	0.060	0.013	0	0	0.031	0.042
B*15	**0.141**	0.067	0.042	**0.130**	0.051	0.080	**0.103**	**0.153**	**0.171**	**0.109**	**0.167**
B*18	0.022	0.029	0.052	0.028	0.013	0.020	0	0.031	0.014	0.031	0.014
B*27	0.044	0	0.010	0.046	0.026	0	0	0	0.014	0	0
B*35	0.044	0.029	0.031	0.056	0.026	0.070	0.039	**0.184**	0.086	0.078	0.083
B*37	0.022	0	0	0.028	0	0	0	0	0	0	0
B*38	0.022	0	0.031	0.037	0	0.030	0.039	0	0	0	0
B*39	0.076	0	0.094	0.028	**0.167**	**0.130**	0.051	0.071	0.014	0.031	0
B*40	0.054	0	0.010	0	0	0.010	0	0	0	0	0
B*41	**0.152**	0.029	0.073	0.083	0.064	0.090	0.026	0.010	0.029	0.016	0.014
B*42	0.022	0	0	0.028	0.013	0.090	0.077	**0.102**	**0.129**	**0.125**	**0.119**

(Continues)

TABLE 2 (Continued)

	Sudan				Chad				Burkina Faso		
	SUD	*RAS*	*BEJ*	NUB	*DAZ*	*BAG*	MAB	DAN	MOS	GRS	GOU
B*44	0.033	0.019	0	0.037	0.013	0.030	0.013	0.031	0	0	0
B*45	0.033	0	0	0.019	**0.103**	0.020	0.013	0.041	0.057	0.063	0.028
B*47	0	0	0.010	0.028	0.013	0	0.013	0.041	0	0	0
B*48	0	0	0	0	0	0	0	0	0	0	0
B*49	0.022	0.039	0.021	0.019	0.026	0.023	0.026	0.020	0	0	0.056
B*50	0.033	**0.260**	**0.219**	0.037	**0.115**	0.057	0	0	0.014	0	0
B*51	0.065	**0.183**	**0.146**	0.083	**0.115**	0.080	0.077	0.010	0.029	0.072	0.064
B*52	0.054	0.010	0	0.009	0	0.020	0	0	0.029	0.084	0.048
B*53	0.022	0.010	0.021	0.056	0.051	0.020	0.039	0.041	**0.214**	**0.156**	**0.208**
B*54	0	0	0	0	0	0	0	0	0	0	0
B*55	0.011	0	0	0.009	0	0	0	0	0	0	0
B*56	0	0	0	0	0.013	0	0.013	0	0	0	0
B*57	0.011	0.010	0.021	0.074	0.039	0.020	0.051	0.041	0	0.016	0
B*58	0.011	**0.125**	0.010	0.009	0.051	0.040	**0.218**	**0.102**	**0.100**	0.063	0
B*67	0	0	0	0	0	0	0	0	0	0	0
B*73	0.011	0	0.010	0	0	0	0	0	0	0	0
B*78	0	0	0	0	0	0	0	0	0.057	0.016	0.028
B*81	0	0	0	0	0	0	0	0	0	0	0
B*82	0	0	0	0	0.013	0.010	0	0.051	0	0	0
B*blank	0	0	0	0	0	0	0	0	0	0	0
k	24	16	20	24	22	23	20	19	17	17	14
He	0.9246	0.8441	0.8906	0.9388	0.9158	0.9318	0.9047	0.9034	0.8804	0.9116	0.8833
HWE	1.000	1.000	0.550	1.000	1.000	1.000	1.000	0.890	1.000	1.000	1.000
BEWS-min	0.033	0.442	0.3	**0.001**	**0.017**	**0.003**	0.03	**0.013**	0.379	**0.009**	0.043
BEWS-max	0.711	0.874	0.835	0.287	0.776	0.374	0.639	0.623	0.909	0.587	0.607

N, number of individuals tested; *k*, number of lineages (1st-field level alleles) detected. He, Estimated heterozygosity; HWE, Hardy–Weinberg equilibrium *p*-value; BEWS-min,-max, bootstrapped Ewens–Watterson–Slatkin neutrality test, minimum and maximum *p*-values. Lineage frequencies above 10% and significant values are in bold. Population names: SUD, Sudanese Arabs; RAS, Rashaayda Bedouins; BEJ, Beja Hadendowa; NUB, Nubians; DAZ, Daza; BAG, Baggara Arabs; MAB, Maba; DAN, Dangaléat; MOS, Mossi; GRS, Gurunsi Kassena; GOU, Gourmantche. (Semi-)nomadic populations are italicized.

analysis to check whether the depression effect was due to the Sahara Desert). As previous HLA population genetic studies performed at continental levels revealed strong correlations of HLA frequencies with other variables than genetic distance, for example, latitude (Di & Sanchez-Mazas, 2011, 2014; Di et al., 2015), we also checked for patterns potentially associated with latitude and longitude. Spearman's ρ and Pearson's *r* tests revealed high (above 0.5) and significant correlations (for at least one of the two coefficients ρ and *r*, see Table S6) between 16 alleles and latitude (positive correlation for A*01, A*24, B*38, B*41, B*50, B*51 and B*55, and negative for A*30, A*36, A*43, A*66, A*74, B*15, B*42, B*58 and B*81) and between five alleles and longitude (negative correlation for A*23, A*33, B*35 and B*78, and positive for B*13); we thus used these alleles as putative predictors of latitude or longitude in two corresponding linear models, respectively. We found that latitude was more strongly associated with allele frequency variation (adjusted $R^2 = 0.806$) than longitude (adjusted $R^2 = 0.522$), and we identified the five alleles A*30, A*43, A*74, B*15 and B*58 and the three alleles A*33, B*13 and B*35 as the main contributors of these two geographic variables, respectively (Table S7 and Figs. S3 and S4).

3.4 | Associations of HLA diversity to environmental parameters

Because HLA genes are known to be targets of natural selection through heterozygous advantage and resistance or susceptibility to many diseases (see Section 1), we further explored possible effects of environmental parameters on the observed HLA genetic patterns. We first checked whether the heterozygosity estimated in the studied populations was correlated to the pathogen richness reported at their corresponding geographic locations, as was found by some studies at the global level (Prugnolle et al., 2005; Qutob et al., 2012; Sanchez-Mazas et al., 2012), but the results were not significant (Spearman's $\rho = -0.029$, $p = .860$ and Pearson's $r = -.149$, $p = .361$, at locus A; Spearman's $\rho = -0.180$, $p = .266$ and Pearson's

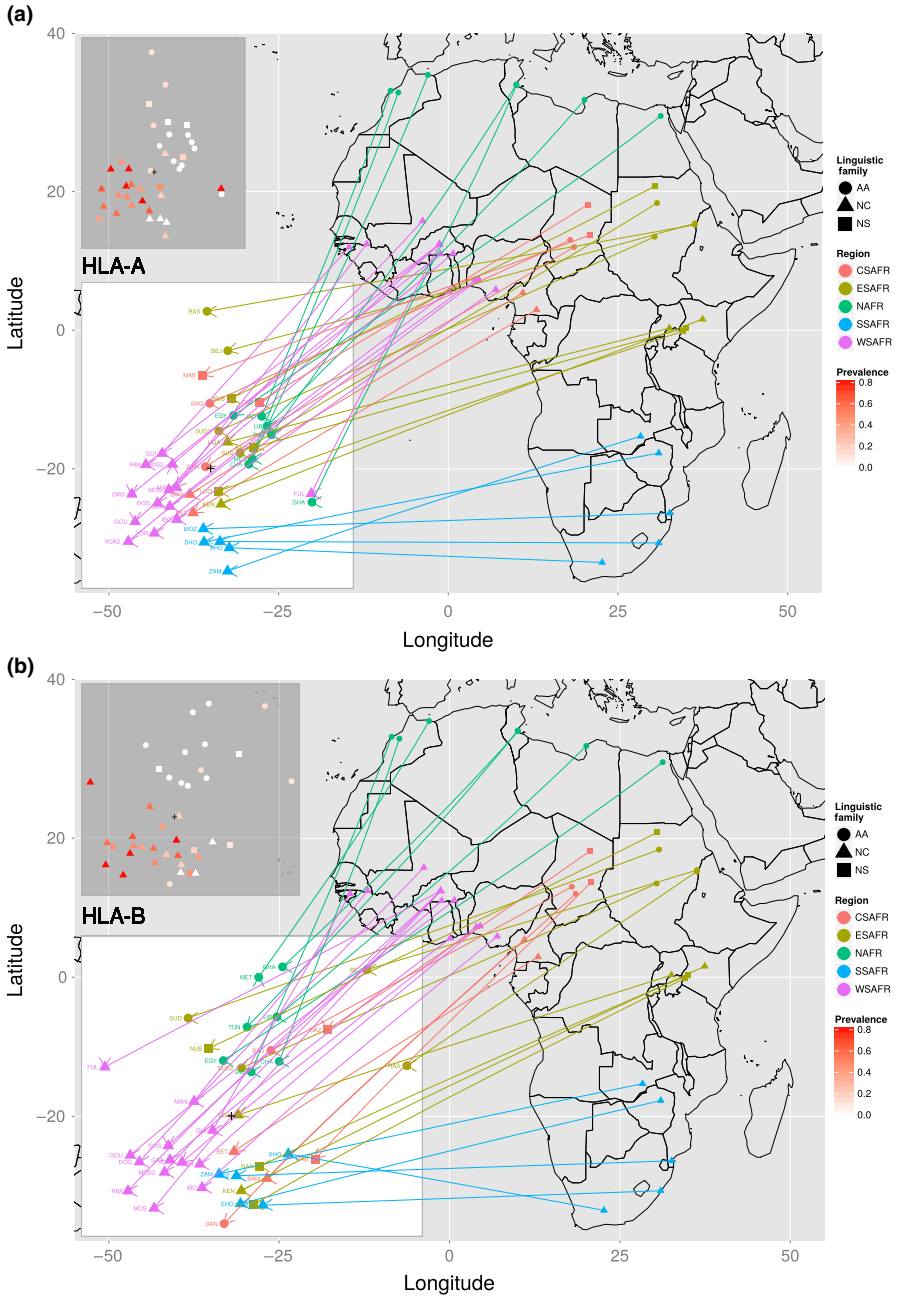

FIGURE 2 Two-dimensional nonmetric multidimensional scaling (2D-NMDS) analysis based on Prevosti's distances computed from HLA-A (a) and HLA-B (b) allele frequencies, respectively. Both NMDS are validated by reasonably good stress values (0.160 at locus HLA-A, 0.159 at locus HLA-B) (Kruskal, 1964). On the larger NMDS plot of each 2D figure (bottom left), populations are linked to their geographic locations on the map of Africa shown on the right, and different colours and symbols indicate their geographic sub-regions and linguistic families, respectively. The smaller NMDS plot of each figure (top left) is identical to the larger one except that a continuous colour code indicates the *PfPR* values assigned to the populations. Population names (see Table S1 for details): CHA, Moroccans Chaouya; JAD, Moroccans El Jadida; MET, Moroccans Metalsa; TUN, Tunisians South; GHA, Tunisians Ghannouch; LIB, Libyans Benghazi; EGY, Egyptians; SUD2, Sudanese; SUD, Sudanese Arabs; RAS, Rashaayda Bedouins; BEJ, Beja Hadendowa; NUB, Nubians; LUO, Luo; NAN, Nandi; KEN, Kenyans; UGA, Ugandans; DAZ, Daza; BAG, Baggara Arabs; MAB, Maba; DAN, Dangaléat; BAM, Bamileke; BET, Beti; IBO, Ibo; YOR, Yoruba; YOR2, Yoruba; GAA, Ga-Adangbe; MOS, Mossi; MOS2, Mossi; RIM, Rimaibé; FUL, Fulani; GRS, Gurunsi Kassena; GOU, Gourmantche; DOG, Dogons; GUI, Guiné-Bissauans; MAN, Mandenka; MOZ, Mozambicans; ZAM, Zambians; SHO, Shona; ZUL, Zulu; XHO, Xhosa

$r = -.180$, $p = .268$, at locus B). Actually, pathogen richness (an average of 228 ± 11 infectious diseases to which the populations represented in our study are potentially exposed) appears to be higher in East Africa (average 232 ± 2), lower in North Africa (average 213 ± 6) and more heterogeneous in West Africa (average 230 ± 11) (NID values, Table S1), whereas heterozygosity values are more heterogeneous across the different geographic regions (He values, Table S2).

TABLE 3 Linear model for PfPR prevalence

	Dependent variable PfPr	
	Initial model coefficient (st. residual)	Retained model coefficient (st. residual)
A*01	0.484 (1.054)	
A*11	−1.418 (2.831)	
A*23	0.725 (1.064)	
A*24	−0.067 (1.727)	
A*33	0.600 (1.131)	
A*36	−0.186 (1.272)	
A*74	2.159 (1.436)	**2.571** ** **(1.036)**
kB	−0.001 (0.014)	
B*08	−0.915 (1.522)	
B*38	0.340 (3.903)	
B*41	−1.577 (1.560)	
B*50	−0.068 (0.730)	
B*53	1.262 (1.015)	**2.310** *** **(0.502)**
B*55	−0.293 (6.678)	
B*78	3.623* (1.96)	**5.366** *** **(1.405)**
Constant	0.071 (0.347)	−0.039 (0.045)
Observations	40	40
R^2	0.773	0.727
Adjusted R^2	0.631	0.705
Residual Std. error	0.179 (*df* = 24)	0.161 (*df* = 36)
F statistic	**5.450*** (*df* = 15; 24)**	**32.016*** (*df* = 3; 36)**

Allele names denote the estimated allele frequencies, kB is the number of alleles observed at locus HLA-B, the coefficients denote how much a 1% increase in frequency changes PfPR (e.g., in the retained model, an increase of 1% of the B*78 frequency is associated with an increase of 5.366% of PfPR, with a standard error of 1.405, the ratio of the estimate to its standard deviation being a *t* test statistic, highly significant in this case). R^2 is the coefficient of determination and the adjusted version compensates for the number of parameters of the model. The initial and retained models are not significantly different. Significant coefficients and *F* statistics of the retained model are in bold and marked with stars. *$p < .1$; **$p < .05$; ***$p < .01$; *df* = degrees of freedom.

Thanks to the detailed map of malaria prevalence (Figure 1) that we used to assign *PfPR* values to the 40 populations studied, we then analysed jointly possible effects of several genetic (heterozygosity, number of alleles and allele frequencies), geographic (latitude, longitude) and environmental (pathogen richness) variables on *P. falciparum* malaria prevalence through linear modelling. The initial and retained models are described in Table 3, where the coefficients represent the estimated increases of *PfPR* with increased allele frequencies. The retained model, which provides a good fit by explaining more than 70% of the *PfPR* variance ($R^2 = 0.727$ and adjusted $R^2 = 0.705$) and having low and well-behaved residuals, reveals that the frequencies of two HLA-B alleles, namely B*53 and B*78, and one HLA-A allele, A*74, are significantly correlated with *PfPR* ($p < .01$ for the two former, and $p < .05$ for the latter), whereas neither of the other parameters have significant effects. Furthermore,

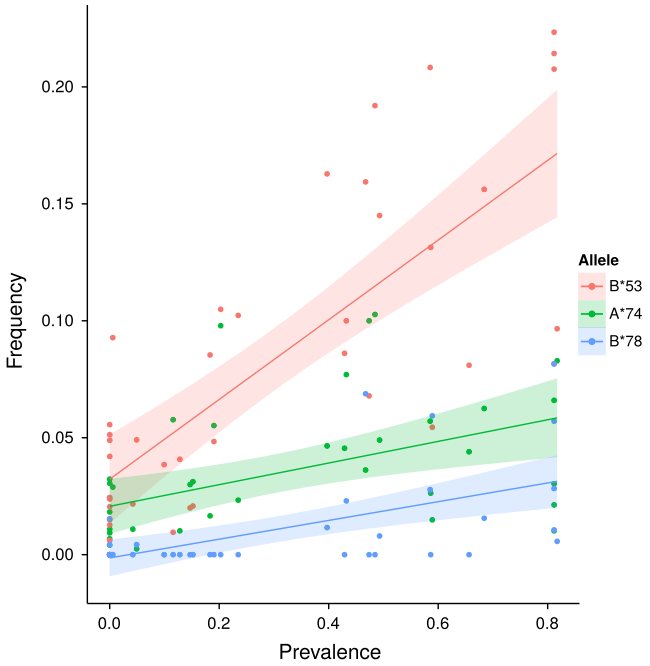

FIGURE 3 Plot of allele frequency as a function of *Plasmodium falciparum* malaria prevalence (estimated by *P. falciparum* parasite rates *PfPR*). The slope of the straight lines corresponds to the regression coefficients of the three alleles B*53, A*74 and B*78 (see Table S8)

the reverse relationship, that is, allele frequency as a function of *PfPR* and allele (Figure 3), confirms the strongest association (steepest slope and largest coefficient, significant interaction) for B*53, which is by far the most frequent in malaria-endemic countries, but very similar, although flatter patterns were observed for B*78 and A*74 (Table S8).

3.5 | Candidate HLA alleles to malaria protection

We finally identified, by additional high-throughput sequencing, the precise molecular variants of the three alleles identified above in all individuals of the new samples tested in this study (*Sahel set*). We found that in all B*53-positive individuals ($N = 55$) but one, B*53 corresponded to the single allele B*53:01:01 (homozygous in three individuals, i.e., one GRS, one MOS and one NUB). The only exception is in one RAS individual, where B*53 consists of a recently discovered allele, B*53:19 (Hernandez-Frederick et al. 2014). Interestingly, by sequencing the HLA-C locus closely linked to HLA-B (both loci being at about 95 kb and 0.2 cM from each other, Litwin, 1989) in all B*53-positive individuals, we found that 34 of the 55 sequenced individuals (i.e., 62%) also carried the C*04:01:01:01 allele (Table S9), although the amount of linkage disequilibrium across the HLA region appears to be low in Africa (Testi et al., 2015) compared to Europe (Bugawan, Klitz, Blair, & Erlich, 2000; Sanchez-Mazas et al., 2000), as generally observed for other genetic markers (Campbell & Tishkoff, 2008). Therefore HLA-B*53:01:01~C*04:01:01:01 (observed at the homozygous state in one GOU individual from Burkina Faso) appears as a putative haplotype associated with *P. falciparum* resistance. Likewise,

all B*78-positive individuals ($N = 14$), all belonging to populations located in malaria-endemic areas, were found to carry the unique allele B*78:01. On the other hand, among all A*74-positive individuals ($N = 24$), about two-thirds ($N = 15$) carry A*74:01 (actually, one ambiguous A*74:01/74:02 result could not be solved due to damaged DNA, but the ambiguity lies outside the peptide-binding region of the HLA molecule) and one-third ($N = 9$) carries A*74:03. Allele A*74:01 is the unique allele found in populations living in malaria-endemic areas (GRS, GOU and MOS), whereas both A*74:01 and A*74:03 are observed in populations residing in nonendemic regions (BEJ and DAZ, for the former; RAS, BEJ, NUB and SUD, for the latter).

4 | DISCUSSION

Thanks to the typing of 11 new population samples located in different regions across the Sahel, the present study fills major gaps in our knowledge on HLA diversity in Africa. Indeed, compared to the two main studies published by Cao et al. (2004) and Spinola et al. (2011), which previously compared the HLA class I allele and haplotype frequencies among several African populations, we included here several populations from Burkina Faso, Sudan and Chad, the latter being a key Central region which was never investigated before for HLA. Our new data also comprise several populations following nomadic or semi-nomadic lifestyles and/or speaking languages of the Nilo-Saharan family which were up to now underrepresented for this polymorphism. Moreover, by analysing these original data combined to those of published studies and to comprehensive records of *P. falciparum* malaria prevalence and infectious diseases in Africa, we provide the first extensive analysis of HLA-A and -B variation in relation to both geography and environment factors in this continent. Owing to several complementary analyses and, more particularly, linear modelling allowing us to explore jointly the contribution of these diverse parameters, our goal was to understand better the different mechanisms that shaped the HLA genetic landscape of Africa. We conclude that in addition to populations' demography and migrations history, which certainly shaped the variation of several HLA allele frequencies across this continent, pathogen-driven selection involving three HLA alleles, and more particularly B*53:01:01 and B*78:01, probably also modelled the observed HLA genetic patterns, as discussed below.

4.1 | HLA and African populations' demography and migrations history

Among the 40 African populations considered in this study, the RAS and the BEJ (both of them newly typed in this study) have shown to exhibit extreme HLA genetic profiles compared to most other populations. This is easily explained by the isolated way of life of these (semi-)nomadic populations where rapid genetic drift likely increased some allele and haplotype frequencies and reduced their genetic diversity. Indeed, the RAS (Arabic speakers, camel herders), who live in the desert along the coast of the Red Sea between Massawa in Eritrea and Port Sudan (Young, 1996), migrated from Saudi Arabia at the end of the 19th century and remained isolated from the Cushitic populations (in particular the Beja) dwelling in the same territory. The BEJ (whose name is related to the Arabic word "Bedawi" meaning Bedouin, and who speak their own language) are composed of several endogamous groups, among which the Hadendowa (camel herders) typed in this study (Gudrun, 2006; Manger, 1996). Interestingly, the FUL (included in the 29 population samples taken from published studies), who also appear as outliers in our analyses and whose genetic divergence has been observed for other genetic markers (Černý et al., 2011), are one of the largest nomadic pastoralist (camel herder) groups of Africa, a way of life that would have maintained them genetically isolated from neighbouring populations. Noteworthy is the genetic similarity, at locus HLA-A (visible in the corresponding NMDS), between the FUL and the GHA (Tunisians from Gannouch), previously reported as having a "Berber substratum" (Hajjej, Almawi, Hattab, El-Gaaied, & Hmida, 2015; Hajjej et al., 2006), which agrees with mtDNA and genomewide analyses supporting a Middle Holocene input to ancestral Fulani population from the Mediterranean area (Kulichová et al., 2017; Triska et al., 2015). Interestingly, it has also been stated that the FUL are less susceptible to malaria than neighbouring populations of Burkina Faso (Modiano et al., 1996), showing a higher immune reactivity which would not be due to the "classical" malaria-resistance genes (Modiano et al., 2001) but to a functional deficit of T regulatory cells (Torcia et al., 2008) and higher levels of cytokines (Bostrom et al., 2012). This would explain why allele HLA-B*53, which probably underwent substantial positive selection in other populations living in malaria-endemic areas (see below), does not show an increased frequency in FUL.

Another noteworthy result of our analyses is the great genetic heterogeneity, found at both HLA loci, of ESAFR and CSAFR populations, which overlap extensively (also with the NAFR) in the NMDS. This suggests that the wide geographic area encompassing Eastern, Central and North Africa has been the scene, in the past, of complex population movements and interactions involving both transhumant and sedentary populations characterized by distinct genetic backgrounds and undergoing heterogeneous levels of gene flow, as previously stressed for East African populations based on other genetic studies (Poloni et al., 2009; Tishkoff et al., 2009). In contrast, the populations located in the WSFAR (except the FUL) and SSAFR regions, here all represented by Niger-Congo (NC) speakers, form two distinct population groups. This is fully consistent with past genetic differentiations of Western and Southern NC populations resulting from two extensive migrations westwards (between 8,000 and 5,000 years ago) and southwards (between 5,000 and 3,000 years ago, Bantu expansion) from an original location situated around Nigeria and Cameroon, as advocated by archaeological and linguistic studies (Diamond & Bellwood, 2003; de Filippo, Bostoen, Stoneking, & Pakendorf, 2012; Li, Schlebusch, & Jakobsson, 2014).

Overall, part of the HLA-A and -B landscape of Africa is thus well explained by populations' demography and migrations history, suggesting that natural selection had moderate (although non-negligible, according to the present study) effects on the evolution of these

polymorphisms, as previously indicated by computer simulations (Currat et al., 2010; Di et al., 2015).

4.2 | Signatures of natural selection on African HLA genetic patterns

Two kinds of natural selection were addressed in this study, that is, balancing selection due to heterozygous advantage (in the general sense), which is expected to maintain high levels of populations' heterozygosity, and directional selection due to higher fitness of genotypes more protective to particular diseases, which is expected to increase the frequencies of specific alleles (see, e.g., Hedrick, 2010; for theoretical backgrounds). The higher level of average heterozygosity (He) and greater number of significant rejections of selective neutrality towards an excess of heterozygotes observed at HLA-B (He = 0.9125 ± 0.0216 and 24 rejections, respectively), which is the most polymorphic HLA gene (Mungall et al., 2003), compared to HLA-A (He = 0.8672 ± 0.0386 and 15 rejections, respectively) on the same set of 40 African populations (Table S2) is consistent with the hypothesis that locus B and is more prone to balancing selection than locus A (Buhler & Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2013; Solberg et al., 2008). However, we did not find any correlation between heterozygosity and pathogen richness at any of the two loci; therefore, our results do not support a model where heterozygous individuals would, in general, be favoured (Prugnolle et al., 2005; Qutob et al., 2012; Sanchez-Mazas et al., 2012). One possible reason is that our analysis lacked power to reveal such a correlation, either because the number of infectious diseases (NID) used in our study as a proxy of pathogen richness is a poor estimator of pathogen diversity, or because we used HLA data defined at the 1st-field level of resolution, which corresponds to specificities defining sets of molecules with possibly distinct peptide-binding properties, likely lowering the sensibility to detect relevant associations. Another likely explanation with regards to our results, however, is that either pathogen-driven positive selection or genetic drift, by increasing the frequencies of specific alleles in peculiar environments (e.g., where malaria is endemic) or populations (e.g., isolated (semi-)nomadic pastoralists), respectively, blurred the signatures of heterozygous advantage (in the general sense) at the continent level. Actually, at local geographic levels heterozygotes for both a malaria-protective allele and an allele that would protect to another prevalent disease in the same environment would probably better survive, and such regional heterozygote advantages would not necessarily be related to the overall pathogen richness prevailing in the corresponding location.

Malaria is due to infection by one of the five known *Plasmodium* species, *P. falciparum* being however largely predominant and the most virulent in sub-Saharan Africa (Snow, Guerra, Noor, Myint, & Hay, 2005). Different kinds of resistance to malaria (reviewed by Hedrick, 2011; Lopez, Saravia, Gomez, Hoebeke, & Patarroyo, 2010) are assumed to be conferred by diverse genetic markers related to haemoglobin disorders (HbS, HbE, HbC, thalassaemia), erythrocyte polymorphisms (ovalocytosis, DARC (see also Demogines, Truong, & Sawyer, 2012)), enzymopathies (e.g., G6pD), and immunogenetic variants (e.g., CR1, HLA). Although attention has recently been given to a putative role of HLA-G in the regulation of the response to *P. falciparum* infection (Garcia et al., 2013; Sabbagh et al., 2013), the strongest relationships between HLA and malaria protection are found for HLA-A and -B (Garamszegi, 2014), and the main effect is explained by the recognition of a nonamer peptide derived from *P. falciparum* liver stage antigen-1 by HLA-B*53 restricted T cells (Hill et al., 1991, 1992). In addition, it has been suggested that B*53:01, carrier of the Bw4 epitope, has the capacity of interacting with NK cells, notably through the receptor KIR3DL1, which would contribute to its protective effects (Norman et al., 2013). The results of our linear modelling are consistent with the hypothesis that a protective effect to *P. falciparum* malaria contributed to shape the HLA-B genetic landscape of Africa by increasing the frequency of the B*53 allele. In support to this idea, we found that all B*53-positive individuals of our new population samples carried the B*53:01:01 allele, the only exception being one B*53:19-positive individual in the RAS population who lives in a region where malaria is not endemic. In addition, one HLA-C allele, C*04:01:01:01, was observed in 62% of the B*53:01:01-positive individuals, suggesting that a hitchhiking effect (Smith & Haigh, 1974) increased the frequency of a putative B*53:01:01~C*04:01:01:01 haplotype.

Interestingly, we also detected a significant correlation with malaria prevalence for another HLA-B allele, B*78, that we further identified as B*78:01. To our knowledge, B*78:01 has never been mentioned as protective to malaria so far. It is possible that B*78:01 and B*53:01:01 (actually B*53:01 if we refer to the corresponding functional HLA protein) share functional similarities with other potentially protective HLA molecules by binding the same or similar peptides derived from *P. falciparum*; alternatively, B*78:01 and B*53:01 might be functionally divergent from each other and confer protection to malaria through the presentation of nonoverlapping sets of peptides. To consider these alternative hypotheses, we built an UPGMA tree based on peptide-binding distances inferred by MHCCLUSTER 2.0 Server (Thomsen, Lundegaard, Buus, Lund, & Nielsen, 2013) for different HLA molecules, besides which we also represented logos of nonamer peptides showing the amino acids that would be most strongly bound to these molecules (Fig. S5). The HLA-A and -B molecules that we included in this figure correspond to the most frequent alleles found in Africa (i.e., having a frequency greater than 15% in at least one population under study), plus a number of other ones supposed to confer protection or susceptibility to the disease. Interestingly, the peptide-binding profile of B*53:01 shows a very strong affinity for peptides with amino acid P at position 2 (accommodated by pocket B of the peptide-binding groove), a characteristic that is shared by a number of other HLA-B alleles among which B*53:19, found in one RAS, B*39:10, previously suggested to share peptide-binding similarities with B*53:01 (Yague, Vazquez, & Lopez de Castro, 1998), and B*78:01, found to be associated with *P. falciparum* malaria prevalence in the present study. The fact that B*53:01 and B*78:01 exhibit similar binding profiles for peptide position 2 thus indicates that the peptide repertories of

the two molecules may overlap to some extent, including peptides derived from *P. falciparum*, although this remains to be formally investigated. Among the other alleles functionally related to B*53:01 based on peptide-binding properties, B*15:13, previously mentioned by Ghosh (2008), and B*35(:01), which has been suggested to be protective to malaria in Ghana (Yamazaki et al., 2011) and Sardinia (Contu et al., 1998), share a strong affinity for amino acid F at peptide position 9 (accommodated by pocket F of the peptide-binding groove), this binding being then also putatively significant against this disease. Overall, the observation that several HLA-B alleles, with either similar or distinct peptide-binding profiles, would be protective to malaria suggests that pathogen-driven selection acting on the HLA-B locus would be of the kind defined as *soft selective sweep* (Hermisson & Pennings, 2005; Messer & Petrov, 2013) by which several alleles from the standing genetic variation (here B*53:01:01 and B*78:01 proposed in this study, and possibly others) would have become beneficial at the onset of a specific infectious disease (in this case *P. falciparum* malaria). Together with different forms of balancing selection (i.e., heterozygous advantage, frequency-dependent selection and selection fluctuating in time and/or space) to which the HLA loci are probably submitted (Hedrick, 2002; Meyer & Thomson, 2001), this mechanism would have also contributed to maintain a high level of polymorphism at this locus, rather than to decrease its diversity as expected from the "classical" form of selective sweep (Smith & Haigh, 1974).

Finally, one HLA-A allele, A*74, was also found to be positively associated with *PfPR*, speaking in favour of this allele being an additional protective factor. This finding is supported by the study of Migot-Nabias et al. (2001) showing that specific immune responses operating at the pre-erythrocytic stage of development of *P. falciparum* would be regulated by HLA-A19, the latter being the broad antigen serotype to which A*74 belongs. By performing additional sequencing of A*74-positive individuals in our new samples, we also found that among the two identified alleles, A*74:01 was the only one observed in malaria-endemic areas (Western Africa) where it could have been positively selected, whereas A*74:03 was predominantly observed in the (semi-)nomadic RAS population from Sudan (six individuals on a total of nine A*74:03 carriers), where its frequency could have increased by genetic drift. On the other hand, the peptide-binding profiles predicted for these two alleles (Fig. S5) show a strong affinity for amino acid R at position 9, that is, the same affinity as that predicted for A*33:01 (and A*33:03) which has been proposed as a susceptibility, rather than a protective factor to a severe form of malaria (cerebral malaria) in Malian children (Lyke et al., 2011). Also, our linear models indicated that the frequency of A*74 was correlated not only with *P. falciparum* malaria prevalence but also with latitude (which was not the case for B*53 and B*78), unveiling a potential confounding effect possibly related to the amalgamation of the two different alleles A*74:01 and A*74:03. The hypothesis that A*74 would be protective to malaria is thus currently less supported than for the two candidate alleles proposed at the HLA-B locus.

To conclude, the present study emphasizes the multiple interests of performing a large-scale population genetic analysis of HLA genes taking jointly into account various genetic, geographic and environmental parameters. In addition to provide a better understanding of the evolutionary mechanisms that shaped the HLA polymorphism, in this case the HLA-B genetic landscape of Africa, our results disclosed two main candidate alleles to malaria protection at this locus, B*53:01:01 and B*78:01, the immunological role of which would now need to be investigated by direct functional analyses focusing on HLA peptide-binding properties. In a broader perspective, we expect further analyses focusing on the numerous highly polymorphic HLA genes to unveil other crucial associations between their molecular diversity and the wide range of environmental conditions constraining the subsistence of human populations.

ACKNOWLEDGEMENTS

This work was supported by grant #31003A_144180 of the Swiss National Science Foundation to ASM and by the Grant Agency of the Czech Republic (13-37998S-P505) to VC. We sincerely thank two anonymous reviewers for their help to improve a first version of this manuscript.

DATA ACCESSIBILITY

All the data used in this study are available within the article and in the Supporting Information.

AUTHOR CONTRIBUTIONS

A.S.M. and J.M.N. conceived and designed the research, performed the statistical analyses and wrote the manuscript. V.C. and E.P. designed and implemented new samples collection. D.D. and S.B. participated in some analyses and produced some figures. E.C. and S.W. participated in data gathering and preliminary analyses. L.B., B.K. and S.B. performed laboratory typings. V.C., D.D., S.B., M.T., M.A., J.M.T. and J.V. participated in discussing the results and in drafting of the manuscript. All authors read and approved the final manuscript.

ORCID

Alicia Sanchez-Mazas http://orcid.org/0000-0002-7714-2432

REFERENCES

Assane, A. A., Fabricio-Silva, G. M., Cardoso-Oliveira, J., Mabunda, N. E., Sousa, A. M., Jani, I. V., ... Porto, L. C. (2010). Human leukocyte antigen-A, -B, and -DRB1 allele and haplotype frequencies in the Mozambican population: A blood donor-based population study. *Human Immunology*, *71*, 1027–1032.

Aubert, M., Pike, A. W., Stringer, C., Bartsiokas, A., Kinsley, L., Eggins, S., ... Grun, R. (2012). Confirmation of a late middle Pleistocene age for the Omo Kibish 1 cranium by direct uranium-series dating. *Journal of Human Evolution*, *63*, 704–710.

Barbujani, G. (2000). Geographic patterns: How to identify them and why. *Human Biology*, *72*, 133–153.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., … Gething, P. W. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, *526*, 207–211.

Bjornstad, O. N. (2015). NCF: *Spatial nonparametric covariance functions*. R package v. 1.1-6. Retrieved from https://CRAN.R-project.org/package=ncf.

Bostrom, S., Giusti, P., Arama, C., et al. (2012). Changes in the levels of cytokines, chemokines and malaria-specific antibodies in response to *Plasmodium falciparum* infection in children living in sympatry in Mali. *The Malaria Journal*, *11*, 109.

Bugawan, T. L., Klitz, W., Blair, A., & Erlich, H. A. (2000). High-resolution HLA class I typing in the CEPH families: Analysis of linkage disequilibrium among HLA loci. *Tissue Antigens*, *56*, 392–404.

Buhler, S., Nunes, J. M., & Sanchez-Mazas, A. (2016). HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. *Immunogenetics*, *68*, 401–416.

Buhler, S., & Sanchez-Mazas, A. (2011). HLA DNA sequence variation among human populations: Molecular signatures of demographic and selective events. *PLoS ONE*, *6*, e14643.

Campbell, M. C., Hirbo, J. B., Townsend, J. P., & Tishkoff, S. A. (2014). The peopling of the African continent and the diaspora into the new world. *Current Opinion in Genetics & Development*, *29*, 120–132.

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, *9*, 403–433.

Cao, K., Moormann, A. M., Lyke, K. E., Masaberg, C., Sumba, O. P., Doumbo, O. K., … Fernandez-Vina, M. A. (2004). Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens*, *63*, 293–325.

Černý, V., Pereira, L., Musilova, E., Kujanova, M., Vasikova, A., Blasi, P., & ăNovelletto, A. (2011). Genetic structure of pastoral and farmer populations in the African Sahel. *Molecular Biology and Evolution*, *28*, 2491–2500.

Contu, L., Carcassi, C., Orru, S., Mulargia, M., Arras, M., Boero, R., … Floris, L. (1998). HLA-B35 frequency variations correlate with malaria infection in Sardinia. *Tissue Antigens*, *52*, 452–461.

Currat, M., Poloni, E. S., & Sanchez-Mazas, A. (2010). Human genetic differentiation across the Strait of Gibraltar. *BMC Evolutionary Biology*, *10*, 237.

Dalgaard, P. (2008). *Introductory statistics with* R, 2nd ed. New York, NY: Springer-Verlag.

de Filippo, C., Bostoen, K., Stoneking, M., & Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 3256–3263.

de Groot, N. G., & Bontrop, R. E. (2013). The HIV-1 pandemic: Does the selective sweep in chimpanzees mirror humankind's future? *Retrovirology*, *10*, 53.

Demogines, A., Truong, K. A., & Sawyer, S. L. (2012). Species-specific features of DARC, the primate receptor for *Plasmodium vivax* and *Plasmodium knowlesi*. *Molecular Biology and Evolution*, *29*, 445–449.

Di, D., & Sanchez-Mazas, A. (2011). Challenging views on the peopling history of East Asia: The story according to HLA markers. *American Journal of Physical Anthropology*, *145*, 81–96.

Di, D., & Sanchez-Mazas, A. (2014). HLA variation reveals genetic continuity rather than population group structure in East Asia. *Immunogenetics*, *66*, 153–160.

Di, D., Sanchez-Mazas, A., & Currat, M. (2015). Computer simulation of human leukocyte antigen genes supports two main routes of colonization by human populations in East Asia. *BMC Evolutionary Biology*, *15*, 240.

Diamond, J., & Bellwood, P. (2003). Farmers and their languages: The first expansions. *Science*, *300*, 597–603.

Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, *256*, 50–52.

Excoffier, L., & Lischer, H. E. L. (2010). ARLEQUIN SUITE ver 3.5: A new series of programs to perform population genetics analyses under LINUX and WINDOWS. *Molecular Ecology Resources*, *10*, 564–567.

Excoffier, L., Pellegrini, P., Sanchez-Mazas, A., Simon, C., & Langaney, A. (1987). Genetics and history of Sub-Saharan Africa. *Yearbook of Physical Anthropology*, *30*, 151–194.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491.

Galgani, A., Mancino, G., Martinez-Labarga, C., Cicconi, R., Mattei, M., Amicosante, M., … Montesano, C. (2013). HLA-A, -B and -DRB1 allele frequencies in Cyrenaica population (Libya) and genetic relationships with other populations. *Human Immunology*, *74*, 52–59.

Garamszegi, L. Z. (2014). Global distribution of malaria-resistant MHC-HLA alleles: The number and frequencies of alleles and malaria risk. *The Malaria Journal*, *13*, 349.

Garcia, A., Milet, J., Courtin, D., Sabbagh, A., Massaro, J. D., Castelli, E. C., … Moreau, P. (2013). Association of HLA-G 3′UTR polymorphisms with response to malaria infection: A first insight. *Infection, Genetics and Evolution*, *16*, 263–269.

Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R., Johnston, G. L., … Hay, S. I. (2011). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *The Malaria Journal*, *10*, 378.

Ghosh, K. (2008). Evolution and selection of human leukocyte antigen alleles by *Plasmodium falciparum* infection. *Human Immunology*, *69*, 856–860.

Gomez, F., Hirbo, J., & Tishkoff, S. A. (2014). Genetic variation and adaptation in Africa: Implications for human evolution and disease. *Cold Spring Harbor Perspectives in Biology*, *6*, a008524.

Gomez-Casado, E., del Moral, P., Martinez-Laso, J., Garcia-Gomez, A., Allende, L., Silvera-Redondo, C., … Arnaiz-Villena, A. (2000). HLA genes in Arabic-speaking Moroccans: Close relatedness to Berbers and Iberians. *Tissue Antigens*, *55*, 239–249.

Gourraud, P. A., Khankhanian, P., Cereb, N., Yang, S. Y., Feolo, M., Maiers, M., … Oksenberg, J. (2014). HLA diversity in the 1000 genomes dataset. *PLoS ONE*, *9*, e97282.

Gudrun, D. (2006). Precolonial Beja: A periphery at the crossroads. *Nordic Journal of African Studies*, *15*, 473–498.

Hajjej, A., Almawi, W. Y., Hattab, L., El-Gaaied, A., & Hmida, S. (2015). HLA Class I and Class II Alleles and Haplotypes confirm the Berber origin of the present day Tunisian population. *PLoS ONE*, *10*, e0136909.

Hajjej, A., Hmida, S., Kaabi, H., Dridi, A., Jridi, A., El Gaaied, A., & Boukef, K. (2006). HLA genes in Southern Tunisians (Ghannouch area) and their relationship with other Mediterraneans. *European Journal of Medical Genetics*, *49*, 43–56.

Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*, *56*, 1902–1908.

Hedrick, P. W. (2010). *Genetics of populations*, 4th ed. Sudbury, MA: Jones & Bartlett Learning.

Hedrick, P. W. (2011). Population genetics of malaria resistance in humans. *Heredity*, *107*, 283–304.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*, 2335–2352.

Hernández-Frederick, C. J., Giani, A. S., Cereb, N., Sauter, J., Silva-González, R., Pingel, J., … Yang, S. Y. (2014). Identification of 2127 new HLA class I alleles in potential stem cell donors from Germany, the United States and Poland. *Tissue Antigens*, *83*, 184–189.

Hijmans, R. J. (2015). *GEOSPHERE: Spherical trigonometry*. R package v. 1.5-1. Retrieved from https://CRAN.R-project.org/package=geosphere.

Hill, A. V., Allsopp, C. E., Kwiatkowski, D., Anstey, N. M., Twumasi, P., Rowe, P. A., … Greenwood, B. M. (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature*, *352*, 595–600.

Hill, A. V., Elvin, J., Willis, A. C., Aidoo, M., Allsopp, C. E., Gotch, F. M., … Whittle, H. C. (1992). Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature*, *360*, 434–439.

Hublin, J. J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., … Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*, *546*, 289–292.

Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, *335*, 167–170.

Kruskal, J. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Kulichová, I., Fernandes, V., Deme, A., Nováčková, J., Stenzl, V., Novelletto, A., … Černý, V. (2017). Internal diversification of non-sub-Saharan haplogroups in Sahelian populations and the spread of pastoralism beyond the Sahara. *American Journal of Physical Anthropology*, *164*, 424–434.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (2016). *Ethnologue: Languages of the world*. Dallas, TX: SIL International. Retrieved from http://www.ethnologue.com

Li, S., Schlebusch, C., & Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20141448.

Litwin, S. D. (1989). *Human immunogenetics, basic principles and clinical relevance*. New York, NY: Marcel Dekker Inc.

Lopez, C., Saravia, C., Gomez, A., Hoebeke, J., & Patarroyo, M. A. (2010). Mechanisms of genetically-based resistance to malaria. *Gene*, *467*, 1–12.

Lyke, K. E., Fernandez-Vina, M. A., Cao, K., Hollenbach, J., Coulibaly, D., Kone, A. K., … Sztein, M. B. (2011). Association of HLA alleles with *Plasmodium falciparum* severity in Malian children. *Tissue Antigens*, *77*, 562–571.

Mack, S. J., Bugawan, T. L., Moonsamy, P. V., Erlich, J. A., Trachtenberg, E. A., Paik, Y. K., … Erlich, H. A. (2000). Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens*, *55*, 383–400.

Mack, S. J., & Erlich, H. A. (2007). 13th International histocompatibility workshop anthropology/human genetic diversity joint report: Population relationships as inferred from classical HLA genes. In J. A. Hansen (Ed.), *Immunobiology of the human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference* (pp. 747–757). Seattle, WA: IHWG Press.

Mahfoudh, N., Ayadi, I., Kamoun, A., Ammar, R., Mallek, B., Maalej, L., … Makni, H. (2013). Analysis of HLA-A, -B, -C, -DR, -DQ polymorphisms in the South Tunisian population and a comparison with other populations. *Annals of Human Biology*, *40*, 41–47.

Manger, L. (1996). *Survival on meagre resources: Hadendowa pastoralism in the Red Sea Hills*. Uppsala: The Nordic Africa Institute.

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, *433*, 733–736.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, *28*, 659–669.

Meyer, D., Single, R. M., Mack, S. J., Erlich, H. A., & Thomson, G. (2006). Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. *Genetics*, *173*, 2121–2142.

Meyer, D., & Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: A review. *Annals of Human Genetics*, *65*, 1–26.

Migot-Nabias, F., Luty, A. J., Minh, T. N., Fajardy, I., Tamouza, R., Marzais, F., … Deloron, P. (2001). HLA alleles in relation to specific immunity to liver stage antigen-1 from *Plasmodium falciparum* in Gabon. *Genes and Immunity*, *2*, 4–10.

Modiano, D., Luoni, G., Sirima, B. S., Lanfrancotti, A., Petrarca, V., Cruciani, F., … Coluzzi, M. (2001). The lower susceptibility to *Plasmodium falciparum* malaria of Fulani of Burkina Faso (west Africa) is associated with low frequencies of classic malaria-resistance genes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *95*, 149–152.

Modiano, D., Petrarca, V., Sirima, B. S., Nebie, I., Diallo, D., Esposito, F., & Coluzzi, M. (1996). Different response to *Plasmodium falciparum* malaria in west African sympatric ethnic groups. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 13206–13211.

Mungall, A. J., Palmer, S. A., Sims, S. K., Edwards, C. A., Ashurst, J. L., Wilming, L., … Beck, S. (2003). The DNA sequence and analysis of human chromosome 6. *Nature*, *425*, 805–811.

Naranbhai, V., & Carrington, M. (2017). Host genetic variation and HIV disease: From mapping to mechanism. *Immunogenetics*, *69*, 489–498.

Norman, P. J., Hollenbach, J. A., Nemat-Gorgani, N., Guethlein, L. A., Hilton, H. G., Pando, M. J., … Parham, P. (2013). Co-evolution of human leukocyte antigen (HLA) class I ligands with killer-cell immunoglobulin-like receptors (KIR) in a genetically diverse population of sub-Saharan Africans. *PLoS Genetics*, *9*, e1003938.

Nunes, J. M. (2015). Using uniformat and gene[rate] to analyze data with ambiguities in population genetics. *Evolutionary Bioinformatics Online*, *11*, 19–26.

Nunes, J. M., Buhler, S., Roessli, D., Sanchez-Mazas, A., & Collaboration HLA-net (2014). The HLA-net GENE[RATE] pipeline for effective HLA data analysis and its application to 145 population samples from Europe and neighbouring areas. *Tissue Antigens*, *83*, 307–323.

Nunes, J. M., Riccio, M. E., Buhler, S., Di, D., Currat, M., Ries, F., … Sanchez-Mazas, A. (2010). 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*, *76*, 18–30.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2016). *VEGAN: Community ecology package*. R package v. 2.3-4. https://CRAN.R-project.org/package=vegan.

Parham, P., & Janeway, C. (2015). *The immune system*, 4th ed. New York, NY: Garland Science, Taylor & Francis Group.

Piancatelli, D., Canossi, A., Aureli, A., Oumhani, K., Del Beato, T., Di Rocco, M., … Adorno, D. (2004). Human leukocyte antigen-A, -B, and -Cw polymorphism in a Berber population from North Morocco using sequence-based typing. *Tissue Antigens*, *63*, 158–172.

Podgorná, E., Diallo, I., Vangenot, C., Sanchez-Mazas, A., Sabbagh, A., Černý, V., & Poloni, E. S. (2015). Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. *BMC Evolutionary Biology*, *15*, 263.

Poloni, E. S., Naciri, Y., Bucho, R., Niba, R., Kervaire, B., Excoffier, L., … Sanchez-Mazas, A. (2009). Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Annals of Human Genetics*, *73*, 582–600.

Prevosti, A., Ocaña, J., & Alonso, G. (1975). Distances between populations of *Drosophila* subobscura based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, *45*, 231–241.

Prugnolle, F., Manica, A., Charpentier, M., Guegan, J. F., Guernier, V., & Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology*, *15*, 1022–1027.

Qutob, N., Balloux, F., Raj, T., Liu, H., Marion de Proce, S., Trowsdale, J.., & Manica, A. (2012). Signatures of historical demography and pathogen richness on MHC class I genes. *Immunogenetics*, *64*, 165–175.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Riccio, M. E., Buhler, S., Nunes, J. M., Vangenot, C., Cuenod, M., Currat, M., … Sanchez-Mazas, A. (2013). 16(th) IHIW: Analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report). *International Journal of Immunogenetics*, *40*, 21–30.

Richter, D., Grun, R., Joannes-Boyau, R., Steele, T. E., Amani, F., Rue, M., … McPherron, S. P. (2017). The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature*, *546*, 293–296.

Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., & Marsh, S. G. (2015). The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Research*, *43*, D423–D431.

Sabbagh, A., Courtin, D., Milet, J., Massaro, J. D., Castelli, E. C., Migot-Nabias, F., … Donadi, E. A. (2013). Association of HLA-G 3′ untranslated region polymorphisms with antibody response against *Plasmodium falciparum* antigens: preliminary results. *Tissue Antigens*, *82*, 53–58.

Sanchez-Mazas, A. (2001). African diversity from the HLA point of view: Influence of genetic drift, geography, linguistics, and natural selection. *Human Immunology*, *62*, 937–948.

Sanchez-Mazas, A. (2007). 13th International histocompatibility workshop anthropology/human genetic diversity joint report. Chapter 7: HLA genetic diversity of the 13th IHWC population data relative to worldwide linguistic families. In J. A. Hansen, & B. Dupont (Eds.), *Immunobiology of the human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference* (pp. 758–766). Seattle, WA: IHWG Press.

Sanchez-Mazas, A., Buhler, S., & Nunes, J. M. (2013). A new HLA map of Europe: Regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. *Human Heredity*, *76*, 162–177.

Sanchez-Mazas, A., Djoulah, S., Busson, M., de Le Monnier Gouville, I., Poirier, J. C., Dehay, C., … Hors, J. (2000). A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families. *European Journal of Human Genetics*, *8*, 33–41.

Sanchez-Mazas, A., Lemaitre, J. F., & Currat, M. (2012). Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 830–839.

Sanchez-Mazas, A., & Tiercy, J.-M. (2007). The HLA molecular diversity of an African population from Eastern Senegal. In J. A. Hansen (Ed.), *Immunobiology of the human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference* (pp. 212–215). Seattle, WA: IHWG Press.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*, 23–35.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., & Hay, S. I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, *434*, 214–217.

Solberg, O. D., Mack, S. J., Lancaster, A. K., Single, R. M., Tsai, Y., Sanchez-Mazas, A., & Thomson, G. (2008). Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Human Immunology*, *69*, 443–464.

Sollid, L. M. (2017). The roles of MHC class II genes and post-translational modification in celiac disease. *Immunogenetics*, *69*, 605–616.

Spinola, H., Bruges-Armas, J., Middleton, D., & Brehm, A. (2005). HLA polymorphisms in Cabo Verde and Guine-Bissau inferred from sequence-based typing. *Human Immunology*, *66*, 1082–1092.

Spinola, H., Couto, A. R., Peixoto, M. J., Anagnostou, P., Destro-Bisol, G., Spedini, G., … Bruges-Armas, J. (2011). HLA class-I diversity in Cameroon: Evidence for a north-south structure of genetic variation and relationships with African populations. *Annals of Human Genetics*, *75*, 665–677.

Sun, H., Yang, Z., Lin, K., Liu, S., Huang, K., Wang, X., … Huang, X. (2015). 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*, *10*, e0134334.

Testi, M., Battarra, M., Lucarelli, G., Isgro, A., Morrone, A., Akinyanju, O., … Sanchez-Mazas, A. (2015). HLA-A-B-C-DRB1-DQB1 phased haplotypes in 124 Nigerian families indicate extreme HLA diversity and low linkage disequilibrium in Central-West Africa. *Tissue Antigens*, *86*, 285–292.

Thomsen, M., Lundegaard, C., Buus, S., Lund, O., & Nielsen, M. (2013). MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics*, *65*, 655–665.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., … Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science*, *324*, 1035–1044.

Torcia, M. G., Santarlasci, V., Cosmi, L., Clemente, A., Maggi, L., Mangano, V. D., … Modiano, D. (2008). Functional deficit of T regulatory cells in Fulani, an ethnic group with low susceptibility to *Plasmodium falciparum* malaria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 646–651.

Torimiro, J. N., Carr, J. K., Wolfe, N. D., Karacki, P., Martin, M. P., Gao, X., … Carrington, M. (2006). HLA class I diversity among rural rainforest inhabitants in Cameroon: Identification of A*2612-B*4407 haplotype. *Tissue Antigens*, *67*, 30–37.

Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., & Pereira, L. (2015). Extensive admixture and selective pressure across the sahel belt. *Genome Biology and Evolution*, *7*, 3484–3495.

Trowsdale, J., & Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics*, *14*, 301–323.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York, NY: Springer.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*, 1–20.

Wickham, H. (2009). *GGPLOT2: Elegant graphics for data analysis*. New York, NY: Springer.

Yague, J., Vazquez, J., & Lopez de Castro, J. A. (1998). A single amino acid change makes the peptide specificity of B*3910 unrelated to B*3901 and closer to a group of HLA-B proteins including the malaria-protecting allotype HLA-B53. *Tissue Antigens*, *52*, 416–421.

Yamazaki, A., Yasunami, M., Ofori, M., Horie, H., Kikuchi, M., Helegbe, G., … Hirayama, K. (2011). Human leukocyte antigen class I polymorphisms influence the mild clinical manifestation of *Plasmodium falciparum* infection in Ghanaian children. *Human Immunology*, *72*, 881–888.

Young, W. C. (1996). *The Rashaayda Bedouin: Arab pastoralists of eastern Sudan*. Fort Worth, TX: Harcourt Brace College Publishers.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Sanchez-Mazas A, Černý V, Di D, 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;00:1–15. https://doi.org/10.1111/mec.14366