Molecular Immunology 112 (2019) 40–50

Contents lists available at ScienceDirect

Molecular Immunology

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

HLAIb worldwide genetic diversity: New *HLA-H* alleles and haplotype structure description

Julien Paganini[a], Laurent Abi-Rached[b,c], Philippe Gouret[a], Pierre Pontarotti[a,b,c], Jacques Chiaroni[d,e], Julie Di Cristofaro[d,e,*]

[a] Xegen, Gemenos, France
[b] Aix Marseille Univ, IRD, APHM, MEPHI, IHU-Mediterranée Infection, Marseille, France
[c] CNRS, Marseille, France
[d] Etablissement Français du Sang PACA Corse, Biologie des Groupes Sanguins, Marseille, France
[e] Aix Marseille Univ, CNRS, EFS, ADES, "Biologie des Groupes Sanguins", Marseille, France

ARTICLE INFO

Keywords:
Genetic diversity
Allele
Haplotype
HLA Ib
Next generation sequencing

ABSTRACT

The classical HLA class I genes (HLA Ia) were extensively studied because of their implication in clinical fields and anthropology. Less is known about worldwide genetic diversity and linkage disequilibrium for non-classical HLA class I genes (HLA Ib) and HLA pseudogenes. Notably, *HLA-H,* which is deleted in a fraction of the population, remains scarcely explored. The aims of this study were 1/ to get further insight into *HLA-H* genetic diversity and into how this variability potentially affects its expression and 2/ to define HLA Ib worldwide allelic diversity and linkage.

Exome sequence data from the 1000 Genomes Project were used to define second field HLA-A, -E, -F, -G and -H typing using PolyPheMe software. Allelic and two-loci haplotype frequencies were estimated using Gene [Rate] software both at worldwide and continental levels.

Eleven novel *HLA-H* alleles identified in exome data were validated by NGS performed on 25 genomic DNA samples from the same cohort. Phylogenetic analysis and frequency distribution of *HLA-H* alleles revealed three clades, each predominantly represented in Admixed American, European and East Asian populations, African populations and South Asian populations. Among these eleven novel alleles, two potentially encode complete transmembrane HLA proteins.

We confirm the high LD between *HLA-H* and *-A,* and between *HLA-H* and *-G*, and show the three genes have distinct worldwide allelic distribution. Conversely, *HLA-E* and *HLA-F* both showed little LD, displayed restricted allelic diversity and practically no difference in their distribution across the planet.

Our work thus reveals an unexpectedly high *HLA-H* genetic diversity, with alleles highly represented in Asia possibly encoding a functional HLA protein. Functional implication of these results remains to be explored, both in physiological and pathological contexts.

1. Introduction

The Major Histocompatibility Complex (MHC) region is the most studied genetic region of the human genome, in large part thanks to the presence of classical HLA class I *(HLA-A, -B* and *-C*) and HLA class II (*HLA-DR* and *-DQ*) genes. Its huge genetic diversity is challenging because of functional implication in many clinical fields such as transplant and graft outcome, or viral escape. However, this diversity also deeply contributed to anthropological science by helping to define *Homo sapiens* evolution and the earliest worldwide migration routes (Parham, 1993). Conversely, less is known about worldwide genetic diversity and Linkage Disequilibrium (LD) of non-classical HLA class I genes (*HLA-E, -F* and *-G*) and of HLA class I pseudogenes. Among these, *HLA-E* and *-G* are the most studied, both at genetic diversity and at functional levels (Hviid and Christiansen, 2005; Kolte et al., 2010). The 1000 Genomes Project (Genomes Project et al., 2015) gives the outstanding opportunity to contribute to studying these loci (Olieslagers et al., 2017; Felicio et al., 2014; Castelli et al., 2014).

Non classical HLA-E, -F and -G (HLA Ib) display specific features compared to classical HLA class I (HLA Ia) such as very low genetic

Abbreviations: LD, linkage disequilibrium; NK, natural killer

* Corresponding author at: Aix Marseille Univ, CNRS, EFS, ADES, "Biologie des Groupes Sanguins", 27 Boulevard Jean Moulin, 13005, Marseille, France.
E-mail address: julie.dicristofaro@efs.sante.fr (J. Di Cristofaro).

https://doi.org/10.1016/j.molimm.2019.04.017
Received 12 February 2019; Received in revised form 9 April 2019; Accepted 23 April 2019
Available online 08 May 2019
0161-5890/

Table 1
DNA samples from the 1000 Genomes Project used as control and validation samples.

Coriell Samples	*HLA-H* Allele	IMGT Submission Number	Genbank Submission Number
HG03673 and NA20524	H*02:07	HWS10053785	MK387860
HG03419 and HG03679	H*02:08	HWS10053787	MK387861
HG03057 and HG03740	H*02:09	HWS10053789	MK387862
NA19143 and NA20510	H*01:03	HWS10053791	MK387859
HG03078 and NA19119	H*02:10	HWS10053793	MK387863
HG03061 and NA20805	H*01:01:02	HWS10053795	MK387856
HG03457 and NA18487	H*01:04	HWS10053797	MK387857
HG03894 and NA20540	H*02:03:02	HWS10053799	MK387864
NA19118	H*02:11	HWS10053801	MK387865
HG02851	H*02:12	HWS10053803	MK387866
HG03548	H*02:13	HWS10053805	MK387867
HG03838	H*02:14	HWS10053807	MK387868
NA20587	H*01:05	HWS10053809	MK387858
NA20521	H*01:02	NA	NA
HG02870	H*02:02	NA	NA
NA20535	H*02:04	NA	NA
HG03076	H*02:05	NA	NA

DNA samples from the 1000 Genomes Project (Coriell Institute, Camden, New Jersey, USA) displaying new *HLA-H* alleles identified in exome sequence data or displaying alleles used as controls (in grey) used in Next Generation Sequencing (NGS) analysis. All samples are hemizygote for *HLA-H*. NA: Not-Applicable.

polymorphism and restricted pattern of antigens presentation. Their role is not to elicit an immune response but rather to inhibit its activation.

HLA-E regulates natural killer cells (NK) and cytotoxic T-lymphocyte cells via its inhibitory receptor CD94/NKG2 (Celik et al., 2015; Allan et al., 2002; Pratheek et al., 2014). HLA-E mRNA is expressed in most tissues (Heinrichs and Orr, 1990) and HLA-E is mobilized at the cell surface by leader peptides of HLA Ia and HLA-G molecules. HLA-E also binds peptide ligands from stress proteins and viruses (Foroni et al., 2014; Kraemer et al., 2014). The two main *HLA-E* alleles, *E*01:01* and *E*01:03*, display similar frequencies worldwide, suggesting an advantage for heterozygous carriers; *E*01:03* is associated with higher expression (Olieslagers et al., 2017; Felicio et al., 2014; Castro et al., 2019; Geraghty et al., 1992a; Pabon et al., 2014; Grimsley and Ober, 1997; Sonon et al., 2018; Ramalho et al., 2017).

HLA-G modulates NK and cytotoxic T-lymphocyte mediated activity as well as B-lymphocyte proliferation and is involved in epithelial cell differentiation (Allan et al., 2002; Rouas-Freiss et al., 1997; Howangyin et al., 2012). Many diseases involving immune tolerance were studied in regards to HLA-G expression variation, especially in pregnancy (Lynge Nilsson et al., 2014; Rebmann et al., 2014). Several studies associated genetic polymorphisms both at coding level and in regulatory regions with inter-individual expression variation (Rebmann et al., 2014). *HLA-G* displays five main alleles with unequal worldwide distribution (Castelli et al., 2014; Castro et al., 2019; Sonon et al., 2018; Oliveira et al., 2018; Carlini et al., 2013, 2016; Castelli et al., 2007, 2017).

HLA-F mRNA is expressed in most cell types and the protein is intracellular, and is mobilized at the cell surface of activated monocytes, NK, B-lymphocyte and T-lymphocyte (Lee et al., 2010; Goodridge et al., 2013a). HLA-F, expressed in an open conformer form and whose function seems independent of peptide loading (Boyle et al., 2006; Goodridge et al., 2013b) is implicated in immune system regulation in pregnancy, infection, autoimmunity and cancer, especially via its interaction with the inhibitory receptor KIR3DS1 (Burian et al., 2016; Garcia-Beltran et al., 2016). *HLA-F* displays four alleles defined at second field resolution, with *F*01:01* representing 90% of allelic diversity (Castro et al., 2019; Carlini et al., 2016; Buttura et al., 2019; Lima et al., 2016).

We formerly reported in a French cohort that *HLA-A* and *-G* were both in highly significant pairwise Global Linkage Disequilibrium (GLD) with the HLA-H locus ($p < 0.001$), but displayed no exclusive association with either *HLA-E* or *HLA-F* (Carlini et al., 2016).

HLA-H (formerly named *HLA-12.4* or *HLA-AR*), located at 55 Kbp away from the telomeric side of *HLA-A,* is more related to HLA-A than to *HLA-B* or *–C* (Malissen et al., 1982; Zemmour et al., 1990) and together with *HLA-J* and *-G* forms a group defined as *HLA-A* related genes (Messer et al., 1992).

HLA-H has orthologs in chimpanzees, bonobos and gorillas, hence the separation between *HLA-H* and other *HLA-A*-related genes predates the divergence of these species (Lawlor et al., 1991; Adams et al., 2001; Hans et al., 2017). A proposed model of evolution suggests the duplication of an ancestral MHC-A/H block 30 million years ago evolving respectively into two genomic blocks containing MHC-A and MHC-H/AL loci (Adams et al., 2000). A second duplication of the MHC-H/AL-containing block generated the MHC-H and MHC-AL blocks seven million years later. The latter was subsequently lost in humans but is retained in ˜50% of chimpanzees by a balancing selection process (Adams et al., 2001, 2000; Gleimer et al., 2011).

Similarly to *HLA-E, -F* and *-G,* limited *HLA-H* polymorphism is documented in the IPD-IMGT/HLA Database 3.33. Indeed, the 12 *HLA-H* allelic variants display less than one hundred SNPs and deletions (Robinson et al., 2015). Sequence preservation can be interpreted by strong selection of primordial biological functions, or as inactivation by deleterious mutation of genes (Zemmour et al., 1990).

Although no transcription signals could be detected upstream from the signal coding sequence in the earliest study describing one *HLA-H* allele, the authors added in their note-added proof that in vitro transcription experiments showed a major transcript initiated upstream of the signal sequence (Malissen et al., 1982). In a recent study based on GeneChip Whole Transcript from Affymetrix, HLA-H transcriptional activity was shown (Aka et al., 2016); these authors analyzed transcript profile modulation on breast cancer cells siRNA and drugs assays and showed a 2.3–2.5 fold upregulation of HLA-H, -F, -G, -E and -A.

One of the first *HLA-H* sequence descriptions showed that all exon/intron junctions were provided with the requisite splicing signal (Malissen et al., 1982). This allele (corresponding to *HLA-H*02:06* in the IPD and IMGT/HLA database) was described to encode 362 amino-acids in the Uniprot database (UniProtKB: P01893), whereas the *HLA-H*01:01:01:01* allele encodes 295 amino-acids. Although the peptide signal and non-cytoplasmic topological domain are encoded by most *HLA-H* alleles described in the IPD and IMGT/HLA database, none but *HLA-H*02:06* displays a full length transmembrane domain. Furthermore all HLA-H alleles lack the cysteine at codon 164 (amino-acid 188 including peptide signal length) critical for the disulfide bond of the α2 domain, impairing the antigen presenting function (Zemmour et al., 1990). They do, however, present three of the four critical cysteines (position 101, 203, 259) (amino-acid 125, 227 and 283 including

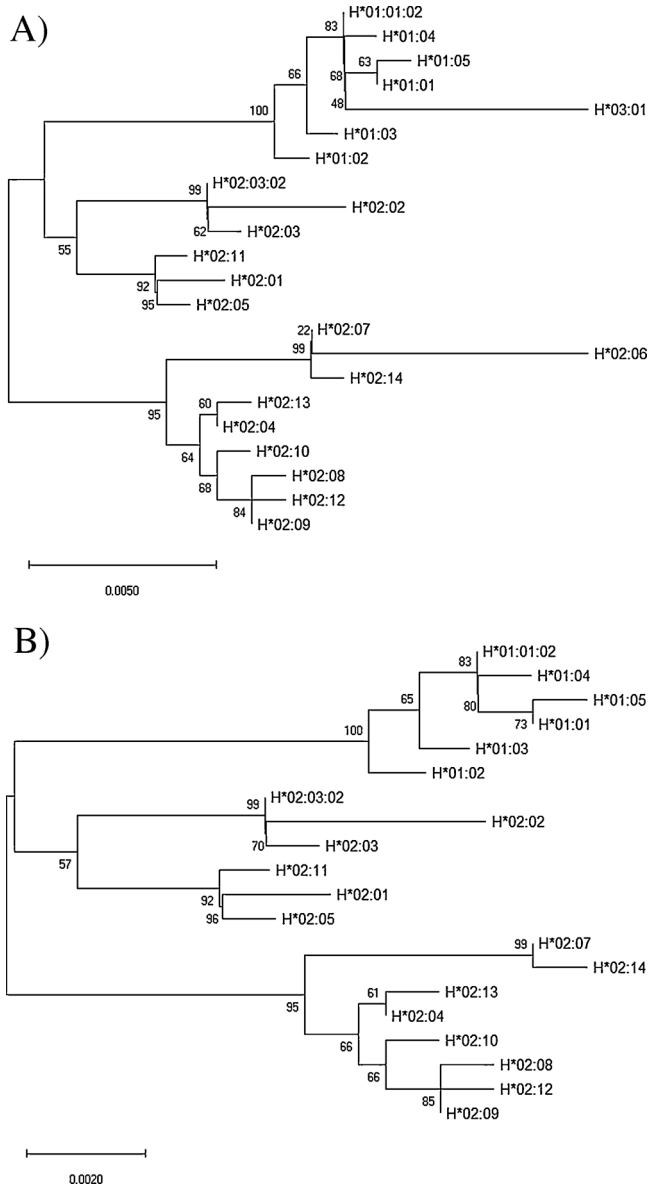

Fig. 1. A and B. Phylogenetic relationship between *HLA-H* alleles defined at second field.

Phylogenetic trees contain either all 22 *HLA-H* allelic sequences from the IPD-IMGT /HLA Database and sequences from new *HLA-H* alleles described in this study (Fig. 1A), or 20 *HLA-H* allelic sequences, excluding *H*03:01* and *H*02:06* alleles not observed in this study (Fig. 1B). Optimal trees are shown (respectively with the sum of branch length = 0.05748059 (Fig. 1A) and 0.04361905 (Fig. 1B)). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). Trees are drawn to scale, with branch lengths in the same units (base differences per site) as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and units are the number of base substitutions per site. There were a total of 1105 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

peptide signal length) and the invariant glycosylation site at position 86 (amino-acid 110 including peptide signal length) (Malissen et al., 1982).

HLA-H was thus defined as a non-functional gene, or pseudogene, reinforced by its genetic deletion in a significant fraction of the population. This deletion of more than 50 kb between *HLA-G* and *HLA-A* loci, spanning *HLA-H*, is in LD with *HLA-A*23/24* and *HLA-G*01:04* alleles (Geraghty et al., 1992a; Carlini et al., 2016; Geraghty et al., 1992b; el Kahloun et al., 1992; Shukla et al., 1991) Therefore, genetic and/or functional studies on HLA Ib have paid little, if any, attention to *HLA-H* genetic diversity.

The aims of this study were 1/ to get further insight into *HLA-H* genetic diversity and its possible implication in expression and 2/ to better define HLA Ib worldwide allelic diversity and haplotypic structure data.

2. Material and methods

2.1. Material

Analyses were performed both on exome sequence data from 2693 individuals forming the 1000 Genomes Project (Genomes Project et al., 2015) and on 25 genomic DNA samples from the same cohort (Coriell Institute, Camden, New Jersey, USA).

2.2. HLA allelic typing from the 1000 genomes project

HLA-A second field allelic typing was from (Abi-Rached et al., 2018). *HLA-E, -F, -G* and *-H* allelic assignment at the second field was performed with the PolyPheMe software v1.2 (Xegen, Gemenos, France) (Abi-Rached et al., 2018) using the IPD-IMGT/HLA Database 3.33 as reference (Robinson et al. (2015)). Briefly, HLA typing first involved isolation of all the reads related to *HLA* loci using Bowtie 2 (Langmead et al., 2009). The reads thus isolated were assigned to each locus targeted (*HLA-E, -F, -G* and *-H*) using an “end to end” mapping step with Bowtie 2 (Langmead et al., 2009). The positive reference datasets included all the *HLA* alleles of each locus investigated. The negative reference dataset included all the alleles from related loci (*HLA* gene and pseudogene sequences except *HLA-E, -F, -G* and *-H*). FASTQ files containing exclusively the specific sequence reads for each locus were generated. For each locus, first field-level types were determined and resolution was then incremented to second field. In this analysis, all loci with at least 300 specific reads identified were analyzed and typed using all the variable positions described in the IPD-IMGT/HLA Database 3.33 (Robinson et al. (2015)).

2.3. HLA-H new allele identification and confirmation, phylogenetic analysis and putative protein prediction

HLA-H typing was not possible for many samples’ exome sequence data, i.e. no allele from the IPD-IMGT/HLA Database 3.33 corresponded to the observed SNP combination. Whole sequence analysis revealed unreported SNP combinations and/or sequence variations in the IPD-IMGT/HLA Database 3.33 (Robinson et al., 2015). These new alleles were thus explored in samples which were hemizygous for *HLA-H*, i.e. displaying one allele *HLA-A*23/24* (Geraghty et al., 1992a; Carlini et al., 2016; Geraghty et al., 1992b; el Kahloun et al., 1992; Shukla et al., 1991).

These new *HLA-H* alleles identified in exome sequence data were confirmed using targeted Next Generation Sequencing (NGS) of 25 genomic DNA samples hemizygous for *HLA-H* from the 1000 Genomes Project (Coriell Institute, Camden, New Jersey, USA). *HLA-H* was amplified by long-range PCR (primer sequences CAAACTCCGTGGGTGASTTT and TGGCTGCTACTCTGGGTTCT, from position -483 to 4418 according to the IPD-IMGT/HLA Database (Robinson et al., 2015)), generating an amplicon of approximately 4900 bp depending on alleles. PCR fragments were sequenced as previously described (Carlini et al., 2017) using a MiSeq NGS platform (Illumina, Eindhoven, The Netherlands). NGS data were analyzed using PolyPheMe software v1.2.

HLA-H cDNA sequences were aligned using the multiple sequence alignment tool MUSCLE in the Molecular Evolutionary Genetics Analysis (MEGA) software version X (Kumar et al., 2018). Evolutionary relationships among *HLA-H* cDNA sequences were inferred using the

Table 2
Putative protein prediction from new *HLA-H* alleles' analysis.

Allele	IMGT Number	Genbank Number	Protein length prediction (Expasy)	Predicted features by Prosite tool (Expasy)	Predicted features by Uniprot (Blast)	Predicted features by Phobius
*A*02:01:01:01*	NA	NA	365	Domain Ig-like: 209-295 Disulfide bound: 227-283	Signal peptide: 1-24 Topological Domain: 25-307 Transmembrane Domain: 308-332 Topological Domain: 333-365	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-307 Transmembrane Domain: 308-332 Cytoplasmic Domain: 333-365
*G*01:01:01:01*	NA	NA	338	Domain Ig-like: 209-287 Disulfide bound: 227-283	Signal peptide: 1-24 Topological Domain: 25-307 Transmembrane Domain: 308-332 Topological Domain: 333-338	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-307 Transmembrane Domain: 308-332 Cytoplasmic Domain: 333-338
*H*01:01:01:01*	NA	NA	295	Domain Ig-like: 209-245 Absent feature: Disulfide bound	No annotation hit	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-295
*H*02:06*	NA	NA	362	Domain Ig-like: 209-297 Disulfide bound: 227-283	Signal peptide: 1-24 Glycosylation: 110 Disulfid bound: 227 283 Topological Domain: 25-307 Transmembrane Domain: 308-332 Topological Domain: 333-362	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-307 Transmembrane Domain: 308-332 Cytoplasmic Domain: 333-362
*H*02:07*	HWS10053785	MK387860	362	Domain Ig-like: 209-297 Disulfide bound: 227-283	Signal peptide: 1-24 Glycosylation: 110 Disulfid bound: 227 283 Topological Domain: 25-307 Transmembrane Domain: 308-332 Topological Domain: 333-362	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-307 Transmembrane Domain: 308-332 Cytoplasmic Domain: 333-362
*H*02:08*	HWS10053787	MK387861	18	No hit	No BLAST hit	No hit
*H*02:09*	HWS10053789	MK387862	18	No hit	No BLAST hit	No hit
*H*01:03*	HWS10053791	MK387859	295	Domain Ig-like: 209-245 Absent feature: Disulfide bound	No annotation hit	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-295
*H*02:10*	HWS10053793	MK387863	215	No hit	Signal peptide: 1-26 Glycosylation: 112 Topological Domain: 27-215	Signal peptide: 1-26 N-Region: 1-8 H-Region:9-21 C-Region: 22-26 Non-Cytoplasmic Domain: 27-215
*H*01:01:02*	HWS10053795	MK387856	295	Domain Ig-like: 209 245 Absent feature: Disulfide bound	No annotation hit	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-295
*H*01:04*	HWS10053797	MK387857	295	Domain Ig-like: 209-245 Absent feature: Disulfide bound	No annotation hit	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-295
*H*02:03:02*	HWS10053799	MK387864	293	Domain Ig-like: 207-243 Absent feature: Disulfide bound	Signal peptide: 1-22 Glycosylation: 108 Topological Domain: 23-293	Signal peptide: 1-22 N-Region: 1-7 H-Region:8-17 C-Region: 18-222

(*continued on next page*)

Table 2 (*continued*)

Allele	IMGT Number	Genbank Number	Protein length prediction (Expasy)	Predicted features by Prosite tool (Expasy)	Predicted features by Uniprot (Blast)	Predicted features by Phobius
						Non-Cytoplasmic Domain: 23-293
*H*02:11*	HWS10053801	MK387865	295	Domain Ig-like: 207-243 Absent feature: Disulfide bound	Signal peptide: 1-24 Glycosylation: 110 Topological Domain: 25-295	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-295
*H*02:12*	HWS10053803	MK387866	18	No hit	No BLAST hit	No hit
*H*02:13*	HWS10053805	MK387867	221	No hit	Signal peptide: 1-26	Signal peptide: 1-26 N-Region: 1-8 H-Region:9-21 C-Region: 22-26 Non-Cytoplasmic Domain: 27-221
*H*02:14*	HWS10053807	MK387868	362	Domain Ig-like: 209-297 Disulfide bound: 227-283	Signal peptide: 1-24 Glycosylation: 110 Disulfid bound: 227 283 Topological Domain: 25-307 Transmembrane Domain: 308-332 Topological Domain: 333-362	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-307 Transmembrane Domain: 308-332 Cytoplasmic Domain: 333-362
*H*01:05*	HWS10053809	MK387858	254	Domain Ig-like: 209-254 Absent feature: Disulfide bound	Signal peptide: 1-24	Signal peptide: 1-24 N-Region: 1-7 H-Region:8-19 C-Region: 20-24 Non-Cytoplasmic Domain: 25-254

Putative protein prediction from new *HLA-H* alleles based on *HLA-H* cDNA sequences using Expasy (Artimo et al., 2012). Specific patterns are shown according to analysis using Prosite, Uniprot and Phobius (Hulo et al., 2008; UniProt Consortium, 2018; Kall et al., 2004). For Uniprot analysis, alignment of the best match is shown. Different *HLA* alleles are shown as examples (in grey: *HLA-A*02:01:01:01, HLA-G*01:01:01:01, HLA-H*01:01:01:01* and *HLA-H*02:06*). NA: Not-Applicable.

Neighbor-Joining method with 1000 bootstrap replicates. Evolutionary distances were computed using the p-distance method and units were the number of base differences per site.

Putative protein prediction from *HLA-H* alleles was performed based on *HLA-H* cDNA sequences using Expasy (Artimo et al., 2012). *HLA-A*02:01:01:01, G*01:01:01:01, H*01:01:01:01* and *H*02:06* alleles were added for reference *HLA-A*02:01:01:01, HLA-G*01:01:01:01, HLA-H*01:01:01:01* alleles were selected because of their high frequency in populations of European descent; they are also commonly used as reference sequences, particularly in the IMGT/ HLA database. *HLA-H*02:06* was selected because it encodes the protein with the longest amino-acid sequence with putative functional patterns. Peptide sequence characteristics for specific domains, motifs or sites were analyzed according to Prosite (https://prosite.expasy.org/), Uniprot (https://www.uniprot.org/) and Phobius (http://phobius.sbc.su.se/). Prosite relies on documentation entries describing protein domains, families and functional sites. Uniprot uses the Basic Local Alignment Search Tool (BLAST) to find regions of local similarity between sequences. Phobius is used for prediction of transmembrane topology and signal peptides (Hulo et al., 2008; UniProt Consortium, 2018; Kall et al., 2004).

2.4. HLA-A, E, -F, -G and -H allelic and two-loci haplotype analysis

Allelic and two-loci haplotype frequencies were estimated based on typing results of each individual at a worldwide population level and at continental level in African populations, Admixed American populations, European populations, East Asian populations and South Asian populations (see (Genomes Project et al., 2015) for population description). Missing data at a locus led to the exclusion of the concerned sample from further analyses. No multiple imputations were used.

Frequencies were estimated using an EM algorithm implemented in the Gene[Rate] computer tools (Nunes, 2014). Two-loci Linkage Disequilibrium (LD) was investigated according to locus proximity, i.e. *HLA-E˜HLA-A, HLA-A˜HLA-H, HLA-H˜HLA-G* and *HLA-G˜HLA-F*. LD was assessed by a likelihood-ratio test on the frequency estimations (Nunes, 2014) and was provided for specific pairs of alleles as a list of standardized residuals for each observed haplotype. Values greater than |2| were considered to be a significant deviation (Nunes et al., 2014).

3. Results

3.1. HLA-H worldwide diversity analysis identifies 11 novel alleles

Eleven *HLA-H* novel alleles identified in exome data were validated by concordance with targeted NGS performed on identical DNA samples (Table 1). *HLA-H* novel alleles, submitted to Genbank and to the IPD-IMGT /HLA Database are described in Appendice Table 1. Official names for the eleven *HLA-H* novel alleles identified in this study submitted to the IPD-IMGT /HLA Database have been officially assigned by the WHO Nomenclature Committee. This followed the agreed policy and was subject to the conditions stated in the most recent Nomenclature Report (Marsh et al., 2010).

3.2. HLA-H phylogenetic analysis defines three distinct groups

All *HLA-H* sequences, i.e. from the IPD-IMGT/HLA Database 3.33 and identified in this study (N = 22), were aligned and evolutionary relationships among these cDNA sequences were inferred using the Neighbor-Joining method. The resulting phylogenetic tree (Fig. 1A) shows three main clades. The first clade contains all the *HLA-H*01* alleles and the *H*03:01* allele; the second clade includes six *HLA-H*02* sequences (*HLA-H*02:01/02/03/03:02/05/11*); the third clade includes nine *HLA-H*02* sequences, seven of which are novel (*HLA-*

Table 3
A, B, C, D and E. *HLA-A, -E, -F, -G* and *-H* allelic frequencies.

HLA-A alleles	ALL (N = 2157)	AFR (N = 592)	EUR (N = 416)	AMR (N = 294)	EAS (N = 446)	SAS (N = 409)
A						
*A*02:01*	**0.139**	**0.115**	**0.307**	**0.222**	**0.076**	**0.032**
*A*11:01*	**0.093**	0.001	**0.058**	**0.041**	**0.234**	**0.160**
*A*24:02*	**0.090**	0.007	**0.073**	**0.124**	**0.171**	**0.124**
*A*01:01*	**0.070**	0.028	**0.127**	**0.043**	0.020	**0.155**
*A*03:01*	**0.065**	**0.054**	**0.150**	**0.086**	0.011	**0.051**
*A*33:03*	**0.049**	**0.058**	0.006	0.005	**0.076**	**0.087**
*A*68:01*	**0.035**	**0.030**	0.030	**0.054**	0.007	**0.066**
*A*23:01*	**0.034**	**0.095**	0.024	0.025	0.000	0.005
*A*26:01*	**0.030**	0.019	**0.035**	0.014	**0.030**	**0.056**
*A*30:02*	0.030	**0.078**	0.016	**0.036**	0.000	0.004
*A*30:01*	0.029	**0.078**	0.011	0.012	0.014	0.009
*A*02:07*	0.026				**0.128**	
*A*31:01*	0.025	0.007	0.026	**0.045**	**0.034**	0.024
*A*68:02*	0.023	**0.070**	0.005	0.024		
*A*32:01*	0.022	0.011	**0.044**	0.019	0.006	**0.036**
*A*29:02*	0.022	0.026	**0.038**	**0.054**		
*A*02:11*	0.018	0.001		**0.032**		**0.075**
*A*74:01*	0.016	**0.056**		0.007		
*A*02:06*	0.015		0.001	0.018	**0.043**	0.019
*A*02:03*	0.015				**0.055**	0.020
*A*02:02*	0.013	**0.040**		0.015		
*A*33:01*	0.012	0.029	0.006	0.019		
*A*34:02*	0.011	**0.040**		0.003		
*A*36:01*	0.011	**0.041**				
*A*02:05*	0.009	0.016	0.006	0.012	0.001	0.009
*A*66:01*	0.008	0.023	0.006	0.003		0.001
*A*11:02*	0.007				**0.034**	
*A*29:01*	0.005	0.001		0.009	0.017	0.004
*A*24:07*	0.005				0.006	0.020
*A*23:17*	0.004	0.015		0.002		
*A*01:02*	0.004	0.014	0.001	0.002		
*A*66:02*	0.004	0.014				
*A*25:01*	0.004	0.000	0.016	0.005		
*A*80:01*	0.004	0.010	0.001	0.003		
*A*03:02*	0.002	0.001	0.002	0.003		0.005
*A*66:03*	0.002	0.006				
*A*02:22*	0.002			0.012		
Total	0.951	0.982	0.989	0.949	0.962	0.960
B						
*E*01:01*	**0.501**	**0.581**	**0.558**	**0.509**	**0.332**	**0.526**
*E*01:03*	**0.465**	**0.405**	**0.391**	**0.462**	**0.661**	**0.433**
*E*01:05*	0.004	0.013	0.004			
*E*01:06*	0.008		**0.036**	0.005		0.004
*E*01:09*	0.001	0.001	0.002	0.002		
Total	0.979	1.000	0.991	0.978	0.992	0.962
C						
*F*01:01*	**0.843**	**0.781**	**0.757**	**0.879**	**0.977**	**0.925**
*F*01:03*	**0.112**	**0.173**	**0.176**	**0.117**	0.011	**0.069**
*F*01:02*	0.005	0.017		0.003		
*F*01:05*	0.002				0.010	
Total	0.963	0.971	0.933	1.000	0.998	0.995
D						
*G*01:01*	**0.658**	**0.567**	**0.825**	**0.706**	**0.682**	**0.612**
*G*01:04*	**0.185**	**0.231**	**0.082**	**0.150**	**0.251**	**0.190**
*G*01:03*	**0.052**	**0.110**	0.028	**0.085**	0.006	0.027
*G*01:06*	**0.046**	0.004	**0.054**	0.028	0.015	**0.152**
*G*01:05N*	0.029	**0.078**	0.011	0.012	0.014	0.009
Total	0.970	0.990	0.999	0.981	0.966	0.990
E						
*H*01:01*	**0.234**	**0.121**	**0.310**	**0.311**	**0.360**	**0.152**
*H*Del*	**0.170**	**0.120**	**0.108**	**0.160**	**0.218**	**0.172**
*H*02:05*	**0.108**	**0.256**	**0.050**	**0.103**	0.019	**0.075**
*H*02:07*	**0.086**	0.001	**0.058**	**0.042**	**0.196**	**0.158**
*H*02:01*	**0.069**	0.021	**0.129**	**0.044**	0.020	**0.158**
*H*02:04*	**0.064**	**0.048**	**0.143**	**0.089**	0.011	**0.057**
*H*01:02*	**0.061**	**0.059**	**0.060**	0.028	**0.046**	**0.105**
*H*02:08*	**0.039**	**0.056**	0.006	0.005	**0.070**	**0.039**
*H*02:09*	**0.037**	**0.047**	0.027	**0.051**	**0.035**	0.026
*H*02:02*	0.027	0.027	**0.038**	**0.063**	0.016	0.004
*H*01:01:02*	0.023	**0.054**	0.015	**0.033**		0.004
*H*01:03*	0.022	**0.056**	0.006	0.028	0.001	0.010

(*continued on next page*)

Table 3 (*continued*)

HLA-A alleles	ALL (N = 2157)	AFR (N = 592)	EUR (N = 416)	AMR (N = 294)	EAS (N = 446)	SAS (N = 409)
*H*02:03:02*	0.021	0.013	**0.043**	0.015	0.003	**0.036**
*H*02:10*	0.017	**0.064**		0.002		
*H*01:04*	0.008	0.025	0.001	0.005		
*H*02:12*	0.006	0.009	0.006	0.016		
*H*02:11*	0.005	0.018		0.003		
*H*02:03*	0.002		0.001	0.003	0.005	0.004
*H*02:13*	0.002	0.006				
Total	0.999	1.000	0.999	1.000	1.000	0.999

HLA-A, -E, -F, -G and *-H* allelic frequencies, observed in at least 0.2%, in worldwide populations (ALL) and in African populations (AFR), Admixed American populations (AMR), European populations (EUR), East Asian populations (EAS) and South Asian populations (SAS) (see (Genomes Project et al., 2015) for population description). Number of individuals included in each group is given in brackets. Frequencies above 3% are in bold.

*H*02:04/06/07/08/09/10/12/13/14*). The first and third clades contain sequences that were not observed in the individuals forming the 1000 Genomes Project (*H*03:01* and *H*02:06* alleles, respectively) and that display divergent sequences (long branches). Thus, a second alignment containing all observed *HLA-H* sequences except the *H*03:01* and *H*02:06* alleles (N = 20) was performed and their evolutionary relationships are shown in Figure X (Fig. 1B). This phylogenetic tree displays the same groupings as the former one, without the divergent sequences.

3.3. Two novel HLA-H alleles display all patterns of full-length HLA protein

Putative protein prediction from the novel *HLA-H* alleles based on *HLA-H* cDNA sequences using Expasy (Artimo et al., 2012) ranged from 18 amino-acid (AA) to 362 A A (Table 2). Peptide sequence analysis using Prosite, Uniprot and Phobius (Hulo et al., 2008; UniProt Consortium, 2018; Kall et al., 2004) are described in Table 2. Reference protein sequences encoded by *HLA-A*02:01:01:01, G*01:01:01:01* and *H*02:06* displayed all patterns of full-length HLA proteins whereas those encoded by *H*01:01:01:01* and *H*02:04* did not show any of them. Specific patterns of transmembrane HLA protein were found in new alleles *HLA-H*02:07* and *HLA-H*02:14:* a peptide signal, a non-cytoplasmic domain, a transmembrane domain, a cytoplasmic domain, a glycosylation site and a disulfide bond. The other alleles newly reported, as well as all *HLA-H* alleles described so far in the IPD-IMGT /HLA Database (Robinson et al., 2015) lacked all or part of these critical domains and/or sites.

3.4. HLA-A, -G and -H display distinct worldwide allelic distribution conversely to HLA-E and –F.

From the exome sequence data for 2693 individuals forming the 1000 Genomes Project, 2160 (80%) were accurately resolved (*i.e.* defined without ambiguity at second field level) for *HLA-A, -E, -H, -F and -G* loci. Typing results for each sample and each locus are given in Appendice Table 2. Their allelic frequencies are given in Table 3A to E. While thirty-seven second field *HLA-A* alleles represent over 95% of all worldwide alleles, the different populations defined at continental level display however a great disparity regarding the most representative *HLA-A* alleles (Table 3A). For instance, *HLA-A*11:01* was barely present in Africa, whereas it was most frequent in East Asian populations and found at intermediate frequencies in European populations.

Five second field *HLA-E* alleles represent more than 97% of frequencies (Table 3B), in concordance with published data showing an equal distribution between *HLA-E*01:01* and *E*01:03* (Olieslagers et al., 2017; Felicio et al., 2014; Castro et al., 2019; Geraghty et al., 1992a; Pabon et al., 2014; Grimsley and Ober, 1997; Sonon et al., 2018; Ramalho et al., 2017); *E*01:06* and *E*01:05* are respectively virtually only observed in Europe and in Africa. *HLA-E*01:09* displays very low frequencies everywhere and was not observed in Asia.

Four second field *HLA-F* alleles represent more than 96% of worldwide alleles (Table 3C), *F*01:01* was the most frequent. *F*01:03* was observed at over 10% in all continents, except in East Asian populations.

Five second field *HLA-G* alleles are observed (Table 3D), in accordance with published data (Castelli et al., 2014; Castro et al., 2019; Sonon et al., 2018; Oliveira et al., 2018; Carlini et al., 2013, 2016; Castelli et al., 2007, 2017; Di Cristofaro et al., 2011, 2013). *HLA-G*01:06* is mostly present in South Asia and to a lesser extent in Europe, whereas *HLA-G*01:05 N* has a higher frequency in Africa.

Nineteen second field *HLA-H* alleles represent 99% of worldwide frequencies, among which 11 novel alleles (Table 3E), displaying an unreported allelic frequency. *HLA-H*01:01* is the most frequent allele except in Africa, where *HLA-H*02:05* is more common. The deletion encompassing *HLA-H* (named *HLA-H*Del*) is observed at over 10% in all continents and is highly represented in East Asia. The new allele *HLA-H*02:07* has a higher frequency both in East and South Asian populations but is absent in Africa. This is in accordance with this allele, referred to as *H*02:04new* and defined by a silent G > A substitution at position 368 compared to *H*02:04,* being observed at ˜9% in a French population (Carlini et al., 2016). Of note, *H*03:01* and *H*02:06* alleles, reported in early studies on *HLA-H* (Malissen et al., 1982; Zemmour et al., 1990) are not observed in the 1000 Genomes Project panel.

3.5. High two-loci linkage disequilibrium between HLA-H and -A, and between HLA-H and -G

Haplotypes estimated between each pair of loci and their combined frequencies are given in Table 4A to D. Strong LD is observed between *HLA-A* and *-H,* and *HLA-H* and *-G* alleles, as all haplotypes with a frequency above 3% were in significant LD with very high standardized residual values.

In contrast, *HLA-E* and *-A,* and *HLA-G* and *-F* alleles display few significant associations with standardized residuals very close to the threshold value for significance; indeed the main *HLA-E* and *-F* alleles (i.e. *E*01:01* and *E*01:03* and *F*01:01* and *F*01:03*) were not observed to be exclusively associated with *HLA-A* and *-G* alleles, with the exception of *E*01:01˜A*01:01* in Europe and South Asia.

The tight associations found between *HLA-A* and *-H,* and *HLA-H* and *-G* are concordant with results previously observed at the scale of a French population (Carlini et al., 2016) and further confirm known associations, like the deletion encompassing *HLA-H* with *A*23* and *A*24* alleles, but also reveal unobserved exclusive associations between *HLA-A* and *-H*, such as that between *HLA-A*11:01* and *-H*02:07.*

4. Discussion

The aim of this study was to investigate worldwide HLA Ib genetic diversity. To achieve this goal, we used public data from exome sequencing and confirmed novel sequences experimentally. This approach allowed to 1/ get further insight into *HLA-H* genetic diversity and how this plasticity can potentially lead to functional diversity and

Table 4
A, B, C and D. *HLA-E¯HLA-A, HLA-A¯HLA-H, HLA-H¯HLA-G* and *HLA-G¯HLA-F* Linkage Disequilibrium (LD) and frequencies.

Populations	ALL (N = 2157)		AFR (N = 592)		AMR (N = 294)		EAS (N = 446)		EUR (N = 416)		SAS (N = 409)	
HLA-E¯HLA-A haplotypes	obs	stdres	obs	stdres	obs	stdres	obs	stdres	obs	stdres	obs	stdres
A												
*E*01:03¯A*02:01*	**0.074**	**2.4**	**0.097**	**7.8**	**0.072**	**2.2**	**0.042**	1.1	**0.128**	0.6	0.016	0.5
*E*01:01¯A*02:01*	**0.065**	1.2	0.016	**6.5**	**0.134**	1.4	0.019	1.1	**0.178**	0.4		
*E*01:01¯A*01:01*	**0.063**	**9.6**	0.015	0.4	**0.034**	1.9	0.020	**4.9**	**0.124**	**5.5**	**0.140**	**5.6**
*E*01:03¯A*11:01*	**0.055**	**3.7**			0.010	1.5	**0.166**	0.8	0.017	1.2	**0.091**	**2.2**
*E*01:03¯A*24:02*	**0.055**	**4.1**	0.004	1.0	**0.079**	**2.1**	**0.111**	0.2	**0.032**	0.6	**0.071**	**2.1**
*E*01:03¯A*03:01*	**0.040**	**3.5**	0.027	1.2	**0.067**	**3.2**	0.011	1.3	**0.082**	**2.7**	**0.034**	**2.3**
*E*01:01¯A*11:01*	**0.038**	**2.6**			0.026	0.9	**0.067**	1.1	**0.041**	1.4	**0.070**	1.3
*E*01:01¯A*33:03*	**0.037**	**5.1**	**0.033**	0.1	0.005	1.2	**0.059**	**6.4**	0.002	0.5	**0.083**	**4.8**
*E*01:01¯A*24:02*	**0.035**	**3.0**			**0.045**	1.7	**0.059**	0.3	**0.041**	0.0	**0.052**	1.4
*E*01:01¯A*30:01*	0.028	**7.2**	**0.074**	**4.5**	0.012	1.8	0.014	**4.0**	0.011	1.8	0.009	1.7
*E*01:01¯A*23:01*	0.027	**4.7**	**0.073**	**2.5**	0.017	1.0			0.015	0.5	0.003	0.3
*E*01:03¯A*02:07*	0.024	**7.7**					**0.116**	**3.8**				
*E*01:01¯A*30:02*	0.022	**3.8**	**0.065**	**3.1**	0.013	0.9			0.005	1.1	0.004	1.1
*E*01:01¯A*68:01*	0.018	0.3	0.008	**2.6**	0.020	1.2	0.005	1.4	0.022	1.2	**0.039**	0.6
*E*01:03¯A*68:02*	0.018	**4.6**	**0.055**	**5.3**	0.016	1.2			0.002	0.2		
*E*01:01¯A*03:01*	0.018	**5.4**	0.027	0.9	0.015	**3.3**			0.029	**5.2**	0.012	**2.5**
*E*01:03¯A*29:02*	0.017	**4.8**	0.017	**2.0**	**0.048**	**3.4**			**0.036**	**4.9**		
*E*01:03¯A*68:01*	0.017	0.4	0.022	**2.9**	**0.035**	1.4			0.008	1.0	0.028	0.1
*E*01:01¯A*32:01*	0.017	**4.1**	0.010	1.5	0.010	0.1	0.006	**2.6**	**0.034**	1.6	0.026	1.5
*E*01:01¯A*26:01*	0.016	0.1	0.012	0.2	0.003	1.4	0.008	0.6	0.015	0.9	**0.037**	1.3
*E*01:01¯A*74:01*	0.015	**4.9**	**0.051**	**3.5**	0.007	1.4						
*E*01:03¯A*26:01*	0.015	0.3	0.007	0.2	0.010	1.0	0.022	0.4	0.019	1.4	0.020	0.8
*E*01:03¯A*02:03*	0.014	**5.2**					**0.048**	1.8			0.020	**3.5**
*E*01:03¯A*02:11*	0.014	**4.1**			0.003	**2.3**					**0.072**	**6.1**
*E*01:03¯A*33:03*	0.012	**4.8**	0.025	0.3			0.016	**4.5**			0.006	**4.7**
*E*01:03¯A*31:01*	0.012	0.3	0.003	0.1	0.014	1.2	0.025	0.5	0.020	**2.6**		
*E*01:03¯A*02:02*	0.010	**3.3**	0.030	**3.8**	0.010	0.9						
*E*01:01¯A*36:01*	0.010	**3.7**	**0.039**	**3.3**								
*E*01:01¯A*33:01*	0.009	**2.5**	0.022	1.4	0.019	**2.3**			0.004	0.3		
*E*01:03¯A*02:06*	0.008	1.0			0.010	0.4	0.024	0.8			0.008	0.0
*E*01:01¯A*34:02*	0.008	**2.2**	0.029	1.2	0.003	1.0						
*E*01:01¯A*31:01*	0.008	**2.8**	0.004	0.0	0.017	0.9	0.008	1.0	0.007	1.9	0.005	1.9
*E*01:03¯A*30:02*	0.008	**3.3**	0.013	**3.5**	0.022	1.0			0.010	1.5		
*E*01:06¯A*03:01*	0.008	**21.6**			0.005	**5.4**			**0.037**	**12.4**	0.004	**7.3**
*E*01:03¯A*11:02*	0.007	**4.2**					**0.033**	**2.1**				
*E*01:03¯A*66:01*	0.007	**3.2**	0.017	**2.9**	0.003	1.1			0.006	**2.2**		
*E*01:03¯A*01:01*	0.007	**9.3**	0.013	0.4	0.009	1.8						
*E*01:01¯A*02:06*	0.007	0.5			0.008	0.3	0.020	1.3			0.008	0.5
*E*01:03¯A*23:01*	0.006	**5.3**	0.009	**5.2**	0.007	1.0			0.009	0.2		
*E*01:01¯A*68:02*	0.005	**4.0**	0.015	**4.3**	0.008	1.0			0.003	0.0		
*E*01:01¯A*02:05*	0.005	0.6	0.008	0.5	0.006	0.1					0.009	1.7
*E*01:01¯A*29:01*	0.005	**3.4**			0.009	1.5	0.017	**4.4**			0.004	1.1
*E*01:03¯A*32:01*	0.004	**3.7**			0.009	0.0			0.011	1.4	0.010	1.3
*E*01:03¯A*02:05*	0.004	0.4	0.008	0.6	0.004	0.6			0.003	0.6		
*E*01:03¯A*24:07*	0.004	**2.1**					0.002	1.0			0.016	**2.4**
*E*01:03¯A*01:02*	0.004	**2.4**	0.014	**3.7**	0.002	0.8						
*E*01:01¯A*25:01*	0.004	**2.8**			0.005	1.2			0.016	**2.1**		
*E*01:01¯A*29:02*	0.004	**4.5**	0.007	**2.1**	0.003	**3.5**						
*E*01:01¯A*80:01*	0.003	2.0	0.009	1.2	0.003	1.0						
*E*01:03¯A*34:02*	0.003	**2.3**	0.011	1.5								
*E*01:01¯A*02:11*	0.003	**4.4**			0.025	1.7						
*E*01:03¯A*33:01*	0.003	**2.2**	0.007	1.6								
*E*01:01¯A*02:02*	0.003	**2.8**	0.010	**3.0**	0.004	1.0						
*E*01:01¯A*66:02*	0.003	1.4	0.010	0.9								
*E*01:03¯A*23:17*	0.002	0.5	0.003	1.1								
*E*01:01¯A*23:17*	0.002	0.7	0.008	0.3	0.002	0.7						
Total	0.930		0.927		0.888		0.918		0.967		0.897	
B												
*A*02:01¯H*01:01*	**0.139**	**38.1**	**0.115**	**29.3**	**0.223**	**13.7**	**0.076**	**8.7**	**0.305**	**18.7**	**0.032**	**11.1**
*A*24:02¯H*Del*	**0.090**	**39.3**	0.007	**7.2**	**0.125**	**17.9**	**0.170**	**20.3**	**0.073**	**21.2**	**0.121**	**19.4**
*A*11:01¯H*02:07*	**0.081**	**53.3**			**0.041**	**23.0**	**0.171**	**17.1**	**0.058**	**27.1**	**0.158**	**23.5**
*A*01:01¯H*02:01*	**0.064**	**56.0**	0.007	**8.9**	**0.043**	**23.1**	0.020	**29.2**	**0.127**	**24.7**	**0.155**	**23.6**
*A*03:01¯H*02:04*	**0.063**	**59.9**	**0.048**	**30.6**	**0.085**	**21.2**	0.011	**29.4**	**0.150**	**25.0**	**0.051**	**25.4**
*A*33:03¯H*02:08*	**0.038**	**54.1**	**0.056**	**31.6**	0.005	**23.4**	**0.067**	**25.2**	0.006	**28.7**	**0.039**	**17.4**
*A*68:01¯H*02:05*	**0.035**	**33.1**	0.030	**8.4**	**0.055**	**15.9**	0.007	**17.3**	0.030	**21.2**	**0.066**	**24.9**
*A*23:01¯H*Del*	**0.034**	**24.4**	**0.095**	**26.8**	0.025	**8.0**			0.024	**12.1**	0.005	**4.0**
*A*26:01¯H*01:02*	0.030	**43.4**	0.018	**17.6**	0.014	**16.8**	**0.030**	**23.1**	**0.035**	**20.7**	**0.056**	**18.6**
*A*30:01¯H*02:05*	0.029	**30.5**	**0.079**	**14.1**	0.012	**7.4**	0.014	**24.6**	0.011	**12.7**	0.009	**9.0**
*A*31:01¯H*02:09*	0.024	**51.3**	0.006	**10.8**	**0.045**	**21.8**	**0.034**	**28.4**	0.026	**28.0**	0.024	**27.0**
*A*02:07¯H*01:01*	0.024	**16.1**					**0.119**	**10.7**				
*A*30:02¯H*01:01:02*	0.023	**55.8**	**0.054**	**26.3**	**0.032**	**22.1**			0.014	**27.1**	0.004	**28.5**

(*continued on next page*)

Table 4 (*continued*)

Populations	ALL (N = 2157)		AFR (N = 592)		AMR (N = 294)		EAS (N = 446)		EUR (N = 416)		SAS (N = 409)	
HLA-E¯HLA-A haplotypes	obs	stdres	obs	stdres	obs	stdres	obs	stdres	obs	stdres	obs	stdres
*A*68:02¯H*02:05*	0.023	**27.0**	**0.070**	**13.3**	0.024	**10.4**			0.005	**8.5**		
*A*29:02¯H*02:02*	0.022	**57.0**	0.026	**33.1**	**0.054**	**21.1**			**0.038**	**27.8**		
*A*32:01¯H*02:03:02*	0.019	**57.8**	0.011	**31.2**	0.015	**21.5**			**0.043**	**27.4**	**0.032**	**24.6**
*A*02:11¯H*01:01*	0.018	**14.1**			**0.032**	**5.3**					**0.076**	**17.0**
*A*74:01¯H*02:05*	0.016	**22.0**	**0.054**	**11.3**	0.007	**5.6**						
*A*02:06¯H*01:01*	0.015	**12.5**			0.016	**3.5**	**0.043**	**6.6**			0.019	**8.5**
*A*02:03¯H*01:01*	0.015	**12.7**					**0.055**	**7.4**			0.020	**8.8**
*A*02:02¯H*01:03*	0.013	**49.0**	**0.040**	**27.3**	0.015	**17.6**						
*A*11:01¯H*01:01*	0.013	**4.0**					**0.063**	**2.1**				
*A*33:03¯H*01:02*	0.012	**10.7**	0.003	0.5			0.009	**2.8**			**0.049**	**11.9**
*A*36:01¯H*02:10*	0.011	**51.3**	**0.041**	**25.9**								
*A*02:05¯H*01:03*	0.009	**40.6**	0.016	**17.4**	0.012	**15.5**			0.006	**28.7**	0.009	**26.5**
*A*66:01¯H*01:02*	0.008	**22.6**	0.023	**20.1**	0.003	**8.2**			0.006	**8.6**		
*A*30:02¯H*01:04*	0.007	**30.4**	0.024	**17.5**	0.003	**5.8**						
*A*11:02¯H*02:07*	0.007	**16.7**					**0.033**	**9.8**				
*A*33:01¯H*02:12*	0.006	**45.9**	0.009	**19.0**	0.015	**21.4**			0.006	**28.7**		
*A*33:01¯H*02:09*	0.006	**16.4**	0.019	**16.1**	0.003	1.9						
*A*34:02¯H*02:09*	0.006	**17.9**	0.022	**15.9**								
*A*01:01¯H*02:10*	0.006	**8.7**	0.021	**15.7**								
*A*34:02¯H*02:11*	0.005	**44.3**	0.018	**21.6**	0.003	**24.2**						
*A*24:07¯H*Del*	0.005	**9.3**					0.006	**3.7**			0.020	**8.0**
*A*29:01¯H*02:02*	0.005	**28.6**			0.009	**8.4**	0.017	**30.1**			0.004	**28.5**
*A*25:01¯H*01:02*	0.004	**15.2**			0.005	**10.1**			0.016	**13.9**		
*A*01:02¯H*02:01*	0.004	**15.1**	0.014	**27.3**								
*A*66:02¯H*02:05*	0.004	**10.8**	0.014	**5.9**								
*A*23:17¯H*Del*	0.004	**8.6**	0.014	**10.5**								
*A*80:01¯H*01:02*	0.003	**13.5**	0.009	**12.3**	0.003	**8.2**						
*A*32:01¯H*02:03*	0.002	**20.9**			0.003	**10.2**	0.005	**26.6**			0.004	**8.8**
*A*02:22¯H*01:01*	0.002	**4.2**			0.012	**3.3**						
*A*03:02¯H*02:04*	0.002	**11.1**			0.003	**4.3**			0.002	**3.2**	0.005	**8.1**
*A*03:01¯H*02:13*	0.002	**9.6**	0.006	**10.8**								
Total	0.948		0.966		0.944		0.947		0.982		0.955	
C												
*H*01:01¯G*01:01*	**0.235**	**12.5**	**0.121**	**6.6**	**0.306**	**3.9**	**0.364**	**6.2**	**0.310**	**3**	**0.152**	**5.2**
*H*Del¯G*01:04*	**0.128**	**35.1**	**0.117**	**18.1**	**0.141**	**18.1**	**0.176**	**15.1**	**0.075**	**20**	**0.143**	**17.2**
*H*02:07¯G*01:01*	**0.082**	**6.9**			**0.041**	1.6	**0.176**	**3.2**	**0.058**	1	**0.156**	**5.1**
*H*02:05¯G*01:01*	**0.081**	**2.2**	**0.178**	**2.7**	**0.090**	1.5	0.006	1.8	**0.039**	0	**0.066**	**2.6**
*H*02:04¯G*01:01*	**0.064**	**6.8**	**0.047**	**4.0**	**0.088**	**2.3**	0.011	1.2	**0.145**	**2**	**0.057**	**3.3**
*H*01:02¯G*01:01*	**0.060**	**6.6**	**0.059**	**4.7**	0.029	1.5	**0.045**	**2.2**	**0.059**	1	**0.106**	**4.6**
*H*02:01¯G*01:06*	**0.046**	**49.8**	0.004	**15.2**	0.028	**18.5**	0.015	**24.8**	**0.054**	**16**	**0.153**	**23.4**
*H*02:08¯G*01:04*	**0.038**	**24.0**	**0.054**	**12.4**	0.005	**3.8**	**0.070**	**11.7**	0.006	**7**	**0.039**	**10.4**
*H*02:05¯G*01:05N*	0.029	**30.3**	**0.078**	**14.0**			0.014	**24.6**				
*H*02:02¯G*01:01*	0.027	**4.6**	0.027	**3.2**	**0.063**	**2.1**	0.016	1.6	**0.039**	1	0.004	0.9
*H*02:09¯G*01:03*	0.025	**33.7**	**0.044**	**18.6**	**0.042**	**13.8**	0.005	**9.2**	0.016	**16**	0.016	**16.4**
*H*01:01:02¯G*01:01*	0.023	**4.1**	**0.054**	**4.5**	**0.033**	1.6			0.015	1	0.004	0.9
*H*02:01¯G*01:01*	0.022	**7.1**	0.015	1.1	0.017	1.9	0.005	**2.4**	**0.075**	**3**		
*H*01:03¯G*01:03*	0.021	**38.4**	**0.052**	**20.2**	0.027	**12.4**	0.001	**13.2**	0.006	**13**	0.010	**16.8**
*H*02:03:02¯G*01:01*	0.021	**3.7**	0.012	1.7	0.015	1.0	0.003	0.7	**0.043**	1	**0.036**	**2.6**
*H*02:10¯G*01:04*	0.015	**13.7**	**0.057**	**11.8**								
*H*02:09¯G*01:01*	0.012	**5.1**	0.003	**5.0**	0.009	**3.4**	0.029	1.0	0.011	**2**	0.010	1.3
*H*Del¯G*01:01*	0.010	**18.9**	0.003	**8.4**			0.013	**9.7**	0.024	**6**	0.018	**7.3**
*H*01:04¯G*01:01*	0.008	**2.2**	0.025	**3.2**	0.003	0.1			0.001	0		
*H*02:12¯G*01:03*	0.006	**20.0**	0.008	**7.9**	0.015	**9.3**			0.006	**13**		
*H*02:11¯G*01:01*	0.005	1.9	0.018	**2.5**	0.003	0.2						
*H*02:03¯G*01:01*	0.002	1.3			0.003	0.2	0.005	0.8	0.001	0	0.003	0.3
*H*02:07¯G*01:04*	0.002	**7.4**			0.002	1.4					0.003	**4.4**
*H*02:13¯G*01:01*	0.002	1.1	0.006	1.5								
Total	0.961		0.980		0.960		0.953		0.981		0.975	
D												
*G*01:01¯F*01:01*	**0.581**	1.5	**0.510**	**2.5**	**0.618**	0.1	**0.665**	0.1	**0.624**	0.1	**0.551**	0.5
*G*01:04¯F*01:01*	**0.133**	**3.5**	**0.083**	**7.2**	**0.122**	0.7	**0.252**	0.3	**0.041**	**2.4**	**0.185**	0.5
*G*01:01¯F*01:03*	**0.058**	**3.6**	0.020	**8.2**	**0.087**	0.4	0.009	0.6	**0.134**	0.8	**0.061**	**2.4**
*G*01:04¯F*01:03*	**0.051**	**13.4**	**0.149**	**18.4**	0.029	**2.0**			**0.041**	**6.4**	0.005	**2.1**
*G*01:03¯F*01:01*	**0.050**	1.8	**0.105**	**2.1**	**0.085**	0.9	0.005	0.4	0.028	1.3	0.027	0.3
*G*01:06¯F*01:01*	**0.044**	1.8	0.003	0.3	0.027	0.4	0.013	0.4	**0.054**	1.8	**0.153**	0.8
*G*01:05N¯F*01:01*	0.029	1.9	**0.077**	0.0	0.011	0.1	0.013	0.2	0.011	0.8		
*G*01:01¯F*01:02*	0.004	1.1	0.015	1.9								
Total	0.951		0.961		0.979		0.956		0.932		0.981	

HLA-E¯HLA-A, HLA-A¯HLA-H, HLA-H¯HLA-G and *HLA-G¯HLA-F* Linkage Disequilibrium (LD) provided for pairs of alleles observed in at least 0.2% in worldwide populations (ALL) and in African populations (AFR), Admixed American populations (AMR), European populations (EUR), East Asian populations (EAS) and South Asian populations (SAS) (see (Genomes Project et al., 2015) for population description). Number of individuals included in each group is given in brackets. Observed frequencies (obs) above 3% and standardized residuals (stdres) for each observed haplotype with values greater than |2| are in bold.

2/ to analyze worldwide *HLA-E, -F* and *-G* allelic distribution and LD. Indeed, exome sequence data from the 1000 Genomes Project (Genomes Project et al., 2015) constitute a priceless public resource allowing genetic diversity studies (Olieslagers et al., 2017; Felicio et al., 2014; Castelli et al., 2014).

Our main result is the description of 11 novel *HLA-H* alleles, and thanks to the 1000 Genomes Project associated DNA collection, their confirmation using targeted NGS. We show in particular that the *HLA-H* locus displays an unexpected diversity with 18 s field alleles at worldwide level with frequency above 0.2%, with unequal distribution across the five continents. We further confirm the high LD between *HLA-H* and *-A,* and between *HLA-H* and *-G,* though at a lower extent. However, *HLA-G* has a greater diversity in its regulatory regions (5'URR and 3'UTR), that are structured in conserved haplotypes and in stronger LD with *HLA-H* (Carlini et al., 2016). We also confirm the observation that *HLA-G* second field allelic distribution, like that of *HLA-H* and *HLA-A,* is different from one part of the globe to another. Conversely, *HLA-E* and *HLA-F* both show little, if any, LD with their loci neighbors, display very restricted allelic diversity and little difference in their worldwide distribution.

Phylogenetic analysis and frequency distribution of *HLA-H* alleles described in the IPD-IMGT /HLA Database and newly identified here, revealed three clades, each predominantly represented in Admixed American, European and East Asian populations (Clade containing *H*01:01*), African populations (Clade containing *H*02:05*) and South Asian populations (Clade containing *H*02:07*), suggesting a strong genetic drift. The genetic deletion encompassing *HLA-H* could not be included in phylogenetic analysis but was present in at least 10% of every population and above 20% in South Asian populations. Due to this deletion of the *HLA-H* locus and to the fact that most *HLA-H* alleles described so far encode truncated proteins, *HLA-H* is considered to be a pseudogene. Some studies however, linked its transcriptional expression with inter-individual variations in immune responsiveness (Aka et al., 2016; Yucesoy et al., 2013; Qin et al., 2017), that might imply an activity for this locus. *HLA-H* allelic diversity and worldwide distribution could also be due to hitchhiking with the *HLA-A* locus (Kulski et al., 2011), reflected by their high LD. The observation of the new alleles *H*02:07* and *H*02:14* that possibly encode a protein with all the patterns of a transmembrane HLA protein, opens up new perspectives for this locus. The fact that these 'complete' HLA-H sequences display three out of four cysteines (lacking the one at codon 164) that are important antigen presenting function (Zemmour et al., 1990) whereas HLA-E, -F and -G displayed all four of them remains to be investigated, as this cysteine was not implicated in the disulfide bond predicted in all of the HLA protein sequences included in this study.

Our results on HLA Ib allelic frequencies and two-loci haplotypes estimation on worldwide populations further supports the LD between *HLA-A, -H* and *-G.* Although their worldwide distribution displays considerable differences, their strong LD is observed in all populations investigated. Conversely, *HLA-E* and *HLA-F,* whose genetic diversity was lower at second field resolution, displayed no LD. This difference could be an echo of their specific expression and function. Indeed, whereas HLA-G is expressed in a tissue-specific manner with documented inter-individual variations and coding polymorphisms (Rebmann et al., 2014; Moreau et al., 2009; Ferreira et al., 2016), HLA-E and HLA-F appear to be mobilized at the cell surface upon activation (Allan et al., 2002; Burian et al., 2016; Garcia-Beltran et al., 2016; Lauterbach et al., 2015), and HLA-E has one high and one low expressing allele. The apparent shared function of HLA Ib, *i.e.* inhibition of immune activation, might thus have evolved and be regulated by two independent pathways, one driven by a population genetics drift and adaptation to specific environments, and the other led by cellular compensation mechanisms.

Furthermore, HLA Ib two-loci haplotypes may be helpful in clinical fields as the involvement of HLA-G and HLA-E, both at genetic polymorphism and expression levels, has been well described in solid organ transplant and grafts. Given that HLA-A typing of recipients and donors is systematically and routinely performed in immunogenetics laboratories, our results on LD results could potentially improve transplant matching strategies.

5. Conclusions

In conclusion, our work reveals an unexpected *HLA-H* genetic diversity with alleles highly represented in Asia potentially encoding a functional HLA protein. This study also contributes to better define worldwide *HLA-E, -F* and *-G* genetic diversity, distribution and LD. Functional implication of these results remains to be explored, both in physiological and pathological contexts.

Formatting of funding sources

The authors of this manuscript have no conflicts of interest to disclose.

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Acknowledgements

The authors thank Justine Buand for providing help with language.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.molimm.2019.04.017.

References

Abi-Rached, L., et al., 2018. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS One 13 (10), e0206512.
Adams, E.J., et al., 2000. Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51 (6), 410–424.
Adams, E.J., Cooper, S., Parham, P., 2001. A novel, nonclassical MHC class I molecule specific to the common chimpanzee. J. Immunol. 167 (7), 3858–3869.
Aka, J.A., Calvo, E.L., Lin, S.X., 2016. Genomic data on breast cancer transcript profile modulation by 17beta-hydroxysteroid dehydrogenase type 1 and 17-beta-estradiol. Data Brief 9, 1000–1012.
Allan, D.S., et al., 2002. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. J. Immunol. Methods 268 (1), 43–50.
Artimo, P., et al., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40 (Web Server issue), W597–603.
Boyle, L.H., et al., 2006. Selective export of HLA-F by its cytoplasmic tail. J. Immunol. 176 (11), 6464–6472.
Burian, A., et al., 2016. HLA-F and MHC-I open conformers bind natural killer cell Ig-like receptor KIR3DS1. PLoS One 11 (9), e0163297.
Buttura, R.V., et al., 2019. HLA-F displays highly divergent and frequent haplotype lineages associated with different mRNA expression levels. Hum. Immunol. 80 (2), 112–119.
Carlini, F., et al., 2013. HLA-G UTR haplotype conservation in the Malian population: association with soluble HLA-G. PLoS One 8 (12), e82517.
Carlini, F., et al., 2016. Association of HLA-a and Non-classical HLA class I alleles. PLoS One 11 (10), e0163570.
Carlini, F., et al., 2017. Bronchial epithelial cells from asthmatic patients display less functional HLA-G isoform expression. Front. Immunol. 8, 6.
Castelli, E.C., Mendes-Junior, C.T., Donadi, E.A., 2007. HLA-G alleles and HLA-G 14 bp polymorphisms in a Brazilian population. Tissue Antigens 70 (1), 62–68. https://doi.org/10.1111/j.1399-0039.2007.00855.x.
Castelli, E.C., et al., 2014. Insights into HLA-G genetics provided by worldwide haplotype diversity. Front. Immunol. 5, 476.
Castelli, E.C., et al., 2017. HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographicaly distinct population samples of Brazil and Cyprus. Mol. Immunol. 83, 115–126.
Castro, M.S., et al., 2019. High-resolution characterization of 12 classical and non-classical HLA loci in Southern Brazilians. HLA 93 (2-3), 80–88.
Celik, A.A., et al., 2015. The diversity of the HLA-E-restricted peptide repertoire explains the immunological impact of the Arg107Gly mismatch. Immunogenetics.
Di Cristofaro, J., et al., 2011. Linkage disequilibrium between HLA-G*0104 and HLA-E*0103 alleles in Tswa Pygmies. Tissue Antigens 77 (3), 193–200.
Di Cristofaro, J., et al., 2013. HLA-G haplotype structure shows good conservation between different populations and good correlation with high, normal and low soluble HLA-G expression. Hum. Immunol. 74 (2), 203–206.

el Kahloun, A., et al., 1992. A continuous restriction map from HLA-E to HLA-F. Structural comparison between different HLA-A haplotypes. Immunogenetics 35 (3), 183–189.
Felicio, L.P., et al., 2014. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. Tissue Antigens 83 (2), 82–93.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39 (4), 783–791.
Ferreira, L.M., et al., 2016. A distant trophoblast-specific enhancer controls HLA-G expression at the maternal-fetal interface. Proc. Natl. Acad. Sci. U. S. A. 113 (19), 5364–5369.
Foroni, I., et al., 2014. HLA-E, HLA-F and HLA-G—The Non-Classical Side of the MHC Cluster, in HLA and Associated Important Diseases. InTech.
Garcia-Beltran, W.F., et al., 2016. Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. Nat. Immunol. 17 (9), 1067–1074.
Genomes Project, C., et al., 2015. A global reference for human genetic variation. Nature 526 (7571), 68–74.
Geraghty, D.E., et al., 1992a. The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. J. Immunol. 149 (6), 1934–1946.
Geraghty, D.E., et al., 1992b. Cloning and physical mapping of the HLA class I region spanning the HLA-E-to-HLA-F interval by using yeast artificial chromosomes. Proc. Natl. Acad. Sci. U. S. A. 89 (7), 2669–2673.
Gleimer, M., et al., 2011. Although divergent in residues of the peptide binding site, conserved chimpanzee Patr-AL and polymorphic human HLA-A*02 have overlapping peptide-binding repertoires. J. Immunol. 186 (3), 1575–1588.
Goodridge, J.P., et al., 2013a. HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors. J. Immunol. 191 (7), 3553–3562.
Goodridge, J.P., et al., 2013b. HLA-F and MHC-I open conformers cooperate in a MHC-I antigen cross-presentation pathway. J. Immunol. 191 (4), 1567–1577.
Grimsley, C., Ober, C., 1997. Population genetic studies of HLA-E: evidence for selection. Hum. Immunol. 52 (1), 33–40.
Hans, J.B., Bergl, R.A., Vigilant, L., 2017. Gorilla MHC class I gene and sequence variation in a comparative context. Immunogenetics 69 (5), 303–323.
Heinrichs, H., Orr, H.T., 1990. HLA non-a,B,C class I genes: their structure and expression. Immunol. Res. 9 (4), 265–274.
Howangyin, K.Y., et al., 2012. Multimeric structures of HLA-G isoforms function through differential binding to LILRB receptors. Cell. Mol. Life Sci.
Hulo, N., et al., 2008. The 20 years of PROSITE. Nucleic Acids Res. 36 (Database issue), D245–9.
Hviid, T.V., Christiansen, O.B., 2005. Linkage disequilibrium between human leukocyte antigen (HLA) class II and HLA-G–possible implications for human reproduction and autoimmune disease. Hum. Immunol. 66 (6), 688–699.
Kall, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338 (5), 1027–1036.
Kolte, A.M., et al., 2010. Study of the structure and impact of human leukocyte antigen (HLA)-G-A, HLA-G-B, and HLA-G-DRB1 haplotypes in families with recurrent miscarriage. Hum. Immunol. 71 (5), 482–488.
Kraemer, T., Blasczyk, R., Bade-Doeding, C., 2014. HLA-E: a novel player for histocompatibility. J. Immunol. Res. 2014, 352160.
Kulski, J.K., Shigenari, A., Inoko, H., 2011. Genetic variation and hitchhiking between structurally polymorphic Alu insertions and HLA-A, -B, and -C alleles and other retroelements within the MHC class I region. Tissue Antigens 78 (5), 359–377.
Kumar, S., et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547–1549.
Langmead, B., et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (3), R25.
Lauterbach, N., et al., 2015. Peptide-induced HLA-E expression in human PBMCs is dependent on peptide sequence and the HLA-E genotype. Tissue Antigens 85 (4), 242–251.
Lawlor, D.A., et al., 1991. Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. J. Exp. Med. 174 (6), 1491–1509.
Lee, N., Ishitani, A., Geraghty, D.E., 2010. HLA-F is a surface marker on activated lymphocytes. Eur. J. Immunol. 40 (8), 2308–2318.
Lima, T.H.A., et al., 2016. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. Hum. Immunol. 77 (10), 841–853.
Lynge Nilsson, L., Djurisic, S., Hviid, T.V., 2014. Controlling the immunological crosstalk during conception and pregnancy: HLA-G in reproduction. Front. Immunol. 5, 198.
Malissen, M., Malissen, B., Jordan, B.R., 1982. Exon/intron organization and complete nucleotide sequence of an HLA gene. Proc. Natl. Acad. Sci. U. S. A. 79 (3), 893–897.
Marsh, S.G., et al., 2010. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. pp. 291–455.
Messer, G., et al., 1992. HLA-J, A second inactivated class I HLA gene related to HLA-G and HLA-A. Implications for the evolution of the HLA-A-related genes. J. Immunol. 148 (12), 4043–4053.
Moreau, P., Flajollet, S., Carosella, E.D., 2009. Non-classical transcriptional regulation of HLA-G: an update. J. Cell Mol. Med. 2973–2989. https://doi.org/10.1111/j.1582-4934.2009.00800.x. England.
Nunes, J.M., 2014. Using uniformat and gene[rate] to analyse data with ambiguities in population genetics. Figshare. https://doi.org/10.6084/m9.figshare.984299.
Nunes, J.M., et al., 2014. The HLA-net GENE[RATE] pipeline for effective HLA data analysis and its application to 145 population samples from Europe and neighbouring areas. Tissue Antigens 83 (5), 307–323.
Olieslagers, T.I., et al., 2017. New insights in HLA-E polymorphism by refined analysis of the full-length gene. HLA 89 (3), 143–149.
Oliveira, M.L.G., et al., 2018. Extended HLA-G genetic diversity and ancestry composition in a Brazilian admixed population sample: implications for HLA-G transcriptional control and for case-control association studies. Hum. Immunol. 79 (11), 790–799.
Pabon, M.A., et al., 2014. Impact of human leukocyte antigen molecules e, f, and g on the outcome of transplantation. Transplant Proc. 46 (9), 2957–2965.
Parham, P., 1993. HLA, anthropology, and transplantation. Transplant Proc. 25 (1 Pt 1), 159–161.
Pratheek, B.M., et al., 2014. Mammalian non-classical major histocompatibility complex I and its receptors: important contexts of gene, evolution, and immunity. Indian J. Hum. Genet. 20 (2), 129–141.
Qin, N., et al., 2017. Fine-mapping the MHC region in Asian populations identified novel variants modifying susceptibility to lung cancer. Lung Cancer 112, 169–175.
Ramalho, J., et al., 2017. HLA-E regulatory and coding region variability and haplotypes in a Brazilian population sample. Mol. Immunol. 91, 173–184.
Rebmann, V., et al., 2014. HLA-G as a tolerogenic molecule in transplantation and pregnancy. J. Immunol. Res. 2014, 297073.
Robinson, J., et al., 2015. The IPD and IPD-IMGT/HLA database: allele variant databases. Nucleic Acids Res. D423–D431.
Rouas-Freiss, N., et al., 1997. Direct evidence to support the role of HLA-G in protecting the fetus from maternal uterine natural killer cytolysis. Proc. Natl. Acad. Sci. U. S. A. 94 (21), 11520–11525.
Shukla, H., et al., 1991. A class I jumping clone places the HLA-G gene approximately 100 kilobases from HLA-H within the HLA-A subregion of the human MHC. Genomics 10 (4), 905–914.
Sonon, P., et al., 2018. HLA-G, -E and -F regulatory and coding region variability and haplotypes in the Beninese Toffin population sample. Mol. Immunol. 104, 108–127.
Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. 101 (30), 11030–11035.
UniProt Consortium, T., 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 46 (5), 2699.
Yucesoy, B., et al., 2013. Genetic variants within the MHC region are associated with immune responsiveness to childhood vaccinations. Vaccine 31 (46), 5381–5391.
Zemmour, J., et al., 1990. HLA-AR, an inactivated antigen-presenting locus related to HLA-A. Implications for the evolution of the MHC. J. Immunol. 144 (9), 3619–3629.