WILEY

HLA
Immune Response Genetics

ORIGINAL ARTICLE OPEN ACCESS

Extensive Analysis of Genetic Diversity in HLA-DMA, HLA-DMB, HLA-DOA and HLA-DOB: Characterisation of 236 Novel Alleles

Viviane Albrecht[1] | Madlen Pahlke[1] | Jürgen Sauter[2] | Christin Paech[1] | Kathrin Putke[1] | Alexander H. Schmidt[1,2] | Vinzenz Lange[1] | Anja Klussmeier[1]

[1]DKMS Life Science Lab, Dresden, Germany | [2]DKMS Group, Tübingen, Germany

Correspondence: Anja Klussmeier (klussmeier@dkms-lab.de)

Received: 4 December 2024 | **Revised:** 19 March 2025 | **Accepted:** 25 April 2025

Keywords: allele | genotype | haplotype | HLA | HLA-DM | HLA-DO | MHC | novel allele

ABSTRACT

HLA-DMA, *-DMB*, *-DOA* and *-DOB* are non-classical HLA Class II genes that play a crucial role in the selection of highly stable HLA Class II/peptide complexes on antigen-presenting cells. Although the genes were initially thought to have a limited diversity with less than 13 alleles per gene documented in the IPD-IMGT/HLA Database in 2022, recent studies suggest a potential impact of certain alleles on the outcome of hematopoietic cell transplantation. To gain a deeper understanding of allelic diversity, we sequenced *HLA-DMA*, *-DMB*, *-DOA* and *-DOB* of 1880 potential stem cell donors from Germany, Poland, Great Britain and Chile, achieving full-gene resolution. Remarkably, we identified 3968 previously undescribed sequences, including 28 distinct novel proteins. The observed allele frequencies were consistent across all studied populations with one dominating protein for each gene: *HLA-DMA*01:01* (> 77%), *HLA-DMB*01:01* (> 63%), *HLA-DOA*01:01* (> 97%) and *HLA-DOB*01:01* (> 77%). Notably, a much higher diversity was observed in full-genomic resolution. Finally, we submitted 51 distinct novel sequences for *HLA-DMA*, 58 for *HLA-DMB*, 80 for *HLA-DOA* and 47 for *HLA-DOB* to the IPD-IMGT/HLA Database. This comprehensive reference database update will not only simplify future genotyping of *HLA-DMA*, *-DMB*, *-DOA* and *-DOB* but will hopefully also enhance our understanding of the complex process of peptide selection and loading to the HLA Class II proteins.

1 | Introduction

HLA-DM and HLA-DO (from here on DM and DO) are HLA Class II-like heterodimers, encoded by the genes *HLA-DMA* and *HLA-DMB* (from here on *DMA* and *DMB*), and *HLA-DOA* and *HLA-DOB* (from here on *DOA* and *DOB*), respectively [1, 2]. While they do not present peptides, they serve as accessory molecules for optimal peptide loading of the classical HLA Class II proteins. DM is ubiquitously expressed in antigen-presenting cells (APCs) and necessary for the efficient displacement of the Class II-associated invariant chain peptide (CLIP) from the peptide binding groove. Likewise, non-optimal peptides are displaced from HLA Class II molecules, thereby selecting immunodominant (DM-resistant) epitopes [3]. In contrast, expression of DO is restricted to B cells, thymic medullary epithelial cells, and subpopulations of dendritic cells [4, 5]. As of today, multiple, partly conflicting models for the modulatory function of DO have been proposed [6]. On one hand, competitive inhibition of DM by DO has been proposed to result in a broader repertoire of (DM-sensitive) epitopes [7–10]. On the other hand, DO has been reported to stabilise the open binding groove of HLA Class II proteins, thereby increasing the stringency of epitope selection by DM [11, 12].

Even though the *DMA*, *DMB*, *DOA* and *DOB* genes are located within the MHC complex between the highly diverse *HLA-DQB1* and *HLA-DPB1* genes, knowledge about DM/DO diversity and its functional impact is sparse [13–15]. The

heterodimer *DMA*01:03/DMB*01:07* has been predicted to be a less efficient peptide-exchange catalyst than *DMA*01:01/DMB*01:01* [16]. Similarly, some *DOA* and *DOB* variants show altered efficiency in vitro in comparison to *DOA*01:01* and *DOB*01:01*. Among them are *DOA*01:02* and *DOA*01:03* with increased efficiency [14, 15]. In vivo, a recent study on haploidentical hematopoietic cell transplantations (HCTs) associated *DMA*01:03* with a higher risk of relapse for the patient [17]. Additionally, DM impacts the permissiveness of *HLA-DPB1* mismatches in HCT. *HLA-DPB1* alleles of the same T-cell receptor group usually present peptides of limited divergence. However, in the absence of DM in APCs in vitro, the divergence of peptides was as high as for nonpermissive *HLA-DPB1* mismatches [18, 19].

With these first indications that the HLA Class II peptide loading machinery might have a clinical impact in at least some HCT settings, the question arose if the incorporation of DM/DO genotyping in donor selection could be beneficial for the patient. Naturally, that would only be warranted if there was a relevant allelic variation in the population. However, due to limited previous studies, it was presumed that the described alleles in the IPD-IMGT/HLA Database, the global reference for donor HLA genotyping, might only touch the surface of the allelic diversity [13]. Consequently, the aim of this study was to gain additional insights into the diversity and allele frequencies of *DMA*, *DMB*, *DOA* and *DOB* in different populations and submit the encountered novel alleles to the IPD-IMGT/HLA Database to generate a broader and more reasonable database foundation for these four genes.

2 | Methods

2.1 | Samples

A total of 1880 volunteers from Germany (40%), the United Kingdom (30%), Poland (15%), Chile (10%) and the United States (5%) provided samples to DKMS for registration as potential stem cell donors. As part of the registration process, the donors are asked to self-assign their ethnic background. The genotyping is within the scope of the consent forms signed at recruitment.

2.2 | PCR and Sequencing

DNA was isolated from buccal swabs as described before [20, 21]. In general, obtained DNA concentrations ranged from 2 to 100 ng/μL and DNA was used without prior normalisation or fragment length analysis. The genes *DMA*, *DMB*, *DOA* and *DOB* were amplified from 5′ UTR to 3′ UTR (complete exon and intron coverage) by PCR with the following primer pairs: GCTGGTC GCTTACAGACTGAG and GAAAAACATCATGTACCGCC AAGAC (*DMA*); CTGAACTCCCGGCATCTTTACAG and GAG AGGCATGGTAGCATCATTGAG (*DMB*); GTGCTCTGACGGC CTTTTCTC and CATCATGGCAGGATACCAGTGG (*DOA*); GG AGTTTCCAATCCTGGGGAAG and CCAGAGAGAGAAGCT TCAGTGAGG (*DOB*) (metabion, Planegg, Germany). Overhangs were attached at primer 5′ ends to provide binding sites for subsequent indexing PCR [22]. Primers for *DMA* and *DMB* (125 nM each) or primers for *DOA* and *DOB* (25 and 50 nM) were used in two PCR reactions containing genomic DNA (range: 2–100 ng), dNTPs (0.4 mM each) (Roche Diagnostics, Mannheim, Germany), 1× Advantage Genomic LA Buffer, and 0.5 U Advantage Genomic LA Polymerase Mix (Takara Bio, Mountain View, CA, USA) in a 10 μL reaction. PCR conditions: 94°C 3 min, 30 cycles: 98°C 10 s/57°C 20 s/66°C 8 min, 72°C 10 min. Amplicon sizes were expected between 3.6 kb (*DOA*) and 6 kb (*DMB*). PCR products were diluted 1:20, and 1 μL was used for a 10-cycle indexing PCR with custom barcodes [22] using the same reaction conditions as above. PCR success was checked by agarose gel electrophoresis. Failed PCR reactions were not repeated.

Barcoded PCR products were pooled and purified using 0.7× SPRIselect beads (Beckman Coulter, Brea, CA, USA). Library preparation was performed with SQK-LSK110 kits. Sequencing was performed on a GridION instrument (7× MinION flowcells 10.3, basecalling: Guppy R10.3 HAC) according to the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK) (Figure 1). Up to 570 samples were loaded on one MinION flowcell.

2.3 | Characterisation and Submission of Novel Alleles

Reads from ONT sequencing were used to identify samples with novel alleles (NGSengine (GenDx, Utrecht, The Netherlands), IPD-IMGT/HLA Database version 3.44). Since novel alleles for IPD-IMGT/HLA Database submission have to be confirmed by a second PCR, the respective genes were subjected to additional Illumina shotgun sequencing. Samples for Illumina sequencing were selected to include all sequences with novel exon variations and frequent intron variations. If possible, three independent samples with an identical novel sequence were used. In addition to these prioritised samples, 96-well plates were filled up with novel sequences that were not prioritised (e.g., intron variations that were detected only once). Finally, 380 samples were subjected to Illumina sequencing (Figure 1). Subsequent to another PCR reaction as described above, fragmentation and adapter ligation were performed according to 'NEBNext Ultra II DNA Library Prep Kit for Illumina' protocol (New England Biolabs, Ipswich, MA, USA). Amplicons were purified with 0.7× SPRIselect beads, and custom barcodes were attached by a seven-cycle-indexing PCR. Finally, 47 or 48 samples were pooled (one PCR positive control per 96-well plate is not sequenced) and subsequently purified using 0.7× SPRIselect beads. After qPCR-based library quantification, pools (190 samples each) were sequenced on a MiSeq instrument (MiSeq Reagent Kit v2 2 × 250) according to the manufacturer's instructions (Illumina, San Diego, CA, USA).

Sequencing reads obtained from ONT and Illumina sequencing were first analysed independently using NGSengine, IPD-IMGT/HLA Database version 3.48 (Figure 1). This release was the current one at the time of sequence analysis and, for DM and DO genes, did not differ from version 3.44, which was used for novel allele identification. The mean read length for ONT reads was 4734 bp (*DMA*), 5434 (*DMB*), 3389 (*DOA*) and

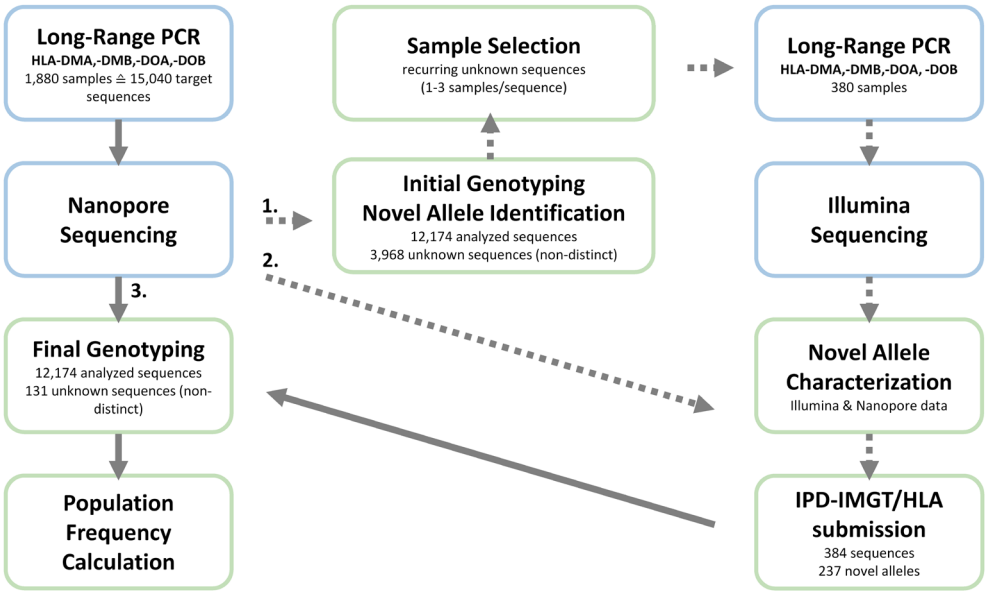

FIGURE 1 | Workflow for DMA, DMB, DOA and DOB genotyping and novel allele characterisation. Nanopore sequencing reads were analysed multiple times. First, to identify samples with novel sequences. Selected samples were then subjected to a second long-range PCR and Illumina sequencing. Second, to characterise the novel alleles by joint analysis of Illumina and Nanopore sequencing reads. Third, for the final genotyping after the novel alleles were named by the IPD-IMGT/HLA Database. Blue boxes depict wet lab processes, and green boxes depict analysis steps. Filled arrows indicate the final genotyping workflow, while dashed arrows indicate the novel allele characterisation workflow.

3648 (*DOB*). As expected, the median base quality score of mapped ONT reads was lower than that of Illumina reads (21 vs. 37). Nevertheless, for most samples, both sequencing methods generated identical and fully phased sequences. These were approved. For samples with questionable or inconsistent results (all *DMB* sequences because of their long homopolymeric region and three sequences that contained insertions and/or deletions (1× *DOA*, 2× *DOB*)), dual redundant reference sequencing (DR2S) was used. Its algorithm combines short and long reads to generate the highest quality error corrected consensus sequences [22–24]. Sequences that were still ambiguous were rejected ($n = 45$). Finally, all approved sequences with novel alleles were submitted to the IPD-IMGT/HLA Database using TypeLoader2 [25, 26]. If possible, a second, independent sample was submitted for each novel allele to confirm the sequence. Additionally, sequence extensions were submitted for existing alleles with only partial sequence information in the database.

Please note that some novel alleles were identified but not characterised and submitted to the IPD-IMGT/HLA Database, mostly because of prioritisation of our resources on exon and/or frequent variations. Furthermore, some sequences could not be characterised because of technical reasons (e.g., PCR failure).

2.4 | Genotyping

After the novel alleles were named by the IPD-IMGT/HLA Database, final genotyping of the *DMA*, *DMB*, *DOA* and *DOB* data sets was performed with NGSengine, IPD-IMGT/HLA Database version 3.51, based on the original Oxford Nanopore sequencing data (Figure 1). Primer regions were ignored for the genotyping calls. The chosen phasing algorithm was 'cluster' with parameters 50 (coverage), 30 (average second base for heterozygous calls) and 20 (noise). If necessary, error-prone bases (e.g., the long T-homopolymer in *DMB*) were edited manually. Before manuscript submission, all leftover novel alleles were rechecked against the newest version of the IPD-IMGT/HLA Database (3.58). Due to the described sequence variations in the UTRs outside of our primer binding sites, the following ambiguities could not be resolved: *DMA*01:01:01:01/01:01:01:41*, *DMB*01:01:01:01/01:01:01:03/01:01:01:46*, *DMB*01:01:01:02/01:01:01:51*, *DMB*01:03:01:02/01:03:01:04/01:03:01:19* and *DOB*01:01:03:02/01:01:03:10*.

2.5 | Haplotypes

Using Hapl-o-Mat, our open-source implementation of an expectation–maximisation (EM) algorithm, haplotypes and their respective frequencies were estimated from unphased genotype data [27, 28]. Since the algorithm operates with the prerequisite of a sufficiently large cohort of a population in Hardy–Weinberg equilibrium, we restricted our analysis to samples with German ethnic backgrounds ($n = 426$). To estimate eight-locus haplotypes (*DMA*, *DMB*, *DOA*, *DOB*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1* and *HLA-DQB1*), we included additional genotyping data that was generated as described before [29, 30]. To exclude the issue of residual typing ambiguities, haplotype resolution for *DMA*, *DMB*, *DOA* and *DOB* was reduced to three fields. The remaining four loci have been considered on 'small-g' resolution. This approach groups alleles with identical and synonymous exon 2 and 3 sequences, including null alleles [31]. To more closely investigate linkage disequilibrium on smaller haplotypes, we also estimated frequencies for subsets with fewer loci.

3 | Results

3.1 | Novel Alleles

Out of 1880 samples, 1637 (87%), 1493 (79%), 1482 (79%) and 1475 (78%) samples were successfully sequenced and analysed for *DMA*, *DMB*, *DOA* and *DOB*, respectively (average sequencing depth: 2068×, 895×, 876× and 1122×). Sample dropouts were predominantly caused by PCR failure, presumably due to insufficient DNA quality for long-range amplification. However, we cannot rule out that single samples failed because of unknown base variations at the primer binding sites. Strikingly, we identified 3968 non-distinct unknown sequences in 1498 samples (*DMA*: 665; *DMB*: 1139; *DOA*: 1336 and *DOB*: 828). Since many of these unknown sequences were encountered multiple times, we identified 311 distinct novel sequences. Among them were 5 novel DMA proteins, 2 novel DMB proteins, 11 novel DOA proteins and 10 novel DOB proteins (Table 1, Figure 2A, Supporting Information 1). These high numbers confirmed our hypothesis that the genomic diversity in the IPD-IMGT/HLA Database has so far been severely underestimated. To improve this situation, we submitted 236 distinct novel sequences to the IPD-IMGT/HLA Database (51 *DMA*, 58 *DMB*, 80 *DOA* and 47 *DOB*). Specifically, we submitted sequences of 5 distinct novel proteins, 1 synonymous exon variation and 45 intron variations of *DMA*. For *DMB*, we submitted 2 novel proteins, 1 synonymous exon variation and 55 intron variations. Similarly, for *DOA*, we submitted 11 novel proteins, 6 synonymous exon variations and 63 intron variations, and for *DOB*, 10 novel proteins, 5 synonymous exon variations and 32 intron variations (Figure 2A). Additionally, we characterised and submitted 148 sequence extensions and confirmations (24 *DMA*, 50 *DMB*, 42 *DOA* and 32 *DOB*) (Figure 2A, Supporting Information 1). These numbers include sequences with which we confirmed our own data (e.g., a second sample of a novel allele) as well as confirmations and sequence extensions of previously submitted sequences from other laboratories. Overall, the number of described alleles increased tremendously from the IPD-IMGT/HLA Database release 3.48 (April 2022) to 3.58 (October 2024): for *DMA* from 7 known alleles to 60, for *DMB* from 13 to 85, for *DOA* from 12 to 113 and for *DOB* from 13 to 71 (Figure 2B).

Since we focused our resources on the characterisation of novel proteins and frequent sequence variations and experienced some PCR failures and/or insufficient sequencing quality, we did not submit every novel allele that we encountered to the IPD-IMGT/HLA Database.

3.2 | Allele Frequencies

Allele frequencies of *DMA*, *DMB*, *DOA* and *DOB* were calculated for the four largest populations in our data set: Germans ($n = 637$), British/Irish ($n = 379$), Poles ($n = 282$) and non-indigenous Chileans ($n = 138$). Overall, the allele frequencies in the four populations were consistent (Figure 3, Supporting Information 2). In our largest cohort, the German population, one major protein (*01:01) with frequencies above 74% was identified in all four genes. In detail, DMA*01:01 has a frequency of 83%, DMA*01:02 of 11% and DMA*01:03 of 4%. DMB*01:01 (75%) is followed by DMB*01:03 (18%) and DMB*01:07 (4%). For DOA, there is only one major protein, DOA*01:01 with 97% allele frequency, followed by DOA*01:02 (2%). DOB*01:01 has 77% allele frequency in the German population, followed by DOB*01:04 (9%), DOB*01:02 (6%), DOB*01:03 (4%) and DOB*01:05 (4%). Albeit still small, the largest differences between the populations were observed in DMB. DMB*01:01 frequencies are higher in Germany and Poland (75% and 82%) compared to GB/Ireland and non-indigenous Chileans (64% and 65%). In contrast, the British/Irish population and non-indigenous Chileans have higher frequencies of DMB*01:03 (26% and 25%) compared to Germans and Poles (18% and 9%). The same is true for DMB*01:04 (3% and 4% compared to 1%) (Figure 3, Supporting Information 2).

In *DMA* and *DMB*, almost no synonymous exon variations (three-field resolution) could be identified (Figure 4). In contrast, the major DOA protein DOA*01:01 (97%) is subdivided into multiple sequences, among them *DOA*01:01:02* (55%), *DOA*01:01:01* (22%), *DOA*01:01:04* (13%), *DOA*01:01:03* (4%) and *DOA*01:01:05* (2%). Likewise, DOB demonstrates some variability at the third field level. By zooming even further into the full-gene resolution (four-field resolution), it becomes evident that the genomic allelic diversity of *DMA*, *DMB*, *DOA*, and *DOB* is actually quite high (Figure 4). Especially, the four major proteins can be subdivided into multiple distinct genomic sequences that were often detected with allele frequencies greater than 1% in all studied populations (Figure 4, Supporting Information 2).

In our data set, we could identify all proteins from IPD-IMGT/HLA Database release 3.58 with the exception of DMB*01:05 [32], DMB*01:06 [33], DOA*01:03, DOA*01:04N [34], DOA*01:16, DOA*01:17, DOB*01:17 and DOB*01:18N. Some of these alleles were originally characterised from samples of Asian heritage, ethnicities that are underrepresented in this study. DOA*01:16, DOA*01:17 and DOB*01:17 were more recently identified in samples from Europe [13]. Since we could not identify them in our predominantly European cohort, we assume that these are rare alleles.

3.3 | Diversity Coverage in the IPD-IMGT/HLA Database

Even though we genotyped over 1500 samples and characterised all novel proteins, the question remained if we caught the most frequent ones of the investigated populations. To get a rough estimation, we compared available SNP data from the gnomAD database (version 4.1.0; 730,947 exomes and 76,215 genomes) [35]. In its current version, 297, 317, 408 and 383 protein variations for DMA, DMB, DOA and DOB are included, respectively (Supporting Information 3). Naturally, these data do not contain phasing information and do not meet the quality criteria for novel allele characterisation, but they might be a rough estimate of how many proteins could be found in the global population. According to gnomAD, the most frequent and still unnamed (IPD-IMGT/HLA Database version 3.58) protein variations are DMA V209M (0.04% global, 0.5% in African Genetic Ancestry Group), DMB M196V (0.08% global, 1% in South Asian Genetic Ancestry Group), DOA V155M (0.01% global, 0.2% in African Genetic Ancestry Group)

TABLE 1 | Characterised and submitted novel alleles of *DMA*, *DMB*, *DOA* and *DOB*, leading to altered protein sequences.

Gene	Allele name	AA position (mature protein)	Reference allele *01:01 > Novel AA	Exon	Samples in data set	Ethnic group(s)
HLA-DMA	*DMA*01:05*	149	F > L	3	1	GB/Ireland
	*DMA*01:06*	80	E > K	2	1	Poland
	*DMA*01:07*	30	S > *N*	2	1	GB/Ireland
	*DMA*01:08*	146	G > E	3	1	Unknown
	*DMA*01:09*	220	V > M	4	1	Germany
HLA-DMB	*DMB*01:08*	187	*R* > Q	3	1	Poland
	*DMB*01:09*	48	F > S	2	9	Europe, Chile
HLA-DOA	*DOA*01:05*	212	V > I	4	1	Germany
	*DOA*01:06*	206	G > V	4	4	Europe
	*DOA*01:07*	124	*R* > C	3	1	Germany
	*DOA*01:08*	114	F > L	3	6	Europe, Chile
	*DOA*01:09*	54	*R* > H	2	5	Europe
	*DOA*01:10*	−20	G > *R*	1	4	India
	*DOA*01:11*	126	G > *R*	3	1	Wales
	*DOA*01:12*	3	A > D	2	1	India
	*DOA*01:13*	189	D > *N*	4	2	Germany
	*DOA*01:14*	150	H > *R*	3	1	Chile
	*DOA*01:15*	147	*R* > C	3	1	Philippines
HLA-DOB	*DOB*01:06*	203	I > T	4	1	GB/Ireland
	*DOB*01:07*	60	Q > K	2	8	Europe, Ukraine, Kosovo
	*DOB*01:08*	93	*R* > K	2	4	Germany, Chile
		208	L > F	4		
	*DOB*01:09*	26	F > S	2	1	Poland
	*DOB*01:10*	−9	*R* > Q	1	1	Poland
		208	L > I	4		
	*DOB*01:11*	57	D > Y	2	1	Russia
		218	V > I	4		
	*DOB*01:12*	208	L > F	4	1	GB/Ireland
		235	*N* > S	5		
	*DOB*01:13*	236	E > K	5	1	Unknown
	*DOB*01:14*	−6	S > T	1	1	Germany
		208	L > F	4		
	*DOB*01:15*	150	*N* > S	3	1	Poland

Note: The amino acid changes, their positions (in the mature protein) and the respective exons in the gene are listed based on the reference proteins DMA*01:01, DMB*01:01, DOA*01:01 and DOB*01:01. Ethnic group(s) refer to the self-assigned heritage of donors. Here, Europe is used as an abbreviation for multiple samples from German, Polish and/or British/Irish ethnic groups. See Table S1 for a complete list of novel alleles.

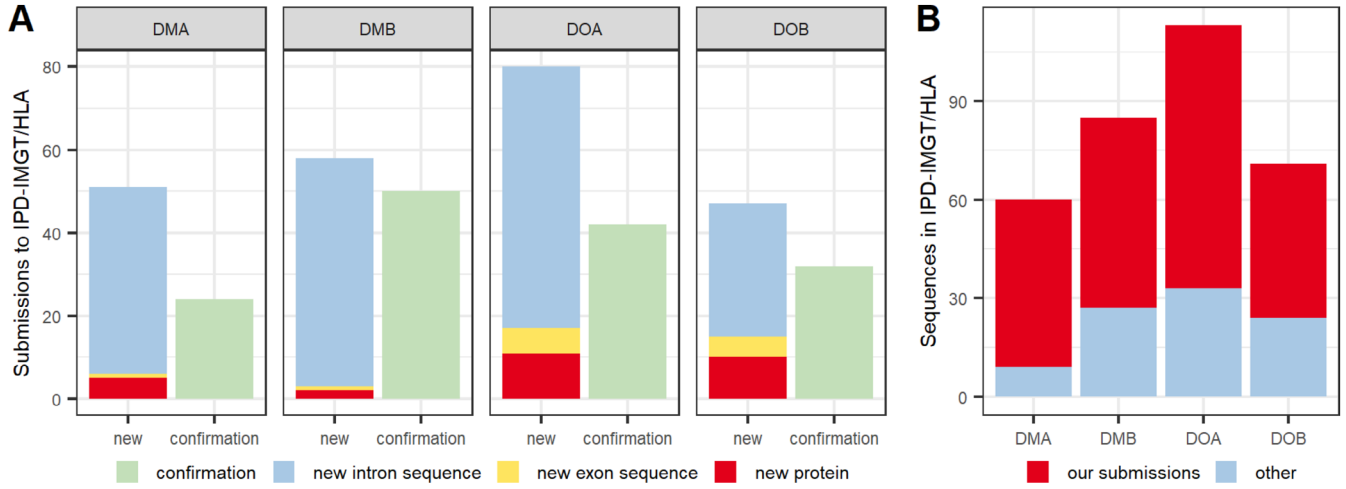

FIGURE 2 | Sequence submissions to the IPD-IMGT/HLA Database. (A) Number of characterised and submitted sequences for *DMA*, *DMB*, *DOA* and *DOB*, respectively. Novel alleles are differentiated into new proteins (red), new synonymous exon variations (yellow), and new intron sequences (blue). In addition, we submitted sequence extensions of only partially described alleles as well as sequence confirmations (summed up in green). (B) Contribution of the submitted sequences (red) to the current IPD-IMGT/HLA Database (release 3.58).

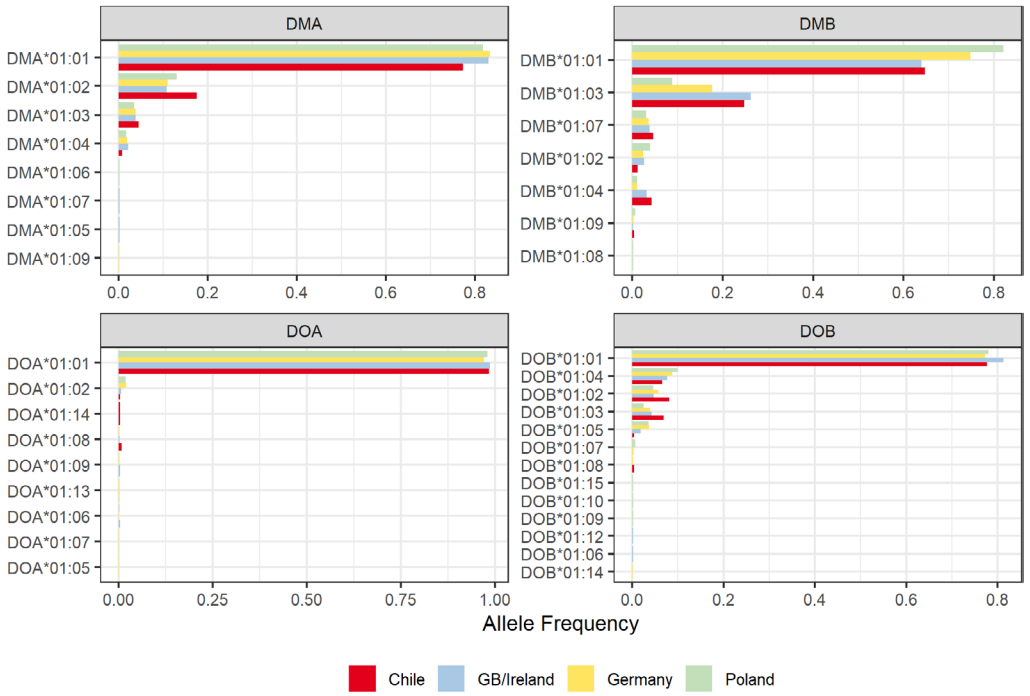

FIGURE 3 | Frequencies of *DMA*, *DMB*, *DOA* and *DOB* alleles at two-field resolution (protein level) in samples from non-indigenous Chilean, British/Irish, German and Polish ethnic groups.

and DOB G226R (0.03% global, 0.03% in European Genetic Ancestry Group) (Supporting Information 3). Reference alleles for these variant calls are DMA*01:01, DMB*01:03, DOA*01:01 and DOB*01:01 (GRCh38). From these data, we estimate that IPD-IMGT/HLA-listed alleles now cover over 99% of the cumulative allele frequency and lack only minor variants at protein resolution. However, this may not be valid for underrepresented populations in gnomAD. On the other hand, some fully characterised and named variations are extremely rare in gnomAD; for example, variation F148L (in DMA*01:05) with a frequency of 0.0002% and DOB F26S (DOB*01:09) with only 0.0006%. Furthermore, some variations (DOA L74V (DOA*01:03), DOA S40N (DOA*01:16), DOA A3D (DOA*01:12) and DOB S-6T (DOB*01:14)) were only detected once in the whole gnomAD data set (Supporting Information 3).

3.4 | Estimation of Major Haplotypes

Based on our largest cohort (German population), we estimated haplotype frequencies using all samples with successful genotyping of all four loci ($n = 426$). *DOB*01:01:01~DMB*01:01:01~DMA*01:01:01~DOA*01:01:02* was identified

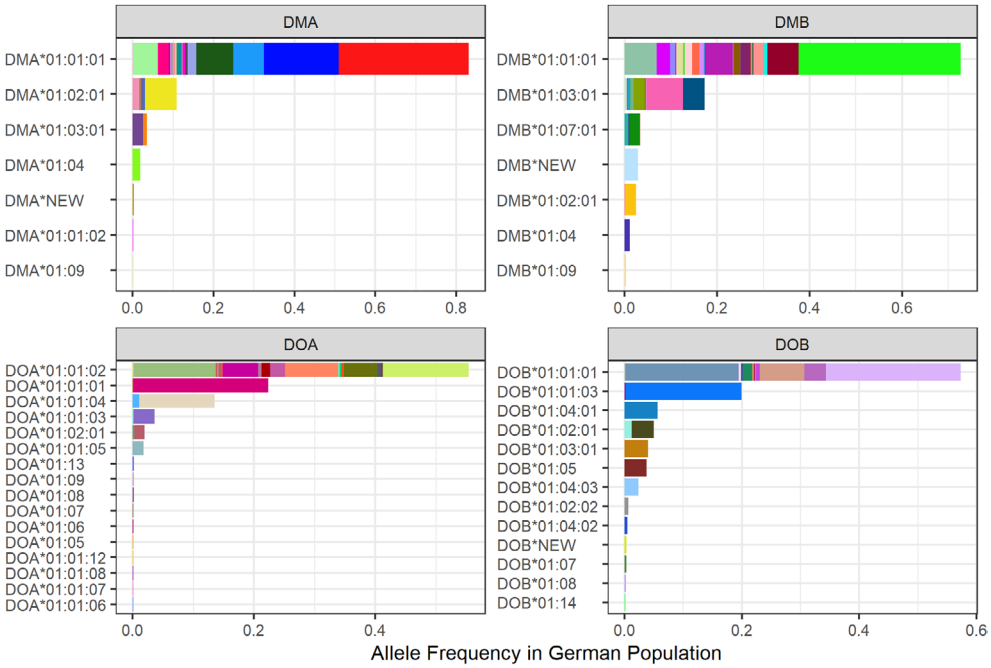

FIGURE 4 | High-resolution frequencies of *DMA*, *DMB*, *DOA* and *DOB* alleles (bars: three-field resolution [exon level]; colours: four-field resolution [full gene]) in the German population. The suffix *NEW is used to sum up all novel alleles that were identified but not characterised and submitted to IPD-IMGT/HLA.

as the major haplotype with a frequency of 16%, followed by *DOB*01:01:03~DMB*01:01:01~DMA*01:01:01~DOA*01:01:02* (12%), *DOB*01:01:01~DMB*01:03:01~DMA*01:01:01~DOA*01:01:02* (9%) and *DOB*01:01:01~DMB*01:01:01~DMA*01:01:01~DOA*01:01:01* (8%). More haplotypes are reported in Supporting Information 4 (only the 60 most frequent are included as appropriate with the limited sample size). The most common extended haplotype in the German population was identified as *DQA1*01:02g~DQB1*06:02g~DOB*01:01:01~DMB*01:03:01~DMA*01:01:01~DOA*01:01:02~DPA1*01:03g~DPB1*04:01g* with a frequency of 4% (Supporting Information 4). Since the *DMB*01:07~DMA*01:03* haplotype might be of further interest for HCT, we investigated if this haplotype could potentially be predicted from standard HLA genotyping. Upon inclusion of the two neighbouring classical HLA genes *HLA-DPB1* and *HLA-DQB1* into the haplotype calculation, the most frequent *DMB*01:07~DMA*01:03* haplotype in the German population is *DQB1*03:01g~DMB*01:07:01~DMA*01:03:01~DPB1*04:01g* (0.8%), followed by *DQB1*05:01g~DMB*01:07:01~DMA*01:03:01~DPB1*04:01g* and *DQB1*03:01g~DMB*01:07:01~DMA*01:03:01~DPB1*04:02g* (0.4% each). However, the same *HLA-DPB1* and *HLA-DQB1* alleles can also be associated with *DMB*01:01~DMA*01:01* haplotypes with even higher frequencies (e.g., *DQB1*03:01g~DMB*01:01:01~DMA*01:01:01~DPB1*04:01g*; 5%). Consequently, with the available allelic resolution for *HLA-DPB1* and *HLA-DQB1*, it does not seem possible to use linkage information to predict DM alleles (Supporting Information 4).

4 | Discussion

While it has long been known that DM and DO are involved in peptide loading to HLA Class II proteins [3–5], most research has been dedicated to the HLA Class II proteins themselves. Yet, recent technological advancements in the field of immunopeptidomics begin to shed a more detailed light on the interplay between specific HLA alleles and the structure of presented peptides [36]. Naturally, the loading of a specific peptide onto a certain HLA Class II molecule might be not only driven by the receptor/peptide combination but also by the peptide loading machinery, for example, DM and/or DO. Indeed, HLA-DPB1 presents a broader peptide repertoire in the absence of DM in vitro [19].

So far, only a few studies looked at the potential impact of allelic differences of DM/DO, one of the reasons surely being that the allelic diversity and allelic frequencies were largely unknown. While IPD-IMGT/HLA Database release 3.48 only contained 3 DOA proteins (and one null allele), available exome sequencing data of 60,000 samples hinted at at least 95 existing missense variations [14]. To the best of our knowledge, our study is the first to investigate the frequencies of DM/DO in a larger set of samples at full genomic resolution. As expected, the diversity of DM and DO proteins is much lower than the diversity of classical HLA proteins. In all studied populations, there is one major protein (*01:01) with allele frequencies over 64% (*DMA*, *DMB* and *DOB*), or even 95% (*DOA*) (Figure 3). Minor differences in the allele frequencies could be observed in a similarly sized study for the Chinese Han population. Here, the two major alleles *DMA*01:01* and *DMB*01:01* have slightly lower allele frequencies: 70% and 53%, in comparison to 83% and 75% in Germany, respectively. Instead, Chinese Hans more often harbour the alleles *DMA*01:02* (28% vs. 11%), *DMB*01:02* (14% vs. 2%) and *DMB*01:03* (32% vs. 18%) [37]. Similar differences might be discovered for populations that are underrepresented in this study, for example, African populations.

There is evidence that DMA*01:03 differs functionally from DMA*01:01. The heterodimer DMA*01:03/DMB*01:01 has been reported to exchange peptides less efficiently than the heterodimer DMA*01:01/DMB*01:01 [16]. Furthermore, in haploidentical HCT, a donor *DMA*01:03* could be associated with a higher risk of relapse [17]. The same effect was observed for *DMB*01:07*, which is in high linkage disequilibrium with *DMA*01:03* (Supporting Information 4). This potential risk haplotype has a frequency of approximately 4% in all studied populations. In DMA*01:04 (allele frequency of 2% in the German population), the same amino acid is modified (position 184 Arg/His in DMA*01:03, Arg/Cys in DMA*01:04). Data from Petersdorf et al. for a donor DMA*01:04 indicate a similar negative effect for patient outcome in haploidentical HCT as DMA*01:03, but the data on relapse rates were not statistically significant. A general DM (mis)match between patient and donor had no effect on outcome or graft-versus-host disease (GVDH) [17]. It remains to be investigated whether donor DMA*01:03 and/or DMA*01:04 would also have an adverse effect in unrelated fully or partially matched allogeneic HCT. In this setting, it might be possible to genotype and exclude unfavourable donors if multiple matches are available.

For DOA and DOB, some amino acid variations with altered DM inhibition were identified in vitro. Among them were DOA*01:02 and DOA*01:03 that showed a gain of function [15]. Other observations affect alleles that were characterised and named in this study: DOB*01:08 and DOB*01:15 with decreased, and DOA*01:14 with increased DM inhibition [14, 15]. Interestingly, DOA*01:08 could not be expressed in vitro [15]. However, all these alleles have very low population frequencies, which will make it rather difficult to gather enough samples for clinical investigations.

In conclusion, we genotyped over 1500 samples for *DMA*, *DMB*, *DOA* and *DOB* and submitted 384 sequences (including confirmations and sequence extensions) to the IPD-IMGT/HLA Database. With these additions, the database should now cover most protein sequences of major European populations (Germany, GB/Ireland and Poland), which will benefit future genotyping. If some of the more frequent variations have an impact on HCT or disease phenotypes remains to be investigated. Since these are limited to only a few genomic positions, genotyping could be performed with feasible effort and costs.

Author Contributions

K.P. designed the primers. M.P. and C.P. performed the sequencing. V.A. and A.K. genotyped the samples. V.A. characterised and submitted novel alleles. A.K. and J.S. analysed frequency haplotype data. A.K. prepared figures and tables. A.K. and V.L. wrote the first draft of the manuscript. A.K., A.H.S. and V.L. conceived and supervised the work. All authors contributed to the manuscript revision and read and approved the submitted version.

Acknowledgements

We are grateful to all members of the DKMS Life Science Lab for their dedicated daily work that was fundamental to the analysis of all the donor samples.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. A. P. Kelly, J. J. Monaco, S. Cho, and J. Trowsdale, "A New Human HLA Class II-Related Locus, DM," *Nature* 353, no. 6344 (1991): 571–573.

2. L. Karlsson, C. D. Surh, J. Sprent, and P. A. Peterson, "A Novel Class II MHC Molecule With Unusual Tissue Distribution," *Nature* 351, no. 6326 (1991): 485–488.

3. L. K. Denzin and P. Cresswell, "HLA-DM Induces Clip Dissociation From MHC Class II αβ Dimers and Facilitates Peptide Loading," *Cell* 82, no. 1 (1995): 155–165.

4. D. C. Douek and D. M. Altmann, "HLA-DO Is an Intracellular Class II Molecule With Distinctive Thymic Expression," *International Immunology* 9, no. 3 (1997): 355–364.

5. T. M. C. Hornell, T. Burster, F. L. Jahnsen, et al., "Human Dendritic Cell Expression of HLA-DO Is Subset Specific and Regulated by Maturation," *Journal of Immunology* 176, no. 6 (2006): 3536–3547.

6. R. A. Welsh and S. Sadegh-Nasseri, "The Love and Hate Relationship of HLA-DM/DO in the Selection of Immunodominant Epitopes," *Current Opinion in Immunology* 64 (2020): 117–123.

7. A. I. Guce, S. E. Mortimer, T. Yoon, et al., "HLA-DO Acts as a Substrate Mimic to Inhibit HLA-DM by a Competitive Mechanism," *Nature Structural & Molecular Biology* 20, no. 1 (2013): 90–98.

8. L. K. Denzin, D. B. Sant'Angelo, C. Hammond, M. J. Surman, and P. Cresswell, "Negative Regulation by HLA-DO of MHC Class II-Restricted Antigen Processing," *Science* 278, no. 5335 (1997): 106–109.

9. P. P. Nanaware, M. M. Jurewicz, J. D. Leszyk, S. A. Shaffer, and L. J. Stern, "HLA-DO Modulates the Diversity of the MHC-II Self-Peptidome," *Molecular & Cellular Proteomics* 18, no. 3 (2019): 490–503.

10. B. Budeus, M. Álvaro-Benito, and P. Crivello, "HLA-DM and HLA-DO Interplay for the Peptide Editing of HLA Class II in Healthy Tissues and Leukemia," *Best Practice & Research Clinical Haematology* 37, no. 2 (2024): 101561.

11. Y. O. Poluektov, A. Kim, I. Z. Hartman, and S. Sadegh-Nasseri, "HLA-DO as the Optimizer of Epitope Selection for MHC Class II Antigen Presentation," *PLoS One* 8, no. 8 (2013): e71228.

12. R. Welsh, N. Song, and S. Sadegh-Nasseri, "What to DO With HLA--DO/H-2O Two Decades Later?," *Immunogenetics* 71, no. 3 (2019): 189–196.

13. J. Robinson, D. J. Barker, X. Georgiou, M. A. Cooper, P. Flicek, and S. G. E. Marsh, "IPD-IMGT/HLA Database," *Nucleic Acids Research* 48, no. D1 (2020): D948–D955.

14. L. K. Denzin, A. A. Khan, F. Virdis, et al., "Neutralizing Antibody Responses to Viral Infections Are Linked to the Non-Classical MHC Class II Gene H2-Ob," *Immunity* 47, no. 2 (2017): 310–322.e7.

15. A. M. Graves, F. Virdis, E. Morrison, et al., "Human Hepatitis B Viral Infection Outcomes Are Linked to Naturally Occurring Variants of HLA-DOA That Have Altered Function," *Journal of Immunology* 205, no. 4 (2020): 923–935.

16. M. Álvaro-Benito, M. Wieczorek, J. Sticht, C. Kipar, and C. Freund, "HLA-DMA Polymorphisms Differentially Affect MHC Class II Peptide Loading," *Journal of Immunology* 194, no. 2 (2015): 803–816.

17. E. W. Petersdorf, C. McKallor, M. Malkki, et al., "HLA Haplotypes and Relapse After Hematopoietic Cell Transplantation," *Journal of Clinical Oncology* 42, no. 8 (2024): 886–897.

18. E. Zino, G. Frumento, S. Marktel, et al., "A T-Cell Epitope Encoded by a Subset of HLA-DPB1 Alleles Determines Nonpermissive Mismatches for Hematologic Stem Cell Transplantation," *Blood* 103, no. 4 (2004): 1417–1424.

19. T. Meurer, P. Crivello, M. Metzing, et al., "Permissive HLA-DPB1 Mismatches in HCT Depend on Immunopeptidome Divergence and Editing by HLA-DM," *Blood* 137, no. 7 (2021): 923–928.

20. G. Schöfl, K. Lang, P. Quenzel, et al., "2.7 Million Samples Genotyped for HLA by Next Generation Sequencing: Lessons Learned," *BMC Genomics* 18 (2017): 161.

21. I. Wagner, D. Schefzyk, J. Pruschke, et al., "Allele-Level KIR Genotyping of More Than a Million Samples: Workflow, Algorithm, and Observations," *Frontiers in Immunology* 9 (2018): 2843.

22. K. Putke, V. Albrecht, C. Paech, et al., "Full-Length Characterization of Novel HLA-DRB1 Alleles for Reference Database Submission," *Methods in Molecular Biology* 2809 (2024): 145–156.

23. S. Klasberg, A. H. Schmidt, V. Lange, and G. Schöfl, "DR2S: An Integrated Algorithm Providing Reference-Grade Haplotype Sequences From Heterozygous Samples," *BMC Bioinformatics* 22, no. 1 (2021): 236.

24. V. Albrecht, C. Zweiniger, V. Surendranath, et al., "Dual Redundant Sequencing Strategy: Full-Length Gene Characterisation of 1056 Novel and Confirmatory HLA Alleles," *HLA* 90, no. 2 (2017): 79–87.

25. B. Schöne, M. Fuhrmann, V. Surendranath, A. H. Schmidt, V. Lange, and G. Schöfl, "TypeLoader2: Automated Submission of Novel HLA and Killer-Cell Immunoglobulin-Like Receptor Alleles in Full Length," *HLA* 93, no. 4 (2019): 195–202.

26. B. Schöne, M. Fuhrmann, V. Surendranath, A. H. Schmidt, V. Lange, and G. Schöfl, "Submitting Novel Full-Length HLA, MIC, and KIR Alleles With TypeLoader2," *Methods in Molecular Biology* 2809 (2024): 157–196.

27. C. Schäfer, A. H. Schmidt, and J. Sauter, "Hapl-o-Mat: Open-Source Software for HLA Haplotype Frequency Estimation From Ambiguous and Heterogeneous Data," *BMC Bioinformatics* 18, no. 1 (2017): 284.

28. U. V. Solloch, A. H. Schmidt, and J. Sauter, "Graphical User Interface for the Haplotype Frequency Estimation Software Hapl-o-Mat," *Human Immunology* 83, no. 2 (2022): 107–112.

29. V. Lange, I. Böhme, J. Hofmann, et al., "Cost-Efficient High-Throughput HLA Typing by MiSeq Amplicon Sequencing," *BMC Genomics* 15, no. 1 (2014): 63.

30. A. Klussmeier, C. Massalski, K. Putke, et al., "High-Throughput MICA/B Genotyping of Over Two Million Samples: Workflow and Allele Frequencies," *Frontiers in Immunology* 11 (2020): 314, https://doi.org/10.3389/fimmu.2020.00314.

31. A. H. Schmidt, D. Baier, U. V. Solloch, et al., "Estimation of High-Resolution HLA-A, -B, -C, -DRB1 Allele and Haplotype Frequencies Based on 8862 German Stem Cell Donors and Implications for Strategic Donor Registry Planning," *Human Immunology* 70, no. 11 (2009): 895–902.

32. T. K. Naruse, H. Kawata, M. Ishihara, et al., "Analysis on Allelic Variation of the HLA-DMB Gene in Japanese by PCR-RFLP as Well as Direct DNA Sequencing and Identification of a New DMB Allele, DMB*0105," *Tissue Antigens* 47, no. 6 (1996): 530–537.

33. C. McTernan, C. Mijovic, C. Cockram, and A. Barnett, "The Nucleotide Sequence of a New DMB Allele, DMB*0106," *Tissue Antigens* 55, no. 5 (2000): 470–472.

34. S. M. Moon, H. Gu, H. J. Ryu, et al., "Identification of Four Novel HLA-DOA Alleles, DOA*010106, DOA*0102, DOA*0103, and DOA*0104N, by Sequence-Based Typing," *Tissue Antigens* 66, no. 3 (2005): 242–245.

35. K. J. Karczewski, L. C. Francioli, G. Tiao, et al., "The Mutational Constraint Spectrum Quantified From Variation in 141,456 Humans," *Nature* 581, no. 7809 (2020): 434–443.

36. I. E. Shapiro and M. Bassani-Sternberg, "The Impact of Immunopeptidomics: From Basic Research to Clinical Implementation," *Seminars in Immunology* 66, no. 1 (2023): 101727, https://doi.org/10.1016/j.smim.2023.101727.

37. M. L. Feng, R. Z. Liu, T. Shen, Y. L. Zhao, Z. Y. Zhu, and D. Z. Liu, "Analysis of HLA-DM Polymorphisms in the Chinese Han Population," *Tissue Antigens* 79, no. 3 (2012): 157–164.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.