UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

Isabelle Mira da Silva

VARIABILIDADE DOS GENES *HLA-DOA* E *HLA-DOB*: POLIMORFISMOS, HAPLÓTIPOS E ASPECTOS EVOLUTIVOS

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu
2024

Isabelle Mira da Silva

VARIABILIDADE DOS GENES *HLA-DOA* E *HLA-DOB*: POLIMORFISMOS, HAPLÓTIPOS E ASPECTOS EVOLUTIVOS

Exame de Defesa de Dissertação de Mestrado apresentado ao Instituto de Biociências (IBB), Campus de Botucatu, UNESP, como requisito para obtenção do título de Mestre em Genética pelo Programa de Pós-graduação em Ciências Biológicas (Genética).

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu

2024

M671v

Mira da Silva, Isabelle
VARIABILIDADE DOS GENES HLA-DOA E HLA-DOB: POLIMORFISMOS, HAPLÓTIPOS E ASPECTOS EVOLUTIVOS / Isabelle Mira da Silva. -- Botucatu, 2024
86 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Botucatu
Orientador: Erick da Cruz Castelli

1. HLA-DOA e HLA-DOB. 2. imunogenética. 3. NGS. 4. populações humanas. 5. alelos novos. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: IMUNOGENÉTICA DOS GENES HLA-DOA E HLA-DOB: VARIABILIDADE, HAPLÓTIPOS E ASPECTOS EVOLUTIVOS

AUTORA: ISABELLE MIRA DA SILVA

ORIENTADOR: ERICK DA CRUZ CASTELLI

Aprovada como parte das exigências para obtenção do Título de Mestra em Ciências Biológicas (Genética), pela Comissão Examinadora:

Prof. Dr. ERICK DA CRUZ CASTELLI (Participaçao Presencial)
Departamento de Patologia / Faculdade de Medicina de Botucatu Unesp

Profa. Dra. YARA COSTA NETTO MUNIZ (Participaçao Virtual)
Departamento de Biologia Celular e do Desenvolvimento / Universidade Federal de Santa Catarina - UFSC

Prof. Dr. ROBSON FRANCISCO CARVALHO (Participaçao Presencial)
Departamento de Biologia Estrutural e Funcional / Instituto de Biociências de Botucatu - UNESP

Botucatu, 27 de novembro de 2024

gov.br Documento assinado digitalmente
FLAVIA DANIELI MARTINS GODINHO
Data: 11/12/2024 14:35:12-0300
Verifique em https://validar.iti.gov.br

FLÁVIA DANIELI MARTINS GODINHO
Assistente Técnico Administrativo I da Seção Técnica de Pós-Graduação
Instituto de Biociências de Botucatu – UNESP

“Um amigo é fruto de uma escolha. É uma opção de amor, é uma descoberta de alma irmã. É a consciência clara e permanente de algo sublime que não está na natureza das coisas perecíveis. É um tesouro sem preço, um gostar sem distância, de alguém presente em nosso caminho, nas horas de dúvidas, de alegria, demais para ser perdido, importante para ser esquecido”

Antoine De Saint-Exupéry

AGRADECIMENTOS

Começo meus agradecimentos primeiramente a todas as pessoas maravilhosas que estiveram em minha vida, como minha mãe e irmão Gustavo que sempre me apoiaram e incentivaram, possibilitando chegar e finalizar essa etapa tão importante. Amo muito vocês. Agradeço a toda minha família, por tudo. Agradeço ao meu grande amor, Karine, por tudo o que você é. Obrigada por tudo que estamos vivendo e vamos viver, pelo apoio e força, te amo.

Agradeço à minha orientadora de Iniciação Científica, Profa. Dra. Flávia Karina Delella, minha amiga do coração, fada e inspiração pessoal e profissional para esse caminho.

Agradeço à todas as minhas amigas e amigos especiais, Luana, Leonardo, Jefferson, Raphaela, Viviane, Gabriela, Joyce, Amanda, Nayane, Heloísa, Marcel, Ícaro e João, por tudo, as risadas, choros, abraços, aprendizados (mais choros, brincadeira) e tudo de incrível que pudemos compartilhar e guardar. Que muito mais desses momentos venham, amo muito vocês.

Agradeço ao meu orientador, Prof. Dr. Erick da Cruz Castelli, por compartilhar um pouco do seu vasto conhecimento, saberes, conversas científicas e incentivo nesse caminho. Agradeço a Profa. Camila, por compartilhar um pouco dos seus conhecimentos imunológicos em nossas conversas científicas, pelo apoio e ajuda.

Agradeço à CNPq pelo apoio financeiro para que esse estudo pudesse ser desenvolvido. Agradeço por todo o apoio e ajuda da Orquestra Filarmônica do Instituto de Biociências (OFIBB), na qual me ajudou angariando os valores para pagamento da inscrição do processo seletivo. Agradeço imensamente ao Programa de Pós-Graduação de Ciências Biológicas (Genética), pelo acolhimento e auxílio em todas as etapas, desde o primeiro até o último dia do mestrado. Agradeço também a Unidade de Pesquisa Experimental (UNIPEX)/FMB pela disponibilização da equipe, espaço e materiais para desenvolvimento do trabalho.

RESUMO

Os genes *HLA-DOA* e *HLA-DOB* codificam as duas subunidades que compõem a molécula HLA-DO, responsáveis por mediar o processamento e carregamento de antígenos em moléculas HLA de classe II. Estes genes exibem menor polimorfismo em comparação com outros genes HLA. No entanto, estes polimorfismos podem influenciar significativamente a suscetibilidade à infecções por patógenos e ao desenvolvimento de doenças autoimunes. Em nosso estudo, avaliamos 5.349 amostras de diferentes populações ao redor do mundo, incluindo o Brasil. Utilizamos um *pipeline* de bioinformática direcionado para a avaliação de genes HLA, o que nos permitiu caracterizar a diversidade genética observada nos genes *HLA-DOA* e *HLA-DOB* e encontrar variantes até então desconhecidas. Apesar de levantados, os alelos relacionados aos genes foram pouco frequentes e restritos a determinadas regiões biogeográficas, como por exemplo o *DOA*01:01:11* e *DOB*01:03:01* presente no continente africado, europeu e americano (sul e norte). Adicionalmente, avaliamos os perfis de frequência alélica para cada região biogeográfica. Além disso, mesmo incluíndo dados de população brasileira (miscigenadas), esses genes apresentaram-se bastante conservados, principalmente à nível de isoformas proteicas, sendo para ambos a presença majoritária de *DOA*01:01* e *DOB*01:01*. Mas além disso, em termos de sequência (de DNA e da proteína codificada) quando comparados aos genes vizinhos, possivelmente devido a forças evolutivas mantendo a estrutura da molécula HLA-DO devido à sua importante função.

Palavras-chave: *HLA-DOA* e *HLA-DOB*, imunogenética, NGS, populações humanas, alelos novos

Abstract

The HLA-DOA and HLA-DOB genes encode the two subunits that comprise the HLA-DO molecule, which plays a critical role in mediating antigen processing and loading in class II HLA molecules. These genes exhibit lower polymorphism compared to other HLA genes. However, these polymorphisms can significantly influence susceptibility to pathogen infections and the development of autoimmune diseases. In this study, we analyzed 5,349 samples from diverse populations worldwide, including Brazil. Using a bioinformatics pipeline tailored for the evaluation of HLA genes, we characterized the genetic diversity of *HLA-DOA* and *HLA-DOB*, identifying previously unknown variants. Both genes presented a few different alleles, which are restricted to specific biogeographical regions. For instance, *DOA*01:01:11* and *DOB*01:03:01* were observed in African, European, and American (both North and South) populations. Additionally, we assessed allele frequency profiles for each biogeographical region. Despite the inclusion of data from the Brazilian population, characterized by extensive admixture, these genes appeared to be highly conserved, particularly at the protein isoform level, with *DOA*01:01* and *DOB*01:01* being the most prevalent for both genes. Moreover, at both the DNA and protein sequence levels, these genes demonstrated high conservation compared to neighboring HLA genes, potentially driven by evolutionary pressures maintaining the structure of the HLA-DO molecule due to its critical biological function.

Keywords: *HLA-DOA* and *HLA-DOB*, genetic variability, linkage disequilibrium, nucleotide diversity.

Lista de Figuras – Revisão Bibliográfica

Lista de Figuras do Artigo

Lista de Figuras dos Materiais Suplementares

Lista de Tabelas

Lista de Tabelas dos Materiais Suplementares

SUMÁRIO

CAPÍTULO I – Revisão da Literatura

1. INTRODUÇÃO

1.1 RESPOSTA IMUNOLÓGICA ADAPTATIVA E A VIA DE APRESENTAÇÃO DE ANTÍGENO POR MOLÉCULAS HLA DE CLASSE II

O Complexo Principal de Histocompatibilidade (do inglês, *Major Histocompatibility Complex* - MHC), com a localização 6p21.3, é a região do genoma humano com maior nível de diversidade alélica e densidade gênica. Os genes do MHC estão, em sua maioria, relacionados com o controle das respostas imunológicas contra microrganismos, atuando principalmente na via de processamento e apresentação de antígenos aos linfócitos T e modulação da resposta imunológica. No processo de invasão/persistência patogênica, ou detecção de células genética ou fisiologicamente alteradas, diferentes vias são requeridas para conter sua proliferação e/ou eliminá-las, sendo umas das principais a MHC II. Tal via possui essa nomenclatura, pois é intermediada por moléculas de MHC de classe II, majoritariamente expressas em células apresentadoras de antígeno (do inglês, *Antigen Presenting Cells* – APC), como Macrófagos (MFs) e Células Dendríticas (CDs)[1].

O fator desencadeador da via MHC II, é a internalização do patógeno, ou proteína liberada pelas células alteradas, através da endocitose ou fagocitose realizada pelo MF e pelas CDs. Após sua retenção, esse componente será pré-processado na vesícula endolisossomal com auxílio dos lisossomos e encaminhada ao compartimento de processamento de antígenos (também conhecido como compartimento de MHC de classe II – MIIC) (**Figura 1**). Nesse local, os fragmentos, ainda não completamente denaturados, são marcados através de mudanças físico-químicas induzidas pelo meio acidificado pelas ATPases vacuolares internalizadoras de Hidrogênio $(H)^{+}$[2,3]. A finalização do processo de denaturação é realizado pela Tiol Redutase Lisossômica Induzível por IFNγ (do inglês, *IFNγ-Inducible Lysosomal Thiol Reductase* – GILT), na qual identifica a marcação e, juntamente com o meio acidificado, auxilia a quebra das pontes de sulfeto[4].

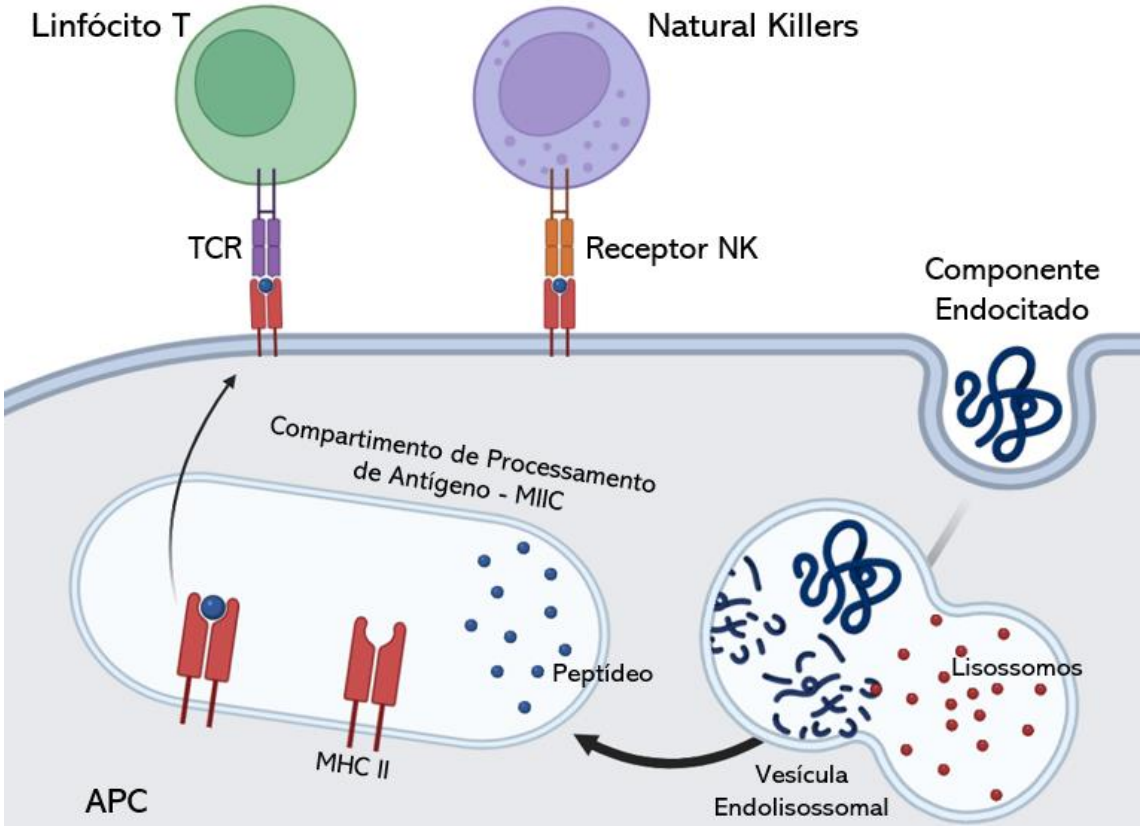

Figura 1. Via MHC de classe II de apresentação de peptídeos. O ponto de partida da via é a internalização de um componente endocitado, que será, posteriormente, processado no Compartimento de Processamento de Antígeno e apresentado na superfície da célula APC com auxílio da molécula MHC II.

Após a etapa finalizada, esses fragmentos são fracionados, através da ação dos lisossomos, em partes menores chamadas de peptídeos ou antígenos que serão acomodados na fenda peptídica das moléculas MHC de classe II. A partir do carregamento do peptídeo no MHC II realizado no compartimento MIIC, formando o complexo MHCII-peptídeo, a molécula é direcionada à superfície da célula APC. Para a modulação e desencadeamento da resposta dos Linfócitos T (LT) CD4+ (ou *helper*), é necessário a interação do MHCII-peptídeo com os receptores das células T (do inglês, *T Cell Receptor* – TCR). Mas além desse contato, outras moléculas co-estimulatórias expressas tanto pela célula APC quanto pelo Linfócito, como CD80, CD86 e CD40 (APC); e CD28 e CD40L (LT) são fundamentais para a ativação dos LT CD4+, juntamente com fatores solúveis e citocinas também liberados pela APC[5,6] (**Figura 2**).

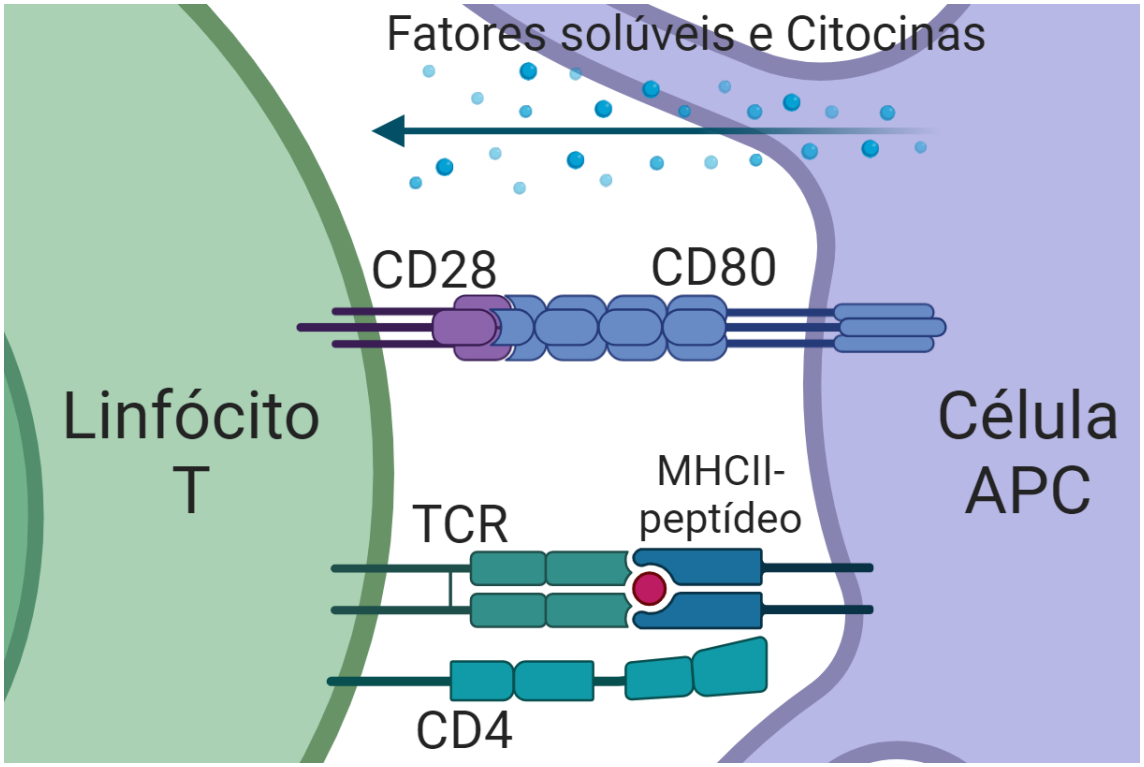

Figura 2. Representação da interação dos receptores da APC e o LT CD4+. Adaptado de "TCR (T Cell Receptor)", por Biorender.com (2024).

1.2 O COMPLEXO PRINCIPAL DE HISTOCOMPATIBILIDADE E OS GENES *HLA*

Entre os genes mais conhecidos e estudados do MHC humano, estão os genes HLA (do inglês, *Human Leucocyte Antigen*)[7,8], envolvidos com apresentação de antígenos e modulação de células Natural Killer (NK). O grupo gênico *HLA* possui cerca de 20 genes, que estão didaticamente divididos em dois grupos ou classes principais (**Figura 3**). Os genes *HLA* de classe I são expressos na superfície da maioria das células nucleadas. Este grupo de genes são subdivididos em clássicos, englobando os genes mais polimórficos do genoma humano e envolvidos com apresentação de antígenos, *HLA-A, HLA-B* e *HLA-C*, e os não-clássicos, *HLA-E, HLA-F* e *HLA-G*, que são menos polimórficos e relacionados à modulação da atividade citotóxica de células T e NK. Já os genes *HLA* de classe II, também são subdivididos em clássicos, *HLA-DRA, HLA-DRB1-DRB9*, *HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DPA1, HLA-DPA2, HLA-DPB1* e *HLA-DPB2*, e envolvidos na apresentação de antígenos; e em não-

clássicos, *HLA-DMA, HLA-DMB, HLA-DOA e HLA-DOB*; envolvidos com o carregamento de antígenos nas moléculas HLA-II[9].

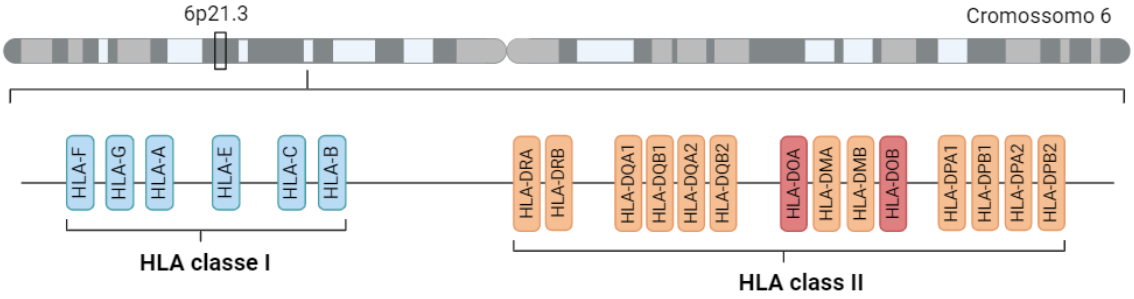

Figura 3. Representação gráfica da ordem dos genes HLA na região do MHC humano, em 6p21.3. No grupo HLA classe II, os genes *HLA-DOA* e *-DOB* estão destacados em vermelho.

As moléculas HLA de classe II são formadas pela combinação de duas subunidades (são heterodímeros), uma α e uma β. As cadeias são originadas por genes diferentes, ou seja, a cadeia α codificada pelo gene *HLA-DRA*, *HLA-DQA1*, *HLA-DPA1*, *HLA-DOA*, etc, enquanto a cadeia β é codificada pelo gene *HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*, *HLA-DOB*, etc. A molécula HLA-DO, por exemplo, é a união das cadeias provenientes de *HLA-DOA* e *HLA-DOB*[10,11]. Ambos genes expressam suas respectivas cadeias α e β e após serem produzidas, as sequências de aminoácidos das mesmas são direcionadas ao Retículo Endoplasmático (RE), onde os resíduos de aminoácidos Arginina (posição 44), Treonina (posição 45), Glutamina (posição 47), Fenilalanina (posição 48), Serina (posição 49), Lisina (posição 50), Fenilalanina (posição 51) e Glicina (posição 53)[12] interagem de maneira não-covalente, formando o heterodímero HLA de classe II, HLA-DO. As moléculas de classe II, bem como a molécula HLA-DO, possuem uma porção transmembranar e citoplasmática, dois domínios α1 e α2 (subunidade α) e β1 e β2 (subunidade β). A fenda de ligação ao antígeno está presente na união dos domínios α1β1. Quando as duas subunidades se unem, a fenda de ligação de peptídeo apresenta um peptídeo denominado CLIP - Peptídeo de Cadeia Invariável (do inglês, *Invariable Chain Peptide*), também conhecido como CD74, para manutenção de sua estabilidade estrutural da molécula de classe II quando armazenadas no RE.

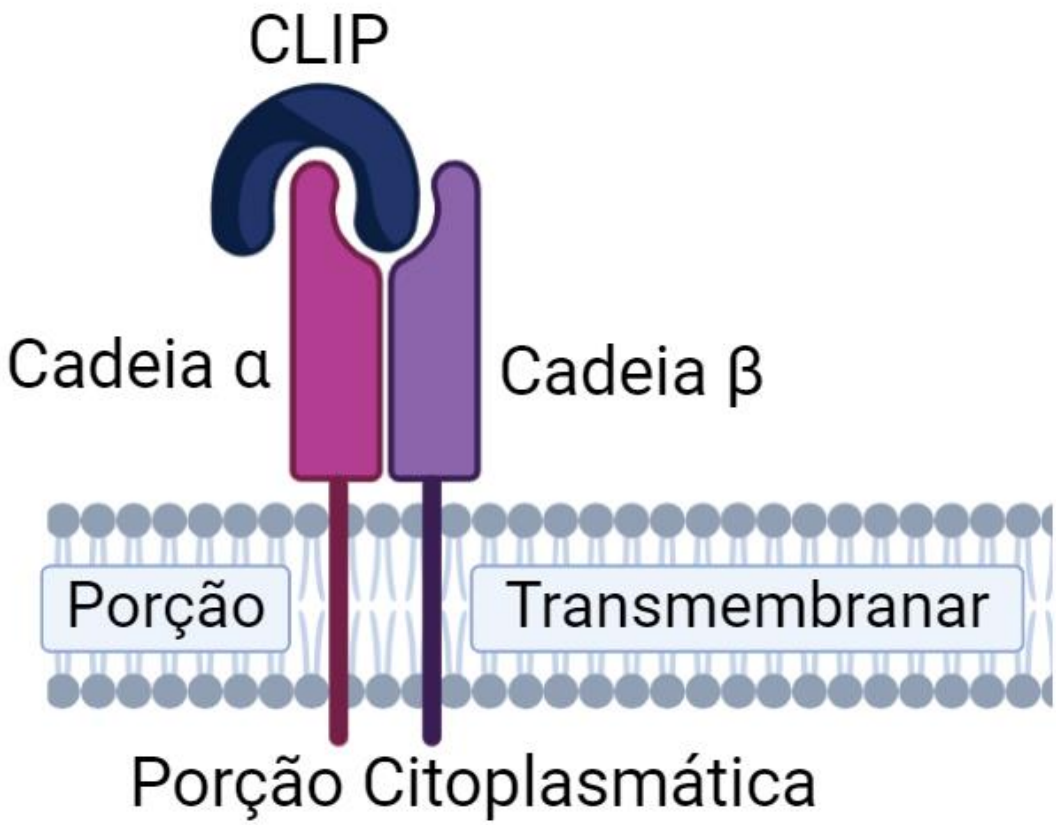

Figura 4. Representação da molécula de HLA de classe II aderida a membrana plasmática do Retículo Endoplasmático (RE).

Os genes HLA de classe II, principalmente os clássicos, estão entre os genes mais polimórficos do genoma humano e contabiliza, aproximadamente, 37 mil alelos descritos no banco de dados IPD-IMGT/HLA, versão 3.56, **Tabela 1** (https://www.ebi.ac.uk/ipd/imgt/hla/release/v356/). No entanto, os genes *HLA-DOA* e *HLA-DOB*, alvos desde estudo, são caracterizados como menos polimórficos (somando-se, juntos, 155 alelos), porém, circundados por genes altamente polimórficos, como o *HLA-DRB1, HLA-DQB1* e *HLA-DPB1* (**Figura 3**, **Tabela 1**).

HLA Classe I		HLA Classe II	
Genes	**Alelos**	**Genes**	**Alelos**
HLA-A	8.288	*HLA-DRB*	4.639
HLA-B	9.877	*HLA-DQA1*	773
HLA-C	8.361	*HLA-DQB1*	2.549
HLA-E	353	*HLA-DPA1*	678
HLA-F	91	*HLA-DPB1*	2.569
HLA-G	160	*HLA-DMA*	59
		HLA-DMB	82
		HLA-DOA	92
		HLA-DOB	63

Tabela 1. Número de alelos descritos para os genes *HLA* atualmente. **Fonte:** IPD-IMGT/HLA database, versão 3.56.0.

1.3 OS GENES *HLA-DOA* E *HLA-DOB* E A FUNÇÃO DE HLA-DO

O gene *HLA-DOA* possui cerca de 5.414pb, com 5 éxons (**Figura 5**). O éxon 1 codifica o peptídeo líder (ou sinal), os éxons 2 e 3 codificam os domínios α1 e α2, respectivamente, e os éxons 4 e 5 codificam a porção transmembranar e citoplasmática. *HLA-DOA* apresenta 92 variantes alélicas catalogadas (https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/) (**Tabela 1**) e 14 isoformas proteicas resultantes de alguns desses alelos. O número de isoformas proteicas é bastante inferior ao número de sequências de DNA conhecidas (os alelos), pois a maioria das trocas de base estão localizadas em íntrons ou configuram alterações sinônimas[13].

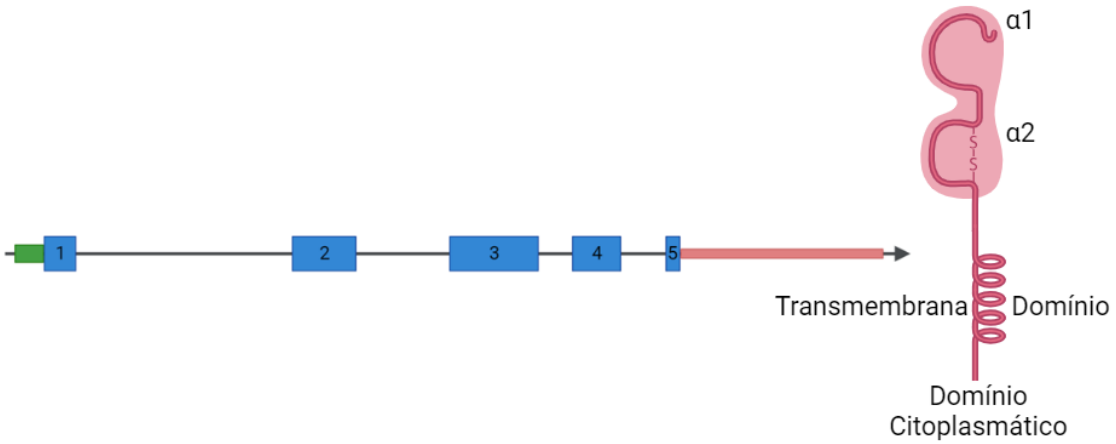

Figura 5. À esquerda, representação esquemática da estrutura do gene *HLA-DOA*, nos quais o bloco verde representa a região 5' não traduzida, os blocos azuis numerados representam os éxons, as linhas pretas representam os íntrons, e o bloco rosa representa a região 3' não-traduzida. À direita, a proteína HLA-DOA. **Fonte:** De autoria própria.

O *HLA-DOB* possui cerca de 4.240pb, com 6 éxons. O éxon 1 codifica o peptídeo líder, os éxons 2 e 3 codificam os domínios β1 e β2, respectivamente, os éxons 4, 5 e 6 codificam a porção transmembranar e citoplasmática (**Figura 6**). *HLA-DOB* é menos polimórfico que *HLA-DOA*, com 63 variantes alélicas reportadas (**Tabela 1**), que codificam 16 isoformas proteicas (https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/).

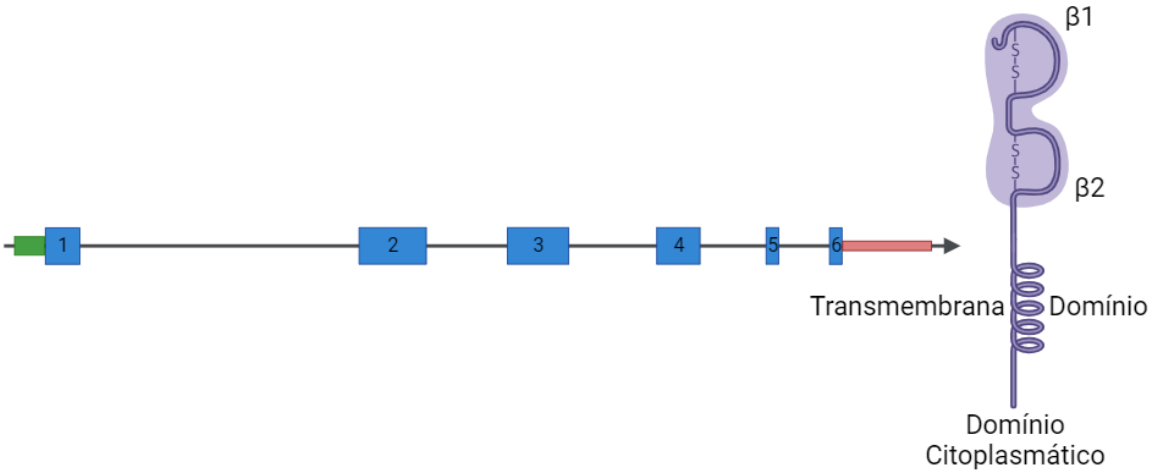

Figura 6. À esquerda, representação esquemática da estrutura do gene *HLA-DOB*, nos quais o bloco verde representa a região 5' não traduzida, os blocos azuis numerados representam os éxons, as linhas pretas representam os íntrons, e o bloco rosa representa a região 3' não-traduzida. À direita, a proteína HLA-DOB. **Fonte:** De autoria própria.

HLA-DOA e *HLA-DOB* codificam as subunidades da molécula HLA-DO (**Figura 7.A**). Após formada, HLA-DO é armazenada no Retículo

Endoplasmático (RE) também associada ao CLIP (**Figura 7.B**)[14], assim como outras moléculas HLA de classe II.

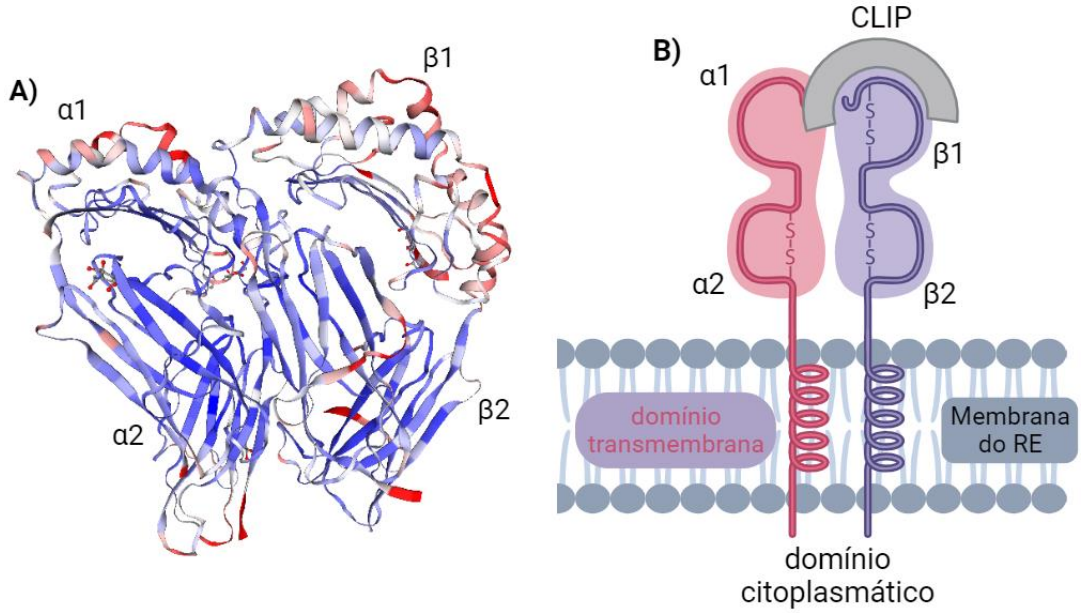

Figura 7. A) Representação tridimensional da estrutura de HLA-DO (P06340) disponível no repositório SWISS-MODEL e **B)** Representação da estrutura molecular do HLA de Classe II, grupo pertencente às proteínas de DOA e DOB, demonstrando seus domínios α e β, transmembrana e citoplasmático. **Fonte:** De autoria própria.

Para exercer sua função no compartimento de processamento de antígenos, HLA-DO é dependente de outra molécula, a HLA-DM, pois somente com a oligomerização e formação do complexo HLA-DO/HLA-DM (complexo DO/DM), HLA-DO adquire a capacidade de emigrar do RE[2]. Alguns estudos apontam HLA-DO como mimético de HLA-II, competindo pelo peptídeo ou inibindo as ações de HLA-DM[15,16]. Porém, outros estudos baseados na ausência de HLA-DO apoiam que essa molécula seja estabilizadora de HLA-DM para seleção, "edição" e troca do CLIP pelo fragmento peptídico instável, conhecido como imunodominante, nas moléculas HLA-II, além de contribuir para aumento da diversidade do repertório peptídico "próprio" e "não-próprio"[17,18].

A construção e diversificação do repertório de peptídeos intermediada pelo complexo DO/DM é fundamental para reconhecimento e diferenciação dos antígenos externos e internos (autoantígenos), bem como, mais especificamente para o segundo caso, para a construção do limiar de tolerância, ou seja, limite

imunológico de não reatividade aos antígenos “próprios” expostos previamente; e treinamento/seleção dos LT e B atuantes na resposta imunológica adaptativa[19]. Diversas etapas para transformação do antígeno fagocitado ou endocitado em peptídeos são necessárias para sua apresentação na superfície celular (**Figura 8**).

As moléculas de HLA-II, ao chegarem no compartimento de processamento de antígeno por meio de informações de classificação na porção terminal NH_2 no CLIP associadas a elas, a molécula internalizada permanece aderida à uma vesícula intraluminal (do inglês, *Intraluminal Vesicule* - ILV) até ser fixada na parte interna da membrana do compartimento[20]. Na membrana, o HLA II relaciona-se com o complexo DO-DM, mais especificamente com HLA-DM, pois é ela quem fará a troca do CLIP pelo peptídeo. Em seguida, HLA II emigrará do compartimento com ajuda, inicialmente de ILV, e posteriormente, por uma vesícula citoplasmática, chegando à membrana plasmática extracelular, expondo-se aos Linfócitos T[21].

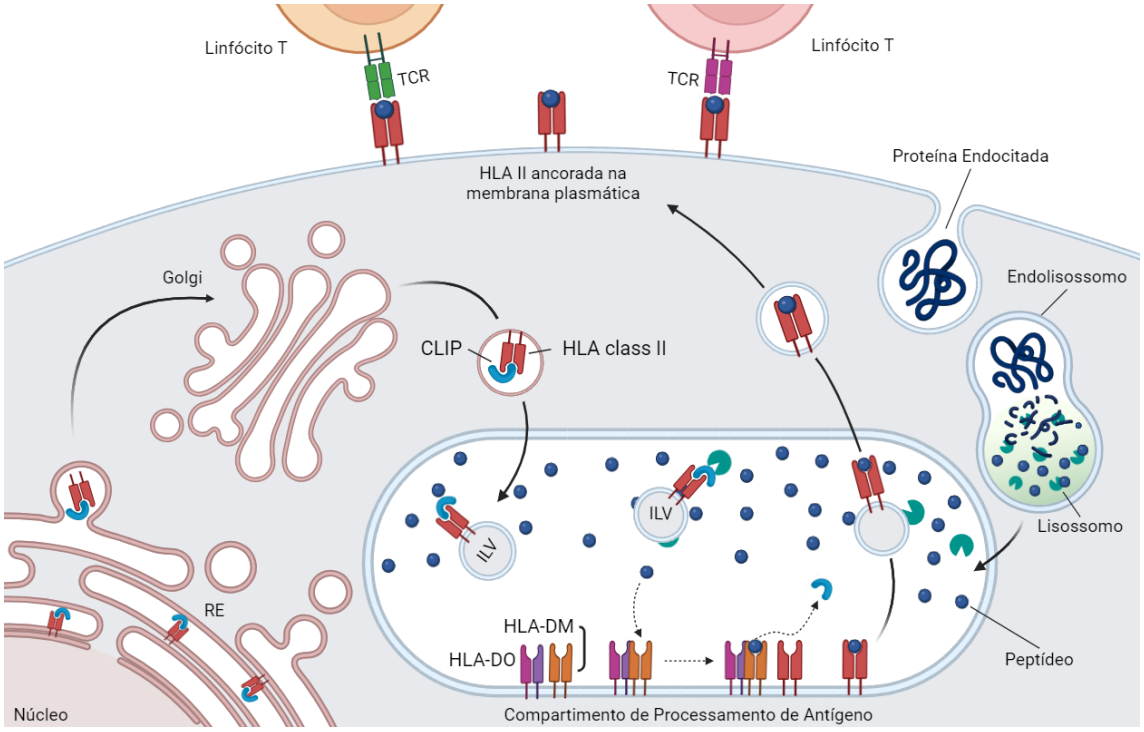

Figura 8. Representação do processamento e carregamento de peptídeos de HLA classe II intermediado por HLA-DO e HLA-DM. **Baseado em:** ROCHE, P., FURUTA, K. 2015.

1.4 A HISTÓRIA EVOLUTIVA DOS *HLA* E A DIVERSIDADE DOS GENES *HLA-DOA* E *HLA-DOB*

Diferentes estudos[22,23] buscaram reconstruir e entender a história evolutiva dos genes *HLA*. Sabe-se que a interação entre o sistema imunológico e os fatores externos, como os patógenos, foi essencial para o aumento da variabilidade alélica, principalmente dos genes *HLA,* responsáveis pela apresentação de peptídeos[24]. A quantidade expressiva de polimorfismos nos genes clássicos *HLA-A, -B* e *-C, HLA-DRB1*, *-DQB1* e *-DPB1*, todos relacionadas com a apresentação de peptídeos[9,25], demonstra a ocorrência de seleção balanceadora ou divergente nos mesmos.

Em contraste, o HLA-DO é uma molécula altamente conservada, quando comparada a outros genes HLA de Classe II, com um perfil de seleção natural completamente diferente do observado para os genes clássicos. Como especulado por He (2020), devido à influência do HLA-DO e HLA-DM no processamento de antígenos e seguindo a perspectiva evolutiva dos HLA Classe II, ambos os genes podem ter passado por processos que levaram ao estabelecimento de suas funções. Nesse caso, há possível processo de eliminação de variantes deletérias, ou seja, ação da seleção purificadora (ou negativa) ou a diminuição da frequência de outros alelos e a sobressalência de um único, no caso da seleção direcional (ou positiva)[26,27].

A partir do final da década de 1960, o grupo de genes HLA foi considerado influente nos processos imunológicos de rejeição aos transplantes, tornando-se foco de investigações para definições de quais genes e alelos seriam essenciais na interação e desencadeamento da resposta que interferia na sobrevida do enxerto[28]. Atualmente, o “padrão ouro” para checar a compatibilidade entre doador e receptor analisa principalmente os genes *HLA-A*, *HLA-B*, *HLA-C* e *HLA-DRB1* (nesse caso, correspondência de 8/8) e, em alguns casos, há inclusão de *HLA-DPB1, -DPA1, HLA-DQA1* e *-DQB1*[29,30]. A definição de genes e alelos a serem considerados na tipagem de HLA para transplantes é baseado no catálogo de Alelos HLA Comuns, Intermediários e Bem-

documentados (do inglês *Common, intermediate and well-documented HLA alleles* – CIWD HLA alleles) organizado e mantido pela Sociedade Americana de Histocompatibilidade e Imunogenética (*American Society for Histocompatibility and Immunogenetics* – ASHI)[31].

Por *HLA-DOA* e *HLA-DOB* não serem considerados um gene relevante para transplantes (pois não é expresso na superfície celular), é possível que a diversidade conhecida para estes genes seja extremamente subestimada. Além disso, raros são os trabalhos que avaliaram *HLA-DOA* e *HLA-DOB* em populações reais, especialmente populações miscigenadas como as observadas na América.

1.5 OS BANCOS DE DADOS INTERNACIONAIS E A POPULAÇÃO BRASILEIRA

A maioria das pesquisas[32,33] destinadas à investigação de polimorfismos utiliza o banco de dados do projeto *1000genomes* - 1000GP e HGDP, como as principais fontes de informações de variantes genéticas comuns e raras em indivíduos saudáveis.

O projeto 1000GP é um banco de dados genômico, especificamente de variantes genéticas humanas, destinado a estudos clínicos e evolutivos. Iniciado em 2007, o projeto piloto[34] abrangia o sequenciamento de genoma completo de baixa cobertura (2-4X) de 180 indivíduos não-relacionados de 5 regiões, Europa, Leste Asiático, Sul da Ásia, Oeste da África e América; e 2 trios relacionados (trios de mãe-pai-filha(o)). Contudo, ao longo das décadas, mais amostras e grupos populacionais foram incluídos, bem como o melhoramento das técnicas de sequenciamento (plataforma Illumina NovaSeq 6000); contabilizando hoje, na última fase do projeto[35], 2.504 amostras não-relacionadas e 698 amostras relacionadas de 26 populações de 5 regiões biogeográficas, totalizando 3.202 amostras sequenciadas de genoma completo de alta cobertura (cerca de 30X)[36]. O dado bruto de sequenciamento está disponível publicamente para *download*.

O projeto *Human Genome Diversity Project* (HGPD) foi idealizado pelo Centro de Estudos de Polimorfismos Humanos (*Centre d'Etude du Polymorphisme Humain* – CEPH). Trata-se de um banco de dados destinado ao armazenamento de informações genéticas (especificamente, Polimorfismos de Nucleotídeo Único (SNPs), Inserções-Deleções (InDels) e variantes de número de cópia (do inglês, *Copy Number Variations - CNVs*) de 54 populações humanas da África, África Central, Oceania e América, contabilizando 929 amostras. A característica populacional desse banco é a maior diversidade em termos de ancestralidade, tendo em vista a presença de populações indígenas, como Surui e Karitiana (região da América do Sul). As amostras depositadas no banco foram provenientes do sequenciamento de genoma completo realizado com a tecnologia Illumina HiSeq X, de cobertura média de 35x. Além disso, o mapeamento das sequências foi baseado na referência GRCh38[37,38].

Semelhante ao 1000GP, no qual incluiu princípios éticos, ou seja, proteção, integridade e finalidades dos dados fornecidos, o HGDP igualmente os incluiu em sua diretriz de proteção de dados, mas também ampliou os critérios para uso das informações disponibilizadas, nesse caso, adicionando e discriminando a proibição da comercialização e exploração das informações abrangidas pelo banco de dados[39].

Em 2015, com o fim do projeto 1000GP, o Centro de Coordenação de Dados do Laboratório Europeu de Biologia Molecular do Instituto de Bioinformática Europeu (do inglês, *European Molecular Biology Laboratory's European Bioinformatics Institute* – EMBL-EBI) recebeu fundos para a reunião das amostras genômicas do 1000GP, HGDP, Projeto de Diversidade do Genoma Simons (do inglês *Simons Genome Diversity Project* - SGDP) e Projeto de Variação do Genoma da Gâmbia (do inglês, *Gambian Genome Variation Project - GGVP)* (https://www.internationalgenome.org/about), para constituição do banco de dados *International Genome Sample Resource* (IGSR) (**Figura 6**).

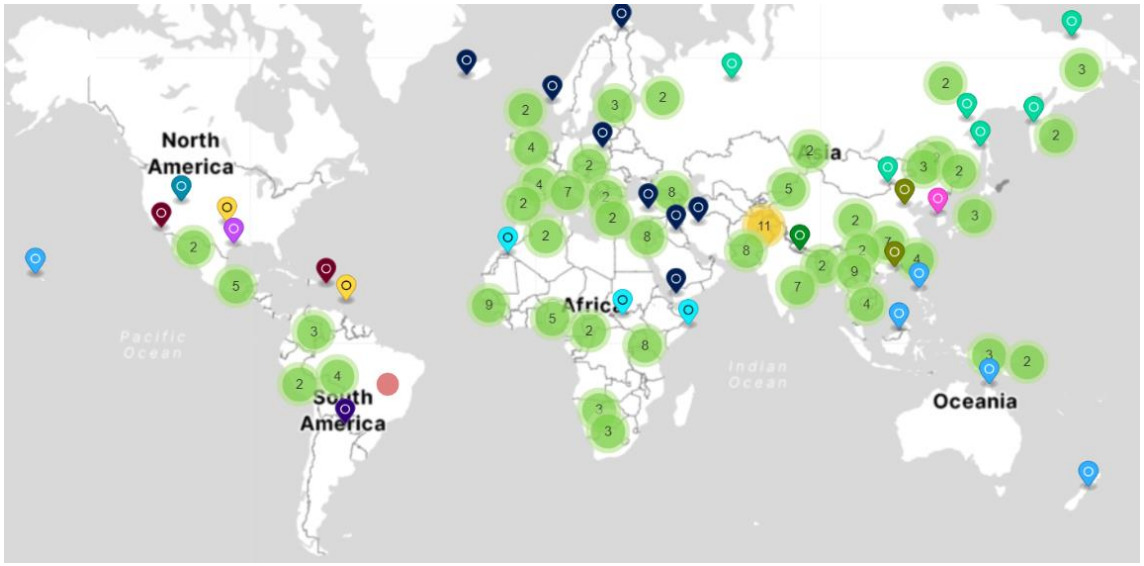

Figura 9. Mapa extraído do *The International Genomes Sample Resource* (https://www.internationalgenome.org), demonstrando a distribuição das amostras populacionais pelos continentes ou regiões biogeográficas. Os números representam as populações relacionadas ao banco de dados 1000GP e os pontos coloridos, são relacionadas àquelas provenientes do HGDP.

Todavia, algo comum a estas fontes de dados é a defasagem ou ausência de amostras advindas da América, ou de populações altamente miscigenadas. Em estudos conduzidos por Popejoy em 2016[40] e Martin e colaboradores em 2017[41], diversas evidências foram levantadas sobre os impactos negativos da falta de representação de populações miscigenadas. Popejoy demonstrou que 81% dos estudos de GWAS baseiam-se majoritariamente em amostras europeias. Por conta disso, há um aumento significativo nas porcentagens ou perpetuações de relações falso positivas de variantes erroneamente inferidas a populações secundárias (com ancestralidade, nesse caso, europeia, mas fora da Europa). Nesse caso, os possíveis impactos dessas variantes na estrutura gênica, variação (aumento ou diminuição) de frequências alélicas de determinados genes, dificuldade no estabelecimento de susceptibilidade ou desencadeadores de diferentes patologias[42], ou atribuições incorretas de variantes ou nomenclaturas de alelos no processo de reconstrução da história evolutiva de diferentes genes, com destaque aos genes *HLA*. Como observado no principal banco de dados para análise de variantes genéticas em diferentes populações, 1000GP, há uma sub-representação amostral de populações multi-ancestrais, como é visto no Brasil (círculo vermelho na **Figura 9**).

Atualmente, o Brasil possui aproximadamente 212 milhões de habitantes com ancestralidade genética compartilhado entre diferentes povos, como nativos sul-americanos, europeus imigrantes e colonizadores (italianos, espanhóis, alemães, poloneses, e principalmente, portugueses lusitanos e açorianos)[43,44], africanos trazidos ao Brasil no período da escravidão (etnia sudanesa - Guiné Equatorial e, majoritariamente, da etnia bantus - Angola)[45,46] e imigrantes asiáticos (sírios, libaneses e principalmente japoneses)[47], como visto na **Figura 10**. Como demonstrado por Escher e colaboradores[48], apesar da existência de bancos de dados disponível que forneçam informações sobre populações com características genéticas semelhantes à brasileira, ou seja, ancestralidade multi-híbrida (tri ou tetra-híbrida), como a dos Estados Unidos, Haiti, Jamaica, México, entre outras; que são importantes para comparações em estudos genéticos entre algumas populações, esses mesmos bancos carecem de informações sobre a brasileira; o que torna ambígua as análises de alelos raros e específicos dessa população. Para preenchimento dessa lacuna, ou seja, defasagem de sequências genômicas de amostras de populações com ancestralidade múltipla sul-americanas, a Organização Pan-Americana de Saúde vinculada a Organização Mundial da Saúde (OPAS/OMG) idealizou e concretizou o primeiro banco de dados específico da população brasileira, o banco SABE.

Figura 10. Mapa mundial demonstrando, nos continentes Africano, Europeu e Asiático mencionados acima, os países de origem das populações/etnias responsáveis pelo povoamento do Brasil[43]. **Fonte:** De autoria própria.

O projeto Saúde, Bem-estar e Envelhecimento (SABE) iniciou-se em 2000 e incluía 1.171 amostras de indivíduos da cidade de São Paulo, todos com mais de 60 anos de idade. Contudo, atualmente, possui 1.324 de indivíduos. Todas as amostras apresentam sequenciamento de genoma completo (do inglês, *Whole-Genome Sequencing* – WGS) de alta cobertura, processadas utilizando a plataforma ilumina HiSeq X[49]. As informações presentes nesses dados possibilitam estudos genômicos diversos, como levantamento de variantes como SNPs e InDels.

A análise dos genes *HLA-DOA* e *HLA-DOB* em populações reais pode revelar a presença de novos alelos e novas associações com doenças.

2. JUSTIFICATIVA

Os genes *HLA-DOA* e *HLA-DOB* possuem funções críticas no processo de processamento e apresentação de antígenos e resposta imunológica mediada por células T. Por não serem expressos na superfície celular, diferentemente dos genes HLA clássicos, eles não são avaliados durante transplantes. Além disso, esses genes não foram avaliados em populações reais, principalmente aquelas com diferentes ancestralidades genéticas. Nesse sentido, a diversidade conhecida para esses genes pode estar extremamente subestimada. Utilizando e combinando os dados de diferentes bancos de dados e uma metodologia com ferramentas mais adequadas, buscamos avaliar e caracterizar a diversidade alélica de ambos os genes em torno do mundo.

3. OBJETIVO

3.1 Objetivo Geral

Avaliar a diversidade genética dos genes *HLA-DOA* e *HLA-DOB* em amostras de populações mundiais (América, Europa, Ásia e África), comparando as frequências observadas em cada região demográfica.

3.2 Objetivos Específicos

- Consolidar a metodologia para avaliação de todo o gene *HLA-DOA* e *HLA-DOB* por sequenciamento massivo paralelo de segunda geração;

- Detectar os SNPs e InDels existentes ao longo dos genes *HLA-DOA* e *HLA-DOB* e os haplótipos formados por essas variantes (alelos HLA-DOA ou HLA-DOB);

- Avaliar as frequências dos alelos nas diferentes regiões demográficas;

- Averiguar o perfil de desequilíbrio de ligação entre os alelos dos genes *HLA-DOA* e *HLA-DOB*;

- Avaliar o perfil de seleção natural atuando nos genes estudados

REFERÊNCIAS BIBLIOGRÁFICAS

1. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* **37**, 724–737 (2016).
2. Jiang, W. *et al.* pH-susceptibility of HLA-DO tunes DO/DM ratios to regulate HLA-DM catalytic activity. *Sci. Rep.* **5**, 17333 (2015).
3. Delamarre, L., Pack, M., Chang, H., Mellman, I. & Trombetta, E. S. Differential Lysosomal Proteolysis in Antigen-Presenting Cells Determines Antigen Fate. *Science* **307**, 1630–1634 (2005).
4. West, L. C. & Cresswell, P. Expanding roles for GILT in immunity. *Curr. Opin. Immunol.* **25**, 103–108 (2013).
5. Chang, Z. L. *et al.* Rewiring T-cell responses to soluble factors with chimeric antigen receptors. *Nat. Chem. Biol.* **14**, 317–324 (2018).
6. Brown, B. *et al.* Immunopathogenesis of Nipah Virus Infection and Associated Immune Responses. *Immuno* **3**, 160–181 (2023).
7. Wieczorek, M. *et al.* Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* **8**, 292 (2017).
8. Douillard, V. *et al.* Approaching Genetics Through the MHC Lens: Tools and Methods for HLA Research. *Front. Genet.* **12**, 774916 (2021).
9. Álvaro-Benito, M. & Freund, C. Revisiting nonclassical HLA II functions in antigen presentation: Peptide editing and its modulation. *HLA* **96**, 415–429 (2020).
10. Wu, Y., Zhang, N., Hashimoto, K., Xia, C. & Dijkstra, J. M. Structural Comparison Between MHC Classes I and II; in Evolution, a Class-II-Like Molecule Probably Came First. *Front. Immunol.* **12**, (2021).

11. Harari, O. & Liao, J. K. Inhibition of MHC II Gene Transcription by Nitric Oxide and Antioxidants. *Curr. Pharm. Des.* **10**, 893–898 (2004).
12. Sarri, C. A. *et al.* Amino acid signatures in the HLA class II peptide-binding region associated with protection/susceptibility to the severe West Nile Virus disease. *PLoS ONE* **13**, e0205557 (2018).
13. Naruse, T. k. *et al.* Limited polymorphism in the HLA-DOA gene. *Tissue Antigens* **53**, 359–365 (1999).
14. Welsh, R., Song, N. & Sadegh-Nasseri, S. What to do with HLA-DO/H-2O two decades later? *Immunogenetics* **71**, 189–196 (2019).
15. Guce, A. I. *et al.* HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. *Nat. Struct. Mol. Biol.* **20**, 90–98 (2013).
16. Mellins, E. D. & Stern, L. J. HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Curr. Opin. Immunol.* **26**, 115–122 (2014).
17. Yin, L., Maben, Z. J., Becerra, A. & Stern, L. J. Evaluating the Role of HLA-DM in MHC Class II–Peptide Association Reactions. *J. Immunol.* **195**, 706–716 (2015).
18. Shapiro, I. E. & Bassani-Sternberg, M. The impact of immunopeptidomics: From basic research to clinical implementation. *Semin. Immunol.* **66**, 101727 (2023).
19. Santambrogio, L. Molecular Determinants Regulating the Plasticity of the MHC Class II Immunopeptidome. *Front. Immunol.* **13**, 878271 (2022).
20. Roche, P. A. & Furuta, K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* **15**, 203–216 (2015).

21. Pishesha, N., Harmand, T. J. & Ploegh, H. L. A guide to antigen processing and presentation. *Nat. Rev. Immunol.* **22**, 751–764 (2022).
22. Silva, N. D. S. B. *et al.* Immunogenetics of HLA-B: SNP, allele, and haplotype diversity in populations from different continents and ancestry backgrounds. *HLA* **101**, 634–646 (2023).
23. Hernandez, P. V. *et al.* HLA-B evolutionary divergence is associated with outcomes after SARS-CoV-2 infection. *Hum. Immunol.* **83**, 803–807 (2022).
24. Quintana-Murci, L. Human Immunology through the Lens of Evolutionary Genetics. *Cell* **177**, 184–199 (2019).
25. Meyer, D., C Aguiar, V. R., Bitarello, B. D., C Brandt, D. Y. & Nunes, K. A genomic perspective on HLA evolution. *Immunogenetics* **70**, 5–27 (2018).
26. Lenz, T. L., Spirin, V., Jordan, D. M. & Sunyaev, S. R. Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. *Mol. Biol. Evol.* **33**, 2555–2564 (2016).
27. van Oosterhout, C. A new theory of MHC evolution: beyond selection on the immune genes. *Proc. R. Soc. B Biol. Sci.* **276**, 657–665 (2009).
28. Flomenberg, N. *et al.* Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood* **104**, 1923–1930 (2004).
29. Lee, S. J. *et al.* High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood* **110**, 4576–4583 (2007).
30. Tiercy, J.-M. How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica* **101**, 680–687 (2016).

31. Hurley, C. K. *et al.* Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. *HLA* **95**, 516–531 (2020).
32. Olieslagers, T. I. *et al.* New insights in HLA-E polymorphism by refined analysis of the full-length gene. *HLA* **89**, 143–149 (2017).
33. Gourraud, P.-A. *et al.* HLA Diversity in the 1000 Genomes Dataset. *PLOS ONE* **9**, e97282 (2014).
34. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
35. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
36. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
37. Greely, H. T. Human genome diversity: what about the other human genome project? *Nat. Rev. Genet.* **2**, 222–227 (2001).
38. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
39. Cavalli-Sforza, L. L. The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* **6**, 333–340 (2005).
40. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
41. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
42. Marian, A. J. Clinical Interpretation and Management of Genetic Variants. *JACC Basic Transl. Sci.* **5**, 1029–1042 (2020).

43. IBGE | Brasil: 500 anos de povoamento | território brasileiro e povoamento. https://brasil500anos.ibge.gov.br/territorio-brasileiro-e-povoamento.html.
44. Souchaud, S. & Fusco, W. População e ocupação do espaço: o papel das migrações no Brasil. *Redes* **17**, 5–17 (2012).
45. Costa, S. A construção sociológica da raça no Brasil. *Estud. Afro-Asiáticos* **24**, 35–61 (2002).
46. Vieira, F. S. da S. As ciências sociais e as etnias africanas no Brasil: análise crítica da construção do projeto de identidade nacional. *Social sciences and African ethnic groups in Brazil: a critical analysis of the construction of the project of national identity* (2010).
47. Truzzi, O. Sírios e libaneses no oeste paulista – décadas de 1880 a 1950. *Rev. Bras. Estud. Popul.* **36**, 1–27 (2019).
48. Escher, L. M. *et al.* Challenges in selecting admixture models and marker sets to infer genetic ancestry in a Brazilian admixed population. *Sci. Rep.* **12**, 21240 (2022).
49. Naslavsky, M. S. *et al.* Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. *Nat. Commun.* **13**, 1004 (2022).

Capítulo II - Artigo

Short communication

HLA-DOA and *HLA-DOB* genetic diversity in worldwide populations

Isabelle Mira da Silva[1,2], Nayane dos Santos Brito Silva[1,2,3], Diogo Meyer[4], Erick C. Castelli[1,2,5]

Filiations

[1]São Paulo State University, Molecular Genetics and Bioinformatics Laboratory – Experimental Research Unity, School of Medicine, Botucatu, State of São Paulo, Brazil. [2]São Paulo State University, Biological Science (Genetics) Program, Bioscience Institute of Botucatu, State of São Paulo, Brazil. [3]Nantes Université, INSERM, Ecole Centrale Nantes, Center for Research in Transplantation and Translational Immunology, UMR 1064, F-44000 Nantes, France. [4]Institute of Biosciences, University of São Paulo (USP), Sao Paulo, SP, Brazil. [5]São Paulo State University (UNESP), Pathology Program, School of Medicine, Botucatu, State of São Paulo, Brazil.

Contact
Erick da Cruz Castelli
Pathology Department, School of Medicine, UNESP – Botucatu, SP
CEP: 18618-687
Phone: 3880 – 1696
E-mail address: erick.castelli@unesp.br

Abstract

The HLA-DOA and HLA-DOB genes encode the two subunits that comprise the HLA-DO molecule, which plays a critical role in mediating antigen processing and loading in class II HLA molecules. These genes exhibit lower polymorphism compared to other HLA genes. However, these polymorphisms can significantly influence susceptibility to pathogen infections and the development of autoimmune diseases. In this study, we analyzed 5,349 samples from diverse populations worldwide, including Brazil. Using a bioinformatics pipeline tailored for the evaluation of HLA genes, we characterized the genetic diversity of *HLA-DOA* and *HLA-DOB*, identifying previously unknown variants. Both genes presented a few different alleles, which are restricted to specific biogeographical regions. For instance, *DOA*01:01:11* and *DOB*01:03:01* were observed in African, European, and American (both North and South) populations. Additionally, we assessed allele frequency profiles for each biogeographical region. Despite the inclusion of data from the Brazilian population, characterized by extensive admixture, these genes appeared to be highly conserved, particularly at the protein isoform level, with *DOA*01:01* and *DOB*01:01* being the most prevalent for both genes. Moreover, at both the DNA and protein sequence levels, these genes demonstrated high conservation compared to neighboring HLA genes, potentially driven by evolutionary pressures maintaining the structure of the HLA-DO molecule due to its critical biological function.

Keywords: *HLA-DOA* and *HLA-DOB*, genetic variability, linkage disequilibrium, nucleotide diversity.

Introduction

The HLA class II antigen presentation pathway is a key feature of the adaptive immune response against microorganisms. In this pathway, foreign antigens are internalized, processed, and lately presented at the cell surface by the classical HLA class II molecules of Antigen-Processing Cells (APCs). The HLA class II molecules + antigens interact with the T cell receptor (TCR) from T lymphocytes and this interaction might activate T cell response. HLA-DO (DO) is a "backstage" molecule, acting together with HLA-DM (DM) in the selection of antigens to be loaded to the classical HLA class II molecules[1,2].

Before peptides were displayed on the cell surface, many processes and cellular components were involved to process and select these peptides. Firstly, when an exogenous protein is endocytosed, it is processed into peptides by lysosomes and subsequently addressed to the Antigen Processing Compartment. There, these peptides encounter nascent HLA class II molecules, that are associated with an Invariant Chain Peptide (CLIP)[3]. A complex formed by the association of HLA-DO and HLA-DM (DO-DMc) removed the CLIP and selected peptides to be loaded into the HLA class II molecule[4], as demonstrated in **Figure 1**.

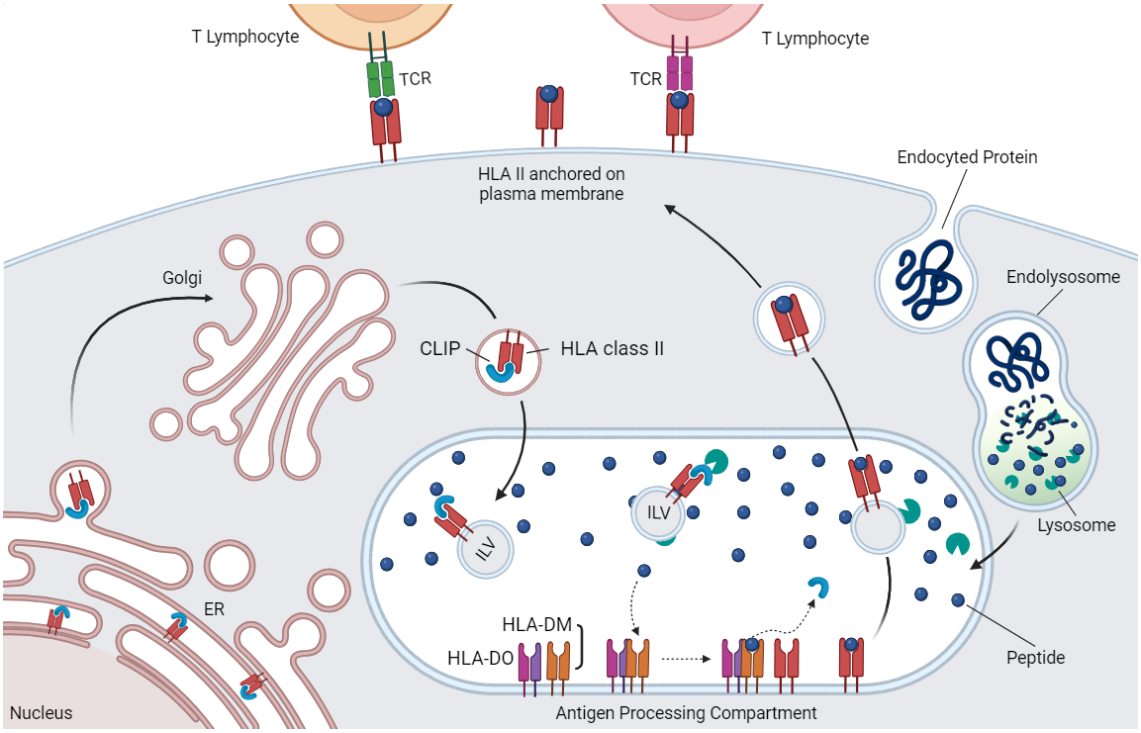

Figure 1. Representation of the processing and loading of HLA class II peptides mediated by HLA-DO and HLA-DM in an APC. **Based on:** ROCHE & FURUTA[3].

The HLA-DO molecule is formed by two subunits: alpha (DOα), encoded by *HLA-DOA*, and beta (DOβ), encoded by *HLA-DOB*. Numerous studies demonstrate a link between certain diseases and the malfunction of the DO-DM complex, particularly HLA-DO, such as viral infections (e.g., Chronic Hepatitis B[5] and COVID-19[6]) and autoimmune diseases (e.g., Type 1 Diabetes[7]). *HLA-DOA* and *HLA-DOB* are encoded at the human Major Histocompatibility Complex (MHC) and are surrounded by highly polymorphic genes, including the ones encoding the HLA class II molecules.

Despite this, both these genes are highly conserved in terms of polymorphisms. In 1999, Naruse and colleagues[8,9] evaluated the allelic diversity for both genes in 37 HLA class II homozygous cells. They detected a few different sequences for both genes, with nucleotide exchanges in exons and introns, and some non-synonymous mutations in HLA-DOB, but none of these modifications altered the HLA-DO function and its interaction with HLA-DM.

According to the IPD-IMGT/HLA Database[10], which reports *HLA* sequences and serves as the standard repository for HLA research, *HLA-DOA* currently

comprises ninety-two allele sequences (version 3.53.0, July 2023), with ninety full-length sequences (from 5'UTR to 3'UTR) and expressing 14 different protein isoforms. *HLA-DOB* has sixty-three allele sequences, with sixty-two being full-length and encoding 12 protein sequences (version 3.53.0, July 2023). As widely known from different studies, HLA genes experience strong pressure to increase the diversity of the molecules to deal with different peptides, a process described as balancing selection[11].

In contrast, HLA-DO is a highly conserved molecule with unclear selective pressures and a lack of information about why and how this protein maintains this characteristic. However, as speculated by He (2022)[12], due to the influence of HLA-DO and HLA-DM in antigen processing and following the evolutionary perspective of HLA Class II, both genes might have undergone processes that led to the establishment of their functions. Despite that, it is still notable that there is a significant gap in the literature regarding the HLA-DO, *HLA-DOA*, and *HLA-DOB* in real populations, particularly in admixed ones such as Brazilians.

Because HLA-DO is not genotyped before transplantation, few studies explored its diversity in real population samples, and the ones available explored mostly exons. In addition, since HLA-DO is not routinely genotyped, laboratories around the world do not explore (and submit) new sequences to the IPD-IMGT/HLA database. Therefore, HLA-DO diversity (and both genes encoding it) might be extremely underestimated, particularly in admixed populations. Here, we applied a bioinformatic pipeline tailored to HLA genes to survey *HLA-DOA* and *HLA-DOB* genetic diversity across the world.

Samples and Methods

We analyzed the polymorphisms across *HLA-DOA* and *HLA-DOB* in 5,349 individuals from different populations available in open access repositories (**Table S1**), which includes 3,197 individuals from the 1000genomes project[13], 828 from the Human Genome Diversity Project (HGDP)[14], and 1,324 Brazilians from São Paulo city (the SABE cohort)[15]. All these samples present high-

coverage 30X whole-genome sequencing (WGS). We downloaded the BAM files from each sample, with reads aligned to the hg38 human reference genome.

The workflow for read alignment, variant and allele calling is represented in **Figure 2**. The starting point for all samples was the downloaded BAM file with reads aligned to the reference genome hg38. To optimize the read alignment for HLA genes and to minimize alignment errors and cross-alignments, we employed *hla-mapper* version 4.1[16], using the BAM files as input. *hla-mapper* is available at www.castelli-lab.net/apps/hla-mapper.

Subsequently, we called variants using the Genome Analysis Toolkit (GATK) HaplotypeCaller (version 4.0.9)[17] in the GVCF (genomic variant call format file) mode, joining all into a single multi-sample GVCF file. Then, we called SNPs and InDels with GATK GenotypeGVCFs to generate a multi-sample VCF (variant call format file), annotating all variants with dbSNP version 150. We post-processed the VCF file to remove false positives and artifacts, using GATK VQSR (Variant Quality Score Recalibration)[18]. To determine the phase among closely located variants from the sequencing data, we utilized WhatsHap[19]. Then, we employed Shapeit4[20] to infer the complete haplotypes, considering the phase sets determined by *WhatsHap*. The outcome is phased VCF file. The genotype workflow is described at https://github.com/erickcastelli/hla-mapper.

To call *HLA-DOA* and *HLA-DOB* alleles, we applied a Perl script that generates the genomic sequences for each gene based on the phased SNPs, one for each chromosome, comparing these sequences with the ones at the IPD-IMGT/HLA database. We exported the complete sequences (exons + introns) and the CDS sequence (only exons, from the translation start codon to the stop codon). The CDS sequences were translated into a protein sequence using EMBOSS. For *HLA-DOA*, we focused on the region chr6:33,002,182-33,011,591 (hg38). For *HLA-DOB*, we focused on the region chr6:32,810,763-32,819,002 (hg38). In both cases, the sequence comprises all exons and introns along with 2,000 nucleotides upstream and downstream each locus.

We tracked the allele frequencies in three levels, genomic sequences (also known as 4-field allele resolution), CDS sequences (3-field allele resolution), and protein sequences (two-field allele resolution), in each biogeographic region: Europe, Africa, Central South Asia, East Asia, South Asia, Middle East, Oceania, and America. We also assessed the Linkage Disequilibrium (LD) pattern across *HLA-DOA* and *HLA-DOB* using *Haploview 4.2*[21] to create LD plots according to the following parameters: sites with minimum allele frequency (MAF) threshold of 2% and Hardy-Weinberg *p-value* cutoff of 0.0001%.

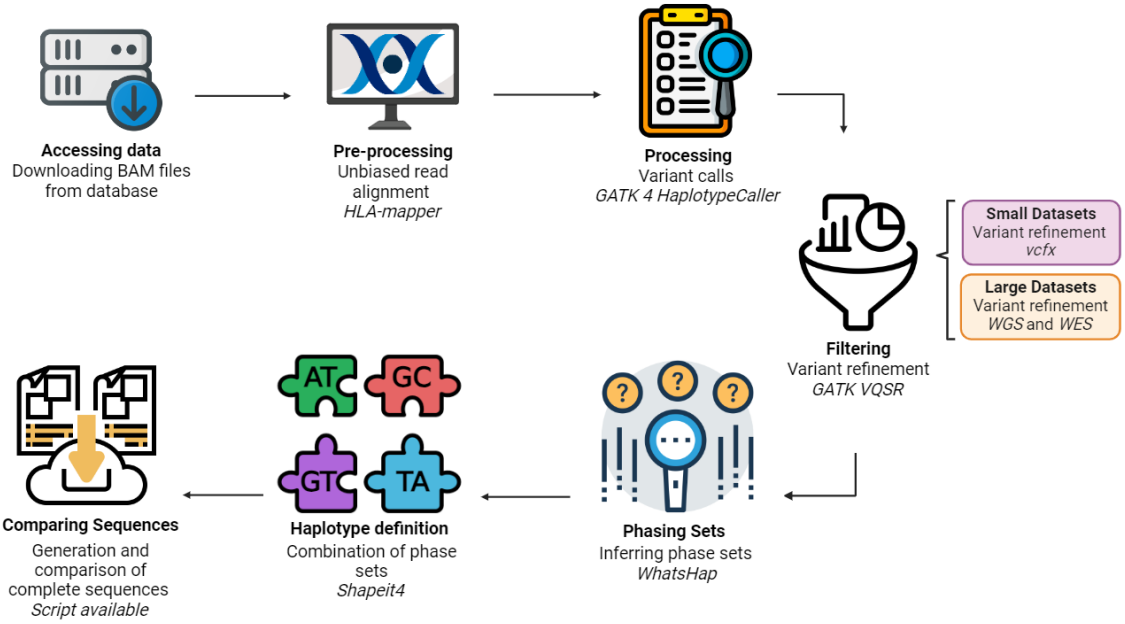

Figure 2. Scheme of the HLA genotyping workflow, from the original BAM file to the phased VCF.

The nucleotide diversity was evaluated using Variscan[22].To do that, we first aligned all sequences (FASTA file) using *mafft*[23] and we transformed the FASTA file into a PHYLIP-format using an in-house script. We established the following parameters: 1) start position included the first nucleotide of the sequence in 5'UTR (StartPos = 0) and the endpoint, the last nucleotide of 3'UTR (EndPos = 0); 2) the type of analysis considered was RunMode equal 12, that included statistical analysis of Pi and Tajima's D; 3) we specify the number of the total number of segregating sites or mutation through UseMuts equal 1; and 4) the window size or WidthSW was 100 with 1 nucleotide as jump (JumpSW and SlidingWindow). Therefore, we evaluated nucleotide diversity in windows of 100 nucleotides and a step size of 1 nucleotide.

Results

We detected 276 SNVs throughout *HLA-DOA*, which 235 are in intronic regions and untranslated regions and 41 in exonic regions (**Figure 3**). Of the 41 SNVs in exons, are described 1 frameshift, 16 non-coding and 24 missense variants with all of them being rare (**Table S1**).

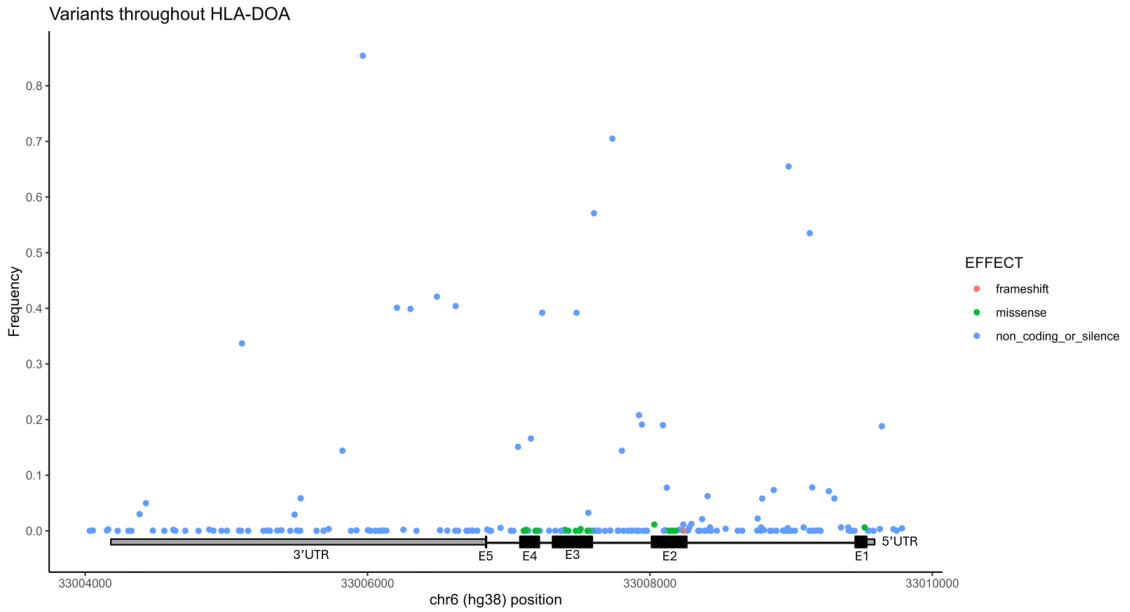

Figure 3. SNVs across *HLA-DOA* gene. Exons 1 to 5 are depicted by black boxes, with intronic regions indicated by black lines between these boxes. The 5'UTR is shown as smaller gray boxes, and 3'UTR region is depicted as larger grey box. Non-synonymous, intronic, or non-coding variants are represented in blue; missense variants are represented in green; and frameshift variants are represented in red.

The haplotypes formed by these SNVs configure the *HLA-DOA* alleles. We first tracked the four-field allele frequencies (exons + introns + untranslated regions). The comparison with the IPD-IMGT/HLA database revealed novel sequences with a frequency of 6.73% of 276 of SNVs. These novel alleles are mostly related to intronic nucleotide exchanges, and they were quite frequent in Africa (14.72%) and in the Middle East (14%) (**Figure 4**). We detected 58 of 276 of sequences were already documented in the IPD-IMGT/HLA database (**Table S3**). Of those, 12 presented a higher frequency than 2% in at least one of the studied populations, with *DOA*01:01:01:01* and *DOA*01:01:02:01*, being the most frequent ones in most regions (Figure 4). There is a clear differentiation

between populations in terms of allele frequencies, and some alleles are population-specific (**Table S1**).

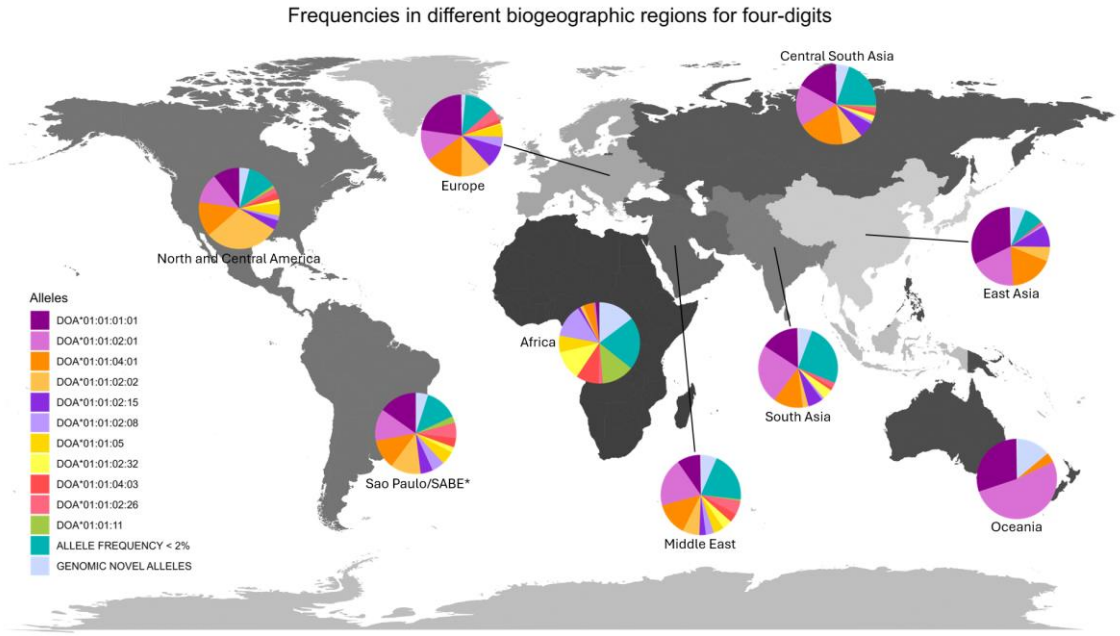

Figure 4. *HLA-DOA* four-field allele frequencies in different biogeographic regions. We plotted Brazil separately because the Brazilian data came from a different cohort (SABE/Brazil).

When tracking the three-field allele frequencies (only exons), we identified 45 different alleles, 19 of them already documented in the IPD-IMGT/HLA database (version 3.53). The most prevalent allele was *DOA*01:01:02* (51.69%) in all regions except in East Asia (**Figure 5** and **Table S4**). It is notable that allele diversity in Africa is higher, with many different alleles. The remaining 26 sequences are potential novel and uncommon alleles, with a summed frequency of 0.68%. However, these novel alleles are frequent in Oceania (2.00%), Africa (1.62%), and South Asia (1.33%). Nevertheless, the low proportion of new alleles considering exons (0.68%) as compared to the complete sequences (6.73%) clearly indicates that most new alleles are related to new intronic variants.

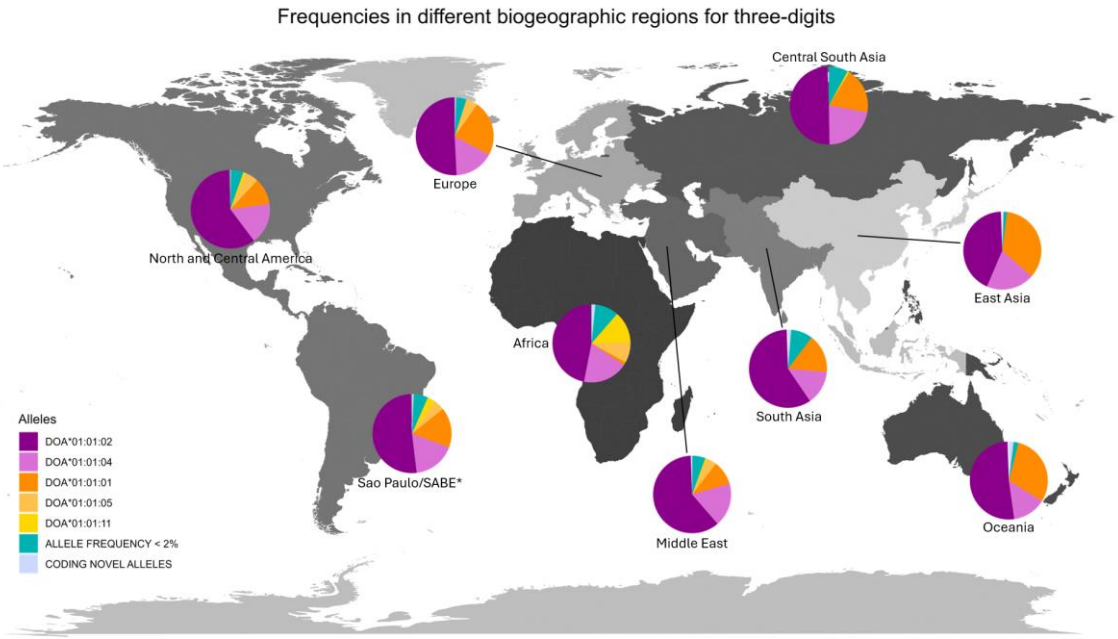

Figure 5. Worldwide frequency of HLA-DOA three-field alleles in different biogeographic regions, Africa, America, Central South Asia, East Asia, Europe, Middle East, Oceania, and South Asia, surveyed from more than 5,000 individuals.

We predicted the HLA-DOA protein sequences (two-field allele) by translating the CDS sequence, and there were 27 different alleles. Of those, 17 are novels and present very low frequencies (**Table S5**). The most frequent allele is *DOA*01:01*, with a summed frequency of 97.19% worldwide (**Figure S1**). Nevertheless, exploring real population samples across the world revealed new HLA-DOA versions, that might be frequent in other populations not explored in this survey.

There is a high LD among variants of the second half of the 3'UTR and among variants from exons 2 to 5, which encode the mature protein (**Table S9** and **Figure S3**). Regarding the r^2 parameter, it is noteworthy that specific allele combinations show a statistically significant association, as shown in **Figure S4**. Therefore, it would be possible to predict exonic variants with intronic ones, which might be explored for imputation purposes.

For *HLA-DOB*, we found 194 SNVs across the gene (**Table S2**), which included 116 in introns and untranslated regions and 30 in exons (1 frameshift, 1 stop-codon gain, and 9 non-coding or silence, and 19 missenses). Most of the missense variants present a frequency lower than 0.5%. Only two missense

variants are frequent (about 9%), rs2071554 in exon 1 and rs2070121 in exon 4 (**Figure 6**).

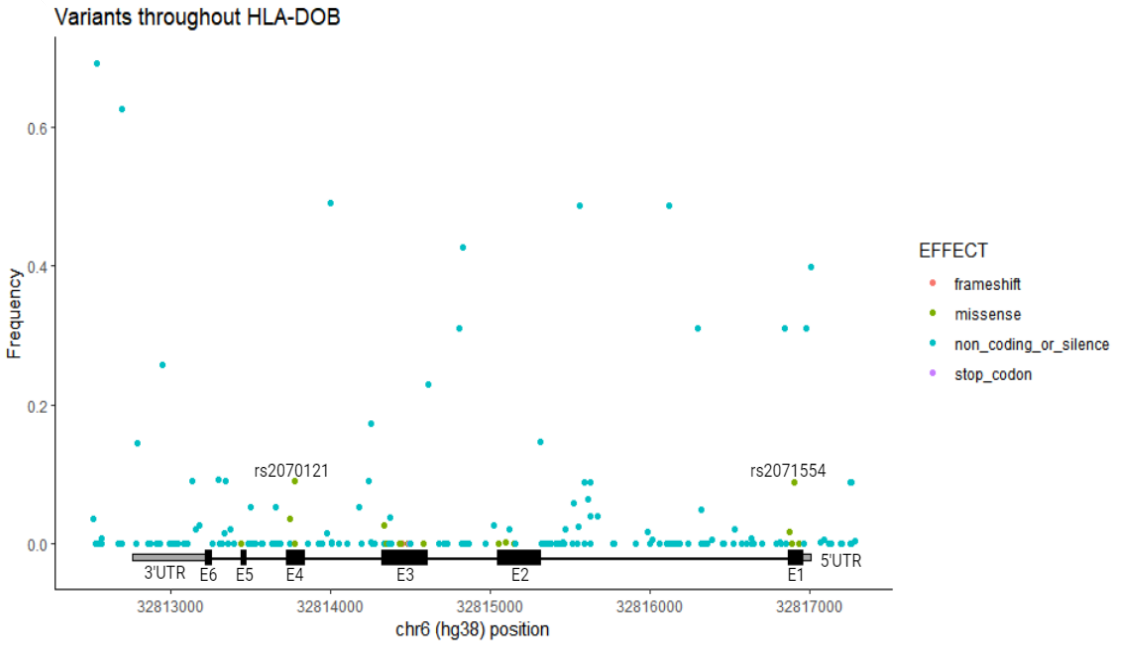

Figure 6. SNVs across *HLA-DOB* gene. Exons 1 to 6 are depicted by black boxes, with intronic regions indicated by black lines between these boxes. The 5'UTR is shown as smaller gray boxes, and 3'UTR region is depicted as larger grey box. Non-synonymous, intronic, or non-coding variants are represented in blue; missense variants are represented in green; and frameshift variants are represented in red.

The haplotypes formed by these SNVs (four-digit alleles) represent 208 alleles, 49 that had already been documented in the IPD-IMGT/HLA database and 159 novels (**Table S6**). Allele *DOB*01:01:01:01* is the most frequent one worldwide (18,03%) and the most prevalent in Africa, Europe, and America. Allele *DOB*01:01:01:24* has a global frequency of 15.23% and is the most prevalent in Oceania and East Asia (**Figure 7**). The summed frequency of all novel alleles is 11.89%, and they were very frequent in Africa (29%). It is not clear why we detected so many novel alleles, but these new alleles are mostly related to new intronic mutations. This high proportion is possibly linked to the lack of studies addressing HLA-DO full variability, particularly in under-represented populations in public datasets, such as Africans and Americans.

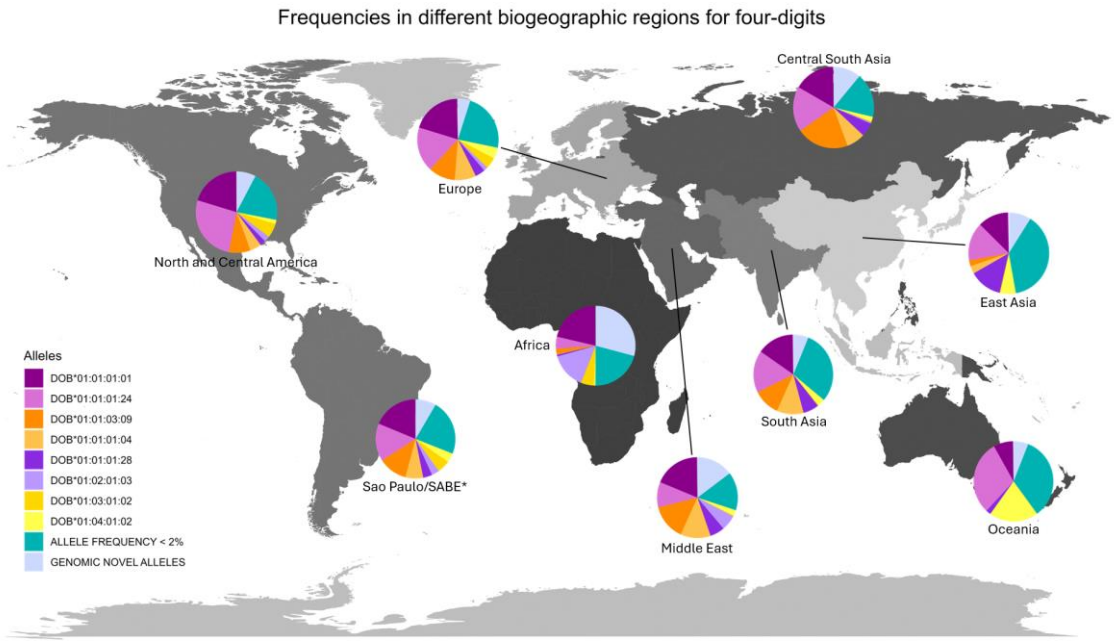

Figure 7. HLA-DOB four-field allele frequencies in different biogeographic regions, Africa, America, Central South Asia, East Asia, Europe, Middle East, Oceania, and South Asia, surveyed from more than 5,000 individuals.

When examining only the CDS sequences (three-field alleles), we identified 34 alleles, 14 of them already documented in the IPD-IMGT/HLA and 20 potentially novel alleles (**Table S7**). Allele *DOB*01:01:01* is the most frequent one (61.95%) followed by *DOB*01:01:03* (12.01%) and *DOB*01:02:01* (6.64%), **Figure 8**. *DOB*01:02:01* encodes a different protein sequence than *DOB*01:01:01* and *DOB*01:01:03*. Because the novel alleles are related to intronic mutations, novel 3-field alleles were not actually frequent.

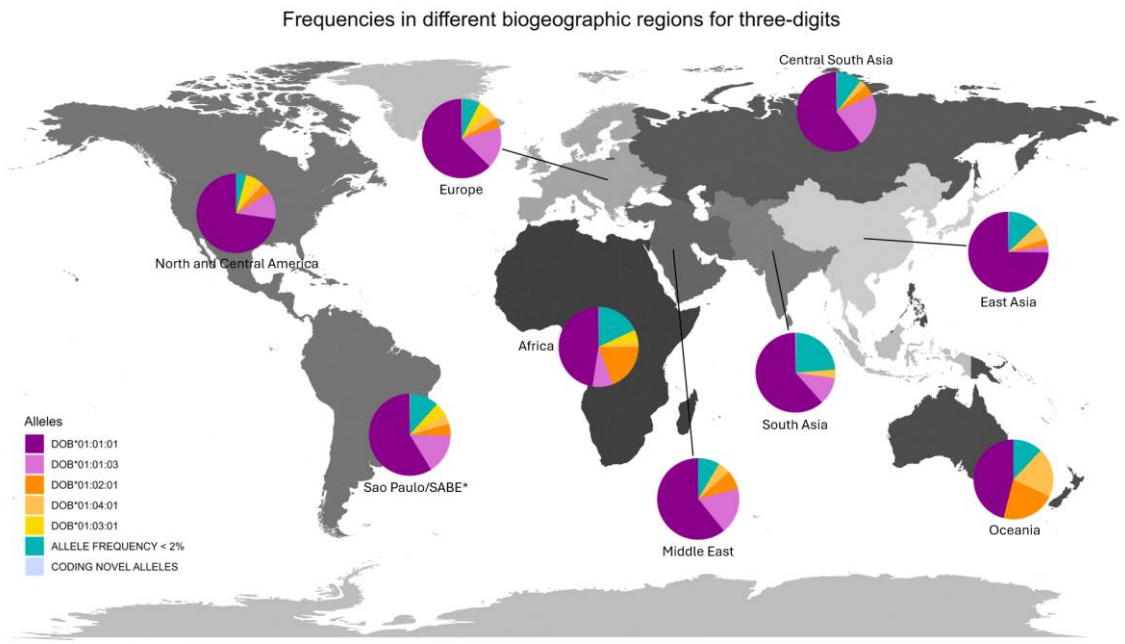

Figure 8. Worldwide frequency of HLA-DOB three-field alleles in different biogeographic regions, Africa, America, Central South Asia, East Asia, Europe, Middle East, Oceania, and South Asia, surveyed from more than 5,000 individuals.

By translating the CDS sequences into protein sequences, we detected only 21 different alleles, 11 of them potentially novel (**Table S8**). *HLA-DOB* is much more polymorphic than *HLA-DOA* in terms of protein sequences. At least five versions are frequent, with a global frequency higher than 1%: DOA*01:01 (the most prevalent everywhere), **01:02*, **01:03*, **01:04*, and **01:05*. Nevertheless, in general, *DOA*01:01* represents about 75% of the alleles. *DOB*01:04* and *DOB*01:02* are also frequent in all regions. *DOB*01:03* is frequent in Africa, América and Europe. DOB*01:05 is frequent in Europe, South Asia and the Middel East (**Figure S2**). New alleles are also rare because the majority of the new SNVs occur in introns.

There is also a strong LD across *HLA-DOB*, particularly among the exonic variants. Contrary to observations with the D' parameter, analysis of the r^2 (shown in **Figure S5** and **Figure S6**) reveals statistically significant associations within the previously described blocks, demonstrating a high or "perfect" correlation between variants across them. For instance, in HLA-DOA, significant correlations are noted between rs13205975 and rs13206015 (block 1), as detailed in **Table S10** and **Figures S5** or **S6**. In HLA-DOB, significant correlations are observed

between rs2070120 and rs17213693 (block 2), rs2070121 and rs16870880, all alleles within block 3, and rs7742120 and rs7762120 (block 4).

The nucleotide diversity across *HLA-DOA* is represented in **Figure 9**. The region intron 2 until intron 3 and a portion of 3'UTR showed relatively high diversity when compared to the average of MHC (the blue line) and the observed for the entire chr6. Therefore, at least for introns, HLA-DOA is highly polymorphic, with a high number of low-frequency variants. For exons, nucleotide diversity is low compared to the rest of the MHC or chromosome 6, particularly for exons 2 and 4. On the other hand, the pattern observed for *HLA-DOB* is different, with very low nucleotide diversity indexes even in introns (**Figure 10**). These results support the claim that HLA-DOA and HLA-DOB are indeed conserved in terms of sequence despite being surrounded by the most polymorphic genes in the human genome.

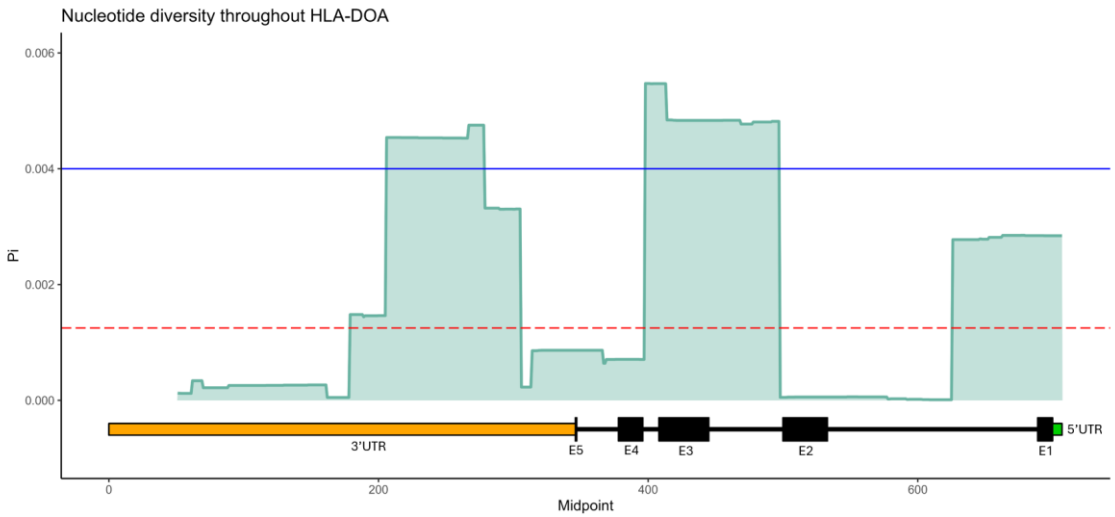

Figure 9. Nucleotide diversity throughout *HLA-DOA* in windows of 100 nucleotides and step-size of 1, for all samples used in this study. The *HLA-DOA* gene structure is represented at the bottom of the plot, following the same orientation in the hg38 genome reference and scale proportion. The 5'UTR and 3'UTR are represented by green and orange squares, respectively. The exonic regions are represented by the tallest black squares (E1-5) and the intronic regions by the lines between the black squares. Nucleotide diversity levels above the blue line are higher than 99.9% of the ones observed at the MHC region. Nucleotide diversity levels above the red line are higher than the average observed for chromosome 6.

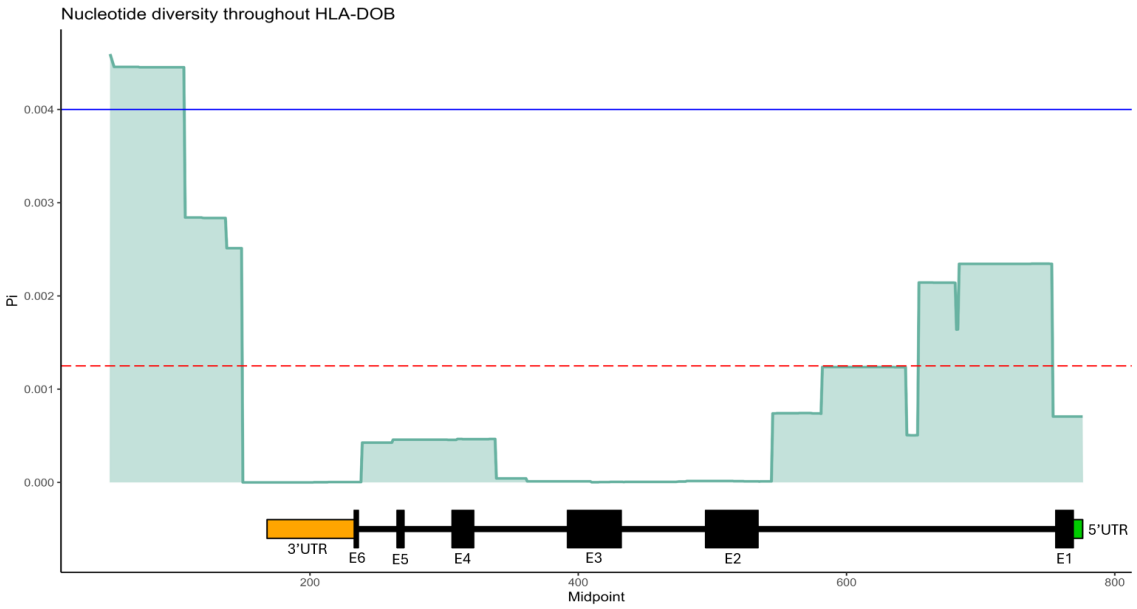

Figure 10. Nucleotide diversity throughout *HLA-DOB* in windows of 100 nucleotides and step-size of 1, for all samples used in this study. The *HLA-DOB* gene structure is represented at the bottom of the plot, following the same orientation in the hg38 genome reference and scale proportion. The 5'UTR and 3'UTR are represented by green and orange squares, respectively. The exonic regions are represented by the tallest black squares (E1-6) and the intronic regions by the lines between the black squares. Nucleotide diversity levels above the blue line are higher than 99.9% of the ones observed at the MHC region. Nucleotide diversity levels above the red line are higher than the average observed for chromosome 6.

Discussion

HLA-DOA and *HLA-DOB* are not evaluated before transplantation, possibly due to their function not being directly linked to peptide presentation. Therefore, their genetic variability might be underestimated. However, after evaluating thousands of samples from different populations, including populations with various genetic ancestries and that have undergone certain evolutionary processes related to their specific environment or social organization, we demonstrated that this gene is highly conserved in terms of sequence and the protein sequence they encode. There are indeed hundreds of new alleles that were not previously described, but the vast majority is related to new intronic mutations that possibly do not interfere with the gene function.

Nonetheless, looking at the genomic sequences, we noticed that a higher number of novel alleles were found, which made us wonder if this result could be

linked to the lack of studies involving the entire gene and real population samples. Then, we can speculate that the novel alleles found correspond to polymorphisms that have not yet been catalogued, and which are potentially related to a variety of samples, especially Brazilian ones, and/or a more adequate *pipeline*, including a different tool *HLA-mapper*, that reduced bias of alignments and cross-mapping of *reads* from HLA genes.

When we analyzed the *HLA-DOA*, we observed a highly conserved sequence that even with a considerable number of polymorphisms, they were majority in the intronic region or synonymous. Though most variants are uncommon, some exonic ones were frequent, such as *rs2070121* and *rs2071554* for *HLA-DOB*, reaching frequencies of approximately 9%. However, for *HLA-DOB*, both variants are non-synonymous, with *rs2071554* related to an amino acid exchange (Arg>Leu) at β2 domain, and *rs2070121* an amino acid exchange (Leu>Phe) in the transmembrane domain. Based on the first possible amino acid alteration, we wondered if it could affect the cytoplasmic trafficking of the molecule to organelles in its pathway, such as Golgi Complex or Antigen Processing Compartment. In case of the second polymorphism, due to the region of occurrence, the possible impact could be on the attachment of the molecule or the interaction of Metalloproteinases (MMPs) that cleave and release HLA-DOB, potentially disarranging HLA class II presentation processes. Despite *DOB*01:01* being the most frequent allele around the world, *DOB*01:04* and *DOB*01:02* presented significant frequency.

Besides different coding alleles being found for both genes, we noticed that *DOA*01:01:02* (more frequent worldwide)*, DOA*01:01:04,* and *DOA*01:01:01* were presented in all biogeographic regions. However, *DOA*01:01:05* was only detected in Africa, America (including Brazil), Europe, and the Middle East, and absent in South Asia, East Asia, and Oceania. Considering the frequency of this allele in Africa and Brazil (represented by “São Paulo/SABE” in **Figure 5**), we postulate that *DOA*01:01:05* was inserted into the Brazilian population by the genome exchange through the Brazil colonization period. The same situation can be observed by *HLA-DOB*, with *DOB*01:01:01* and *DOB*01:01:03* present in all populations analyzed. On the other hand, *DOB*01:03:01* was found with higher

frequencies in Africa, Europe, North and Central America and Brazil. Based on this, DOB**01:03:01* could be an allele presented in almost all heterozygote individuals, which facilitated the process of insertion or increase of frequency of this allele in America populations.

When analyzed according to biogeographic regions, Africa and Oceania demonstrated higher frequency, with approximately 20%. As well known, *DOB*01:02* presents an amino acid exchange at the peptide leader, that potentially interferes with protein transit inside the cell, in this case preventing the molecule departure from Endoplasmic Reticulum. Nevertheless, assuming a potential modification in HLA-DOB function due to the presence of these different isoforms, most of the individuals present at least one copy of an allele encoding *DOB*01:01* and most of the time two copies of an allele encoding HLA-*DOA*01:01*, therefore maintaining the HLA-DO function.

Conclusion

In conclusion, our comprehensive analysis of *HLA-DOA* and *HLA-DOB* on a global scale has shed light on the inner genetic diversity of both genes. Therefore, it became evident that the HLA-DOA is highly conserved in terms of different protein sequences since almost every individual expresses the same HLA-DOA molecule, unless some alternative splicing is in place. Using high-coverage whole-genome sequencing samples from diverse populations and advanced bioinformatics tools, we uncovered previously unrecognized aspects of the genetic and protein sequences of these genes, but we also revealed hundreds of potentially novel alleles. Our findings demonstrated the remarkable conservation of HLA-DOA and HLA-DOB in terms of nucleotide and protein sequences, even in admixed populations such as Brazilians.

Reference

1. YIN, L., MABEN, Z., BECERRA, A., STERN, L. J. 2015. Evaluating the role of HLA-DM in MHC II-peptide association reactions. **J. Immunol.** 195(2): 706-719. DOI: https://doi.org/10.4049/jimmunol.1403190.

2. SHAPIRO, I. E., BASSANI-STERNBERG, M. 2023. The impact of immunopeptidomics: From basic research to clinical implementation. 66: 101727. DOI: https://doi.org/10.1016/j.smim.2023.101727.

3. ROCHE, P., FURUTA, K. 2015. The ins and outs of MHC class II-mediated antigen processing and presentation. **Nat Rev Immunol**. 15, 203-216. DOI: https://doi.org/10.1038/nri3818.

4. GUCE, I. A. et al. 2012. HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. **Nat Struct Mol Biol**. 20, 90-98. DOI: https://doi.org/10.1038/nsmb.2460.

5. LIU, B., SHAO, Y., FU, R. 2021. Current research status of HLA in immune-related diseases. **Immunity, Inflammation and Disease**. 9(2), p. 340-350. DOI: https://doi.org/10.1002/iid3.416.

6. CASTELLI, E. C. et al. 2022. MUC22, HLA, and HLA-DOB variants and COVID-19 in resilient super-agers from Brazil. **Front. Immunol**. DOI: https://doi.org/10.3389/fimmu.2022.975918.

7. STICHT, J., ÁLVARO-BENITO, M., KONIGORSKI, S. 2021. Type 1 Diabetes and the HLA Region: Genetic Association Besides Classical HLA Class II Genes. **Front. Genet.** Vol. 12. DOI: https://doi.org/10.3389/fgene.2021.683946.

8. NARUSE, T.K. et al. 1999. Limited polymorphism in the HLA-DOA gene. **Tissue Antigens**. 53(4): 359-65. DOI: https://doi.org/10.1034/j.1399-0039.1999.530406.x.

9. NARUSE, T.K. et al. 2002. The HLA-DOB gene displays limited polymorphism with only one amino acid substitution. **Tissue Antigens.** 59(6): 512-9. DOI: https://doi.org/10.1034/j.1399-0039.2002.590608.x.

10. BARKER, D. J. et al. 2023. The IPD-IMGT/HLA Database. **Nucleic Acids Res.** 51(D1): D1053-D1060. DOI: https://doi.org/10.1093%2Fnar%2Fgkac1011.

11. ÁLVARO-BENITO, M.; FREUND, C. 2020. Revisiting nonclassical HLA II functions in antigen presentation: Peptide editing and its modulation. **HLA**. Vol. 96, Issue 4, p. 415-429. DOI: https://doi.org/10.1111/tan.14007.

12. HE, J. et al. 2022. Association of HLA-DM and HLA class II genes with antibody response induced by inactivated Japanese encephalitis vaccine. **HLA**. Vol. 99, Issue 4, p. 357-367. DOI: https://doi.org/10.1111/tan.14575.

13. The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. **Nature**. 526, 68-74. DOI: https://doi.org/10.1038/nature15393.

14. CANN, H. M. et al. 2002. A Human Genome Diversity Cell Line Panel. **Science**. Vol. 296 (5566), pp. 261-262. DOI: https://doi.org/10.1126/science.296.5566.261b.

15. NASLAVKSY, M.S., SCLIAR, M.O., YAMAMOTO, G.L. et al. 2022. Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. Nat Commun 13, 1004. https://doi.org/10.1038/s41467-022-28648-3.

16. CASTELLI, E.C., PAZ, M.A., SOUZA, A.S., RAMALHO, J., MENDES-JUNIOR, C.T. 2018. *Hla-mapper*: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. Hum Immunol. 79(9): 678–84. DOI: https://doi.org/10.1016/j.humimm.2018.06.010.

17. AUWERA, G.A.V der, O'CONNOR, B.D. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. **O'Reilly Media, Inc**. Available in: https://www.oreilly.com/library/view/genomics-in-the/9781491975183/.

18. MARTIN, M. et al. 2016. WhatsHap: fast and accurate read-based phasing. **Europe PMC**. DOI: https://doi.org/10.1101/085050.

19. DELANEAU, O. et al. 2019. Accurate, scalable and integrative haplotype estimation. **Nature Communications**. 10, 5436. DOI: https://doi.org/10.1038/s41467-019-13225-y.

20. 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. **Human Immunology.** 77(10), pages 841-853. DOI: https://doi.org/10.1016/j.humimm.2016.07.231.

21. BARRETT, J.C., FRY, B. MALLER, J. & DALY, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**. 21(2): 263-265. DOI: https://doi.org/10.1093/bioinformatics/bth457.

22. HUTTER, S., VILELLA, A. J., ROZAS, J. 2006. Genome-wide DNA polymorphism analyses using VariScan. **BMC Bioinformatics**. 7, 409. DOI: https://doi.org/10.1186/1471-2105-7-409.

23. KATOH, K., MISAWA, K., KUMA, K-C., MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, 30(14), 3059-3066. DOI: https://doi.org/10.1093/nar/gkf436.

Supplementary Materials

Table S1. List of samples used in this study

Table S2. HLA-DOA SNP frequency comparing nucleotide of the reference and alternative according to a position

Position	Reference	Alternative	Effect	Group
33009785	0.995512	4.49E-03	non_coding_or_silence	upstream_gene_variant
33009758	0.49308	5.07E-01	non_coding_or_silence	upstream_gene_variant
33009748	0.999719	2.81E-04	non_coding_or_silence	upstream_gene_variant
33009741	0.999158	8.42E-04	non_coding_or_silence	upstream_gene_variant
33009725	0.997101	2.90E-03	non_coding_or_silence	upstream_gene_variant
33009681	0.999906	9.40E-05	non_coding_or_silence	upstream_gene_variant
33009643	0.812418	1.88E-01	non_coding_or_silence	upstream_gene_variant
33009642	0.994202	5.80E-03	non_coding_or_silence	upstream_gene_variant
33009629	0.996821	3.18E-03	non_coding_or_silence	upstream_gene_variant
33009611	0.999906	9.40E-05	non_coding_or_silence	5_prime_UTR_variant
33009584	0.999906	9.40E-05	non_coding_or_silence	5_prime_UTR_variant
33009571	0.999158	8.42E-04	non_coding_or_silence	5_prime_UTR_variant
33009550	0.999719	2.81E-04	non_coding_or_silence	5_prime_UTR_variant
33009548	0.999532	4.68E-04	non_coding_or_silence	5_prime_UTR_variant
33009544	0.999906	9.40E-05	non_coding_or_silence	5_prime_UTR_variant
33009538	0.826725	1.73E-01	non_coding_or_silence	5_prime_UTR_variant
33009521	0.993922	6.08E-03	missense	missense_variant
33009449	0.999906	9.40E-05	non_coding_or_silence	splice_region_variant&intron_variant
33009447	0.999813	1.87E-04	non_coding_or_silence	splice_region_variant&intron_variant
33009427	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33009422	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33009409	0.999626	3.74E-04	non_coding_or_silence	intron_variant
33009405	0.993922	6.08E-03	non_coding_or_silence	intron_variant
33009354	0.993922	6.08E-03	non_coding_or_silence	intron_variant
33009307	0.94193	5.81E-02	non_coding_or_silence	intron_variant
33009267	0.928839	7.12E-02	non_coding_or_silence	intron_variant
33009209	0.999906	9.40E-05	non_coding_or_silence	intron_variant

33009200	0.999345	6.55E-04	non_coding_or_silence	intron_variant
33009182	0.999532	4.68E-04	non_coding_or_silence	intron_variant
33009159	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33009158	0.999626	3.74E-04	non_coding_or_silence	intron_variant
33009150	0.921919	7.81E-02	non_coding_or_silence	intron_variant
33009133	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33009132	0.465121	5.35E-01	non_coding_or_silence	intron_variant
33009130	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33009089	0.993922	6.08E-03	non_coding_or_silence	intron_variant
33009027	0.999813	1.87E-04	non_coding_or_silence	intron_variant
33009000	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008993	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008982	0.344586	6.55E-01	non_coding_or_silence	intron_variant
33008976	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008975	0.995044	4.96E-03	non_coding_or_silence	intron_variant
33008953	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008951	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008943	0.999439	5.61E-04	non_coding_or_silence	intron_variant
33008895	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008877	0.926594	7.34E-02	non_coding_or_silence	intron_variant
33008862	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008856	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008851	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008807	0.998691	1.31E-03	non_coding_or_silence	intron_variant
33008795	0.94193	5.81E-02	non_coding_or_silence	intron_variant
33008788	0.993735	6.26E-03	non_coding_or_silence	intron_variant
33008787	0.999813	1.87E-04	non_coding_or_silence	intron_variant
33008763	0.977838	2.22E-02	non_coding_or_silence	intron_variant
33008760	0.999532	4.68E-04	non_coding_or_silence	intron_variant
33008654	0.999719	2.81E-04	non_coding_or_silence	intron_variant
33008622	0.999813	1.87E-04	non_coding_or_silence	intron_variant
33008536	0.99626	3.74E-03	non_coding_or_silence	intron_variant
33008474	0.999719	2.81E-04	non_coding_or_silence	intron_variant
33008433	0.999906	9.40E-05	non_coding_or_silence	intron_variant

33008426	0.993735	6.26E-03	non_coding_or_silence	intron_variant
33008408	0.937629	6.24E-02	non_coding_or_silence	intron_variant
33008402	0.999626	3.74E-04	non_coding_or_silence	intron_variant
33008387	0.999719	2.81E-04	non_coding_or_silence	intron_variant
33008370	0.978867	2.11E-02	non_coding_or_silence	intron_variant
33008354	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008342	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33008294	0.987563	1.24E-02	non_coding_or_silence	intron_variant
33008279	0.993922	6.08E-03	non_coding_or_silence	intron_variant
33008271	0.999158	8.42E-04	non_coding_or_silence	intron_variant
33008265	0.999906	9.40E-05	non_coding_or_silence	splice_region_variant&intron_variant
33008236	0.988685	1.13E-02	non_coding_or_silence	synonymous_variant
33008235	0.999719	2.81E-04	frameshift	frameshift_variant
33008209	0.998036	1.96E-03	non_coding_or_silence	synonymous_variant
33008181	0.999906	9.40E-05	missense	missense_variant
33008168	0.999813	1.87E-04	missense	missense_variant
33008150	0.999813	1.87E-04	missense	missense_variant
33008136	0.999906	9.40E-05	missense	missense_variant
33008119	0.922386	7.76E-02	non_coding_or_silence	synonymous_variant
33008108	0.999065	9.35E-04	missense	missense_variant
33008098	0.999906	9.40E-05	non_coding_or_silence	synonymous_variant
33008092	0.810174	1.90E-01	non_coding_or_silence	synonymous_variant
33008031	0.988685	1.13E-02	missense	missense_variant
33007979	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007976	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007955	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007943	0.809332	1.91E-01	non_coding_or_silence	sequence_feature
33007925	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007922	0.792033	2.08E-01	non_coding_or_silence	sequence_feature
33007918	0.999813	1.87E-04	non_coding_or_silence	sequence_feature
33007914	0.999813	1.87E-04	non_coding_or_silence	sequence_feature
33007910	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007903	0.999906	9.40E-05	non_coding_or_silence	sequence_feature

33007874	0.999719	2.81E-04	non_coding_or_silence	sequence_feature
33007869	0.999813	1.87E-04	non_coding_or_silence	sequence_feature
33007845	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007814	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007801	0.85562	1.44E-01	non_coding_or_silence	sequence_feature
33007779	0.999719	2.81E-04	non_coding_or_silence	sequence_feature
33007772	0.999719	2.81E-04	non_coding_or_silence	sequence_feature
33007734	0.294651	7.05E-01	non_coding_or_silence	sequence_feature
33007719	0.999439	5.61E-04	non_coding_or_silence	sequence_feature
33007682	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007640	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007628	0.999532	4.68E-04	non_coding_or_silence	sequence_feature
33007619	0.999906	9.40E-05	non_coding_or_silence	sequence_feature
33007604	0.429119	5.71E-01	non_coding_or_silence	sequence_feature
33007603	0.999813	1.87E-04	non_coding_or_silence	sequence_feature
33007584	0.999906	9.40E-05	missense	missense_variant
33007576	0.999813	1.87E-04	non_coding_or_silence	synonymous_variant
33007565	0.999906	9.40E-05	missense	missense_variant
33007564	0.967645	3.24E-02	non_coding_or_silence	synonymous_variant
33007560	0.999906	9.40E-05	missense	missense_variant
33007559	0.999813	1.87E-04	missense	missense_variant
33007557	0.999906	9.40E-05	missense	missense_variant
33007509	0.996727	3.27E-03	missense	missense_variant
33007502	0.999906	9.40E-05	missense	missense_variant
33007480	0.608285	3.92E-01	non_coding_or_silence	synonymous_variant
33007474	0.999906	9.40E-05	missense	missense_variant
33007422	0.999906	9.40E-05	missense	missense_variant
33007402	0.999906	9.40E-05	non_coding_or_silence	synonymous_variant
33007400	0.998223	1.78E-03	missense	missense_variant
33007386	0.999439	5.61E-04	missense	missense_variant
33007375	0.999906	9.40E-05	non_coding_or_silence	synonymous_variant
33007374	0.999813	1.87E-04	missense	missense_variant
33007369	0.999906	9.40E-05	non_coding_or_silence	synonymous_variant
33007330	0.999813	1.87E-04	non_coding_or_silence	synonymous_variant

33007286	0.999813	1.87E-04	non_coding_or_silence	sequence_feature
33007237	0.608192	3.92E-01	non_coding_or_silence	sequence_feature
33007220	0.999906	9.40E-05	non_coding_or_silence	splice_region_variant&intron_variant
33007203	0.999719	2.81E-04	missense	missense_variant
33007189	0.999906	9.40E-05	missense	missense_variant
33007157	0.8343	1.66E-01	non_coding_or_silence	synonymous_variant
33007137	0.999439	5.61E-04	missense	missense_variant
33007130	0.99841	1.59E-03	non_coding_or_silence	synonymous_variant
33007121	0.999906	9.40E-05	non_coding_or_silence	synonymous_variant
33007120	0.99841	1.59E-03	missense	missense_variant
33007106	0.999906	9.40E-05	missense	missense_variant
33007066	0.848513	1.51E-01	non_coding_or_silence	intron_variant
33007032	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33007015	0.999719	2.81E-04	non_coding_or_silence	intron_variant
33006942	0.994576	5.42E-03	non_coding_or_silence	intron_variant
33006874	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33006873	0.999906	9.40E-05	non_coding_or_silence	intron_variant
33006866	0.999813	1.87E-04	non_coding_or_silence	intron_variant
33006863	0.999532	4.68E-04	non_coding_or_silence	intron_variant
33006847	0.997662	2.34E-03	non_coding_or_silence	splice_region_variant&intron_variant
33006777	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33006774	0.862259	1.38E-01	non_coding_or_silence	3_prime_UTR_variant
33006746	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006741	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006740	0.999345	6.55E-04	non_coding_or_silence	3_prime_UTR_variant
33006726	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33006714	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006703	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33006700	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006650	0.998317	1.68E-03	non_coding_or_silence	3_prime_UTR_variant
33006641	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33006627	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant

33006624	0.596129	4.04E-01	non_coding_or_silence	3_prime_UTR_variant
33006623	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006621	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33006566	0.950439	4.96E-02	non_coding_or_silence	3_prime_UTR_variant
33006564	0.999626	3.74E-04	non_coding_or_silence	3_prime_UTR_variant
33006541	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006513	0.998878	1.12E-03	non_coding_or_silence	3_prime_UTR_variant
33006492	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006491	0.578829	4.21E-01	non_coding_or_silence	3_prime_UTR_variant
33006373	0.99626	3.74E-03	non_coding_or_silence	3_prime_UTR_variant
33006346	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006345	0.995137	4.86E-03	non_coding_or_silence	3_prime_UTR_variant
33006304	0.600711	3.99E-01	non_coding_or_silence	3_prime_UTR_variant
33006291	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006254	0.997943	2.06E-03	non_coding_or_silence	3_prime_UTR_variant
33006227	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33006208	0.599027	4.01E-01	non_coding_or_silence	3_prime_UTR_variant
33006148	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006133	0.999626	3.74E-04	non_coding_or_silence	3_prime_UTR_variant
33006109	0.998504	1.50E-03	non_coding_or_silence	3_prime_UTR_variant
33006108	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33006101	0.146905	8.53E-01	non_coding_or_silence	3_prime_UTR_variant
33006099	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006092	0.999439	5.61E-04	non_coding_or_silence	3_prime_UTR_variant
33006080	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006064	0.766692	2.33E-01	non_coding_or_silence	3_prime_UTR_variant
33006055	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006022	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33006021	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33006017	0.849261	1.51E-01	non_coding_or_silence	3_prime_UTR_variant
33006002	0.998878	1.12E-03	non_coding_or_silence	3_prime_UTR_variant
33005971	0.989433	1.06E-02	non_coding_or_silence	3_prime_UTR_variant
33005966	0.146437	8.54E-01	non_coding_or_silence	3_prime_UTR_variant
33005941	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant

33005922	0.999158	8.42E-04	non_coding_or_silence	3_prime_UTR_variant
33005893	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33005882	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005876	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005822	0.85562	1.44E-01	non_coding_or_silence	3_prime_UTR_variant
33005736	0.405648	5.94E-01	non_coding_or_silence	3_prime_UTR_variant
33005725	0.996821	3.18E-03	non_coding_or_silence	3_prime_UTR_variant
33005724	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33005691	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33005689	0.999345	6.55E-04	non_coding_or_silence	3_prime_UTR_variant
33005688	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005646	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005638	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005575	0.959323	4.07E-02	non_coding_or_silence	3_prime_UTR_variant
33005526	0.941463	5.85E-02	non_coding_or_silence	3_prime_UTR_variant
33005525	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005523	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33005522	0.998691	1.31E-03	non_coding_or_silence	3_prime_UTR_variant
33005518	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005504	0.517767	4.82E-01	non_coding_or_silence	3_prime_UTR_variant
33005499	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33005486	0.38227	6.18E-01	non_coding_or_silence	3_prime_UTR_variant
33005483	0.970918	2.91E-02	non_coding_or_silence	3_prime_UTR_variant
33005468	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005452	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005397	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005390	0.998971	1.03E-03	non_coding_or_silence	3_prime_UTR_variant
33005361	0.990275	9.72E-03	non_coding_or_silence	3_prime_UTR_variant
33005360	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33005358	0.991865	8.14E-03	non_coding_or_silence	3_prime_UTR_variant
33005309	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005293	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33005288	0.999345	6.55E-04	non_coding_or_silence	3_prime_UTR_variant
33005276	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant

33005262	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005212	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005154	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005126	0.887601	1.12E-01	non_coding_or_silence	3_prime_UTR_variant
33005110	0.662521	3.37E-01	non_coding_or_silence	3_prime_UTR_variant
33005105	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33005104	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005100	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33005089	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33005060	0.999439	5.61E-04	non_coding_or_silence	3_prime_UTR_variant
33005004	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33004971	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004965	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004913	0.819805	1.80E-01	non_coding_or_silence	3_prime_UTR_variant
33004908	0.999719	2.81E-04	non_coding_or_silence	3_prime_UTR_variant
33004886	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004879	0.997756	2.24E-03	non_coding_or_silence	3_prime_UTR_variant
33004865	0.843931	1.56E-01	non_coding_or_silence	3_prime_UTR_variant
33004802	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004794	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004709	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004669	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004638	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004627	0.867309	1.33E-01	non_coding_or_silence	3_prime_UTR_variant
33004624	0.998036	1.96E-03	non_coding_or_silence	3_prime_UTR_variant
33004598	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004560	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004545	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33004479	0.999532	4.68E-04	non_coding_or_silence	3_prime_UTR_variant
33004436	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004430	0.950346	4.97E-02	non_coding_or_silence	3_prime_UTR_variant
33004386	0.948569	5.14E-02	non_coding_or_silence	3_prime_UTR_variant
33004385	0.96989	3.01E-02	non_coding_or_silence	3_prime_UTR_variant
33004347	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant

33004327	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004317	0.999813	1.87E-04	non_coding_or_silence	3_prime_UTR_variant
33004306	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004273	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004230	0.999906	9.40E-05	non_coding_or_silence	3_prime_UTR_variant
33004168	0.999345	6.55E-04	non_coding_or_silence	downstream_gene_variant
33004163	0.997382	2.62E-03	non_coding_or_silence	downstream_gene_variant
33004155	0.999065	9.35E-04	non_coding_or_silence	downstream_gene_variant
33004154	0.999252	7.48E-04	non_coding_or_silence	downstream_gene_variant
33004153	0.999813	1.87E-04	non_coding_or_silence	downstream_gene_variant
33004054	0.999439	5.61E-04	non_coding_or_silence	downstream_gene_variant
33004053	0.999626	3.74E-04	non_coding_or_silence	downstream_gene_variant
33004032	0.999906	9.40E-05	non_coding_or_silence	downstream_gene_variant
33003999	0.999626	3.74E-04	non_coding_or_silence	downstream_gene_variant

Table S3. HLA-DOB SNP frequency comparing nucleotide of the reference and alternative according to a position

Position	Reference	Alternative	Effect	Group
32817279	0.995979	0.004021	non_coding_or_silence	upstream_gene_variant
32817258	0.910791	0.089209	non_coding_or_silence	upstream_gene_variant
32817257	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817252	0.910885	0.089115	non_coding_or_silence	upstream_gene_variant
32817249	0.999813	0.000187	non_coding_or_silence	upstream_gene_variant
32817197	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817195	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817190	0.999626	0.0003740	non_coding_or_silence	upstream_gene_variant
32817131	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817120	0.999813	0.000187	non_coding_or_silence	upstream_gene_variant
32817086	0.993922	0.006078	non_coding_or_silence	upstream_gene_variant
32817068	0.998691	0.001309	non_coding_or_silence	upstream_gene_variant
32817006	0.601272	0.398728	non_coding_or_silence	5_prime_UTR_variant
32816976	0.690294	0.309706	non_coding_or_silence	5_prime_UTR_premature_start_codon_gain_variant
32816964	0.999439	0.000561	non_coding_or_silence	5_prime_UTR_variant

32816961	0.999906	9,40E+09	non_coding_or_silence	5_prime_UTR_variant
32816933	0.999813	0.000187	missense	missense_variant
32816899	0.911633	0.088367	missense	missense_variant
32816886	0.999813	0.000187	missense	missense_variant
32816885	0.999906	9,40E+09	missense	missense_variant
32816868	0.983542	0.016458	missense	missense_variant
32816849	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32816843	0.690294	0.309706	non_coding_or_silence	downstream_gene_variant
32816816	0.999813	0.000187	non_coding_or_silence	intron_variant
32816813	0.997008	0.002992	non_coding_or_silence	intron_variant
32816789	0.998971	0.001029	non_coding_or_silence	intron_variant
32816701	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816646	0.999719	0.000281	non_coding_or_silence	intron_variant
32816635	0.999345	0.000655	non_coding_or_silence	intron_variant
32816632	0.9928	0.007110	non_coding_or_silence	intron_variant
32816599	0.999439	0.000561	non_coding_or_silence	intron_variant
32816576	0.999626	0.000374	non_coding_or_silence	intron_variant
32816527	0.979334	0.020666	non_coding_or_silence	intron_variant
32816523	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816463	0.999813	0.000187	non_coding_or_silence	intron_variant
32816454	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816387	0.994763	0.005237	non_coding_or_silence	intron_variant
32816355	0.999813	0.000187	non_coding_or_silence	intron_variant
32816353	0.999813	0.000187	non_coding_or_silence	intron_variant
32816331	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816321	0.950814	0.049186	non_coding_or_silence	intron_variant
32816320	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816296	0.690294	0.309706	non_coding_or_silence	intron_variant
32816235	0.999719	0.000281	non_coding_or_silence	intron_variant
32816187	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816159	0.999813	0.000187	non_coding_or_silence	intron_variant
32816142	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816125	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816119	0.512998	0.487002	non_coding_or_silence	intron_variant

32816105	0.999719	0.000281	non_coding_or_silence	intron_variant
32816058	0.999532	0.000468	non_coding_or_silence	intron_variant
32816013	0.993361	0.006639	non_coding_or_silence	intron_variant
32815998	0.999065	0.000935	non_coding_or_silence	intron_variant
32815986	0.98214	0.01786	non_coding_or_silence	intron_variant
32815907	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815775	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815774	0.999813	0.000187	non_coding_or_silence	intron_variant
32815766	0.999813	0.000187	non_coding_or_silence	intron_variant
32815673	0.959697	0.040303	non_coding_or_silence	intron_variant
32815629	0.959604	0.040396	non_coding_or_silence	intron_variant
32815628	0.911726	0.088274	non_coding_or_silence	intron_variant
32815626	0.999252	0.000748	non_coding_or_silence	intron_variant
32815610	0.936039	0.063961	non_coding_or_silence	intron_variant
32815588	0.910791	0.089209	non_coding_or_silence	intron_variant
32815585	0.999439	0.000561	non_coding_or_silence	intron_variant
32815560	0.512998	0.487002	non_coding_or_silence	intron_variant
32815551	0.976061	0.023939	non_coding_or_silence	intron_variant
32815547	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815518	0.941182	0.058818	non_coding_or_silence	intron_variant
32815512	0.999719	0.000281	non_coding_or_silence	intron_variant
32815467	0.979521	0.020479	non_coding_or_silence	intron_variant
32815464	0.999345	0.000655	non_coding_or_silence	intron_variant
32815456	0.998597	0.001403	non_coding_or_silence	intron_variant
32815442	0.999813	0.000187	non_coding_or_silence	intron_variant
32815439	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815421	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815407	0.999252	0.000748	non_coding_or_silence	intron_variant
32815382	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815366	0.999626	0.000374	non_coding_or_silence	intron_variant
32815347	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815328	0.999626	0.000374	non_coding_or_silence	intron_variant
32815317	0.999813	0.000187	non_coding_or_silence	splice_region_variant&intron_variant

32815309	0.852721	0.147279	non_coding_or_silence	synonymous_variant
32815156	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32815145	0.999906	9,40E+09	missense	missense_variant
32815120	0.978399	0.021601	non_coding_or_silence	synonymous_variant
32815097	0.99841	0.001590	missense	missense_variant
32815049	0.999532	0.000468	missense	missense_variant
32815018	0.972976	0.027024	non_coding_or_silence	intron_variant
32814971	0.999719	0.000281	non_coding_or_silence	intron_variant
32814868	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814840	0.999532	0.000468	non_coding_or_silence	intron_variant
32814828	0.573406	0.426594	non_coding_or_silence	intron_variant
32814821	0.999252	0.000748	non_coding_or_silence	intron_variant
32814819	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814805	0.690481	0.309519	non_coding_or_silence	intron_variant
32814729	0.999626	0.000374	non_coding_or_silence	intron_variant
32814727	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814718	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814707	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814677	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814610	0.769684	0.230316	non_coding_or_silence	intron_variant
32814583	0.999906	9,40E+09	missense	missense_variant
32814501	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814493	0.999906	9,40E+09	stop_codon	stop_gained
32814482	0.999906	9,40E+09	missense	missense_variant
32814447	0.999719	0.000281	frameshift	frameshift_variant
32814443	0.999906	9,40E+09	missense	missense_variant
32814442	0.999906	9,40E+09	missense	missense_variant
32814436	0.999719	0.000281	missense	missense_variant
32814381	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814372	0.962128	0.037872	non_coding_or_silence	splice_region_variant&synonymous_variant
32814369	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814352	0.999906	9,40E+09	missense	missense_variant
32814336	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant

32814335	0.974285	0.025715	missense	missense_variant
32814279	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32814260	0.999158	0.000842	non_coding_or_silence	downstream_gene_variant
32814255	0.998691	0.001309	non_coding_or_silence	downstream_gene_variant
32814253	0.826725	0.173275	non_coding_or_silence	downstream_gene_variant
32814250	0.826725	0.173275	non_coding_or_silence	downstream_gene_variant
32814241	0.910043	0.089957	non_coding_or_silence	downstream_gene_variant
32814195	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32814177	0.947915	0.052085	non_coding_or_silence	downstream_gene_variant
32814107	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32814049	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32814006	0.997288	0.002712	non_coding_or_silence	downstream_gene_variant
32814005	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813999	0.509445	0.490555	non_coding_or_silence	downstream_gene_variant
32813978	0.983823	0.016177	non_coding_or_silence	downstream_gene_variant
32813944	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813929	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813926	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32813855	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813779	0.999906	9,40E+09	missense	missense_variant
32813777	0.910043	0.089957	missense	missense_variant
32813748	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32813747	0.963531	0.036469	missense	missense_variant
32813684	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813676	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813657	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813656	0.947821	0.052179	non_coding_or_silence	downstream_gene_variant
32813638	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813572	0.999158	0.000842	non_coding_or_silence	downstream_gene_variant
32813568	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32813567	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813530	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813504	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813501	0.947915	0.052085	non_coding_or_silence	downstream_gene_variant

32813492	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813484	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813442	0.999906	9,40E+09	missense	missense_variant&splice_region_variant
32813398	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813395	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813371	0.978399	0.021601	non_coding_or_silence	downstream_gene_variant
32813361	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813344	0.908921	0.091079	non_coding_or_silence	downstream_gene_variant
32813335	0.983823	0.016177	non_coding_or_silence	downstream_gene_variant
32813320	0.999439	0.000561	non_coding_or_silence	downstream_gene_variant
32813306	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813299	0.907518	0.092482	non_coding_or_silence	downstream_gene_variant
32813264	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813180	0.973069	0.026931	non_coding_or_silence	3_prime_UTR_variant
32813158	0.978306	0.021694	non_coding_or_silence	3_prime_UTR_variant
32813157	0.978399	0.021601	non_coding_or_silence	3_prime_UTR_variant
32813137	0.908921	0.091079	non_coding_or_silence	3_prime_UTR_variant
32813103	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813084	0.999626	0.000374	non_coding_or_silence	3_prime_UTR_variant
32813049	0.999626	0.000374	non_coding_or_silence	splice_region_variant&non_coding_transcript_exon_variant
32813029	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813014	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812999	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812992	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812947	0.741631	0.258369	non_coding_or_silence	3_prime_UTR_variant
32812933	0.999813	0.000187	non_coding_or_silence	3_prime_UTR_variant
32812927	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812909	0.999252	0.000748	non_coding_or_silence	3_prime_UTR_variant
32812873	0.999345	0.000655	non_coding_or_silence	3_prime_UTR_variant
32812872	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812863	0.999813	0.000187	non_coding_or_silence	3_prime_UTR_variant

32812861	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812795	0.854404	0.145596	non_coding_or_silence	3_prime_UTR_variant
32812788	0.999719	0.000281	non_coding_or_silence	3_prime_UTR_variant
32812694	0.999439	0.000561	non_coding_or_silence	downstream_gene_variant
32812693	0.373574	0.626426	non_coding_or_silence	downstream_gene_variant
32812674	0.998784	0.001216	non_coding_or_silence	downstream_gene_variant
32812570	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32812569	0.9928	0.007110	non_coding_or_silence	downstream_gene_variant
32812558	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32812540	0.307182	0.692818	non_coding_or_silence	downstream_gene_variant
32812538	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32812531	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32812528	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32812517	0.963531	0.036469	non_coding_or_silence	downstream_gene_variant
32817279	0.995979	0.004021	non_coding_or_silence	upstream_gene_variant
32817258	0.910791	0.089209	non_coding_or_silence	upstream_gene_variant
32817257	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817252	0.910885	0.089115	non_coding_or_silence	upstream_gene_variant
32817249	0.999813	0.000187	non_coding_or_silence	upstream_gene_variant
32817197	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817195	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817190	0.999626	0.000374	non_coding_or_silence	upstream_gene_variant
32817131	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant
32817120	0.999813	0.000187	non_coding_or_silence	upstream_gene_variant
32817086	0.993922	0.006078	non_coding_or_silence	upstream_gene_variant
32817068	0.998691	0.001309	non_coding_or_silence	upstream_gene_variant
32817006	0.601272	0.398728	non_coding_or_silence	5_prime_UTR_variant
32816976	0.690294	0.309706	non_coding_or_silence	5_prime_UTR_premature_start_codon_gain_variant
32816964	0.999439	0.000561	non_coding_or_silence	5_prime_UTR_variant
32816961	0.999906	9,40E+09	non_coding_or_silence	5_prime_UTR_variant
32816933	0.999813	0.000187	missense	missense_variant
32816899	0.911633	0.088367	missense	missense_variant
32816886	0.999813	0.000187	missense	missense_variant

32816885	0.999906	9,40E+09	missense	missense_variant
32816868	0.983542	0.016458	missense	missense_variant
32816849	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32816843	0.690294	0.309706	non_coding_or_silence	downstream_gene_variant
32816816	0.999813	0.000187	non_coding_or_silence	intron_variant
32816813	0.997008	0.002992	non_coding_or_silence	intron_variant
32816789	0.998971	0.001029	non_coding_or_silence	intron_variant
32816701	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816646	0.999719	0.000281	non_coding_or_silence	intron_variant
32816635	0.999345	0.000655	non_coding_or_silence	intron_variant
32816632	0.9928	0.007110	non_coding_or_silence	intron_variant
32816599	0.999439	0.000561	non_coding_or_silence	intron_variant
32816576	0.999626	0.000374	non_coding_or_silence	intron_variant
32816527	0.979334	0.020666	non_coding_or_silence	intron_variant
32816523	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816523	0.999813	0.000187	non_coding_or_silence	intron_variant
32816463	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816454	0.994763	0.005237	non_coding_or_silence	intron_variant
32816387	0.999813	0.000187	non_coding_or_silence	intron_variant
32816355	0.999813	0.000187	non_coding_or_silence	intron_variant
32816353	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816331	0.950814	0.049186	non_coding_or_silence	intron_variant
32816321	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816320	0.690294	0.309706	non_coding_or_silence	intron_variant
32816296	0.999719	0.000281	non_coding_or_silence	intron_variant
32816235	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816187	0.999813	0.000187	non_coding_or_silence	intron_variant
32816159	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816142	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32816125	0.512998	0.487002	non_coding_or_silence	intron_variant
32816119	0.999719	0.000281	non_coding_or_silence	intron_variant
32816105	0.999532	0.000468	non_coding_or_silence	intron_variant
32816058	0.993361	0.006639	non_coding_or_silence	intron_variant
32816013	0.999065	0.000935	non_coding_or_silence	intron_variant

32815998	0.98214	0.01786	non_coding_or_silence	intron_variant
32815986	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815907	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815775	0.999813	0.000187	non_coding_or_silence	intron_variant
32815774	0.999813	0.000187	non_coding_or_silence	intron_variant
32815766	0.959697	0.040303	non_coding_or_silence	intron_variant
32815673	0.959604	0.040396	non_coding_or_silence	intron_variant
32815629	0.911726	0.088274	non_coding_or_silence	intron_variant
32815628	0.999252	0.000748	non_coding_or_silence	intron_variant
32815626	0.936039	0.063961	non_coding_or_silence	intron_variant
32815610	0.910791	0.089209	non_coding_or_silence	intron_variant
32815588	0.999439	0.000561	non_coding_or_silence	intron_variant
32815585	0.512998	0.487002	non_coding_or_silence	intron_variant
32815560	0.976061	0.023939	non_coding_or_silence	intron_variant
32815551	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815547	0.941182	0.058818	non_coding_or_silence	intron_variant
32815518	0.999719	0.000281	non_coding_or_silence	intron_variant
32815512	0.979521	0.020479	non_coding_or_silence	intron_variant
32815467	0.999345	0.000655	non_coding_or_silence	intron_variant
32815464	0.998597	0.001403	non_coding_or_silence	intron_variant
32815456	0.999813	0.000187	non_coding_or_silence	intron_variant
32815442	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815439	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815421	0.999252	0.000748	non_coding_or_silence	intron_variant
32815407	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815382	0.999626	0.000374	non_coding_or_silence	intron_variant
32815366	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32815347	0.999626	0.000374	non_coding_or_silence	intron_variant
32815328	0.999813	0.000187	non_coding_or_silence	splice_region_variant&intron_variant
32815317	0.852721	0.147279	non_coding_or_silence	synonymous_variant
32815309	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32815156	0.999906	9,40E+09	missense	missense_variant
32815145	0.978399	0.021601	non_coding_or_silence	synonymous_variant

32815120	0.99841	0.001510	missense	missense_variant
32815097	0.999532	0.000468	missense	missense_variant
32815049	0.972976	0.027024	non_coding_or_silence	intron_variant
32815018	0.999719	0.000281	non_coding_or_silence	intron_variant
32814971	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814868	0.999532	0.000468	non_coding_or_silence	intron_variant
32814840	0.573406	0.426594	non_coding_or_silence	intron_variant
32814828	0.999252	0.000748	non_coding_or_silence	intron_variant
32814821	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814819	0.690481	0.309519	non_coding_or_silence	intron_variant
32814805	0.999626	0.000374	non_coding_or_silence	intron_variant
32814729	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814727	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814718	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814718	0.999906	9,40E+09	non_coding_or_silence	intron_variant
32814707	0.769684	0.230316	non_coding_or_silence	intron_variant
32814677	0.999906	9,40E+09	missense	missense_variant
32814610	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814583	0.999906	9,40E+09	stop_codon	stop_gained
32814501	0.999906	9,40E+09	missense	missense_variant
32814493	0.999719	0.000281	frameshift	frameshift_variant
32814482	0.999906	9,40E+09	missense	missense_variant
32814447	0.999906	9,40E+09	missense	missense_variant
32814443	0.999719	0.000281	missense	missense_variant
32814442	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814436	0.962128	0.037872	non_coding_or_silence	splice_region_variant&synonymous_variant
32814381	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814372	0.999906	9,40E+09	missense	missense_variant
32814369	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32814352	0.974285	0.025715	missense	missense_variant
32814336	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32814335	0.999158	0.000842	non_coding_or_silence	downstream_gene_variant
32814279	0.998691	0.001309	non_coding_or_silence	downstream_gene_variant

32814260	0.826725	0.173275	non_coding_or_silence	downstream_gene_variant
32814255	0.826725	0.173275	non_coding_or_silence	downstream_gene_variant
32814253	0.910043	0.089957	non_coding_or_silence	downstream_gene_variant
32814250	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32814241	0.947915	0.052085	non_coding_or_silence	downstream_gene_variant
32814195	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32814177	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32814107	0.997288	0.002712	non_coding_or_silence	downstream_gene_variant
32814049	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32814006	0.509445	0.490555	non_coding_or_silence	downstream_gene_variant
32814005	0.983823	0.016177	non_coding_or_silence	downstream_gene_variant
32813999	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813978	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813944	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32813929	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813926	0.999906	9,40E+09	missense	missense_variant
32813855	0.910043	0.089957	missense	missense_variant
32813779	0.999906	9,40E+09	non_coding_or_silence	synonymous_variant
32813777	0.963531	0.036469	missense	missense_variant
32813748	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813747	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813684	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813676	0.947821	0.052179	non_coding_or_silence	downstream_gene_variant
32813657	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813656	0.999158	0.000842	non_coding_or_silence	downstream_gene_variant
32813638	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32813572	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813568	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813567	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813530	0.947915	0.052085	non_coding_or_silence	downstream_gene_variant
32813504	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813504	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813501	0.999906	9,40E+09	missense	missense_variant&splice_region_variant

32813492	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813484	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813442	0.978399	0.021601	non_coding_or_silence	downstream_gene_variant
32813398	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813395	0.908921	0.091079	non_coding_or_silence	downstream_gene_variant
32813371	0.983823	0.016177	non_coding_or_silence	downstream_gene_variant
32813361	0.999439	0.000561	non_coding_or_silence	downstream_gene_variant
32813344	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32813335	0.907518	0.092482	non_coding_or_silence	downstream_gene_variant
32813320	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32813306	0.973069	0.026931	non_coding_or_silence	3_prime_UTR_variant
32813299	0.978306	0.021694	non_coding_or_silence	3_prime_UTR_variant
32813264	0.978399	0.021601	non_coding_or_silence	3_prime_UTR_variant
32813180	0.908921	0.091079	non_coding_or_silence	3_prime_UTR_variant
32813158	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813157	0.999626	0.000374	non_coding_or_silence	3_prime_UTR_variant
32813137	0.999626	0.000374	non_coding_or_silence	splice_region_variant&non_coding_transcript_exon_variant
32813103	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813084	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813049	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813029	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32813014	0.741631	0.258369	non_coding_or_silence	3_prime_UTR_variant
32812999	0.999813	0.000187	non_coding_or_silence	3_prime_UTR_variant
32812992	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812947	0.999252	0.000748	non_coding_or_silence	3_prime_UTR_variant
32812933	0.999345	0.000655	non_coding_or_silence	3_prime_UTR_variant
32812927	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812909	0.999813	0.000187	non_coding_or_silence	3_prime_UTR_variant
32812873	0.999906	9,40E+09	non_coding_or_silence	3_prime_UTR_variant
32812872	0.854404	0.145596	non_coding_or_silence	3_prime_UTR_variant
32812863	0.999719	0.000281	non_coding_or_silence	3_prime_UTR_variant
32812861	0.999439	0.000561	non_coding_or_silence	downstream_gene_variant

32812795	0.373574	0.626426	non_coding_or_silence	downstream_gene_variant
32812788	0.998784	0.001216	non_coding_or_silence	downstream_gene_variant
32812694	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32812693	0.9928	0.007110	non_coding_or_silence	downstream_gene_variant
32812674	0.999719	0.000281	non_coding_or_silence	downstream_gene_variant
32812570	0.307182	0.692818	non_coding_or_silence	downstream_gene_variant
32812569	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32812558	0.999906	9,40E+09	non_coding_or_silence	downstream_gene_variant
32812540	0.999813	0.000187	non_coding_or_silence	downstream_gene_variant
32812538	0.963531	0.036469	non_coding_or_silence	downstream_gene_variant
32812531	0.995979	0.004021	non_coding_or_silence	upstream_gene_variant
32812528	0.910791	0.089209	non_coding_or_silence	upstream_gene_variant
32812517	0.999906	9,40E+09	non_coding_or_silence	upstream_gene_variant

Table S4. *HLA-DOA* four-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	**1958**	**1080**	**2644**	**362**	**1556**	**1540**	**306**	**50**	**1198**	**10694**
DOA*01:01:01:01	0.0123	0.1046	0.1513	0.1685	0.3207	0.2260	0.0948	0.3000	0.1553	0.1566
DOA*01:01:02:01	0.0082	0.1185	0.1278	0.1685	0.1883	0.1234	0.1961	0.5200	0.2412	0.1310
DOA*01:01:04:01	0.0414	0.1398	0.1191	0.1878	0.1774	0.1474	0.1340	0.0400	0.1210	0.1221
DOA*01:01:02:02	0.0138	0.3000	0.1214	0.0829	0.0598	0.1195	0.0686	-	0.0234	0.0961
DOA*01:01:02:15	0.0097	0.0380	0.0495	0.0608	0.0893	0.0877	0.0261	-	0.0609	0.0531
DOA*01:01:02:08	0.1328	0.0194	0.0507	0.0110	0.0096	0.0403	0.0327	-	0.0092	0.0483
DOA*01:01:05	0.0669	0.0509	0.0541	-	-	0.0487	0.0425	-	-	0.0390
DOA*01:01:02:32	0.1185	0.0148	0.0212	0.0249	0.0006	0.0052	0.0392	-	0.0426	0.0360
DOA*01:01:04:03	0.0919	0.0259	0.0378	0.0138	0.0045	0.0149	0.0359	-	0.0125	0.0345
DOA*01:01:02:26	0.0163	0.0213	0.0609	0.0193	0.0051	0.0506	0.0556	-	0.0217	0.0329
DOA*01:01:11	0.1292	0.0120	0.0265	0.0083	-	-	0.0065	-	-	0.0319
DOA*01:01:02:20	0.0010	0.0167	0.0231	0.0387	0.0109	0.0286	0.0294	-	0.0459	0.0206
DOA*01:01:02:30	0.0296	0.0065	0.0125	0.0304	0.0174	0.0071	0.0425	-	0.0142	0.0166
DOA*01:01:03:01	0.0235	0.0083	0.0200	0.0028	0.0019	0.0227	0.0163	-	0.0025	0.0145
DOA*01:01:08	0.0429	0.0046	0.0095	-	-	-	0.0065	-	-	0.0108
DOA*01:01:02:22	-	0.0074	0.0068	0.0193	0.0039	0.0097	0.0261	-	0.0334	0.0095
DOA*01:01:04:05	0.0347	0.0019	0.0030	-	-	0.0006	0.0098	-	-	0.0077
DOA*01:02:01:02	-	0.0148	0.0042	0.0221	0.0045	0.0097	-	-	0.0167	0.0072
DOA*01:01:02:03	0.0015	0.0139	0.0053	-	-	0.0149	0.0065	-	-	0.0053
DOA*01:01:02:38	0.0005	0.0009	0.0061	0.0055	-	0.0006	0.0196	-	0.0192	0.0047
DOA*01:01:02:10	-	0.0028	0.0019	0.0028	0.0019	0.0032	0.0033	-	0.0259	0.0046
DOA*01:01:03:05	0.0174	0.0019	0.0038	-	-	-	-	-	-	0.0043
DOA*01:10	-	-	0.0004	0.0331	0.0006	-	-	-	0.0259	0.0042
DOA*01:01:02:36	0.0194	-	0.0015	-	-	-	-	-	-	0.0039
DOA*01:01:02:05	-	0.0009	0.0015	0.0055	0.0013	0.0006	0.0163	-	0.0192	0.0036
DOA*01:02:01:03	-	0.0009	0.0034	0.0055	0.0006	-	0.0065	-	0.0184	0.0035
DOA*01:08	-	0.0019	0.0019	0.0055	-	0.0039	0.0033	-	0.0159	0.0033
DOA*01:01:02:28	0.0066	-	0.0061	0.0028	-	0.0013	-	-	-	0.0030
DOA*01:01:02:29	0.0153	0.0009	0.0004	-	-	-	-	-	-	0.0030
DOA*01:01:02:12	0.0005	-	0.0004	0.0110	0.0071	0.0013	0.0033	-	0.0083	0.0028
DOA*01:01:01:04	-	-	0.0004	-	0.0167	-	-	-	-	0.0025
DOA*01:01:02:13	0.0020	-	0.0030	0.0055	-	0.0019	-	-	0.0058	0.0022
DOA*01:01:04:06	0.0072	0.0009	0.0019	-	-	-	-	-	-	0.0019
DOA*01:01:04:10	-	-	0.0008	0.0083	0.0084	-	-	-	-	0.0017
DOA*01:14	-	0.0130	-	0.0028	0.0019	-	-	-	-	0.0017
DOA*01:01:02:09	0.0056	0.0009	-	-	0.0013	0.0006	0.0033	-	-	0.0015
DOA*01:01:02:06	-	0.0028	0.0019	-	-	0.0019	-	-	-	0.0010
DOA*01:01:02:14	-	0.0074	0.0004	-	-	-	-	-	0.0008	0.0009
DOA*01:09	-	0.0009	0.0008	-	-	0.0026	0.0065	-	0.0008	0.0009
DOA*01:01:02:16	-	0.0009	0.0023	-	-	0.0013	-	-	-	0.0008
DOA*01:01:02:31	-	0.0019	0.0008	-	-	0.0013	-	-	-	0.0006
DOA*01:02:01:01	-	-	0.0019	-	-	-	0.0033	-	-	0.0006
DOA*01:06	-	-	0.0015	-	-	0.0013	-	-	-	0.0006

DOA*01:01:01:06	-	-	0.0008	-	0.0019	-	-	-	-	0.0005
DOA*01:01:02:34	0.0015	-	0.0004	-	-	-	-	-	0.0008	0.0005
DOA*01:01:02:25	-	-	0.0004	-	-	0.0019	-	-	-	0.0004
DOA*01:01:01:07	-	-	-	-	-	0.0019	-	-	-	0.0003
DOA*01:01:02:19	-	0.0019	0.0004	-	-	-	-	-	-	0.0003
DOA*01:01:02:46	0.0010	-	-	-	-	-	-	-	-	0.0002
DOA*01:01:01:02	-	-	-	-	0.0006	-	-	-	-	0.0001
DOA*01:01:01:03	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:02:11	0.0005	-	-	-	-	-	-	-	-	0.0001
DOA*01:01:02:37	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:04:08	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:04:09	-	-	-	-	-	0.0006	-	-	-	0.0001
DOA*01:01:07	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:12	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:13	-	-	-	-	-	0.0006	-	-	-	0.0001
GENOMIC POTENTIALLY NOVEL	0.1472	0.04	0.0495	0.0527	0.0626	0.0149	0.0655	0.14	0.0578	0.0673

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions.

Table S5. *HLA-DOA* three-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	1958	1080	2644	362	1556	1540	306	50	1198	10694
DOA*01:01:02	0.4663	0.6000	0.5306	0.5000	0.4325	0.5104	0.6111	0.5200	0.5935	0.5169
DOA*01:01:04	0.1879	0.1704	0.1664	0.2210	0.2005	0.1649	0.1830	0.1400	0.1419	0.1750
DOA*01:01:01	0.0133	0.1102	0.1581	0.1906	0.3470	0.2318	0.1013	0.3000	0.1603	0.1652
DOA*01:01:05	0.0868	0.0537	0.0624	-	-	0.0487	0.0425	-	-	0.0450
DOA*01:01:11	0.1313	0.0120	0.0265	0.0083	-	-	0.0065	-	-	0.0323
DOA*01:01:03	0.0546	0.0130	0.0250	0.0028	0.0019	0.0234	0.0196	-	0.0025	0.0221
DOA*01:02:01	-	0.0157	0.0095	0.0276	0.0051	0.0097	0.0098	-	0.0351	0.0112
DOA*01:01:08	0.0429	0.0056	0.0095	-	-	-	0.0065	-	-	0.0109
DOA*01:10	-	-	0.0004	0.0387	0.0013	-	-	0.0200	0.0367	0.0058
DOA*01:08	-	0.0019	0.0019	0.0055	-	0.0039	0.0033	-	0.0159	0.0033
DOA*01:01:06	0.0005	0.0009	0.0030	0.0028	0.0026	0.0019	0.0098	-	-	0.0020
DOA*01:14	-	0.0130	-	0.0028	0.0019	-	-	-	-	0.0017
DOA*01:09	-	0.0009	0.0008	-	-	0.0026	0.0065	-	0.0008	0.0009
DOA*01:06	-	-	0.0015	-	-	0.0013	-	-	-	0.0006
DOA*01:04N	-	-	-	-	0.0019	-	-	-	-	0.0003
DOA*01:01:07	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:10	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:01:12	-	-	0.0004	-	-	-	-	-	-	0.0001
DOA*01:13	-	-	-	-	-	0.0006	-	-	-	0.0001
CODING POTENTIALLY NOVEL	0.0162	0.0028	0.0036	0	0.005	0.0006	0	0.02	0.0133	0.0068

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions.

Table S6. *HLA-DOA* two-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	**1958**	**1080**	**2644**	**362**	**1556**	**1540**	**306**	**50**	**1198**	**10694**
DOA*01:01	0.9872	0.9657	0.9826	0.9254	0.9871	0.9812	0.9804	0.9600	0.9098	0.9719
DOA*01:02	-	0.0157	0.0095	0.0276	0.0051	0.0097	0.0098	-	0.0351	0.0112
DOA*01:10	-	-	0.0004	0.0387	0.0013	-	-	0.0200	0.0367	0.0058
DOA*01:08	-	0.0019	0.0019	0.0055	-	0.0039	0.0033	-	0.0159	0.0033
DOA*01:14	-	0.0130	-	0.0028	0.0019	-	-	-	-	0.0017
DOA*01:05	0.0077	-	0.0008	-	-	-	-	-	-	0.0016
DOA*01:09	-	0.0009	0.0008	-	-	0.0026	0.0065	-	0.0008	0.0009
DOA*01:06	-	-	0.0015	-	-	0.0013	-	-	-	0.0006
DOA*01:04N	-	-	-	-	0.0019	-	-	-	-	0.0003
DOA*01:13	-	-	-	-	-	0.0006	-	-	-	0.0001
PROTEIN POTENCIALLY NOVEL	0.005	0.0028	0.0028	0	0.0025	0.0006	0	0.02	0.0016	0.0028

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions.

Table S7. *HLA-DOB* four-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	**1958**	**1080**	**2644**	**362**	**1556**	**1540**	**306**	**50**	**1198**	**10694**
DOB*01:01:01:01	0.2114	0.2019	0.1868	0.1657	0.1234	0.2000	0.1863	0.0800	0.1511	0.1803
DOB*01:01:01:24	0.0485	0.2676	0.1543	0.1796	0.1523	0.1786	0.1013	0.3000	0.1711	0.1515
DOB*01:01:03:09	0.0230	0.0861	0.1180	0.2099	0.0238	0.1071	0.1438	-	0.1093	0.0844
DOB*01:01:01:04	0.0015	0.0454	0.0722	0.0746	0.0308	0.0851	0.1209	-	0.1093	0.0577
DOB*01:01:01:28	0.0072	0.0222	0.0374	0.0552	0.1311	0.0396	0.0588	0.0200	0.0643	0.0484
DOB*01:02:01:03	0.1471	0.0231	0.0318	0.0055	-	0.0169	0.0621	-	-	0.0415
DOB*01:03:01:02	0.0526	0.0565	0.0518	0.0028	0.0032	0.0442	0.0033	-	0.0033	0.0355
DOB*01:04:01:02	0.0066	0.0167	0.0378	0.0193	0.0617	0.0442	0.0229	0.2000	0.0301	0.0332
DOB*01:01:01:03	0.0026	0.0602	0.0269	0.0110	0.0450	0.0377	0.0098	-	0.0083	0.0267
DOB*01:05	0.0036	0.0083	0.0193	0.0221	0.0148	0.0357	0.0261	-	0.0918	0.0253
DOB*01:01:03:02	0.0225	0.0222	0.0371	-	-	0.0571	0.0098	-	-	0.0240
DOB*01:02:02	-	0.0083	0.0113	0.0552	0.0231	0.0039	0.0131	-	0.1027	0.0213
DOB*01:01:01:11	0.0143	0.0296	0.0337	-	-	0.0435	0.0033	-	-	0.0203
DOB*01:01:01:09	-	0.0074	0.0064	0.0166	0.1073	0.0039	0.0033	0.0600	-	0.0195
DOB*01:01:01:05	0.0209	0.0306	0.0363	-	-	0.0078	0.0131	-	0.0008	0.0175
DOB*01:02:01:01	0.0112	0.0102	0.0121	0.0387	0.0289	0.0201	0.0065	0.1600	0.0017	0.0156
DOB*01:01:01:27	-	-	0.0019	0.0028	0.0906	-	-	-	0.0109	0.0150
DOB*01:01:01:23	0.0598	0.0056	0.0117	-	-	-	0.0065	-	-	0.0146
DOB*01:01:05	-	0.0019	0.0034	0.0083	0.0675	-	0.0033	0.1200	0.0225	0.0143
DOB*01:01:01:08	0.0587	0.0028	0.0076	-	-	0.0019	0.0065	-	-	0.0134
DOB*01:01:01:21	0.0066	0.0093	0.0095	-	-	0.0143	0.0229	-	-	0.0072
DOB*01:01:01:26	-	-	0.0008	0.0138	0.0019	-	0.0033	-	0.0376	0.0052
DOB*01:01:01:07	-	-	-	-	0.0019	-	-	-	0.0150	0.0020
DOB*01:16	0.0046	-	0.0026	0.0028	-	-	-	-	-	0.0016
DOB*01:01:01:10	-	0.0019	0.0008	-	-	0.0032	0.0163	-	-	0.0013
DOB*01:01:01:07	-	-	-	-	0.0013	-	-	-	0.0050	0.0007
DOB*01:01:01:11	-	0.0037	-	-	-	-	-	-	-	0.0004
DOB*01:03:01:02	0.0010	-	0.0008	-	-	-	-	-	-	0.0004
DOB*01:08	-	0.0019	0.0008	-	-	-	-	-	-	0.0004
DOB*01:01:01:12	-	-	0.0008	-	-	0.0006	-	-	-	0.0003
DOB*01:01:01:16	-	-	-	-	-	0.0006	0.0065	-	-	0.0003
DOB*01:01:01:24	-	-	-	-	-	-	-	-	0.0025	0.0003
DOB*01:03:01:02	-	-	0.0004	-	-	0.0013	-	-	-	0.0003
DOB*01:15	-	-	0.0011	-	-	-	-	-	-	0.0003
DOB*01:01:01:24	-	-	-	-	-	0.0006	-	-	0.0008	0.0002
DOB*01:01:03:05	0.0010	-	-	-	-	-	-	-	-	0.0002
DOB*01:01:03:06	0.0010	-	-	-	-	-	-	-	-	0.0002
DOB*01:01:05	-	-	-	0.0055	-	-	-	-	-	0.0002
DOB*01:02:01:03	0.0010	-	-	-	-	-	-	-	-	0.0002
DOB*01:01:01:05	-	-	0.0004	-	-	-	-	-	-	0.0001
DOB*01:01:01:22	-	-	-	-	-	-	0.0033	-	-	0.0001
DOB*01:01:01:23	-	-	0.0004	-	-	-	-	-	-	0.0001
DOB*01:01:01:24	-	-	-	-	0.0006	-	-	-	-	0.0001

DOB*01:01:01:24	-	-	-	-	-	0.0006	-	-	-	0.0001
DOB*01:01:01:24	-	-	0.0004	-	-	-	-	-	-	0.0001
DOB*01:01:03:09	-	-	0.0004	-	-	-	-	-	-	0.0001
DOB*01:01:03:09	-	-	-	-	-	0.0006	-	-	-	0.0001
DOB*01:02:01:01	-	-	-	-	-	0.0006	-	-	-	0.0001
DOB*01:02:01:03	0.0005	-	-	-	-	-	-	-	-	0.0001
GENOMIC POTENCIALLY NOVEL	0.2919	0.077	0.0841	0.1106	0.0896	0.049	0.1472	0.06	0.0616	0.1189

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions.

Table S8. *HLA-DOB* three-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	**1958**	**1080**	**2644**	**362**	**1556**	**1540**	**306**	**50**	**1198**	**10694**
DOB*01:01:01	0.4755	0.7278	0.6108	0.6050	0.7481	0.6253	0.6078	0.4600	0.6160	0.6195
DOB*01:01:03	0.0822	0.1130	0.1641	0.2182	0.0257	0.1688	0.1830	-	0.1102	0.1201
DOB*01:02:01	0.1925	0.0407	0.0492	0.0442	0.0289	0.0403	0.0752	0.2200	0.0017	0.0664
DOB*01:04:01	0.0158	0.0204	0.0408	0.0249	0.0643	0.0442	0.0458	0.2000	0.0309	0.0373
DOB*01:03:01	0.0536	0.0565	0.0537	0.0028	0.0032	0.0461	0.0033	-	0.0033	0.0365
DOB*01:04:02	0.0945	0.0102	0.0197	0.0083	0.0096	0.0045	0.0065	-	0.0109	0.0269
DOB*01:05	0.0046	0.0083	0.0193	0.0221	0.0148	0.0357	0.0261	-	0.0935	0.0257
DOB*01:04:03	0.0746	0.0111	0.0197	-	0.0006	0.0279	0.0327	-	0.0017	0.0249
DOB*01:02:02	-	0.0083	0.0113	0.0552	0.0231	0.0039	0.0163	-	0.1043	0.0216
DOB*01:01:05	-	0.0019	0.0038	0.0138	0.0790	-	0.0033	0.1200	0.0242	0.0165
DOB*01:16	0.0046	-	0.0026	0.0028	-	-	-	-	-	0.0016
DOB*01:08	-	0.0019	0.0008	-	-	0.0006	-	-	-	0.0005
DOB*01:15	-	-	0.0011	-	-	-	-	-	-	0.0003
DOB*01:13	-	-	-	-	-	0.0006	-	-	-	0.0001
CODING POTENCIALLY NOVEL	0.002	0	0.0032	0.0028	0.0024	0.0018	0	0	0.0033	0.0024

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions

Table S9. *HLA-DOB* two-field resolution allele in different biogeographic regions. *Brazil group represents samples from SABE database

Allele	AFR	AMR	BRAZIL*	CSA	EAS	EUR	MEA	OCE	SAS	Global
2N	**1958**	**1080**	**2644**	**362**	**1556**	**1540**	**306**	**50**	**1198**	**10694**
DOB*01:01	0.5582	0.8426	0.7791	0.8370	0.8541	0.7942	0.7941	0.5800	0.7529	0.7567
DOB*01:04	0.1849	0.0417	0.0802	0.0331	0.0752	0.0766	0.0850	0.2000	0.0434	0.0892
DOB*01:02	0.1930	0.0491	0.0605	0.0994	0.0521	0.0448	0.0915	0.2200	0.1060	0.0882
DOB*01:03	0.0536	0.0565	0.0537	0.0028	0.0032	0.0461	0.0033	-	0.0033	0.0365
DOB*01:05	0.0046	0.0083	0.0193	0.0221	0.0148	0.0357	0.0261	-	0.0935	0.0257
DOB*01:16	0.0046	-	0.0026	0.0028	-	-	-	-	-	0.0016
DOB*01:08	-	0.0019	0.0008	-	-	0.0006	-	-	-	0.0005
DOB*01:15	-	-	0.0011	-	-	-	-	-	-	0.0003
DOB*01:13	-	-	-	-	-	0.0006	-	-	-	0.0001
PROTEIN POTENTIALLY NOVEL	0.001	0	0.0028	0.0028	0.0006	0.0012	0	0	0.0008	0.0014

Africa (AFR), America (AMR), Central South Asia (CSA), East Asia (EAS), Europe (EUR), Middle East (MEA), Oceania (OCE), and South Asia (SAS). (-) represents allele not found in that determine biogeographic regions

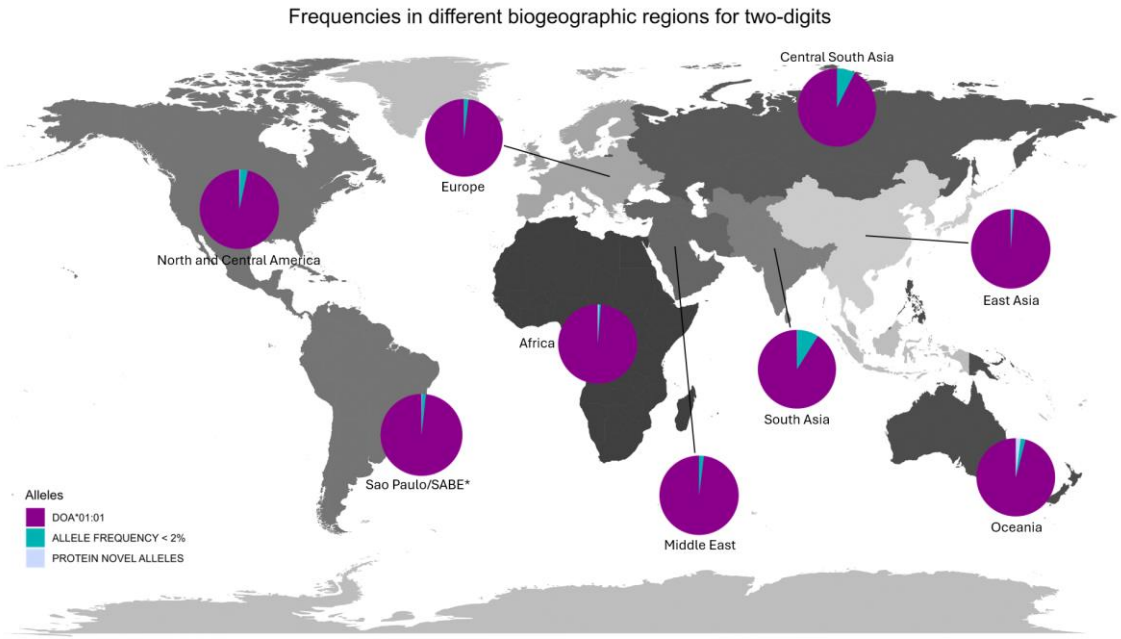

Figure S1. Worldwide frequency of HLA-DOA two-field alleles in different biogeographic regions, Africa, America, Central South Asia, East Asia, Europe, Middle East, Oceania, and South Asia, surveyed from more than 5,000 individuals.

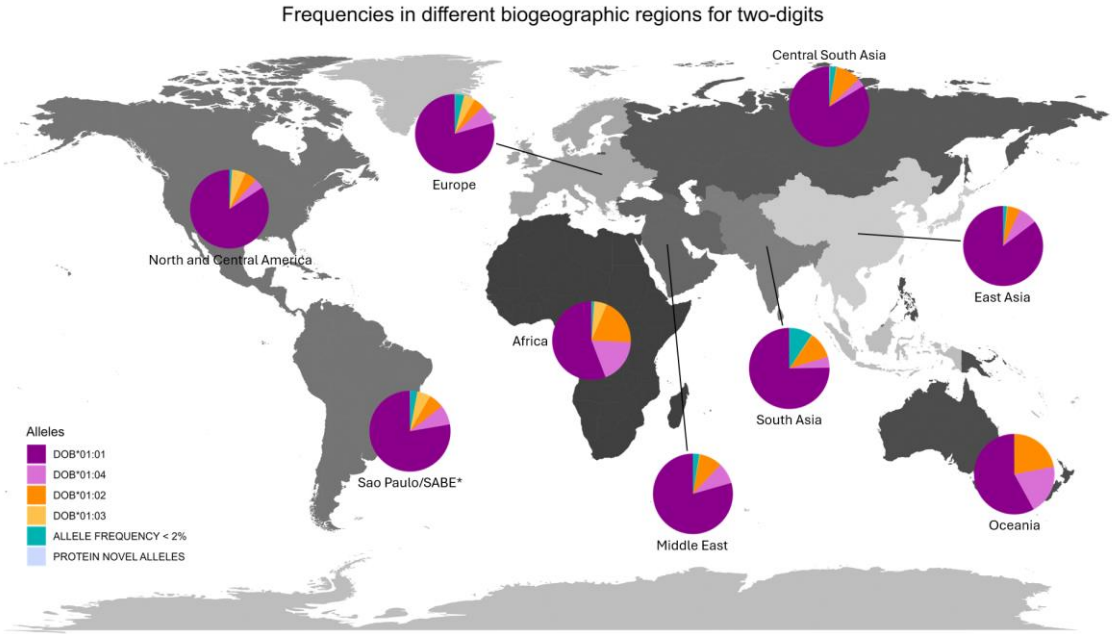

Figure S2. Worldwide frequency of HLA-DOB two-field alleles in different biogeographic regions, Africa, America, Central South Asia, East Asia, Europe, Middle East, Oceania, and South Asia, surveyed from more than 5,000 individuals.

Table S10. First and second SNP combination in *HLA-DOA*, location of occurrence, and significant value of the relationship expressed by D' and r^2. Frequency selection was allele upper than 2%.

Location	First SNP	second SNP	D'	r^2
Block 1	rs13205975	rs13206015	1.0	1.0
Block 1	rs17214106	rs13191715	1.0	0.999
Block 1	rs206763	rs3130602	1.0	0.940
Block 4	rs3129304	rs3129303	0.999	0.995
Block 4	rs9276975	rs9276976	0.982	0.917
Block 5	rs399604	rs365066	1.0	0.999
Block 5	rs369150	rs375256	0.999	0.993
Block 5	rs417812	rs2581	0.992	0.965
Block 8	rs6457694	rs16871338	1.0	0.997
Block 8	rs114209907	rs16871338	1.0	0.904
Block 8	rs6457694	rs114209907	1.0	0.901
Block 9	rs6911639	rs9276982	1.0	0.999

Table S11. First and second SNP combination in *HLA-DOB*, location of occurrence, and significant value of the relationship expressed by D' and r^2. Frequency selection was allele upper than 2%.

Location	First SNP	second SNP	D'	r^2
Block 1	rs2857113	rs2621332	1.0	0.999
Block 1	rs2857113	rs2621331	1.0	0.742
Block 1	rs2621332	rs2621331	0.999	0.742
Block 2	rs2070120	rs17213693	1.0	1.0
Block 2	rs2070121	rs16870880	1.0	1.0
Block 2	rs11967235	rs17213728	1.0	0.998
Block 3	rs2859579	rs2071472	1.0	1.0
Block 3	rs2859579	rs2071470	1.0	1.0
Block 3	rs2071472	rs2071470	1.0	1.0
Block 4	rs7742120	rs7762120	1.0	1.0
Block 4	rs2071547	rs7742120	1.0	0.999
Block 4	rs2071547	rs2071468	1.0	0.999

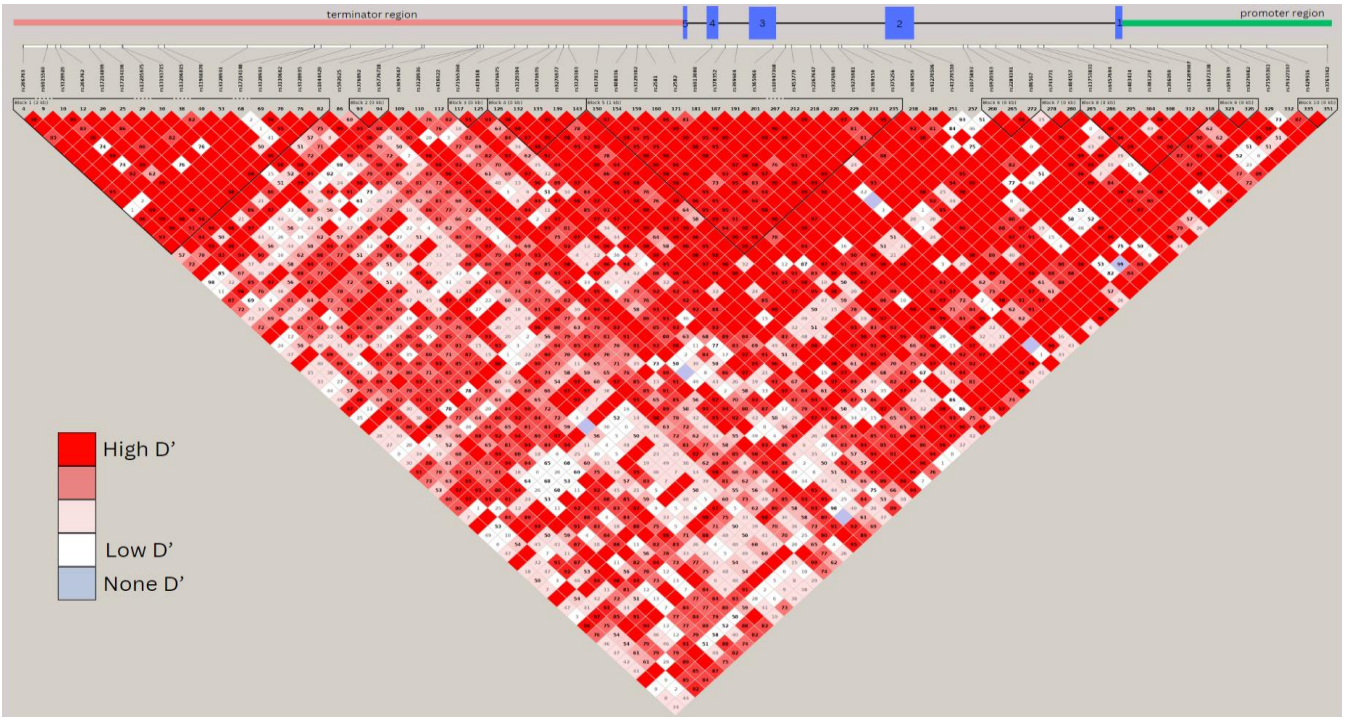

Figure S3. *HLA-DOA* LD plot of allele association measure (D'). Above, there is the *HLA-DOA* sequence in the promoter region, represented by a green square, exons by blue squares, lines between them are introns, and the terminator region is represented by a red square. In the triangle and legend, redder squares indicate higher D', while whiter squares indicate lower D' and the light blue D' relation is nonexistent, with frequency upper than 2%.

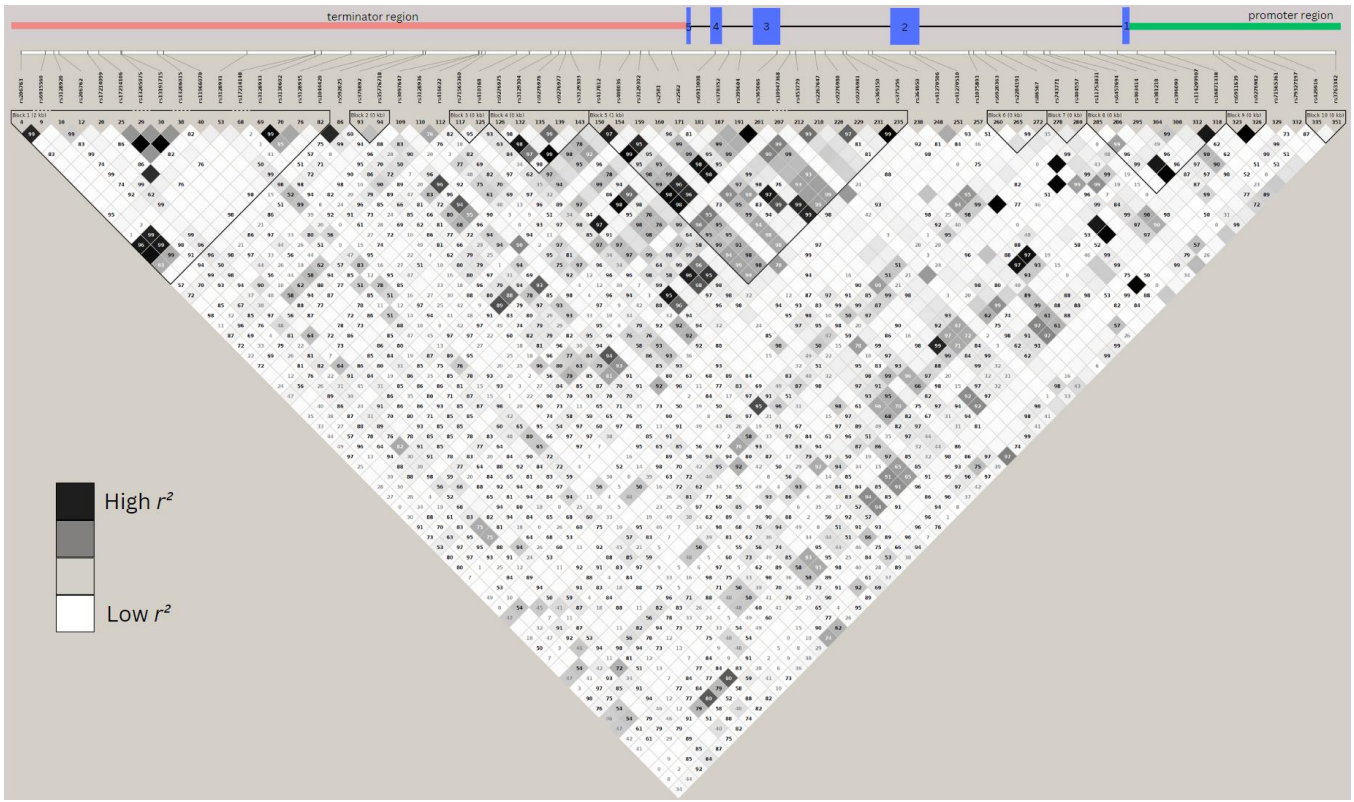

Figure S4. *HLA-DOA* LD plot of allele association in different *loci* measure (r^2). Above, there is the *HLA-DOA* sequence in the promoter region, represented by a green square, exons by blue squares, lines between them are introns, and the terminator region is represented by a red square. In the triangle and legend, blacker squares indicate higher r^2, while grayner squares indicate lower r^2, as all frequency upper than 2%.

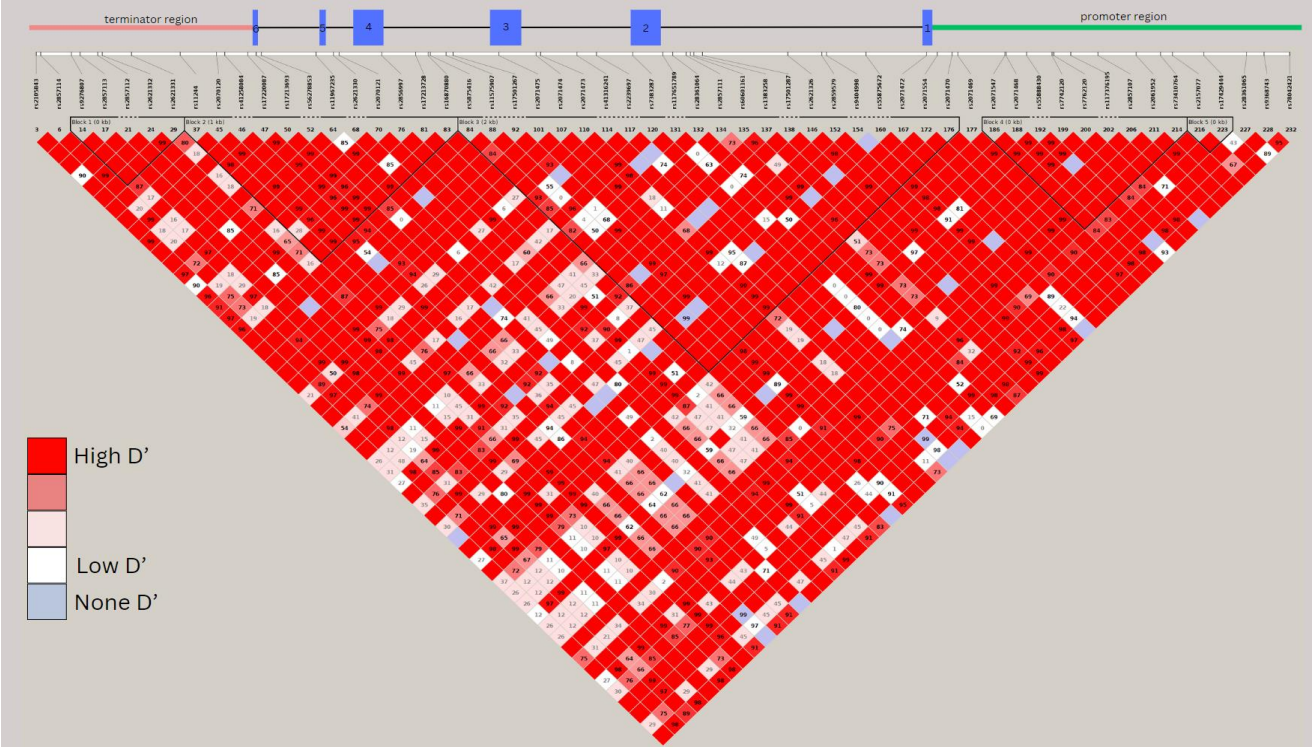

Figure S5. *HLA-DOB* LD plot of allele association measure (D'). Above, there is the HLA-DOB sequence in the promoter region, represented by a green square, exons by blue squares, lines between them are introns, and the terminator region is represented by a red square. In the triangle and legend, redder squares indicate higher D', while whiter squares indicate lower D' and the light blue D' relation is nonexistent, with frequency upper than 2%.

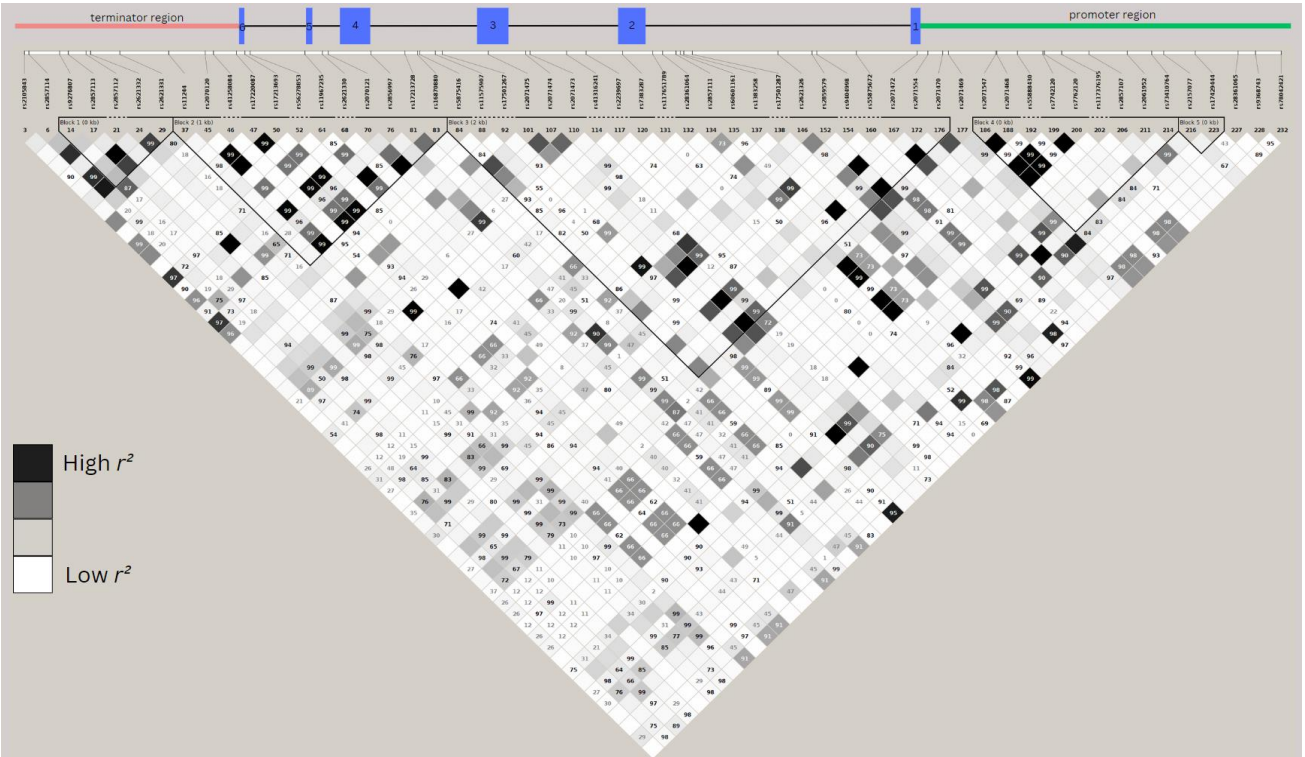

Figure S6. LD plot of allele association in different *loci* measure (r^2). Above, there is the *HLA-DOB* sequence in the promoter region, represented by a green square, exons by blue squares, lines between them are introns, and the terminator region is represented by a red square. In the triangle and legend, blacker squares indicate higher r^2, while grayner squares indicate lower, as all frequency upper than 2%.