

RESSALVA

Atendendo solicitação da
autora, o texto completo desta
dissertação, será
disponibilizado somente a
partir de 27/05/2026.



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
Imunogenética dos genes HLA-DOA e HLA-DOB: variabilidade,
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.

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 (Participação Presencial)
Departamento de Patologia / Faculdade de Medicina de Botucatu Unesp

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

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

Botucatu, 27 de novembro de 2024



Documento assinado digitalmente

FLÁVIA 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 africano, europeu e americano (sul e norte). Adicionalmente, avaliamos os perfis de frequência alélica para cada região biogeográfica. Além disso, mesmo incluindo 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

Figura 1. Representação da via MHC de classe II de apresentação de peptídeos	12
Figura 2. Representação da interação dos receptores da APC e o LT CD4+	13
Figura 3. Representação gráfica da ordem dos genes HLA na região do MHC humano	14
Figura 4. Representação da molécula de HLA de classe II	15
Figura 5. Representação esquemática da estrutura do gene <i>HLA-DOA</i> ...	17
Figura 6. Representação esquemática da estrutura do gene <i>HLA-DOB</i> ...	17
Figura 7. Representação tridimensional da estrutura de HLA-DO disponível no SWISS-MODEL	18
Figura 8. Representação do processamento e carregamento de peptídeo de HLA de classe II	19
Figura 9. Mapa extraído do <i>The International Genomes Sample Resource</i>	23
Figura 10. Mapa mundial demonstrando os países de origem das populações/etnias responsáveis pelo povoamento do Brasil	25

Lista de Figuras do Artigo

Figure 1. Representation of HLA Class II antigen processing and presentation	36
Figure 2. Scheme of the HLA genotyping workflow	39
Figure 3. Representation of frameshift, missense, and non-coding variants across the <i>HLA-DOA</i> gene	40
Figure 4. <i>HLA-DOA</i> four-field allele frequencies in different biogeographic regions	41
Figure 5. Worldwide frequency of <i>HLA-DOA</i> three-field alleles in different biogeographic regions	42
Figure 6. Representation of frameshift, missense, and non-coding variants across the <i>HLA-DOB</i> gene	43
Figure 7. <i>HLA-DOB</i> four-field allele frequencies in different biogeographic regions	44
Figure 8. Worldwide frequency of <i>HLA-DOB</i> three-field alleles in different biogeographic regions	45
Figure 9. Nucleotide diversity throughout <i>HLA-DOA</i>	46
Figure 10. Nucleotide diversity throughout <i>HLA-DOB</i>	47

Lista de Figuras dos Materiais Suplementares

Figure S1. Worldwide frequency of <i>HLA-DOA</i> two-field alleles in different biogeographic regions	81
Figure S2. Worldwide frequency of <i>HLA-DOB</i> two-field alleles in different biogeographic regions	81
Figure S3. HLA-DOA LD plot of allele association measure (D')	84
Figure S4. HLA-DOA LD plot of allele association in different <i>loci</i> measure (r^2)	84
Figure S5. HLA-DOB LD plot of allele association measure (D')	85
Figure S6. HLA-DOB LD plot of allele association in different <i>loci</i> measure (r^2)	85

Lista de Tabelas

Tabela 1. Número de alelos descritos para os genes HLA atualmente	16
--	----

Lista de Tabelas dos Materiais Suplementares

Table S1. List of samples used in this study	54
Table S2. HLA-DOA SNP frequency	54
Table S3. HLA-DOB SNP frequency	62
Table S4. HLA-DOA four-field resolution allele in different biogeographic regions	75
Table S5. HLA-DOA three-field resolution allele in different biogeographic regions	76
Table S6. HLA-DOA two-field resolution allele in different biogeographic regions	77
Table S7. HLA-DOB four-field resolution allele in different biogeographic regions	78
Table S8. HLA-DOB three-field resolution allele in different biogeographic regions	79
Table S9. HLA-DOB two-field resolution allele in different biogeographic regions	80
Table S10. First and second SNP combination in HLA-DOA	82
Table S11. First and second SNP combination in HLA-DOB	83

SUMÁRIO

Capítulo I – Revisão da Literatura

1. INTRODUÇÃO

1.1 Resposta imunológica adaptativa e a via de apresentação de antígeno de classe II	11
1.2 O complexo principal de histocompatibilidade	13
1.3 Os genes <i>HLA-DOA</i> e <i>HLA-DOB</i> e a função de HLA-DO	16
1.4 A história evolutiva dos HLA e a diversidade dos genes <i>HLA-DOA</i> e <i>HLA-DOB</i>	20
1.5 População brasileira e os bancos de dados internacionais	21

2. JUSTIFICATIVA	26
------------------------	----

3. OBJETIVOS GERAIS

3.1 Objetivos Gerais	26
3.2 Objetivos Específicos	26

REFERÊNCIAS BIBLIOGRÁFICAS	27
----------------------------------	----

Capítulo II – Artigo

Abstract	34
Introduction	35
Samples and Methods	37
Results	40
Discussion	47
Conclusion	49
Reference	51

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-STERMBERG, M. 2023. The impact of immuno-peptidomics: 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>.