

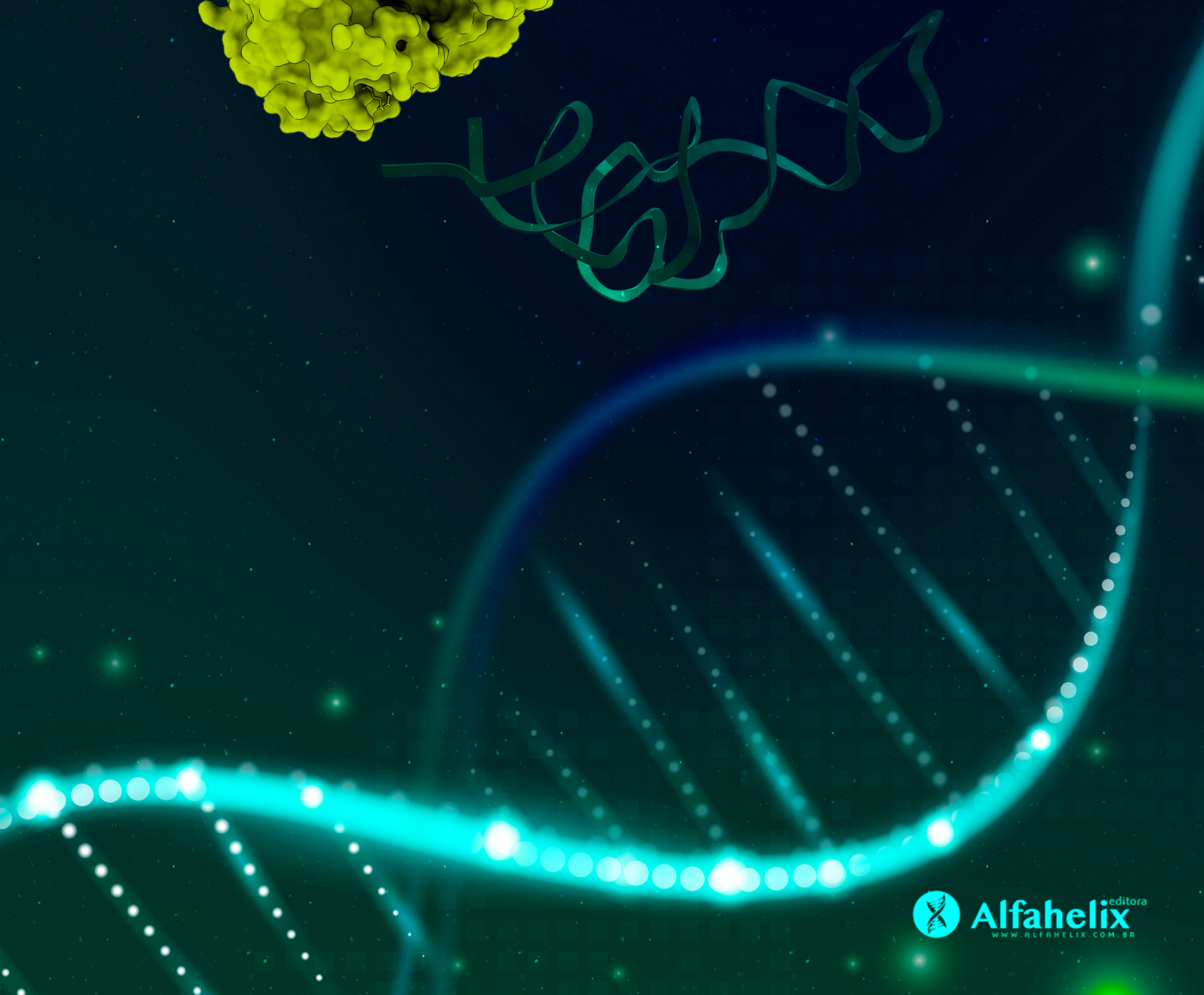
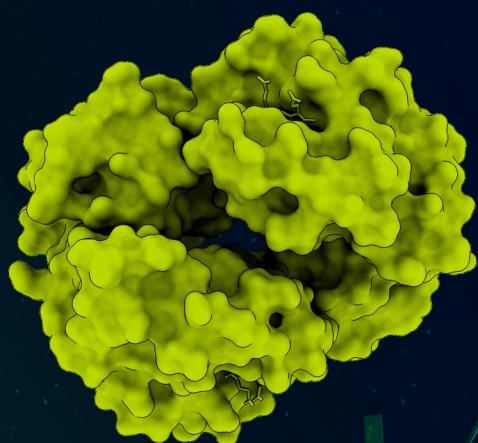


bioinfo

www.bioinfo.com.br

Revista Brasileira de Bioinformática
e Biologia Computacional

Edição especial



BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional

ISBN: 978-65-84391-00-0 | doi: 10.51780/978-65-84391-00-0

www.bioinfo.com.br

Vol. 6 - Fev. 2026 - Edição especial

Organização & Revisão

 Ariany Rosa Gonçalves

Doutora em Genética e Melhoramento de Plantas - IF Goiano

 Bibiana Fam

Doutora em Genética e Biologia Molecular - UFRGS

 Lucianna Helene Santos

Doutora em Biologia Computacional e Sistemas / Pesquisadora *Institut Pasteur de Montevideo* (Uruguai)

 Marcos Antonio Nobrega de Sousa

Doutor em Genética - UFCG

 Mira Raya Paula de Lima

Doutora em Química Biológica - IFCE

 Rafael Eduardo Oliveira Rocha

Doutor em Bioquímica - UFMG / Pesquisador *Universität Hamburg* (Alemanha)

 Savio Costa

Doutor em Genética e Biologia Molecular - UFPA

 Sheila Cruz Araújo

Doutora em Química - UFABC

 Tatiane Senna Bialves

Doutora em Ciências Fisiológicas - FURG

 Thiago Jesus Sousa

Doutor em Bioinformática - UFMG / Pesquisador Lacen-ES


Comitê editorial

Diego Mariano 

Residente pós-doutoral | Departamento de Ciência da Computação (UFMG) | Editor-in-chief Alfahelix


Alessandra Cioletti 

Doutora em química - UFMG | Departamento de Ciência da Computação (UFMG)


Alessandra Lima da Silva 

Residente pós-doutoral | Departamento de Ciência da Computação (UFMG)

Supervisão

Flavia Figueira Aburjaile 

Professora | Programa Interunidades de Pós-graduação em Bioinformática (UFMG)

Raquel Cardoso de Melo-Minardi 

Professora | Programa Interunidades de Pós-graduação em Bioinformática (UFMG)

Ficha catalográfica

Sandro Alex Batista CRB6/2433 - Bibliotecário

Diagramação

Alessandra Lima

Diego Mariano

Lucas Moraes

Publicação

Editora Alfahelix, CNPJ: 37.524.984/0001-10

Lagoa Santa, MG, Brasil

www.alfahelix.com.br

Capa

Adaptado de rawpixel.com/Freepik. Estruturas da hemoglobina e do tRNA foram obtidas no PDB e renderizadas com ChimeraX.

Agradecimentos

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais

UFMG - Universidade Federal de Minas Gerais

ICEx - Instituto de Ciências Exatas da UFMG

DCC - Departamento de Ciência da Computação da UFMG

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Rede BIOINFO.com - a rede de divulgação científica em bioinformática - Fundação de Amparo à Pesquisa do Estado de Minas Gerais (APQ-02690-22).

Apoio



**PROGRAMA INTERUNIDADES
DE PÓS-GRADUAÇÃO EM
BIOINFORMÁTICA DA UFMG**

M333b MARIANO, Diego (org.) et al

BIOINFO #06: Revista Brasileira de Bioinformática e Biologia Computacional / Diego Mariano; Ariany Rosa Gonçalves et al (org.) et al. Lagoa Santa, MG: Alfahelix, Fev de 2026. V.6. (Série BIOINFO).
186p. il.: Imagens (colors).

ISBN: 978-65-84391-00-0

DOI: <https://doi.org/10.51780/978-65-84391-00-0>

1. Bioinformática. 2. Computação. 3. Biologia. 4. GONÇALVES. Ariany Rosa. I.
Título.

CDD: 006.5765

CDU: 004.89/576

Ficha Catalográfica – Bibliotecário: Sandro Alex Batista CRB6/2433



BIOINFO: *free & open access*

Sem custos para autores. Sem custos para leitores.

Este livro compila os artigos publicados na Revista BIOINFO Vol. 6 (2026) - Edição Especial em parceria com o Programa Interunidades de Pós-Graduação em Bioinformática da UFMG. Todos os direitos autorais pertencem aos autores de cada respectivo artigo. Todavia, os autores concordam em compartilhar gratuitamente o conteúdo deste livro e incentivam sua livre distribuição (desde que os autores dos respectivos capítulos sejam corretamente citados e/ou que as leis de uso justo sejam respeitadas). Este livro está compartilhado sob a licença Creative Commons Atribuição 4.0 Internacional - Não comercial (CC BY-NC 4.0). Você pode utilizar qualquer conteúdo aqui apresentado, desde que cite:

Cite este artigo 0.1

Mariano (org.); et al. BIOINFO #06 - Revista Brasileira de Bioinformática e Biologia Computacional. 6. Ed. Vol. 6. Lagoa Santa: Alfahelix, 2026. doi: 10.51780/978-65-84391-00-0

Esta é uma obra digital. Note que fazemos uso de links e outras propriedades de hipertexto, o que pode limitar a qualidade de edições impressas deste manuscrito. Entretanto, sinta-se à vontade caso deseje imprimir este conteúdo. A cópia por meio impresso ou digital (PDF) é permitida, sendo exclusivamente vedada a venda visando lucro. Detalhes sobre a licença de uso estão disponíveis em: <https://bioinfo.com.br/licenca-de-uso/>

Artigos & Autores - BIOINFO #06 (Fev. 2026)

ID	Título	Autores	DOI
0	Editorial	Lucas Moraes dos Santos, Adenilson Arcanjo de Moura Junior, Luana Luiza Bastos, Diego Mariano, Flavia Figueira Aburjaile, Raquel Cardoso de Melo Minardi	10.51780/bioinfo-06-00
1	Bioinformática e o Combate à Desinformação em Saúde: Como as Ciências Ômicas Podem Reforçar a Confiança Pública na Pesquisa Biomédica	Ruan Vitor Santos Silva, Gabriel Quintanilha Peixoto, Aristóteles Góes-Neto	10.51780/bioinfo-06-01
2	Índices de correlação entre a expressão de Genes Condutores e a Desdiferenciação Celular em Câncer	Leonardo Henrique da Silva, Milenna Machado Pirovani, Raquel Cardoso de Melo Minardi	10.51780/bioinfo-06-02
3	Influência de Variáveis Climáticas na Dinâmica Vetorial de Arboviroses no Brasil	Fernanda Elen Silva Dos Santos, Ronison Alves Guimarães, Vagner Fonseca, Joicymara Santos Xavier	10.51780/bioinfo-06-03
4	Vigilância Genômica e Predição: Impactos e Desafios	Ronison Alves Guimarães, Fernanda Elen Silva Dos Santos, Vagner Fonseca, Joicymara Santos Xavier	10.51780/bioinfo-06-04
5	Síntese Quantitativa de Modelos Matemáticos, Estatísticos e de Inteligência Artificial para Biomarcadores, Mecanismos e Reposicionamento Terapêutico	Rossana de Oliveira Souza, Wellington Francisco Rodrigues, Aristóteles Góes-Neto, Marcos Augusto dos Santos	10.51780/bioinfo-06-05
6	Explorando a Patogenicidade e Resistência de <i>C. aurimucosum</i>	Luan Freitas Lana et al.	10.51780/bioinfo-06-06
7	Caracterização Genômica e Fenotípica de <i>Corynebacterium pseudotuberculosis</i> : utilização da Bioinformática no Controle da Linfadenite Caseosa	Maria Leticia Carneiro Rodrigues et al.	10.51780/bioinfo-06-07
8	Desvendando a resistência em <i>Corynebacterium striatum</i> : uma investigação multidisciplinar do genoma ao desenho de inibidores	Diogo Luiz de Carvalho Castro et al.	10.51780/bioinfo-06-08
9	Os genes das estruturas reprodutivas de mamíferos: quão ancestrais são suas origens?	Kadu Penuela Sanches Estevam et al.	10.51780/bioinfo-06-09

ID	Título	Autores	DOI
10	Estudo integrativo genômico e de virulência de <i>Corynebacterium auriscanis</i> : elucidando a patogenicidade	Evandro Bento Rodrigues et al.	10.51780/bioinfo-06-10
11	Genômica comparativa de <i>Ornithorhynchus anatinus</i> revela genes mais próximos de tartarugas do que de outros mamíferos	Ryan Fernandes Vieira de Souza, J. Miguel Ortega	10.51780/bioinfo-06-11
12	Do Modelo ao Movimento: AlphaFold e CABS-flex na Era da Bioinformática Estrutural	Hugo Martins, Adenilson Arcanjo de Moura Junior, Raquel Melo Minardi	10.51780/bioinfo-06-12
13	Por que ensinar Bioinformática no Ensino Médio?	Helena Lott Costa et al.	10.51780/bioinfo-06-13
14	Pseudogenes podem ser peças-chave na diversidade de anticorpos de cavalos	Juliana Edelvacy Lima Pinto et al.	10.51780/bioinfo-06-14
15	A toxicologia in silico no desenvolvimento de fármacos: um caminho ético e inovador na avaliação de segurança	Anne Josiele de Lima Vital et al.	10.51780/bioinfo-06-15
16	As conexões invisíveis da vida: o papel das interações proteína-RNA e suas aplicações tecnológicas	Luana Luiza Bastos, Tatiane Senna Bialves, Raquel Melo Minardi	10.51780/bioinfo-06-16
17	Desvendando o risco renal sob coexposição química: do expossoma às redes biológicas	Bárbara Amaral et al.	10.51780/bioinfo-06-17
18	Valorizando o semiárido mineiro através da utilização de inulinas da biomassa de Agave spp.	Eduardo Horta Santos et al.	10.51780/bioinfo-06-18
19	Detecção de peptídeos bioativos em dermocosméticos: abordagens computacionais	Milenna Machado Pirovani, Raquel Melo Minardi	10.51780/bioinfo-06-19

ID	Título	Autores	DOI
20	Fusão de Descritores Biológicos em Grafos de Resíduos para Predição da Oncogenicidade de Variantes	Lucas Moraes dos Santos, Tatiane Senna Bialves, Raquel Melo Minardi	10.51780/bioinfo-06-20
21	Da Molécula ao Algoritmo: Utilizando Aprendizado de Máquina para Desvendar a O-Glicosilação de Proteínas	Adenilson Arcanjo de Moura Junior et al.	10.51780/bioinfo-06-21
22	Bioinformática Translacional: Integração Multi-Ômica e Inteligência Artificial Aplicada a Periodontite	Wellington Francisco Rodrigues et al.	10.51780/bioinfo-06-22
23	O desafio de trocar muitos resíduos de uma proteína mantendo-a estável	Juliana Rodrigues Pereira Silva et al.	10.51780/bioinfo-06-23
24	Um Passo Rumo a Novos Fármacos contra Doenças Negligenciadas: Mapeamento Computacional Estrutural da Proteína KSRP	Leandro Moraes et al.	10.51780/bioinfo-06-24
25	Bioinformática na sanidade animal: avanços para a saúde única e produção sustentável de organismos aquáticos	Rebeca Dias Serafim-Lima et al.	10.51780/bioinfo-06-25
26	Desafios e impactos na modulação da microbiota intestinal e no tratamento de doenças inflamatórias intestinais	Eduarda Guimarães Sousa et al.	10.51780/bioinfo-06-26
27	Ilhas genômicas: pontos isolados em um oceano de informações	Diego Lucas Neres Rodrigues et al.	10.51780/bioinfo-06-27
28	Como a probiogenômica pode revelar novos probióticos: o exemplo de <i>Streptococcus thermophilus</i>	Juan Carlos Ariute et al.	10.51780/bioinfo-06-28
29	Implicações sustentáveis do Sequenciamento de Nova Geração para a caracterização de linhagens isoladas de <i>Bacillus paralicheniformis</i> com interesse biotecnológico	Gabriel Camargos Gomes et al.	10.51780/bioinfo-06-29
30	Estudo pangenômico de <i>Corynebacterium rouxii</i> e <i>Corynebacterium ulcerans</i> com foco em determinantes genéticos de virulência, resistência antimicrobiana e impacto zoonótico	Fernanda Prates et al.	10.51780/bioinfo-06-30
31	Desafios e Impactos da Pesquisa na Sociedade: Formando Bioinformatas para uma Década de Transformações	Catarina Rodrigues Gomes, Arun Kumar Jaiswal, Vasco Azevedo	10.51780/bioinfo-06-31

SUMÁRIO

CAPÍTULO 1

**BIOINFORMÁTICA E O COMBATE À DESINFORMAÇÃO EM SAÚDE:
COMO AS CIÊNCIAS ÔMICAS PODEM REFORÇAR A CONFIANÇA
PÚBLICA NA PESQUISA BIOMÉDICA** _____ **PÁGINA 20** _____

CAPÍTULO 2

**ÍNDICES DE CORRELAÇÃO ENTRE A EXPRESSÃO DE GENES
CONDUTORES E A DESDIFERENCIAÇÃO CELULAR EM CÂNCER**
PÁGINA 27 _____

CAPÍTULO 3

**INFLUÊNCIA DE VARIÁVEIS CLIMÁTICAS NA DINÂMICA VETORIAL DE
ARBOVIROSES NO BRASIL** _____ **PÁGINA 31** _____

CAPÍTULO 4

VIGILÂNCIA GENÔMICA E PREDIÇÃO: IMPACTOS E DESAFIOS
PÁGINA 36 _____

CAPÍTULO 5

**SÍNTESE QUANTITATIVA DE MODELOS MATEMÁTICOS, ESTATÍSTICOS
E DE INTELIGÊNCIA ARTIFICIAL PARA BIOMARCADORES,
MECANISMOS E REPOSICIONAMENTO TERAPÊUTICO** _____ **PÁGINA 41** _____

CAPÍTULO 6

**EXPLORANDO A PATOGENICIDADE E RESISTÊNCIA DE
Corynebacterium aurimucosum POR ABORDAGENS GENÔMICAS**
PÁGINA 45 _____

CAPÍTULO 7

CARACTERIZAÇÃO GENÔMICA E FENOTÍPICA DE *Corynebacterium pseudotuberculosis*: UTILIZAÇÃO DA BIOINFORMÁTICA NO CONTROLE DA LINFADENITE CASEOSA _____ **PÁGINA 51** _____

CAPÍTULO 8

DESVENDANDO A RESISTÊNCIA EM *Corynebacterium striatum*: UMA INVESTIGAÇÃO MULTIDISCIPLINAR DO GENOMA AO DESENHO DE INIBIDORES _____ **PÁGINA 56** _____

CAPÍTULO 9

OS GENES DAS ESTRUTURAS REPRODUTIVAS DE MAMÍFEROS: QUÃO ANCESTRAIS SÃO SUAS ORIGENS? _____ **PÁGINA 63** _____

CAPÍTULO 10

ESTUDO INTEGRATIVO GENÔMICO E DE VIRULÊNCIA DE *Corynebacterium auriscanis*: ELUCIDANDO A PATOGENICIDADE
PÁGINA 70 _____

CAPÍTULO 11

GENÔMICA COMPARATIVA DE *Ornithorhynchus anatinus* REVELA GENES MAIS PRÓXIMOS DE TARTARUGAS DO QUE DE OUTROS MAMÍFEROS _____ **PÁGINA 76** _____

CAPÍTULO 12

DO MODELO AO MOVIMENTO: ALPHA FOLD E CABS-FLEX NA ERA DA BIOINFORMÁTICA ESTRUTURAL _____ **PÁGINA 82** _____

CAPÍTULO 13

POR QUE ENSINAR BIOINFORMÁTICA NO ENSINO MÉDIO? _____ **PÁGINA 87** _____

CAPÍTULO 14

PSEUDOGENES PODEM SER PEÇAS-CHAVE NA DIVERSIDADE DE ANTICORPOS DE CAVALOS _____ **PÁGINA 91** _____

CAPÍTULO 15

A TOXICOLOGIA *in silico* NO DESENVOLVIMENTO DE FÁRMACOS: UM CAMINHO ÉTICO E INOVADOR NA AVALIAÇÃO DE SEGURANÇA
PÁGINA 98 _____

CAPÍTULO 16**AS CONEXÕES INVISÍVEIS DA VIDA: O PAPEL DAS INTERAÇÕES****PROTEÍNA-RNA E SUAS APLICAÇÕES TECNOLÓGICAS** _____ **PÁGINA 103** _____**CAPÍTULO 17****DESVENDANDO O RISCO RENAL SOB COEXPOSIÇÃO QUÍMICA: DO****EXPOSSOMA ÀS REDES BIOLÓGICAS** _____ **PÁGINA 111** _____**CAPÍTULO 18****VALORIZANDO O SEMIÁRIDO MINEIRO ATRAVÉS DA UTILIZAÇÃO DE****INULINAS DA BIOMASSA DE *Agave spp.*** _____ **PÁGINA 117** _____**CAPÍTULO 19****DETECÇÃO DE PEPTÍDEOS BIOATIVOS EM DERMOCOSMÉTICOS:****ABORDAGENS COMPUTACIONAIS** _____ **PÁGINA 122** _____**CAPÍTULO 20****FUSÃO DE DESCRITORES BIOLÓGICOS EM GRAFOS DE RESÍDUOS****PARA PREDIÇÃO DA ONCOGENICIDADE DE VARIANTES** _____ **PÁGINA 126** _____**CAPÍTULO 21****DA MOLÉCULA AO ALGORITMO: UTILIZANDO APRENDIZADO DE****MÁQUINA PARA DESVENDAR A O-GLICOSILAÇÃO DE PROTEÍNAS****PÁGINA 131** _____**CAPÍTULO 22****BIOINFORMÁTICA TRANSLACIONAL: INTEGRAÇÃO MULTIÔMICA E****INTELIGÊNCIA ARTIFICIAL APLICADA A PERIODONTITE** _____ **PÁGINA 136** _____**CAPÍTULO 23****O DESAFIO DE TROCAR MUITOS RESÍDUOS DE UMA PROTEÍNA****MANTENDO-A ESTÁVEL** _____ **PÁGINA 142** _____**CAPÍTULO 24****UM PASSO RUMO A NOVOS FÁRMACOS CONTRA DOENÇAS****NEGLIGENCIADAS: MAPEAMENTO COMPUTACIONAL ESTRUTURAL****DA PROTEÍNA KSRP** _____ **PÁGINA 147** _____

CAPÍTULO 25

BIOINFORMÁTICA NA SANIDADE ANIMAL: AVANÇOS PARA A SAÚDE ÚNICA E PRODUÇÃO SUSTENTÁVEL DE ORGANISMOS AQUÁTICOS
PÁGINA 151

CAPÍTULO 26

DESAFIOS E IMPACTOS NA MODULAÇÃO DA MICROBIOTA INTESTINAL E NO TRATAMENTO DE DOENÇAS INFLAMATÓRIAS INTESTINAIS **PÁGINA 156**

CAPÍTULO 27

ILHAS GENÔMICAS: PONTOS ISOLADOS EM UM OCEANO DE INFORMAÇÕES **PÁGINA 161**

CAPÍTULO 28

COMO A PROBIOGENÔMICA PODE REVELAR NOVOS PROBIÓTICOS: O EXEMPLO DE *Streptococcus thermophilus* **PÁGINA 166**

CAPÍTULO 29

IMPLICAÇÕES SUSTENTÁVEIS DO SEQUENCIAMENTO DE NOVA GERAÇÃO PARA A CARACTERIZAÇÃO DE LINHAGENS ISOLADAS DE *Bacillus paralicheniformis* COM INTERESSE BIOTECNOLÓGICO
PÁGINA 171

CAPÍTULO 30

ESTUDO PANGENÔMICO DE *Corynebacterium rouxii* E *Corynebacterium ulcerans* COM FOCO EM DETERMINANTES GENÉTICOS DE VIRULÊNCIA, RESISTÊNCIA ANTIMICROBIANA E IMPACTO ZONÓTICO **PÁGINA 175**

CAPÍTULO 31

DESAFIOS E IMPACTOS DA PESQUISA NA SOCIEDADE: FORMANDO BIOINFORMATAS PARA UMA DÉCADA DE TRANSFORMAÇÕES
PÁGINA 182

Editorial BIOINFO #06 – Bioinformática 2035: Desafios e Formação de Talentos para a Próxima Década

Autores 0.1

Lucas Moraes dos Santos , Adenilson Arcanjo de Moura Junior , Luana Luiza Bastos , Diego Mariano , Flávia Figueira Aburjaile , Raquel Cardoso de Melo Minardi 

Revisão: Alessandra Lima 

Resumo 0.1

A 6ª edição da Revista BIOINFO, organizada em colaboração com o Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, reúne 31 artigos, além deste editorial. Os trabalhos publicados nesta edição foram apresentados no evento “Bioinformática 2035: Desafios e Formação de Talentos para a Próxima Década”. Este editorial apresenta uma síntese do evento, destacando seus principais temas, discussões e conclusões.

Introdução

Nos dias 24 e 25 de novembro de 2025, foi realizado o evento “Bioinformática 2035: Desafios e Formação de Talentos para a Próxima Década”, nas dependências da Universidade Federal de Minas Gerais (UFMG), no Centro de Atividades Didáticas 3 (CAD3), auditório B101/102, em Belo Horizonte, Minas Gerais, Brasil. O evento foi organizado pelo Programa Interunidades de Pós-Graduação em Bioinformática da UFMG (PPG-BIOINFO).

O evento contou com 374 inscritos, distribuídos entre 135 estudantes de graduação, 77 mestrandos, 83 doutorandos, 23 pós-doutorandos, 25 docentes e 31 profissionais atuantes em diferentes setores tecnológicos, além dos palestrantes e do comitê de organização.

Em parceria com a Revista BIOINFO, o evento abriu uma chamada para submissão de trabalhos destinados a uma edição especial dedicada às pesquisas desenvolvidas no âmbito do PPG-BIOINFO. Os estudantes foram convidados a escrever um *short paper* sobre o tema “Desafios e Impactos da Pesquisa na Sociedade”. Neste manuscrito, os autores deveriam apresentar uma breve introdução contextualizando sua área de estudo, descrever um problema científico ou tecnológico que sua pesquisa busca solucionar e discutir as estratégias e abordagens empregadas para enfrentá-lo. Os artigos deveriam ter aproximadamente três páginas, com estrutura livre. Solicitou-se, ainda, que os autores evitassem incluir resultados inéditos que pudessem ser submetidos posteriormente a periódicos científicos de alto impacto.

Os trabalhos aprovados pelo comitê editorial da revista foram convidados a apresentar suas pesquisas em pôsteres durante o evento. Ao todo, 34 trabalhos foram apresentados nessa modalidade. Além disso, seis trabalhos de estudantes de mestrado e seis de doutorado foram selecionados para apresentação oral. Estudantes não vinculados ao PPG-BIOINFO também puderam submeter trabalhos sobre o tema “A importância da sua pesquisa”, que foram publicados na quinta edição da Revista BIOINFO.

No total, a Revista BIOINFO recebeu 70 submissões, das quais 56 foram aprovadas para publicação. Dentre essas, 25 artigos de autores não vinculados ao PPG-BIOINFO foram publicados na quinta edição da revista [1,2]. Nesta edição, apresentamos 31 artigos de estudantes de mestrado e doutorado do PPG-BIOINFO. Este editorial também sintetiza as palestras e as principais conclusões do evento.

Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

O Programa de Pós-Graduação em Bioinformática da UFMG é um dos mais antigos e consolidados do país, resultado de uma iniciativa pioneira no início dos anos 2000 em resposta ao edital BIOMICRO da CAPES. Criado em 2002 pelos Departamentos de Bioquímica e Imunologia (DBI) e de Ciência da Computação (DCC), em parceria com outros cinco departamentos dos Institutos de Ciências Biológicas, Ciências Exatas e da Escola de Engenharia, o programa teve sua primeira turma em 2003 e está sediado no Instituto de Ciências Biológicas (ICB), que concentra a maior parte de seu corpo docente. Com conceito CAPES 7, consolidou-se como referência nacional na formação de recursos humanos altamente qualificados em uma área estratégica que integra Ciências Biológicas e Ciência da Computação, atuando em frentes como análise de

sequências de DNA e proteínas, bioinformática estrutural, biologia de sistemas, ômicas, inteligência artificial, visualização e análise de imagens e sinais biológicos.

Ao longo de sua trajetória, o PPG em Bioinformática formou 142 mestres e 186 doutores, muitos dos quais hoje ocupam posições de liderança em universidades, empresas, hospitais e órgãos de governo. O corpo docente é fortemente interdisciplinar, reunindo profissionais das áreas de biologia, computação, medicina, medicina veterinária, estatística e farmácia, com sólida capacidade de captação de recursos e coordenação de redes nacionais e internacionais. A produção científica é expressiva: no último quadriênio, foram mais de 1.600 artigos publicados (segundo a base Scopus), totalizando mais de 14.724 citações, incluindo contribuições relevantes em temas de grande impacto, como a emergência do coronavírus SARS-CoV-2, que levou à pandemia de COVID-19. Tudo isso evidencia o protagonismo do programa em pesquisas de fronteira.

A internacionalização é outro pilar do PPG, com cerca de 60% dos artigos publicados com a participação de colaboradores internacionais, oportunidades regulares de doutorado sanduíche e participação em projetos de pesquisa transformadores. O programa também investe fortemente em divulgação científica e formação ampliada, por meio de iniciativas como o Online Bioinfo [3], o Curso de Férias em Bioinformática [4,5] e a Revista BIOINFO. Apesar dos resultados robustos, permanecem desafios estruturais importantes, como a necessidade de renovação e ampliação do quadro docente e o enfrentamento das demandas crescentes de escalabilidade computacional, fundamentais para sustentar e expandir a excelência acadêmica em um cenário de dados biológicos em rápida expansão.

Nas próximas seções, iremos apresentar alguns dos principais destaques do evento Bioinformática 2035, que contou com diversas palestras organizadas em painéis temáticos.

IA transformando a Saúde, o Agronegócio e a Sustentabilidade

Este painel abordou a interseção entre bioinformática, inteligência artificial (IA) e suas implicações para a saúde, o agronegócio e a sustentabilidade. O Prof. Virgílio Almeida (DCC/UFMG) enfatizou a relevância de uma conduta ética na aplicação de algoritmos, alertando para o potencial aumento das disparidades sociais. Ele ressaltou a importância de preparar a sociedade para as transformações tecnológicas rápidas. Além disso, ele enfatizou a responsabilidade e a transparência no uso de dados, bem como

a necessidade de fomentar a formação de cidadãos com senso crítico em relação ao impacto da tecnologia.

A professora Ana Maria Benko Isepon (UFPE) ressaltou a importância da proteômica e da metabolômica para a identificação de alérgenos, ao mesmo tempo em que alertou sobre os riscos associados à agricultura orgânica. A professora sugeriu o uso de IA para monitorar a produção agrícola e, assim, reduzir a pegada de carbono decorrente do transporte de alimentos. Outros tópicos abordados incluíram a memória adaptativa das plantas e a preocupação com a erosão genética nas culturas. Ela salientou a importância de valorizar espécies nativas e adaptadas ao contexto brasileiro, destacando o papel fundamental da agricultura familiar nesse contexto.

Dr. Guilherme Corrêa Oliveira, diretor científico do Instituto Tecnológico Vale (ITV), enfatizou a importância da inovação tecnológica para assegurar a segurança alimentar e ecológica. Ele destacou, ainda, a urgência de coletar dados sobre biodiversidade e a necessidade de transparência corporativa quanto aos seus impactos ambientais. O palestrante ressaltou, por fim, que uma colaboração eficaz entre a academia, o setor empresarial e a sociedade civil é crucial.

A professora Santuza Maria Ribeiro Teixeira (ICB/UFMG) compartilhou sua experiência no desenvolvimento de vacinas e terapias. A pesquisadora destacou a importância das parcerias entre saúde e instituições, detalhando a criação de vacinas terapêuticas e a relevância dos dados clínicos para o avanço das abordagens terapêuticas.

Por fim, o professor Eduardo Martín Tarazona Santos (ICB/UFMG) encerrou o painel, com foco nos desafios atuais da bioinformática. Ele destacou a urgência de gerenciar o grande volume de dados genéticos e ressaltou o papel crucial dos cientistas da computação no desenvolvimento de novas estruturas de dados. Além disso, o professor destacou as problemáticas da IA, como os vieses existentes e a complexidade inerente aos desafios que a tecnologia se propõe a resolver.

Iniciativas de Capacitação em Bioinformática

Na parte final do encontro, discutiu-se a formação da próxima geração de bioinformatas no Brasil, com a participação de coordenadores de importantes programas de pós-graduação. Raquel Cardoso de Melo Minardi (DCC/UFMG), Eduardo Moraes Rego Reis (IQ/USP) e Rodrigo Juliani Siqueira Dalmolin (Instituto MetrÓpole Digital/UFRN) contribuíram para o debate, que também contou com a participação de Ana Tereza Ribeiro de Vasconcelos (LNCC/MCTI), Wagner Meira Jr. (DCC/UFMG) e Valdir de

Queiroz Balbino (CCB/UFPE), coordenador adjunto da área de Ciências Biológicas na CAPES.

A discussão destacou a proposta de um curso itinerante de capacitação em bioinformática, voltado à democratização do ensino em todas as regiões do país, com valorização de docentes locais e engajamento de jovens pesquisadores na organização das atividades. Também foi ressaltada a importância de integrar ética, tecnologia e sustentabilidade aos currículos, discutir trajetórias profissionais e enfrentar os desafios dos programas de pós-graduação por meio de soluções inovadoras.

A aproximação com o setor privado foi apontada como estratégica, em função da alta empregabilidade dos egressos, reforçando a necessidade de atualização curricular, incluindo conteúdos de inteligência artificial, e de colaboração interinstitucional. Houve consenso de que a formação deve ir além do aspecto técnico, priorizando o desenvolvimento do senso crítico diante de rápidas transformações tecnológicas.

Foram debatidos ainda os desafios da educação em biodiversidade, a necessidade de abordar questões éticas, sociais e ambientais na formação e a carência de recursos tecnológicos adequados. A Profa. Raquel Minardi apresentou o Programa de Pós-Graduação em Bioinformática da UFMG, destacando sua estrutura interunidades, seus pontos fortes e desafios, bem como a necessidade de formar profissionais com sólida base em ciências exatas e biológicas.

O debate reforçou a preocupação com a obsolescência do conhecimento, a necessidade de práticas interdisciplinares, de parcerias com empresas e de equilíbrio entre fundamentos teóricos e aplicação prática. Destacou-se também a importância de criar um ambiente que estimule a curiosidade, a inovação e a capacidade crítica dos estudantes.

Por fim, o professor Valdir Balbino apresentou reflexões sobre o futuro da avaliação dos programas de pós-graduação na CAPES. Foram discutidos os desafios enfrentados pelos avaliadores, a concentração regional dos programas, a importância da qualidade dos dados e a necessidade de maior compromisso com a formação discente e a gestão dos cursos existentes. Foram mencionadas iniciativas de expansão para as regiões do Cerrado e da Amazônia, a relevância da colaboração interestadual e a necessidade de relatórios mais alinhados à realidade dos programas. Balbino também apresentou a evolução das ferramentas Lampião e ESPIA, utilizadas na avaliação da produção acadêmica, e destacou a importância de um sistema organizacional que integre comunicação, investimento estratégico na formação social e articulação institucional para fortalecer os programas.

Apresentações dos estudantes

O evento contou com uma sessão de pôsteres destinada a todos os trabalhos inscritos, além de sessões de apresentações orais, destinadas a mestrandos e doutorandos vinculados ao PPG-BIOINFO. Os trabalhos foram avaliados por professores e pós-doutorandos convidados. Além disso, uma categoria de voto popular possibilitou a participação direta do público na escolha dos destaques na apresentação de pôsteres.

Na sessão de pôsteres, o trabalho “A bioinformática como aliada no enfrentamento da resistência antimicrobiana”, de Paloma da Silva Teixeira e colaboradores, foi premiado como o melhor pôster do evento, com base na avaliação dos pareceristas. Ainda nessa modalidade, os trabalhos “Índices de correlação entre a expressão de Genes Condutores e a Desdiferenciação Celular em Câncer”, de Leonardo Henrique da Silva e colaboradores, e “Vigilância Genômica e Predição: Impactos e Desafios”, de Ronison Alves Guimarães e colaboradores, foram contemplados com o prêmio de voto popular.

Na categoria de apresentação oral de mestrado, o trabalho “Por que ensinar Bioinformática no Ensino Médio?”, de Helena Lott Costa e colaboradores, recebeu o prêmio de melhor apresentação. Por fim, na categoria de apresentação oral de doutorado, destacou-se o trabalho “Pseudogenes podem ser peças-chave na diversidade de anticorpos de cavalos”, de Juliana Edelvacy Lima Pinto e colaboradores, vencedor da respectiva modalidade.

Conclusão

O evento “Bioinformática 2035: Desafios e Formação de Talentos para a Próxima Década” constituiu uma importante oportunidade para estudantes do PPG-BIOINFO e de outras instituições apresentarem seus trabalhos, trocarem experiências e discutirem criticamente suas pesquisas, contribuindo para o aprimoramento científico e a maturação de suas propostas.

Esta edição da Revista BIOINFO reúne e sintetiza os principais trabalhos apresentados durante o evento, oferecendo uma visão abrangente das pesquisas desenvolvidas no âmbito da UFMG e de instituições parceiras. O conteúdo evidencia a diversidade temática e a qualidade científica da produção na área de Bioinformática.

Em janeiro de 2026, o Programa Interunidades de Pós-Graduação em Bioinformática da UFMG recebeu nota 7 na avaliação da CAPES, a pontuação máxima possível. Esse resultado atesta a excelência acadêmica do programa e reafirma seu compromisso

contínuo com a formação de recursos humanos altamente qualificados e com o avanço da Bioinformática no Brasil.

Nota de transparência em relação ao uso de IA. Este artigo foi revisado usando as seguintes ferramentas de IA: Grammarly, LanguageTool e ChatGPT. A ferramenta Read AI foi utilizada para realizar anotações das palestras e sumarizar as principais conclusões. O texto foi lido e revisado por todos os autores.

Referências

[1] Moura Junior, Adenilson Arcanjo de (org.) et al. BIOINFO #05 - Revista Brasileira de Bioinformática e Biologia Computacional. 5. Ed. Vol. 5. ISBN: 978-65-84391-01-7. Lagoa Santa: Alfahelix, 2025. doi: 10.51780/978-65-84391-01-7

[2] Mariano, DCB. Editorial BIOINFO #05: perspectivas sobre o uso de inteligência artificial na escrita acadêmica. BIOINFO, (5), 0 (2025). doi: 10.51780/bioinfo-05-00

[3] de Melo-Minardi, Raquel C., Eduardo C. de Melo, and Luana L. Bastos. "Onlinebioinfo: Leveraging the teaching of programming skills to life science students through learning analytics." Frontiers in Education. Vol. 7. Frontiers Media SA, 2022.

[4] Mariano, D., et al. "Uma estratégia para engajamento de participantes de eventos online." BIOINFO-Rev. Bras Bioinformática e Biol. Comput 10 (2021): 978-6.

[5] DA SILVA, Alessandra Lima et al. From in-person to the online world: insights into organizing events in bioinformatics. Frontiers in Bioinformatics, v. 1, p. 711463, 2021.

Cite este artigo 0.2

Moraes, L; et al. **Editorial BIOINFO #06 – Bioinformática 2035: Desafios e Formação de Talentos para a Próxima Década.** BIOINFO. Vol. 6. p.0 (2026). doi: 10.51780/bioinfo-06-00




Saiba mais 0.1

O editorial está disponível em <https://bioinfo.com.br/editorial-bioinfo-06/>

1

Bioinformática e o combate à desinformação em saúde: como as ciências ômicas podem reforçar a confiança pública na pesquisa biomédica

Autores 1.1

Ruan Vítor Santos Silva , Gabriel Quintanilha-Peixoto , Aristóteles Góes-Neto 

Revisão: Ariany Rosa Gonçalves , Diego Mariano , Savio Costa 




Cite este artigo 1.1

Silva, RV; Quintanilha-Peixoto, G; Góes-Neto, A. **Bioinformática e o Combate à Desinformação em Saúde: Como as Ciências Ômicas Podem Reforçar a Confiança Pública na Pesquisa Biomédica.** BIOINFO. Vol. 6. p.01 (2026). doi: 10.51780/bioinfo-06-01

Saiba mais 1.1

Este artigo está disponível em <https://bioinfo.com.br/bioinformatica-e-o-combate-a-desinformacao-em-saude-como-as-ciencias-omicas-podem-reforcar-a-confianca-publica-na-pesquisa-biomedica/>

Bioinformática e o combate à desinformação em saúde: como as ciências ômicas podem reforçar a confiança pública na pesquisa biomédica

Ruan Vítor Santos Silva^{1*} , Gabriel Quintanilha-Peixoto² , Aristóteles Góes-Neto¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Centro de Biociências e Biotecnologia da UENF, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Rio de Janeiro, Campos dos Goytacazes, Brasil.

* ruanvitor@ufmg.br

Resumo. A desinformação em saúde tornou-se um dos maiores desafios da sociedade contemporânea, impulsionada pela disseminação rápida de conteúdos sem base científica nas redes sociais. Pesquisas sobre genética, microbiomas e medicina personalizada são frequentemente distorcidas, gerando falsas expectativas sobre diagnósticos, terapias e resultados. Esta revisão discute o papel da bioinformática e das ciências ômicas como ferramentas estratégicas para o enfrentamento da desinformação biomédica, promovendo transparência, reprodutibilidade e acesso público a evidências confiáveis. Ao integrar grandes volumes de dados genômicos, proteômicos e metagenômicos, a bioinformática possibilita análises robustas que fortalecem a comunicação científica e a confiança social na pesquisa. Além disso, estratégias de ciência aberta, bancos de dados interoperáveis e práticas FAIR (Findable, Accessible, Interoperable, Reusable) contribuem para a verificação independente dos achados e para o engajamento informado da população. Argumenta-se que, na era digital, a alfabetização em dados biológicos e a colaboração entre cientistas, comunicadores e plataformas tecnológicas são fundamentais para mitigar os impactos da desinformação em saúde e consolidar uma ciência mais ética, participativa e transparente.

Palavras-chave: Ciência aberta; Ética em dados biomédicos; Literacia científica; Transparência científica.

1. Introdução

Entende-se ciências biomédicas pelo conjunto de áreas que estudam os processos biológicos relacionados à saúde, às doenças e ao funcionamento do corpo humano. Nesse contexto, a bioinformática pode ser entendida como uma área estratégica pelo fornecimento de ferramentas computacionais necessárias para analisar, integrar e interpretar os dados de genômica, proteômica, transcriptômica, metagenômica, entre outros [1, 2, 3]. O avanço das ciências ômicas permitiu a evolução das ciências biomédicas, proporcionando análises integradas de dados ômicos de amostras humanas, por exemplo [3]. Tendo em vista disso, doenças complexas, como câncer, ganharam melhorias para prognósticos e tratamentos, já que não basta conhecer as variantes genéticas; também é necessário entender quais genes são expressos e

como essas proteínas codificadas se comportam, além de como interagem com o microbioma humano, possibilitando modulação da imunidade ou respostas à tratamentos [4].

Paralelamente, a divulgação científica enfrenta limitações ao não conseguir estabelecer, de forma clara, uma comunicação plena e acessível entre a ciência e a sociedade [5]. A literatura sugere que a comunicação institucional nem sempre alcança o público de maneira efetiva [5], que a mediação jornalística pode gerar ruídos e dificultar a compreensão [6] e que a ausência de estratégias amplas de educação científica pode favorecer a entrada de desinformação em ambientes digitais [7]. Assim, pode-se entender que essas limitações podem ter ampliado o distanciamento entre especialistas e público, criando condições para a disseminação de informações imprecisas ou completamente falsas. À medida que pesquisadores divulgam avanços no desenvolvimento de novos tratamentos e terapias, muitas vezes ainda experimentais, a desinformação tende a distorcer esses achados, exagerando resultados preliminares. Esse tipo de conteúdo é frequentemente favorecido por algoritmos que priorizam postagens com forte apelo emocional ou sensacionalista, ampliando sua visibilidade [8].

Elementos como as redes sociais, *big data* derivados de seu uso e seus algoritmos baseados em inteligência artificial (IA) desempenham agora papéis duplos, permitindo tanto a disseminação quanto a detecção de informações enganosas sobre saúde, passando a existir interpretações simplificadas demais e/ou distorcidas. Entre os exemplos pertinentes às ômicas, pode-se citar promessas de “dietas baseadas no DNA” ou alegações sobre “genes da felicidade” [9]. Essas narrativas, muitas vezes descontextualizadas, enfraquecem a confiança pública na pesquisa e distorcem o papel da ciência biomédica, promovendo um status de desinformação que hoje já se nomeia como “infodemia” [10]. A bioinformática surge como área estratégica para enfrentar esse cenário, não apenas por sua capacidade de analisar grandes volumes de dados, mas por possibilitar a verificação, reprodutibilidade e comunicação transparente de evidências científicas [9]. Este trabalho discute como a integração entre bioinformática, ciência aberta e comunicação científica pode contribuir para o combate à desinformação em saúde, destacando os impactos sociais e éticos dessa interação.

2. Desenvolvimento

A desinformação biomédica é um fenômeno multifatorial, resultante da combinação entre complexidade científica, falta de literacia em dados e disseminação irresponsável de informações [7, 10]. Um estudo identificou padrões de hesitação vacinal nos EUA que correlacionam diretamente a exposição à desinformação online e redução nas taxas de vacinação infantil [11]. Ainda, surtos de sarampo foram associados à circulação de informações falsas sobre segurança de vacinas, principalmente em redes sociais. Entre as principais consequências estão o aumento de hospitalizações, maior custo para o sistema de saúde e risco aumentado para imunocomprometidos [11].

No contexto da pandemia de COVID-19, as crenças em tratamentos ineficazes, como ivermectina e hidroxiquina, resultaram em atraso no tratamento adequado e aumento da mortalidade [12]. Substancialmente, as plataformas digitais foram como uma força motriz, a possibilidade de disseminação de informações errôneas [12]. Outrossim, a exposição contínua

a teorias conspiratórias sobre saúde aumenta a ansiedade, estresse, sensação de ameaça e probabilidade de rejeitar tratamentos médicos embasados [11, 12, 13]. Continuamente, pesquisas sobre desinformação em HIV/AIDS demonstram que crenças equivocadas diminuem a adesão ao tratamento antirretroviral [14].

A bioinformática oferece ferramentas essenciais para mitigar esses problemas. A adoção de padrões FAIR (*Findable, Accessible, Interoperable, Reusable*) permite que dados e metadados estejam acessíveis e verificáveis, possibilitando que diferentes grupos de pesquisa reproduzam análises e confirmem resultados [9]. Da mesma forma, o uso de pipelines reprodutíveis em plataformas como Nextflow e Snakemake promove a rastreabilidade das etapas analíticas, evitando erros metodológicos e interpretações enviesadas, mostrando assim um passo a passo lógico, que promove o aumento da confiabilidade [8, 15]. Nesse sentido, a bioinformática e a ciência aberta trabalham em um ciclo de confiança que, ao promover o rigor técnico e a transparência (FAIR), permitir a detecção e o combate à desinformação em saúde, fortalecendo a credibilidade científica e a comunicação pública, onde pode-se visualizá-lo em um esquema representativo na figura 1.

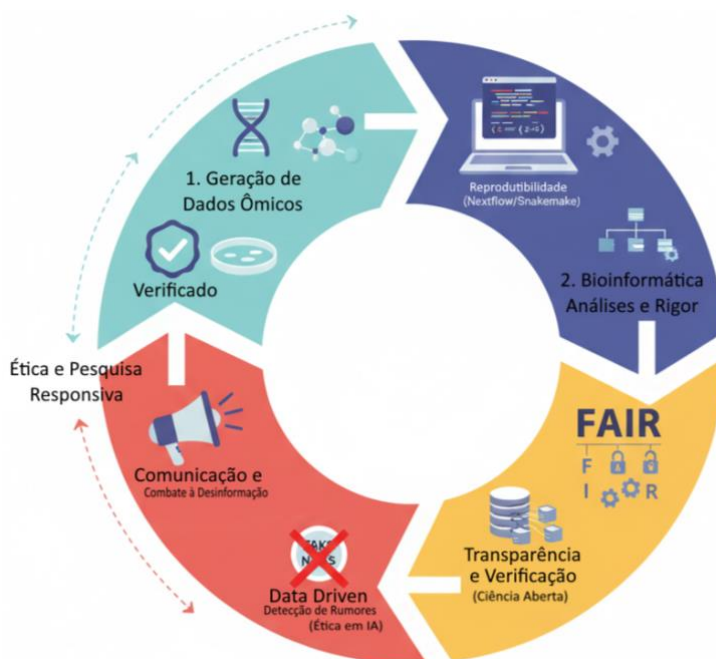


Figura 1. O Ciclo da Confiança: Bioinformática, Ciência Aberta e o Combate à Desinformação em Saúde.

Além do aspecto técnico, há uma dimensão social fundamental: a democratização do acesso à informação científica. Repositórios abertos, como o Gene Expression Omnibus e o European Nucleotide Archive, permitem que jornalistas, educadores e influenciadores científicos baseiem suas comunicações em dados verificados [17]. Dessa forma, a aproximação entre a ciência e o público leigo pode diminuir, fortalecendo a confiança nas instituições científicas aos olhos da sociedade.

Adicionalmente, o uso de inteligências artificiais permite detectar padrões de desinformação em redes sociais, avaliando a propagação de conteúdos falsos sobre terapias genéticas, biotecnologias emergentes e outros conceitos pertinentes à bioinformática [13]. Tais abordagens comumente combinam análise textual e mineração de dados para orientar estratégias de comunicação pública mais eficazes e baseadas em evidências.

Enquanto boatos sobre saúde podem alterar o comportamento populacional, a bioinformática permite modelar impacto de fake news em curvas epidêmicas, cenários contrafactuais e regiões vulneráveis à queda de cobertura vacinal [16]. Dessa maneira, pode permitir que órgãos de saúde (como Institutos Nacionais de Doenças, Centros de Controle e Prevenção de Doenças e Ministérios da Saúde e Informação) intervenham antes que maiores danos ocorram [16, 18].

O uso ético dessas tecnologias requer atenção. A abertura de dados genômicos deve respeitar princípios de privacidade e consentimento, garantindo que a transparência científica não comprometa direitos individuais [19]. Assim, o desafio não é apenas técnico, mas também moral e político: como comunicar a complexidade biológica sem reduzir sua precisão científica.

3. Conclusão

O combate à desinformação em saúde exige mais do que políticas de checagem de fatos; requer uma transformação cultural baseada em ciência aberta, ética de dados e comunicação responsável. Assim como existem sistemas de vigilância epidemiológica, tornar o processo de averiguação da verdade nas redes sociais, melhorias de divulgação científica e contribuição para a propagação das informações de desenvolvimento tecnológico dentro das universidades é uma forma de combater veementemente a desinformação e também tornar o propósito científico acadêmico transparente e acessível. Comparar estudos em larga escala (meta-análises computacionais), verificar a reprodutibilidade de dados e fluxos de trabalho e detectar discrepâncias ou interpretações incorretas pode permitir que algoritmos identifiquem inconsistências em dados genômicos ou clínicos frequentemente usados em boatos [16].

Com o monitoramento da evolução de vírus como o Sars-CoV-2 e MPox, assim como suas variantes e surtos, muitos boatos surgem de incertezas sobre a origem dessas variantes, mutações perigosas, transmissibilidade, e desinformação sobre as vacinas contra cepas novas, dentre outros, por meio da desinformação [8]. Tendo em vista disso, a genômica pode promover o rastreamento de mutações em tempo real [20], calcular impacto funcional, como escape imune e prever comportamento epidemiológico [21]. Dessa forma, pode ser possível gerar respostas públicas mais rápidas, amparadas em dados. Ainda, promover a instalação de ética em IA, é uma das formas possíveis de se combater a infodemia, por meio da desmistificação das avalanches produzidas pela desinformação, tornando a checagem de fatos e detecção dos super espalhadores das inverdades, supramencionadas na presente pesquisa [22].

Ao integrar práticas de reprodutibilidade, dados abertos e literacia científica, a bioinformática contribui para restaurar a credibilidade da ciência e empoderar a sociedade diante da avalanche de informações imprecisas. O futuro da pesquisa biomédica se ampara na

capacidade de unir rigor técnico, clareza comunicativa e compromisso ético, elementos essenciais para que a ciência permaneça um bem público, acessível e confiável.

4. Referências

[1] ALBERTS, Bruce et al. *Molecular Biology of the Cell*. 6. ed. New York: Garland Science, 2014. Disponível em: <https://openlibrary.org/books/OL21074168M>. Acesso em: 03 out. 2025.

[2] MOUNT, David W. *Bioinformatics: Sequence and Genome Analysis*. 2. ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2004. Disponível em: <https://academic.oup.com/clinchem/article/51/11/2219/5629817>. Acesso em: 04 out. 2025.

[3] CARPENTER, Anne E. et al. *A field guide to cultivating computational biology*. arXiv preprint, 2021. Disponível em: <https://arxiv.org/pdf/2104.11364>. Acesso em: 05 out. 2025.

[4] MANZONI, C. et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, v. 19, n. 2, p. 286–302, 2018. doi: 10.1093/bib/bbw114.

[5] HUNTER, P. The communications gap between scientists and public: More scientists and their institutions feel a need to communicate the results and nature of research with the public. *EMBO Reports*, v. 17, n. 11, p. 1513–1515, 2016. doi: 10.15252/embr.201643379.

[6] BELDA, Francisco. Credibilidade jornalística em notícias de ciência: desafios contemporâneos na mediação científica. Assembleia Legislativa do Estado de São Paulo, 2019. Disponível em: <https://www3.al.sp.gov.br/repositorio/ilp/anexos/4985>. Acesso em: 06 out. 2025.

[7] FERREIRA, Ana Paula; OLIVEIRA, Ana Luísa. Desinformação e crise de confiança na ciência: desafios para a comunicação científica. *Ciência da Informação Express*, v. 2, n. 1, p. 1–15, 2021. doi: 10.60144/v2i.2021.28.

[8] PAPANIKOU, V. et al. Health Misinformation in Social Networks: A Survey of Information Technology Approaches. *Future Internet*, v. 17, p. 129, 2025. doi: 10.3390/fi17030129.

[9] WILKINSON, M. D. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, v. 3, p. 160018, 2016. doi: 10.1038/sdata.2016.18.

[10] TASCHNER, N. P.; ORSI, C.; ALMEIDA, P.; PILATI, R. The impact of personal pseudoscientific beliefs in the pursuit for non-evidence-based health care. *Journal of Evidence-Based Healthcare*, v. 3, p. e3516-e3516, 2021.

[11] TUCKERMAN, J.; KAUFMAN, J.; DANCHIN, M. Effective Approaches to Combat Vaccine Hesitancy. *Pediatric Infectious Disease Journal*, v. 41, n. 5, p. e243–e245, 2022. doi: 10.1097/INF.0000000000003499.

[12] FERREIRA CACERES, M. M. et al. The impact of misinformation on the COVID-19 pandemic. *AIMS Public Health*, v. 9, n. 2, p. 262–277, 2022. doi: 10.3934/publichealth.2022018.

[13] VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. *Science*, v. 359, n. 6380, p. 1146–1151, 2018. doi: 10.1126/science.aap9559.




[14] SMALLWOOD, S.; PARKS, F. Quanto mais as coisas mudam, mais elas permanecem as mesmas: mitos e desinformação sobre HIV/AIDS nos Estados Unidos rurais. *Health Promotion Practice*, v. 25, p. 707–716, 2023. doi: 10.1177/15248399231180592.

- [15] DI TOMMASO, P. et al. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, v. 35, n. 4, p. 316–319, 2017. doi: 10.1038/nbt.3820.
- [16] FENG, X. et al. Health Misinformation Detection: Approaches, Challenges and Opportunities. *Inquiry*, v. 62, p. 469580251384784, 2025. doi: 10.1177/00469580251384784.
- [17] EDGAR, R.; DOMRACHEV, M.; LASH, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, v. 30, n. 1, p. 207–210, 2002. doi: 10.1093/nar/30.1.207.
- [18] PADALKO, H. et al. A Novel Comprehensive Framework for Detecting and Understanding Health-Related Misinformation. *Information*, v. 16, n. 3, p. 175, 2025. doi: 10.3390/info1603017.
- [19] BRASIL. Lei n.º 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei n.º 12.965, de 23 de abril de 2014 (Marco Civil da Internet). *Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018*. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 out. 2025.
- [20] RUECA, M. et al. Genomic and Epidemiologic Surveillance of SARS-CoV-2 in the Pandemic Period: Sequencing Network of the Lazio Region, Italy. *Viruses*, v. 15, n. 11, p. 2192, 2023. doi: 10.3390/v15112192.
- [21] BLEHER, Michael et al. Topological data analysis identifies emerging adaptive mutations in SARS-CoV-2. arXiv:2106.07292 [q-bio.PE], 25 ago. 2023. Disponível em: <https://doi.org/10.48550/arXiv.2106.07292>. Acesso em: 18 out. 2025.
- [22] PARK, M. K.; ASHWOOD, N.; CAPES, N. Ethics of Artificial Intelligence in Medicine. *Cureus*, 2025. DOI: 10.7759/cureus.83567. Disponível em: <https://doi.org/10.7759/cureus.83567>. Acesso em: 22 out. 2025.

2

Índices de correlação entre a expressão de genes condutores e a desdiferenciação celular em câncer

Autores 2.1

Leonardo Henrique da Silva , Milenna Machado Pirovani , Raquel Cardoso de Melo-Minardi 

Revisão: Alessandra Lima , Bibiana Fam , Marcos Sousa , Savio Costa , Tatiane Senna Bialves 




Cite este artigo 2.1

Silva, L. H. da; Pirovani, M. M.; Melo-Minardi, R. C. de. **Índices de correlação entre a expressão de genes condutores e a desdiferenciação celular em câncer**. BIOINFO. Vol. 6. p.02 (2026). doi: 10.51780/bioinfo-06-02

Saiba mais 2.1

Este artigo está disponível em <https://bioinfo.com.br/indices-de-correlacao-entre-a-expressao-de-genes-condutores-e-a-desdiferenciacao-celular-em-cancer/>

Índices de correlação entre a expressão de genes condutores e a desdiferenciação celular em câncer

Leonardo Henrique da Silva^{1*} , Milenna Machado Pirovani¹ , Raquel Cardoso de Melo-Minardi¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* Contato pelo e-mail: lhsilva@ufmg.br

Resumo. *O câncer explora o estado de desdiferenciação celular para promover a progressão tumoral e resistência a tratamentos. Genes com potencial de influenciar a formação do câncer são chamados condutores. Não há uma definição clara de como tais genes, oncogenes e supressores de tumores, agem interferindo no processo de desdiferenciação tumoral. Neste trabalho, analisaremos o impacto de 741 genes condutores definidos pelo Cancer Gene Census (COSMIC) na desdiferenciação celular em 33 tipos de tumores do TCGA. Investigaremos aqui, de modo preliminar, dados de correlação entre a desdiferenciação celular tumoral e dados de expressão, metilação, mutação e variação do número de cópias de genes condutores de câncer. Definindo um limiar exploratório encontramos que pelo menos 177 genes condutores possuem correlações diretas ou inversas entre expressão gênica e desdiferenciação celular em ao menos um tipo de tumor, além de um total de 290 correlações distintas.*

Palavras-chave: *Câncer; Genes condutores; Desdiferenciação celular.*

1. Introdução

Um dos mecanismos mais estudados na origem do câncer é a presença de conjuntos de mutações capazes de alterar o equilíbrio entre proliferação e morte celular. A proliferação celular é bem caracterizada em células-tronco, células embrionárias e células desdiferenciadas. No câncer, esse estado de desdiferenciação é explorado de forma a contribuir para evitar a vigilância imunológica, resistir a tratamentos e avançar a progressão tumoral [1]. Genes com o potencial de influenciar o processo de formação do câncer são conhecidos como condutores, geralmente se encaixando em subclasses de oncogenes ou de supressores de tumores.

Nos últimos anos, o estudo desses genes têm sido impulsionados pelos grandes volumes de dados ômicos gerados. Nesse contexto, duas das maiores fontes de dados sobre o câncer atualmente são o *Catalogue Of Somatic Mutations In Cancer* (COSMIC) [3] e o *The Cancer Genome Atlas* (TCGA). A base de dados do COSMIC, ou especificamente sua aba de *Cancer Gene Census* (CGC), contém uma lista de genes curados como condutores de câncer mantida por meio de curadoria manual, baseada na literatura científica – que exige pelo menos dois estudos independentes para a inclusão de um gene. Por outro lado, o TCGA é um projeto que disponibiliza dados clínicos de estudos de câncer, incluindo informações genéticas e fenotípicas, a partir das quais se pode fazer inferências por diferentes estratégias de análises de dados, como correlações estatísticas ou predição por *Machine Learning*. A integração dos dados

das duas plataformas é uma ótima maneira de filtrar alvos e obter percepções sobre a biologia molecular do câncer.

Em 2018, MALTA et al. [2] criaram uma metodologia para prever correlações entre expressão gênica e desdiferenciação celular no contexto de diferentes cânceres usando *Machine Learning*. Essa mesma abordagem pode ser utilizada para diferentes conjuntos de dados, como de metilação, variação do número de cópias e número de mutações. Além de um escopo menor de alvos para filtrar percepções relevantes a diferentes perguntas.

2. Dados e Metodologia

A partir dos dados do CGC disponíveis no COSMIC, definimos uma lista de 741 genes condutores. Utilizando a biblioteca do *UCSCXenaShiny* [4] em R, e a função baseada no método de regressão logística (introduzido por [2]), investigaremos os dados de expressão, metilação, mutação e alteração do número de cópias presentes no TCGA - buscando criar índices de correlação (spearman) variando entre 1 e -1, entre variáveis de genes condutores e a observação de desdiferenciação celular em cada um dos 33 tipos de tumores do TCGA. De maneira exploratória, definimos um limiar de correlação de $\rho > 0.7$ ou $\rho < -0.7$ para reduzir os resultados somente aos genes mais fortemente associados.

3. Resultados Parciais

Encontramos que 177 genes, dos 741 genes condutores observados, possuem índices de correlação entre expressão gênica e desdiferenciação celular com um $\rho > 0.7$ ou $\rho < -0.7$ em pelo menos um tipo de tumor. Alguns atingindo mais de um tipo de câncer e totalizando 290 correlações diferentes, das quais 254 eram correlações negativas e 36 positivas.

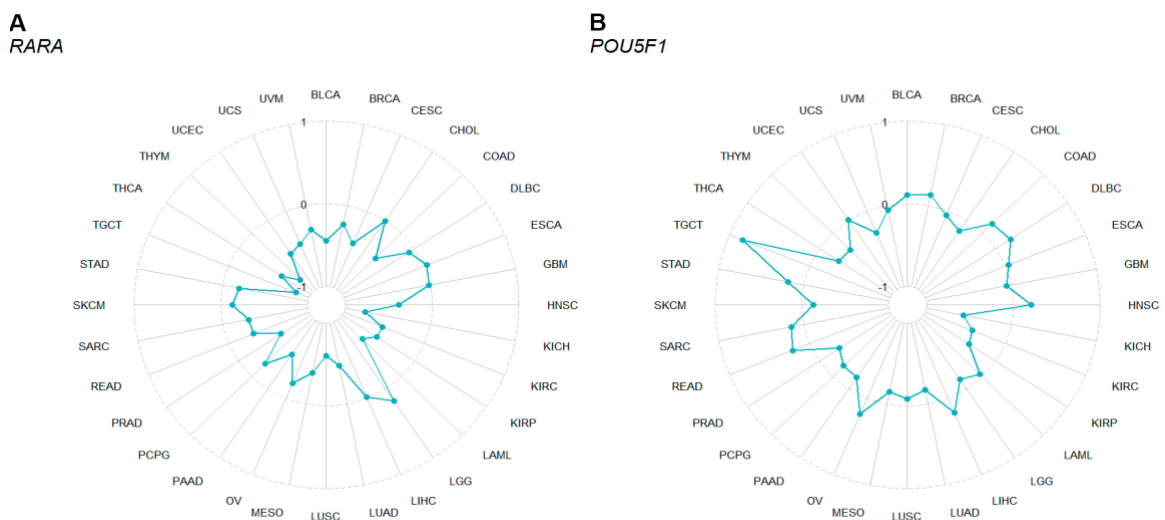


Figura 1. Plot de radar indicando a correlação entre a expressão gênica e a observação de desdiferenciação celular em 33 tipos de tumores do TCGA. Os dados de correlação estão definidos entre um índice 1 e -1, quanto mais próxima os pontos estiverem da margem externa do círculo, maior a correlação entre a expressão gênica e os níveis de desdiferenciação celular no câncer indicado, quanto mais próxima do centro, menor a correlação. Na figura A, perfil de correlação da expressão do gene *RARA* difundidamente negativo. Na figura B, perfil de correlação de *POU5F1*, com uma correlação altamente notável com o tumor de TGCT (*Tenosynovial Giant Cell Tumor*).

Futuramente, esses resultados podem ser expandidos para correlações envolvendo dados de metilação, mutação e variação do número de cópias. Após a obtenção dos resultados completos, esperamos identificar a existência de subconjuntos de genes condutores com características genéticas ou epigenéticas associadas aos tumores em que suas mutações condutoras foram previamente observadas. Bem como descobrir novos tumores em que essas observações ainda não tenham sido feitas, confirmando achados prévios e descobrindo novos.

4. Conclusão

Acreditamos que essas observações podem ajudar a progredir o conhecimento sobre as bases moleculares do câncer, identificar novos tumores associados com genes condutores conhecidos, e somar evidências na identificação de potenciais alvos terapêuticos de diferentes tipos tumorais.

Agradecimentos. Os autores agradecem a rede de pesquisa do TCGA (<https://www.cancer.gov/tcga>.) pelos dados sob os quais os resultados aqui apresentados se baseiam em parte. Além disso, agradecemos às agências de fomento à pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig).


4. Referências

- [1] CARNERO, A. *et al.* The cancer stem-cell signaling network and resistance to therapy. **Cancer treatment reviews**, Netherlands, v. 49, p. 25–36, 2016.
- [2] MALTA, Tathiane M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. **Cell**, [s. l.], v. 173, n. 2, p. 338-354.e15, 2018.
- [3] TATE, John G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. **Nucleic acids research**, England, v. 47, n. D1, p. D941–D947, 2019.
- [4] WANG, Shixiang *et al.* UCSCXenaShiny: an R/CRAN package for interactive analysis of UCSC Xena data. **Bioinformatics (Oxford, England)**, England, v. 38, n. 2, p. 527–529, 2022.

3

Influência de variáveis climáticas na dinâmica vetorial de arboviroses no Brasil

Autores 3.1

Fernanda Elen Silva dos Santos , Ronison Alves Guimarães , Wagner Fonseca , Joicymara Santos Xavier 

Revisão: Diego Mariano , Lucianna Helene Santos 





Cite este artigo 3.1

Santos, F. E. S. dos; Guimarães, R. A.; Fonseca, V.; Xavier, J. S. **Influência de variáveis climáticas na dinâmica vetorial de arboviroses no Brasil**. BIOINFO. Vol. 6. p.03 (2026). doi: 10.51780/bioinfo-06-03

Saiba mais 3.1

Este artigo está disponível em <https://bioinfo.com.br/influencia-de-variaveis-climaticas-na-dinamica-vetorial-de-arboviroses-no-brasil/>

Influência de variáveis climáticas na dinâmica vetorial de arboviroses no Brasil

Fernanda Elen Silva dos Santos¹ , Ronison Alves Guimarães¹ , Vagner Fonseca² ,
Joicymara Santos Xavier^{1,3} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Ciências Exatas e da Terra, Universidade do Estado da Bahia (UNEB), Salvador, Brasil

³ Divisão de Ciência da Computação, Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, Brasil

fernandaelenbio13@gmail.com

Resumo. As arboviroses, como dengue, zika, febre amarela e chikungunya, representam desafios significativos à saúde pública e são transmitidas principalmente por mosquitos do gênero *Aedes*, com destaque para *Aedes aegypti* e *Aedes albopictus*. A relação entre o clima e a proliferação desses vetores, influencia diretamente a ocorrência de surtos em diversas regiões. A análise de variáveis como temperatura e umidade revela como condições ambientais favorecem o ciclo de vida dos mosquitos e a disseminação viral. O *A. aegypti* predomina em áreas urbanas tropicais, enquanto o *A. albopictus* se adapta a climas mais variados, incluindo regiões temperadas. No Brasil, o Instituto Nacional de Meteorologia (INMET) fornece dados essenciais por meio de sua rede de estações meteorológicas para monitorar esses fatores, auxiliando na análise de padrões climáticos associados a surtos. Evidências recentes mostraram que mudanças climáticas amplificam o risco de doenças causadas por mosquitos, especialmente em contextos de alta densidade populacional. A biologia e o comportamento desses vetores propõe uma reflexão sobre a importância do monitoramento epidemiológico para antecipar surtos, destacando a relevância de estratégias preventivas baseadas em observações ambientais.

Palavras-chave: Arboviroses; INMET; Bioinformática.

1. Introdução

As arboviroses são doenças virais transmitidas por artrópodes, sendo os mosquitos e carrapatos seus principais vetores. Mosquitos do gênero *Aedes*, como *Aedes aegypti* e *Aedes albopictus*, são responsáveis pela disseminação dos vírus da dengue, zika, febre amarela e chikungunya. Esses mosquitos prosperam em condições ambientais específicas, e as mudanças climáticas têm ampliado suas áreas de atuação. Estudos indicam que fatores como alterações climáticas, crescimento populacional, mobilidade humana e urbanização agravaram a carga da dengue, tanto pelo risco de infecção em áreas endêmicas quanto pela expansão do alcance do vetor primário para novas áreas [1]. Entender a dinâmica de transmissão de doenças correlacionadas

com mudanças climáticas é essencial para a proteção da saúde pública, especialmente em um cenário de crescente vulnerabilidade ambiental.

2. Biologia e Distribuição dos vetores *Aedes aegypti* e *Aedes albopictus*

O *Aedes aegypti* foi descrito originalmente no Egito, e acompanhou o homem em sua longa migração pelo mundo. É considerado um mosquito cosmopolita, com ocorrência nas regiões tropicais e subtropicais. Sua introdução no Brasil ocorreu durante o período colonial, provavelmente na época do tráfico de escravos [2]. Os mosquitos dessa espécie são adaptados à vida urbana, mas também são encontrados em áreas rurais [2]. Em relação à sua biologia, os criadouros do *A. aegypti* incluem recipientes artificiais, como latas e pneus, embora também possam ser encontradas larvas em plantas como bromélias ou em cavidades de árvores. Sua distribuição concentra-se em regiões quentes, onde a temperatura ideal (25-30°C) acelera seu desenvolvimento [2]. Outro mosquito de importância sanitária é o *Aedes albopictus*, também vetor de arboviroses, como dengue e febre amarela [2]. Descrito originalmente na Índia, destaca-se pela versatilidade, pois resiste a climas mais frios e se estabelece em áreas periurbanas e rurais. Essa espécie chegou à Europa e às Américas por meio de comércio internacional, especialmente pelo transporte de pneus usados [2]. No Brasil, sua introdução ocorreu em maio de 1986, nos estados de Minas Gerais e Rio de Janeiro [2]. Supõe-se que tenha entrado no país através de portos do Espírito Santo e se interiorizado, via estrada de ferro, no vale do rio Doce (transporte de minério de ferro) [2,3].

2.1 Influência do Clima nos Surtos, INMET e a Bioinformática

O clima desempenha um papel crucial na proliferação de mosquitos do gênero *Aedes*. Temperaturas elevadas reduzem o tempo de incubação viral no vetor, enquanto chuvas intensas favorecem a formação de criadouros [1,2]. Além disso, a umidade relativa influencia a sobrevivência dos ovos, especialmente do *A. albopictus*, capaz de hibernar em condições adversas [3]. No Brasil, o Instituto Nacional de Meteorologia (INMET) desempenha papel fundamental nesse contexto, com medições contínuas de temperatura, umidade, precipitação e outras variáveis desde 1961 [4]. O banco de dados meteorológicos do INMET disponibiliza séries históricas diárias que possibilitam análises detalhadas dos padrões climáticos relacionados à dinâmica dos vetores e aos surtos de arboviroses. Variáveis como temperatura máxima e mínima, umidade relativa e precipitação são essenciais para compreender a influência do clima na dinâmica vetorial. Cruzando esses dados meteorológicos com registros entomológicos espaciais e temporais, é possível analisar a predominância e o deslocamento dos mosquitos vetores em função das condições climáticas, contribuindo assim para um melhor entendimento dos fatores ambientais que afetam a transmissão de doenças transmitidas por mosquitos.

A bioinformática oferece uma abordagem interessante para o monitoramento dessas doenças, possibilitando a correlação entre variáveis climáticas e dados de localização dos vetores e, posteriormente, o uso de algoritmos de inteligência artificial para prever possíveis surtos. Um exemplo concreto dessa abordagem é o estudo de Heath et al. (2025) [5] que

combinou modelos matemáticos e algoritmos de aprendizado de máquina para correlacionar dados climáticos, urbanos e populacionais, utilizando tanto séries temporais quanto dados geoespaciais, conseguindo prever áreas e períodos de maior risco para o *Aedes aegypti* e transmissão de arboviroses no Brasil.

A integração desses dados podem auxiliar na mitigação dos impactos das arboviroses, especialmente em áreas mais vulneráveis. Modelos de aprendizado de máquina podem ser uma alternativa eficiente para identificar padrões espaciais relevantes. Os dados climáticos e entomológicos podem ser implementados usando algoritmos de *Machine Learning* como *Random Forest* e *Gradient Boosting*, para analisar os padrões nos dados coletados. Esses modelos permitirão prever a disseminação de arbovírus com base nas variáveis climáticas e epidemiológicas, ajudando na identificação precoce de surtos. Além disso, ferramentas de monitoramento como dashboards interativos, oferecem suporte visual à vigilância sanitária e ajudam na disseminação de informações, facilitando ações coordenadas e mais precisas na contenção das arboviroses.

A imagem abaixo é uma representação das estações meteorológicas do INMET em todo território brasileiro:



Figura 1. Estações meteorológicas do INMET no Brasil. Fonte: Instituto Nacional de Meteorologia - 2025. Os círculos verdes indicam as estações automáticas, ou seja, representam estações meteorológicas automatizadas, que medem e registram variáveis como temperatura, umidade, precipitação e vento continuamente, sem intervenção humana direta. Os pontos azuis indicam estações meteorológicas tradicionais, normalmente operadas manualmente, responsáveis por observações diárias feitas por profissionais capacitados, com medições como chuva, temperatura e vento. Além de possuir estações meteorológicas em áreas terrestres, o INMET também mantém estações em pontos estratégicos, como ilhas oceânicas e regiões litorâneas. Como é possível observar os

pontos deslocados à direita representam as estações Fernando de Noronha - PE e Ilha de Trindade - RJ. Os pontos em lilás correspondem aos pluviômetros que são dispositivos automáticos dedicados exclusivamente à medição da quantidade de chuva, registrando os dados em tempo real para monitoramento contínuo.

3. Conclusão

A relação entre clima, os mosquitos *Aedes* e arboviroses evidencia a necessidade de atenção contínua às variáveis ambientais, com o suporte de instituições como o INMET para a obtenção de dados precisos de monitoramento. A compreensão de como o aumento da temperatura, a precipitação e as chuvas intensas influenciam a disseminação dessas doenças permite antecipar riscos e planejar ações preventivas [5]. Essa abordagem não apenas enriquece o conhecimento científico, mas também oferece suporte prático para a proteção das comunidades afetadas. Investir em monitoramento e educação, representa um caminho essencial para reduzir os impactos dessas doenças a curto e longo prazo. Entretanto existem limitações, principalmente pela variabilidade regional dos fatores socioambientais. Ademais, os dados do INMET podem cobrir apenas certas regiões do Brasil, e as dinâmicas dos vetores podem variar em locais com diferentes padrões de urbanização, saneamento e clima.

Agradecimentos. Os autores agradecem às agências de fomento: CNPq pelo apoio financeiro, ao Projeto NAVIO e a Rede Saúde Única pela colaboração e ao Programa de Pós-Graduação em Bioinformática da UFMG.

4. Referências

- [1] O'Neill SL, et al. New solutions against the dengue global threat: opportunities for *Wolbachia* interventions. *Trends Parasitol.* ISSN: 1471-4922. Vol. 41. p. 35013 (2025). doi: [10.1016/j.ijid.2025.107923](https://doi.org/10.1016/j.ijid.2025.107923)
- [2] Consoli, R. A. G. B., & Oliveira, R. L. (1994). Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro: Editora Fiocruz. 228 p. ISBN: 85-85676-03-5.)
- [3] Caminade C, et al. A climate and population dependent diffusion model forecasts the spread of *Aedes albopictus* mosquitoes in Europe. *Commun. Earth Environ.* ISSN: 2662-4435. Vol. 6. Art. 2199 (2025). doi: <https://doi.org/10.1038/s43247-025-02199-z>
- [4] Instituto Nacional de Meteorologia (INMET). Banco de Dados Meteorológicos para Fins Exclusivos de Pesquisa (BDMEP). Disponível em: <https://bdmep.inmet.gov.br/> (2025)
- [5] Heath K, Muniz Alves L, Bonsall MB. Climate change, urbanisation and transmission potential: *Aedes aegypti* mosquito projections forecast future arboviral disease hotspots in Brazil. *PLOS Neglected Tropical Diseases.* 2025 Sep 18;19(9):e0013415. doi: <https://doi.org/10.1371/journal.pntd.0013415>

4

Vigilância genômica e predição: impactos e desafios

Autores 4.1

Ronison Alves Guimarães , Fernanda Elen dos Santos , Vagner Fonseca ,
Joicymara Santos Xavier 

Revisão: Diego Mariano , Sheila Cruz , Tatiane Senna Bialves 





Cite este artigo 4.1

Guimarães, R. A.; Santos, F. E. dos; Fonseca, V.; Xavier, J. S. **Vigilância genômica e predição: impactos e desafios**. BIOINFO. Vol. 6. p.04 (2026). doi: 10.51780/bioinfo-06-04

Saiba mais 4.1

Este artigo está disponível em <https://bioinfo.com.br/vigilancia-genomica-e-predicao-impactos-e-desafios/>

Vigilância genômica e predição: impactos e desafios

Ronison Alves Guimarães^{1*} , Fernanda Elen dos Santos¹ , Vagner Fonseca² ,
Joicymara Santos Xavier³ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Ciências Exatas e da Terra, Universidade do Estado da Bahia (UNEB), Salvador, Brasil

³ Divisão de Ciência da Computação, Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, Brasil

* ronialvesart@gmail.com

Resumo. *A vigilância genômica busca compreender as mudanças genéticas que podem representar riscos à saúde pública e impactos econômicos. O avanço das tecnologias de sequenciamento de nova geração (NGS) ampliou significativamente as possibilidades da bioinformática aplicada à vigilância epidemiológica. Entretanto, a aplicação dessas conquistas em doenças endêmicas, como as arboviroses, ainda enfrenta desafios consideráveis. Observa-se uma lacuna substancial na disponibilidade de dados genômicos, caracterizada pelo volume reduzido de sequências depositadas nos bancos de dados, pela inconsistência temporal nos depósitos e pela disparidade acentuada entre países com diferentes capacidades econômicas. A heterogeneidade dos dados e a multiplicidade de repositórios tornam os processos de normalização e integração complexos. Iniciativas recentes baseadas em aprendizado de máquina buscam corrigir vieses de vigilância, enquanto esforços de integração de bases de dados visam criar repositórios validados e normalizados. Tais iniciativas propõem estratégias que não apenas buscam prever surtos, mas fazê-lo em tempo hábil para permitir respostas efetivas.*

Palavras-chave: Arboviroses; Bioinformática; Ciência de Dados; Saúde Pública.

1. Introdução

A vigilância genômica constitui uma abordagem estratégica para o monitoramento global de variantes genéticas de patógenos, permitindo a detecção precoce de mudanças evolutivas com potencial impacto na saúde pública [1].

Apesar dos avanços tecnológicos em sequenciamento e das lições aprendidas com emergências recentes, persiste um desafio crítico na aplicação da vigilância genômica: a predição confiável de surtos, especialmente para doenças endêmicas e reemergentes, como as arboviroses. Essas doenças sofrem com escassez informacional, heterogeneidade de dados e inconsistência temporal nos depósitos em bancos de dados públicos [2,3]. Essa assimetria na disponibilidade e qualidade dos dados compromete o desenvolvimento de modelos preditivos robustos, essenciais para antecipar surtos e orientar respostas de saúde pública. Compreender como superar essas limitações e viabilizar a predição efetiva em contextos de dados limitados constitui, portanto, um dos principais desafios contemporâneos da vigilância genômica aplicada

à saúde pública. Este artigo apresenta uma visão panorâmica sobre a vigilância genômica, seu impacto e seus desafios metodológicos neste campo, com foco em arboviroses.

2. Desenvolvimento

A pandemia de SARS-CoV-2 (COVID-19) demonstrou de forma incontestável a importância estratégica da vigilância genômica para o controle de emergências em saúde pública [4]. Durante este período, o elevado número de casos, aliado ao intenso esforço de pesquisa, resultou em uma produção massiva de dados genômicos, permitindo compreender, em alta resolução espaço-temporal, a dinâmica das mutações virais, identificar novas variantes e antecipar respostas de controle epidemiológico [5].

O compartilhamento extensivo de dados genômicos entre instituições e países foi determinante para mitigar os impactos da pandemia iniciada em 2019. O avanço das tecnologias de sequenciamento de nova geração (*Next Generation Sequencing*) expandiu consideravelmente as possibilidades de aplicação da bioinformática na vigilância epidemiológica. As tecnologias NGS, atualmente em sua terceira geração, transformaram processos que anteriormente demandavam semanas ou meses em procedimentos executáveis em poucas horas [6,7]. As arboviroses constituem um grupo de doenças transmitidas por artrópodes (mosquitos e carrapatos) com particular interesse para aplicação das metodologias tecnológicas obtidas. Estas apresentam riscos significativos à saúde pública e geram impactos econômicos consideráveis [8]. Sua proliferação é intensificada pelas mudanças climáticas globais [9].

Doenças negligenciadas caracterizam-se, entre outros aspectos, pela limitada atenção em pesquisa e vigilância [10]. Consequentemente, observa-se volume reduzido de dados genômicos disponíveis para essas doenças, o que representa um desafio significativo para a predição de surtos. A qualidade e o volume dos dados são fatores cruciais para o desenvolvimento de modelos preditivos robustos. No entanto, essas doenças caracterizam-se por volume reduzido de sequências depositadas em bancos de dados públicos e pela inconsistência temporal no fluxo de depósitos [3,11].

Nações com maior capacidade econômica frequentemente priorizam a vigilância genômica para mitigar a disseminação dessas doenças, especialmente diante das mudanças climáticas que tornam novos ambientes favoráveis à proliferação dos vetores. Consequentemente, esses países geram volumes desproporcionais de dados genômicos [3]. Essa concentração de dados em determinadas regiões contrasta com a distribuição geográfica real da carga de doença, criando vieses na vigilância global.

A heterogeneidade dos dados constitui outro problema relevante: múltiplas bases de dados com diferentes estruturas, mudanças frequentes nos repositórios e variação nos padrões de metadados tornam os processos de normalização e integração extremamente desafiadores [2,3].

Diversas iniciativas têm sido implementadas para superar essas limitações. Modelos baseados em aprendizado de máquina, como o *Random Forest*, que utiliza múltiplas árvores de

decisão trabalhando em conjunto para fazer previsões mais robustas e identificar padrões complexos nos dados, vêm sendo propostos para corrigir vieses de vigilância e compensar lacunas na cobertura geográfica e temporal dos dados [3]. Para o exemplo em questão, são citadas abordagens que se baseiam em dois modelos principais: o *Nested Surveillance Model* combinado com o *Ecological Niche Model*. O primeiro modelo reconhece que nem todas as regiões têm a mesma capacidade de detectar e reportar casos de doenças. Usando dados de casos de infecções virais, ele mapeia onde os sistemas de saúde são mais eficientes em identificar casos, considerando fatores como renda, acesso a hospitais e infraestrutura de saúde. Essa informação é usada para calibrar o segundo modelo, distinguindo locais com poucos casos reportados por falta de vigilância daqueles com baixo risco.

Já o segundo aproveita o fato de que dengue, chikungunya, Zika e febre amarela compartilham os mesmos mosquitos transmissores (*Aedes aegypti* e *Aedes albopictus*) e respondem de forma similar a fatores ambientais como temperatura e urbanização. Ao analisar milhares de registros dessas doenças em conjunto, o modelo identifica onde as condições são favoráveis para a sua transmissão. Esses dois modelos em conjunto demonstraram potencial para mitigar vieses de vigilância e heterogeneidade de amostragem, melhorando assim a identificação de áreas de alto risco.

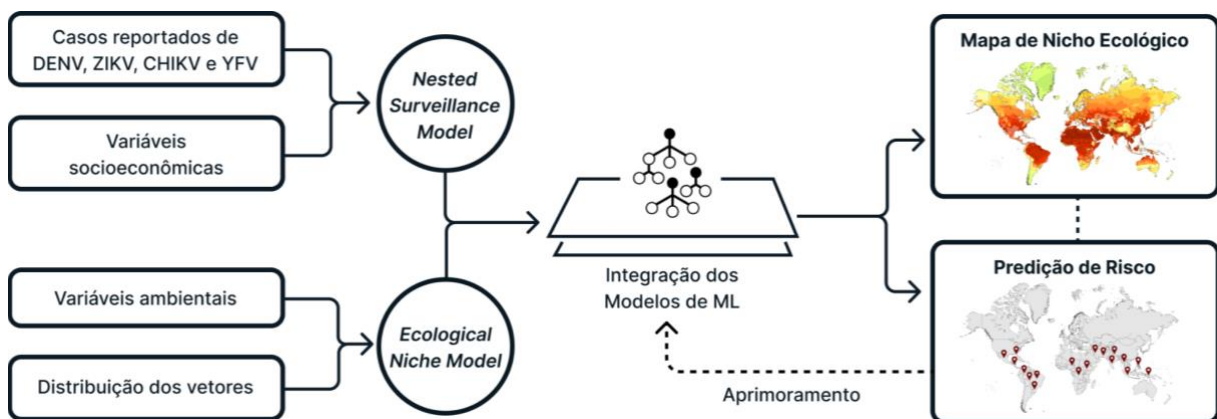


Figura 1: Esquema visual da abordagem de vigilância integrada baseada em Aprendizado de Máquina (Adaptado de Lim et al., 2025). A imagem ilustra como o processamento de dados via *ML* ajuda a diminuir vieses de vigilância epidemiológica. Fonte: próprio autor.

Paralelamente, outros esforços concentram-se na integração de múltiplas bases de dados, com objetivo de criar repositórios de referência validados, normalizados e mais robustos [2]. A correlação entre dados genômicos e metadados epidemiológicos, climáticos e socioeconômicos constitui uma abordagem promissora para compreender os determinantes da ocorrência de surtos e desenvolver modelos preditivos mais precisos.

3. Conclusão

Nas doenças endêmicas e reemergentes, persistem desafios significativos relacionados à predição a partir de dados genômicos e metadados com características heterogêneas e lacunas

de cobertura. Ainda assim, as iniciativas atuais apresentam indicadores promissores [3] de que soluções efetivas podem ser alcançadas em um futuro próximo.

O esforço atual concentra-se em identificar as variáveis que determinam a ocorrência de novos surtos, estabelecer suas correlações e validações estatísticas, e integrá-las às tecnologias disponíveis, que evoluem em escala exponencial. Estudos recentes apontam para a viabilidade de não apenas prever surtos, mas fazê-lo em tempo hábil para implementação de respostas eficazes. Espera-se alcançar elevado grau de confiança nessas previsões, contribuindo assim para o fortalecimento dos sistemas de vigilância em saúde pública globalmente.

Agradecimentos. Os autores agradecem o *Centre for Epidemic Response and Innovation (CERI)*, o Consórcio Climade, o Projeto NAVIO e a Rede Saúde Única pela colaboração. Além disso, agradecemos à agência de fomento à pesquisa Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), ao projeto *INFORM Africa* pelo apoio financeiro, e ao Programa de Pós-Graduação em Bioinformática da UFMG.






4. Referências

1. Lo SW, Jamroz D. Genomics and epidemiological surveillance. *Nat. Rev Microbiol.* 2020; 18:478.
2. Clarke J, Lim A, Gupte P, et al. A global dataset of publicly available dengue case count data. *Sci Data* 2024; 11:296.
3. Lim A, Shearer FM, Sewalk K, et al. The overlapping global distribution of dengue, chikungunya, Zika and yellow fever. *Nat Commun.* 2025; 16:3418.
4. Xavier JS, Moir M, Tegally H, et al. SARS-CoV-2 Africa dashboard for real-time COVID-19 information. *Nat Microbiol.* 2023; 8:1-4.
5. Tegally H, San JE, Cotten M, et al. The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science.* 2022; 378:eabq5358.
6. Satam H, Joshi K, Mangrolia U, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)* 2023; 12:997.
7. Freitas AS, Barboza Pinto H. Sequenciamento NGS: status e perspectivas. *BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional* 2021; 1:38-46.
8. Messina JP, Brady OJ, Golding N, et al. The current and future global distribution and population at risk of dengue. *Nat Microbiol.* 2019; 4:1508-1515.
9. Baes Pereira S, Conrad Bohm B, Dos Reis Gomes A, et al. Emergence and spatiotemporal incidence of dengue in Rio Grande do Sul, Brazil. *Sci Rep.* 2025; 15:18933.
10. Horstick O, Tozan Y, Wilder-Smith A. Reviewing dengue: still a neglected tropical disease? *PLoS Negl Trop Dis.* 2015; 9:e0003632.
11. Kraemer MUG, Tsui JL-H, Chang SY, et al. Artificial intelligence for modelling infectious disease epidemics. *Nature* 2025; 638:623-635.

5

Síntese quantitativa de modelos matemáticos, estatísticos e de inteligência artificial para biomarcadores, mecanismos e reposicionamento terapêutico

Autores 5.1

Rossana O. Souza , Wellington Francisco Rodrigues , Aristóteles Goés-Neto , Siomar de Castro Soares , Marcos A. dos Santos 

Revisão: Alessandra Lima , Diego Mariano , Tatiane Senna Bialves 

Cite este artigo 5.1

Souza, RO *et al.* **Síntese quantitativa de modelos matemáticos, estatísticos e de inteligência artificial para biomarcadores, mecanismos e reposicionamento terapêutico.** *BIOINFO.* Vol. 6. p.05 (2026). doi: 10.51780/bioinfo-06-05

Saiba mais 5.1

Este artigo está disponível em <https://bioinfo.com.br/sintese-quantitativa-de-modelos-matematicos-estatisticos-e-de-inteligencia-artificial-para-biomarcadores-mecanismos-e-reposicionamento-terapeutico/>

Síntese quantitativa de modelos matemáticos, estatísticos e de inteligência artificial para biomarcadores, mecanismos e reposicionamento terapêutico

Rossana O. Souza¹ , Wellington Francisco Rodrigues¹ , Aristóteles Goés-Neto¹ ,
Siomar de Castro Soares , Marcos A. dos Santos¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

rossanasouza@ufmg.br

Resumo. *A interface entre informática e ciências da saúde consolidou um ecossistema de modelagem capaz de transformar dados biomédicos em evidência acionável, embora ainda marcado por heterogeneidade metodológica e lacunas de validação. Este trabalho aborda esse cenário por meio de uma síntese quantitativa das aplicações de modelos matemáticos, estatísticos e de inteligência artificial (IA) em três frentes: identificação de biomarcadores, elucidação de mecanismos de doença e (re)posicionamento/desenvolvimento terapêutico. A abordagem incluiu busca semântica em repositórios acadêmicos amplos, triagem por elegibilidade (modelagem avançada, foco em saúde humana, desfechos mensuráveis, descrição metodológica e validação) e extração padronizada de métodos, focos, áreas terapêuticas e estratégias de validação, seguida de harmonização e análise de co-ocorrências. Os resultados mostram repurposing como foco predominante (26% das menções), seguido por mecanismos (16%) e biomarcadores (14%). A diversidade algorítmica é ampla: ML clássico/ensembles e deep learning dominam as aplicações, enquanto modelos lógicos/booleanos e biologia de sistemas (ODE/PDE) são centrais para estudos mecanísticos. As validações combinam cross-validation, ensaios experimentais (in vitro/in vivo) e evidência do mundo real (EHR/claims), ainda com heterogeneidade de reporte. A síntese oferece um panorama objetivo das práticas atuais e destaca oportunidades de padronização de métricas, expansão de validações externas e integração multi-ômica, contribuindo para acelerar a translação para a medicina de precisão.*

Palavras-chave: *inteligência artificial em saúde; biomarcadores; mecanismos de doença; reposicionamento de fármacos; síntese quantitativa.*

1. Introdução

A interface entre informática e ciências da saúde consolidou um ecossistema de modelagem matemática, estatística e de IA capaz de transformar dados biomédicos em evidência acionável, integrando multi-ômicas, registros clínicos e conhecimento biomolecular [1]. Métodos que vão de regressão e *ensembles* a *deep learning* têm sido aplicados para identificar biomarcadores, elucidar mecanismos e acelerar a (re)descoberta de fármacos, com avanços em oncologia, neurodegeneração, mapeamento de vias causais e priorização de candidatos a reposicionamento, validados computacionalmente, experimentalmente e com dados do mundo

real (EHR/claims) [2, 3]. Apesar do progresso, persistem desafios: heterogeneidade de dados e métricas, escassez de validação experimental, vieses de amostragem, limitações de reprodutibilidade e dificuldades de interpretabilidade clínica [4]. Assim, este trabalho realizou uma revisão guiada por busca semântica e extração padronizada para mapear aplicações de modelos matemáticos, estatísticos e de IA voltadas à identificação de biomarcadores, elucidação de mecanismos e reposicionamento de fármaco, quantificando distribuição por foco, métodos, áreas de doença e estratégias de validação, a fim de evidenciar oportunidades e limitações com impacto potencial na qualidade de vida [5].

2. Metodologia

Foi realizada uma síntese quantitativa de aplicações de modelos matemáticos, estatísticos e de IA em saúde. A estratégia incluiu busca semântica em repositórios indexados (e.g., *Semantic Scholar*, *OpenAlex*) focada em biomarcadores, mecanismos e reposicionamento terapêutico, seguida de triagem com critérios a priori (modelagem além de estatística descritiva, dados de saúde humana, descrição metodológica verificável e validação). Foram incluídos estudos originais, revisões sistemáticas e meta-análises.

A extração de dados seguiu protocolo padronizado com apoio de modelo de linguagem para raspar métodos, foco, área, fontes de dados, validações e achados-chave, com checagem humana amostral para consistência. A harmonização incluiu limpeza textual, padronização de abreviações, expansão de campos multivalorados, remoção de valores vazios e normalização de rótulos. Métodos foram agrupados por palavras-chave em classes (*Deep Learning*; *ML*; *Lógico*; *Bayesian*; etc), e duplicidades foram evitadas por deduplicação por par estudo-rótulo.

As análises consistiram em contagens, proporções e ocorrências (doença \times foco; validação \times foco), visualizadas por barras, *bubble plot* e *heatmaps*. A robustez decorreu de padronização, expansão de listas, deduplicação e clareza de que as métricas representam menções.

3. Conclusão

A síntese quantitativa (Fig. 1) indica que o reposicionamento de fármacos foi o foco predominante (26%), seguido por mecanismos de doença (16%) e biomarcadores (14%). Temas adicionais incluíram resposta a fármacos (8%), interações fármaco-fármaco e descoberta de alvos (4% cada), além de tópicos esparsos ligados à aplicabilidade e medicina de precisão. Houve ampla diversidade algorítmica: *Random Forest* foi mais frequente (7,5%), seguida por *SVM* (4,5%). Entre grupos, destacaram-se *Other/Hybrid* (43,3%), *ML* (14,9%) e *Deep Learning* (13,4%), enquanto modelagem lógica (7,5%) e biologia de sistemas (*ODE/PDE*; 6,0%) foram relevantes para mecanismos e alvos; *Bayesian* e *graph* e testes estatísticos completaram o panorama (~4,5% cada).

O painel doença \times foco mostra predominância de oncologia, seguida por neurodegeneração e casos pontuais em cardiovascular, COVID-19, psoríase, osteoporose, saúde reprodutiva e transtornos de splicing. O reposicionamento concentrou-se em câncer e

neurodegeneração, enquanto mecanismos foram analisados com redes, lógica e modelos *ODE/PDE*. Quanto à validação, observou-se uso de *cross-validation* (5-, 10-, 25-*fold* e variantes *two-tiered*), ensaios experimentais (in vitro/in vivo, dose–response, modelos murinos) e evidência do mundo real (*EHR/claims*). Houve ainda comparações com *SOTA* e validações externas; parte dos estudos, porém, não detalhou claramente a estratégia.

No geral, o reposicionamento é a aplicação mais recorrente de modelos estatísticos em saúde, impulsionada por *ML ensembles* e *deep learning*, enquanto modelos lógicos e de sistemas seguem essenciais para elucidar mecanismos e alvos. A diversidade de métodos e validações é ampla e o corpo de evidências reforça o potencial translacional desses enfoques para identificar biomarcadores, compreender vias patológicas e priorizar reposicionamentos; próximos passos incluem padronização de métricas, validações externas ampliadas e integração multiômica para acelerar aplicações em medicina de precisão.

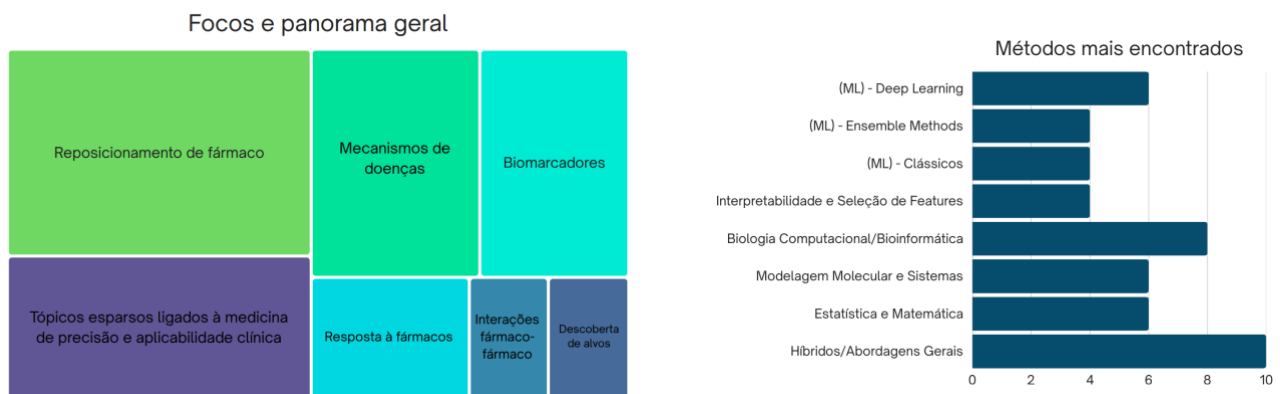


Figura 1. Aplicações de modelos computacionais em ciências da saúde: síntese quantitativa dos estudos incluídos. (A) Focos dos estudos (primário + secundário). (B) Métodos computacionais mais citados. (C) Mapa de bolhas doença × foco (Top-10). (D) Heatmap validação × foco. Notas técnicas: dados padronizados.

Disponibilidade dos dados. O material suplementar e imagens de alta resolução deste manuscrito estão disponíveis em https://github.com/rossanaoliveirasouza/bioinformatica_2035/invitations.


4. Referências

- [1] Raschka T, Sood M, Schultz B, Altay A, Ebeling C, Fröhlich H. AI reveals insights into link between CD33 and cognitive impairment in Alzheimer's Disease. *PLoS Comput Biol*. 2023;19(2):e1009894.
- [2] Aamer N, Asim MN, Dengel A. COMIC: Explainable Drug Repurposing via Contrastive Masking for Interpretable Connections. *bioRxiv*. 2025:2025.03.26.645428.
- [3] Montagud A, Béal J, Tobalina L, Traynard P, Subramanian V, Szalai B, et al. Patient-specific Boolean models of signaling networks guide personalized treatments. *bioRxiv*. 2021:2021.07.28.454126.
- [4] Maeser D, Gruener RF, Galvin R, Lee A, Koga T, Grigore FN, et al. Integration of Computational Pipeline to Streamline Efficacious Drug Nomination and Biomarker Discovery in Glioblastoma. *Cancers (Basel)*. 2024;16(9).
- [5] Voutouri C, Nikmaneshi MR, Hardin CC, Patel AB, Verma A, Khandekar MJ, et al. In silico dynamics of COVID-19 phenotypes for optimizing clinical management. *Proc Natl Acad Sci U S A*. 2021;118(3).

6

Explorando a patogenicidade e resistência de *Corynebacterium aurimucosum* por abordagens genômicas

Autores 6.1

Luan Freitas Lana , Max Roberto Batista Araújo , Diogo Luiz de Carvalho Castro , Genrikh Ashniev Alfredovich , Rafael Augusto Ribeiro , Vinícius Dias do Carmo Costa , Evandro Bento Rodrigues , Maria Letícia Carneiro Rodrigues , Vasco Azevedo 

Revisão: Alessandra Lima , Bibiana Fam , Diego Mariano , Sheila Cruz 

Cite este artigo 6.1

Lana, L. F. *et al.* Explorando a patogenicidade e resistência de *Corynebacterium aurimucosum* por abordagens genômicas. *BIOINFO*. Vol. 6. p.06 (2026). doi: 10.51780/bioinfo-06-06

Saiba mais 6.1

Este artigo está disponível em <https://bioinfo.com.br/explorando-a-patogenicidade-e-resistencia-de-corynebacterium-aurimucosum-por-abordagens-genomicas/>

Explorando a patogenicidade e resistência de *Corynebacterium aurimucosum* por abordagens genômicas

Luan Freitas Lana^{1†} , Max Roberto Batista Araújo^{1,2†} , Diogo Luiz de Carvalho Castro¹ , Genrikh Ashniev Alfredovich³ , Rafael Augusto Ribeiro¹ , Vinícius Dias do Carmo Costa¹ , Evandro Bento Rodrigues^{1,2} , Maria Letícia Carneiro Rodrigues¹ , Vasco Azevedo^{1*} 

†Estes autores contribuíram igualmente para este trabalho.

1 Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

2 Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil

3 Instituto de Problemas de Transmissão de Informação da Academia Russa de Ciências, Moscou, Rússia

*vascoariston@gmail.com

Resumo. O gênero *Corynebacterium* inclui espécies comensais e oportunistas, como *C. aurimucosum*, frequentemente negligenciadas, mas associadas a infecções graves, incluindo bacteremias e outras. A identificação precisa apresenta dificuldades por métodos fenotípicos tradicionais, sendo a MALDI-TOF MS uma ferramenta promissora, mas dependente de bases de dados atualizadas. A escassez de estudos genômicos e funcionais limita a compreensão da patogenicidade, resistência aos antimicrobianos e diversidade genética do gênero. Esse trabalho apresenta um pipeline para análises genômicas de linhagens de *C. aurimucosum*, com avaliação da susceptibilidade antimicrobiana, biofilme, virulência *in vivo*, além de sequenciamento genômico, pangenômica e predição de genes de virulência e resistência. Com os dados obtidos, é possível identificar determinantes moleculares, fortalecer a vigilância epidemiológica e orientar estratégias de controle, que passam a ser mais robustas.

Palavras-chave: *Corynebacterium aurimucosum*, resistência antimicrobiana, pangenoma, virulência, biofilme, MALDI-TOF MS.

1. Introdução

O gênero *Corynebacterium*, inclui diversas espécies caracterizadas e válidas [1], tradicionalmente conhecido pela patogenicidade de *Corynebacterium diphtheriae*, tem revelado uma diversidade crescente de espécies comensais que podem atuar como patógenos oportunistas, principalmente em pacientes imunocomprometidos. O aumento de relatos de infecções graves envolvendo-o tem levantado preocupações sobre a adequação dos métodos tradicionais de identificação [2-4], sugerindo que equívocos na distinção entre colonização e infecção ativa possam ocorrer.

A revolução trazida pela espectrometria de massa por ionização e dessorção a laser assistida por matriz com análise por tempo de voo (MALDI-TOF MS) na microbiologia clínica

foi um marco na identificação rápida e precisa de microrganismos, mas sua utilidade está diretamente relacionada à constante atualização das bases de dados com novas espécies, muitas das quais ainda são subnotificadas ou mal caracterizadas [5]. Novos estudos têm contribuído para a identificação de espécies até então negligenciadas, como *Corynebacterium aurimucosum*, que, apesar de descrita há mais de uma década, continua diagnosticada de forma insuficiente devido à confiabilidade limitada de métodos fenotípicos [6-8].

Essa espécie já foi associada a uma variedade de quadros clínicos, incluindo infecções urinárias, bacteremias e infecções de pele, reforçando a necessidade de investigações mais detalhadas sobre seu papel como patógeno [9]. A escassez de estudos robustos sobre *Corynebacterium* spp. em bacteremias, frequentemente limitados a relatos de casos esporádicos, como em um estudo que relatou 30 casos de bacteremia verdadeira, destaca a urgência de uma análise mais sistemática dessas infecções [10,11]

Modificações taxonômicas no gênero *Corynebacterium* têm sido impulsionadas pelo acesso ao sequenciamento genômico e ferramentas de bioinformática, que permitem comparações globais de cepas bacterianas. A abordagem pangenômica tem revelado a diversidade genética dos patógenos deste gênero, facilitando o monitoramento de fenótipos clínicos e a compreensão das bases moleculares da resistência antimicrobiana, virulência e metabolismo [12,13].

Assim, buscamos aqui demonstrar como ferramentas atuais de Bioinformática podem auxiliar a suprir lacunas de identificação, apoiar a vigilância epidemiológica e ampliar o entendimento sobre a diversidade e o risco clínico desse microrganismo.

2. Contextualização

A compreensão sobre o papel clínico das espécies do gênero *Corynebacterium* ainda é limitada. Embora algumas delas, como *C. aurimucosum*, já tenham sido associadas a infecções graves, muitas passam despercebidas no diagnóstico de rotina. Essa negligência ocorre, em grande parte, porque faltam estudos que revelem como esses microrganismos se comportam, quais mecanismos utilizam para causar doença e de que forma podem resistir a antibióticos. Sem informações genômicas e funcionais detalhadas, torna-se difícil reconhecer sinais de alerta e monitorar a emergência de linhagens potencialmente mais virulentas.

É justamente nesse cenário que a Bioinformática desponta como um instrumento importante. Ao reunir dados de sequenciamento e aplicar pipelines analíticos, é possível comparar genomas, identificar genes-chave e mapear mecanismos de virulência e resistência que não seriam detectados apenas por testes laboratoriais tradicionais. Esse tipo de análise amplia de forma significativa a compreensão da espécie em nível molecular e revela padrões evolutivos que ajudam a explicar o comportamento clínico observado.

Um estudo que integre essas abordagens tem potencial para preencher lacunas importantes. Ele pode aprimorar o diagnóstico, apoiar a identificação precoce de linhagens preocupantes e fornecer pistas para o desenvolvimento de estratégias mais eficazes de tratamento e controle. Dessa forma, a Bioinformática não apenas complementa a vigilância

epidemiológica, como também impulsiona a capacidade de antecipar riscos e responder melhor às infecções por *Corynebacterium*.

3. A Bioinformática como Ferramenta de Precisão Clínica

A crescente detecção de *C. aurimucosum* em amostras clínicas, como sangue, urina, lesões cutâneas e até fluidos abdominais, tem sido um alerta de que mais estudos devem ser realizados. Apesar da identificação rápida por métodos como MALDI-TOF MS, ainda sabemos pouco sobre os mecanismos que permitem a esse microrganismo sobreviver, causar infecção ou desenvolver resistência a antimicrobianos. É justamente nessa lacuna de conhecimento que a Bioinformática se torna essencial.

O sequenciamento genômico de isolados clínicos oferece uma oportunidade de enxergar além do que testes fenotípicos tradicionais conseguem mostrar. A partir dos dados brutos de DNA, algoritmos especializados permitem reconstruir genomas completos com alta precisão e avaliar sua qualidade. Uma vez que esses genomas são validados, ferramentas de anotação funcional entram em cena para vasculhar bancos globais como CARD (<https://card.mcmaster.ca/>) e VFDB (<https://www.mgc.ac.cn/VFs/>), verdadeiras bibliotecas de genes de resistência e virulência, respectivamente. Essa etapa revela os elementos genéticos que caracterizam uma bactéria como um agente infeccioso mais adaptado e mais persistente.

Estudos pangenômicos possibilitam observar “a espécie como um todo”, identificando quais genes são amplamente compartilhados e quais aparecem apenas em alguns isolados, indicando adaptação a nichos específicos ou emergência de linhagens mais virulentas. Em paralelo, ensaios *in vitro* e *in vivo*, como capacidade de formação de biofilme e ensaios de virulência em *Galleria mellonella*, ajudam a confrontar a genética com o comportamento real da bactéria.

Ao combinar sequenciamento, análise computacional e experimentação, a Bioinformática não apenas esclarece a identidade do microrganismo, mas revela sua diversidade, sua evolução e seus potenciais riscos clínicos. Nessa medida, podemos fornecer respostas que hoje são urgentes: o que torna certos isolados mais perigosos? Como prever perfis de resistência? Quais características podem orientar diagnósticos mais precisos e terapias mais eficazes? É justamente nessa interligação de dados, tecnologia e microbiologia que surgem as estratégias mais promissoras para compreender e enfrentar patógenos emergentes como *C. aurimucosum*.

4. Conclusão

A caracterização integrada de *C. aurimucosum* poderá fornecer insights sobre os mecanismos de resistência e virulência, aprimorando diagnósticos, monitoramento epidemiológico e manejo clínico de infecções oportunistas, além de subsidiar o desenvolvimento de intervenções terapêuticas mais eficazes e políticas de controle hospitalar.

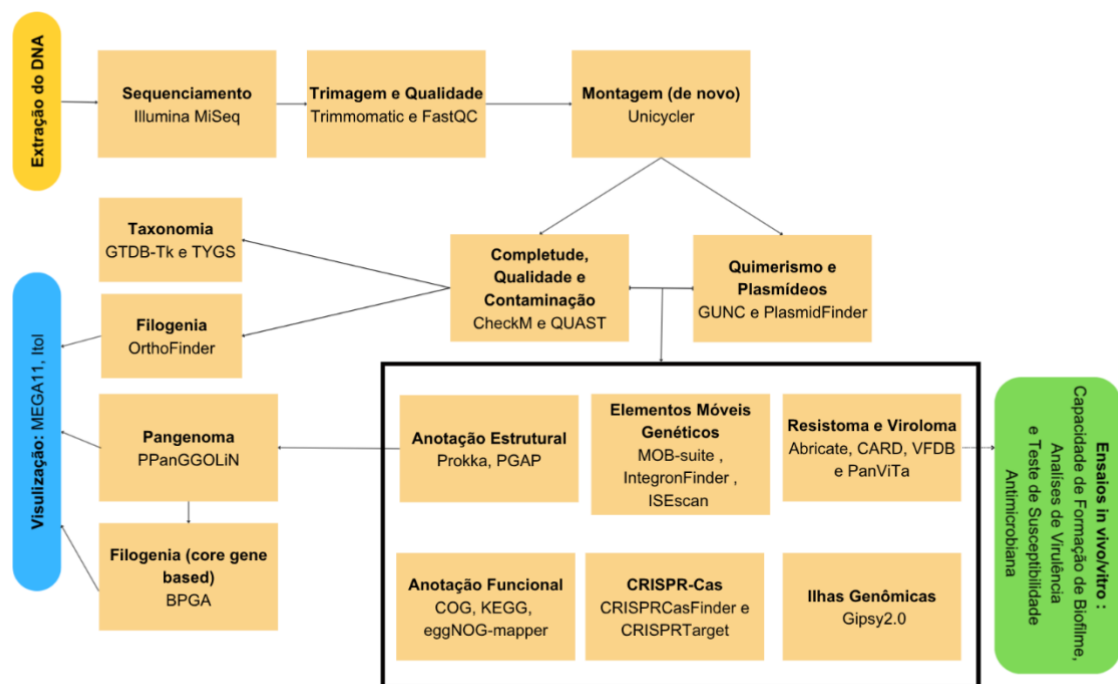


Figura 1. Visão geral do fluxo de trabalho. Cada caixa representa um componente correspondente a uma análise de dados, indicada em negrito. Os softwares utilizados em cada análise estão indicados em baixo.

5. Referências

- [1] Parte, AC et al. **List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ.** *Int J Syst Evol Microbiol.* 2020;70(11):5607-5612. <https://doi.org/10.1099/ijsem.0.004332>
- [2] Sharma, NC et al. **Diphtheria.** *Nat Rev Dis Primers.* 2019;5(1):81. <https://doi.org/10.1038/s41572-019-0131-y>
- [3] Wagner, KS et al. **Diphtheria in the United Kingdom, 1986–2008: the increasing role of *Corynebacterium ulcerans*.** *Epidemiol Infect.* 2010;138(11):1519-1530. <https://doi.org/10.1017/S0950268810001895>
- [4] Indumathi VA, Shikha R, Suryaprakash DR. **Diphtheria-like illness in a fully immunised child caused by *Corynebacterium pseudodiphtheriticum*.** *Indian J Med Microbiol.* 2014;32(4):443-445. <https://doi.org/10.4103/0255-0857.142250>
- [5] Araújo MRB, Lana LF, Seabra LF. **Use of MALDI-TOF MS in diagnostic microbiology for identification of species that conventional methods usually fail to identify.** *Clin Biomed Res.* 2019;39(2):171-174. <https://doi.org/10.4322/2357-9730.90933>
- [6] Badell, E et al. ***Corynebacterium rouxii* sp. nov., a novel member of the diphtheriae species complex.** *Res Microbiol.* 2020;171(3-4):122-127. <https://doi.org/10.1016/j.resmic.2020.02.003>
- [7] de Oliveira Sant'Anna, L et al. ***Corynebacterium guaraldiae* sp. nov.: a new species of *Corynebacterium* from human infections.** *Braz J Microbiol.* 2023;54(2):779-790. <https://doi.org/10.1007/s42770-023-00938-y>
- [8] Cappelli, EA et al. **Expanding the bacterial diversity of the female urinary microbiome: description of eight new *Corynebacterium* species.** *Microorganisms.* 2023;11(2):388. <https://doi.org/10.3390/microorganisms11020388>
- [9] Lefèvre, CR et al. **Clinical relevance and antimicrobial susceptibility profile of the unknown human pathogen *Corynebacterium aurimucosum*.** *J Med Microbiol.* 2021;70(3):001334.



<https://doi.org/10.1099/jmm.0.001334>

[10] Rezaei, BK et al. **Corynebacterium jeikeium endocarditis: A case report and comprehensive review of an underestimated infection.** IDCases. 2017;11:e00462. <https://doi.org/10.1016/j.idcr.2017.11.004>

[11] Rasmussen, M et al. **Corynebacterium ulcerans: An emerging pathogen causing diphtheria-like disease.** Eur J Clin Microbiol Infect Dis. 2019;38:2039-2047. <https://doi.org/10.1007/s10096-019-0369-86>

[12] Trost E, Blom J, Soares SC, et al. **Comparative genomics of Corynebacterium diphtheriae reveals insights into evolution, pathogenicity, and population structure.** J Bacteriol. 2012;194(16):4463-4477.

<https://doi.org/10.1128/JB.00183-12>

[13] Jesus HNR, da Silva DM, da Cunha P, et al. **Genomic insights into Corynebacterium striatum isolates from hospital outbreaks.** Funct Integr Genomics. 2022;22:511-526. <https://doi.org/10.1007/s10142-022-00932-x>

7

Caracterização genômica e fenotípica de *Corynebacterium pseudotuberculosis*: utilização da bioinformática no controle da linfadenite caseosa

Autores 7.1

Maria Leticia Carneiro Rodrigues , Max Roberto Batista Araújo , Virgínia Braga da Silva , Evandro Bento Rodrigues , Luan Freitas Lana , Diogo Luiz de Carvalho Castro , Genrikh Ashniev Alfredovich , Rafael Augusto Ribeiro , Vinícius Dias do Carmo Costa , Patrícia Yoshida Faccioli-Martins , Vasco Azevedo 

Revisão: Bibiana Fam , Diego Mariano , Rafael Eduardo Oliveira Rocha , Savio Costa , Sheila Cruz Araújo 












Cite este artigo 7.1

Rodrigues, M. L. C. *et al.* Caracterização genômica e fenotípica de *Corynebacterium pseudotuberculosis*: utilização da bioinformática no controle da linfadenite caseosa. *BIOINFO.* Vol. 6. p.07 (2026). doi: 10.51780/bioinfo-06-07

Saiba mais 7.1

Este artigo está disponível em <https://bioinfo.com.br/caracterizacao-genomica-e-fenotipica-de-corynebacterium-pseudotuberculosis-utilizacao-da-bioinformatica-no-controle-da-lynfadenite-caseosa/>

Caracterização genômica e fenotípica de *Corynebacterium pseudotuberculosis*: utilização da bioinformática no controle da linfadenite caseosa

Maria Leticia Carneiro Rodrigues^{1†} , Max Roberto Batista Araújo^{1,2†} , Virgínia Braga da Silva¹ , Evandro Bento Rodrigues^{1,2} , Luan Freitas Lana¹ , Diogo Luiz de Carvalho Castro² , Genrikh Ashniev Alfredovich³ , Rafael Augusto Ribeiro¹ , Vinícius Dias do Carmo Costa¹ , Patrícia Yoshida Faccioli-Martins⁴ , Vasco Azevedo^{1*} 

†Estes autores contribuíram igualmente para este trabalho.

¹ Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil;

² Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil;

³ Instituto de Problemas de Transmissão de Informação da Academia Russa de Ciências, Moscou, Rússia;

⁴ Pesquisadora da Embrapa Caprinos e Ovinos, Sobral, Ceará, Brasil

* vascoariston@gmail.com

Resumo. A Linfadenite Caseosa (LC), causada por *Corynebacterium pseudotuberculosis*, é uma doença crônica que afeta caprinos e ovinos, gerando abscessos em linfonodos e prejuízos econômicos significativos, além de representar risco zoonótico sob a perspectiva da Saúde Única. A crescente resistência antimicrobiana (RAM) a β -lactâmicos, macrolídeos e tetraciclina compromete o controle da doença, enquanto os mecanismos genéticos subjacentes permanecem pouco compreendidos. Este estudo propõe uma abordagem integrada combinando análises fenotípicas (disco-difusão, concentração inibitória mínima e teste de Elek para toxina diftérica) e genômicas, incluindo sequenciamento de nova geração, montagem de novo, anotação e predição de genes de resistência e elementos genéticos móveis. Ferramentas de bioinformática como RAST, Prokka, Resfinder, PlasmidFinder e CRISPRCasFinder serão aplicadas, com análises de ANI, AAI e filogenia. Os resultados esperados fornecerão insights sobre diversidade genética, mecanismos de resistência e virulência, subsidiando estratégias de manejo, controle sanitário e terapêutico sustentável da LC.

Palavras-chave: *Corynebacterium pseudotuberculosis*, Linfadenite Caseosa, resistência antimicrobiana, genômica, bioinformática, toxina diftérica, elementos genéticos móveis, CRISPR.

1. Introdução

A Linfadenite Caseosa (LC) é uma doença infectocontagiosa crônica que afeta principalmente caprinos e ovinos, causando abscessos em linfonodos e, ocasionalmente, comprometimento de órgãos internos [1]. O agente etiológico, *Corynebacterium pseudotuberculosis*, é uma bactéria Gram-positiva, pleomórfica e intracelular facultativa [2]. Pertencente ao complexo *Corynebacterium diphtheriae*, um grupo de espécies geneticamente relacionadas que compartilham características fenotípicas e fatores de virulência, compartilha elementos como a possibilidade de produção da toxina diftérica (TD), codificada pelo gene *tox* de origem bacteriofágica, que inibe a síntese proteica e leva à morte celular [3–5].

No Brasil, a LC representa um desafio sanitário e econômico, afetando grande parte dos rebanhos caprinos e ovinos, com implicações na produtividade, na condenação de carcaças e na saúde ocupacional humana [6-8]. Além disso, a resistência aos antimicrobianos (RAM) de *C. pseudotuberculosis* a β -lactâmicos, macrolídeos e tetraciclina compromete o controle da infecção [9].

É relevante a caracterização das linhagens de *C. pseudotuberculosis* por meio de testes fenotípicos e sequenciamento genômico, visando identificar genes de resistência, elementos genéticos móveis e a presença do gene *tox*, além de comparações intra e interespecíficas com o complexo *C. diphtheriae*. A abordagem integrada pode fornecer informações sobre diversidade genética, mecanismos de resistência e fatores de virulência, subsidiando estratégias de controle da LC e práticas sustentáveis em saúde animal, humana e ambiental.



Figura 1. Abscesso formado por Linfadenite Caseosa em linfonodo pré-escapular em bode Saanen.

2. Contextualização

O *C. pseudotuberculosis* ao provocar abscessos (Figura 1) e demais manifestações clínicas nos animais afetados, gera como resultados graves perdas econômicas na caprinocultura e ovinocultura, além de representar risco zoonótico e impacto sob a ótica da Saúde Única. Considerando-se a elevada prevalência da doença, a resistência crescente aos antimicrobianos utilizados e as lacunas no conhecimento dos mecanismos genéticos envolvidos, torna-se fundamental caracterizar isolados clínicos e realizar análises genômicas a fim de subsidiar estratégias de manejo, controle sanitário e terapêutico eficazes.

3. Estratégias de análises e Desenvolvimento

A LC afeta rebanhos em diferentes regiões do Brasil, sendo o tratamento desafiador, e comumente não realizado com o uso de antibióticos. A escolha pelo uso de antissépticos não impede que ocorra o desenvolvimento e transmissão de resistência aos antimicrobianos. Para além, os fatores de virulência envolvidos também possuem grande importância que ultrapassam a preocupação com os fatores clássicos, como a TD e a fosfolipase D. Para um melhor

entendimento da patogênese, pesquisadores costumam trabalhar com diversas amostras da bactéria provenientes de rebanhos reais. Em estudos recentes, foram analisados isolados de regiões como o Nordeste e Minas Gerais, incluindo materiais provenientes de coleções oficiais. A partir desses isolados, investiga-se como cada cepa responde a diferentes antibióticos e quais padrões de susceptibilidade ou resistência predominam na população bacteriana. Testes laboratoriais como o antibiograma por disco-difusão e/ou a determinação da concentração inibitória mínima (CIM) permitem observar, de forma prática, o quanto a bactéria é capaz de tolerar medicamentos.

Essas informações são essenciais, mas têm limitações: mostram o comportamento das bactérias diante dos antibióticos, mas não explicam o porquê elas se tornam resistentes. É aqui que a bioinformática entra como uma ferramenta relevante. A partir do DNA extraído das bactérias, é possível sequenciar e analisar todo o seu genoma, um mapa completo que revela genes, mutações, elementos móveis, toxinas e fatores de virulência. Plataformas de montagem e anotação genética, como SPAdes, RAST e Prokka, ajudam a reconstruir e interpretar esse material, enquanto ferramentas comparativas como ANI (Identidade Média de Nucleotídeos), dDDH (Hibridização DNA-DNA digital) e TYGS confirmam a identidade taxonômica e mostram o quão “próximas ou distantes” são as diferentes linhagens.

Ao examinar o genoma, pesquisadores conseguem identificar genes de resistência (por exemplo, por meio de ResFinder ou RGI), elementos genéticos móveis que favorecem sua disseminação (como plasmídeos, ilhas genômicas, transposons e fagos, detectados por PlasmidFinder, GIPSY, ISEScan e PHASTEST), além de sistemas CRISPR-Cas que representam um “sistema imune bacteriano”.

Unindo microbiologia, genética e bioinformática, podemos correlacionar o que é observado no laboratório com o que está presente no genoma. Por exemplo, quando uma cepa mostra resistência elevada a um antibiótico específico, ferramentas ajudam a confirmar, *in silico*, se existe um gene responsável por esse comportamento ou se a resistência se deve a alterações estruturais, como mutações pontuais, ou mesmo indicar que possa ser por mecanismos ainda pouco conhecidos.

É possível compreender melhor os mecanismos biológicos envolvidos tanto na resistência como na virulência, tornando-se essencial para o desenvolvimento de estratégias de controle mais eficazes. A bioinformática permite explorar o genoma das bactérias com rapidez, precisão e profundidade, abrindo caminho para diagnósticos mais assertivos, terapias mais adequadas e decisões sanitárias mais informadas. Ao revelar o que está por trás da patogênese, a bioinformática se consolida como uma aliada indispensável para a saúde animal, para a produção sustentável e para a prevenção de surtos que podem comprometer toda a cadeia produtiva.

Conclusão

Combinando análises laboratoriais tradicionais com ferramentas modernas de bioinformática, torna-se possível revelar os mecanismos genéticos que sustentam a resistência aos antimicrobianos, bem como identificar genes-chave e compreender a diversidade das linhagens

circulantes. Esses recursos ampliam nossa capacidade de interpretar o comportamento bacteriano e orientam estratégias mais eficazes de diagnóstico, controle e tratamento. A integração entre microbiologia e análise genômica surge como um caminho essencial para enfrentar problemas de forma mais precisa e informada.










4. Referências

- [1] EMBRAPA CAPRINOS E OVINOS. Centro de Inteligência e Mercado de Caprinos e Ovinos. Produção mundial. [S.l.: s.n.], v. 11, 2020. Disponível em: <https://www.embrapa.br/cim-inteligencia-e-mercado-de-caprinos-e-ovinos>. Acesso em: 29 out. 2025.
- [2] DORELLA, F. A.; PACHECO, L. G. C.; OLIVEIRA, S. C.; MIYOSHI, A.; AZEVEDO, V. *Corynebacterium pseudotuberculosis*: microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Veterinary Research*, v. 37, n. 2, p. 201–218, 2006. doi:10.1051/vetres:2005056.
- [3] SHARMA, N. C. et al. Diphtheria. *Nature Reviews Disease Primers*, v. 5, p. 81, 2019. doi:10.1038/s41572-019-0131-y.
- [4] PRYGIEL, M. et al. New *Corynebacterium* species with the potential to produce diphtheria toxin. *Pathogens*, v. 11, n. 11, p. 1264, 2022. doi:10.3390/pathogens11111264.
- [5] DANGEL, A.; BERGER, A.; KONRAD, R.; SING, A. NGS-based phylogeny of diphtheria-related pathogenicity factors in different *Corynebacterium spp.* implies species-specific virulence transmission. *BMC Microbiology*, v. 19, n. 1, p. 28, 2019. doi:10.1186/s12866-019-1402-1.
- [6] ALVES, J. R. A. et al. Seroepidemiological study of caseous lymphadenitis in sheep from the Northeast region of Brazil using an indirect ELISA. *Tropical Animal Health and Production*, v. 52, n. 4, p. 1945–1952, 2020. doi:10.1007/s11250-020-02214-9.
- [7] CRUZ, G. R. B. et al. Aspectos sanitários na produção de caprinos e ovinos de produtores familiares no Semiárido Paraibano. *Revista Conexão UEPG*, v. 15, n. 2, p. 129–134, 2019. doi:10.5212/Rev.Conexao.v.15.i2.0001.
- [8] GALLARDO, A. A. et al. *Corynebacterium pseudotuberculosis* biovar ovis: evaluación de la sensibilidad antibiótica in vitro. *Revista Argentina de Microbiología*, v. 51, n. 4, p. 334–338, 2019. doi:10.1016/j.ram.2018.12.001.
- [9] JUDSON, R.; SONGER, J. G. *Corynebacterium pseudotuberculosis*: in vitro susceptibility to 39 antimicrobial agents. *Veterinary Microbiology*, v. 27, n. 2, p. 145–150, 1991. doi:10.1016/0378-1135(91)90005-Z.
- [10] SOUSA, T. de J.; PARISE, D.; PROFETA, R.; et al. Resequenciamento e mapeamento óptico revelam montagens incorretas e inversões reais em genomas de *Corynebacterium pseudotuberculosis*. *Scientific Reports*, v. 9, p. 16387, 2019. DOI: 10.1038/s41598-019-52695-4.

8

Desvendando a resistência em *Corynebacterium striatum*: uma investigação multidisciplinar do genoma ao desenho de inibidores

Autores 8.1

Diogo Luiz de Carvalho Castro , Max Roberto Batista Araújo , Genrikh Ashniev Alfredovich , Luan Freitas Lana , Rafael Augusto Ribeiro , Vinícius Dias do Carmo Costa , Evandro Bento Rodrigues , Maria Letícia Carneiro Rodrigues , Vasco Azevedo 

Revisão: Ariany Rosa Gonçalves , Diego Mariano , Rafael Eduardo Oliveira Rocha , Savio Costa 










Cite este artigo 8.1

Castro D. L. C. *et al.* Desvendando a resistência em *Corynebacterium striatum*: uma investigação multidisciplinar do genoma ao desenho de inibidores. *BIOINFO*. Vol. 6. p.08 (2026). doi: 10.51780/bioinfo-06-08

Saiba mais 8.1

Este artigo está disponível em <https://bioinfo.com.br/desvendando-a-resistencia-em-corynebacterium-striatum-uma-investigacao-multidisciplinar-do-genoma-ao-desenho-de-inibidores/>

Desvendando a resistência em *Corynebacterium striatum*: uma investigação multidisciplinar do genoma ao desenho de inibidores

Diogo Luiz de Carvalho Castro^{1†} , Max Roberto Batista Araújo^{1,2†} , Genrikh Ashniev Alfredovich³ , Luan Freitas Lana¹ , Rafael Augusto Ribeiro¹ , Vinícius Dias do Carmo Costa¹ , Evandro Bento Rodrigues^{1,2†} , Maria Letícia Carneiro Rodrigues¹ , Vasco Azevedo^{1*} 

†Estes autores contribuíram igualmente para este trabalho.

¹ Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

² Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil

³ Instituto de Problemas de Transmissão de Informação da Academia Russa de Ciências, Moscou, Rússia

* vascoariston@gmail.com

Resumo. A resistência aos antimicrobianos é um dos maiores desafios da saúde pública e, *Corynebacterium striatum* vem se destacando como um microrganismo oportunista capaz de causar infecções hospitalares difíceis de tratar. Isso ocorre porque este microrganismo utiliza diferentes estratégias para escapar da ação dos antibióticos, incluindo as bombas de efluxo, estruturas que expulsam o medicamento antes que ele cumpra sua função. Neste artigo, discutimos porque essa espécie se tornou tão problemática, como as bombas de efluxo contribuem para a resistência e de que maneira ferramentas de bioinformática podem ajudar na busca por novas alternativas terapêuticas, especialmente peptídeos bioativos capazes de inibir esses mecanismos. Ao aproximar genética, microbiologia e computação, mostramos como o estudo digital de microrganismos pode acelerar o desenvolvimento de soluções inovadoras contra bactérias multirresistentes.

Palavras-chave: Resistência aos antibióticos, *Corynebacterium striatum*, Bombas de efluxo, Ferramentas de Bioinformática, Triagem in silico de peptídeos

1. Introdução

A resistência aos antimicrobianos (RAM) e as infecções hospitalares são desafios globais, responsáveis por milhões de casos e mortes anuais [1–3]. Em 2019, a RAM esteve associada a 4,95 milhões de óbitos e pode atingir 10 milhões por ano até 2050, com grandes perdas econômicas [4,5]. No Brasil, estima-se cerca de 34 mil mortes diretas, anuais, destacando a urgência no desenvolvimento de novas terapias [6,7].

Nos últimos anos, *Corynebacterium striatum* deixou de ser um microrganismo quase irrelevante para se tornar uma preocupação real em hospitais. Embora faça parte da nossa própria pele, essa bactéria pode causar infecções graves quando encontra pacientes debilitados ou submetidos a longos tratamentos com antibióticos. O problema é ampliado por sua capacidade de resistir a diversas classes de medicamentos, o que torna o tratamento mais difícil e prolongado. Essa resistência ocorre por diferentes mecanismos, mas um dos mais importantes é a presença de bombas de efluxo, estruturas que atuam como “portas de emergência químicas”,

expulsando rapidamente os antibióticos do interior da célula. Com isso, mesmo medicamentos fortes podem se tornar ineficazes [7-13].

Para compreender por qual razão *C. striatum* se tornou tão resistente, precisamos olhar para o seu genoma, o conjunto de instruções que controla o funcionamento da célula. Alterações genéticas podem ativar mecanismos de defesa, aumentar a virulência e até favorecer a transmissão em ambientes hospitalares. Porém, devido à complexidade dessas informações, compreender a resistência aos antimicrobianos exige cada vez mais o uso de ferramentas computacionais capazes de analisar volumes maiores de dados de forma rápida e integrada.

2. A contribuição da bioinformática para desvendar a resistência

A bioinformática se tornou indispensável no estudo da resistência bacteriana porque permite transformar grandes quantidades de dados genéticos em conhecimento útil. Em vez de depender apenas de experimentos longos e caros em laboratório, pesquisadores podem utilizar o computador para identificar genes de resistência, prever o comportamento de proteínas e até simular o efeito de moléculas terapêuticas. Isso acelera a pesquisa, reduz custos e amplia a capacidade de investigação.

Por exemplo, ferramentas como ARIBA (<https://www.sanger.ac.uk/tool/ariba/>) e ResFinderFG (<https://genepi.food.dtu.dk/resfinder>) conseguem identificar automaticamente genes ligados à resistência antimicrobiana diretamente a partir das sequências genômicas. Outras plataformas, como eggNOG-mapper (<https://eggnog-mapper.embl.de/>), ajudam a entender a função desses genes, enquanto FastMLST (<https://anaconda.org/bioconda/fastmlst>), PHASTEST (<https://phastest.ca/>) e CRISPRCasFinder (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>) revelam características da linhagem que podem influenciar sua virulência ou adaptação. Para comparar diferentes isolados clínicos, programas como PYANI (<https://pypi.org/project/pyani/>), PPanGGOLiN (<https://github.com/labgem/PPanGGOLiN>), PHYLOViZ (<https://www.phyloviz.net/>) e PopPUNK (<https://www.sanger.ac.uk/tool/poppunk/>) mostram como as bactérias se organizam em grupos e como evoluem ao longo do tempo.

Essas ferramentas permitem compreender de forma muito mais profunda como *C. striatum* se adapta ao ambiente hospitalar e quais mecanismos utiliza para sobreviver aos antibióticos. Ao revelar o funcionamento interno da bactéria, a bioinformática se transforma em uma aliada fundamental no combate à resistência.

Como surge a resistência antimicrobiana

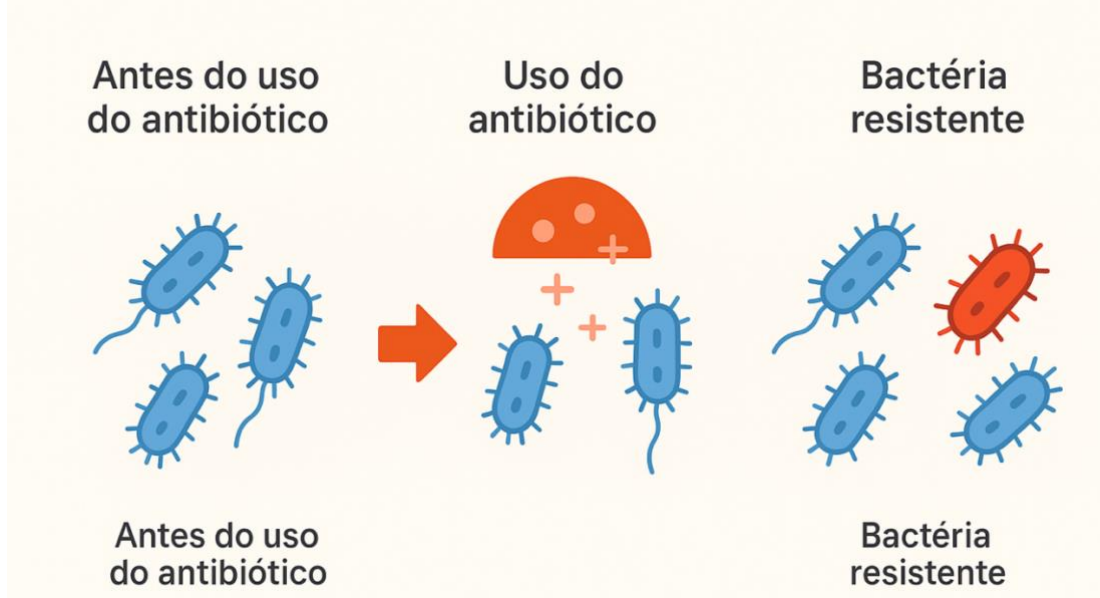


Figura 1. Ilustração simplificada mostrando como o uso de antibióticos seleciona bactérias resistentes, que passam a sobreviver e se multiplicar

3. Bombas de efluxo: foco central e alvo terapêutico promissor

Entre todos os mecanismos identificados por meio dessas análises computacionais, as bombas de efluxo ganham atenção especial. Elas são proteínas localizadas na membrana da bactéria e funcionam literalmente como bombas, expulsando antibióticos antes que consigam agir. Alguns genes já conhecidos, como *tetA*, *tetB* e *cmx*, estão diretamente associados ao funcionamento dessas bombas em *C. striatum* [14].

Com a ajuda da bioinformática, é possível não apenas identificar essas bombas, mas também prever sua estrutura tridimensional utilizando ferramentas como GalaxyWeb (<https://status.galaxy.seoklab.org/>) e AlphaFold (<https://github.com/google-deepmind/alphafold>). Isso permite visualizar, no computador, como essas proteínas se organizam e onde moléculas poderiam se ligar para bloqueá-las. Em seguida, softwares de simulação, como GROMACS (<https://www.gromacs.org/>), permitem observar o movimento dessas bombas ao longo do tempo, quase como assistir a um vídeo microscópico funcionando em velocidade real. Esses recursos são essenciais para compreender como a proteína interage com compostos e onde um potencial inibidor poderia atuar.

A bioinformática transforma bombas de efluxo, antes apenas um conceito abstrato, em alvos terapêuticos muito mais tangíveis e estratégicos.

4. Peptídeos bioativos e a busca computacional por novos inibidores

Com o avanço das ferramentas digitais, tornou-se possível testar virtualmente moléculas que possam bloquear as bombas de efluxo, restaurando a ação de antibióticos já existentes. Nesse

contexto, peptídeos bioativos se destacam como candidatos promissores. São pequenas moléculas compostas por aminoácidos que podem se ligar a regiões específicas das bombas, impedindo seu funcionamento.

A seleção desses peptídeos pode ser totalmente guiada por bioinformática. Bancos de dados e métodos de mineração de texto, como aqueles utilizados em dbAMP, permitem encontrar peptídeos já descritos na literatura. Em seguida, softwares como MODELLER (<https://salilab.org/modeller/>), GalaxyWeb (<https://status.galaxy.seoklab.org/>) e AlphaFold (<https://github.com/google-deepmind/alphafold>) podem prever a estrutura desses peptídeos, enquanto ferramentas de docking, como HPEPDOCK (http://huanglab.phys.hust.edu.cn/hpepdock/hpepdock_test.tgz), simulam a interação entre peptídeos e bombas de efluxo. Depois, análises de interação molecular por PyMOL (<https://www.pymol.org/>) e PLIP (<https://github.com/pharmai/plip>) revelam quais ligações são mais fortes ou estáveis.

Modelos preditivos com redes neurais convolucionais (CNNs) e pacotes como CD-HIT ajudam a avaliar milhares de peptídeos ao mesmo tempo, descartando os menos promissores e destacando aqueles com maior potencial terapêutico. Por fim, moléculas selecionadas podem ser testadas em simulações mais longas em GROMACS, que estimam sua estabilidade e eficiência como possíveis inibidores.

Essa integração mostra como a bioinformática pode acelerar a busca por novas terapias que combatam a resistência bacteriana sem depender exclusivamente de descobertas tradicionais em laboratório.

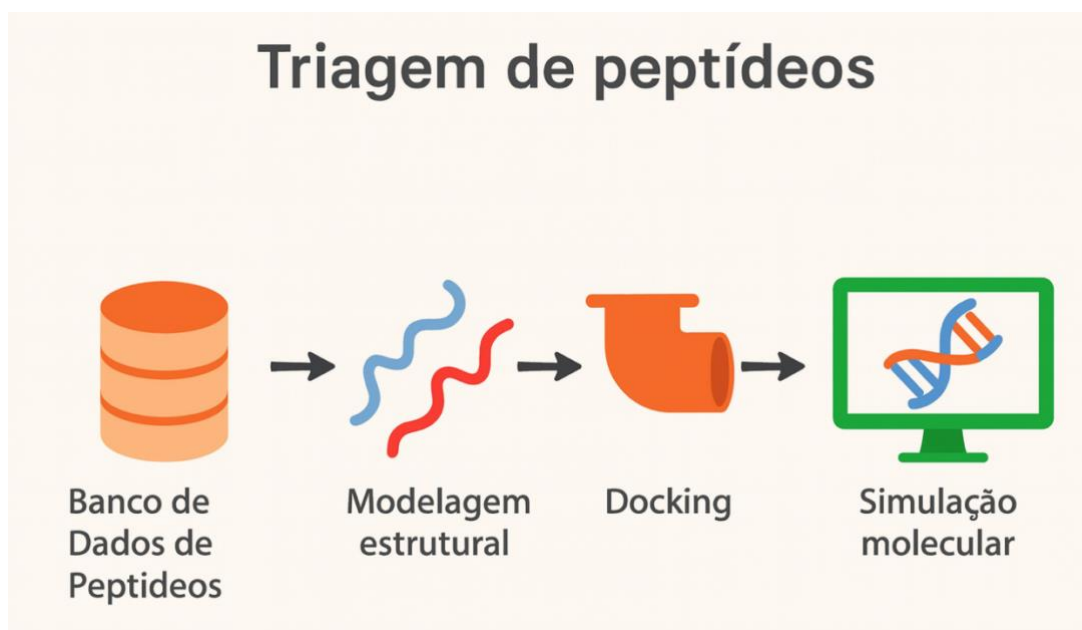


Figura 2. Etapas básicas da triagem computacional de peptídeos candidatos a inibir bombas de efluxo

5. Problema, caminho e solução

A resistência antimicrobiana em *C. striatum* representa um desafio crescente, especialmente devido à ação das bombas de efluxo, que tornam os antibióticos comuns muito menos eficazes. Esse problema tem impactos diretos na saúde pública, aumentando o tempo de internação e reduzindo as opções terapêuticas disponíveis para pacientes hospitalizados.

A bioinformática surge como uma ponte entre a complexidade biológica e a necessidade de respostas rápidas. Ao decodificar o genoma da bactéria e prever como suas proteínas funcionam, ela permite enxergar oportunidades terapêuticas que não seriam visíveis a olho nu. As ferramentas mencionadas, de identificação genética, comparação populacional, modelagem estrutural, docking e simulação molecular, formam um arsenal moderno capaz de transformar milhares de dados biológicos em conhecimento acionável.

A combinação entre o problema clínico, a investigação digital e a triagem de moléculas criam um caminho concreto para o desenvolvimento de inibidores de bombas de efluxo, especialmente peptídeos bioativos. Essa abordagem integrada torna possível enfrentar bactérias multirresistentes de maneira mais rápida, precisa e inovadora, e oferece novas esperanças para o controle de infecções hospitalares.

Referências








- [1] Friedrich A. W. (2019). **Control of hospital acquired infections and antimicrobial resistance in Europe: the way to go.** Wiener medizinische Wochenschrift (1946), 169(Suppl 1), 25–30. <https://doi.org/10.1007/s10354-018-0676-5>
- [2] Richet H. (2012). **Seasonality in Gram-negative and healthcare-associated infections.** Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 18(10), 934–940. <https://doi.org/10.1111/j.1469-0691.2012.03954.x>
- [3] O'Neill, J. (2016). **Tackling drug-resistant infections globally: Final report and recommendations** (84 pp.). Wellcome Trust.
- [4] Antimicrobial Resistance Collaborators (2022). **Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.** Lancet (London, England), 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- [5] World Health Organization. **Antimicrobial resistance. 2021.** Available from: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- [6] Yanai, M et al. **Retrospective evaluation of the clinical characteristics associated with Corynebacterium species bacteremia.** Braz J Infect Dis. 2018 Jan-Feb;22(1):24-29. <https://doi.org/10.1016/j.bjid.2017.12.002>
- [7] Bernard K. **The genus corynebacterium and other medically relevant coryneform-like bacteria.** J Clin Microbiol. 2012 Oct;50(10):3152-8. <https://doi.org/10.1128/JCM.00796-12>
- [8] Milosavljevic, MN et al. **Antimicrobial treatment of Corynebacterium striatum invasive infections: a systematic review.** Rev Inst Med Trop Sao Paulo. 2021 Jun 18;63:e49. <https://doi.org/10.1590/S1678-9946202163049>
- [9] Lee PP, Ferguson DA Jr, Sarubbi FA. **Corynebacterium striatum: an underappreciated community and nosocomial pathogen.** J Infect. 2005 May;50(4):338-43. <https://doi.org/10.1016/j.jinf.2004.05.005>
- [10] Martinez-Martinez, L et al (1997). **Clinical significance of Corynebacterium striatum isolated from human samples.** Clin. Microbiol. Infect. 3, 634–639. <https://doi.org/10.1111/j.1469-0691.1997.tb00470.x>






- [11] Verroken, A., Bauraing, C., Deplano, A., Bogaerts, P., Huang, D., Wauters, G., et al. (2014). **Epidemiological investigation of a nosocomial outbreak of multidrug-resistant *Corynebacterium striatum* at one Belgian university hospital.** *Clin. Microbiol. Infect.* 20, 44–50. [https://doi: 10.1111/1469-0691.12197](https://doi.org/10.1111/1469-0691.12197)
- [12] Chen, FL et al. (2012). ***Corynebacterium striatum* bacteremia associated with central venous catheter infection.** *J. Microbiol. Immunol. Infect.* 45, 255–258. [https://doi: 10.1016/j.jmii.2011.09.016](https://doi.org/10.1016/j.jmii.2011.09.016)
- [13] Zhang, H et al. (2023). **Targeted antibiotics for lower respiratory tract infection with *corynebacterium striatum*.** *Infect. Drug Resist.* 16, 2019–2028. [https://doi: 10.2147/IDR.S404855](https://doi.org/10.2147/IDR.S404855)
- [14] Ramos, JN et al. **Genome sequence of a multidrug-resistant *Corynebacterium striatum* isolated from bloodstream infection from a nosocomial outbreak in Rio de Janeiro, Brazil.** *Mem Inst Oswaldo Cruz.* 2018 Jul 10;113(9):e180051. [https://doi: 10.1590/0074-02760180051](https://doi.org/10.1590/0074-02760180051)

9

Os genes das estruturas reprodutivas de mamíferos: quão ancestrais são suas origens?

Autores 9.1

Kadu Penuela S. Estevam , Elis de Aguiar Jacquemin , Emily Nicolau Koubayache , Gabriela Oliveira Tonidandel , Italo Candelária , Ryan Fernandes Vieira de Souza , J. Miguel Ortega 

Revisão: Ariany Rosa Gonçalves , Bibiana Fam , Diego Mariano , Savio Costa , Tatiane Senna Bialves 








Cite este artigo 9.1

Estevam K. P. S. *et al.* **Os genes das estruturas reprodutivas de mamíferos: quão ancestrais são suas origens?**. *BIOINFO*. Vol. 6. p.09 (2026). doi: 10.51780/bioinfo-06-09

Saiba mais 9.1

Este artigo está disponível em <https://bioinfo.com.br/os-genes-das-estruturas-reprodutivas-de-mamiferos-quao-ancestrais-sao-suas-origens/>

Os genes das estruturas reprodutivas de mamíferos: quão ancestrais são suas origens?

Kadu Penuela S. Estevam ^{1,2,*} , Elis de Aguiar Jacquemin ² , Emily Nicolau Koubayache ² ,
 Gabriela Oliveira Tonidandel ² , Italo Candelária ² , Ryan Fernandes Vieira de Souza ^{1,2} ,
 J. Miguel Ortega ² 

¹ Programa Interunidades de Pós-Graduação em Bioinformática, UFMG

² Laboratório de Biodados - Depto. Bioquímica e Imunologia, ICB, UFMG

* kadupse@ufmg.br

Resumo. *Estruturas biológicas complexas, como os marcos reprodutivos de mamíferos, são tradicionalmente estudadas através de sua morfologia. No entanto, a origem morfológica dessas estruturas reflete necessariamente a origem de seus genes constituintes? Para responder a essa questão, validamos um pipeline computacional integrado à análise manual para reconstruir a ancestralidade gênica de novidades evolutivas. Utilizamos as ferramentas in house GO-Genesis e TaxOnTree para rastrear a história profunda dos genes associados a termos do Gene Ontology (GO). A aplicação desta metodologia revela padrões evolutivos distintos: enquanto certas estruturas recrutam maquinários genéticos muito antigos (anteriores à própria origem dos mamíferos), outras dependem de inovações gênicas recentes. Ao dissociar a idade do gene da idade da estrutura, nosso trabalho demonstra como a evolução atua via cooptação molecular. O Laboratório de Biodados investiga esta ancestralidade, elaborando vias biológicas compreensivas que elucidam as interações entre os genes-chave de cada processo, oferecendo uma nova perspectiva sobre a evo-devo de mamíferos.*

Palavras-chave: *evolução de genes; filogenia; ontologia; mamíferos; evo-devo.*

1. Introdução

A biologia evolutiva do desenvolvimento (evo-devo) busca decifrar os mecanismos genéticos e moleculares que governam a origem de novas formas ao longo da história da vida. Um dos desafios centrais desta área é entender como inovações morfológicas complexas, a exemplo dos marcos reprodutivos de mamíferos, evoluíram.

Essas estruturas representam adaptações cruciais que caracterizam suas linhagens, e podemos investigar a origem dos genes que as orquestram. A suposição clássica de que a novidade morfológica e a novidade genética compartilham um ponto de origem coincidente tem sido desafiada [1]. A hipótese central deste trabalho propõe que muitas inovações dos mamíferos não resultam primariamente do surgimento de novos genes, em vez disso, decorrem da cooptação: o recrutamento de genes ancestrais, antes associados a outras funções, para exercer novos papéis. Esse processo é complementado pela incorporação de genes de origem mais recente.

No entanto, rastrear a origem filogenética de redes gênicas inteiras, especialmente aquelas associadas a funções biológicas específicas, apresenta um desafio metodológico significativo. O volume de dados genômicos disponíveis exige abordagens computacionais, mas a precisão biológica requer curadoria humana. O Laboratório de Biodados da UFMG desenvolveu um processo bioinformático robusto para reconstruir a origem evolutiva de genes associados a novidades morfológicas. A abordagem integra as ferramentas GO-Genesis e TaxOnTree, desenvolvidas internamente, com análises filogenéticas e de interação em bancos públicos. Deste modo, investigamos a origem dos genes implicados na reprodução de mamíferos. Nosso objetivo é duplo: (1) testar a hipótese de que as origens genéticas das estruturas reprodutivas de mamíferos são muito anteriores ao clado Theria; e (2) validar esta metodologia como um recurso computacional acessível e eficaz.

2. Metodologia

O ponto de partida de nossa metodologia é a seleção de processos biológicos de interesse no *Gene Ontology* (GO) [2], um sistema que fornece um vocabulário controlado e hierárquico para descrever funções gênicas. Um termo GO funciona como um marcador digital associado a diversos genes que participam de um mesmo processo, localização celular ou função. O termo "desenvolvimento da placenta" (GO:0001890), por exemplo, está associado a genes que a literatura científica vincula a esta via biológica. A partir dessa compilação, os termos de interesse são submetidos a duas etapas analíticas distintas:

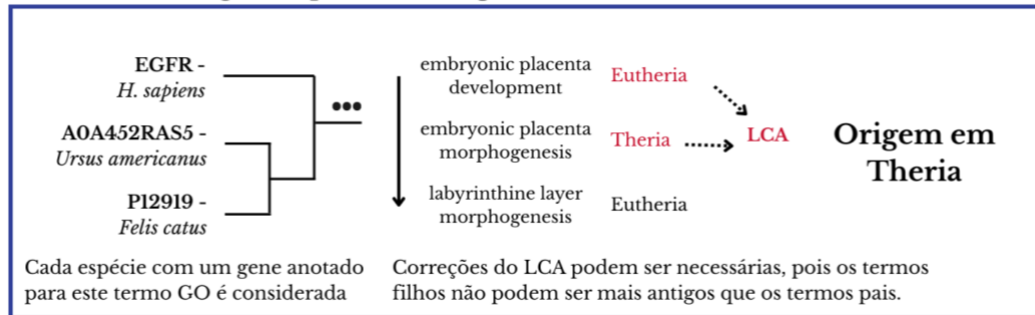
A primeira etapa do *pipeline* utiliza o GO-Genesis [3] para a análise sistemática destes conjuntos gênicos. A ferramenta rastreia a origem evolutiva do processo biológico ao identificar o "Último Ancestral Comum" (*Last Common Ancestor*, LCA) para o conjunto de proteínas anotadas [4]. Isso fornece uma estimativa inicial da origem do processo baseada na distribuição taxonômica atual de todos os genes implicados, corrigindo eventuais incongruências observadas na ontologia, como genes anotados erroneamente em termos "filhos" cuja ancestralidade inferida é anterior à dos "pais" [5].

A segunda etapa foca em determinar a origem individual de cada gene por reconstrução filogenética. Para isso, utilizamos o TaxOnTree [6]. Esta ferramenta automatiza uma sequência de análises: (1) realiza buscas por homólogos de cada gene no banco de dados *UniProt Reference Proteomes* [7] usando BLASTp, assegurando uma amostragem taxonômica ampla; (2) gera alinhamentos múltiplos das sequências obtidas com o algoritmo MUSCLE, estabelecendo a correspondência posicional dos resíduos; e (3) constrói as árvores filogenéticas (hipóteses evolutivas) utilizando o FastTree para inferir topologias por Máxima Verossimilhança, método adequado para grandes conjuntos de dados.

Crucialmente, o TaxOnTree anota as árvores geradas com as informações taxonômicas do NCBI e da Taxallnomy, permitindo a análise manual refinada (*e.g.*, em FigTree) para determinar o LCA de cada gene. Finalmente, para contextualizar as redes genéticas, as proteínas são analisadas em bancos de interação proteína-proteína, como o STRING-DB. Essa etapa permite o mapeamento de *clusters* funcionais, ampliando a identificação de proteínas

associadas ao processo biológico de interesse que, porventura, ainda não tenham sido capturadas pela curadoria do *Gene Ontology*.

GO-Genesis Origem do processo biológico



TaxOnTree Origem dos genes individuais implicados neste processo

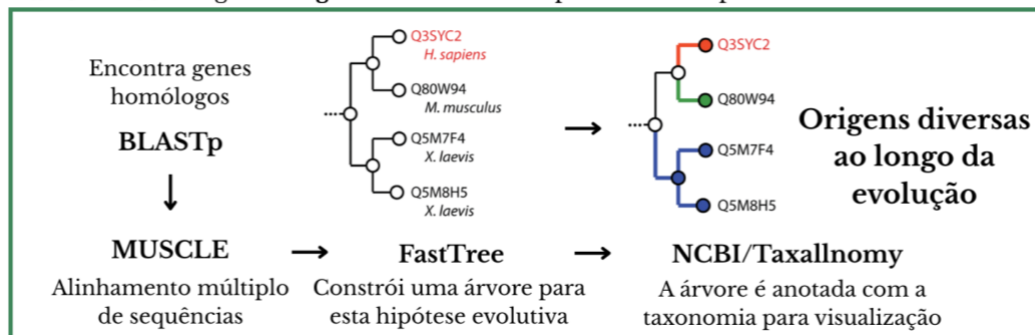


Figura 1. Representação esquemática do pipeline baseado em GO-Genesis e TaxOnTree. O uso de ambas as aplicações permite inferir não somente a origem evolutiva de um processo ou estrutura completa, mas também de cada um de seus genes constituintes.

A aplicação deste *pipeline* integrado nos permite dissecar a história evolutiva de inovações morfológicas, retratando a origem da estrutura (o processo GO) e de seus componentes genéticos individuais.

3. Discussão e perspectivas

A integração dos dados obtidos via GO-Genesis, QuickGO [8] e STRING-DB [9], com a curadoria manual no TaxOnTree, revelou que a origem de uma estrutura reprodutiva raramente corresponde à idade de seus componentes moleculares. Ao analisarmos os perfis filogenéticos, identificamos um claro mosaico evolutivo, e semelhanças e diferenças moleculares que não necessariamente condizem com a intuição inicial esperada ao olharmos apenas para os nós de especiação.

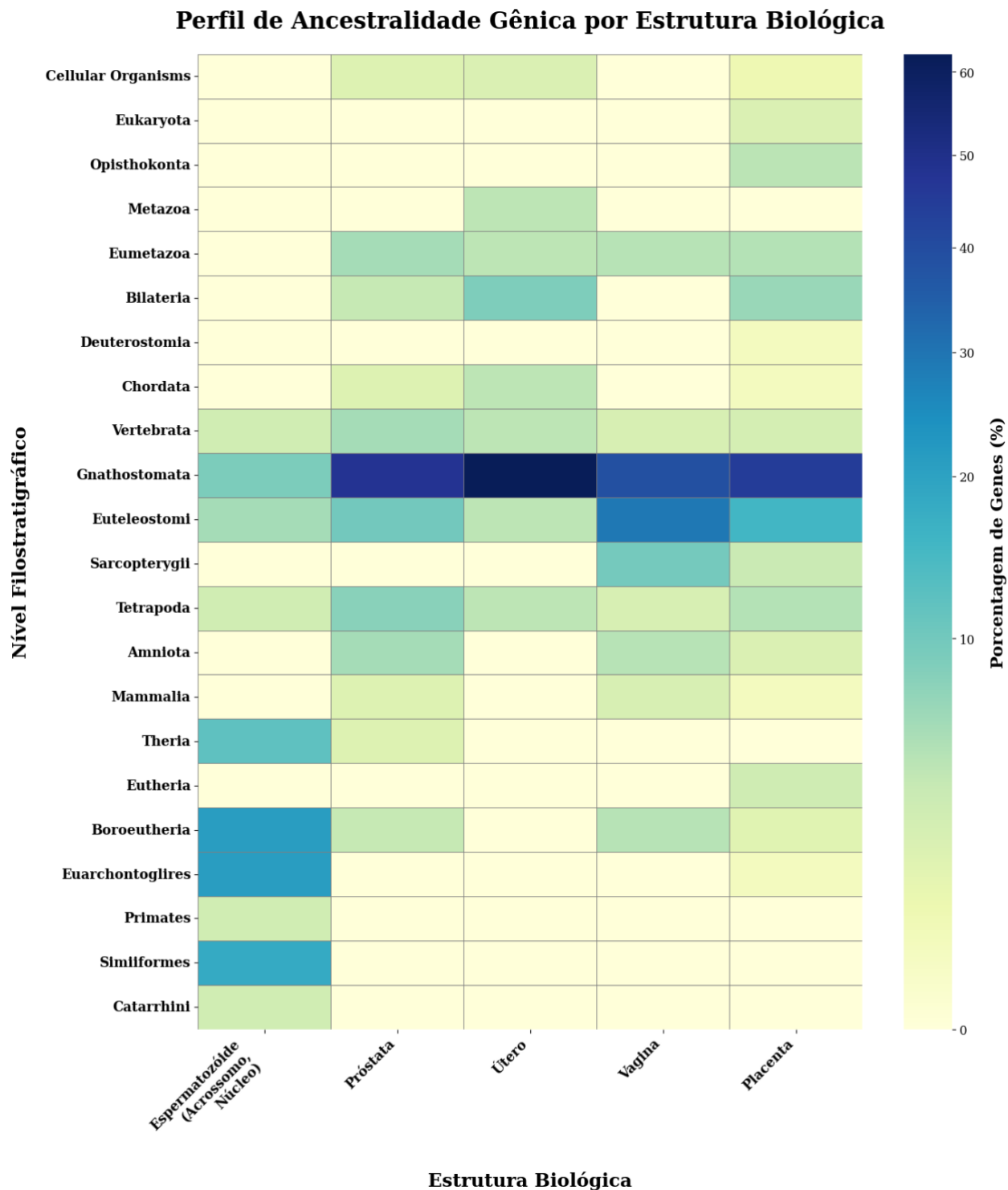


Figura 2. Representação em *heat map* do percentual de genes humanos implicados nas estruturas do útero, placenta, próstata, espermatozoide, e vagina, compartilhados com cada ancestral. Nota-se a grande disparidade temporal entre o aparecimento da estrutura e a origem de seus genes componentes, além da grande representação do nó Gnathostomata, nosso ancestral com tubarões. Intensidade da coloração ajustada com $\gamma = 0.5$ para visualização.

Nossos dados indicam uma predominância significativa de genes ancestrais na composição de estruturas que só surgiriam milhões de anos depois, como é exemplificado pelo útero: a análise revela que mais de 80% dos genes anotados para o desenvolvimento desta

estrutura já estavam presentes em Gnathostomata (nosso ancestral comum com tubarões). Isso sugere que a evolução da viviparidade mamífera não exigiu a invenção de centenas de novos genes, mas sim a reorganização de um *toolkit* ancestral. Vias de sinalização e receptores nucleares, estabelecidos antes mesmo da conquista do ambiente terrestre, foram cooptados para orquestrar a gestação. A curadoria manual permitiu identificar padrões moleculares cruciais. Observamos, por exemplo, que a arquitetura básica desses órgãos é sustentada por receptores de estrogênio conservados desde os peixes ósseos (Euteleostomi). A inovação real, muitas vezes, reside em adições pontuais, mas decisivas, como o recrutamento de genes da família das Citoquinas (ex: CSF2) na placenta, essenciais para a imunomodulação materna, uma função que não era necessária em nossos ancestrais ovíparos.

Metodologicamente, a validação manual se mostrou indispensável. Ferramentas automáticas frequentemente falham em distinguir a ausência real de um gene de uma falha de anotação em genomas não-modelo. Nossa abordagem com o TaxOnTree permitiu corrigir a ancestralidade de diversos alvos, identificando parálogos (genes duplicados) que assumiram funções reprodutivas novas não presentes em suas cópias ancestrais. Isso reforça que, para decifrar a evo-devo de mamíferos, é preciso olhar além da simples presença ou ausência do gene, focando na história de suas duplicações e interações. Mais criticamente, é possível desafiar a dicotomia entre convergência e homologia profunda, como no caso da conservação observada de um conjunto de genes na placenta análoga de tubarões Carcharhiniformes.

O laboratório agora explora essas trajetórias comparativamente, com o objetivo de construir vias metabólicas compreensivas e investigar as hipóteses inferidas pelas proximidades moleculares observadas e eventos como a perda secundária de grupos de genes em linhagens próximas, como as aves. Elucidar estas vias biológicas demonstra como o recrutamento de genes antigos, aliados à origem de novos genes, orquestra a novidade morfológica, respondendo à questão: os genes das estruturas reprodutivas de mamíferos, quão ancestrais são suas origens?

Agradecimentos. Os autores agradecem a Carlos A. X. Gonçalves (SENAI) e a Tetsu Sakamoto (UFRN) pelo desenvolvimento dos pipelines empregados neste projeto e às agências de fomento à pesquisa: CAPES, CNPq e Fapemig.

Disponibilidade dos dados. As ferramentas TaxOnTree e GO-Genesis podem ser acessadas em versão *web* disponível em: <http://bioinfo.icb.ufmg.br/>.

4. Referências










- [1] WAGNER, G. P. Homology, Genes, and Evolutionary Innovation. Princeton: Princeton University Press, 2014.
- [2] ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. Nature Genetics, New York, v. 25, n. 1, p. 25-29, 2000. DOI: 10.1038/75556.
- [3] GONÇALVES, C. A. X.; ORTEGA, J. M. GO-genesis: a web tool to trace back the origin of biological processes and molecular functions. Disponível em: <http://biodados.icb.ufmg.br/go-genesis/>. Acesso em: 20 nov. 2025
- [4] HUNTLEY, R. P. et al. The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Research, Oxford, v. 43, n. Database issue, p. D1057–D1063, 2015.

- [5] ORTEGA, J. M.; GONÇALVES, C. A. X. Revealing the clade of origin of Gene Ontology terms related to developmental processes. *Trends in Developmental Biology*, v. 12, p. 67, 2019.
- [6] SAKAMOTO, T.; ORTEGA, J. M. TaxOnTree: a tool that generates trees annotated with taxonomic information. *bioRxiv*, [S. l.], 24 dez. 2020. DOI: 10.1101/2020.12.24.424364.
- [7] THE UNIPROT CONSORTIUM. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, Oxford, v. 53, n. D1, p. D609–D617, 2025.
- [8] BINNS, D. et al. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, Oxford, v. 25, n. 22, p. 3045–3046, 2009.
- [9] SZKLARCZYK, D. et al. STRING v12: protein–protein interaction networks with increased coverage and new features. *Nucleic Acids Research*, Oxford, v. 51, n. D1, p. D638–D646, 2023.

10

Estudo integrativo genômico e de virulência de *Corynebacterium auriscanis*: elucidando a patogenicidade

Autores 10.1

Evandro Bento Rodrigues , Max Roberto Batista Araújo , Luan Freitas Lana , Diogo Luiz de Carvalho Castro , Genrikh Ashniev Alfredovich , Rafael Augusto Ribeiro , Vinícius Dias do Carmo Costa , Maria Letícia Carneiro Rodrigues , Vasco Azevedo 

Revisão: Diego Mariano , Lucianna Helene Santos , Sheila Cruz Araújo 










Cite este artigo 10.1

Rodrigues E. B. *et al.* Estudo integrativo genômico e de virulência de *Corynebacterium auriscanis*: elucidando a patogenicidade. *BIOINFO.* Vol. 6. p.10 (2026). doi: 10.51780/bioinfo-06-10

Saiba mais 10.1

Este artigo está disponível em <https://bioinfo.com.br/estudo-integrativo-genomico-e-de-virulencia-de-corynebacterium-auriscanis-elucidando-a-patogenicidade/>

Estudo integrativo genômico e de virulência de *Corynebacterium auriscanis*: elucidando a patogenicidade

Evandro Bento Rodrigues^{1,2†} , Max Roberto Batista Araújo^{1,2†} , Luan Freitas Lana¹ ,
 Diogo Luiz de Carvalho Castro² , Genrikh Ashniev Alfredovich³ , Rafael Augusto
 Ribeiro¹ , Vinícius Dias do Carmo Costa¹ , Maria Letícia Carneiro Rodrigues¹ , Vasco
 Azevedo^{1*} 

†Estes autores contribuíram igualmente para este trabalho.

¹ Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

² Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil

³ Instituto de Problemas de Transmissão de Informação da Academia Russa de Ciências, Moscou, Rússia

* vascoariston@gmail.com

Resumo. *Corynebacterium auriscanis* é uma espécie pouco estudada, isolada principalmente de cães e gatos, mas ocasionalmente associada a infecções humanas. Seu papel como potencial patógeno zoonótico, entretanto, permanece pouco compreendido. A identificação laboratorial ainda depende de técnicas moleculares como a espectrometria de massas MALDI-TOF, que são precisas, mas pouco acessíveis em muitas rotinas diagnósticas. Nesse contexto, ferramentas de bioinformática representam um caminho para ampliar o conhecimento sobre diversidade genômica, fatores de virulência e mecanismos de resistência. Este artigo discute como análises genômicas, predição de genes e identificação de sistemas CRISPR-Cas, dentre outros, podem contribuir para compreender o potencial patogênico de *C. auriscanis*, além de destacar, de forma concisa, como modelos animais como *Caenorhabditis elegans* podem complementar investigações funcionais. O foco é apresentar como a integração entre sequenciamento, bioinformática e ensaios biológicos pode apoiar ações de vigilância e prevenção, especialmente diante da importância da interface animal-humano na disseminação de patógenos.

Palavras-chave: *Corynebacterium auriscanis*, genômica, virulência, resistência antimicrobiana, CRISPR-Cas, zoonose

1. Introdução

O gênero *Corynebacterium* inclui espécies amplamente conhecidas pela relevância clínica, como *C. diphtheriae* [1] ainda isolado e endêmico no Brasil [2-5], mas também um grande conjunto de espécies não diftéricas capazes de causar infecções oportunistas em humanos e animais. Entre elas, *Corynebacterium auriscanis* permanece uma das menos estudadas [6,7]. Descrita originalmente a partir de isolados de cães, a espécie já foi associada a otites, infecções cutâneas e casos clínicos esporádicos em humanos [8]. Apesar disso, pouco se sabe sobre sua diversidade, fatores de virulência e mecanismos de adaptação ao hospedeiro.

Temos clara uma questão importante: como compreender o risco zoonótico e o potencial patogênico de uma espécie tão pouco caracterizada? A falta de dados genômicos, a carência de estudos funcionais e a dificuldade de identificação laboratorial dificultam a vigilância e retardam o entendimento de possíveis rotas de transmissão animal-humano.

Entra, assim, em cena a bioinformática ao oferecer meios para preencher lacunas essenciais, permitindo interpretar rapidamente grandes volumes de informação genética e conectá-los a dados clínicos e experimentais.

2. Discussão

O primeiro desafio enfrentado no estudo de *C. auriscanis* é a identificação laboratorial. Embora métodos fenotípicos ainda sejam comuns, a espectrometria de massas MALDI-TOF tornou-se a ferramenta de referência para diferenciar espécies próximas com maior precisão. A técnica analisa padrões proteicos específicos e possibilita identificar bactérias de forma rápida e confiável. No caso de espécies pouco estudadas como *C. auriscanis*, essa precisão é crucial para evitar classificações incorretas e garantir que isolados sejam corretamente incluídos em estudos epidemiológicos.

Uma vez identificada a espécie, o sequenciamento genômico amplia enormemente as possibilidades de investigação. A partir dele, ferramentas de bioinformática permitem responder questões fundamentais, tais como: Quais genes de virulência estão presentes? Existem determinantes de resistência que podem impactar tratamentos? Como a espécie se relaciona evolutivamente com outras do gênero? Há sistemas CRISPR-Cas que indiquem exposição prévia a fagos e plasmídios, sendo um sistema imune bacteriano?

As ferramentas de bioinformática desempenham um papel central na interpretação dos genomas de espécies bacterianas. O processo começa pela montagem genômica de novo, realizada por softwares como Unicycler (<https://github.com/rrwick/Unicycler>), que reconstrói o genoma a partir das leituras de sequenciamento. Após a montagem, a qualidade do genoma precisa ser avaliada: CheckM2 (<https://github.com/chklovski/CheckM2>) estima completude e contaminação; GUNC (https://github.com/grp-bork/gunc_workflow) identifica possíveis quimeras; Barnapp (<https://github.com/tseemann/barnapp>) verifica a presença de genes ribossomais; e QUAST (<https://github.com/ablab/quast>) resume estatísticas da montagem. Cada uma dessas ferramentas responde a perguntas básicas, mas fundamentais: o genoma está completo? confiável? livre de artefatos? Pré-requisitos para qualquer análise posterior.

A partir desse genoma validado, outras ferramentas ajudam a definir a identidade e o posicionamento evolutivo da espécie. GTDB-Tk (<https://github.com/GenomeTaxonomy/Ecogenomics/GTDBTk>) e TYGS (<https://tygs.dsmz.de/>) permitem classificar o isolado com precisão comparando-o a genomas de referência. A anotação funcional, realizada pelo PGAP (https://www.ncbi.nlm.nih.gov/refseq/annotation_prok/), traduz as sequências brutas em genes com funções putativas. Esses passos combinados respondem a uma questão fundamental no estudo: quem é esse organismo e como ele se relaciona com seus parentes próximos?

Ferramentas de análise comparativa aprofundam esse panorama. A avaliação de similaridade genômica por PyANI (<https://github.com/widdowquinn/pyani>) e a confirmação da similaridade global por dDDH via TYGS ajudam a definir fronteiras entre espécies e linhagens. Já bancos especializados como a VFDB (<https://www.mgc.ac.cn/VFs/main.htm>) permitem identificar genes associados a virulência, enquanto a PanViTa (<https://github.com/dlnrodrigues/panvita>) e o CARD (<https://card.mcmaster.ca/>) ajudam a localizar determinantes de resistência antimicrobiana. Essas análises revelam o “arsenal molecular” da bactéria, informações essenciais para avaliar risco clínico, impacto terapêutico e potencial zoonótico.

Outros elementos importantes para compreender evolução e adaptação também podem ser identificados. PlasmidFinder (<https://github.com/genomicepidemiology/plasmidfinder>) e MOB-suite (<https://github.com/phac-nml/mob-suite>) caracterizam plasmídios que podem carrear genes de resistência. IntegronFinder (https://github.com/gem-pasteur/Integron_Finder) identifica integrons, estruturas associadas à captura de genes adaptativos. Já sistemas CRISPR-Cas, indicadores de interações com fagos e plasmídios, são detectados por CRISPRCasFinder (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>) e comparados com possíveis alvos usando CRISPRTarget (https://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html). Essas informações ajudam a entender como o microrganismo se defende de elementos genéticos invasores e como isso molda sua evolução.

A análise de elementos genéticos móveis amplia ainda mais essa visão. ISEScan (<https://github.com/xiezhq/ISEScan>) detecta sequências de inserção que podem promover rearranjos genômicos, enquanto GIPSy2 (<https://zenodo.org/records/14969252>) identifica possíveis ilhas genômicas, incluindo regiões associadas a virulência, resistência ou adaptação ambiental. A anotação funcional dessas ilhas com EggNOG-mapper (<https://eggnog-mapper.embl.de/>) permite classificar funções e relacioná-las a fenótipo e patogenicidade.

Para visualizar similaridades e diferenças entre genomas, o BRIG (<https://beatsonlab.com/software/brig/>) gera mapas circulares que destacam regiões conservadas, genes ausentes, ilhas genômicas e determinantes de resistência ou virulência. É uma forma intuitiva de comunicar dados complexos, essencial em vigilância molecular.

A análise comparativa entre diversos genomas é reforçada pelo uso de Panaroo (<https://github.com/gtonkinhill/panaroo>), que identifica genes essenciais (core), acessórios e exclusivos, categorizando-os segundo bases como COG [9] e KEGG [10]. Já a conservação da ordem dos genes, a sintenia, é avaliada por Mauve (<https://darlinglab.org/mauve/mauve.html>), revelando eventos de recombinação, inversões e blocos genômicos compartilhados.

Por fim, análises evolutivas aprofundam a relação entre linhagens. OrthoFinder (<https://github.com/davidemms/OrthoFinder>) identifica genes ortólogos e reconstrói filogenias com IQ-TREE (<https://www.iqtree.org/>), enquanto sinais de seleção positiva são detectados por POTION (<http://www.lmb.cnptia.embrapa.br/share/POTION/>). Essas ferramentas revelam quais genes evoluem rapidamente, indicando adaptações que podem influenciar virulência, sobrevivência ou interação com hospedeiros.

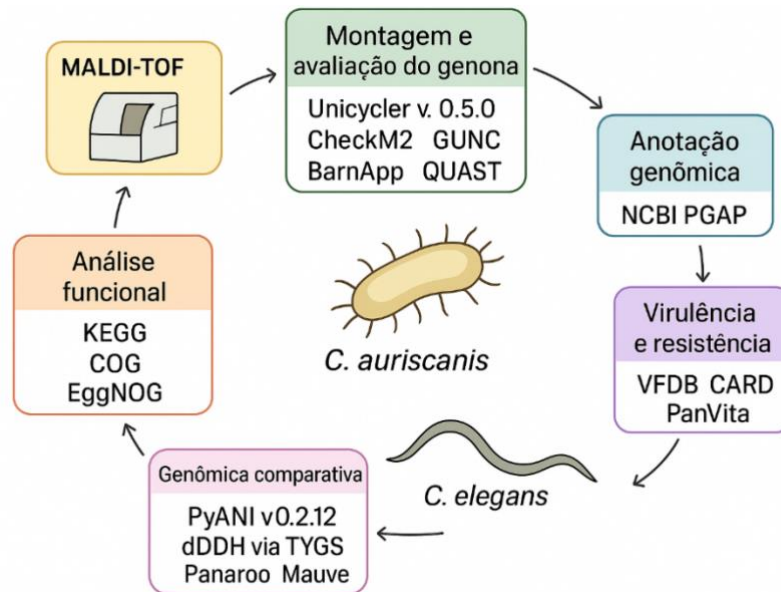


Figura 1. Visão geral das abordagens utilizadas para caracterização de *C. auriscanis*. A identificação por MALDI-TOF MS, seguida pela montagem e avaliação do genoma (Unicycler, CheckM2, GUNC, BarnApp, QUAST). A anotação com NCBI PGAP, e genes de virulência e resistência identificados com VFDB, CARD e PanViTa. A genômica comparativa envolveu PyANI, dDDH, Panaroo e Mauve. A análise funcional com KEGG, COG e EggNOG. Ensaios de virulência em *C. elegans*.

3. Conclusão

Ferramentas de bioinformática têm papel central na compreensão de espécies pouco estudadas como *C. auriscanis*. Elas permitem identificar genes-chave, explorar a diversidade genômica e propor hipóteses sobre virulência e resistência, informações indispensáveis para vigilância molecular, saúde pública e medicina veterinária. Aliadas ao MALDI-TOF MS e a modelos experimentais como *C. elegans*, essas abordagens constituem um arcabouço capaz de orientar ações de prevenção, diagnóstico e monitoramento de potenciais zoonoses. À medida que novos dados são incorporados, análises integrativas tornam-se essenciais para antecipar riscos e apoiar estratégias de controle na interface animal-humano.

4. Referências

- [1] Bernard K. The genus corynebacterium and other medically relevant coryneform-like bacteria. J Clin Microbiol. 2012 Oct;50(10):3152-8. doi: 10.1128/JCM.00796-12
- [2] Batista Araújo MR, Bernardes Sousa MÂ, Seabra LF, Caldeira LA, Faria CD, Bokermann S, Sant'Anna LO, Dos Santos LS, Mattos-Guaraldi AL. Cutaneous infection by non-diphtheria-toxin producing and penicillin-resistant *Corynebacterium diphtheriae* strain in a patient with diabetes mellitus. Access Microbiol. 2021 Nov 30;3(11):000284. doi: 10.1099/acmi.0.0002844.
- [3] Araújo MRB, Ramos JN, de Oliveira Sant'Anna L, Bokermann S, Santos MBN, Mattos-Guaraldi AL, Azevedo V, Prates FD, Rodrigues DLN, Aburjaile FF, Sacchi CT, Campos KR, Alvim LB, Vieira VV, Camargo CH, Dos Santos LS. Phenotypic and molecular characterization and complete genome sequence of a Corynebacterium

- diphtheriae strain isolated from cutaneous infection in an immunized individual. *Braz J Microbiol.* 2023 Sep;54(3):1325-1334. doi: 10.1007/s42770-023-01086-z
- [4] Ramos JN, Araújo MRB, Sant'Anna LO, Bokermann S, Camargo CH, Prates FD, Sacchi CT, Vieira VV, Campos KR, Santos MBN, Viana MVC, Azevedo V, Aburjaile FF, de Mattos-Guaraldi AL, Dos Santos LS. Molecular characterization and whole-genome sequencing of *Corynebacterium diphtheriae* causing skin lesion. *Eur J Clin Microbiol Infect Dis.* 2024 Jan;43(1):203-208. doi: 10.1007/s10096-023-04706-6
- [5] Araújo MRB, Prates FD, Ramos JN, Sousa EG, Bokermann S, Sacchi CT, de Mattos-Guaraldi AL, Campos KR, Sousa MÂB, Vieira VV, Santos MBN, Camargo CH, de Oliveira Sant'Anna L, Dos Santos LS, Azevedo V. Infection by a multidrug-resistant *Corynebacterium diphtheriae* strain: prediction of virulence factors, CRISPR-Cas system analysis, and structural implications of mutations conferring rifampin resistance. *Funct Integr Genomics.* 2024 Aug 28;24(5):145. doi: 10.1007/s10142-024-01434-8
- [6] Bygott JM, Malnick H, Shah JJ, Chattaway MA, Karas JA. First clinical case of *Corynebacterium auriscanis* isolated from localized dog bite infection. *J Med Microbiol.* 2008 Jul;57(Pt 7):899-900. doi: 10.1099/jmm.0.47780-0
- [7] Vinhal ALO, de Araújo MRB, Rodrigues EB, Castro DLC, Pereira CR, Custódio DAC, Dorneles EMS, Aburjaile FF, Brenig B, Azevedo V, Viana MVC. First comparative genomics analysis of *Corynebacterium auriscanis*. *Mem Inst Oswaldo Cruz.* 2024 Oct 25;119:e240156. doi: 10.1590/0074-02760240156
- [8] Collins MD, Hoyles L, Lawson PA, Falsen E, Robson RL, Foster G. Phenotypic and phylogenetic characterization of a new *Corynebacterium* species from dogs: description of *Corynebacterium auriscanis* sp. nov. *J Clin Microbiol.* 1999 Nov;37(11):3443-7. doi: 10.1128/JCM.37.11.3443-3447.1999
- [9] Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D274-D281. doi: 10.1093/nar/gkaa1018
- [10] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D353-D361. doi: 10.1093/nar/gkw1092

11

Genômica comparativa de *Ornithorhynchus anatinus* revela genes mais próximos de tartarugas do que de outros mamíferos

Autores 11.1

Ryan Fernandes Vieira de Souza , José Miguel Ortega 

Revisão: Alessandra Lima , Diego Mariano , Lucianna Helene Santos , Thiago Jesus Sousa 



Cite este artigo 11.1

Souza, R. F. V. de; Ortega, J. M. **Genômica comparativa de *Ornithorhynchus anatinus* revela genes mais próximos de tartarugas do que de outros mamíferos**. *BIOINFO*. Vol. 6. p.11 (2026). doi: 10.51780/bioinfo-06-11

Saiba mais 11.1

Este artigo está disponível em <https://bioinfo.com.br/genomica-comparativa-de-ornithorhynchus-anatinus-revela-genes-mais-proximos-de-tartarugas-do-que-de-outros-mamiferos/>

Genômica comparativa de *Ornithorhynchus anatinus* revela genes mais próximos de tartarugas do que de outros mamíferos

Ryan Fernandes Vieira de Souza¹ , José Miguel Ortega^{1,2*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

ryanfvs@ufmg.br

Resumo. *O Ornithorhynchus anatinus, popularmente conhecido como ornitorrinco, é um mosaico de características por conter o bico semelhante ao de pato, a postura de ovos e a produção de leite sem a saída por mamilos. Ao comparar o seu proteoma com outros organismos, é possível investigar pistas da origem de seus genes e a proximidade deles de outros grupos. Dessa forma, o presente estudo tem como objetivo explorar a genômica comparativa do ornitorrinco, através da obtenção e análise de proteomas completos de organismos eucariontes. Para isso, foram utilizados 2.765 proteomas de referência obtidos do UniProt, incluindo o do ornitorrinco. O conjunto foi utilizado na construção de uma base de dados no software DIAMOND, e o proteoma do ornitorrinco foi alinhado utilizando a função BLASTp, com aplicação de diferentes faixas do parâmetro –top para análise do LCA dos genes mais similares aos dele. Os resultados indicaram que maior parte dos genes apresentaram como LCA a própria espécie, refletindo o efeito do self-score elevado. À medida que o intervalo top aumentou, observou-se que a incorporação de genes com ancestrais mais profundos, como a linhagem Gnathostomata. Além disso, a classe que reúne as tartarugas apresenta-se como a mais ocorrente para aqueles genes que tiveram hits contendo apenas o ornitorrinco como representante dos mamíferos. As análises de enriquecimento funcional desse subconjunto, bem como suas reconstruções filogenéticas, identificaram termos associados à formação da casca do ovo e confirmaram que as tartarugas são mais próximas do ornitorrinco para esse subconjunto, refletindo um caso potencial de homologia profunda e evidenciando a natureza mosaica do genoma dos monotremados.*

Palavras-chave: LCA; genômica; evolução.

1. Introdução

O Ornithorhynchus anatinus (Figura 1), junto com a equidna (*Tachyglossus aculeatus*), são os únicos mamíferos que põem ovos (ovíparos) e são membros da ordem Monotremata [1]. Com uma peculiar mistura de características, o ornitorrinco tem uma pelagem adaptada ao habitat aquático; as fêmeas têm a capacidade de amamentar e botar ovos; e os machos têm um veneno em seu bico semelhante ao dos répteis [2].



Figura 1. Ilustração de um de ornitorrinco (*Ornithorhynchus anatinus*). **Fonte:** Wikimedia Commons (CC BY-SA 4.0).

Os monotremados por ocuparem uma posição basal entre os Amniotas não mamíferos e os Therianos (marsupiais e placentários), constituem um grupo-chave para compreender a transição evolutiva até os mamíferos modernos [3]. Nesse âmbito, a genômica comparativa tem um importante papel em estudar os genes desses organismos buscando compreender a evolução de suas características compartilhadas por ancestralidade comum [4]. Entre os conceitos fundamentais da área está o de homologia profunda (*deep homology*), que é um fenômeno no qual genes, redes regulatórias ou módulos de desenvolvimento extremamente conservados são reutilizados ao longo da evolução em linhagens muito distantes. Assim, estruturas ou funções aparentemente inovadoras podem, na realidade, emergir de arcaicos genéticos mais ancestral que se imagina [5]. Em organismos como o ornitorrinco, que retém tanto características típicas de mamíferos quanto traços mais próximos de aves e répteis, a homologia profunda fornece uma possível explicação quando genes específicos apresentam maior similaridade com linhagens distantes ao invés de outros organismos da mesma linhagem, nesse caso os mamíferos. Apesar de sua forte importância, estudos de genômica comparativa enfrentam limitações que podem influenciar a interpretação dos resultados. Uma delas é a ausência de genomas completos e de alta qualidade para uma boa parte da biodiversidade eucariótica, o que faz com que se crie lacunas filogenômicas que distorcem análises e inferências do último ancestral comum (LCA - *Least Common Ancestor*). Por exemplo, até 2011 estimava-se que entre 1,5 e 1,9 milhão de espécies eucariontes haviam sido formalmente descritas [6]. No entanto, até o final de 2024, apenas cerca de 3.039 a 4.000 espécies tiveram seus genomas sequenciados no âmbito do *Earth BioGenome Project* e de iniciativas associadas [7]. Essa

escassez gera vieses de amostragem, nos quais o LCA passa a refletir artificialmente o último ancestral comum que simplesmente reúne os organismos com proteomas completos disponíveis no banco. Como consequência, a classificação tende a ser aproximada e não exata, aumentando uma probabilidade significativa de uma predição exagerada de atribuições filogenéticas inflacionadas [8]. Dessa forma, embora com ressalvas pela necessidade de bancos de dados mais abrangentes e representativos, a investigação da distribuição de LCA dos genes do ornitorrinco em múltiplos níveis de alinhamento torna-se uma abordagem ainda relevante para compreender a proximidade evolutiva desses genes ao longo da linhagem dos eucariotos e sustentar interpretações sobre o mosaico genômico característico desse monotremado.

2. Desenvolvimento

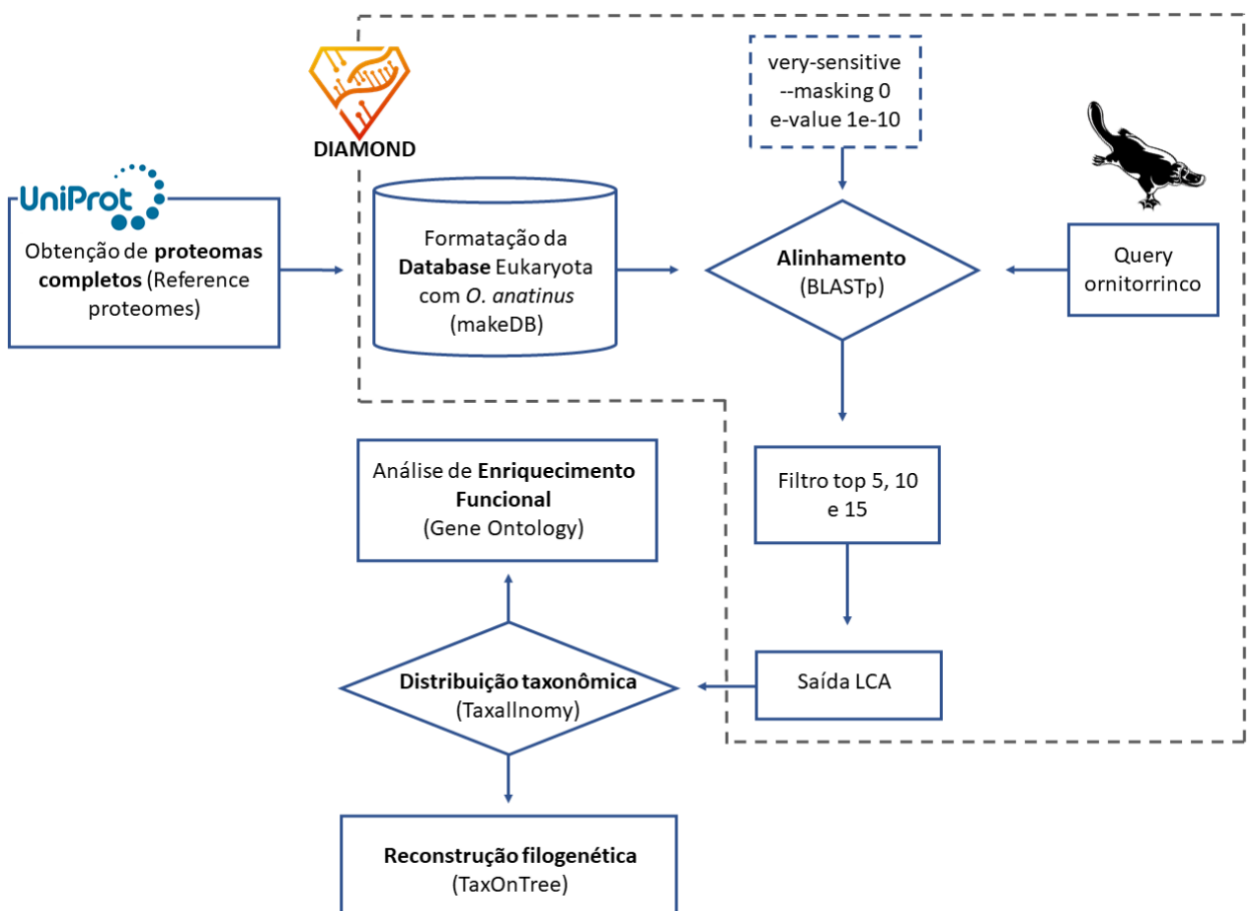


Figura 2: Fluxograma das etapas utilizadas para investigar a evolução gênica do ornitorrinco. O diagrama apresenta todas as etapas do processo, iniciando pela obtenção dos proteomas completos da divisão *Reference proteomes* (UniProt) seguido pela formatação da base e os subsequentes passos feitos através da ferramenta DIAMOND. Os alinhamentos são feitos aplicando diferentes faixas de top hit (5%, 10% e 15%), gerando as saídas em formato LCA. Esses resultados são então integrados ao banco Taxallnomy para recuperar a distribuição taxonômica dos LCAs, seguida por análises de enriquecimento funcional e reconstrução filogenética.

Para explorar a evolução gênica do ornitorrinco, é essencial reunir um grande conjunto de proteínas. Essas informações estão disponíveis em bancos de dados públicos, como o UniProt (<https://uniprot.org>), que reúne sequências de milhares de organismos. Entre suas coleções, destaca-se a seção *Reference Proteomes*, que reúne, no final do ano de 2025, mais de 35 mil proteomas completos e manualmente revisados para cada espécie. Dentre eles, mais de 2.700 pertencem ao domínio Eukaryota, totalizando quase 50 milhões de sequências disponíveis para estudos. No caso do ornitorrinco, o UniProt disponibiliza mais de 17 mil sequências de proteínas completas, o que permite compará-lo com diversos outros organismos e contribuindo para compreender melhor a evolução e as particularidades biológicas desse animal tão singular. Com esse conjunto de dados em mãos, utilizamos ferramentas modernas capazes de comparar proteínas em larga escala de forma rápida e eficiente. Uma das mais citadas é o software DIAMOND, desenvolvido para encontrar rapidamente similaridades entre sequências, mantendo uma sensibilidade comparável à do tradicional BLAST, do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Essa ferramenta compara cada proteína do organismo do nosso interesse, nesse caso, o ornitorrinco, com um vasto conjunto de proteínas de outras espécies, aqui representadas por organismos do domínio Eukaryota. Ela também disponibiliza diferentes formatos de saída para os alinhamentos. Um deles é o formato LCA acessado pela flag -f 102, que identifica qual é o ancestral comum mais recente compartilhado pelos resultados para cada query (com escore acima de determinada % do melhor escore) dentro de um limiar de similaridade definida pelo usuário, com a flag -top %. No presente estudo, aplicamos três diferentes faixas de top (5%, 10% e 15%) e analisamos a distribuição de LCAs entre elas. Essa saída retorna o TaxId do NCBI correspondente ao LCA e recuperar as informações taxonômicas desse identificador, como por exemplo: “Qual é a classe pertencente a esse LCA?”, a ferramenta Taxallnomy (<http://bioinfo.icb.ufmg.br/taxallnomy/>) é uma excelente opção, pois permite que seus dados em formato de tabelas sejam integrados a algum banco de dados relacional local, como o MariaDB, baseado na linguagem SQL. Esta ferramenta complementa e corrige inconsistências da *Taxonomy* do NCBI, interpolando categorias taxonômicas ausentes, como o nível “Cla_of_Testudinata” para a classe de tartarugas.

3. Resultados e Discussão

As análises iniciais revelaram um padrão esperado no conjunto gênico do ornitorrinco. A maior parte das sequências apresentou o LCA na própria espécie, sobretudo nos limiares mais restritivos de alinhamento. Esse comportamento decorre do fato de que o filtro top é baseado no score do *best hit*, o que gera um viés quando o proteoma de entrada também está presente na base de comparação. Nessas condições, o self-score elevado tende a excluir, nos intervalos de top mais estreitos, hits com scores inferiores, justamente aqueles mais divergentes, porém essenciais para investigar a história evolutiva de genes menos conservados, pois esses genes podem marcar inovações específicas da linhagem quando analisados mais a fundo. Por outro lado, à medida que os intervalos de top foram ampliados, surgiu um número crescente de genes cujo ancestral comum mais antigo remonta à linhagem Gnathostomata, nosso ancestral com tubarões. Esse comportamento reflete a inclusão de hits menos similares, que tendem a recuperar relações filogenéticas mais distantes. Um resultado particularmente intrigante emergiu quando focamos nos genes cujos alinhamentos não recuperaram outros representantes

de Mammalia além do ornitorrinco. Nesses casos, a classe que reúne as tartarugas surgiu como o grupo mais recorrente. A análise funcional com Gene Ontology deste subconjunto identificou vários genes, entre os quais haviam genes associados a processos ligados à formação da casca do ovo. As reconstruções filogenéticas manuais dessas proteínas reforçaram essa observação, indicando que a classe que reúne as tartarugas aparece consistentemente como o táxon de maior proximidade filogenética relacionado a esses genes, em comparação aos demais mamíferos. Esse padrão não deve ser interpretado apenas como uma afinidade filogenética entre monotremados e tartarugas, mas sim como possível evidência de uma homologia profunda, ou seja, a persistência e reutilização de repertórios genéticos ancestrais relacionados à reprodução ovípara. No caso dos monotremados, alguns componentes moleculares associados à produção, deposição mineral ou organização estrutural da casca do ovo podem ter sido conservados de maneira mais próxima ao padrão dos ovos das tartarugas, enquanto os mamíferos posteriores aos monotremados, por terem perdido completamente a oviparidade, abandonaram ou modificaram profundamente essas vias. Esses resultados, portanto, revelam que, embora o ornitorrinco pertença à classe Mammalia, uma fração específica de seu repertório gênico, especialmente genes relacionados à formação de ovos, apresenta maior proximidade evolutiva com Testudines do que com outros mamíferos, reforçando o papel da homologia profunda como mecanismo explicativo para a retenção de características genéticas ancestrais em linhagens filogeneticamente distantes.

Agradecimentos. A Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Programa Interunidades de Pós-Graduação em Bioinformática - ICB/UFMG.




5. Referências

- [1] Griffiths, M. **The platypus**. Scientific American, Vol. 258, p. 84-91 (1988). doi: <https://www.jstor.org/stable/10.2307/24989089>
- [2] Warren, W C et al. **Platypus genome analysis reveals unique evolution signatures**. Nature, V. 453 p. 175–183 (2008). doi: <https://www.nature.com/articles/nature06936>
- [3] Rheede T v et al. **The Platypus Is in Its Place: Nuclear Genes and Indels Confirm the Sister Group Relation of Monotremes and Therians**. Molecular Biology and Evolution, V. 23 Issue 3 p. 587–597 (2006). doi: <https://doi.org/10.1093/molbev/msj064>
- [4] Zhang D et al. **Homoplasy or plesiomorphy? Reconstruction of the evolutionary history of mitochondrial gene order rearrangements in the subphylum Neodermata**. International Journal for Parasitology, V 49 p 819-829 (2019). doi: <https://doi.org/10.1016/j.ijpara.2019.05.010>
- [5] Shubin, N et al. **Deep homology and the origins of evolutionary novelty**. Nature V 457 p 818–823 (2009). doi: <https://doi.org/10.1038/nature07891>
- [6] Sweetlove, L. **Number of species on Earth tagged at 8.7 million**. Nature (2011). <https://doi.org/10.1038/news.2011.498>
- [7] Eric Meunier E. **The genome of 1.8 million species is being sequenced**. infOGM - Citizen information watchdog on GMOs and seeds (2024). Disponível em: <https://infogm.org/en/the-genome-of-1-8-million-species-is-being-sequenced/#sdendnote4sym>. Acesso em: 25/11/2025.
- [8] Huson, D H et al. **MEGAN analysis of metagenomic data**. Genome Research V 17 p 377-386 (2007). doi: <https://doi.org/10.1101/gr.5969107>

12

Do modelo ao movimento: AlphaFold e CABS-flex na era da bioinformática estrutural

Autores 12.1

Hugo Guimarães Martins , Adenilson Arcanjo , Raquel C. de Melo-Minardi 

Revisão: Bibiana Fam , Savio Costa , Sheila Cruz Araújo , Tatiane Senna Bialves 

Cite este artigo 12.1

Martins, H. G.; Arcanjo, A.; Melo-Minardi, R. C. de **Do modelo ao movimento: AlphaFold e CABS-flex na era da bioinformática estrutural**. BIOINFO. Vol. 6. p.12 (2026). doi: 10.51780/bioinfo-06-12

Saiba mais 12.1

Este artigo está disponível em <https://bioinfo.com.br/do-modelo-ao-movimento-alphafold-e-cabs-flex-na-era-da-bioinformatica-estrutural/>

Do modelo ao movimento: AlphaFold e CABS-flex na era da bioinformática estrutural

Hugo Guimarães Martins^{1*} , Adenilson Arcanjo^{1,2} , Raquel C. de Melo-Minardi¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Campus Sobral, Sobral, Brasil

* hugoguimaraesmartins@ufmg.br

Resumo. O avanço do poder de processamento e a implementação de algoritmos têm ajudado a comunidade acadêmica a atingir novas fronteiras de conhecimento. Na bioinformática, a adoção de aplicações para a análise e geração de representações visuais gráficas moleculares, acarreta a adição de paradigmas computacionais ao desenvolvimento científico da área. Aplicações como AlphaFold e CABS-flex têm um papel relevante para a modelagem de conformações tridimensionais de biomoléculas e a execução de suas dinâmicas moleculares, fundamentais para a compreensão da biologia molecular e, conseqüentemente, desenvolvimento de diversas áreas econômicas relacionadas. Este trabalho, então, busca introduzir, de forma objetiva, uma análise comparativa com teor biológico e computacional, métodos e técnicas implementados por estes algoritmos, além de problemas enfrentados nestas aplicações.

Palavras-chave: Proteínas; Liberdade conformacional; Dinâmica molecular.

1. Introdução

As proteínas, cadeias de aminoácidos, são constituintes orgânicos fundamentais para a estrutura e atuam como catalisadores de reações químicas em todos os seres vivos, desempenhando papel crucial para a compreensão dos diversos sistemas biológicos e desafios atuais para o avanço da ciência. Sua função e ubiquidade nos diversos sistemas orgânicos despertam interesse econômico e científico, o que é refletido na criação de produtos, como metodologias experimentais, aplicações computacionais, fármacos, alimentos e periódicos acadêmicos [1].

A caracterização em nível atômico é realizada por técnicas experimentais como: difração de raios X, ressonância magnética nuclear, microscopia crioeletrônica. Estas metodologias possuem diferenças em suas execuções específicas e fornecem dados detalhados e confiáveis sobre a conformação molecular. Entretanto, estes métodos compartilham problemas fundamentais: o elevado custo operacional, a complexidade e diversidade molecular, a demanda de tempo para a execução destas técnicas e a incerteza de se obter resultados satisfatórios ao fim da realização de cada experimento.

Diante desses desafios, a bioinformática tem desenvolvido diversas técnicas e aplicações computacionais. A resolução destes empecilhos, agrega eficiência ao processo de catalogação, análise e aplicação da compreensão de biomoléculas, em especial, as proteínas.

Este artigo irá abordar uma breve análise de técnicas e aplicações voltadas à obtenção da conformação molecular e simulações de dinâmicas proteicas no contexto atual, especificamente o AlphaFold e CABS-flex, amplamente utilizados pela comunidade científica.

2. Base de dados de proteínas

A metodologia clássica experimental é, notadamente, uma abordagem fundamental para garantir resultados confiáveis na criação e no gerenciamento de uma base de dados da diversidade das proteínas. A criação de diversos repositórios de acesso livre e de colaboração mundial, como o *Protein Data Bank* (PDB), *The Universal Protein Resource* (UniProt) e *AlphaFold Protein Structure Database*, resultou em uma transparência quanto ao panorama global da compreensão de biomoléculas, devido a suas naturezas *open source*.

3. Parâmetros relacionados

A liberdade conformacional de uma proteína é definida, simplificada, pela sua capacidade de se movimentar sem o rompimento de suas ligações covalentes. Esta flexibilidade garante a ela variações conformacionais, além da capacidade de realizar atividades metabólicas fundamentais, como transporte de moléculas para diferentes ambientes e possibilidade de ligações químicas com outros agentes bioquímicos. Este fenômeno é calculado pela definição dos parâmetros geométricos e propriedades intrínsecas da proteína e dos fatores energéticos e eletrônicos do sistema. Outros fatores, como temperatura, pH e ligantes presentes, são levados em consideração neste sistema [2].

Com este considerável número de agentes conflituosos e constantes, em modelagem estrutural, é possível concluir que a proteína não é uma estrutura imóvel ou estática, tornou-se necessário o desenvolvimento de simulações de dinâmicas moleculares aplicadas a estas moléculas. Semelhantes aos desafios e problemas da obtenção de modelos conformacionais, as simulações de dinâmicas conformacionais são computacionalmente demandantes e, em sua maioria, resultam em poucos microssegundos ou *frames* da movimentação molecular, mesmo em computadores com alta capacidade de processamento e memória. Esta limitação na caracterização da dinâmica da proteína resulta em desconhecimento de funções e participações em processos metabólicos da proteína analisada, tendo em vista a relação direta entre a liberdade conformacional e a exposição ou o ocultamento de sítios ativos, destas aplicações, que interagem com outras moléculas.

4. Técnicas e aplicações

O servidor *web* CABS-flex é uma aplicação direcionada para a realização de simulações de flexibilidade conformacional de proteínas. Ele serve como uma alternativa rápida às dinâmicas moleculares (MD) tradicionais, que, embora mais detalhadas, são computacionalmente mais custosas. Fundamentalmente, o cálculo de conformações e dinâmicas moleculares, nesta aplicação, é realizado pela execução de equações pré-definidas e de caráter biofísico e químico, relacionadas a interações eletrostáticas e forças intermoleculares. O CABS-flex utiliza uma

abordagem *coarse-grained*, simplificando a representação atômica para permitir simulações rápidas de flexibilidade proteica [3]. Um exemplo do resultado de uma análise com CABS-flex pode ser visto na Figura 1.

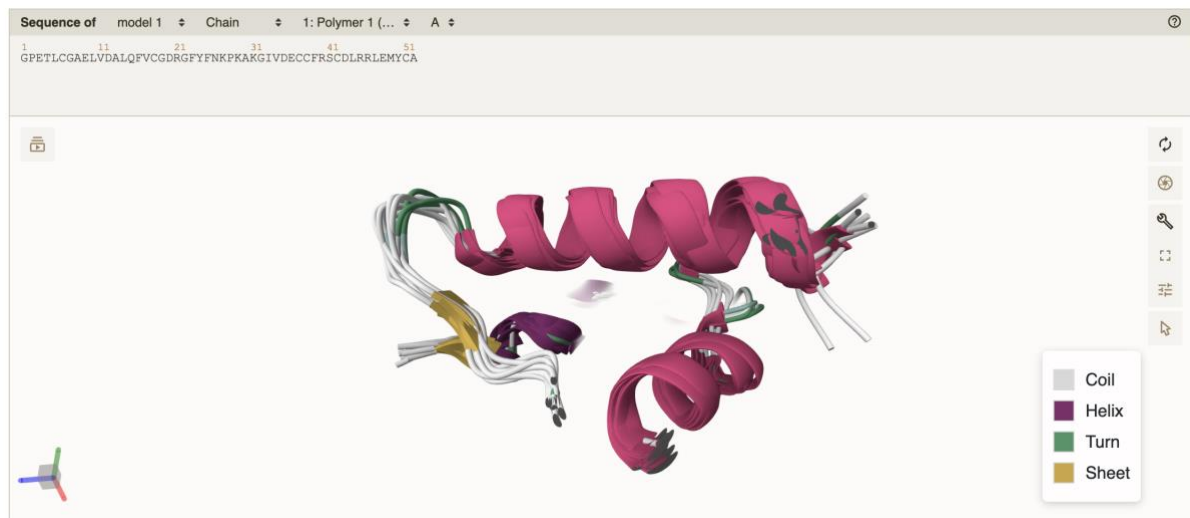


Figura 1. Modelo de uma proteína mini-IGF-1 (PDB 1tgr) obtido na execução do CABS-flex. Disponível em <https://lcbio.pl/cabsflex3/job/6e9db9e8217b9a/>.

Atualmente, uma das aplicações mais adotadas para modelagem molecular e com resultados bastante promissores é o AlphaFold, um modelo de aprendizado de máquina profundo que é capaz de prever estruturas de complexos biomoleculares diversos. Diferente do CABS-flex, o método central para obter as conformações e dinâmicas é realizado pela implementação de aprendizado de máquina. O aprendizado de máquina vem ganhando destaque e participação em diversos contextos. Resultados promissores e sua eficiência derivam-se da necessidade de um grande volume de dados previamente tratados, em que o algoritmo é capaz de criar e estabelecer equações e relacionamentos entre os dados, que, em novas execuções, aplicam a lógica previamente obtida [4].

5. Conclusão

A implementação de aprendizado de máquina, pelo AlphaFold, resultou em avanços significativos para o contexto, demonstrados pela quantidade e qualidade de dados obtidos desde sua criação, em que antes existiam centenas de milhares de estruturas em banco de dados proteicos e, após o AlphaFold, houve um avanço para milhões [5]. Entretanto, é necessária uma análise crítica da aplicação adotada, devido a sua natureza de caixa-preta. A definição e construção de equações pelo algoritmo é ocultado, como consequência disto insegurança e análise, das mesmas, não pode ser estudada pelos seus usuários. Uma de suas consequências mais comuns, não apenas nesta área, é a alucinação, este fenômeno ocorre quando o algoritmo cria dados para solucionar problemas referentes a sua execução. Outros problemas estatísticos são levantados e trabalhados nestes algoritmos, como o *overfitting* e *underfitting*, estes são relacionados ao número de dados fornecidos durante a etapa de treinamento do algoritmo e podem gerar tendências incorretas e, conseqüentemente, resultados também errados. Em

contrapartida, a dinâmica molecular clássica, utilizada no CABS-flex, possui, em seu alicerce, conceitos e relacionamentos transparentes e fundamentados por anos de trabalhos científicos, elevando a confiabilidade e compreensão dos resultados, mesmo não apresentando resultados tão eficientes e promissores em comparação com o AlphaFold.

A adoção dessas aplicações, em qualquer projeto, tem que ser criteriosa e seus usuários necessitam de compreender os métodos e técnicas selecionados; ignorar estes limites ou problemas pode introduzir vieses ou erros em trabalhos científicos e seus resultados.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig).






5. Referências

- [1] Batool, M. et al. **A Structure-Based Drug Discovery Paradigm**. Int. J. Mol. Sci. 2019, 20, 2783. doi: <https://doi.org/10.3390/ijms20112783>
- [2] Hollingsworth, Scott A. et al. **Molecular Dynamics Simulation for All**. Neuron. 2018 Sep 19;99(6):1129-1143. doi: <https://doi.org/10.1016/j.neuron.2018.08.011>
- [3] Aleksander, K. et al. **CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures**, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W338–W343, <https://doi.org/10.1093/nar/gky356>
- [4] Abramson, J. et al. **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. Nature 630, 493–500 (2024). <https://doi.org/10.1038/s41586-024-07487-w>
- [5] Mihaly, V et al. **AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models**. *Nucleic Acids Research*, Volume 50, Issue D1, 7 January 2022, Pages D439–D444, <https://doi.org/10.1093/nar/gkab1061>

13

Por que ensinar Bioinformática no Ensino Médio?

Autores 13.1

Helena Lott Costa , Diego Mariano , Leandro Moraes de Oliveira ,
Tarcísio José Domingos Coutinho , Raquel Cardoso de Melo-Minardi 

Revisão: Lucianna Helene Santos , Tatiane Senna Bialves , Thiago Jesus Sousa 






Cite este artigo 13.1

Costa H. L. *et al.* **Por que ensinar Bioinformática no Ensino Médio?**. BIOINFO. Vol. 6. p.13 (2026).
doi: 10.51780/bioinfo-06-13

Saiba mais 13.1

Este artigo está disponível em <https://bioinfo.com.br/por-que-ensinar-bioinformatica-no-ensino-medio/>

Por que ensinar Bioinformática no Ensino Médio?

Helena Lott Costa¹ , Diego Mariano¹ , Leandro Morais de Oliveira¹ , Tarcísio José Domingos Coutinho² , Raquel Cardoso de Melo-Minardi^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Laboratório de Bioinformática Aplicada a Saúde Única - LaBIASU, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB, Princesa Isabel, Brasil

* raquelcm@dcc.ufmg.br

Resumo. A rápida integração da tecnologia ao ensino tem transformado a forma como aprendemos e aberto espaço para novas áreas, como a bioinformática. Este artigo discute a importância de introduzir a Bioinformática no Ensino Médio como estratégia para desenvolver o pensamento computacional e tornar o aprendizado de Biologia mais dinâmico e interdisciplinar. É apresentado um projeto-piloto implementado em Belo Horizonte (MG), no qual estudantes de 15 a 18 anos aprenderam lógica de programação com as linguagens Scratch e Python, aplicando conceitos de biologia molecular em atividades práticas e investigativas. Os resultados indicam que o ensino de Bioinformática estimula o interesse dos alunos pelas ciências, promove autonomia e fortalece competências essenciais para as carreiras em Ciência, Tecnologia, Engenharia e Matemática (STEM).

Palavras-chave: Ensino; Scratch; Python.

1. Introdução

Vivemos em uma era em que a tecnologia redefine não apenas o modo como nos comunicamos, mas também o modo como aprendemos e entendemos o mundo. Nos últimos anos, essa transformação também alcançou as salas de aula, o que impulsionou mudanças profundas nos currículos escolares com a reformulação do Ensino Médio a partir da criação dos Itinerários Formativos, possibilitando o aprofundamento de áreas de interesse pelos alunos por meio de eletivas [1]. Tal movimento também é reforçado pela criação de um documento contendo habilidades e competências computacionais, como a inserção da lógica da programação, que devem ser alcançadas ao longo do Ensino Básico, como um complemento à Base Nacional Comum Curricular (BNCC) [2].

Entretanto, ensinar programação no Ensino Médio ainda é um desafio. Faltam professores com formação específica e materiais didáticos, o que torna a introdução desses conteúdos um processo lento e desigual. Ainda assim, o potencial dessa área é enorme, especialmente quando unimos o raciocínio lógico da computação à curiosidade natural das ciências biológicas.

É nesse ponto que a bioinformática se destaca. Ela conecta Biologia, Matemática, Computação e até mesmo a Língua Inglesa em uma só disciplina, permitindo que os estudantes

usem códigos e algoritmos para resolver problemas biológicos reais. Imagine, por exemplo, criar um pequeno programa para identificar padrões em sequências de DNA ou simular mutações em proteínas, tudo isso ainda no Ensino Médio! Essa abordagem desperta o interesse dos alunos, desenvolve o raciocínio científico e estimula o aprendizado interdisciplinar.

Além disso, metodologias ativas, como a aprendizagem baseada em investigação (*Inquiry-Based Learning*) e a gamificação, tornam o processo mais dinâmico e envolvente. Ao invés de apenas consumir conteúdo, os estudantes criam, testam e exploram, desenvolvendo autonomia e senso crítico. Por fim, em sintonia com os Itinerários Formativos, o ensino de Bioinformática abre portas para carreiras nas áreas STEM (Ciência, Tecnologia, Engenharia e Matemática), cada vez mais demandadas no mercado de trabalho [4][5].

2. O projeto-piloto em Belo Horizonte, MG

Com o objetivo de colocar essas ideias em prática, foi criada uma disciplina eletiva de Bioinformática dentro do Itinerário Formativo de Ciências da Natureza e Suas Tecnologias. O projeto foi aplicado por dois anos consecutivos no Colégio Nossa Senhora do Monte Calvário, em Belo Horizonte (MG), e contou com a participação de estudantes entre 15 e 18 anos.

A proposta começou de forma simples e acessível: o primeiro contato com a lógica de programação foi feito por meio da linguagem Scratch, uma plataforma gratuita e visual que usa blocos de comandos em vez de código digitado [3]. Isso reduziu a dificuldade inicial e permitiu que os alunos dessem um maior enfoque no raciocínio por trás da programação. Em seguida, eles avançaram para a linguagem Python, uma das mais usadas no mundo científico e na própria bioinformática.

Durante as aulas, os estudantes enfrentaram desafios inspirados em problemas reais da Biologia Molecular, aplicando estruturas de controle como “*if/else*” para simular processos biológicos e criar soluções computacionais. A cada projeto, a curiosidade crescia: compreender o DNA, criar jogos educativos ou até montar programas para analisar sequências genéticas tornava-se uma aventura científica (Figura 1).



Figura 1. Jogo criado pelos alunos: o usuário controla a enzima RNA polimerase para transcrever o DNA, coletando ribonucleotídeos no labirinto e evitando inibidores de transcrição.

Mais do que aprender a programar, os alunos desenvolveram pensamento crítico, resolução de problemas e criatividade, competências essenciais para qualquer carreira científica ou tecnológica do futuro.

3. Um olhar para o futuro

Expandir projetos como este para outras escolas do Brasil seria um passo fundamental para modernizar o ensino de Ciências e preparar os estudantes para os desafios de uma sociedade cada vez mais tecnológica. A introdução da Bioinformática no Ensino Médio estimula a curiosidade científica, o pensamento crítico e o raciocínio lógico, ao mesmo tempo em que conecta os conteúdos escolares a aplicações reais e atuais. Além disso, iniciativas desse tipo contribuem para reduzir a distância entre o ensino básico e a pesquisa científica, despertando vocações em áreas estratégicas como biotecnologia, ciência de dados e inteligência artificial.

4. Conclusão

A experiência em Belo Horizonte mostrou que ensinar Bioinformática no Ensino Médio é possível e transformador. Associando a programação à Biologia, o aprendizado deixa de ser abstrato e passa a fazer sentido para o aluno, que percebe como a tecnologia pode ser usada para compreender a vida. Usando linguagens acessíveis, como Scratch e Python, e metodologias ativas que incentivam a investigação, o projeto estimulou a interdisciplinaridade e o protagonismo estudantil. Também cumpriu com um dos grandes objetivos do Novo Ensino Médio: aproximar a Educação Básica ao mercado de trabalho, facilitando a tomada de decisão futura dos alunos pela experimentação da área proposta.

O trabalho evidencia a vocação interdisciplinar da Bioinformática, mostrando aos estudantes que é possível integrar diferentes campos do conhecimento e atuar em mais de uma área. Essa abordagem tem potencial para despertar o interesse de novos talentos para a Bioinformática, além de contribuir para a formação de professores e para o desenvolvimento de materiais didáticos que viabilizem sua aplicação em outros contextos educacionais.

Agradecimentos. os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

4. Referências

- [1] BRASIL. Lei nº 13.415, de 16 de fevereiro de 2017. 2017. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Lei/L13415.htm. Acesso em: 4 jul. 2025.
- [2] BRASIL. Base Nacional Comum Curricular: Computação — complemento à BNCC. 2022. Disponível em: <https://www.gov.br/mec/pt-br/assuntos/noticias/bncc>. Acesso em: 4 jul. 2025.
- [3] COSTA, H. L.; MAIA, G. C. F.; CHAVES, L. G. S.; MARIANO, D.; MELO-MINARDI, R. C. Teaching bioinformatics programming in high school: a case report. In: SIMPÓSIO BRASILEIRO DE BIOINFORMÁTICA (BSB), 17., 2024, Vitória. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 202-210. ISSN 2316-1248. DOI: 10.5753/bsb.2024.245609.
- [4] LEWITTER, F.; BOURNE, P. Teaching bioinformatics at the secondary school level. *PLoS Computational Biology*, v. 7, n. 10, e1002241, 2011. DOI: 10.1371/journal.pcbi.1002241.
- [5] MARQUES, I.; ALMEIDA, P.; ALVES, R. et al. Bioinformatics projects supporting life-sciences learning in high schools. *PLoS Computational Biology*, v. 10, n. 1, e1003460, 2014. DOI: 10.1371/journal.pcbi.1003460.

14

Pseudogenes podem ser peças-chave na diversidade de anticorpos de cavalos

Autores 14.1

Juliana Edelvacy Lima Pinto , João Henrique Diniz Brandão Gervásio ,
Joseph Chi-fung Ng , Carlena Navas , Liza Figueiredo Felicori 

Revisão: Alessandra Lima , Diego Mariano , Tatiane Senna Bialves 

Cite este artigo 14.1

Pinto J. E. L. *et al.* **Pseudogenes podem ser peças-chave na diversidade de anticorpos de cavalos.**
BIOINFO. Vol. 6. p.14 (2026). doi: 10.51780/bioinfo-06-14

Saiba mais 14.1

Este artigo está disponível em <https://bioinfo.com.br/pseudogenes-podem-ser-pecas-chave-na-diversidade-de-anticorpos-de-cavalos/>

Pseudogenes podem ser peças-chave na diversidade de anticorpos de cavalos

Juliana Edelvacy Lima Pinto¹ , João Henrique Diniz Brandão Gervásio² , Joseph Chifung Ng³, Carlena Navas⁴ , Liza Figueiredo Felicori^{1,4*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Unidade Genômica Evolutiva Baseada em Modelos, Instituto de Ciência e Tecnologia de Okinawa (OIST), Street, Okinawa, 904-0412, Japão.

³ Departamento de Pesquisa em Biologia Estrutural e Molecular, University College London, Londres, 610101, Reino Unido.

⁴ Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

*liza@icb.ufmg.br

Resumo. A resposta imune humoral depende de um repertório diversificado de anticorpos gerados pela recombinação somática de segmentos do gene da imunoglobulina (Ig) nos linfócitos B. Os anticorpos consistem tipicamente em duas cadeias leves (L) e duas cadeias pesadas (H) ligadas por pontes dissulfeto, cada uma contendo regiões constantes (C) e variáveis (V). As regiões variáveis geralmente são responsáveis por interações específicas com antígenos. Além da recombinação V(D)J, existem diferentes mecanismos que geram a diversidade dos anticorpos, como a hipermutação somática, a conversão gênica e a recombinação de troca de classe, que aumentam ainda mais a diversidade de anticorpos com a enzima citidina desaminase induzida por ativação (AID) desempenhando um papel central. Dentre esses processos, a conversão gênica é um dos menos estudados. Esse mecanismo depende da introdução de segmentos de pseudogenes em genes funcionais. Com o intuito de identificar eventos de conversão gênica no repertório de imunoglobulinas da região variável da cadeia pesada (IGHV) de cavalos, dados de sequenciamento de Ig podem ser utilizados na identificação desses eventos em larga escala. Esse tipo de estudo pode trazer uma nova perspectiva na atuação de pseudogenes como geradores de diversidade em anticorpos, além de promover um melhor entendimento da resposta imunológica equina.

Palavras-chave: Conversão gênica, imunoglobulina, repertório de anticorpos, diversidade, pseudogenes.

1. Introdução

No final do século XIX (1890), Behring e Kitasato foram pioneiros ao usar soro de animais como coelhos e cavalos imunizados para desenvolver antitoxinas contra a difteria e o tétano, respectivamente. Em 1891, Behring alcançou o primeiro tratamento bem-sucedido de uma criança contra difteria, uma descoberta que lhe rendeu o primeiro Prêmio Nobel de Fisiologia e Medicina em 1901 [1,2]. Desde então, os cavalos continuam indispensáveis para a produção de antivenenos no Brasil e foram também empregados durante a pandemia da COVID-19 para a produção de soros terapêuticos em países como a Argentina [3]. Seu uso extensivo é

justificado principalmente pelo alto volume sanguíneo desses animais e pela grande capacidade neutralizante de seus anticorpos, que permite a coleta repetida de sangue para o isolamento de grandes quantidades de anticorpos potentes com o intuito de utilizá-los em aplicações terapêuticas [4]. No entanto, apesar de sua relevância biotecnológica de longa data, estudos com foco no repertório de imunoglobulinas equinas permanecem limitados.

Em relação à organização do locus desta espécie, sabe-se atualmente que o locus da cadeia pesada da imunoglobulina equina (IGH) está localizado no cromossomo 24 e compreende 104 genes da região variável IGHV (agrupados em 7 subgrupos), 44 genes IGHD, 9 genes IGHJ e 11 genes IGHC (constantes). Dos 104 genes variáveis, apenas cerca de 21 genes IGHV são funcionais, seguidos por 9 genes com potencial de serem traduzidos -frames de leitura abertos (ORFs)- mas com alterações nos sítios de splicing, nos sinais de recombinação e/ou em elementos regulatórios, e/ou mudanças em aminoácidos conservados que levam ao dobramento incorreto, e/ou a entidade é um orfão e 74 pseudogenes [5, 6]. Curiosamente, os pseudogenes superam em número os genes funcionais na IGHV equina. Além disso, trabalhos anteriores do nosso grupo mostraram que, dos 21 genes funcionais, apenas 3 são usados em mais de 80% do repertório de IGHV em cavalos, sugerindo que o rearranjo gênico pode não ser o principal mecanismo gerador da diversidade de anticorpos nesta espécie [7]. Portanto, como existem estudos que sugerem que os pseudogenes podem contribuir para a diversidade de imunoglobulinas em outras espécies [8], começamos a nos perguntar se esses pseudogenes poderiam desempenhar um papel significativo na diversidade de anticorpos em cavalos.

Os pseudogenes (PGs) são sequências genômicas bastante semelhantes a genes funcionais, mas com funcionalidade alterada ou perdida, presentes em todas as formas de vida [9]. Embora sua relevância biológica tenha sido debatida por muito tempo, evidências crescentes apoiam seus papéis funcionais [10–12]. Um deles, particularmente relevante para imunoglobulinas, é a transferência de segmentos de PGs para genes funcionais por meio de recombinação não alélica, conhecida como conversão gênica [13]. Esse mecanismo tem se mostrado a principal forma pela qual a diversidade de anticorpos é gerada em galinhas e coelhos, integrando fragmentos de PGs em segmentos V recombinados de cadeias pesadas e leves de imunoglobulinas [14–16]. No entanto, embora a conversão gênica tenha sido bem documentada em galinhas e coelhos, sua ocorrência em imunoglobulinas equinas não é bem compreendida, mesmo que esses animais tenham a maquinaria necessária para esse processo [17].

Além disso, nos últimos anos, abordagens bioinformáticas foram desenvolvidas para identificar eventos de conversão gênica, todavia, apesar da variedade de métodos disponíveis, a maioria não é especificamente voltada para o estudo de imunoglobulinas e apresenta desempenho limitado ao lidar com grandes volumes de dados [18-20]. Isso destaca a necessidade de desenvolver novos métodos ou aprimorar os existentes para identificar eventos de conversão gênica que sejam adaptados ao grande volume de dados gerados pelas mais recentes tecnologias de sequenciamento de imunoglobulinas e capazes de serem validados. Assim, este estudo tem como objetivo apresentar formas para realizar a identificação de eventos de conversão gênica nos segmentos gênicos de IGHV de cavalos. A elucidação da conversão

gênica em imunoglobulinas equinas pode revelar novos mecanismos de diversidade de anticorpos, com aplicações potenciais no desenvolvimento de vacinas e na produção de anticorpos terapêuticos, semelhantes aos avanços já alcançados com imunoglobulinas de galinha [21].

2. Abordagem Integrada para Estudo da Diversidade de Imunoglobulinas Equinas

A compreensão dos mecanismos que geram diversidade no repertório de imunoglobulinas em equinos requer uma abordagem metodológica integrada, que combina técnicas avançadas de biologia molecular com ferramentas bioinformáticas. O sequenciamento de nova geração (NGS) emerge como ferramenta fundamental neste contexto, permitindo a caracterização em larga escala das sequências da região variável da cadeia pesada de anticorpos (IGHV). A relevância desta pesquisa reside na interação entre a geração massiva de dados transcriptômicos (sequenciamento de transcritos de células B, ou BCR-seq) e o desenvolvimento de pipelines bioinformáticos especializados para detectar eventos moleculares específicos, particularmente a conversão gênica.

A identificação dos eventos inicia-se com a coleta de amostras sanguíneas de cavalos (Figura 1), por meio do isolamento de células mononucleares seguido pelo isolamento do transcriptoma (Sequências de RNA) de linfócitos B, que é convertido em DNA complementar (cDNA). A amplificação seletiva via PCR multiplex, utilizando primers específicos para os segmentos IGHV, assegura a especificidade do conjunto de dados, direcionando o sequenciamento especificamente para o gene de interesse. Assim, as sequências de cDNA são preparadas para o sequenciamento Illumina MiSeq, por meio da ligação de adaptadores específicos. Essa técnica de sequenciamento gera um volume substancial de leituras (reads) pareadas (paired-end), constituindo uma representação abrangente do repertório de anticorpos em cada cavalo. A transformação desses dados brutos em informações biologicamente significativas exige um pipeline computacional robusto. O framework Immcantation [22] pode ser empregado para etapas críticas de pré-processamento, controle de qualidade e montagem das sequências, garantindo a integridade dos dados subsequentes. A anotação das sequências pode ser realizada através de comparação com bancos de dados curados como o IMGT/HighV-QUEST [23], permitindo a atribuição de cada sequência de anticorpo a sua provável origem germinativa.

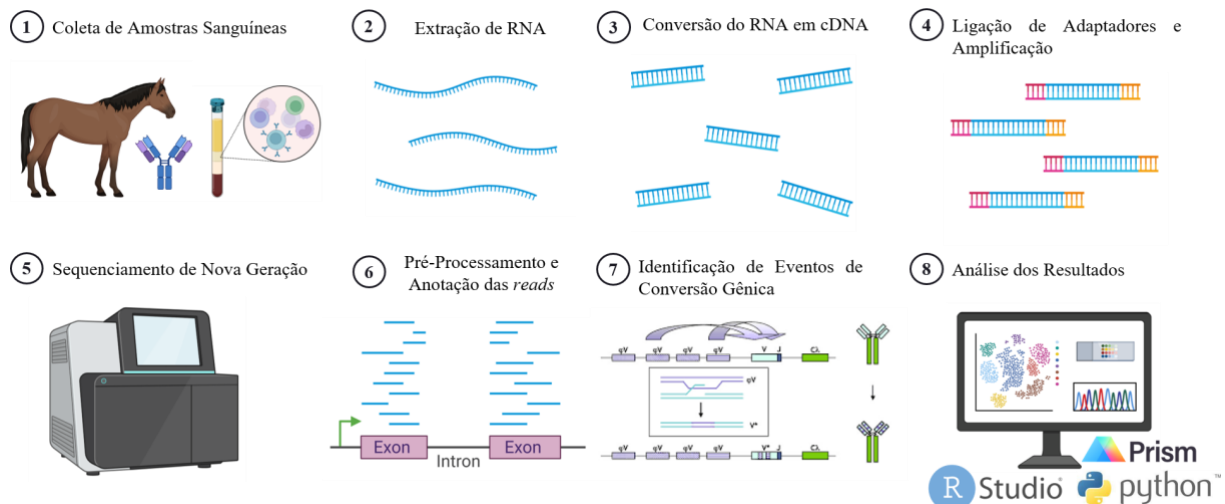


Figura 1. Identificação de Eventos de Conversão Gênica em Anticorpos de Cavalos. (1) A identificação de eventos de conversão se inicia com a coleta de amostras sanguíneas e separação de células mononucleares. (2) Essas amostras têm o seu RNA isolado. (3) O RNA é convertido em DNA complementar (cDNA) e amplificado utilizando *primers* específicos. (4) O DNA amplificado será preparado para o sequenciamento por meio da ligação de adaptadores e amplificação. (4) As sequências serão sequenciadas em sequenciadores de nova geração, como Illumina. (5) As sequências serão pré-processadas e anotadas utilizando bancos de dados públicos como o IMGT. (6) Com as sequências anotadas, estas terão os eventos de conversão gênica identificados por meio do uso de softwares específicos. (7) Os dados são analisados usando programas na linguagem R e Python e os gráficos podem ser representados no programa Prism.

A identificação de eventos de conversão gênica pode ser realizada mediante comparação sistemática de cada sequência recombinada com o conjunto completo de genes e pseudogenes da IGHV do genoma equino, utilizando algoritmos como o BrepConvert [24]. A detecção de segmentos nucleotídicos com maior identidade a pseudogenes do que aos seus genes funcionais correspondentes constitui evidência primária para inferência de eventos de conversão gênica. As etapas subsequentes de caracterização, incluindo quantificação da frequência relativa dos eventos e identificação de pseudogenes doadores preferenciais, podem ser realizadas utilizando scripts na linguagem R e Python e visualizadas utilizando ferramentas como o GraphPad Prism.

3. Conclusão

Em síntese, esta abordagem metodológica integrada demonstra como a combinação entre técnicas avançadas de sequenciamento e análise bioinformática permite converter dados transcriptômicos brutos em evidências validadas sobre mecanismos moleculares específicos. Cada evento de conversão gênica identificado representa uma possível fonte de diversidade em anticorpos equinos. Essa inferência é suportada pela utilização de dados de sequenciamento de alta resolução e pipelines adequados para a pergunta do estudo, contribuindo para o entendimento de mecanismos fundamentais na imunobiologia equina e que podem ser transpostos para identificação destes eventos em outras espécies.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e Fapemig.

4. Referências

- [1] Amanna IJ, Carlson NE, Slifka MK. Duration of humoral immunity to common viral and vaccine antigens. *New England Journal of Medicine*. 2007;357(19):1903–1915.
- [2] Alt FW, Oltz EM, Young F, Gorman J, Taccioli G, Chen J. VDJ recombination. *Immunology today*. 1992;13(8):306–314.
- [3] Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature biotechnology*. 2014;32(2):158–168.
- [4] Bashford-Rogers RJ, Smith KG, Thomas DC. Antibody repertoire analysis in polygenic autoimmune diseases. *Immunology*. 2018;155(1):3–17.
- [5] Putnam FW, Liu Y, Low T. Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain. *Journal of Biological Chemistry*. 1979;254(8):2865–2874.
- [6] Sela-Culang I, Kunik V, Ofra Y. The structural basis of antibody-antigen recognition. *Frontiers in immunology*. 2013;4:302.
- [7] Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*. 2020;160(3):233–247.
- [8] Grundbacher FJ. Behring's discovery of diphtheria and tetanus antitoxins. *Immunology today*. 1992;13(5):188–190.
- [9] Walther S, Rusitzka TV, Diesterbeck US, Czerny CP. Equine immunoglobulins and organization of immunoglobulin genes. *Developmental & Comparative Immunology*. 2015;53(2):303–319.
- [10] Zylberman V, Sanguineti S, Pontoriero AV, Higa SV, Cerutti ML, Morrone Seijo SM, et al. Development of a hyperimmune equine serum therapy for COVID-19 in Argentina. *Medicina (Buenos Aires)*. 2020;80:1–6.
- [11] Manteca Vilanova X, De Briyne N, Beaver B, Turner PV. Horse welfare during equine chorionic gonadotropin (eCG) production. *Animals*. 2019;9(12):1053.
- [12] Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic acids research*. 2015;43(D1):D413–D422.
- [13] Wibmer CK, Mashilo P. Exploiting V-Gene Bias for Rapid, High-Throughput Monoclonal Antibody Isolation from Horses. *Viruses*. 2022;14(10):2172.
- [14] Navas C, Manso T, Martins F, Minto L, Moreira R, Minozzo J, et al. The major role of junctional diversity in the horse antibody repertoire. *Molecular Immunology*. 2022;151:231–241.
- [15] Meyer A, Parng CL, Hansal SA, Osborne BA, Goldsby RA. Immunoglobulin gene diversification in cattle. *International reviews of immunology*. 1997;15(3-4):165–183.
- [16] Cheetham SW, Faulkner GJ, Dinger ME. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics*. 2020;21(3):191–201.
- [17] Clegg J. Gene conversions in the horse alpha-globin gene complex. *Molecular biology and evolution*. 1987;4(5):492–503.
- [18] Tang ES, Martin A. Immunoglobulin gene conversion: synthesizing antibody diversification and DNA repair. *DNA repair*. 2007;6(11):1557–1571.
- [19] Reynaud CA, Anquez V, Grimal H, Weill JC. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell*. 1987;48(3):379–388.









- [20] Patrinos GP, F'erec C, Wells RD, et al. Gene conversion causing human inherited disease: evidence for involvement of non-B- DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. *Human mutation*. 2009;30(8):1189–1198.
- [21] Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols*. 2010;2010(6):pdb–prot5439.
- [22] Gabernet G, Marquez S, Bjornson R, Peltzer A, Meng H, Aron E, et al. nf-core/airrflow: An adaptive immune receptor repertoire analysis workflow employing the Immcantation framework. *PLOS Computational Biology*. 2024;20(7):e1012265.
- [23] Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc MP. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome research*. 2012;8(1):26.
- [24] Mallaby J, Mwangi W, Ng J, Stewart A, Dorey-Robinson D, Kipling D, et al. Diversification of immunoglobulin genes by gene conversion in the domestic chicken (*Gallus gallus domesticus*). *Discovery immunology*. 2023;2(1):kyad002.

15

A toxicologia *in silico* no desenvolvimento de fármacos: um caminho ético e inovador na avaliação de segurança

Autores 15.1

Anne Josiele de Lima Vital , Larissa Camila Ribeiro de Souza , Barbara Moreira Amaral , Nathalia Stephanie Oliveira Nascimento , João Paulo Ataíde Martins , Carlos Tagliati 

Revisão: Ariany Rosa Gonçalves , Diego Mariano , Tatiane Senna Bialves 







Cite este artigo 15.1

Vital A. J. de L. *et al.* A toxicologia *in silico* no desenvolvimento de fármacos: um caminho ético e inovador na avaliação de segurança. *BIOINFO*. Vol. 6. p.15 (2026). doi: 10.51780/bioinfo-06-15

Saiba mais 15.1

Este artigo está disponível em <https://bioinfo.com.br/a-toxicologia-in-silico-no-desenvolvimento-de-farmacos-um-caminho-etico-e-inovador-na-avaliacao-de-seguranca/>

A toxicologia *in silico* no desenvolvimento de fármacos: um caminho ético e inovador na avaliação de segurança

Anne Josiele de Lima Vital¹ , Larissa Camila Ribeiro de Souza^{1,2}  Barbara Moreira Amaral¹ , Nathalia Stephanie Oliveira Nascimento³ , João Paulo Ataíde Martins⁴ , Carlos Tagliati^{1,3*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, United Kingdom

³ Programa de Pós-Graduação em Análises Clínicas e Toxicológicas da Faculdade Farmácia da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

⁴ Laboratório de Química Computacional Aplicada ao Planejamento de Fármacos da UFMG (LQCAPF – DQ-ICEEx), Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* carlostagliati@gmail.com

Resumo. A toxicologia desempenha um papel essencial no desenvolvimento de fármacos, e precisa estar alinhada às exigências regulatórias. As ferramentas computacionais representam um avanço na toxicologia contemporânea ao fornecer novos caminhos na previsão de risco e superação das limitações éticas e técnicas quanto aos modelos clássicos com animais. Métodos baseados em mecanismos oferecem maior interpretabilidade a partir da associação com mecanismos de toxicidade, constituindo um caminho para suprir as exigências de diretrizes regulatórias que buscam robustez comprovada das abordagens computacionais. Nesse contexto, a toxicologia *in silico* tem fornecido as ferramentas necessárias para a previsão do potencial tóxico dos compostos de forma estratégica, possibilitando a priorização de compostos com os melhores perfis de segurança. As abordagens baseadas na estrutura 3D, na relação entre a estrutura e a atividade e em grandes volumes de dados produzem modelos preditivos robustos, enquanto “métodos orientados por especialistas” fornecem inferências especializadas baseadas no conhecimento prévio do perfil das moléculas. Porém, os modelos *in silico* ainda possuem limitações, como a variabilidade dos dados e complexidade. Apesar disso, a toxicologia *in silico* vem se consolidando como uma área essencial, promovendo métodos alternativos ao uso de animais e inovações metodológicas, suprimindo demandas éticas, científicas e regulatórias.

Palavras-chave: toxicologia preditiva; NAMs; 3Rs.

1. Introdução

A escala de desenvolvimento de fármacos é composta por uma série de etapas, desde o descobrimento de uma nova entidade molecular até a liberação comercial de um medicamento [1]. Na avaliação da segurança destes, o modelo clássico com animais apresenta limitações éticas e técnicas, como a limitada reprodutibilidade e elevados custos, justificando o uso de abordagens inovadoras. Nesse contexto, as ferramentas computacionais são importantes aliadas

na fase atual da toxicologia, que abrange diversos métodos *in silico* na busca por performances mais preditivas [2], constituindo o objeto de discussão neste trabalho.

2. A toxicologia *in silico* e a avaliação de segurança de fármacos

A bioinformática é uma área fundamental para a toxicologia contemporânea, especialmente aplicada ao desenvolvimento de fármacos, pois disponibiliza uma ampla gama de ferramentas que auxiliam na solução de limitações relativas às ferramentas tradicionais. Este novo cenário se pauta na predição do risco, objetivo central da toxicologia preditiva, que surgiu como uma área de pesquisa necessária frente às dificuldades envolvendo modelos clássicos utilizando animais [3]. Apesar de tais abordagens integrarem protocolos bem estabelecidos na avaliação de segurança, fundamentando diretrizes internacionais, apresentam lacunas relativas à longa duração de experimentos, altos custos por composto e em relação à sua preditividade [4].

Dificuldades na extrapolação entre espécies são bem documentadas, por exemplo, para a nefrotoxicidade de fármacos, onde modelos animais frequentemente não reproduzem de forma precisa o dano renal observado em humanos [5]. Ainda nesse contexto, no estudo da lesão hepática induzida por fármacos, existem preocupações em relação ao processo de metabolização, uma vez que muitos modelos animais comuns não possuem homólogos enzimáticos com a mesma especificidade de substrato, além de apresentarem diferenças relevantes na função e expressão de outras enzimas metabólicas [6]. Além disso, questões éticas envolvendo o uso de animais na pesquisa vêm sendo debatidas globalmente, fazendo com que surgissem os princípios dos 3Rs – *Reduction, Refinement, Replacement* – que buscam promover a redução, o refinamento e a substituição do uso de animais na pesquisa, impulsionando abordagens éticas e explorando alternativas mais eficazes [7].

As diversas aplicações e técnicas referentes aos métodos computacionais em toxicologia estão representadas na Figura 1.

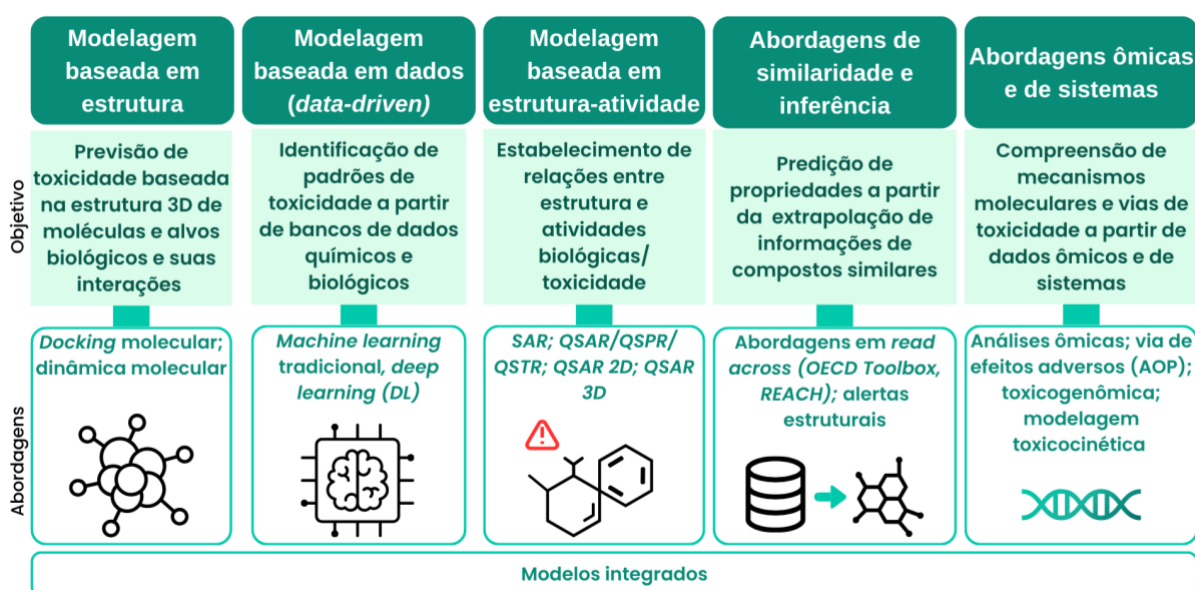


Figura 1. Diferentes técnicas empregadas em toxicologia computacional: métodos estruturais, *data-driven*, (Q)SAR, *read-across*, análises ômicas e sistêmicas [8],[12],[13].

Abordagens *data-driven* e baseadas em estrutura-atividade utilizam grandes volumes de dados toxicológicos para criar modelos robustos, capazes de identificar padrões e relações entre propriedades químicas e respostas biológicas [8]. A iniciativa CATMoS (*Collaborative Acute Toxicity Modeling Suite*) ilustra esse potencial ao fornecer estimativas de toxicidade aguda oral estruturadas, seguindo critérios que favorecem sua integração em contextos regulatórios [9]. Os alertas estruturais e *read-across*, conhecidos como “métodos orientados por especialistas”, fornecem inferências confiáveis, sendo formalmente aceitas por agências como a OECD (*Organisation for Economic Co-operation and Development*) e a ECHA (*European Chemicals Agency*) [8]. Também é possível integrar diferentes técnicas na busca pela otimização do desempenho, como a aplicação de aprendizagem de máquina avançada em modelos de relação estrutura-atividade quantitativos (QSAR) [10],[11]. A modelagem orientada a mecanismos e o conceito de via de efeitos adversos (AOP) buscam trazer uma representação concreta dos processos de toxicidade, fornecendo interpretações mecanicistas exigidas por marcos regulatórios [8],[12],[13]. A diretriz ICH M7 (R2) reconhece oficialmente o uso de técnicas *in silico* como ferramentas válidas para a avaliação de mutagenicidade, desde que sejam aplicadas duas abordagens independentes e complementares em QSAR: (a) baseada em regras; (b) baseada em estatística [14]. Desta forma, a priorização de compostos químicos na escala de desenvolvimento de fármacos é orientada, buscando estabelecer o equilíbrio entre a substância e suas propriedades de segurança [3],[8].

Dado o seu potencial, é importante reconhecer que essas ferramentas apresentam algumas limitações. A qualidade, curadoria e propriedades de dados disponíveis são pontos cruciais para o sucesso da aplicação dos métodos *in silico*. A complexidade de algumas estratégias frequentemente dificulta a interpretabilidade mecanicista [8]. A aplicação de modelos onde o funcionamento interno não é totalmente compreendido, também conhecidos como “caixas-pretas”, restringe a aceitação regulatória, que requer robustez e confiabilidade comprovada [13],[15].

Desta forma, a toxicologia *in silico* é um campo emergente, que desempenha um papel crucial na avaliação de segurança de fármacos, com impactos positivos na pesquisa e desenvolvimento. Trata-se de uma área que promove o uso de métodos alternativos ao uso de animais e de novas abordagens metodológicas (NAMs), sendo uma alternativa inovadora na solução de lacunas existentes, suprimindo a demanda ética e regulatória cada vez mais forte na atualidade.

Agradecimentos. Os autores agradecem à agência de fomento à pesquisa CAPES.

3. Referências

- [1] SUN, Duxin *et al.* Why 90% of clinical drug development fails and how to improve it?. **Acta Pharmaceutica Sinica. B**, [s. l.], v. 12, n. 7, p. 3049–3062, 2022.
- [2] WANG, Ningning *et al.* Data-driven toxicity prediction in drug discovery: Current status and future directions. **Drug Discovery Today**, [s. l.], v. 29, n. 11, p. 104195, 2024.

- [3] ZHANG, Jiangyan *et al.* Computational toxicology in drug discovery: applications of artificial intelligence in ADMET and toxicity prediction. **Briefings in Bioinformatics**, [s. l.], v. 26, n. 5, p. bbaf533, 2025.
- [4] ROMANO, Joseph D.; HAO, Yun; MOORE, Jason H. Improving QSAR Modeling for Predictive Toxicology using Publicly Aggregated Semantic Graph Data and Graph Neural Networks. **Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing**, [s. l.], v. 27, p. 187–198, 2022.
- [5] CONNOR, Skylar; ROBERTS, Ruth A; TONG, Weida. Drug-induced kidney injury: challenges and opportunities. **Toxicology Research**, [s. l.], v. 13, n. 4, p. tfae119, 2024.
- [6] MCGILL, Mitchell R.; JAESCHKE, Hartmut. ANIMAL MODELS OF DRUG-INDUCED LIVER INJURY. **Biochimica et biophysica acta. Molecular basis of disease**, [s. l.], v. 1865, n. 5, p. 1031–1039, 2019.
- [7] POH, Wen Tsin; STANSLAS, Johnson. The new paradigm in animal testing – “3Rs alternatives”. **Regulatory Toxicology and Pharmacology**, [s. l.], v. 153, p. 105705, 2024.
- [8] BUESO-BORDILS, Jose I. *et al.* Overview of Computational Toxicology Methods Applied in Drug and Green Chemical Discovery. **Journal of Xenobiotics**, [s. l.], v. 14, n. 4, p. 1901–1918, 2024.
- [9] MANSOURI, Kamel *et al.* CATMoS: Collaborative Acute Toxicity Modeling Suite. **Environmental Health Perspectives**, [s. l.], v. 129, n. 4, p. 47013, 2021.
- [10] ROE, Hannah M. *et al.* A Systematic Analysis of Read-Across Adaptations in Testing Proposal Evaluations by the European Chemicals Agency. **ALTEX**, [s. l.], v. 42, n. 1, p. 22–38, 2025.
- [11] TROPSHA, Alexander *et al.* Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. **Nature Reviews Drug Discovery**, [s. l.], v. 23, n. 2, p. 141–155, 2024.
- [12] CIALLELLA, Heather L.; ZHU, Hao. Advancing Computational Toxicology in the Big Data Era by Artificial Intelligence: Data-Driven and Mechanism-Driven Modeling for Chemical Toxicity. **Chemical Research in Toxicology**, [s. l.], v. 32, n. 4, p. 536–547, 2019.
- [13] HEMMERICH, Jennifer; ECKER, Gerhard F. In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. **WIREs Computational Molecular Science**, [s. l.], v. 10, n. 4, p. e1475, 2020.
- [14] INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. *ICH M7(R2): Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk*. Geneva: ICH, 2023.
- [15] POLISHCHUK, Pavel. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. **Journal of Chemical Information and Modeling**, [s. l.], v. 57, n. 11, p. 2618–2639, 2017.

16

As conexões invisíveis da vida: o papel das interações proteína-RNA e suas aplicações tecnológicas

Autores 16.1

Luana Luiza Bastos , Tatiane Senna Bialves , Raquel Minardi 

Revisão: Lucianna Helene Santos , Thiago Jesus Sousa 

Cite este artigo 16.1

Bastos, L. L.; Bialves, T. S.; Minardi, R. **As conexões invisíveis da vida: o papel das interações proteína-RNA e suas aplicações tecnológicas**. BIOINFO. Vol. 6. p.16 (2026). doi: 10.51780/bioinfo-06-16

Saiba mais 16.1

Este artigo está disponível em <https://bioinfo.com.br/as-conexoes-invisiveis-da-vida-o-papel-das-interacoes-proteina-rna-e-suas-aplicacoes-tecnologicas/>

As conexões invisíveis da vida: o papel das interações proteína-RNA e suas aplicações tecnológicas

Luana Luiza Bastos^{1,*} , Tatiane Senna Bialves¹ , Raquel Minardi¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* luizabastos.luana9@gmail.com

Resumo. *As interações entre proteínas e RNA são um dos pilares da regulação celular. A compreensão dessas interações é essencial para desvendar os mecanismos de reconhecimento molecular e as funções que essas macromoléculas realizam. O estudo estrutural dos complexos proteína-RNA, têm sido impulsionado por avanços em técnicas experimentais. Entretanto, a limitação de dados estruturais disponíveis, bem como custo e dificuldades de lidar com a flexibilidade do RNA, ainda representa um desafio para a consolidação do conhecimento nessa área. Discutir a importância do estudo dessas interações e a aplicação de metodologias de bioinformática é fundamental para o desenvolvimento de metodologias que ajudem a compreender os mecanismos da regulação celular. Além disso, esse entendimento ajuda a embasar pesquisas voltadas a patologias associadas à disfunção de complexos proteína-RNA e o desenvolvimento de novas terapias.*

Palavras-chave: RNA, proteínas, RBPs.

1. Introdução

O bom funcionamento celular depende da interação constante entre moléculas. Essas interações formam uma rede de comunicações invisível, que conduz e regula inúmeros processos. Essas interações são fundamentais para, por exemplo, determinar quando um gene será ativado ou quando a resposta a um estímulo deve começar. Entre os protagonistas dessa rede de interação, encontramos as proteínas e o ácido ribonucleico (RNA), duas macromoléculas essenciais para o equilíbrio celular [1].

Proteínas são moléculas fundamentais para a vida. Elas possuem múltiplas funções, como dar suporte e controlar processos celulares. Um grupo específico é especializado em interagir com as moléculas de RNA, são elas as proteínas de ligação do RNA (RNA-binding proteins ou RBPs) [2]. O RNA, por sua vez, também possui diferentes funções. Embora sua atribuição mais famosa seja a de mensageiro, sabemos que o RNA também atua como um coordenador, regulando diversas atividades celulares. Podemos dividir os RNAs em dois grupos principais: os construtores e os reguladores [3].

- **Construtores:** incluem o mRNA (RNA mensageiro), a molécula que atua como molde contendo as instruções para a tradução em proteína. O tRNA (RNA transportador), que tem como função transportador de aminoácidos até o ribossomo para a tradução. E por

fim, o rRNA (RNA ribossômico), que é um componente principal da estrutura do ribossomo [3].

- **Reguladores:** Os RNAs reguladores são RNAs que participam do controle celular. Existem vários RNAs nessa classe, mas podemos citar como exemplo os miRNAs (microRNAs), que atuam como interruptores, ou seja, desligam a produção de proteínas específicas. E os lncRNAs (RNAs longos não codificantes), que atuam como coordenadores, ajudando a organizar quando e como diferentes genes devem funcionar [3].

Sabe-se que muitas funções tanto do RNA, quanto da proteína são dependentes da interação dessas macromoléculas ocorrerem de forma correta. Diversos estudos atuais demonstraram que a falha nessas interações tem correlação direta com algumas doenças, como o câncer, o que torna esses complexos alvos promissores para o desenvolvimento de novos fármacos. E nesse sentido, o estudo estrutural dessas interações é fundamental para o entendimento desses processos biológicos, bem como o desenvolvimento de novas terapias direcionadas [4-5].

Apesar de técnicas modernas como a criomicroscopia nos fornecerem uma visão tridimensional (3D) dessas interações moleculares, os métodos experimentais enfrentam gargalos significativos. A realização de análises experimentais são processos de alto custo e com alta demanda de tempo [6-7]. Além disso, existem dificuldades associadas à natureza desses complexos. Por exemplo, os RNAs são altamente dinâmicos e flexíveis. Eles podem apresentar inúmeros dobramentos alternativos, o que muitas vezes dificulta a sua resolução por métodos experimentais. Ademais, complexos proteína-RNA são sensíveis à degradação por RNases (ribonucleases), enzimas que hidrolisam RNA e apresentam características como a facilidade de precipitação, o que dificulta a realização de análises experimentais [5-7].

É neste cenário que a bioinformática se torna uma ferramenta de grande utilidade para o entendimento desses complexos. Através da bioinformática estrutural, é possível realizar simulações e modelagens computacionais, que ajudam a prever interações ainda não observadas experimentalmente [8-9].

Além disso, modelos de inteligência artificial têm contribuído significativamente para a área. Esses modelos são capazes de aprender padrões complexos, utilizando grandes volumes de dados, para prever funções e comportamentos moleculares com alta precisão. Logo, essas abordagens multidisciplinares podem ajudar a preencher as lacunas do conhecimento no entendimento de complexos proteína-RNA [8-9].

2. Desenvolvimento

As RBPs são um conjunto de proteínas que regulam processos celulares vitais ao interagir com o RNA. Atualmente, elas formam um grupo de mais de 2.000 proteínas identificadas, que interagem com o RNA através de motivos de ligação presentes na estrutura da proteína [4-6].

O mau funcionamento de sua interação com o RNA mostrou-se diretamente ligado a diversas doenças, como câncer, distúrbios neurodegenerativos e cardiovasculares [4-6]. Nesse contexto, o estudo das interações proteína-RNA torna-se fundamental para o desenvolvimento de novos tratamentos [4-6]. No entanto, as RBPs são consideradas alvos terapêuticos complexos. Isso ocorre devido à dificuldade de bloquear as interações isoladas sem afetar outras vias essenciais à célula, desencadeando um efeito cascata [4-6]. Esse aspecto é fundamental, uma vez que muitas dessas proteínas e complexos atuam em mais de uma via de sinalização e regulação [5,6].

Nesse sentido, diversas estratégias inovadoras vêm sendo pesquisadas para superar esses desafios. Um exemplo de estratégia desenvolvida recentemente é o uso de esponjas moleculares. Em vez de propor metodologias para bloquear a proteína, essa abordagem utiliza RNAs terapêuticos (como os RNAs circulares) para sequestrar a RBP-alvo. Esses RNAs são projetados para capturar e neutralizar uma RBP específica, como já foi demonstrado com sucesso na proteção do coração em modelos animais [10].

Um outro exemplo é a utilização do RNA como um inibidor direto. Nesse caso, RNAs como os aptâmeros ligam-se a uma proteína para neutralizar sua função. Estas moléculas de RNA são projetadas para se dobrarem em estruturas 3D complexas, funcionando como se fossem uma espécie de anticorpos de RNA. Essas moléculas já demonstraram ser capazes de bloquear proteínas patogênicas com alta especificidade [11-13].

Para exemplificar, temos o complexo da Figura 1. Ele foi obtido em um estudo que investigava a estrutura do complexo entre a proteína ribossômica bacteriana S8 e um RNA artificial (aptâmero), utilizando o RNA como um inibidor (uma molécula que impediria o ribossomo de realizar sua função). Neste trabalho, foi observado que o aptâmero, ao se ligar à proteína, reorganiza drasticamente sua própria forma para imitar o sítio de ligação do ligante natural dessa proteína. Esses achados demonstraram a notável flexibilidade estrutural do RNA e o potencial terapêutico dessas estratégias [13].

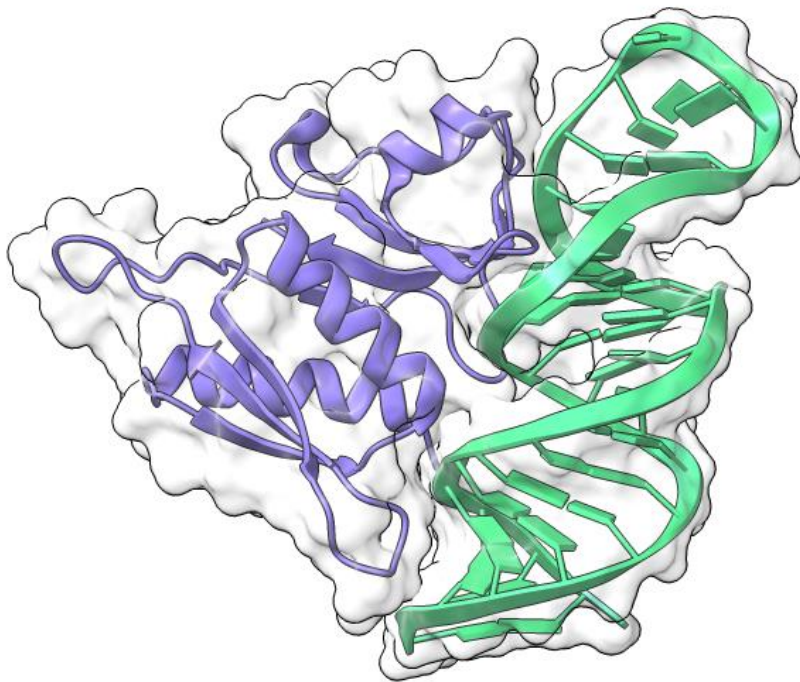


Figura 1. Complexo entre a proteína S8 e um aptâmero de RNA [13]. Em verde observamos a estrutura do aptâmero e em roxo a estrutura da proteína S8

Nesse cenário, a bioinformática tem se tornado uma aliada essencial para entender como proteínas e RNA se reconhecem e se conectam dentro das células. Ferramentas como o HDOCK [14] e o RnaX [15] permitem realizar o acoplamento molecular (*docking*) entre proteínas e RNAs, ajudando a simular e visualizar como essas moléculas interagem. Outras ferramentas, como o RBPsuite [16], utilizam inteligência artificial para prever quais trechos do RNA são reconhecidos por proteínas específicas. Além disso, existem recursos como o DSSR [17], projetado para identificar modificações presentes no RNA e prever as suas possíveis conformações alternativas. Ademais, muitos dos estudos *in silico* desses complexos se apoiam em bancos de dados, como o PRIDB [18] e o NPIDB [19], que reúnem estruturas e interações já observadas experimentalmente, servindo de base para novas descobertas. Um avanço recente veio com o AlphaFold3 [20], desenvolvido pela DeepMind. A ferramenta demonstrou ser capaz de modelar não apenas proteínas isoladas, mas também complexos formados por proteínas, RNAs, DNAs e pequenas moléculas [21-22].

As interações entre proteínas e RNA são extremamente diversas e dinâmicas. Logo, compreender completamente como elas acontecem e o que ocorre quando falham continua sendo um dos grandes desafios da biologia moderna e da bioinformática estrutural [20-22]. Como dito anteriormente, do ponto de vista experimental, características como alta flexibilidade, facilidade de degradação e precipitação dificultam a cristalização ou a captura

desses complexos por técnicas como criomicroscopia eletrônica, cristalografia e ressonância magnética nuclear [6-8].

Quando realizamos análises de complexos proteína-RNA *in silico*, também observamos inúmeros desafios. A capacidade do RNA de adotar conformações alternativas, sua alta flexibilidade e a presença de modificações são grandes pontos de atenção. Essas características tornam difícil prever com precisão suas conformações possíveis e suas interações com proteínas utilizando ferramentas computacionais [6-8,14-20]. Além disso, a superfície de interação de complexos proteína-RNA costuma ser extensa e altamente dependente do contexto celular. Essas características, somadas à alta flexibilidade do RNA, limitam o desempenho de métodos de *docking*, por exemplo. As ferramentas de predição de estrutura, por sua vez, enfrentam limitações relacionadas à escassez de dados experimentais de alta qualidade. Ademais, o desequilíbrio entre os diferentes tipos de complexos disponíveis nos bancos de dados, bem como a pouca disponibilidade de informações de afinidade experimental, dificultam a criação de modelos de inteligência artificial [14-20].

Além disso, a eficiência e a reprodutibilidade das análises computacionais são diretamente dependentes da curadoria e padronização dos dados estruturais. Atualmente, a fragmentação da informação e a ausência de anotações consistentes dificultam a construção de modelos preditivos robustos. A criação de bases de dados curadas e voltadas para esses complexos é, portanto, crucial para aumentar a acurácia desses métodos, fortalecer a integração com dados experimentais e viabilizar o desenvolvimento de novas terapias baseadas em RNA [21-23].

3. Conclusão

As interações entre proteínas e RNA são pilares da regulação gênica e de inúmeros processos celulares. A sua desregulação tem se mostrado relacionada a doenças como o câncer, o que as torna alvos terapêuticos de grande potencial. Contudo, o avanço no desenvolvimento de novas terapias, incluindo as baseadas em RNA, é afetado diretamente pelas limitações na realização de experimentos de bancada, pela dificuldade de lidar com a alta flexibilidade do RNA *in silico* e experimentalmente, bem como pela falta de dados precisos sobre esses complexos [4,6,21-22].

Portanto, investir na geração de dados experimentais por meio de técnicas como a criomicroscopia, aliado ao desenvolvimento de *pipelines* de bioinformática para curadoria, integração e análise estrutural, é fundamental para o entendimento desses complexos. Da mesma forma, a criação de bases de dados especializadas em complexos proteína-RNA torna-se essencial para consolidar e disponibilizar esse conhecimento. Por fim, o estudo aprofundado dessas interações permite compreender melhor inúmeros mecanismos moleculares e, ao mesmo tempo, impulsiona o desenvolvimento de medicamentos mais precisos, eficazes e seguros [4,6,21-23].

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

4. Referências

- [1] Alberts, B. et al. **Fundamentos da Biologia Celular**. 4ª edição. Artmed (2017).
- [2] Wuyun, Q. et al. **Recent Progress of Protein Tertiary Structure Prediction**. *Molecules*. Vol. 29. p.832 (2024). doi: <https://doi.org/10.3390/molecules29040832>
- [3] Polisenio, L; Lanza, M; Pandolfi, P.P. **Coding, or Non-Coding, That Is the Question**. *Cell Res*. Vol. 34. p.609–629 (2024). doi: <https://doi.org/10.1038/s41422-024-00975-8>
- [4] Fierro-Monti, I. **RBPs: An RNA Editor’s Choice**. *Front. Mol. Biosci*. Vol. 11 (2024). doi: <https://doi.org/10.3389/fmolb.2024.1454241>
- [5] Lunde, B.M. et al. **RNA-Binding Proteins: Modular Design for Efficient Function**. *Nat. Rev. Mol. Cell Biol*. Vol. 8. p.479–490 (2007). doi: <https://doi.org/10.1038/nrm2178>
- [6] Smith, P.R; Campbell, Z.T. **RNA-Binding Proteins in Pain**. *WIREs RNA*. Vol. 15. p.e1843 (2024). doi: <https://doi.org/10.1002/wrna.1843>
- [7] Steinmetz, B. et al. **Protein–RNA Interactions: From Mass Spectrometry to Drug Discovery**. *Essays Biochem*. Vol. 67. p.175–186 (2023). doi: <https://doi.org/10.1042/EBC20220177>
- [8] Ramanathan, M; Porter, D.F; Khavari, P.A. **Methods to Study RNA–Protein Interactions**. *Nat. Methods*. Vol. 16. p.225–234 (2019). doi: <https://doi.org/10.1038/s41592-019-0330-1>
- [9] Krautwurst, S; Lamkiewicz, K. **RNA-protein interaction prediction without high-throughput data: An overview and benchmark of in silico tools**. *Comput. Struct. Biotechnol. J*. Vol. 23. p.4036–4046 (2024). doi: <https://doi.org/10.1016/j.csbj.2024.11.015>
- [10] Acharya, P. et al. **RNA Binding Proteins as Mediators of Pathological Cardiac Remodeling**. *Front. Cell Dev. Biol*. Vol. 12 (2024). doi: <https://doi.org/10.3389/fcell.2024.1368097>
- [11] Niazi, S. et al. **Aptamer Functionalized Gold Nanoclusters as an Emerging Nanoprobe in Biosensing, Diagnostic, Catalysis and Bioimaging**. *Talanta*. Vol. 268. p.125270 (2024). doi: <https://doi.org/10.1016/j.talanta.2023.125270>
- [12] Yan, B. et al. **Aptamer-Mediated Therapeutic Strategies Provide a Potential Approach for Cancer**. *Int. Immunopharmacol*. Vol. 136. p.112356 (2024). doi: <https://doi.org/10.1016/j.intimp.2024.112356>
- [13] Davlieva, M. et al. **Structure analysis of free and bound states of an RNA aptamer against ribosomal protein S8 from Bacillus anthracis**. *Nucleic Acids Res*. Vol. 42. p.10795-808 (2014). doi: <https://doi.org/10.1093/nar/gku743>
- [14] Yan, Y. et al. **HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy**. *Nucleic Acids Research*. Vol. 45. p.W365–W373 (2017). doi: <https://doi.org/10.1093/nar/gkx407>
- [15] Blanco, J.D. et al. **Protein-assisted RNA fragment docking (RnaX) for modeling RNA-protein interactions using ModelX**. *Proc Natl Acad Sci U S A*. Vol. 116. p.24568-24573 (2019). doi: <https://doi.org/10.1073/pnas.1910999116>
- [16] Pan, X. et al. **RBPsuite: RNA-protein binding sites prediction suite based on deep learning**. *BMC Genomics*. Vol. 21. p.884 (2020). doi: <https://doi.org/10.1186/s12864-020-07291-6>

- [17] Lu, X.-J. & Olson, W. K. (2015). **DSSR: an integrated software tool for dissecting the spatial structure of RNA**. *Nucleic Acids Research*, 43(21), e142. <https://doi.org/10.1093/nar/gkv716>.
- [18] Lewis, B.A. et al. **PRIDB: A Protein–RNA Interface Database**. *Nucleic Acids Res.* Vol. 39. p.D277 (2011). doi: <https://doi.org/10.1093/nar/gkq1108>
- [19] Cook, K.B. et al. **RBPDB: A Database of RNA-Binding Specificities**. *Nucleic Acids Res.* Vol. 39. p.D301 (2011). doi: <https://doi.org/10.1093/nar/gkq1069>
- [20] Abramson, J. et al. **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature*. Vol. 630. p.493–500 (2024). doi: <https://doi.org/10.1038/s41586-024-07487-w>
- [21] Treger, M.; Westhof, E. **Statistical analysis of atomic contacts at RNA–protein interfaces**. *J. Mol. Recognit.* Vol. 14. p.199 (2001). doi: <https://doi.org/10.1002/jmr.534>
- [22] Zuo, Y. et al. **Research Progress on Prediction of RNA-Protein Binding Sites in the Past Five Years**. *Anal. Biochem.* Vol. 691. p.115535 (2024). doi: <https://doi.org/10.1016/j.ab.2024.115535>
- [23] Medina-Munoz, H.C. et al. **Expanded Palette of RNA Base Editors for Comprehensive RBP-RNA Interactome Studies**. *BioRxiv.* (2023). doi: <https://doi.org/10.1101/2023.09.25.558915>

17

Desvendando o risco renal sob coexposição química: do expossoma às redes biológicas

Autores 17.1

Barbara Moreira Amaral , Nathalia Stephanie Oliveira Nascimento , Anne Josiele Lima Vital , Carlos Alberto Tagliati 

Revisão: Alessandra Lima , Diego Mariano , Rafael Eduardo Oliveira Rocha 





Cite este artigo 17.1

Amaral B. M. *et al.* **Desvendando o risco renal sob coexposição química: do expossoma às redes biológicas**. *BIOINFO*. Vol. 6. p.17 (2026). doi: 10.51780/bioinfo-06-17

Saiba mais 17.1

Este artigo está disponível em <https://bioinfo.com.br/desvendando-o-risco-renal-sob-coexposicao-quimica-do-expossoma-as-redes-biologicas/>

Desvendando o risco renal sob coexposição química: do expossoma às redes biológicas

Barbara Moreira Amaral¹ , Nathalia Stephanie Oliveira Nascimento² , Anne Josiele Lima Vital¹ , Carlos Alberto Tagliati^{1,2*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Programa de Pós-graduação em Análises Clínicas e Toxicológicas da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

*carlostagliati@gmail.com

Resumo. *A toxicidade renal causada por medicamentos é desafio crescente à saúde pública, principalmente com o aumento da complexidade das exposições químicas concomitantes, que compõem o expossoma humano, conjunto total de exposições químicas a que somos expostos ao longo da vida. Essas coexposições podem modificar a resposta biológica dos rins, incluindo vias imunológicas, e alterar o risco de lesão renal, fenômeno ainda pouco explorado pela toxicologia tradicional. Nosso estudo apresenta um modelo mecanístico computacional, baseado em Bioinformática de Redes, para identificar potenciais interações entre fármacos e agentes químicos externos. A proposta integra dados ômicos, informações de vias metabólicas e toxicogenômica, permitindo mapear vias moleculares associadas à modulação do potencial nefrotóxico. O modelo busca compreender como exposições ambientais crônicas influenciam vias biológicas renais, alterando mecanismos de toxicidade e limiares de suscetibilidade. Essa abordagem, ao converter informações sistêmicas em métricas mecanísticas mensuráveis, alinhada às Novas Abordagens Metodológicas (NAMs) e aos princípios da Toxicologia Mecanicística, fornece uma estrutura reprodutível, biologicamente interpretável e auxilia na tomada de decisões em segurança farmacológica e regulatória, minimizando riscos à saúde pública.*

Palavras-chave: NAMs; Nefrotoxicidade; Bioinformática de Redes; Machine Learning; Expossoma.

1. Introdução

A toxicidade renal induzida por fármacos representa um importante problema de saúde pública, contribuindo para o agravamento de condições clínicas crônicas e elevando os índices de morbidade e mortalidade. A nefrotoxicidade medicamentosa é uma das principais causas de disfunção renal adquirida. Embora mecanismos clássicos, como lesão tubular e inflamação, estejam bem descritos, o contexto real de exposição humana é mais complexo [1].

Diversos são os casos relatados de indivíduos que apresentam histórico único de exposições ambientais crônicas, parte do que se conhece como expossoma, que modulam vias moleculares e imunológicas, influenciando a resposta renal a medicamentos e alterando o limiar de toxicidade individual. Essa modulação ainda é pouco explorada, mas evidências recentes apontam o papel de poluentes e contaminantes emergentes na progressão de doenças renais

crônicas [3]. O desafio consiste na questão de que os métodos tradicionais de avaliação de segurança analisam substâncias isoladamente, o que limita a previsão de respostas mais complexas e torna mais difícil o desenvolvimento de modelos mecanísticos mais robustos.

Diante disso, a Toxicologia Mecanicística fornece a estrutura para compreender a sequência de eventos que levam a um desfecho adverso. Já a Toxicologia de sistemas (*Systems Toxicology*) aumenta esse cenário quando une dados moleculares e funcionais provenientes de vários níveis biológicos [5]. A Toxicologia de Redes (*Network Toxicology*) surge dessa estrutura, uma vez que permite mapear interações, como por exemplo, redes de genes, proteínas e vias de sinalização, e seus efeitos sob condições de coexposição, ou seja, oferece uma perspectiva do corpo como um sistema interligado e demonstra de que forma variadas exposições químicas reconfiguram essas conexões [2,4,6,7]. A partir dessa estratégia, é possível unirmos informações de diferentes áreas da biologia molecular para gerar modelos de sistemas biológicos capazes de identificar rapidamente vias-alvo e modos de ação de fármacos e outras substâncias químicas.

Com o auxílio dessas plataformas, esta pesquisa utiliza a Bioinformática de Sistemas para apresentar um modelo computacional que pode explorar como exposições externas modulam vias renais e imunológicas, estimulando os avanços em toxicologia preditiva, a partir de modelos mecanísticos baseados em dados e contribuindo para práticas mais seguras em saúde pública e no uso de fármacos.

2. Abordagem Metodológica em Toxicologia de Redes

A metodologia proposta foi desenvolvida para enfrentar um dos principais desafios da toxicologia contemporânea: a dificuldade de prever a nefrotoxicidade induzida por fármacos em cenários de múltiplas exposições, uma das dimensões centrais do expossoma. Diferentemente de métodos tradicionais, que buscam avaliar a segurança através da análise das substâncias isoladamente e não de interações sistêmicas. Para isso, o trabalho integra princípios da toxicologia de sistemas e, de forma mais específica, da toxicologia de redes que analisa como diferentes exposições químicas afetam redes moleculares interconectadas. Essa abordagem amplia a visão clássica da toxicologia ao permitir a análise em diferentes níveis dos efeitos que possam ser gerados nas células e a identificação de nós críticos de toxicidade renal. Métodos baseados em rede têm se mostrado promissores na predição dos efeitos de fármacos e xenobióticos, elucidar modos de ação e identificar biomarcadores relevantes [5].

1. Curadoria e integração dos dados

Para criar o Atlas de Assinaturas Moleculares Renais, é realizada uma curadoria e integração de dados multiômicos e toxicogenômicos de fontes públicas (GEO, CTD, LINCS, ToxCast, entre outros), principalmente genômicos e transcriptômicos, os dados priorizados no estudo são com qualidade estatística controlada e padronizados, a combinação de dados multiômicos para construir redes biomoleculares capazes de predizer alvos e vias tóxicas de substâncias exógenas

é um dos pontos que fundamenta a toxicologia de redes, que ao ser empregada, possibilita a prevenção dos efeitos que surgem da interação de diferentes substâncias químicas [4, 6].

O Atlas será um recurso essencial para mapear vias biológicas afetadas e permitir análises comparativas de risco, auxiliando na identificação do potencial nefrotóxico em diferentes contextos de coexposição. A incorporação deste Atlas com técnicas de aprendizado de máquina é crucial, pois estudos recentes mostram que integrar aprendizado de máquina com dados ômicos tem melhorado a capacidade preditiva do risco de lesões renais de maneira mais precisa [8,9,10].

2. Construção da Rede e geração da estimativa de métricas de toxicidade

Os perfis moleculares são representados em redes de interação proteína-proteína (PPI), e torna-se possível identificar e medir as alterações funcionais e moleculares causadas pela interação de múltiplos agentes. Com base nessas alterações, geramos uma estimativa métrica (índice) da modulação tóxica que avalia a resposta celular e mostra, através de dados quantificáveis, o nível da modulação tóxica em vias renais. Esse índice será utilizado como uma variável de entrada para um modelo preditivo baseado em aprendizado de máquina, empregado de forma exploratória para estimar a modulação do risco nefrotóxico em diferentes cenários de coexposição.

3. Robustez e validação do Modelo

Para garantir que o modelo tenha robustez, confiabilidade e uma boa adaptação, duas ideias principais serão avaliadas:

a) Critério de seleção de dados:

Os bancos de dados priorizados serão públicos e consolidados na área de bioinformática e biologia de sistemas. Serão incluídos apenas estudos que possuem controle estatístico adequado e padronização de métricas, para ajudar a reduzir ruídos e imprecisões em experimentos diferentes.

b) Validação:

O conjunto de dados é dividido em grupos de treinamento e grupos de testes, para garantir que o algoritmo aprenda padrões reais e apresente resultados de forma eficiente. Técnicas como validação cruzada serão adotadas para assegurar que os resultados sejam reprodutíveis e robustos, verificando se ele consegue aplicar o que foi inserido.

4. Impacto social da abordagem

O fluxo metodológico do trabalho está detalhado na figura 1. Ele representa um progresso significativo para as Novas Abordagens Metodológicas e para a ciência translacional. Ao transformar interações biológicas complexas em previsões quantitativas, ele fortalece a

segurança de fármacos e contribui para a proteção da saúde pública, oferecendo ferramentas inovadoras para prevenir efeitos adversos renais e orientar triagens pré-clínicas. Desse modo, a metodologia une rigor científico à importância para a sociedade, evidenciando como a bioinformática pode apoiar práticas mais seguras e preditivas no desenvolvimento de fármacos.

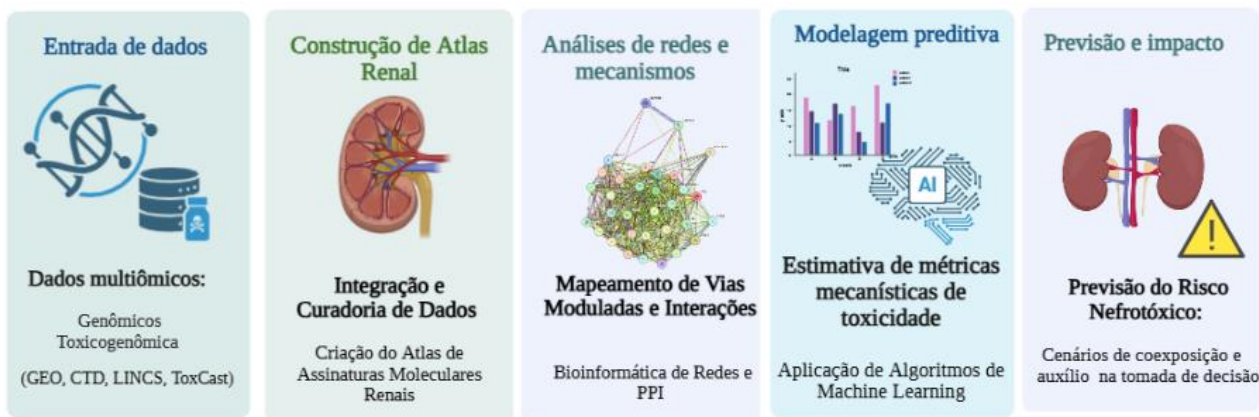


Figura 1. Fluxograma conceitual do modelo preditivo do risco nefrotóxico em coexposição química. O esquema ilustra a metodologia em cinco fases: (1) Curadoria e integração de dados multiômicos e toxicogenômicos para a (2) Construção de um Atlas de assinaturas moleculares renais. (3) A Bioinformática de Redes é utilizada para o mapeamento de vias noduladas e interações PPI. (4) A complexidade da rede é quantificada por uma estimativa de métricas mecánísticas de toxicidade, que vai alimentar algoritmos de Machine Learning para a (5) Previsão do risco nefrotóxico e suporte à decisão regulatória.

3. Considerações Finais

O presente estudo apresenta um novo modelo mecánístico capaz de avaliar a modulação biológica do risco nefrotóxico em cenários de exposição múltipla. Integrando dados ômicos e redes moleculares, a abordagem oferece previsões *in silico* éticas e reprodutíveis, ampliando a capacidade de antecipar efeitos adversos renais e elevando os padrões de segurança farmacológica. Ao demonstrar o potencial da bioinformática na análise de interações complexas, o modelo fortalece as NAMs e contribui para práticas mais seguras e preditivas em saúde e toxicologia translacional.

4. Referências

- [1] Ali, S.; Shahzad, M.; Aftab, M.; Ashraf, S. M. **Review article – Drug-induced kidney disease: epidemiology, pathophysiology, risk factors, diagnosis and prevention.** *International Journal of Pharmaceutical Sciences Review and Research*, v. 84, n. 2, p. 52–60, 2024
- [2] Del Giudice, G.; Serra, A.; Pavel, A.; Torres Maia, M.; Saarimäki, L. A.; Fratello, M.; Federico, A.; Alenius, H.; Fadeel, B.; Greco, D. **A network toxicology approach for mechanistic modelling of nanomaterial hazard and adverse outcomes.** *Advanced Science*, Weinheim, v. 11, n. 32, e2400389, 2024. DOI: 10.1002/advs.202400389..
- [3] Liang, K. H.; Colombijn, J. M. T.; Verhaar, M. C.; Ghannoum, M.; Timmermans, E. J.; Vernooij, R. W. M. **The general external exposome and the development or progression of chronic kidney disease: A systematic review and meta-analyses.** *Environmental Pollution*, v. 358, p. 124509, 2024. DOI: <https://doi.org/10.1016/j.envpol.2024.124509>.

- [4] Lin, Y.; Tao, Z.; Xiao, X. **Network toxicology and its application in studying exogenous chemical toxicity.** *Journal of Environmental and Occupational Medicine*, v. 32, n. 3, p. 275–278, 2025. DOI: 10.11836/JEOM24348.
- [5] Sturla, S. J. et al. **Systems toxicology: from basic research to risk assessment.** *Chemical Research in Toxicology*, v. 27, n. 3, p. 314–329, 2014. DOI: 10.1021/tx400410s.
- [6] Valls-Margarit, J.; Piñero, J.; Füzi, B.; Cerisier, N.; Taboureau, O.; Furlong, L. I. **Assessing network-based methods in the context of system toxicology.** *Frontiers in Pharmacology*, v. 14, p. 1225697, 12 jul. 2023. DOI: 10.3389/fphar.2023.1225697.
- [7] Zhang, W. **Network toxicology: a new science.** *Computational Ecology and Software*, v. 6, n. 2, p. 31–40, 2016. Disponível em: <http://www.iaees.org/publications/journals/ces/onlineversion.asp>.
- [8] Gong, Y. et al. **In silico prediction of potential drug-induced nephrotoxicity with machine learning methods.** *Frontiers in Pharmacology*, v. 14, p. 1222453, 2023. DOI: 10.3389/fphar.2023.1222453.
- [9] Wang, W. et al. **Artificial Intelligence and Machine Learning Models for Predicting Drug-Induced Kidney Injury in Small Molecules.** *Molecules*, v. 27, n. 23, 8436, 2022. DOI: 10.3390/molecules27238436.
- [10] Kim, M. T.; Allen, B. C.; Clewell, H. J. **The utility of new approach methodologies (NAMs) in human health risk assessment.** *Current Opinion in Toxicology*, v. 20–21, p. 100344, 2020. DOI: 10.1016/j.cotox.2020.10.005.

18

Valorizando o semiárido mineiro através da utilização de inulinas da biomassa de *Agave spp.*

Autores 18.1

Eduardo Horta Santos , Bruno Andrade , Maria Gabriela Bello Koblitz ,
Aristóteles Góes-Neto 

Revisão: Diego Mariano , Sheila Cruz Araújo 





Cite este artigo 18.1

Santos E. H. *et al.* **Valorizando o semiárido mineiro através da utilização de inulinas da biomassa de *Agave spp.***. BIOINFO. Vol. 6. p.18 (2026). doi: 10.51780/bioinfo-06-18

Saiba mais 18.1

Este artigo está disponível em <https://bioinfo.com.br/valorizando-o-semiarido-mineiro-atraves-da-utilizacao-de-inulinas-da-biomassa-de-agave-spp/>

Valorizando o semiárido mineiro através da utilização de inulinas da biomassa de *Agave spp.*

Eduardo Horta Santos^{1*} , Bruno Andrade^{1,2,3} , Maria Gabriela Bello Koblitz⁴ ,
Aristóteles Góes-Neto¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Laboratório de Bioinformática e Química Computacional, Universidade Estadual do Sudoeste da Bahia, Jequié, Brasil

³ Institut National de Recherche en Agriculture, Alimentation et Environnement (INRAE), Institut Agro, STLO, Rennes, França

⁴ Programa de Pós-Graduação em Alimentos e Nutrição, Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, Brasil

* eduardohdsant@gmail.com

Resumo. *A intensa demanda global por energia fomenta a busca por fontes renováveis para mitigar a dependência em combustíveis fósseis e reduzir as emissões de gases do efeito estufa. O bioetanol, vital para o setor de transportes no Brasil, quando oriundo de sacarose ou amido de biomassa de primeira geração (ex.: cana-de-açúcar, milho, beterraba), geralmente compete com o uso alimentar em solos agricultáveis. Uma exceção promissora seria a inulina, um carboidrato de reserva rico em frutose e abundante em plantas do gênero Agave. O cultivo de Agave já é bem estabelecido no Brasil e não compete com terras agrícolas, valorizando ecológica e economicamente áreas semiáridas, mesmo as atualmente degradadas. Entretanto, a eficiência da conversão da inulina em monossacarídeos fermentescíveis, majoritariamente frutose, é limitada, visto que a atividade catalítica e a termoestabilidade das inulinases (família GH32 de glicosídeo hidrolases) estão aquém do necessário para processos industriais. Portanto, é proposta a otimização de inulinases fúngicas reunidas no banco de dados FUNIN, dada a preferência industrial por seu maior rendimento quando comparada às enzimas bacterianas. Computacionalmente, modelos de linguagem de larga-escala de proteínas (ESM-3) serão usados para gerar hipóteses que serão inicialmente testadas por modelagem estrutural (DeepGlycanSite, HADDOCK, GROMACS). Em seguida, as enzimas candidatas passarão por validação biofísica experimental, incluindo dicróismo circular, SEC-MALS e ressonância magnética nuclear. Objetiva-se aumentar a eficiência da hidrólise e prevenir a reação de transfructosilação, maximizando a produção de frutose para a geração de bioetanol a partir de Agave e impulsionando a sustentabilidade energética.*

Palavras-chave: inulinases; bioetanol; inteligência artificial; GH32.

1. Introdução

Dentre os agravantes das mudanças climáticas, destaca-se a dependência energética em combustíveis fósseis, os quais atenderam 80% da demanda energética global em 2023 [1]. O

bioetanol é uma forte alternativa no Brasil, com 21% de participação nas demandas do transporte rodoviário em 2024 [1]. Contudo, a biomassa de primeira geração na produção de bioetanol geralmente enfrenta resistência devido ao desvio de uso de terras agrícolas (e.g. cana-de-açúcar, milho, beterraba) [2]. Uma solução é ofertada pela inulina, carboidrato de reserva rico em frutose, presente nas raízes e tubérculos de diversas plantas, muitas de regiões semiáridas, como o *Agave*. O cultivo de *Agave* pode se aproveitar de terras inadequadas ao plantio de alimentos, contornando a maior limitação das biomassas de primeira geração. Apesar disso, a biomassa de *Agave* ainda é muito pouco explorada. Por exemplo, apenas 4% das folhas de *Agave sisalana* (sisal) originam a fibra comercializada, verificando-se a produção de 24 toneladas de resíduos orgânicos por tonelada de fibra [3]. Ademais, o sisal já se encontra bem estabelecido no Brasil, concentrado em regiões onde poucas culturas comerciais são viáveis, particularmente no estado da Bahia, o qual contribui com 90% à produção nacional [3].

Nesse cenário, destacam-se as inulinases, enzimas da família das glicosídeo hidrolases 32 (GH32), que são responsáveis pela hidrólise da inulina. No entanto, a degradação enzimática da inulina requer avanços consideráveis para rendimento industrial com as inulinases, visando elevar suas propriedades críticas de eficiência catalítica e termoestabilidade. Trata-se, porém, de problemas multifatoriais que até o momento têm sido abordados de forma pontual, com estudos de caracterização e otimização de poucas enzimas por vez. A abordagem do presente trabalho (Figura 1) envolve o uso do banco de dados FUNIN de inulinases fúngicas, o que permitirá uma visão ampla do grupo essencial para entender as possibilidades de otimização dessas enzimas por inteligência artificial (IA), modelagem estrutural e validação biofísica.

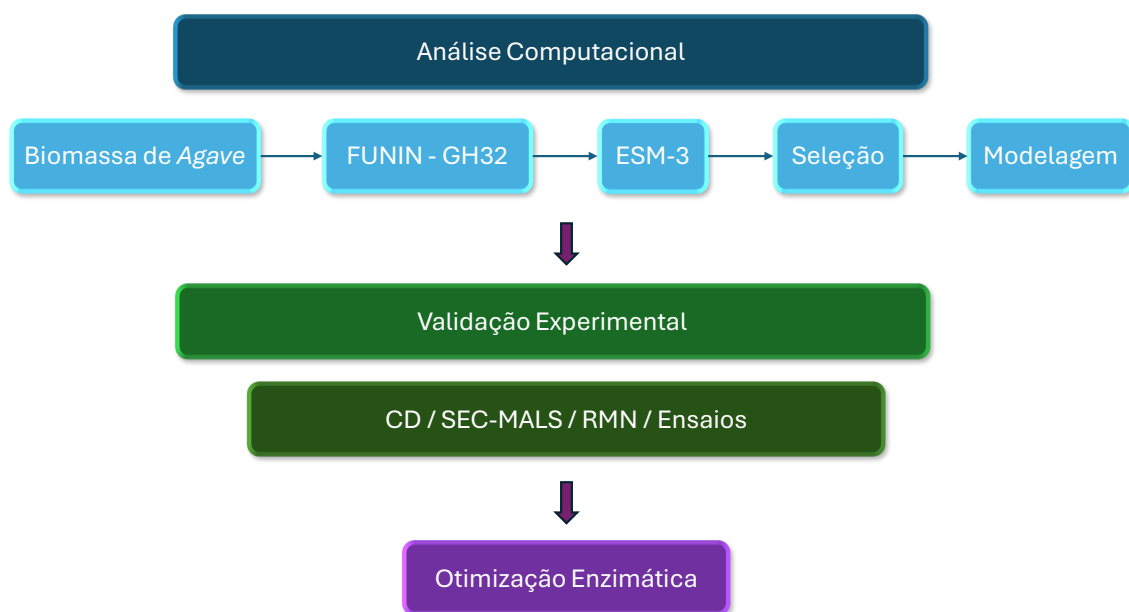


Figura 1. Metodologia resumida para otimização de inulinases.

Dessa forma, o modelo de linguagem de larga-escala de proteínas ESM-3 será usado como base para a geração de hipóteses a partir das inulinases no FUNIN, evidenciando padrões ocultos que podem informar a escolha de mutações interessantes, sejam elas novas ou

complementares às descritas na literatura, assim como apontar quais as melhores enzimas para otimização. O ESM-3 acessa informação para além de padrões diretos nas sequências de entrada, sendo capaz de explicitar relações não-arentes entre sequência, estrutura e função aprendidas a partir do treinamento com bilhões de proteínas advindas de diversos bancos de dados de sequência e estrutura [4]. O método organiza o conhecimento em um espaço interno onde tendências evolutivas, aspectos funcionais e motivos estruturais se reúnem, tornando esses dados mais acessíveis. Isso permite ultrapassar similaridades e padrões óbvios, podendo recuperar, através de seus impactos estruturais e funcionais, restrições bioquímicas que a literatura ainda não descreveu adequadamente nas GH32. Logo, o ESM-3 fornece um ponto de partida objetivo e inovador para definir mutações relevantes e selecionar enzimas com melhor potencial de otimização. Além disso, o modelo produz resultados que facilitam o uso posterior de ferramentas estruturais, ao filtrar os conjuntos de enzimas e mutações quanto à sua coerência geométrica e funcional, acelerando o processo em direção à modelagem estrutural dos candidatos. Primeiramente, ferramentas especializadas na predição de sítios de ligação a glicanos, como o DeepGlycanSite [5], fornecerão informação adicional para o acoplamento (HADDOCK [6]) e simulação de dinâmica molecular (GROMACS [7]) de complexos enzima-substrato. Com base nesses resultados, o conjunto filtrado de inulinases e variantes passará por validação com técnicas de biofísica experimental, almejando confirmar e entender de forma mecanística os comportamentos dos complexos modelados. Será realizada uma validação hierárquica, conforme o estudo de Jendrusch et al. de 2025 [8], implicando no uso de técnicas como dicroísmo circular (CD), Size Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS), ressonância magnética nuclear (RMN) e ensaios funcionais.

2. Resultados Esperados

São previstos avanços na sustentabilidade da produção de *Agave*, associando o tratamento dos resíduos com a geração de bioetanol. Paralelamente, o uso de IA na engenharia de proteínas vem demonstrando resultados promissores em trabalhos recentes: a geração de uma nova proteína fluorescente razoavelmente distante de qualquer outra conhecida (identidade máxima de 58%) [4], além do design *de novo* de inibidores da defesa bacteriana a fagos [8]. É razoável que abordagens nessa interface sustentem a geração de novas alternativas enzimáticas à hidrólise de carboidratos, sendo este trabalho um pioneiro na área. Finalmente, a modelagem de complexos proteína-carboidrato implica em desafios inerentes à grande flexibilidade do carboidrato e quantidade comparativamente baixa de dados experimentais disponíveis para esses complexos [9]. Assim, inovações metodológicas podem ser necessárias para permitir uma visão mais integrada dos mecanismos e relações inerentes aos padrões encontrados, uma demanda que será atendida ao longo do projeto. Crucialmente, o ESM-3 figura como mecanismo de aprofundamento na extração de padrões não-evidentes da diversidade das GH32, muito frequentemente inalcançáveis por métodos tradicionais como alinhamento, predição estrutural ou identificação de motivos descritos.

3. Conclusão

Conforme exposto, a proposta deste trabalho engendra ganhos não só para a valorização da biomassa de *Agave*, mas também inovação metodológica no estudo das inulinas e GHs como um todo. Por meio da integração entre modelos de IA e modelagem estrutural, é possível orientar etapas experimentais, primariamente biofísicas, que validem as hipóteses fundamentadas nas relações ressaltadas computacionalmente. Por fim, a estratégia estabelece um caminho claro para a otimização das inulinas, promovendo sua aplicabilidade industrial mediante aprofundamento científico, assim como o enriquecimento regional e a economia circular.

Agradecimentos. Os autores agradecem ao programa de pós-graduação em Bioinformática da UFMG.

4. Referências

1. IEA. World Energy Outlook 2024 [Internet]. Paris: IEA; 2024 [cited 2025 Jul 16]. Available from: <https://www.iea.org/reports/world-energy-outlook-2024>
2. Igwebuike CM, Awad S, Andrès Y. Renewable Energy Potential: Second-Generation Biomass as Feedstock for Bioethanol Production. *Molecules*. 2024 Jan;29(7):1619.
3. Broeren MLM, Dellaert SNC, Cok B, Patel MK, Worrell E, Shen L. Life cycle assessment of sisal fibre – Exploring how local practices can influence environmental performance. *J Clean Prod*. 2017 Apr 15;149:818–27.
4. Hayes T, Rao R, Akin H, Sofroniew NJ, Oktay D, Lin Z, et al. Simulating 500 million years of evolution with a language model [Internet]. bioRxiv; 2024 [cited 2025 Jul 18]. p. 2024.07.01.600583. Available from: <https://www.biorxiv.org/content/10.1101/2024.07.01.600583v2>
5. He X, Zhao L, Tian Y, Li R, Chu Q, Gu Z, et al. Highly accurate carbohydrate-binding site prediction with DeepGlycanSite. *Nat Commun*. 2024 Jun 17;15(1):5163.
6. Ranaudo A, Giuliani M, Pelissou Ayuso A, Bonvin AMJJ. Modeling Protein–Glycan Interactions with HADDOCK. *J Chem Inf Model*. 2024 Oct 14;64(19):7816–25.
7. Trapala J, González-Andrade M, Olvera C, Cayetano-Cruz M, Sanz-Aparicio J, Jimenez-Ortega E, et al. Relevance of aromatic and polar amino acids in the specificity of Inulinase ISO3 from *Kluyveromyces marxianus*: A molecular dynamics approach with an experimental verification. *Int J Biol Macromol*. 2023 Jul 1;242:124734.
8. Jendrusch MA, Yang ALJ, Cacace E, Bobonis J, Voogdt CGP, Kaspar S, et al. AlphaDesign: a de novo protein design framework based on AlphaFold. *Mol Syst Biol*. 2025 Jun 17;21(9):1166–89.
9. Nieto-Fabregat F, Lenza MP, Marseglia A, Carluccio CD, Molinaro A, Silipo A, et al. Computational toolbox for the analysis of protein–glycan interactions. *Beilstein J Org Chem*. 2024 Aug 22;20(1):2084–107.

19

Detecção de peptídeos bioativos em dermocosméticos: abordagens computacionais

Autores 19.1

Milenna Machado Pirovani , Raquel Cardoso de Melo-Minardi 

Revisão: Ariany Rosa Gonçalves , Lucianna Helene Santos , Thiago Jesus Sousa 

Cite este artigo 19.1

Pirovani, M. M.; Melo-Minardi, R. C. de **Detecção de peptídeos bioativos em dermocosméticos: abordagens computacionais**. BIOINFO. Vol. 6. p.19 (2026). doi: 10.51780/bioinfo-06-19

Saiba mais 19.1

Este artigo está disponível em <https://bioinfo.com.br/deteccao-de-peptideos-bioativos-em-dermocosmeticos-abordagens-computacionais/>

Detecção de peptídeos bioativos em dermocosméticos: abordagens computacionais

Milenna Machado Pirovani¹ , Raquel Cardoso de Melo-Minardi¹ 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* milennapirovani@ufmg.br

Resumo. O crescente interesse por peptídeos bioativos, devido aos seus efeitos anti-idade, antioxidantes e anti-inflamatórios, impulsiona o desenvolvimento de dermocosméticos mais seguros. Este artigo apresenta uma abordagem integrativa de bioinformática e modelagem molecular para identificar padrões estruturais e funcionais, visando prever novas sequências com potencial biotecnológico. A ideia sugere utilizar bases de dados públicas e algoritmos de aprendizado de máquina, seguidos por análises de docking e dinâmica molecular. A integração entre inteligência artificial e modelagem estrutural busca acelerar a descoberta racional de peptídeos bioativos para aplicação dermocosmética.

Palavras-chave: Bioinformática Estrutural; Dermocosméticos; Bioquímica.

1. Introdução

A indústria cosmética tem demonstrado crescente interesse em peptídeos bioativos devido às suas ações específicas na pele, como efeitos anti-idade, antioxidante e anti-inflamatório [1]. Essas moléculas, formadas pela união de aminoácidos, atuam na regulação de vias celulares relacionadas à renovação tecidual e à síntese da matriz extracelular. Entre os principais exemplos destacam-se o GHK-Cu (tripeptídeo de cobre) [2] e os peptídeos palmitoilados, amplamente utilizados em formulações antirrugas e estimuladoras de colágeno [3]. Apesar dos avanços, a ausência de uma correlação sistematizada entre estrutura e função limita o desenvolvimento racional de novos peptídeos bioativos. Tal lacuna representa um desafio científico, pois a criação de cosméticos eficazes, seguros e economicamente viáveis requer o entendimento dos mecanismos moleculares que conferem bioatividade [4–6].

Neste contexto, a presente pesquisa propõe identificar e caracterizar padrões estruturais e funcionais em peptídeos empregados em dermocosméticos, utilizando abordagens computacionais para prever novas sequências com potencial biotecnológico. Parte-se da hipótese de que peptídeos com ação cosmética compartilham motivos estruturais e propriedades bioquímicas específicas, passíveis de modelagem *in silico*, promovendo inovação científica e contribuindo para o avanço da biotecnologia cosmética [1,4].

2. Fonte de Dados e Curadoria

Para embasar a modelagem computacional, primeiramente deve ser construída uma base de dados de peptídeos cosméticos. Isso inclui compilar todas as sequências de peptídeos

disponíveis, usando bancos como *CosIng* (*Cosmetic Ingrediente Database*) [7], Propedia [8], *Peptide Atlas* [9], SATPdb [10], BIOPEP-UWN [11] e complementá-las com peptídeos relatados na literatura científica. Esta etapa de curadoria é fundamental para modelos de aprendizado de máquina, que dependem de conjuntos confiáveis de exemplos. Assim, cada peptídeo será caracterizado por suas propriedades de sequência (composição de aminoácidos, peso molecular, hidrofobicidade, dentre outros) e, quando possível, anotações funcionais associadas.

3. Triagem base-sequência

Algoritmos de aprendizado de máquina, como classificadores baseados em *Support Vector Machine* (SVM) [12], *Random Forest* [13] e redes neurais profundas, podem ser empregados no desenvolvimento de modelos preditivos a partir de sequências de peptídeos previamente caracterizados. Esses modelos são treinados com representações quantitativas, nas quais vetores de características físico-químicas descrevem as propriedades dos peptídeos. Metodologias de *virtual screening*, por sua vez, partem do princípio de que sequências com similaridade estrutural ou química tendem a compartilhar atividades biológicas semelhantes. Essa etapa gera um escore de probabilidade para cada sequência candidata, permitindo a seleção racional de peptídeos com maior potencial para análises estruturais subsequentes [14].

4. Modelagem Estrutural, Docking e Dinâmica Molecular

Inicialmente, as estruturas tridimensionais dos peptídeos serão obtidas por meio de ferramentas de predição estrutural, como o AlphaFold. De forma análoga, serão modelados os alvos biológicos, incluindo proteínas associadas à fisiologia cutânea e fosfolipídeos de membrana. Em seguida, será realizado o acoplamento molecular (*docking*) com o objetivo de prever o modo de ligação dos peptídeos aos respectivos alvos. Após essa etapa, os complexos com melhores pontuações serão submetidos a simulações de dinâmica molecular (*Molecular Dynamics*, MD). Essa abordagem permite avaliar a estabilidade temporal das interações em condições próximas às fisiológicas, considerando a mobilidade e a flexibilidade das moléculas envolvidas [4].

5. Desafios, Perspectivas e Conclusões

O projeto enfrenta desafios como a escassez de dados rotulados e a alta diversidade estrutural e funcional dos peptídeos, que dificultam a criação de modelos preditivos robustos. A representação adequada das sequências e o desequilíbrio entre classes de peptídeos ativos e inativos também comprometem o desempenho dos algoritmos de aprendizado de máquina. Na etapa estrutural, a flexibilidade dos peptídeos e a falta de estruturas experimentais dos alvos aumentam a incerteza nas predições de docking. Embora a dinâmica molecular ofereça uma análise mais realista das interações, seu custo computacional limita o número de candidatos testados. Ainda assim, a proposta integra bioinformática e modelagem estrutural para descobrir novos peptídeos dermatológicos, combinando inteligência artificial, docking e simulações. Essa

integração visa acelerar a triagem e priorização de compostos promissores para aplicação dermocosmética.

Agradecimentos. Às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

6. Referências

- [1] Ngoc, LTN et al. Insights into Bioactive Peptides in Cosmetics. *Cosmetics*. ISSN: 2673-9496. Vol. 10. p.111 (2023). doi: <https://doi.org/10.3390/cosmetics10040111>
- [2] Mortazavi, SM et al. Topically applied GHK as an anti-wrinkle peptide: Advantages, problems and prospective. *Bioimpacts*. ISSN: 2228-5660. Vol. 14. p.27719 (2024). doi: <https://doi.org/10.34172/bi.2024.27719>
- [3] Pinteá, A et al. Peptides: Emerging Candidates for the Prevention and Treatment of Skin Senescence: A Review. *Biomolecules*. ISSN: 2218-273X. Vol. 15. p.88 (2025). doi: <https://doi.org/10.3390/biom15010088>
- [4] Agoni, C et al. Molecular Modelling in Bioactive Peptide Discovery and Characterisation. *Biomolecules*. ISSN: 2218-273X. Vol. 15. p.524 (2025). doi: <https://doi.org/10.3390/biom15040524>
- [5] Li, S et al. Discovery of Highly Bioactive Peptides through Hierarchical Structural Information and Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*. ISSN: 1549-9596. Vol. 64. p.8164 (2024). doi: <https://doi.org/10.1021/acs.jcim.4c01965>
- [6] Du, Z et al. UniDL4BioPep: a universal deep learning architecture for binary classification in peptide bioactivity. *Briefings in Bioinformatics*. ISSN: 1477-4054. Vol. 24. p.bbád135 (2023). doi: <https://doi.org/10.1093/bib/bbad135>
- [7] European Commission. Cosmetic Ingredient Database (CosIng). European Commission. [Base de Dados Online]. (2024). Disponível em: https://single-market-economy.ec.europa.eu/sectors/cosmetics/cosmetic-ingredient-database_en
- [8] Martins, PM et al. Propedia: a database for protein-peptide identification based on a hybrid clustering algorithm. *BMC Bioinformatics*. ISSN: 1471-2105. Vol. 22. p.1 (2021). doi: <https://doi.org/10.1186/s12859-020-03882-y>
- [9] Desiere, F et al. The PeptideAtlas Project. *Nucleic Acids Research*. ISSN: 0305-1048. Vol. 34. p.D655 (2006). doi: <https://doi.org/10.1093/nar/gkj040>
- [10] Singh, S et al. SATPdb: a database of structurally annotated therapeutic peptides. *Nucleic Acids Research*. ISSN: 0305-1048. Vol. 44. p.D1119 (2015). doi: <https://doi.org/10.1093/nar/gkv1114>
- [11] Minkiewicz, P et al. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. *International Journal of Molecular Sciences*. ISSN: 1422-0067. Vol. 20. p.5978 (2019). doi: <https://doi.org/10.3390/ijms20235978>
- [12] Brereton, RG et al. Support Vector Machines for classification and regression. *The Analyst*. ISSN: 0003-2654. Vol. 135. p.230 (2010). doi: <https://doi.org/10.1039/B918972F>
- [13] Breiman, L. Random Forests. *Machine Learning*. ISSN: 0885-6125. Vol. 45. p.5 (2001). doi: <https://doi.org/10.1023/A:1010933404324>
- [14] Vincenzi, M et al. Virtual Screening of Peptide Libraries: The Search for Peptide-Based Therapeutics Using Computational Tools. *International Journal of Molecular Sciences*. ISSN: 1422-0067. Vol. 25. p.1798 (2024). doi: <https://doi.org/10.3390/ijms25031798>

20

Fusão de descritores biológicos em grafos de resíduos para predição da oncogenicidade de variantes

Autores 20.1

Lucas Moraes dos Santos , Tatiane Senna Bialves , Raquel Cardoso de Melo-Minardi 

Revisão: Bibiana Fam , Rafael Eduardo Oliveira Rocha , Savio Costa , Sheila Cruz Araújo 




Cite este artigo 20.1

Santos, L. M. dos; Bialves, T. S.; Melo-Minardi, R. C. de **Fusão de descritores biológicos em grafos de resíduos para predição da oncogenicidade de variantes**. *BIOINFO*. Vol. 6. p.20 (2026). doi: 10.51780/bioinfo-06-20

Saiba mais 20.1

Este artigo está disponível em <https://bioinfo.com.br/fusao-de-descritores-biologicos-em-grafos-de-residuos-para-predicao-da-oncogenicidade-de-variantes/>

Fusão de descritores biológicos em grafos de resíduos para predição da oncogenicidade de variantes

Lucas Moraes dos Santos ^{1,2} , Tatiane Senna Bialves ^{1,2} , Raquel Cardoso de Melo-Minardi ^{1,2} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* lucas.santos@dcc.ufmg.br

Resumo. O aprendizado multimodal tem se consolidado como abordagem promissora para lidar com a heterogeneidade e a complexidade dos dados biomédicos, especialmente em doenças como o câncer. Ao integrar descritores de múltiplas modalidades, esses modelos aprendem representações latentes mais expressivas e robustas a mudanças na distribuição dos dados, inclusive em cenários com poucos rótulos. Neste trabalho, discutimos fundamentos de aprendizado de representação multimodal e o papel de arquiteturas modernas, como Transformers e redes neurais de grafos, na modelagem de dependências de longo alcance e estruturas relacionais. Em particular, comparamos duas técnicas centrais de representação: (i) representação conjunta e (ii) representação coordenada. Por fim, destacamos aplicações recentes e desafios em interpretabilidade e robustez.

Palavras-chave: Multimodalidade; Redes Neurais de Grafos; Câncer; Variantes.

1. Aprendizado Multimodal

Tecnologias transformadoras como o sequenciamento de nova geração (NGS) têm revolucionado a bioinformática ao viabilizar a geração em larga escala de dados biológicos complexos [1,2]. Conjuntos de dados heterogêneos demandam métodos multimodais que combinam diferentes tipos de informação para elucidar sistemas e processos biológicos [3,4]. Nesse contexto, o aprendizado multimodal é um subcampo do aprendizado de máquina cujo objetivo é desenvolver modelos capazes de processar e relacionar descritores de diversas modalidades, para aprimorar o desempenho preditivo [3]. Um exemplo é a integração de dados de morfologia histopatológica (imagens) com os de transcriptômica (e.g., dados de expressão gênica), resultando em técnicas como a transcriptômica espacialmente resolvida, que possibilitam a análise da expressão gênica em seu contexto espacial [5].

Uma abordagem comum no aprendizado profundo é aprender representações numéricas, ou *embeddings*, em um espaço latente de baixa dimensionalidade e, no contexto multimodal, alinhá-las para que elementos semanticamente equivalentes ocupem regiões próximas nesse espaço [3,5]. Para viabilizar esse alinhamento em dados sequenciais, mecanismos de atenção, introduzidos por Bahdanau et al. (2014), permitem que o modelo ajuste dinamicamente o foco em partes da entrada [6]. Os Transformers (Vaswani et al., 2017) generalizam esse princípio por meio de autoatenção (*self-attention*), dispensando a recorrência das redes neurais recorrentes e modelando dependências de longo alcance [7].

Além disso, em cenários onde relações entre modalidades assumem uma estrutura de rede (e.g., interações em proteínas), as redes neurais de grafos (RNG) oferecem uma abordagem expressiva e agnóstica para modelar interdependências em dados multimodais [8]. As RNGs aprendem representações de nós, arestas, subgrafos e grafos inteiros por meio de estratégias de

propagação de mensagens (message passing) [4]. Nesse sentido, o aprendizado multimodal em grafos combina diferentes modalidades e aprende dependências intermodais a partir da estrutura do grafo. Modelos como o AlphaFold [9], uma rede neural profunda que representou um avanço significativo na predição estrutural, preveem estruturas 3D usando grafos de resíduos a partir de dados de homologia de sequência [4].

No contexto biomédico, doenças multifatoriais como o câncer apresentam acentuada heterogeneidade em nível tecidual, tornando crítica a integração de dados [5]. A literatura sugere que padrões de sequência e propriedades estruturais em genes reguladores modulam sua elevada flexibilidade conformacional e, conseqüentemente, sua maior tolerância ao acúmulo de mutações [10]. Essas alterações, ao comprometerem a função e a estabilidade proteicas, estariam envolvidas no processo de carcinogênese. Assim, hipotetizamos que abordagens de fusão multimodal são promissoras não apenas para prever mutações clinicamente relevantes com maior precisão [2], mas também para fornecer um esquema inicial para a compreensão mecanicista subjacente ao desenvolvimento tumoral.

2. Representação Multimodal

Nas últimas décadas, o aprendizado profundo (*deep learning*) tem mostrado grande potencial para identificar características e relações em grandes volumes de dados, oferecendo *insights* biológicos relevantes e aprimorando a caracterização de patologias [1,2]; ainda assim, a heterogeneidade intrínseca de doenças complexas, como o câncer, é frequentemente tratada de forma limitada por modelos unimodais [2]. Nesse contexto, o aprendizado multimodal busca integrar descritores heterogêneos provenientes de diferentes modalidades a fim de desenvolver modelos mais generalizáveis, robustos a mudanças nas distribuições de dados e eficientes em cenários com poucos dados rotulados [4].

A integração dessas informações é viabilizada pelo aprendizado de representação (*representation learning*), que aprende representações eficazes e seus mapeamentos diretamente a partir dos dados, formando espaços latentes expressivos nos quais relações semânticas podem ser inferidas [1,3]. No contexto multimodal, essa abordagem integra informações de múltiplas modalidades (por exemplo, imagens e texto) [9]. Nesse cenário, arquiteturas *Transformer* têm se destacado por sua escalabilidade e pela capacidade de modelar diferentes modalidades com menos suposições arquitetonicas específicas [8].

As abordagens de representação podem ser categorizadas em duas estratégias centrais:

- I. A representação conjunta (*joint representation*) refere-se à projeção de representações unimodais em um espaço multimodal [8], sendo utilizada em tarefas nas quais dados multimodais estão disponíveis no treinamento e na inferência. Geralmente, esse processo envolve a concatenação das modalidades (ou características) de entrada antes da projeção (fusão antecipada) [5]. No campo da bioinformática, por exemplo, avanços recentes integram informações de sequência e de estrutura usando modelos de linguagem de proteínas pré-treinados e redes neurais profundas geométricas [11]. Essas abordagens têm sido empregadas para avaliar a qualidade de modelos [11], pontuar poses de acoplamento (*docking*) [12], determinar a afinidade de ligação de ligantes [11] e prever o efeito funcional de variantes *missense* [13];
- II. A representação coordenada (*coordinated representation*) mapeia cada modalidade em seu respectivo espaço latente por meio de um codificador (*encoder*) dedicado [3]. Nesse caso, as representações são alinhadas por meio de uma função de perda contrastiva

(aprendizagem contrastiva), cujo objetivo é minimizar a distância entre as representações de pares semanticamente correspondentes (positivos) e maximizá-la entre pares não correspondentes (negativos), preservando a independência de cada projeção unimodal [9]. Um exemplo é treinar ativamente o modelo a mapear a região de pixels que forma um 'rosto sorridente' em uma imagem diretamente aos tokens exatos da palavra 'sorriso' na legenda, resultando em uma representação onde ambos os conceitos são mapeados para regiões muito próximas no espaço latente.

3. Conclusão

O aprendizado multimodal representa uma alternativa promissora à heterogeneidade de dados biomédicos, superando limitações de abordagens unimodais e impulsionando tarefas críticas em bioinformática. A representação conjunta favorece cenários com dados multimodais disponíveis no treino e na inferência, enquanto a representação coordenada mantém especificidades por modalidade e habilita alinhamento para pareamento entre domínios. Na bioinformática, essas estratégias têm demonstrado eficácia na avaliação de modelos, *docking*, estimação de afinidade e predição de efeitos de variantes. Como perspectivas, destacam-se: (i) o desenvolvimento de *benchmarks* que cubram diferentes combinações de modalidades; (ii) o aprimoramento da interpretabilidade e da validação biológica das representações; (iii) o pré-treino autossupervisionado em larga escala; e (iv) protocolos de robustez e generalização a mudanças de distribuição — condição comum em dados biomédicos.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e Fapemig.

4. Referências

- [1] Min, S; Lee, B; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* ISSN: 1467-5463. Vol. 18. p. 851–869 (2016). doi: <https://doi.org/10.1093/bib/bbw068>
- [2] Steyaert, S; Pizurica, M; Nagaraj, D. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat. Mach. Intell.* ISSN: 2522-5839. Vol. 5. p. 351–362 (2023). doi: <https://doi.org/10.1038/s42256-023-00633-5>
- [3] Akkus, C; Chu, L; Djakovic, V. Multimodal deep learning. *arXiv [Preprint]*. (2023). doi: [g/10.48550/arXiv.2301.04856](https://doi.org/10.48550/arXiv.2301.04856)
- [4] Ektefaie, Y; Dasoulas, G; Noori, A. Multimodal learning with graphs. *Nat. Mach. Intell.* ISSN: 2522-5839. Vol. 5, n. 4. p. 340–350 (2023). doi: <https://doi.org/10.1038/s42256-023-00624-6>
- [5] Acosta, JN; Falcone, GJ; Rajpurkar, P. Multimodal biomedical AI. *Nat. Med.* ISSN: 1078-8956. Vol. 28, n. 9. p. 1773–1784 (2022). doi: <https://doi.org/10.1038/s41591-022-01981-2>
- [6] Vaswani A et al. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* 30:5998–6008 (2017). <https://arxiv.org/abs/1706.03762>
- [7] Xu, P; Zhu, X; Clifton, DA. Multimodal learning with Transformers: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* Vol. 45, n. 10. p. 12113–12132 (2023). doi: <https://doi.org/10.1109/TPAMI.2023.3275156>
- [8] Baltrušaitis, T; Ahuja, C; Morency, L-P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* Vol. 41, n. 2 p. 423–443 (2019). doi: <https://doi.org/10.1109/TPAMI.2018.2798607>
- [9] Jumper, J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). doi: <https://doi.org/10.1038/s41586-021-03819-2>
- [10] Chillón-Pino, D et al. Protein structural context of cancer mutations reveals molecular mechanisms and candidate driver genes. *Cell Rep.* 43, 114905 (2024). doi: <https://doi.org/10.1016/j.celrep.2024.114905>



- [11] Wu, F et al. Integration of pre-trained protein language models into geometric deep learning networks. *Commun. Biol.* 2023;6(1):876. <https://doi.org/10.1038/s42003-023-05133-1>
- [12] Xu, X; Bonvin, AMJJ. DeepRank-GNN-esm: A graph neural network for scoring protein–protein models using protein language model. *Bioinformatics Adv.* 2024;4(1):vbad191. <https://doi.org/10.1093/bioadv/vbad191>
- [13] Zhang, H et al. Predicting functional effect of missense variants using graph attention neural networks. *Nat Mach Intell.* 2022;4(11):1017–1028. doi: <https://doi.org/10.1038/s42256-022-00561-w>

21

Da molécula ao algoritmo: utilizando aprendizado de máquina para desvendar a O-glicosilação de proteínas

Autores 21.1

Adenilson Arcanjo , Hugo Martins , Diego Mariano , Raquel C. de Melo-Minardi 

Revisão: Alessandra Lima , Lucianna Helene Santos , Savio Costa 

Cite este artigo 21.1

Arcanjo A. *et al.* Da molécula ao algoritmo: utilizando aprendizado de máquina para desvendar a O-glicosilação de proteínas. *BIOINFO*. Vol. 6. p.21 (2026). doi: 10.51780/bioinfo-06-21

Saiba mais 21.1

Este artigo está disponível em <https://bioinfo.com.br/da-molecula-ao-algoritmo-utilizando-aprendizado-de-maquina-para-desvendar-a-o-glicosilacao-de-proteinas/>

Da molécula ao algoritmo: utilizando aprendizado de máquina para desvendar a O-glicosilação de proteínas

Adenilson Arcanjo^{*1,2} , Hugo Martins² , Diego Mariano² , Raquel Melo-Minardi² 

¹ Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Campus Sobral, Sobral, Brasil

² Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

*adenilson.junior@ifce.edu.br

Resumo. A vida, em toda a sua diversidade, é fundamentalmente baseada na célula e seus constituintes, incluindo as proteínas. Estas moléculas realizam as mais variadas funções e podem sofrer modificações, como a O-Glicosilação com N-acetilglicosamina. A pesquisa apresentada concentra-se em compreender quando essa modificação ocorre, utilizando técnicas de aprendizado de máquina, métodos capazes de resolver problemas complexos que não são ditados por regras simples. O estudo desta pequena modificação, que a princípio pode parecer um detalhe, aponta para soluções de problemas importantes, como a doença de Alzheimer.

Palavras-chave: Proteínas; O-Glicosilação; Aprendizado de Máquina.

1. Introdução

Araras-azuis escolhem um parceiro e passam toda a vida ao seu lado, camaleões mudam de cor utilizando nanocristais em sua pele, bolor cresce sobre um pão velho, árvores permanecem vivas por milênios, cianobactérias produzem oxigênio e arqueias vivem em fendas no fundo do oceano Atlântico onde as temperaturas ultrapassam cem graus Celsius.

A vida apresenta uma variedade fascinante, organismos exuberantes e surpreendentes, em seu comportamento, aparência e modos de vida. Entretanto, todos os organismos vivos, sem exceção, têm algo em comum: são compostos pela mesma unidade fundamental, a célula. Alguns organismos são compostos por uma única célula, outros por trilhões. Porém, em todos os casos, são compostos por células.

Células são compostas, em sua maior parte, por água. Apesar disso, os seres vivos são muito mais complexos do que uma coleção de saquinhos de água. Além da água, as células contêm DNA, RNA, carboidratos, lipídios, proteínas e uma coleção de outras moléculas menores. Essa porção, que não é água, é o que torna os seres vivos tão fantásticos e tão diferentes de uma simples coleção de saquinhos de água.

De maneira simplificada, podemos dizer que o DNA guarda o manual de instruções da vida, o RNA transmite essas instruções, os carboidratos fornecem a energia para a execução das instruções, os lipídios armazenam essa energia e, também, delimitam cada célula, as moléculas menores executam papéis especializados e, por fim, as proteínas, obtidas pela tradução de informações do RNA, são, de fato, as moléculas efetoras e estruturais.

As proteínas têm as mais diversas funções: fazem os vaga-lumes brilharem no escuro, transportam oxigênio dos nossos pulmões ao resto do nosso corpo, quebram moléculas de

açúcar para obter energia, constituem cabelos, unhas, penas e chifres e atuam como maquinário para a movimentação de bactérias e músculos.

Apesar dessa variedade funcional, proteínas são constituídas, em quase todos os organismos, por apenas vinte aminoácidos e, além disso, estes aminoácidos são dispostos de maneira linear, ou seja, como numa corrente em que cada elo se liga ao próximo. Para cada nova sequência, uma nova proteína e, possivelmente, uma nova molécula extraordinária e única, talvez um pigmento que reage com a luz, ou um veneno letal, ou um remédio capaz de prevenir infartos.

As proteínas, após sintetizadas, podem sofrer pequenas modificações, adições de algumas moléculas à sua estrutura, o que aumenta consideravelmente o alfabeto inicial de vinte aminoácidos. Por exemplo, a adição de um grupo fosfato ($-\text{PO}_4^{3-}$) ou de um grupo metil ($-\text{CH}_3$). Na verdade, existem centenas destas modificações. Uma delas é a O-Glicosilação com N-acetilglicosamina, que acontece quando uma molécula de N-acetilglicosamina ($\text{C}_8\text{H}_{15}\text{NO}_6$) é adicionada a uma hidroxila de uma serina ou treonina, dois dos vinte aminoácidos que podem compor proteínas.

Finalmente, chegamos ao assunto da minha pesquisa: compreender quando um aminoácido de uma proteína pode receber uma N-acetilglicosamina [1]. Por que estudar algo tão específico? Antes de responder a esta pergunta, eu gostaria de descrever minha abordagem para enfrentar este problema.

2. Metodologia

Os programas de computador, conjuntos de regras pré-definidas por um desenvolvedor, permitiram desde enviar o homem à Lua a simular ataques alienígenas na tela de um videogame. Porém, existem tarefas tão complexas que os humanos nunca conseguiram criar um conjunto de regras capaz de resolvê-las. Um exemplo simples, e notavelmente difícil, é a identificação de gatos em imagens.

Uma imagem digital de 15 cm x 15 cm, em um monitor comum, é representada por cerca de um milhão de números, que definem as intensidades de vermelho, verde e azul dos seus pixels. Um programa que identifica gatinhos deve receber como entrada esse um milhão de números, realizar operações matemáticas com esses números, e devolver como saída 0 (não há gatinho) ou 1 (há gatinho).

Ninguém nunca conseguiu criar um conjunto de regras capaz de transformar esse milhão de números na resposta correta. Porém, os computadores são capazes, por meio de um método chamado aprendizado de máquina. Neste método, inicia-se com vários parâmetros, a princípio, escolhidos ao acaso. Esses parâmetros são utilizados para manipular os dados de entrada e obter a saída. Por meio de técnicas matemáticas, estes parâmetros são pouco a pouco ajustados, cada vez tornando o algoritmo mais certo.

Nas últimas décadas, o aprendizado de máquina floresceu, cada vez mais poderoso, tornou possível não apenas identificar gatinhos, mas também identificar câncer de pele melhor do que médicos com anos de experiência e foi capaz de superar, no Go, um jogo de tabuleiro milenar, os melhores jogadores do mundo [2]. Essa poderosa técnica tem sido aplicada aos mais diversos tipos de problemas, que, até então, nem humanos nem algoritmos criados por humanos foram capazes de resolver.

Identificar quais aminoácidos de uma proteína podem ser modificados por O-Glicosilação com N-acetilglicosamina é um problema difícil. Apesar de décadas de estudo, não

foi possível encontrar um conjunto de regras, simples ou complexo, que permitisse identificar quais aminoácidos podem sofrer esta modificação. Ou seja, um problema ideal para o aprendizado de máquina: esta é a técnica que aplico ao problema.

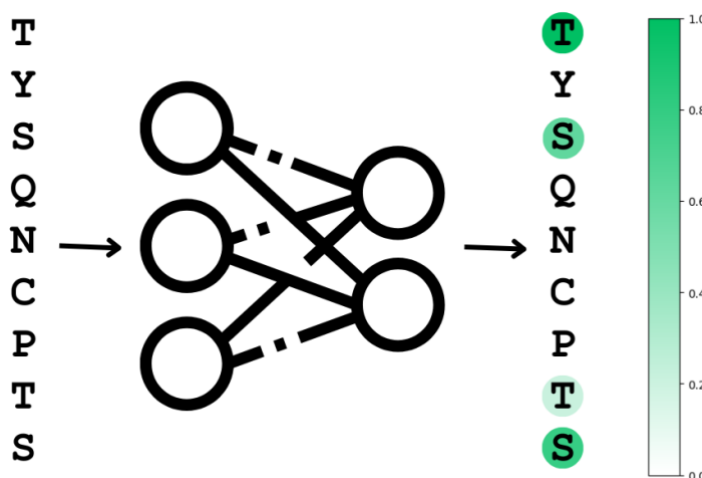


Figura 1. O algoritmo concebido recebe como entrada a sequência de aminoácidos de uma proteína e retorna como saída a probabilidade de que cada serina ou treonina da sequência seja um sítio de O-Glicosilação com N-acetilglicosamina. As serinas são representadas pela letra S, as treoninas são representadas pela letra T, a probabilidade de um aminoácido ser modificado é representada graficamente por um gradiente que vai do branco, representando probabilidade zero, ao verde escuro, representando probabilidade 1.

3. Conclusão

No início do texto, fomos da escala de organismo à escala atômica. Agora, faremos o caminho reverso. Considere o aminoácido que ocupa a ducentésima segunda posição de uma proteína chamada Tau, que tem quatrocentos e onze aminoácidos. Esta proteína é abundante nos neurônios do nosso cérebro.

Este aminoácido é uma serina; sabe-se que, quando esse aminoácido não sofre adição de N-acetilglicosamina, aumenta a chance de sofrer adição de um grupo fosfato. Esta adição, por sua vez, muda o formato e a carga da proteína Tau, o que faz com que proteínas Tau tenham afinidade umas pelas outras, o que faz com que formem agregados, que, por sua vez, prejudicam o funcionamento saudável dos neurônios, causando a doença de Alzheimer [3].

No momento, estão sendo testados medicamentos capazes de aumentar a adição de N-acetilglicosamina em proteínas e, conseqüentemente, impedir o aparecimento ou o avanço da doença de Alzheimer. Possivelmente, uma pequena modificação de alguns aminoácidos de uma proteína específica de um certo grupo de células do nosso corpo guarda o segredo para a cura da doença de Alzheimer, uma doença grave que afeta duramente a vida de milhões de pessoas e seus familiares. Este é apenas um exemplo, dentre muitos outros, de como esta pequena modificação pode ser importante.

Como primeira contribuição do meu projeto de pesquisa de doutorado, realizada durante o primeiro trimestre de trabalho, publicamos o algoritmo com a melhor performance já obtida para prever O-Glicosilação com N-acetilglicosamina em proteínas [1]. Essa ferramenta (Figura 1), disponível abertamente para outros pesquisadores, pode ser utilizada na elucidação dos mais diversos fenômenos biológicos. As implicações da utilização dessa ferramenta, assim como o estudo de outras modificações, darão continuidade ao projeto.

Assim funciona a ciência: nos debruçamos, muitas vezes, sobre uma molécula, de uma célula, de um tecido, de um organismo, comumente utilizando técnicas profundas, desenvolvidas por meio do acúmulo de séculos de conhecimento, para compreender um fenômeno que pode parecer pequeno ou insignificante, mas que pode ser a chave para a solução de problemas fundamentais para a humanidade.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.





4. Referências

- [1] Arcanjo, A., Mariano, D., Bastos, L., Bastos, A., Pirovani, M., Melo-Minardi, R. Using Protein Language Models Embeddings to predict O-GlcNAc glycosylation sites. 28th Brazilian Symposium on Bioinformatics.
- [2] Ludermir, T. B. (2021). Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. Estudos Avançados, 35(101), 85–94. <https://doi.org/10.1590/s0103-4014.2021.35101.007>.
- [3] Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W., & Gong, C.-X. (2004). O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease. Proceedings of the National Academy of Sciences, 101(29), 10804–10809. <https://doi.org/10.1073/pnas.0400348101>.

22

Bioinformática translacional: integração multiômica e inteligência artificial aplicada a periodontite

Autores 22.1

Wellington Francisco Rodrigues , Raquel Cardoso de Melo Minardi ,
Aristóteles Goés-Neto , Siomar de Castro Soares 

Revisão: Bibiana Fam , Diego Mariano , Marcos Antonio Nobrega de Sousa , Thiago
Jesus Sousa 





Cite este artigo 22.1

Rodrigues, W. F. *et al.* **Bioinformática translacional: integração multiômica e inteligência artificial aplicada a periodontite**. BIOINFO. Vol. 6. p.22 (2026). doi: 10.51780/bioinfo-06-22

Saiba mais 22.1

Este artigo está disponível em <https://bioinfo.com.br/bioinformatica-translacional-integracao-multi-omica-e-inteligencia-artificial-aplicada-a-periodontite/>

Bioinformática translacional: integração multi-ômica e inteligência artificial aplicada a periodontite

Wellington Francisco Rodrigues^{1*}, Raquel Cardoso de Melo Minardi¹, Aristóteles Goés-Neto¹, Siomar de Castro Soares^{1,2}

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Microbiologia, Imunologia e Parasitologia da Universidade Federal do Triângulo Mineiro, Uberaba, Brasil

* wfr2023@ufmg.br

Resumo. A periodontite é uma doença inflamatória crônica que compromete os tecidos de suporte dos dentes, podendo causar dor, sangramento, perda dentária e impactos relevantes na qualidade de vida. Apesar de sua alta prevalência e de sua importância em saúde pública, sobretudo em países de baixa e média renda, ainda há desafios para compreender plenamente seus mecanismos e para implementar estratégias de diagnóstico e monitoramento mais precisas e acessíveis. Evidências atuais indicam que a doença está fortemente associada à disbiose do microbioma subgengival e à resposta inflamatória persistente do hospedeiro, com possíveis repercussões sistêmicas. Neste cenário, o avanço das tecnologias ômicas (como metagenômica, transcriptômica, proteômica e metabolômica) ampliou a capacidade de investigar, em alta resolução, tanto a composição microbiana quanto as alterações moleculares dos tecidos periodontais. Entretanto, a complexidade e o volume desses dados exigem integração e interpretação robustas. Assim, este trabalho discute, em linguagem de divulgação científica, como a bioinformática translacional e a inteligência artificial podem organizar e integrar dados multi-ômicos, reconhecer padrões biológicos associados à saúde e à doença, e apoiar a identificação de assinaturas moleculares com potencial aplicação em diagnóstico, triagem e acompanhamento clínico. Por fim, são discutidos os principais desafios para a implementação dessas abordagens, incluindo padronização, validação clínica e adaptação a contextos com infraestrutura limitada, destacando oportunidades para soluções escaláveis e de maior impacto populacional.

Palavras-chave: periodontite; bioinformática translacional; multi-ômica; inteligência artificial; biomarcadores.

1. Introdução

A periodontite é uma doença inflamatória crônica que afeta os tecidos de suporte dos dentes e pode levar à dor, sangramento gengival, perda dentária e impactos na qualidade de vida. Trata-se de uma condição altamente prevalente em todo o mundo e que representa um importante desafio para os sistemas de saúde, especialmente em países de baixa e média renda [1–3]. Além das manifestações locais, evidências recentes indicam que a periodontite pode se associar a alterações sistêmicas, reforçando sua relevância em saúde pública.

O desenvolvimento da periodontite está intimamente relacionado ao desequilíbrio do microbioma subgengival. Nesse processo, microrganismos normalmente presentes na cavidade oral passam a se organizar de forma desarmônica, favorecendo espécies associadas à inflamação e reduzindo microrganismos comensais. Esse desarranjo microbiano sustenta uma resposta inflamatória persistente, capaz de promover destruição tecidual e manutenção da doença [4]. Entre os diferentes tipos de amostras estudadas, o fluido crevicular gengival destaca-se por permitir a avaliação simultânea de marcadores da resposta do hospedeiro e de componentes microbianos [5].

Nas últimas décadas, o avanço das tecnologias ômicas revolucionou a investigação da periodontite. Técnicas de sequenciamento genético, análises de expressão gênica, estudos proteômicos e metabolômicos permitem examinar, em grande detalhe, tanto a composição microbiana quanto a resposta molecular dos tecidos periodontais. Nesse cenário, a bioinformática tornou-se essencial para integrar grandes volumes de dados biológicos, identificar padrões ocultos e construir modelos explicativos sobre os mecanismos envolvidos na doença [6].

Paralelamente, a incorporação de métodos de inteligência artificial e aprendizado de máquina ampliou a capacidade de análise desses dados complexos. Essas ferramentas permitem reconhecer perfis moleculares associados à presença, progressão ou resposta ao tratamento da periodontite, abrindo caminho para estratégias mais precisas de diagnóstico, monitoramento e prognóstico. Apesar desses avanços, ainda existem desafios relacionados à padronização metodológica e à validação clínica dessas abordagens, especialmente quando se considera sua aplicação em contextos com recursos limitados [7].

Diante desse panorama, o objetivo deste artigo é apresentar de forma integrada como diferentes abordagens ômicas, associadas à bioinformática e à inteligência artificial, contribuem para o entendimento dos mecanismos da periodontite e para o desenvolvimento de novas possibilidades diagnósticas e translacionais, com potencial impacto em saúde pública.

2. Desenvolvimento

O entendimento da periodontite passou por uma transformação significativa nas últimas décadas. Tradicionalmente, essa condição era interpretada como resultado da ação direta de bactérias patogênicas específicas. Entretanto, estudos recentes demonstraram que a doença está associada principalmente ao desequilíbrio do microbioma subgengival como um todo, no qual comunidades microbianas organizam-se de forma disfuncional, favorecendo processos inflamatórios persistentes e destruição dos tecidos periodontais [1–4]. Essa nova perspectiva ampliou o interesse por abordagens capazes de investigar simultaneamente microrganismos, tecidos do hospedeiro e suas interações.

Nesse contexto, as tecnologias ômicas tornaram-se ferramentas centrais para a pesquisa em periodontite. Técnicas de sequenciamento genético permitem identificar a composição microbiana presente na cavidade oral, enquanto análises de expressão gênica, proteômica e metabolômica possibilitam investigar como células do hospedeiro e microrganismos

respondem às alterações do ambiente periodontal. Essas abordagens revelaram que a periodontite não é apenas uma infecção localizada, mas um processo biológico complexo que envolve múltiplas camadas moleculares em constante interação [2–6].

A grande quantidade de dados gerada por essas tecnologias trouxe um novo desafio: como integrar e interpretar informações tão diversas. É nesse ponto que a bioinformática assume papel essencial. Por meio de ferramentas computacionais, torna-se possível organizar grandes conjuntos de dados, identificar padrões biológicos relevantes e construir modelos que ajudam a compreender os mecanismos envolvidos na instalação e progressão da periodontite. Além disso, a integração de diferentes tipos de dados, como informações microbianas, genéticas e proteicas, permite uma visão mais completa da doença, favorecendo a identificação de potenciais biomarcadores e novos alvos terapêuticos [6,7].

Paralelamente, a aplicação de inteligência artificial e aprendizado de máquina vem ampliando ainda mais o potencial dessas análises. Esses métodos são capazes de reconhecer padrões complexos que dificilmente seriam percebidos por análises tradicionais. No contexto da periodontite, modelos computacionais têm sido empregados para classificar perfis moleculares associados à saúde e à doença, sugerir assinaturas biológicas úteis para diagnóstico precoce e indicar possíveis trajetórias de progressão ou resposta ao tratamento. Dessa forma, a inteligência artificial atua como uma ponte entre dados biológicos complexos e aplicações clínicas mais precisas.

Outro aspecto relevante é a busca por matrizes biológicas de coleta simples e minimamente invasiva. Amostras como saliva e fluido crevicular gengival permitem acessar informações tanto do microbioma quanto da resposta inflamatória do hospedeiro. Isso abre perspectivas para o desenvolvimento futuro de testes rápidos e acessíveis, capazes de auxiliar o diagnóstico e o monitoramento da periodontite em ambientes clínicos e comunitários [5–7].

Apesar dos avanços, ainda existem desafios importantes. A diversidade de metodologias empregadas nos estudos, a falta de padronização de protocolos e a necessidade de validação clínica em diferentes populações limitam, por enquanto, a ampla implementação dessas tecnologias na prática odontológica cotidiana. Superar essas barreiras será essencial para transformar descobertas laboratoriais em soluções realmente aplicáveis, especialmente em contextos com restrições de infraestrutura e recursos.

De modo geral, a integração entre abordagens multi-ômicas, bioinformática e inteligência artificial representa uma das fronteiras mais promissoras da pesquisa em periodontite. Ao permitir uma compreensão mais profunda dos mecanismos biológicos envolvidos e ao abrir caminho para estratégias diagnósticas e prognósticas mais precisas, essas ferramentas têm potencial para impactar positivamente a prevenção, o tratamento e o monitoramento da doença em escala global (Figura 1).

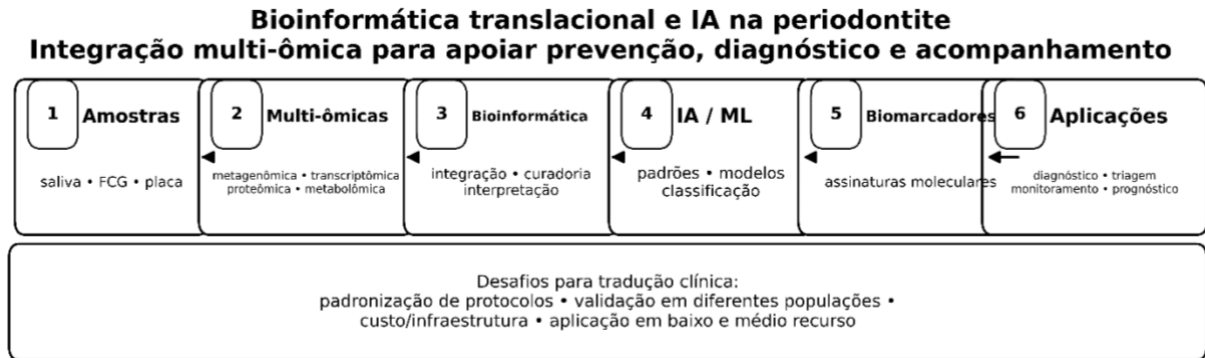


Figura 1. Bioinformática translacional e inteligência artificial aplicadas à periodontite. Fluxo conceitual integrando (1) matrizes biológicas de coleta acessível, como saliva, fluido crevicular gengival e placa subgengival; (2) abordagens multi-ômicas, incluindo metagenômica, transcriptômica, proteômica e metabolômica; (3) processamento e integração dos dados por ferramentas de bioinformática; (4) aplicação de modelos de inteligência artificial e aprendizado de máquina para reconhecimento de padrões biológicos; (5) identificação de biomarcadores moleculares; e (6) desenvolvimento de aplicações clínicas voltadas ao diagnóstico, triagem, monitoramento e prognóstico da periodontite. O painel inferior destaca desafios para a tradução clínica dessas abordagens, incluindo padronização metodológica, validação em diferentes populações, custo e infraestrutura, especialmente em contextos de baixo e médio recurso.

3. Conclusão

A periodontite deixou de ser compreendida apenas como uma infecção localizada para ser reconhecida como uma condição biológica complexa, resultante da interação contínua entre microbioma, resposta imunológica e fatores ambientais. Esse novo entendimento tem sido possível graças ao avanço das tecnologias ômicas e à capacidade da bioinformática em integrar e interpretar grandes volumes de dados biológicos.

A incorporação de métodos de inteligência artificial ampliou ainda mais esse cenário, permitindo a identificação de padrões moleculares associados à saúde e à doença e abrindo perspectivas para diagnósticos mais precoces, monitoramento individualizado e estratégias terapêuticas mais direcionadas. Ao mesmo tempo, o uso de matrizes biológicas de fácil obtenção, como saliva e fluido crevicular gengival, aproxima essas inovações da prática clínica e da realidade dos serviços de saúde.

Apesar dos avanços expressivos, a consolidação dessas abordagens ainda depende de esforços para padronização metodológica, validação em diferentes populações e adaptação a contextos com infraestrutura limitada. Superar esses desafios será essencial para que descobertas laboratoriais se convertam em soluções acessíveis e aplicáveis em larga escala.

Em síntese, a integração entre bioinformática, abordagens multi-ômicas e inteligência artificial representa um caminho promissor para transformar o cuidado em periodontite, fortalecendo ações de prevenção, diagnóstico e acompanhamento, com potencial impacto positivo na saúde bucal e sistêmica da população.

Agradecimentos. A FAPEMIG - “Redes Estruturantes de Pesquisa Científica ou de Desenvolvimento Tecnológico”, no projeto “Aplicações da Genômica no Contexto de OneHealth” (ID: RED-00181-23).

4. Referências





- [1] Abusleme, L; Dupuy, AK; Dutzan, N; Silva, N; Burleson, JA; Strausbaugh, LD; et al. **The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation.** ISME J. Vol. 7, n. 5, p. 1016-25 (2013). doi: <https://doi.org/10.1038/ismej.2012.174>
- [2] LaMonte, MJ; Genco, RJ; Zheng, W; McSkimming, DI; Andrews, CA; Hovey, KM; et al. **Substantial Differences in the Subgingival Microbiome Measured by 16S Metagenomics According to Periodontitis Status in Older Women.** Dent J (Basel). Vol. 6, n. 4 (2018). doi: <https://doi.org/10.3390/dj6040060>
- [3] Genco, RJ; LaMonte, MJ; McSkimming, DI; Buck, MJ; Li, L; Hovey, KM; et al. **The Subgingival Microbiome Relationship to Periodontal Disease in Older Women.** J Dent Res. Vol. 98, n. 9, p. 975-84 (2019). doi: <https://doi.org/10.1177/0022034519854464>
- [4] Camelo-Castillo, AJ; Mira, A; Pico, A; Nibali, L; Henderson, B; Donos, N; et al. **Subgingival microbiota in health compared to periodontitis and the influence of smoking.** Front Microbiol. Vol. 6, p. 119 (2015). doi: <https://doi.org/10.3389/fmicb.2015.00119>
- [5] Field, CA; Gidley, MD; Preshaw, PM; Jakubovics, N. **Investigation and quantification of key periodontal pathogens in patients with type 2 diabetes.** J Periodontal Res. Vol. 47, n. 4, p. 470-8 (2012). doi: <https://doi.org/10.1111/j.1600-0765.2011.01454.x>
- [6] Hu, H; Leung, WK. **Mass Spectrometry-Based Proteomics for Discovering Salivary Biomarkers in Periodontitis: A Systematic Review.** Int J Mol Sci. Vol. 24, n. 19 (2023). doi: <https://doi.org/10.3390/ijms241914871>
- [7] Bostanci, N; Selevsek, N; Wolski, W; Grossmann, J; Bao, K; Wahlander, A; et al. **Targeted Proteomics Guided by Label-free Quantitative Proteome Analysis in Saliva Reveal Transition Signatures from Health to Periodontal Disease.** Mol Cell Proteomics. Vol. 17, n. 7, p. 1392-409 (2018). doi: <https://doi.org/10.1074/mcp.RA117.000454>

23

O desafio de trocar muitos resíduos de uma proteína mantendo-a estável

Autores 23.1

Juliana Rodrigues Pereira Silva , Estela Mariana Guimarães Lourenço ,
Raquel Cardoso de Melo-Minardi , José Miguel Ortega 

Revisão: Diego Mariano , Marcos Antonio Nobrega de Sousa , Rafael Eduardo Oliveira
Rocha , Sheila Cruz Araújo 

Cite este artigo 23.1

Silva J. R. P. *et al.* O desafio de trocar muitos resíduos de uma proteína mantendo-a estável.
BIOINFO. Vol. 6. p.23 (2026). doi: 10.51780/bioinfo-06-23

Saiba mais 23.1

Este artigo está disponível em <https://bioinfo.com.br/o-desafio-de-trocar-muitos-residuos-de-uma-proteina-mantendo-a-estavel/>

O desafio de trocar muitos resíduos de uma proteína mantendo-a estável

Juliana Rodrigues Pereira Silva¹ , Estela Mariana Guimarães Lourenço¹ , Raquel Cardoso de Melo-Minardi¹ , José Miguel Ortega^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* miguel@ufmg.br

Resumo. *Enriquecer proteínas alternativas com aminoácidos essenciais e semiessenciais podem se tornar cruciais para a saúde humana, especialmente em populações vulneráveis (crianças e idosos). A deficiência desses aminoácidos causa graves prejuízos ao desenvolvimento e à saúde, como a desnutrição crônica. Utilizando ferramentas de bioinformática (inicialmente BLOSUM62 e mCSM), iremos enriquecer proteínas com um único tipo ou um conjunto de aminoácidos essenciais mantendo sua estrutura globular. Selecionando proteínas menores e bem resolvidas para simular substituições de aminoácidos, buscando maximizar o enriquecimento sem perder a estabilidade estrutural (dobramento correto). Modelagens e simulações de dinâmica molecular (Gromacs/NAMD3) estão sendo usadas para avaliar a estabilidade destas proteínas. Resultados iniciais com a mioglobina mostram que é possível atingir altas porcentagens de substituição (até 55,3% para leucina) mantendo sua estabilidade até o tempo de simulação testado. O estudo continuará explorando outros indicadores e posições estruturais sensíveis para otimizar o enriquecimento proteico.*

Palavras-chave: *proteínas alternativas; SCP; aminoácidos essenciais; dinâmica molecular.*

1. Introdução

O acesso a quantidades de alimentos adequados e de qualidade é fundamental para manter a integridade metabólica e fisiológica, sendo uma das bases materiais para a própria dignidade da existência humana [1], [2]. Mas a fome e a insegurança alimentar permanecem altas, em grande parte por mudanças climáticas e conflitos sociopolíticos, apesar da capacidade global de produção de alimentos serem suficientes para a necessidade atual [3]. Ainda assim, essa demanda irá aumentar à medida que a população cresce. Neste contexto, proteínas alternativas tem sido uma opção necessária para os desafios de segurança alimentar. Fontes nutricionais de proteínas que não vem da agricultura tradicional, podem ser de origem microbiana, vegetal, de carnes cultivadas ou até mesmo de insetos, representando uma oportunidade de atenuação e resiliência climática [4].

Proteínas microbianas também conhecidas como proteínas de células únicas (“*single-cell proteins*” ou SCP), são provenientes da fermentação de microrganismos unicelulares, como fungos, leveduras, microalgas e bactérias [4]. Sua produção ocupa espaços reduzidos e bem

controlados, como os setores de fermentação industrial, não dependendo das condições climáticas, sazonais ou de fertilidade do solo [4]. Seu potencial em usar subprodutos e resíduos da agroindústria reduz o impacto ambiental e os custos de fabricação, convertendo o problema do descarte em uma solução relevante [4], [5]. Embora os SCPs apresentem muitas vantagens em comparação as fontes de proteína convencionais, existem alguns desafios relacionados a sua produção. A possibilidade de substâncias prejudiciais à saúde humana, como toxinas, ácidos nucleicos (RNA), alérgenos, patógenos, entre outros, torna essencial a adoção de medidas preventivas [4]. Sendo uma estratégia em relação, por exemplo, aos ácidos nucleicos, selecionar um microrganismo mais apropriado ou métodos que os removam ou o degradem [4]. Aspectos sensoriais, como aparência, sabor, textura e aroma podem ser uma barreira, fora o ceticismo do consumidor com um alimento produzido biotecnologicamente, gerando resistência atribuído a preocupações com segurança e matéria-prima utilizada [6].

Em um artigo de 2005 [7], foi proposto expressar a proteína de reserva (vitelogenina) em leveduras recombinantes, a fim de melhorar a oferta proteica para artêmias e, posteriormente, estas artêmias, usadas na alimentação de tilápias. Esse trabalho motivou a idealizar metodologias de bioinformática para enriquecer quaisquer proteínas alternativas com aminoácidos essenciais sem a perda de sua estrutura globular, originando este projeto de doutorado. Aminoácidos essenciais são aminoácidos que animais não são capazes de sintetizar e precisam ser adquiridas pela alimentação [8]. Eles são representados por valina (V), isoleucina (I), leucina (L), fenilalanina (F), metionina (M), triptofano (W), histidina (H), lisina (K) e treonina (T) [8]. Já os semiessenciais ou condicionalmente essenciais, são aminoácidos que são sintetizados, mas que em uma maior demanda como por exemplo, atividade física intensa ou traumas, o corpo não consegue produzir em quantidades suficientes [9]. Sendo estes aminoácidos a tirosina (Y), arginina (R), histidina (H) e cisteína (C) [9].

A relevância destes aminoácidos estaria em ser uma opção mais nutritiva a diversas proteínas alternativas que possam vir a ser usadas na alimentação ou suplementação. Deficiência em aminoácidos essenciais pode causar grandes prejuízos a saúde, comprometendo o progresso físico e intelectual, reduzindo a produtividade econômica [1], [2]. Por exemplo, o atrofiamento de crianças com desnutrição crônica (um quarto das crianças menores de 5 anos em todo o mundo), está associado a um padrão alimentar de proteínas com baixo teor de aminoácido essenciais [10]. É preciso destacar que o consumo de alimentos de origem animal é limitado entre crianças de países em desenvolvimento, sendo menos de 5% em muitos países da África subsaariana [10]. Aminoácidos essenciais também se provaram eficientes na construção muscular deste grupo, tendo benefício anabólico [11]. E para síntese proteica muscular em idosas, uma dose baixa de aminoácidos essenciais enriquecidas com leucina se mostrou tão eficaz quanto uma dose muito maior de *whey protein* [12].

2. Estratégias utilizadas

A pergunta central deste trabalho é até quanto podemos enriquecer proteínas com um único tipo ou um conjunto de aminoácidos essenciais mantendo sua estrutura globular. E para responder essa pergunta estamos usando a bioinformática, testando proteínas que teriam uma

porcentagem de seus resíduos originais substituídos por aminoácidos essenciais ou semiessenciais através de trocas planejadas, mantendo sua característica globular, o que é importante, já que proteínas sem dobramentos ou mal dobradas podem levar a agregações, formação de corpos de inclusões, apoptose ou danos a organelas celulares [13], [14]. Brevemente, as estratégias para enriquecer proteínas com aminoácidos essenciais vão se basear ou no escore da matriz de substituição BLOSUM ou, no caso de mCSM, nas predições de substituições que menos desestabilizam a proteína, e os resultados serão publicados futuramente. A estratégia para definir a proporção de enriquecimento com BLOSUM foi iniciar com escores que representam trocas observadas em proteínas homólogas (trocas conservadas) e com mCSM, com base em diferenças entre ΔG s em posição específica, reduzindo a proporção quando ela não permanecer globular no tempo de simulação testado.

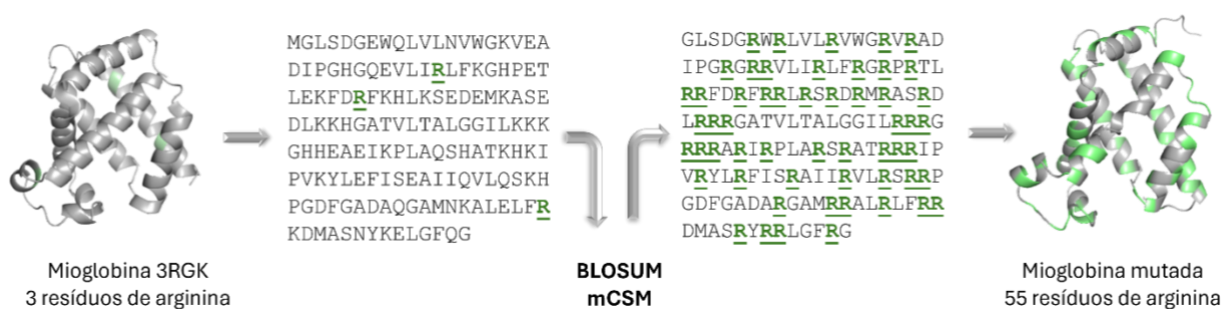


Figura 1. Visão geral da estratégia

No presente momento, alcançamos valores de porcentagem de substituição de resíduos para mioglobina que geraram simulações com estabilização, com RMSD inferior a quatro angstroms, com substituições entre 22 a 43% baseada em BLOSUM62, com resultados iniciais com Gromacs corroborando com NAMD3. Em paralelo, para substituições com base nas predições de estabilidade com mCSM conseguimos percentagens de enriquecimento entre 25 a 55%. Outras métricas como exposição a solvente, globularidade, contatos, estão sendo mensuradas como expressão da instabilidade ao longo da simulação e serão alvo de publicação futura. Preliminarmente, estamos tendo sucesso ao reduzir o tempo de simulação de 200-1000 nanossegundos para apenas 5 nanossegundos realizando a simulação em alta pressão atmosférica, o que nos permitirá avaliar o potencial de enriquecimento de muitas outras proteínas com reduzido tempo computacional. Como a abordagem utilizada gerou um conjunto de mutantes com instabilidades em grau variável, trabalhos futuros podem revelar condições de avaliação rápida de instabilidade.

Agradecimentos: Agências de fomento CAPES, FAPEMIG e as prestadoras de serviços LCC CENAPAD-MG e CELAM.

3. Referências






1. Joint WHO/FAO/UNU Expert Consultation. Protein and amino acid requirements in human nutrition. World Health Organ Tech Rep Ser. 2007;(935):1-265, back cover. PMID: 18330140.

2. Ayala A, Meier BM. A human rights approach to the health implications of food and nutrition insecurity. *Public Health Rev.* 2017 Mar 9;38:10. doi: 10.1186/s40985-017-0056-5. PMID: 29450082; PMCID: PMC5810069.
3. Marc J. Cohen and Ellen Messer. *Food Wars - Conflict, Hunger, and Globalization*, 2023. Oxfam. 2024.
4. Bajić *et al.* Biotechnological Production of Sustainable Microbial Proteins from Agro-Industrial Residues and By-Products. *Foods.* 2022 Dec 25;12(1):107. doi: 10.3390/foods12010107. PMID: 36613323; PMCID: PMC9818480.
5. Kobayashi *et al.* Life-cycle assessment of yeast-based single-cell protein production with oat processing side-stream. *Sci Total Environ.* 2023 May 15;873:162318. doi: 10.1016/j.scitotenv.2023.162318. Epub 2023 Feb 18. PMID: 36805067.
6. Onyeaka *et al.* 2022, 'Single Cell Protein for Foods and Feeds: A Review of Trends', *The Open Microbiology Journal*, vol. 16, no. 1, e187428582206160. <https://doi.org/10.2174/18742858-v16-e2206160>
7. Lim EH, Lam TJ, Ding JL. Single-cell protein diet of a novel recombinant vitellogenin yeast enhances growth and survival of first-feeding tilapia (*Oreochromis mossambicus*) larvae. *J Nutr.* 2005 Mar;135(3):513-8. doi: 10.1093/jn/135.3.513. PMID: 15735086.
8. ROSE WC. The role of the amino acids in human nutrition. *Proc Am Philos Soc.* 1947 Feb;91(1):112-6. PMID: 20293557.
9. Hendrickson *et al.* Conditionally Essential Amino Acid Supplementation Reduces Postoperative Complications and Muscle Wasting After Fracture Fixation: A Randomized Controlled Trial. *J Bone Joint Surg Am.* 2022 May 4;104(9):759-766. doi: 10.2106/JBJS.21.01014. Epub 2022 Mar 14. PMID: 35286282.
10. Semba *et al.* Perspective: The Potential Role of Essential Amino Acids and the Mechanistic Target of Rapamycin Complex 1 (mTORC1) Pathway in the Pathogenesis of Child Stunting. *Adv Nutr.* 2016 Sep 15;7(5):853-65. doi: 10.3945/an.116.013276. PMID: 27633102; PMCID: PMC5015042.
11. Volpi *et al.* Essential amino acids are primarily responsible for the amino acid stimulation of muscle protein anabolism in healthy elderly adults. *Am J Clin Nutr.* 2003 Aug;78(2):250-8. doi: 10.1093/ajcn/78.2.250. PMID: 12885705; PMCID: PMC3192452.
12. Bukhari *et al.* Intake of low-dose leucine-rich essential amino acids stimulates muscle anabolism equivalently to bolus whey protein in older women at rest and after exercise. *Am J Physiol Endocrinol Metab.* 2015 Jun 15;308(12):E1056-65. doi: 10.1152/ajpendo.00481.2014. Epub 2015 Mar 31. PMID: 25827594.
13. Martinotti, S.; Ranzato, E. Targeting the Unfolded Protein Response with Natural Products: Therapeutic Potential in ER Stress-Related Diseases. *Int. J. Mol. Sci.* 2025, 26, 8814. <https://doi.org/10.3390/ijms26188814>.
14. Rao RV, Bredesen DE. Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr Opin Cell Biol.* 2004 Dec;16(6):653-62. doi: 10.1016/j.ceb.2004.09.012. PMID: 15530777; PMCID: PMC3970707.

24

Um passo rumo a novos fármacos contra doenças negligenciadas: mapeamento computacional estrutural da proteína KSRP

Autores 24.1

Leandro Morais de Oliveira , Sheila C. Araújo , Helena Lott Costa ,
Leonardo Henrique Franca de Lima , Raquel C. de Melo-Minardi 

Revisão: Lucianna Helene Santos , Marcos Antonio Nobrega de Sousa , Thiago Jesus Sousa 

Cite este artigo 24.1

Oliveira L.M. de *et al.* **Um passo rumo a novos fármacos contra doenças negligenciadas: mapeamento computacional estrutural da proteína KSRP.** *BIOINFO.* Vol. 6, p.24 (2026). doi: 10.51780/bioinfo-06-24

Saiba mais 24.1

Este artigo está disponível em <https://bioinfo.com.br/um-passo-rumo-a-novos-farmacos-contra-doencas-negligenciadas-mapeamento-computacional-estrutural-da-proteina-ksrp/>

Um passo rumo a novos fármacos contra doenças negligenciadas: mapeamento computacional estrutural da proteína KSRP

Leandro Morais de Oliveira^{1*} , Sheila C. Araújo^{1,2} , Helena Lott Costa¹ , Leonardo Henrique Franca de Lima^{1,2} , Raquel C. de Melo-Minardi^{1,2} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Ciências Exatas e Biológicas, Universidade Federal de São João Del Rei, Campus Sete Lagoas, MG, Brasil

* leandrozno@gmail.com

Resumo. As doenças negligenciadas causadas por cinetoplastídeos, como a Doença de Chagas, afetam milhões de pessoas, e os tratamentos atuais são limitados por toxicidade e baixa eficácia. A busca por novos alvos terapêuticos é, portanto, urgente. A Proteína Ribossomal Específica de Cinetoplastídeos (KSRP) surge como um alvo promissor por ser exclusiva do parasito e essencial para sua sobrevivência. Contudo, suas vulnerabilidades estruturais permanecem pouco exploradas. Para solucionar este problema, este trabalho propõe uma metodologia computacional integrada para identificar e caracterizar sítios de interação na KSRP, visando guiar o desenvolvimento racional de futuros inibidores. A estratégia combina dados estruturais com ferramentas de bioinformática, como docking molecular e análise de flexibilidade, para mapear regiões funcionalmente relevantes. Este estudo fornece hipóteses estruturais para direcionar futuros ensaios de validação.

Palavras-chave: KSRP; Bioinformática Estrutural; Docking; Doenças Negligenciadas.

1. Introdução

Os cinetoplastídeos são um grupo de protozoários responsáveis por doenças de alto impacto na saúde pública, como a Doença de Chagas, a leishmaniose e a doença do sono [1]. A Doença de Chagas, causada pelo *Trypanosoma cruzi*, é a principal parasitose da América Latina e, apesar de sua relevância, continua sendo uma enfermidade negligenciada, com tratamentos limitados, tóxicos e de eficácia variável. Diante deste cenário, a identificação de novos alvos moleculares, que sejam cruciais para o parasito, mas distintos do hospedeiro humano, é uma estratégia fundamental para o desenvolvimento de fármacos mais seguros e eficazes [2].

Nesse contexto, a proteína KSRP (*Kinetoplastid-Specific Ribosomal Protein*) se destaca como um alvo terapêutico de alto potencial. Ela é um componente exclusivo e essencial para a integridade do ribossomo do parasito, a maquinaria celular responsável pela síntese de proteínas [3]. Estudos anteriores demonstraram que a supressão da KSRP é letal para o *Trypanosoma brucei*, o que a valida como um "calcanhar de Aquiles". No entanto, a falta de dados detalhados sobre sua estrutura tridimensional e seus pontos de interação molecular dificulta o planejamento de inibidores [4].

Para superar esse desafio, este trabalho propõe uma abordagem de bioinformática estrutural como uma prospecção inicial. A estratégia utiliza um conjunto de ferramentas computacionais para mapear a superfície da KSRP e gerar hipóteses estruturais que possam guiar, de forma racional, futuros ensaios experimentais, otimizando o processo de descoberta de fármacos. É importante ressaltar que esta abordagem não busca apresentar resultados finais, mas sim construir um alicerce de conhecimento para investigações futuras.

2. Metodologia Computacional Integrada

O ponto de partida da nossa análise foi a estrutura tridimensional da proteína KSRP de *Trypanosoma cruzi*, disponível no Protein Data Bank (PDB ID: 5OPT), que pode ser visualizada na Figura 1. A metodologia foi desenhada como um fluxo de trabalho integrado, combinando diferentes abordagens computacionais para aumentar a confiabilidade das predições.

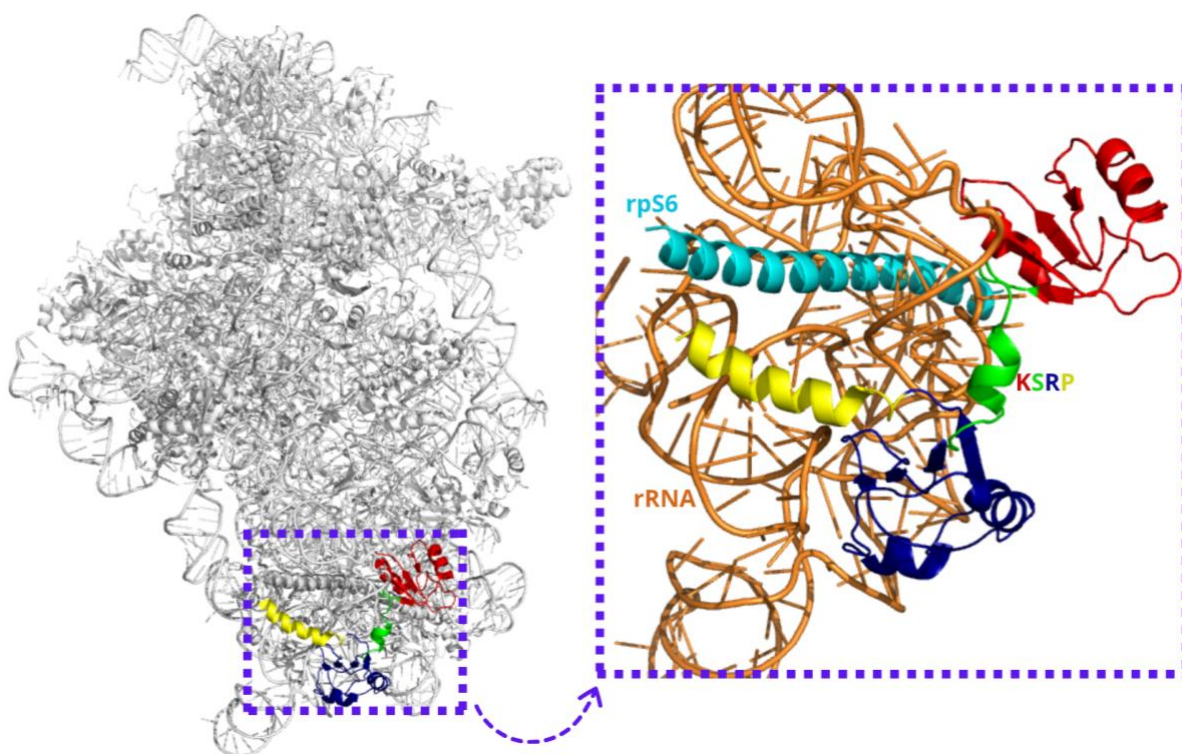


Figura 1. Representação estrutural da localização da KSRP na subunidade ribossomal de *Trypanosoma cruzi* (PDB ID: 5OPT).

Para investigar a KSRP, seguimos uma metodologia em várias etapas. Primeiro, realizou-se uma análise de frequência e conservação para descobrir quais aminoácidos são os mais importantes, ou seja, aqueles que aparecem com frequência em locais de interação e que se mantiveram os mesmos ao longo da evolução. Em seguida, utilizou-se o *docking* molecular, uma simulação que testa as melhores orientações de um ligante em um receptor. Com os modelos prontos, fizemos uma análise dos resíduos em contato para ver de perto quais partes da proteína estavam realmente interagindo com o peptídeo. Também fizemos uma análise da superfície exposta ao solvente para entender quais partes da KSRP ficam na superfície e a aptas

a interagirem com um ligante. Por fim, uma análise de flexibilidade com simulações de dinâmica molecular simplificada nos permitiu observar como a proteína se movimentava, nos dando pistas sobre como a ligação poderia afetar seu funcionamento em outras regiões, que não necessariamente teriam contato direto com o ligante.

3. Conclusão

A principal contribuição deste trabalho é a proposição de uma metodologia computacional para a prospecção de alvos terapêuticos. A aplicação desta abordagem na proteína KSRP sugere a existência de regiões com alto potencial de relevância funcional. A partir das simulações, infere-se que a KSRP possui sítios estruturalmente potencialmente relevantes, que poderiam, hipoteticamente, ser explorados para o desenho de futuros inibidores. As análises de flexibilidade levantam a hipótese de que a interação com moléculas externas possa induzir um efeito alostérico, modulando a função da proteína à distância. Se confirmadas experimentalmente, tais inferências poderiam abrir uma via inovadora para a inibição da KSRP. Portanto, este estudo não apresenta conclusões definitivas, mas sim um conjunto de hipóteses computacionais que servem como um guia racional para o direcionamento de futuros ensaios de validação, buscando otimizar o complexo processo de descoberta de fármacos.

Agradecimentos. Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq (166320/2023-4) e FAPEMIG (APQ-05883-24).





4. Referências

- [1] HANOUT, Wessal; QVIT, Nir. Antiparasitic therapeutic peptidomimetics. In: Peptide and Peptidomimetic Therapeutics. Academic Press, 2022. p. 371-415. <https://doi.org/10.1016/b978-0-12-820141-1.00015-7>
- [2] STANAWAY, Jeffrey D.; ROTH, Gregory. The burden of Chagas disease: estimates and challenges. *Global Heart*, v. 10, n. 3, p. 139-144, 2015. <https://doi.org/10.1016/j.gheart.2015.06.001>
- [3] QUERIDO, Jailson Brito et al. The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein. *Structure*, v. 25, n. 12, p. 1785-1794. e3, 2017. <https://doi.org/10.1016/j.str.2017.09.014>
- [4] HELLMAN, Kristina M. et al. Two trypanosome-specific proteins are essential factors for 5S rRNA abundance and ribosomal assembly in *Trypanosoma brucei*. *Eukaryotic Cell*, v. 6, n. 10, p. 1766-1772, 2007. <https://doi.org/10.1128/ec.00119-07>

25

Bioinformática na sanidade animal: avanços para a saúde única e produção sustentável de organismos aquáticos

Autores 25.1

Rebeca Dias Serafim-Lima , Marcela Laryssa Oliveira Duarte , Victor Silveira Caricatte De Araújo , Angelo Carlo Chaparro Barbanti , Flávia Figueira Aburjaile 

Revisão: Alessandra Lima , Diego Mariano , Rafael Eduardo Oliveira Rocha , Tatiane Senna Bialves 



Cite este artigo 25.1

Serafim-Lima, R. D. *et al.* **Bioinformática na sanidade animal: avanços para a saúde única e produção sustentável de organismos aquáticos**. *BIOINFO*. Vol. 6. p.25 (2026). doi: 10.51780/bioinfo-06-25

Saiba mais 25.1

Este artigo está disponível em <https://bioinfo.com.br/bioinformatica-na-sanidade-animal-avancos-para-a-saude-unica-e-producao-sustentavel-de-organismos-aquaticos/>

Bioinformática na sanidade animal: avanços para a saúde única e produção sustentável de organismos aquáticos

Rebeca Dias Serafim-Lima¹ , Marcela Laryssa Oliveira Duarte², Victor Silveira Caricatte De Araújo³, Angelo Carlo Chaparro Barbanti¹, Flávia Figueira Aburjaile^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Programa de Pós-Graduação em Ciência Animal da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

³ Graduação em Aquicultura da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

*serafim.rebecadias@gmail.com

Resumo. *A aquicultura é responsável pela produção controlada de organismos aquáticos. No entanto, esta atividade vem sofrendo com problemas sanitários causados por bactérias, que por sua vez representam um problema dentro do setor animal. Na produção de animais aquáticos existem patógenos conhecidos por causarem surtos de doenças levando a grandes perdas econômicas no cenário mundial. Além disso, alguns destes patógenos também são relatados em casos de infecções em humanos, representando um alerta quanto a transmissão zoonótica. Adicionalmente têm crescido os relatos de resistência a diferentes classes de antibióticos relacionados à produção aquícola, que representa risco não só aos animais cultivados, mas também para os humanos que entram em contato com bactérias altamente virulentas e resistentes. Neste sentido a bioinformática tem papel fundamental na realização de estudos no âmbito da genômica bacteriana, sobretudo na investigação de características de virulência, resistência antimicrobiana, de plasticidade, além de aspectos biológicos e evolutivos que ajudam a entender as características fenotípicas apresentadas pelos organismos estudados, contribuindo assim para o monitoramento de patógenos importantes. O conhecimento aprofundado do patógeno permite traçar novas perspectivas para o futuro da sanidade animal, dentro da perspectiva da saúde única, como a produção de novas vacinas, alternativas para o controle de infecções nos sistemas produtivos e para promover o trabalho de conscientização da população sobre as consequências do uso indiscriminado de antibióticos na produção animal e seus riscos para a continuidade da atividade e para saúde pública.*

Palavras-chave: *patogenicidade; resistência antimicrobiana; aquicultura.*

1. Introdução

A aquicultura é a atividade responsável pelo cultivo controlado de organismos aquáticos, como peixes, moluscos, crustáceos, anfíbios e algas, para fins de consumo, conservação ou mercado ornamental. Assim como em outras atividades agropecuárias, a intensificação dos sistemas produtivos tem levado ao aumento da propagação de doenças. Em especial, as infecções bacterianas têm sido responsáveis por grandes perdas econômicas [1], devido aos seus impactos

sobre a sanidade animal. Nesse sentido, o controle de bactérias patogênicas é crucial para a manutenção e continuidade da atividade aquícola.

O controle de surtos de infecções bacterianas começa com ações preventivas, que incluem o estudo do patógeno. Os fatores de virulência, por exemplo, são determinantes para que as bactérias causem doenças e podem atuar individualmente ou conjuntamente no processo de infecção [2]. Entre esses fatores estão os sistemas de secreção, adesinas, toxinas, entre outros. Por essa razão, a investigação genômica desempenha papel fundamental na compreensão dessas características patogênicas. A genômica comparativa, por sua vez, permite a análise integrada de dois ou mais genomas, a fim de identificar características semelhantes ou distintas [3], aprofundando assim o entendimento sobre a biologia de patógenos de interesse, como os relatados na aquicultura.

Dentre as áreas de atuação da Bioinformática, destaca-se a investigação de bactérias patogênicas de animais aquáticos, especialmente aquelas associadas a doenças em peixes, que representam a maior parcela da produção aquícola brasileira. Além do foco na prevenção e na sanidade animal, há também a preocupação com a saúde única, visto que existem diversos patógenos de animais aquáticos relatados em casos de infecções em humanos, o que gera grande alerta à saúde pública.

2. A bioinformática no monitoramento de patógenos

Dentre as abordagens utilizadas para monitoramento de patógenos está a investigação de genes de virulência e resistência que são fundamentais no contexto do surgimento das infecções e na sua prevalência. Nesse contexto, realiza-se a caracterização genômica de patógenos a partir do sequenciamento do genoma completo, o controle da qualidade dos dados, identificação taxonômica, estudos filogenômicos para avaliar diversidade e eventos evolutivos, como também estudos de pangenoma, panviruloma e panresitoma, estudos de mobiloma, dentre outras análises de bioinformática. Estas abordagens, permitem compreender características expressas fenotipicamente em experimentos de bancada e a campo. A investigação de bactérias ubíquas de ambientes aquáticos tem contribuído sobretudo para o entendimento da plasticidade genômica, em especial quanto à aquisição de genes relacionados à sobrevivência em ambientes com exposição a metais e a antimicrobianos. Além disso, a investigação genômica de bactérias isoladas dentro do território brasileiro ajuda a entender sobre a evolução genômica não só localmente, mas possibilita comparar o perfil genômico de cepas brasileiras em relação a cepas de outras localidades, que é importante para aumentar o conhecimento a respeito dos fatores presentes no ambiente que influenciam a aquisição de características de patogenicidade por um determinado grupo.

Em virtude disso, o Laboratório de Bioinformática Integrativa (LBI/UFMG) busca manter uma rede de colaboradores de diferentes universidades composta por pesquisadores das áreas de biologia, veterinária, aquicultura, microbiologia, genética entre outras áreas do conhecimento, para reunir informações relevantes para o estudo dos organismos de interesse. Além disso, o nosso grupo de pesquisa é composto por alunos de bioinformática da UFMG com experiência em diferentes áreas, compondo uma rede de conhecimento diversa e integrada a

estudantes de outros programas. Dentre os patógenos causadores de doenças em peixes alvo do grupo de pesquisa estão bactérias dos gêneros *Aeromonas* [4]. Os autores identificaram uma grande variedade de genes de virulência, além de genes de resistência a metais pesados e a diferentes classes de antibióticos em genomas brasileiros de *Aeromonas jandaei* isoladas de tilápia do Nilo, sugerindo aquisição por transferência horizontal de genes que conferem vantagens adaptativas a este patógeno. Além disso, já foram publicados trabalhos de investigação genômica de outras espécies dos gêneros *Aeromonas* [5] e *Plesiomonas* [6] juntamente com outros gêneros que afetam a sanidade de animais aquáticos contribuindo assim para compreensão das infecções bacterianas na aquicultura. A figura 1 demonstra um exemplo de fluxo de análises de patógenos em peixes.



Figura 1. (1) Tanque de produção de peixes, (2) coleta de indivíduos para isolar o patógeno, (3) isolamento bacteriano, (4) extração do DNA do patógeno, (5) sequenciamento do genoma completo, (6) investigação do genoma para identificar sequências de interesse na área (fármacos e vacinas), (7) estudos comparativos, (8) divulgação das informações para a comunidade científica e para o setor produtivo.

3. Conclusão

Conclui-se que o monitoramento de patógenos em atividades agropecuárias tornou-se uma necessidade urgente para o futuro da produção sustentável, sobretudo em países que possuem posição relevante neste cenário, como é o caso do Brasil. Dessa forma, os trabalhos desenvolvidos no laboratório são extremamente relevantes e úteis para o estudo de patógenos de interesse aquícola, veterinário e no contexto de saúde única. Os achados encontrados pelo grupo de pesquisa corroboram na prevenção de surtos de doenças em cultivos de animais aquáticos para o combate a “pandemia silenciosa” de resistência antimicrobiana e reforçam as perspectivas futuras de controle de patógenos sob a perspectiva de saúde única.

Agradecimentos. Agradecimentos. CAPES, CNPq, FAPEMIG e Rede RECOM.





4. Referências

- [1] MAEZONO, Miho; NIELSEN, Rasmus; BUCHMANN, Kurt; NIELSEN, Max. The current state of knowledge of the economic impact of diseases in global aquaculture. *Reviews in Aquaculture*, v. 17, n. 3, p. e70039, 2025. DOI: <https://doi.org/10.1111/raq.70039>
- [2] WU, Hsing-Ju; WANG, Andrew H. J.; JENNINGS, Michael P. Discovery of virulence factors of pathogenic bacteria. *Current Opinion in Chemical Biology*, v. 12, n. 1, p. 93-101, 2008. DOI: <https://doi.org/10.1016/j.cbpa.2008.01.023>
- [3] WEI, L.; LIU, Y. ; DUBCHAK, I.; SHON, J.; PARK, J. Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics*, v. 35, n. 2, p. 142–150, abr. 2002. DOI: 10.1016/S1532-0464(02)00506-3. DOI: [https://doi.org/10.1016/S1532-0464\(02\)00506-3](https://doi.org/10.1016/S1532-0464(02)00506-3)
- [4] DUARTE, M. L. O.; *et al.* In silico characterization of resistance and virulence genes in *Aeromonas jandaei* strains isolated from *Oreochromis niloticus* in Brazil. *Microorganisms*, v. 13, n. 5, p. 1094, 08 May 2025. DOI: <https://doi.org/10.3390/microorganisms13051094>
- [5] MAIA, J. C. S.; *et al.* Genomic characterization of *Aeromonas veronii* provides insights into taxonomic assignment and reveals widespread virulence and resistance genes throughout the world. *Antibiotics*, v. 12, n. 6, p. 1039, 2023. DOI: <https://doi.org/10.3390/antibiotics12061039>
- [6] ABURJAILE, F. F.; *et al.* Genomic islands and mobilome demonstrate a potential pathogen, the *Plesiomonas shigelloides*, as in Nile tilapia (*Oreochromis niloticus*). *The Microbe*, v. 8, p. 100441, 2025. DOI: <https://doi.org/10.1016/j.microb.2025.100441>

26

Desafios e impactos na modulação da microbiota intestinal e no tratamento de doenças inflamatórias intestinais

Autores 26.1

Eduarda Guimarães Sousa , Éric Guédon , Siomar de Castro Soares ,
Vasco Azevedo 

Revisão: Alessandra Lima , Diego Mariano , Marcos Antonio Nobrega de Sousa , Rafael
Eduardo Oliveira Rocha 





Cite este artigo 26.1

Sousa E. G. *et al.* **Desafios e impactos na modulação da microbiota intestinal e no tratamento de doenças inflamatórias intestinais.** BIOINFO. Vol. 6. p.26 (2026). doi: 10.51780/bioinfo-06-26

Saiba mais 26.1

Este artigo está disponível em <https://bioinfo.com.br/desafios-e-impactos-na-modulacao-da-microbiota-intestinal-e-no-tratamento-de-doencas-inflamatorias-intestinais/>

Desafios e impactos na modulação da microbiota intestinal e no tratamento de doenças inflamatórias intestinais

Eduarda Guimarães Sousa ¹ , Éric Guédon ² , Siomar de Castro Soares ¹ ,
 Vasco Azevedo ^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² STLO, INRA, *Agrocampus Ouest*, Rennes, France

* vascoariston@ufmg.br

Resumo. *A microbiota intestinal é essencial para a saúde, influenciando processos digestivos, metabólicos e imunológicos. Quando desestabilizada por fatores como antibióticos ou quimioterapia, podem surgir condições como disbiose e mucosite, que comprometem a mucosa intestinal. Este estudo investiga o potencial terapêutico de *Bacillus velezensis* 1273, uma bactéria probiótica resistente a pH ácido e sais biliares, capaz de liberar vesículas extracelulares (EVs) com biomoléculas bioativas. A pesquisa integra abordagens multiômicas — genômica, transcriptômica, proteômica e metabolômica — para compreender os mecanismos moleculares e metabólicos associados às propriedades imunomoduladoras do probiótico e de suas EVs. Essa estratégia, ancorada na probiogenômica, conecta dados *in vitro*, *in vivo* e *in silico*, traduzindo o conhecimento acadêmico em possíveis aplicações terapêuticas. O trabalho representa um avanço científico e social ao propor alternativas seguras e acessíveis para o tratamento de doenças inflamatórias intestinais, fortalecendo a ponte entre ciência, tecnologia e inovação biotecnológica.*

Palavras-chave: *Microbiota intestinal; *Bacillus velezensis*; imunomodulação; vesículas extracelulares; disbiose.*

1. Introdução

A microbiota intestinal, composta por trilhões de microrganismos, é fundamental para o metabolismo e a imunidade do hospedeiro. O desequilíbrio dessa comunidade, especialmente após o uso de antibióticos ou quimioterapia, resulta em disbiose intestinal e inflamação. A mucosite induzida por quimioterapia é uma dessas consequências, afetando gravemente a integridade da mucosa intestinal e a qualidade de vida dos pacientes [1].

Nesse contexto, o uso de probióticos surge como uma estratégia promissora para restaurar o equilíbrio microbiano e reduzir a inflamação. *Bacillus velezensis* 1273 destaca-se por sua capacidade de esporulação, que garante sobrevivência em condições adversas e potencial de colonização intestinal. Além disso, suas vesículas extracelulares (EVs) carregam proteínas, lipídios, RNAs e metabólitos bioativos capazes de atuar na comunicação celular e na modulação imune [2–5].

O projeto busca compreender os mecanismos terapêuticos de *Bacillus velezensis* e de suas EVs por meio de uma caracterização probiogenômica. A integração de diferentes camadas

analíticas — genômica, transcriptômica, proteômica e metabolômica — permite compreender como genes e metabólitos se conectam para promover a homeostase intestinal e reduzir processos inflamatórios.

2. Desenvolvimento

A força desta pesquisa está em sua abordagem integrativa. Ela une microbiologia, biotecnologia e bioinformática, exigindo domínio de técnicas experimentais e computacionais. Essa jornada científica não envolve apenas experimentos, mas também aprendizado contínuo e colaboração entre grupos de pesquisa, mostrando que a ciência se constrói de forma coletiva.

A probiogenômica ocupa o centro dessa integração, permitindo relacionar a informação genética de *Bacillus velezensis* com seus efeitos observados *in vivo* e *in vitro*. A análise genômica fornece a base da segurança da cepa, confirmando a ausência de genes de virulência e a presença de vias relacionadas à produção de compostos bioativos. Por sua vez, a transcriptômica, obtida por RNA-Seq, revela como a bactéria responde a estímulos do hospedeiro e a condições de estresse, elucidando mecanismos de regulação gênica e imunomodulação. A proteômica, via espectrometria de massa, identifica proteínas envolvidas em adesão, comunicação e resposta inflamatória, especialmente nas EVs — pequenas estruturas que carregam sinais biológicos capazes de modular o sistema imune mesmo sem a presença da célula bacteriana. A metabolômica complementa essas análises ao identificar metabólitos com propriedades anti-inflamatórias e antioxidantes, ligando os efeitos moleculares às respostas fisiológicas observadas. Essa integração multiômica oferece uma visão abrangente dos mecanismos de ação de *Bacillus velezensis*, permitindo correlacionar dados genéticos, proteicos e metabólicos a efeitos terapêuticos concretos.

Mais do que resultados técnicos, esta pesquisa reflete a vivência de aprender, errar e dominar diferentes metodologias, do cultivo bacteriano ao sequenciamento e à análise de dados de alta complexidade. Trabalhar com técnicas tão diversas requer resiliência e, sobretudo, colaboração. As parcerias acadêmicas foram fundamentais para o avanço deste projeto, unindo competências em genômica, metabolômica e bioinformática em um esforço multidisciplinar. Essa rede de colaboração permitiu ir além do conhecimento básico e pensar em aplicações práticas. O estudo de *Bacillus velezensis* e de suas EVs abre caminho para o desenvolvimento de produtos biotecnológicos baseados em probióticos e pós-bióticos, as quais são formulações que associam eficácia, estabilidade e segurança. Assim, o trabalho vai da bancada ao mundo real, transformando o conhecimento em uma oportunidade concreta de impacto social, especialmente na área de saúde intestinal e terapias complementares ao tratamento oncológico o qual se encontra simplificado na figura 1.

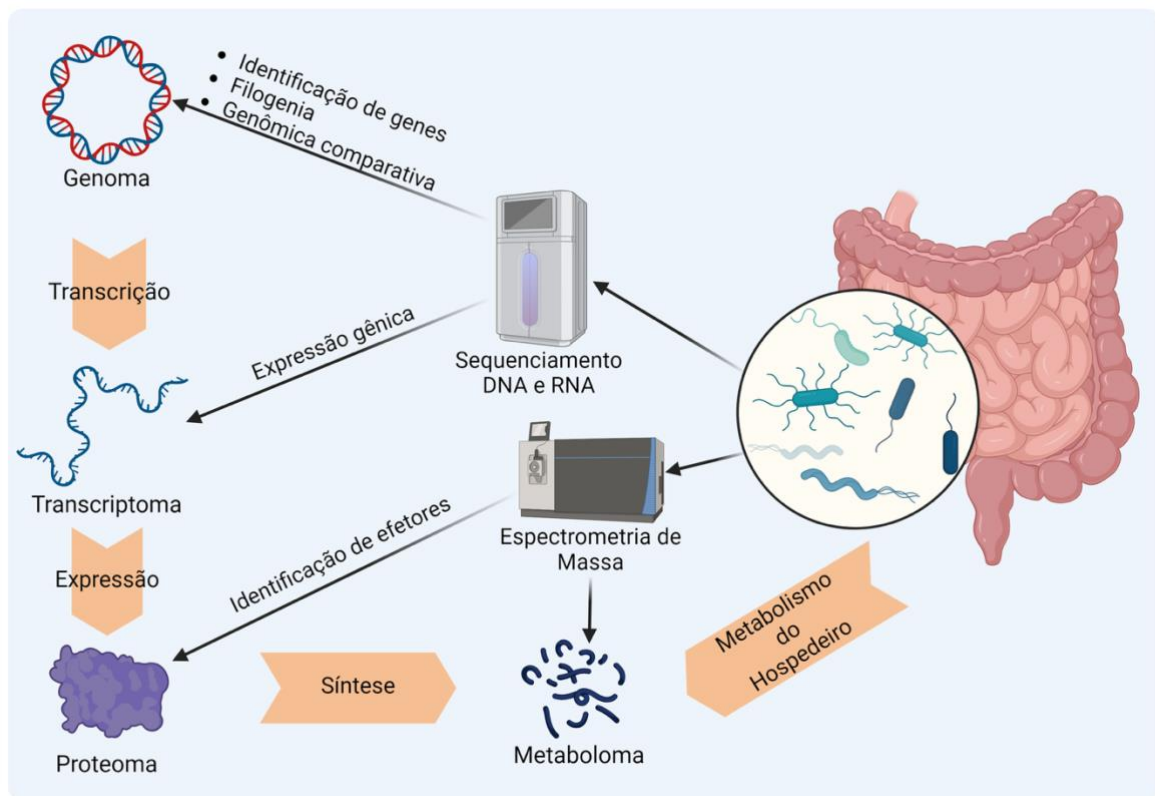


Figura 1. Integração probiogenômica e multiômica aplicada à caracterização de *B. velezensis* 1273. Fonte: próprio autor.

3. Conclusão

O estudo de *Bacillus velezensis* 1273 e de suas vesículas extracelulares representa um avanço significativo na interface entre microbiologia, bioinformática e saúde. A abordagem probiogenômica, sustentada por dados multiômicos, possibilita compreender de forma ampla os mecanismos de modulação imunológica e de restauração da microbiota intestinal. Mais do que uma pesquisa de bancada, este projeto demonstra a importância das parcerias acadêmicas e do aprendizado interdisciplinar. Ele reafirma que a ciência, quando guiada pela curiosidade e colaboração, tem potencial de gerar conhecimento transformador — e de aproximar o universo acadêmico da inovação biotecnológica voltada ao bem-estar humano.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

4. Referências

1. De Jesus LCL, Drumond MM, de Carvalho A, Santos SS, Martins FS, Ferreira Ê, et al. Protective effect of *Lactobacillus delbrueckii* subsp. *Lactis* CIDCA 133 in a model of 5-Fluorouracil-Induced intestinal mucositis. *J Funct Foods*. 2019 Feb 1;53:197–207.
2. Sousa EG, Campos GM, Viana MVC, Gomes GC, Rodrigues DLN, Aburjaile FF, et al. The research on the identification, taxonomy, and comparative genomics analysis of nine *Bacillus velezensis* strains significantly contributes to microbiology, genetics, bioinformatics, and biotechnology. *Front Microbiol* [Internet]. 2025 Mar 19 [cited 2025 Oct 13];16:1544934. Available from: <https://quast.sourceforge.net/>



3. Li X, Gao X, Zhang S, Jiang Z, Yang H, Liu XD, et al. Characterization of a *Bacillus velezensis* with antibacterial activity and inhibitory effect on common aquatic pathogens. *Aquaculture* [Internet]. 2020 Jun 30 [cited 2025 Oct 13];523:735165. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0044848619324925>
4. Zhang X, Wang Y, E Q, Naveed M, Wang X, Liu Y, et al. The biological activity and potential of probiotics-derived extracellular vesicles as postbiotics in modulating microbiota-host communication. *J Nanobiotechnology* [Internet]. 2025 Dec 1 [cited 2025 Oct 13];23(1):349. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12082936/>
5. Shin HH, Kim JH, Jung YJ, Kwak MS, Sung MH, Imm JY. Postbiotic potential of *Bacillus velezensis* KMU01 cell-free supernatant for the alleviation of obesity in mice. *Heliyon* [Internet]. 2024 Mar 15 [cited 2025 Oct 13];10(5):e25263. Available from: <https://www.sciencedirect.com/science/article/pii/S2405844024012945>

27

Ilhas genômicas: pontos isolados em um oceano de informações

Autores 27.1

Diego Rodrigues , Vasco Azevedo , Siomar Soares , Flávia Aburjaile 

Revisão: Ariany Rosa Gonçalves , Diego Mariano , Rafael Eduardo Oliveira Rocha ,
Sheila Cruz Araújo 

Cite este artigo 27.1

Rodrigues D. *et al.* **Ilhas genômicas: pontos isolados em um oceano de informações.** BIOINFO.
Vol. 6. p.27 (2026). doi: 10.51780/bioinfo-06-27

Saiba mais 27.1

Este artigo está disponível em <https://bioinfo.com.br/ilhas-genomicas-pontos-isolados-em-um-oceano-de-informacoes/>

Ilhas genômicas: pontos isolados em um oceano de informações

Diego Rodrigues¹ , Vasco Azevedo¹ , Siomar Soares^{1,2} , Flávia Aburjaile^{1,3*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² Departamento de Desenvolvimento de Pesquisa e Inovação Tecnológica, Escola de Veterinária, Universidade Federal do Triângulo Mineiro, Uberaba, Brasil

³ Departamento de Medicina Veterinária Preventiva, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* faburjaile@gmail.com

Resumo. *O avanço exponencial das tecnologias de sequenciamento tem gerado uma quantidade crescente de dados genômicos, exigindo ferramentas de bioinformática especializadas e capazes de extrair conhecimento biológico destes. Entre os elementos mais intrigantes estão as ilhas genômicas. Essas ilhas são fundamentais para compreender a evolução e a adaptação bacteriana, mas sua detecção continua sendo um desafio metodológico. Diferentes organismos exibem assinaturas genômicas variadas, tornando imprecisos os critérios universais de predição. Essa variabilidade compromete a padronização entre ferramentas e aumenta as taxas de falso positivo ou negativo. Além disso, muitos métodos clássicos baseiam-se em heurísticas e observações empíricas que não capturam adequadamente a diversidade dos mecanismos de transferência horizontal. Nesse cenário, abordagens baseadas em aprendizado de máquina emergem como alternativas promissoras. Nesse contexto, o desenvolvimento da ferramenta GIPSy (Genomic Island Prediction Software) propõe superar parte dessas dificuldades, unindo evidências composicionais, funcionais e comparativas em um sistema integrado. A nova versão em desenvolvimento busca não apenas aprimorar a acurácia das predições, mas também consolidar o conhecimento acumulado sobre a plasticidade genômica bacteriana, contribuindo para uma compreensão mais robusta da evolução microbiana.*

Palavras-chave: *mobilômica; elementos móveis; patogenicidade.*

1. Introdução

O rápido avanço tecnológico gerado pela humanidade veio com ganhos igualmente majestosos. Nos mais diversos campos da ciência uma quantidade cada vez maior de dados é produzida a um custo que se reduz a cada passo. Isso não poderia ser diferente nos campos da genética e bioinformática. Hoje em dia, compreender um genoma é uma tarefa árdua não apenas pela complexidade natural que é intrínseca à biologia, mas também pelo gigantesco volume de dados produzido. Isso tudo leva a uma demanda constante no que tange ao desenvolvimento de novos recursos e ferramentas computacionais para revelar informações implícitas no interior de genomas inteiros.

Em termos gerais, um dos campos de estudo existentes dentro das ciências ômicas (em especial, genômica bacteriana) é a plasticidade genômica, ou seja, capacidade de adaptação e modulação apresentada por determinados organismos. Essa característica é por vezes

demarcada pela habilidade singular de agregar material genético exógeno ao próprio DNA e produzir proteínas que a princípio não estariam presentes no repertório genético de um organismo, mas que na condição atual promovem alguma vantagem adaptativa. Com o passar do tempo e com o avanço das gerações, determinadas regiões do genoma podem acumular uma quantidade significativa desses elementos e se diferenciar estruturalmente das demais regiões que as tangenciam. Por vezes essas regiões se tornam tão distintas que todo o conteúdo genético em seu interior aparenta ser advindo de organismos completamente não correlacionados e tornam-se pontos isolados no genoma, conforme mostrado na Figura 1 [1]. Devido a essas diferenças, essas regiões de aquisição gênica são comumente conhecidas como ilhas genômicas e são cruciais para o entendimento do perfil e nicho biológico de alguns organismos [2].

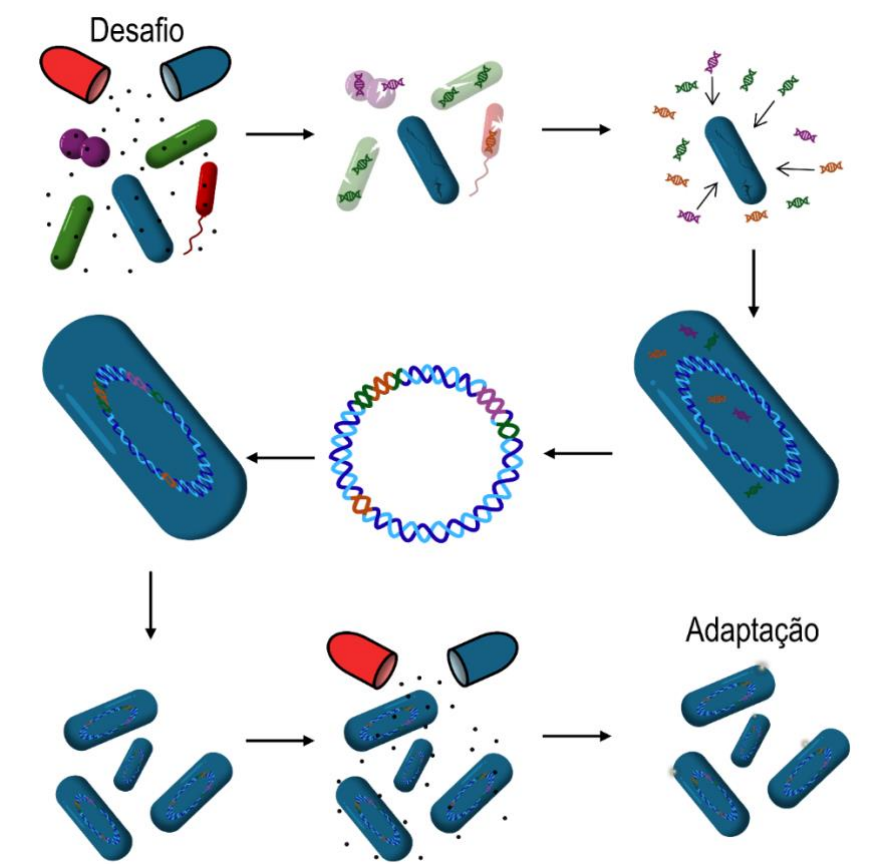


Figura 1. Esquematização simplificada da aquisição de material genético exógeno durante a transformação bacteriana.

Ao longo dos anos, diferentes grupos se propuseram a desenvolver ferramentas de bioinformática capazes de inferir a presença e localização de ilhas genômicas. Cada novo software contribuiu significativamente, permitindo que a comunidade científica aprendesse a identificar diferentes padrões evolutivos como possíveis indicadores de mobilidade gênica, bem como marcadores específicos de determinados grupos de organismos. Todavia, a lacuna metodológica de padronização dos métodos permanece aberta. Isso porque ainda existe uma alta variabilidade entre os resultados obtidos por diferentes softwares que colocam pesos desiguais sobre distintos padrões de assinatura para genes exógenos.

Nesse contexto, o desenvolvimento de um novo preditor de ilhas genômicas deve buscar reunir as informações anteriormente trazidas à luz por outras ferramentas desenvolvidas com o mesmo propósito. Ainda assim, é importante que revise de forma crítica e científica os fundamentos que propiciam a detecção desses elementos genômicos.

2. Desafios conceituais e metodológicos

Mesmo que por via de regra, o termo “ilha genômica” já seja amplamente referenciado por campos da genética e microbiologia, sua definição permanece situacional e fluida. Dessa forma, diferentes autores e ferramentas abordam a temática de forma distinta, dando mais ou menos importância a determinados critérios de predição a depender do contexto biológico. Isso não é apenas considerado normal, como também é necessário para atender às demandas específicas apresentadas por diferentes organismos. Por exemplo, para determinados grupos a presença de alterações demarcadas de conteúdo GC em regiões específicas é um forte indicador de mobilidade genética [2]. Todavia, em algumas bactérias, essas irregularidades podem ser relativamente constantes ao longo do genoma [3]. Se por um lado, ter noção dessas características individuais dificulta a propagação de erros pela predição de falsos positivos, por outro, isso também significa que existe uma redução na precisão de preditores que se baseiam majoritariamente nesse dado para realizar sua inferência.

Essas ambiguidades dificultam o desenvolvimento de softwares com esse propósito ao impor um desafio singular. Afinal, é extremamente difícil desenvolver um algoritmo capaz de capturar resultados verdadeiros, quando o próprio conceito de ilha genômica se torna ambíguo.

Nesse contexto, sabe-se que os primeiros softwares capazes de detectar esses elementos eram, em seu cerne, baseados em observações empíricas. Portanto, mesmo que funcionassem bem a nível de linhagem ou espécie, podiam apresentar alta taxa de erro em contextos taxonômicos mais abrangentes. Além disso, nem todos os mecanismos de transferência horizontal de genes são ubíquos ou totalmente descritos, o que adiciona um ponto a mais de complexidade na interpretação dos resultados.

Dessa forma, modelos de aprendizado de máquina têm se apresentado como uma solução alternativa, por tratar os padrões de assinatura de ilhas genômicas como *features*, e utilizar abordagens baseadas em modelos estatísticos. Todavia, estes métodos estatísticos também apresentam suas limitações conceituais em relação à plausibilidade biológica.

3. Contribuição

De fato, a predição de um elemento tão singular quanto uma ilha genômica é uma tarefa simultaneamente complexa e simplória. Ainda não existe uma abordagem metodológica que consiga garantir a predição desse elemento genômico com precisão perfeita. Todavia, novas abordagens são importantes para atualizações nesse campo. Nesse contexto, a ferramenta *Genomic Island Prediction Software* (GIPSy) [4] visa contribuir com a sociedade científica, permitindo que o conhecimento gerado pelo estudo da plasticidade genômica possa ser acessível a todo o público de pesquisadores interessados. Dessa forma, existe a proposta da

publicação de uma nova versão dessa consolidada ferramenta, denominada GIPSy2 que visa suprir os desafios em aberto.

De forma mais ampla, o projeto representa um esforço para aproximar a bioinformática de sua essência, transformando dados genômicos em conhecimento biológico significativo. Espera-se ainda que as melhorias implementadas no GIPSy2 fortaleçam o entendimento sobre os mecanismos de mobilidade gênica e sirvam de base para o desenvolvimento de futuras abordagens integrativas.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

4. Referências

- [1] Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep* 2001;2:376–81. <https://doi.org/10.1093/embo-reports/kve097>.
- [2] Zhang R, Ou H-Y, Gao F, Luo H. Identification of Horizontally-transferred Genomic Islands and Genome Segmentation Points by Using the GC Profile Method. *Curr Genomics* 2014;15:113–21. <https://doi.org/10.2174/1389202915999140328163125>.
- [3] Bohlin J, Snipen L, Hardy SP, Kristoffersen AB, Lagesen K, Dønsvik T, et al. Analysis of intra-genomic GC content homogeneity within prokaryotes. *BMC Genomics* 2010;11:464. <https://doi.org/10.1186/1471-2164-11-464>.
- [4] Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, et al. GIPSy: Genomic island prediction software. *Journal of Biotechnology* 2016;232:2–11. <https://doi.org/10.1016/j.jbiotec.2015.09.008>.

28

Como a probiogenômica pode revelar novos probióticos: o exemplo de *Streptococcus thermophilus*

Autores 28.1

Juan Carlos Ariute , Vasco Azevedo , Eric Guédon , Flávia Figueira Aburjaile 

Revisão: Alessandra Lima , Ariany Rosa Gonçalves , Diego Mariano , Marcos Antonio Nobrega de Sousa , Sheila Cruz Araújo 



Cite este artigo 28.1

Aurite J. C. *et al.* Como a probiogenômica pode revelar novos probióticos: o exemplo de *Streptococcus thermophilus*. BIOINFO. Vol. 6. p.28 (2026). doi: 10.51780/bioinfo-06-28

Saiba mais 28.1

Este artigo está disponível em <https://bioinfo.com.br/como-a-probiogenomica-pode-revelar-novos-probioticos-o-exemplo-de-streptococcus-thermophilus/>

Como a probiogenômica pode revelar novos probióticos: o exemplo de *Streptococcus thermophilus*

Juan Carlos Ariute^{1,2} , Vasco Azevedo¹, Eric Guédon², Flávia Figueira Aburjaile^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² UMR 1253 STLO - *Science et Technologie du Lait et d'oeuf*, INRAE, Rennes, França

* faburjaile@ufmg.br

Resumo. Os probióticos são descritos como microrganismos capazes de conferir efeitos benéficos à saúde e bem-estar de um indivíduo, sendo consumidos em formulações específicas ou presente naturalmente em alimentos fermentados. As bactérias ácido-láticas representam um grupo heterogêneo de bactérias que são conhecidas por suas propriedades probióticas. Entretanto, o potencial probiótico de uma espécie não é inerente a todas as linhagens, e mesmo linhas probióticas de uma mesma espécie podem apresentar efeitos distintos. Sendo assim, a probiogenômica representa um domínio da genômica que se concentra em não apenas elucidar os mecanismos de probiose a nível molecular, mas também prever novas linhagens probióticas em potencial a partir de seus respectivos genomas, quando disponíveis. O presente estudo foca na prospecção do potencial probiótico de *Streptococcus thermophilus*, uma BAL utilizada amplamente como cultura iniciadora na indústria de laticínios. A anotação funcional de uma coleção diversa de genomas de *S. thermophilus* revela que os genes acessórios e únicos são responsáveis pela maior parte do potencial probiótico. Além disso, de um ponto de vista de biossegurança, genes descritos como fatores de virulência na verdade podem atuar positivamente num contexto de probiose, representando um viés dos bancos de dados. Finalmente, observa-se que a espécie possui grande potencial probiótico e linhagens podem ser triadas para utilização em produtos probióticos específicos.

Palavras-chave: genômica; probióticos; laticínios.

1. Introdução

O estilo de vida de um indivíduo possui impacto direto nos quadros patológicos que podem ser desenvolvidos pelo mesmo. A Organização Mundial de Saúde (OMS) estipulou em 2002 que, até 2020, pelo menos $\frac{3}{4}$ de todas as mortes no mundo seriam causadas por doenças crônicas, sendo necessários esforços governamentais para que medidas econômicas e sociais minimizem os impactos na saúde pública [1]. Desta mesma forma, associado a hábitos de vida mais saudáveis, a alimentação é tida como um meio eficaz para a prevenção e melhoramentos dos impactos negativos à saúde desencadeados por tais doenças [2]. Dentre eles, os produtos fermentados são considerados alimentos processados funcionais cujos benefícios à saúde não estão restritos à aquisição de nutrientes e energia, se estendendo a alívio de quadros diarreicos, aumento de diversidade da microbiota intestinal, melhora na digestão e inflamação locais, e outros [3–5].

A fermentação é em grande parte decorrente da ação de microrganismos específicos que ocorrem naturalmente no ambiente e/ou são adicionados durante o preparo, tais como as bactérias ácido-láticas (BALs), um grupo heterogêneo de bactérias Gram-positivas que são amplamente aplicadas nas indústrias alimentícia e farmacêutica pelo seu alto grau de segurança (GRAS/QPS). Muitas linhagens bacterianas dentre as espécies de BALs são consideradas como probióticos, microrganismos vivos que, quando consumidos em doses adequadas diárias, podem conferir benefícios à saúde de seus hospedeiros [6]. Ressalta-se, todavia, que apesar de ser frequentemente observado em certas espécies, o efeito probiótico é linhagem específico; ademais, duas linhagens probióticas de uma mesma espécie podem apresentar efeitos distintos por possuírem mecanismos diferentes de ação.

Na bioinformática, a probiogenômica é um domínio de estudos genômicos voltados para associação de características responsáveis pelo efeito probiótico observado nestas linhagens [7]. Adicionalmente, a probiogenômica também pode avaliar a biossegurança de uma linhagem probiótica buscando por possíveis genes de resistência e virulência, para assegurar o consumo dos isolados em produtos. Tendo em vista que o fenótipo é intrinsecamente um produto do arcabouço genômico de um organismo, a aplicação de tais “pipelines” a um conjunto em larga escala de genomas com linhagens que pertencem a um mesmo grupo taxonômico pode auxiliar não apenas na compreensão dos mecanismos atrelados à probiose, mas também na prospecção de novas linhagens que possam apresentar um potencial probiótico a serem utilizadas pela indústria.

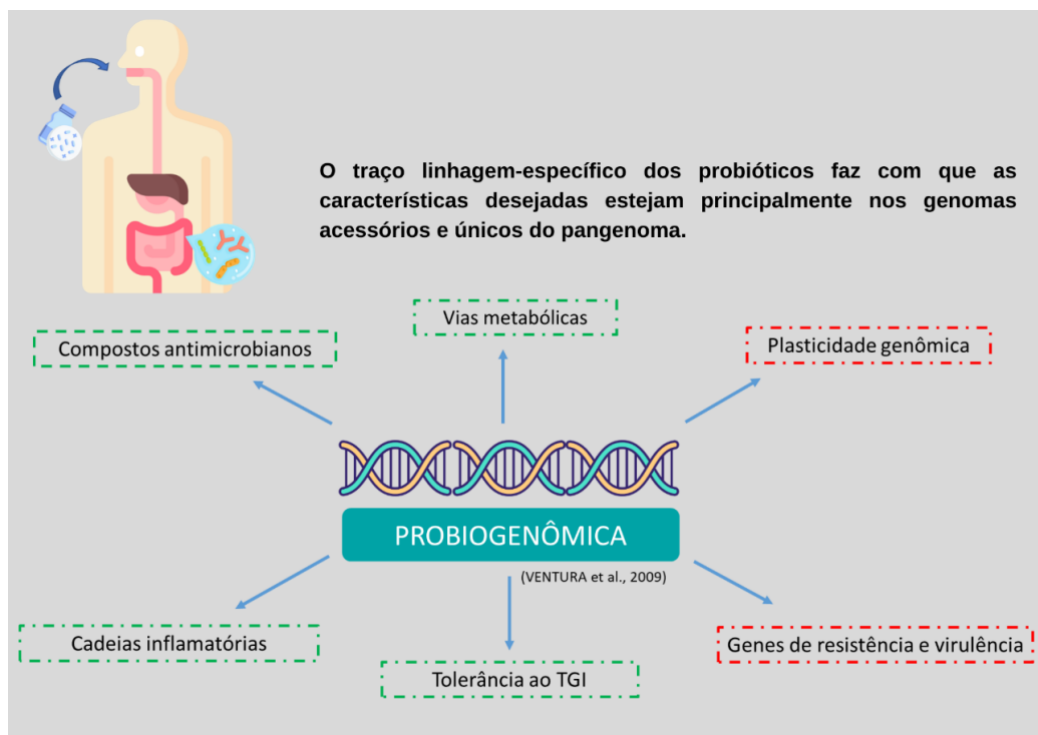


Figura 1. Esquema representativo dos objetivos de análises dentro da probiogenômica. Em verde, estão características relacionadas à probiose e em vermelho, características que merecem atenção a um nível de biossegurança. Fonte: Ventura e colaboradores [7].

2. O potencial probiótico de culturas iniciadoras

Ao realizar anotação funcional do pangenoma de 109 genomas de *Streptococcus thermophilus*, uma cultura homofermentativa amplamente utilizada na indústria láctea para a produção de iogurte natural, leite fermentado, e diversos tipos de queijos, percebe-se que características atreladas à probiose estão presentes como genes acessórios e únicos da espécie, pois não são compartilhados por todos os genomas. Isso corrobora com o postulado de que o potencial probiótico é linhagem específico e que diferentes linhagens apresentam potenciais diferentes em suas aplicações.

Especificamente na categoria de metabólitos secundários, foram preditas diversas sequências que potencialmente produzem compostos com efeito antimicrobiano, importante na competitividade por exclusão durante atuação no intestino contra microrganismos nocivos, por exemplo. A gama variada de mecanismos antimicrobianos preditos indica que algumas linhagens possuem a capacidade de interagir de maneira antagonista contra alguns microrganismos específicos. Isso também, pode ser um reflexo da pressão seletiva sofrida por linhagens em matrizes mais competitivas.

Finalmente, a predição do panviruloma e panresistoma da coleção corrobora com o perfil seguro da espécie, como demonstrado pela ausência de fatores de resistência que são comuns a todos os isolados. Interessantemente, os dois prováveis fatores de virulência comuns a todos os genomas estão atrelados à aderência à mucosa intestinal e imunomodulação, que também são funções importantes em um contexto de probiose.

3. Conclusão

Observa-se que existe um grande potencial probiótico contido em culturas iniciadoras como *S. thermophilus*. Por estar presente em produtos comuns da dieta, pesquisas que focam na busca de novas linhagens probióticas proporcionam o desenvolvimento de novos produtos focados em alimentos funcionais. Além disso, a probiogenômica se apresenta como uma ferramenta indispensável para as análises de segurança, visando que as linhagens não causem risco à saúde.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e FAPEMIG.

4. Referências

1. WHO, W.H.O. *Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation*; World Health Organization, 2003; ISBN 978-92-4-120916-8.
2. Milner, J.A. Functional Foods and Health Promotion. *J. Nutr.* **1999**, *129*, 1395S-1397S, doi:10.1093/jn/129.7.1395S.
3. Guandalini, S. Probiotics for Prevention and Treatment of Diarrhea. *J. Clin. Gastroenterol.* **2011**, *45*, S149, doi:10.1097/MCG.0b013e3182257e98.
4. Vonk, R.J.; Reckman, G.A.R.; Harmsen, H.J.R.; Priebe, M.G. Probiotics and Lactose Intolerance. In *Probiotics*; Rigobelo, E., Ed.; InTech, 2012 ISBN 978-953-51-0776-7.



5. Abraham, B.P.; Quigley, E.M.M. Probiotics in Inflammatory Bowel Disease. *Gastroenterol. Clin.* **2017**, *46*, 769–782, doi:10.1016/j.gtc.2017.08.003.
6. *Probiotics in Food: Health and Nutritional Properties and Guidelines for Evaluation*; FAO, F. and A.O. of the U.N., WHO, W.H.O., Eds.; FAO food and nutrition paper; Food and Agriculture Organization of the United Nations : World Health Organization: Rome, 2006; ISBN 978-92-5-105513-7.
7. Ventura, M.; O’Flaherty, S.; Claesson, M.J.; Turrone, F.; Klaenhammer, T.R.; van Sinderen, D.; O’Toole, P.W. Genome-Scale Analyses of Health-Promoting Bacteria: Probiogenomics. *Nat. Rev. Microbiol.* **2009**, *7*, 61–71, doi:10.1038/nrmicro2047.

29

Implicações sustentáveis do Sequenciamento de Nova Geração para a caracterização de linhagens isoladas de *Bacillus paralicheniformis* com interesse biotecnológico

Autores 29.1

Gabriel Camargos Gomes , Eduarda Guimarães Sousa , Eric Guédon ,
Vasco Azevedo 

Revisão: Diego Mariano , Lucianna Helene Santos , Savio Costa 





Cite este artigo 29.1

Gomes G. C. *et al.* **Implicações sustentáveis do Sequenciamento de Nova Geração para a caracterização de linhagens isoladas de *Bacillus paralicheniformis* com interesse biotecnológico.** BIOINFO. Vol. 6. p.29 (2026). doi: 10.51780/bioinfo-06-29

Saiba mais 29.1

Este artigo está disponível em <https://bioinfo.com.br/implicacoes-sustentaveis-do-sequenciamento-de-nova-geracao-para-a-caracterizacao-de-linhagens-isoladas-de-bacillus-paralicheniformis-com-interesse-biotecnologico/>

Implicações sustentáveis do Sequenciamento de Nova Geração para a caracterização de linhagens isoladas de *Bacillus paralicheniformis* com interesse biotecnológico

Gabriel Camargos Gomes ¹ , Eduarda Guimarães Sousa ¹ , Eric Guédon ² , Vasco Azevedo ^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

² STLO, INRA, Agrocampus Ouest, Rennes, France

* vascoariston@gmail.com

Resumo. A bioinformática tem se consolidado como uma ferramenta essencial para a prospecção e caracterização de microrganismos com potencial biotecnológico. No contexto agrícola, a crescente demanda por alimentos convive com práticas que impactam negativamente o meio ambiente, enquanto, no contexto farmacêutico, busca-se a produção de biomoléculas eficazes contra patógenos cada vez mais resistentes aos medicamentos tradicionais. A descoberta de novas bactérias que auxiliem nesses processos é, portanto, fundamental, mas as abordagens tradicionais mostram-se ineficazes, exigindo grandes quantidades de tempo e recursos, dando luz ao Sequenciamento de Nova Geração (NGS). Diante desse cenário, neste trabalho foram conduzidas análises comparativas e funcionais de genomas sequenciados de *Bacillus paralicheniformis*, uma espécie bacteriana emergente, produtora de metabólitos com potencial biotecnológico e reconhecida como promotora do crescimento de plantas (PGPB). O estudo teve como foco entender melhor a espécie do ponto de vista filogenômico e comparativo, além da caracterização de isolados, com identificação de genes relacionados à síntese de metabólitos secundários, mecanismos de defesa e metabolismo do nitrogênio. Além disso, buscou-se evidenciar como a bioinformática se mostra cada vez mais confiável na identificação e triagem de novos microrganismos, além da contribuição para o avanço de práticas agrícolas sustentáveis e para o enfrentamento de desafios globais, como a fome e a degradação ambiental.

Palavras-chave: Agricultura; NGS; Genômica; *Bacillus*; Biotecnologia;

1. Introdução

A crescente demanda por práticas agrícolas sustentáveis tem impulsionado a busca por alternativas biotecnológicas que reduzam a dependência de fertilizantes e defensivos químicos. Nesse contexto, microrganismos do gênero *Bacillus* destacam-se como promotores de crescimento vegetal (PGPR), capazes de sintetizar metabólitos que favorecem a nutrição e a defesa das plantas [1]. Entretanto, devido à alta similaridade genômica entre espécies do gênero, métodos tradicionais, como o MALDI-TOF, apresentam limitações na diferenciação em nível de espécie e subespécie. Em contrapartida, o Sequenciamento de Nova Geração (NGS) tem se mostrado uma ferramenta de destaque na identificação e caracterização de novas linhagens, impulsionado pela redução de custos e pela crescente disponibilidade de genomas depositados

em bancos de dados públicos, como o *National Center for Biotechnology Information* (NCBI) [2]. Essa expansão tem permitido o uso de softwares comparativos que integram dados evolutivos e genômicos, fornecendo classificações taxonômicas mais precisas.

Uma espécie emergente desse gênero, *Bacillus paralicheniformis*, tem despertado interesse por sua diversidade genética e potencial de aplicação na agricultura e na indústria. Com o avanço das ferramentas de bioinformática e do sequenciamento genômico, tornou-se possível explorar o potencial funcional de novas linhagens com maior precisão, revelando genes associados a vias metabólicas de interesse agrônômico e farmacêutico.

2. Desenvolvimento da pesquisa

Sintetizado na Figura 1, foram identificadas e posteriormente caracterizadas duas linhagens isoladas de *Bacillus paralicheniformis* (Denominadas como BAC30 e BAC220) a partir do NGS, e depois incluídas com outros genomas da espécie, obtidos pelo NCBI para análises comparativas. As análises filogenômicas e pangenômicas indicaram elevada similaridade entre as linhagens, salientando a dificuldade de identificação de exemplares do gênero *Bacillus*. Além disso, análises pangenômicas mostraram um padrão genético com alta porcentagem de genes essenciais e singulares, e poucos genes compartilhados, além de um pangenoma aberto, evidenciando um padrão cosmopolita da espécie, hipotetizando um comportamento de co-evolução com hospedeiros, como plantas e animais. Quanto aos isolados, foi notada a presença de genes relacionados à biossíntese de vitaminas do complexo B e antibióticos naturais, como a riboflavina e a bacitracina. A anotação funcional identificou módulos completos de resistência a β -lactâmicos, rotas de assimilação de nitrogênio e síntese de siroheme, além de clusters de genes envolvidos na produção de metabólitos não ribossomais. Esses resultados reforçam a versatilidade metabólica da espécie e sua capacidade de adaptação a diferentes nichos ecológicos. Do ponto de vista aplicado, a caracterização genômica de cepas de *Bacillus paralicheniformis* contribui para a prospecção de novas linhagens com potencial uso em biofertilizantes, biocontroladores e biorremediadores, reduzindo impactos ambientais e promovendo maior eficiência agrícola.

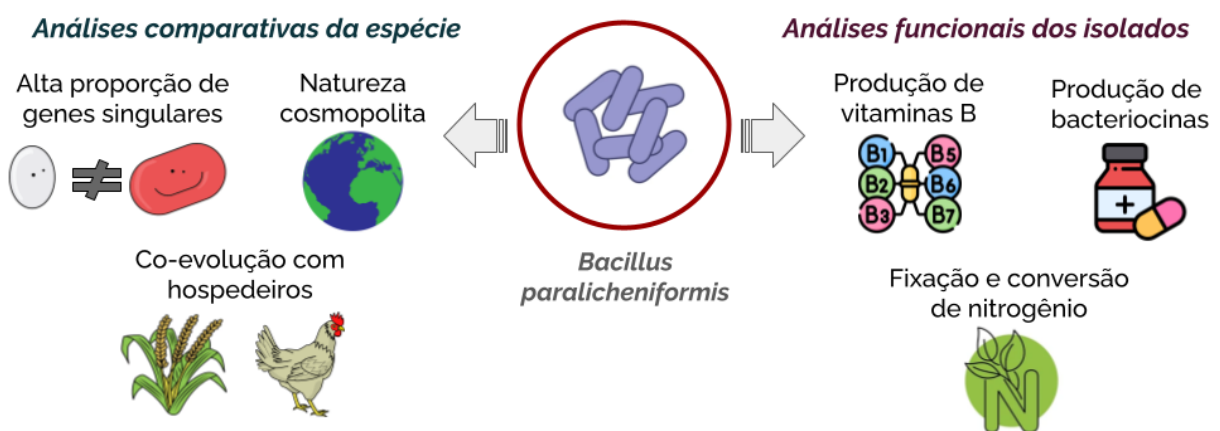


Figura 1. Principais resultados à partir das análises comparativas e funcionais dos isolados BAC30 e BAC220

3. Implicações do estudo para práticas sustentáveis

A identificação de genes relacionados à fixação de nitrogênio implica em um maior aproveitamento e conversão de nitratos e nitritos no solo, reduzindo a dependência de fertilizantes químicos sintéticos, um dos principais vetores de poluição agrícola e de emissões de gases de efeito estufa [3]. Além disso, a presença de genes associados à síntese de bacitracina e outros peptídeos não ribossômicos (NRPs) demonstra o potencial antimicrobiano natural dessas linhagens, favorecendo o controle biológico de patógenos vegetais e reduzindo a necessidade de pesticidas e antibióticos artificiais. Quanto à produção de siroheme, sua síntese por bactérias representa uma alternativa livre de crueldade animal às fontes tradicionais obtidas de sangue [4], com aplicações promissoras nas indústrias alimentícia e farmacêutica. A combinação desses fatores, aliada à produção de vitaminas do complexo B, evidencia o potencial biotecnológico de *Bacillus paralicheniformis* como uma espécie capaz de contribuir para a agricultura regenerativa, reduzindo o impacto ambiental causado por fertilizantes e pesticidas, além de viabilizar o desenvolvimento de bioprodutos éticos e economicamente sustentáveis. Assim, este estudo reforça a importância do NGS não apenas como padrão-ouro para a identificação de novas linhagens de *B. paralicheniformis*, mas também como ferramenta essencial para a caracterização e aplicação racional de microrganismos em soluções biotecnológicas voltadas à sustentabilidade global. Entretanto, como a abordagem adotada baseia-se exclusivamente em análises *in silico*, estudos complementares *in vitro* e *in vivo* ainda são necessários para confirmar a expressão funcional destes genes e validar o potencial biotecnológico das linhagens BAC30 e BAC220.

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e Fapemig.




4. Referências





- [1] Du Y, Ma J, Yin Z, Liu K, Yao G, Xu W, Fan L, Du B, Ding Y, Wang C (2019) **Comparative genomic analysis of *Bacillus paralicheniformis* MDJK30 with its closely related species reveals an evolutionary relationship between *B. paralicheniformis* and *B. licheniformis*.** BMC Genomics 20(1):283. <https://doi.org/10.1186/s12864-019-5646-9>
- [2] Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., & Ussery, D. W. (2015). **Insights from 20 years of bacterial genome sequencing.** *Functional & Integrative Genomics*, 15(2), 141–161. <https://doi.org/10.1007/s10142-015-0433-4>
- [3] Ahmed M, Rauf M, Mukhtar Z, Saeed NA (2017) **Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health.** Environ Sci Pollut Res 24(35):26983–26987. <https://doi.org/10.1007/s11356-017-0589-7>
- [4] Yang S, Wang A, Li J, Shao Y, Sun F, Li S, Cao K, Liu H, Xiong P, Gao Z (2023) **Improved biosynthesis of heme in *Bacillus subtilis* through metabolic engineering assisted fed-batch fermentation.** Microb Cell Fact 22(1):102. <https://doi.org/10.1186/s12934-023-02077-3>

30

Estudo pangenômico de *Corynebacterium rouxii* e *Corynebacterium ulcerans* com foco em determinantes genéticos de virulência, resistência antimicrobiana e impacto zoonótico

Autores 30.1

Fernanda Diniz Prates , Max Roberto Batista Araújo , Louisy Sanches dos Santos , Vasco Ariston de Carvalho Azevedo 

Revisão: Diego Mariano , Marcos Antonio Nobrega de Sousa , Rafael Eduardo Oliveira Rocha , Savio Costa 





Cite este artigo 30.1

Prates F D. *et al.* Estudo pangenômico de *Corynebacterium rouxii* e *Corynebacterium ulcerans* com foco em determinantes genéticos de virulência, resistência antimicrobiana e impacto zoonótico. *BIOINFO*. Vol. 6. p.30 (2026). doi: 10.51780/bioinfo-06-30

Saiba mais 30.1

Este artigo está disponível em <https://bioinfo.com.br/estudo-pangenomico-de-corynebacterium-rouxii-e-corynebacterium-ulcerans-com-foco-em-determinantes-geneticos-de-virulencia-resistencia-antimicrobiana-e-impacto-zoonotico/>

Estudo pangenômico de *Corynebacterium rouxii* e *Corynebacterium ulcerans* com foco em determinantes genéticos de virulência, resistência antimicrobiana e impacto zoonótico

Fernanda Diniz Prates^{1,2*} , Max Roberto Batista Araújo^{1,2} , Louisy Sanches dos Santos³ , Vasco Ariston de Carvalho Azevedo² 

¹ Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil

² Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

³ Laboratório de Difteria e Corinebactérias de Importância Clínica, Departamento de Microbiologia, Imunologia e Parasitologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil

* fernandaprates3@hotmail.com

Resumo. A difteria é uma doença infecciosa aguda e potencialmente fatal, cujo principal agente etiológico é *Corynebacterium diphtheriae*. Apesar da vacinação, casos da doença ainda são registrados em todo mundo, inclusive em indivíduos imunizados. Nas últimas décadas, outras espécies do gênero também foram identificadas como potenciais produtoras da toxina diftérica e foram agrupadas no chamado complexo *C. diphtheriae*. Dentro deste complexo, as espécies emergentes *Corynebacterium rouxii* e *Corynebacterium ulcerans*, têm se destacado devido ao seu potencial de transmissão zoonótica, com relevância crescente em saúde pública. Nesse contexto, a obtenção de dados genômicos dessas espécies em circulação no Brasil é imprescindível para melhor entendimento do impacto zoonótico e para realizar a vigilância epidemiológica molecular destes patógenos. Além disso, a comparação dos dados genômicos através da abordagem pangenômica ajuda a compreender a aquisição de determinantes de virulência e resistência a antimicrobianos apresentados por estas espécies, assim como no desenvolvimento de novas estratégias diagnósticas e terapêuticas.

Palavras-chave: *Corynebacterium rouxii*; *Corynebacterium ulcerans*; pangenômica, virulência, resistência antimicrobiana, zoonose.

1. Introdução

O gênero *Corynebacterium* inclui 168 espécies válidas, algumas delas patogênicas para humanos e animais [1]. A espécie tipo *Corynebacterium diphtheriae* é o principal agente etiológico da difteria, uma doença potencialmente fatal, cujas principais manifestações locais e sistêmicas são decorrentes da ação da toxina diftérica (TD), uma exotoxina produzida pelo microrganismo quando lisogenizado por fagos portadores do gene *tox* [2]. Apesar da diminuição mundial do número de casos da doença reportados após as campanhas de vacinação, a difteria permanece endêmica em algumas regiões e ainda é relatada mesmo entre indivíduos imunizados [3,4].

Além de *C. diphtheriae*, outras espécies do gênero (*Corynebacterium belfantii*, *Corynebacterium pseudotuberculosis*, *Corynebacterium rouxii*, *Corynebacterium silvaticum* e

Corynebacterium ulcerans) podem carrear o gene *tox* e, atualmente, formam o chamado complexo *C. diphtheriae* [5-8].

C. rouxii foi descrito em 2020 com base na taxonomia genômica de seis cepas atípicas maltose negativas, isoladas de humanos e um cão, previamente identificadas como *C. diphtheriae* biovar Belfanti [9]. Neste mesmo estudo, cepas atípicas do biovar Belfanti isoladas de gatos nos EUA [7] também foram reclassificadas como *C. rouxii*. Até o momento, os casos mais recentes de infecção por *C. rouxii* foram relatados nos EUA [10], Alemanha [11], Espanha [12] e Brasil [13].

C. ulcerans, por sua vez, é um patógeno zoonótico emergente que coloniza tanto animais domésticos [14] quanto selvagens [15]. Casos em humanos relacionados ao contato com animais infectados têm sido cada vez mais reportados, reforçando a necessidade de uma vigilância integrada entre saúde humana e veterinária [16]. Tanto cepas toxigênicas quanto atoxigênicas têm sido relatados com crescente frequência em todo o mundo, sendo que as primeiras costumam causar doenças semelhantes à difteria.

Embora estudos recentes tenham explorado a caracterização genômica de *C. rouxii* e *C. ulcerans*, ainda há poucos genomas disponíveis em bancos públicos e uma representatividade geográfica limitada, especialmente para *C. rouxii*, o que restringe a compreensão da real diversidade e distribuição global dessas espécies. Além disso, pouco se sabe sobre a persistência ambiental e sobre os mecanismos de virulência e resistência antimicrobiana que sustentam o potencial patogênico de ambas as espécies. Por fim, o impacto dessas espécies na saúde pública brasileira ainda não foi plenamente dimensionado. Diante disto, este artigo explora como a abordagem da pangenômica tem sido utilizada para estudar as espécies *C. rouxii* e *C. ulcerans* no Brasil, contribuindo para preencher essas lacunas de conhecimento.

2. Discussão

O conceito de pangenoma foi inicialmente proposto para descrever a diversidade genômica de microrganismos, especialmente bactérias, cujos genomas apresentam elevada plasticidade devido à ocorrência frequente de mutações, eventos de recombinação e transferência horizontal de genes. Essa abordagem possibilita a identificação de genes que persistem no genoma de várias linhagens de uma mesma espécie e que apresentam mutações mínimas ao longo do tempo, assim como de genes frequentemente associados à virulência, à resistência a antimicrobianos e à adaptação a diferentes tipos de nichos [17].

O pangenoma é geralmente dividido em três componentes principais: o genoma central (*core genome*), formado por genes compartilhados por todas as linhagens de uma mesma espécie e que geralmente codificam proteínas que desempenham papéis significativos em vários processos celulares críticos e essenciais para a sua sobrevivência [18]; o genoma compartilhado (*shared*), constituído por genes presentes em algumas linhagens de uma determinada espécie, mas não tem todas, e que podem codificar proteínas envolvidas em funções suplementares que conferem vantagens seletivas; e os genes exclusivos (*singletons*), restritos a uma única linhagem, frequentemente relacionados a eventos recentes de aquisição horizontal ou mutações pontuais (Figura 1) [17,19].

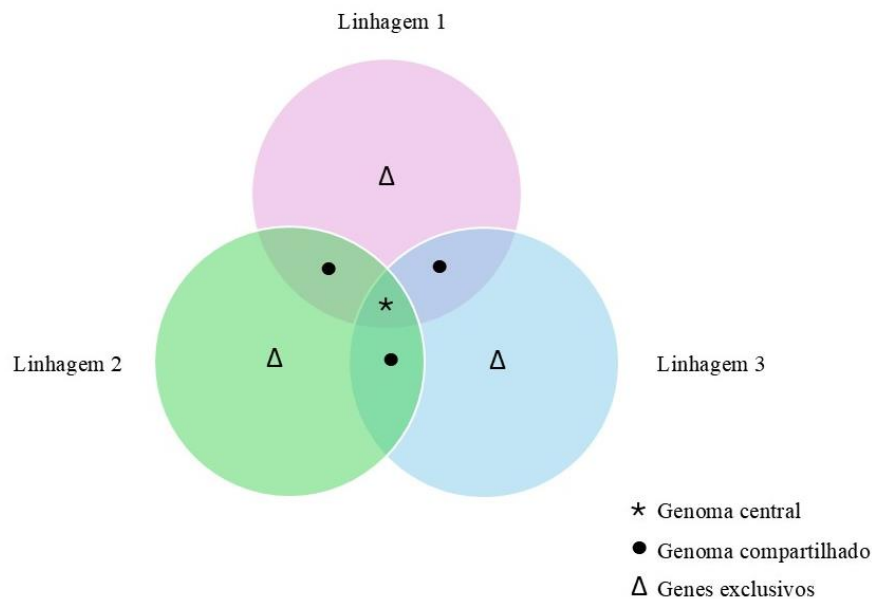


Figura 1. Diagrama de Venn representando os componentes de um pangenoma: genoma central (*), genoma compartilhado (●) e genes exclusivo (Δ).

Diversos pipelines têm sido desenvolvidos para realizar análises pangenômica, cada um com funcionalidades específicas. Como exemplo temos o Roary, uma ferramenta eficiente para genomas anotados pelo Prokka, que realiza a identificação de genes ortólogos via BLASTP e geração de matrizes de presença/ausência de genes (<https://sanger-pathogens.github.io/Roary/>) [20] e o Panaroo, que utiliza grafos de associação para identificar genes ortólogos com maior precisão, corrigindo erros de anotação e fragmentação de genes (<https://github.com/gtonkinhill/panaroo>) [21]. Há também o pipeline BPGA (*Bacterial Pan Genome Analysis*), que permite o agrupamento de genes por similaridade, classificação funcional com base em COGs (*Clusters of Orthologous Groups*), além de vários outros módulos de funcionalidade (**Figura 2**) [22].

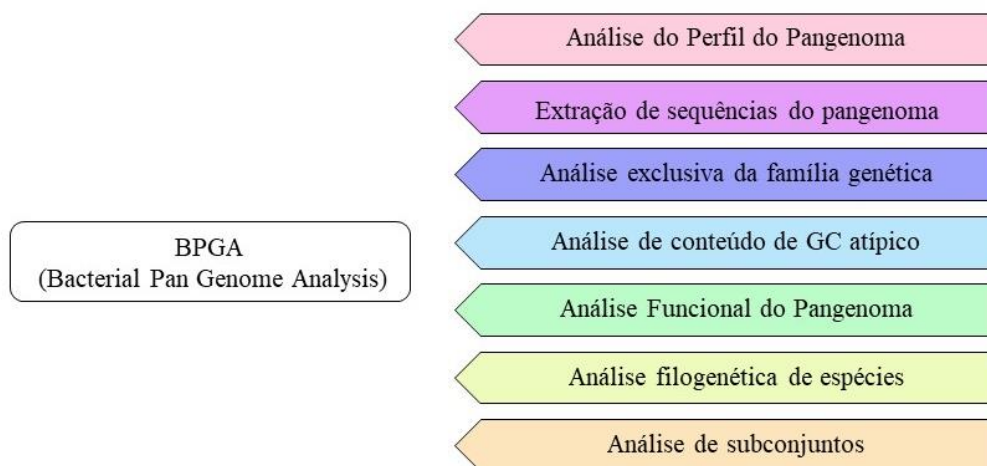


Figura 2. Módulos de funcionalidade disponível no BPGA.

A aplicação da pangenômica vai além da análise genômica básica. Essa categorização permite não apenas compreender os mecanismos evolutivos que moldam o repertório gênico de

uma espécie, mas também identificar potenciais alvos candidatos para o desenvolvimento de novas vacinas e métodos diagnósticos mais assertivos. A abordagem tem sido ainda muito utilizada na vigilância epidemiológica, auxiliando na detecção precoce de surtos, no rastreamento de rotas de transmissão e na diferenciação entre linhagens endêmicas e emergentes, mesmo em contextos de alta diversidade genômica [23]. Também tem sido muito utilizada na análise da interação patógeno-hospedeiro, contribuindo na caracterização de fatores de virulência, genes de resistência a antimicrobianos e genes que possam estar relacionados com a modulação da resposta imune [24].

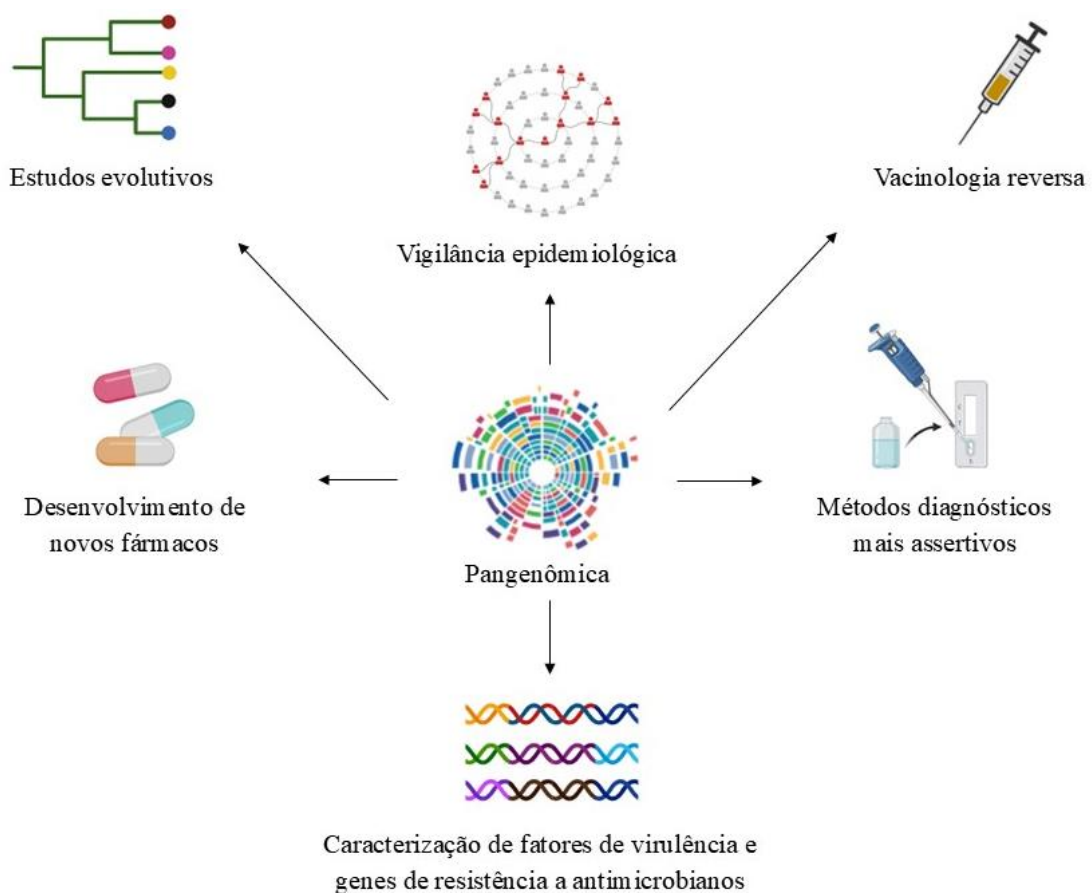


Figura 3. Principais aplicações da pangenômica na genômica de microrganismos.

Devido à sua versatilidade, a abordagem pangenômica é de grande utilidade nos estudos envolvendo *C. rouxii* e *C. ulcerans*. Ela tem revelado, por exemplo, determinantes genéticos relevantes para virulência, adaptação a diferentes ambientes e hospedeiros, e resistência antimicrobiana nestas espécies. Combinada com outras análises computacionais e laboratoriais, a pangenômica tem auxiliado na compreensão de eventos evolutivos nestas espécies, além de ser uma potencial ferramenta para o entendimento do seu impacto zoonótico. No contexto da Saúde Única (*One Health*), que reconhece a conexão entre humanos, animais e o meio ambiente, o conhecimento adquirido sobre estas espécies é essencial para orientar políticas de controle e promover ações preventivas contra a difteria e outras infecções relacionadas a estas espécies.

3. Conclusão

A integração entre a saúde humana, animal e ambiental é fundamental para enfrentar os desafios impostos pelas doenças infecciosas, pois permite compreender a origem e a dinâmica dessas doenças, prevenir surtos e monitorar a disseminação da resistência antimicrobiana. Nesse contexto, a pangenômica tem se destacado como uma ferramenta útil para caracterizar a diversidade genética e identificar genes associados à adaptação e ao potencial zoonótico de espécies como *C. rouxii* e *C. ulcerans*. Quando combinada com outras metodologias computacionais e integrada a dados experimentais e epidemiológicos, a pangenômica tende a fortalecer a vigilância epidemiológica, apoiar a identificação de novos alvos para diagnóstico e intervenção terapêutica, e fornecer informações que podem subsidiar políticas públicas baseadas mais eficazes.

4. Referências

- [1] Parte, AC et al. **List of Prokaryotic Names with Standing in Nomenclature (LPSN) Moves to the DSMZ.** *Int J Syst Evol Microbiol.* Vol. 70. p.5607–5612 (2020). doi:10.1099/ijsem.0.004332.
- [2] Sharma, NC et al. **Diphtheria.** *Nat Rev Dis Primers.* Vol. 5. p.81 (2019). doi:10.1038/s41572-019-0131-y
- [3] Batista, MR et al. **Cutaneous Infection by Non-Diphtheria-Toxin Producing and Penicillin-Resistant *Corynebacterium Diphtheriae* Strain in a Patient with Diabetes Mellitus.** *Access Microbiol.* Vol. 3 (2021). doi:10.1099/acmi.0.000284
- [4] Batista, MR et al. **Phenotypic and Molecular Characterization and Complete Genome Sequence of a *Corynebacterium Diphtheriae* Strain Isolated from Cutaneous Infection in an Immunized Individual.** *Brazilian Journal of Microbiology.* Vol. 54. p.1325-1334 (2023). doi:10.1007/s42770-023-01086-z
- [5] Dangel, A et al. ***Corynebacterium Silvaticum* Sp. Nov., a Unique Group of NTTB *Corynebacteria* in Wild Boar and Roe Deer.** *Int J Syst Evol Microbiol.* Vol. 70. p.3614-3624 (2020). doi:10.1099/ijsem.0.004195
- [6] Viana, MC et al. **Analysis of *Corynebacterium Silvaticum* Genomes from Portugal Reveals a Single Cluster and a Clade Suggested to Produce Diphtheria Toxin.** *PeerJ.* Vol. 11. (2023). doi:10.7717/peerj.14895
- [7] Dazas, M et al. **Taxonomic Status of *Corynebacterium Diphtheriae* Biovar Belfanti and Proposal of *Corynebacterium Belfantii* Sp. Nov.** *Int J Syst Evol Microbiol.* Vol. 68. p.3826-3831 (2018). doi:10.1099/ijsem.0.003069.
- [8] Prates, FD et al. **First Pangenome of *Corynebacterium Rouxii*, a Potentially Toxigenic Species of *Corynebacterium Diphtheriae* Complex.** *Bacteria.* Vol. 3. p.99-117 (2024). doi:10.3390/bacteria3020007
- [9] Badell, E et al. ***Corynebacterium rouxii* sp. nov., a novel member of the diphtheriae species complex.** *Res. Microbiol.* Vol. 171. p.122-127 (2020). doi: 10.1016/j.resmic.2020.02.003
- [10] Tyler, R et al. **Toxigenic *Corynebacterium diphtheriae* Infection in Cat, Texas, USA.** *Emerg. Infect. Vol.* 28. p.1686-1688 (2022). doi: 10.3201/eid2808.220018
- [11] Schlez, K et al. ***Corynebacterium rouxii*, a recently described member of the *C. diphtheriae* group isolated from three dogs with ulcerative skin lesions.** *Antonie Van Leeuwenhoek.* Vol. 114. p.1361-1371 (2021). doi: 10.1007/s10482-021-01605-8

- [12] Hoefler, A et al. **Molecular and Epidemiological Characterization of Toxigenic and Nontoxigenic *Corynebacterium diphtheriae*, *Corynebacterium belfantii*, *Corynebacterium rouxii*, and *Corynebacterium ulcerans* Isolates Identified in Spain from 2014 to 2019.** J. Clin. Microbiol. Vol. 59. p.1-15 (2021). doi: 10.1128/JCM.02410-20
- [13] Ramos, JN et al. **Molecular characterization and phylogenetic analysis of the first *Corynebacterium rouxii* strains isolated in Brazil: A recent member of *Corynebacterium diphtheriae* complex.** BMC Genom. Data. Vol. 24. (2023). doi: 10.1186/s12863-023-01167-w
- [14] Katsukawa, C et al. **Prevalence of *Corynebacterium Ulcerans* in Dogs in Osaka, Japan.** J Med Microbiol. Vol. 61. p.266-273 (2012). doi:10.1099/jmm.0.034868-0
- [15] Contzen, M et al. ***Corynebacterium Ulcerans* from Diseased Wild Boars.** Zoonoses Public Health. Vol. 58. p.479-488 (2011). doi:10.1111/j.1863-2378.2011.01396.x
- [16] De Zoysa, A et al. **Characterization of Toxigenic *Corynebacterium Ulcerans* Strains Isolated from Humans and Domestic Cats in the United Kingdom.** J Clin Microbiol. Vol. 43. p.4377-4381 (2011). doi:10.1128/JCM.43.9.4377-4381.2005
- [17] Tettelin, H et al. **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome.** Proc Natl Acad Sci U S A. Vol. 102. p.13950-13955 (2015). 10.1073/pnas.0506758102
- [18] Rosconi, F et al. **A bacterial pan-genome makes gene essentiality strain-dependent and evolvable.** Nat Microbiol. Vol. 10. p.1580-1592 (2022). 10.1038/s41564-022-01208-7
- [19] Rouli, L et al. **The bacterial pangenome as a new tool for analysing pathogenic bacteria.** New Microbes and New Infections. Vol. 7. p.72-85 (2015). 10.1016/j.nmni.2015.06.005
- [20] Page, A et al. **Roary: rapid large-scale prokaryote pangenome analysis.** Bioinformatics. Vol. 31. p.3691-3693 (2015). 10.1093/bioinformatics/btv421
- [21] Tonkin-Hill, H et al. **Producing polished prokaryotic pangenomes with the Panaroo pipeline.** Genome Biol. Vol. 21 (2020). 10.1186/s13059-020-02090-4
- [22] Chaudhari, N et al. **BPGA- an ultra-fast pan-genome analysis pipeline.** Sci Rep. Vol. 6. (2016). 10.1038/srep24373
- [23] Asnicar, F et al. **Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0.** Nat Commun. Vol. 11 (2020). 10.1038/s41467-020-16366-7
- [24] Chekole, W et al. **Genomic insights into antimicrobial resistance and virulence of *E. coli* in central Ethiopia: a one health approach.** Front Microbiol. Vol. 16 (2025). 10.3389/fmicb.2025.1597580

31

Desafios e impactos da pesquisa na sociedade: formando bioinformatas para uma década de transformações

Autores 31.1

Catarina Rodrigues Gomes , Arun Kumar Jaiswal , Vasco Azevedo 

Revisão: Diego Mariano , Lucianna Helene Santos , Thiago Jesus Sousa 

Cite este artigo 31.1

Gomes, C. R.; Jaiswal, A. K.; Azevedo, V. **Desafios e impactos da pesquisa na sociedade: formando bioinformatas para uma década de transformações**. BIOINFO. Vol. 6. p.31 (2026). doi: 10.51780/bioinfo-06-31

Saiba mais 31.1

Este artigo está disponível em <https://bioinfo.com.br/desafios-e-impactos-da-pesquisa-na-sociedade-formando-bioinformatas-para-uma-decada-de-transformacoes/>

Desafios e impactos da pesquisa na sociedade: formando bioinformatas para uma década de transformações

Catarina Rodrigues Gomes¹ , Arun Kumar Jaiswal¹ , Vasco Azevedo^{1*} 

¹ Programa Interunidades de Pós-Graduação em Bioinformática da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

* vascoariston@gmail.com

Resumo. A bioinformática, hoje, é uma disciplina central na interseção entre a biologia, a computação e a ciência de dados, com impacto crescente à medida que grandes bancos de dados de informação biológica se expandem. O campo já impacta as áreas da saúde, agricultura, meio ambiente e indústria, sendo motor de inovação e descoberta científica. No entanto, seu crescimento, impulsionado pelas eras do Big Data e Inteligência Artificial, traz desafios estruturais e éticos. Buscamos, por meio deste artigo, discutir como preparar uma nova geração de bioinformatas treinados para navegar a próxima década da bioinformática de forma crítica. Destacamos a necessidade de repensar a formação acadêmica no campo, enfatizando as bases biológicas e computacionais do conhecimento, e de apoiar programas de pós-graduação na área, marcados por precarização e escassez de recursos. Argumentamos que a IA, por mais que seja promissora como ferramenta de apoio a análises complexas, não substitui o pensamento crítico e a compreensão fundamental de dados biológicos. Propomos estratégias pedagógicas interdisciplinares, parcerias com o setor privado, e fortalecimento da bioinformática no âmbito da política pública para garantir a formação de novos bioinformatas com perspectivas de futuro. Este artigo, portanto, busca chamar a atenção da comunidade científica e das agências financiadoras de pesquisa para a urgência de formar e valorizar profissionais de bioinformática capazes de impulsionar o Brasil até 2035.

Palavras-chave: Bioinformática; Inteligência Artificial; Big Data; Formação de Talentos; Política Científica.

1. Introdução

A bioinformática, área que busca unir a biologia e a ciência da computação, vem se tornando ferramenta essencial para decifrar os segredos da vida, com efeitos tangíveis nos campos ambientais e da saúde [1]. Nos últimos anos, com o crescimento revolucionário de métodos laboratoriais e de sequenciamento de DNA, conseguimos produzir informações genéticas em velocidades e quantidades que antes seriam inimagináveis. Com isso, nós entramos na era que chamamos de *Big Data*, com enormes bancos de dados repletos de informação biológica, e o novo desafio é interpretá-la [2]. No Brasil, esbarramos ainda em uma dificuldade concreta: a formação de profissionais que dominem ao mesmo tempo a biologia e a ciência de dados e possam interpretar e trabalhar com esse volume de informação.

Diante desse cenário, esta comunicação tem como objetivo discutir caminhos para desenvolver a bioinformática de forma sustentável no país. A alta demanda por profissionais qualificados colide com a fragilidade das instituições que buscam formar pesquisadores e a

escassez de recursos dedicados à pesquisa. Em paralelo, a Inteligência Artificial (IA) figura cada vez mais fortemente no campo, prometendo tornar análises complexas mais acessíveis, mas podendo também mascarar sérias lacunas conceituais na formação de um pesquisador. Enfrentar essas questões estruturais e éticas na bioinformática será fundamental para que o Brasil participe dessa fronteira do conhecimento pelos próximos 10 anos.

2. A evolução da bioinformática e o impacto da IA

Com a chegada da era *Big Data*, a bioinformática passou por uma grande transformação, deixando de ser ferramenta auxiliar a outras disciplinas científicas para ser um grande motivador de descobertas. Essa mudança foi acelerada pela IA e o aprendizado de máquina. Hoje, a bioinformática é essencial para pesquisar e desenvolver novos medicamentos e vacinas, criar cultivos resistentes à seca e às pragas, e monitorar doenças [3, 4]. A IA acelera diversas descobertas, permitindo analisar uma quantidade imensa de dados de forma rápida, como com o AlphaFold [5], que prevê a estrutura de proteínas, ou ferramentas capazes de identificar mutações e variantes que levem a maiores riscos de doenças no genoma anos antes de seus sintomas aparecerem [6].

No entanto, a parceria entre a bioinformática e a IA é uma faca de dois gumes. Por um lado, a IA democratiza o acesso. Ferramentas modernas são intuitivas e acessíveis, permitindo até não especialistas a realizar análises complexas de genoma. Por outro, há um risco: a facilidade de obter respostas com um clique não requer a compreensão do que está sendo feito ou gerado. Se confiamos cegamente nos resultados dessas caixas-pretas sem entender a fundamentação biológica por trás dos processos, o pensamento crítico pode ficar em segundo plano. Outro aspecto potencialmente nocivo dessa integração são os modelos de linguagem e IAs generativas. O que por um lado pode ser uma ferramenta em potencial, também pode ter efeitos no esforço cognitivo e pensamento crítico de pesquisadores, devendo ser usada com o devido cuidado [7].

3. Desafios na formação de bioinformatas

A formação desses novos cientistas no Brasil esbarra em um desafio, a desconexão entre as disciplinas. A graduação produz biólogos sem bases na programação e análise de dados, e a formação de cientistas da computação não abarca questões biológicas. Logo, os alunos de ambos os lados têm que complementar as próprias formações de modo que abarque ambos os campos. É como se a bioinformática buscasse integrar duas áreas que falam línguas diferentes. E mais, além da lacuna educacional, persistem também desafios estruturais. A carreira acadêmica sofre com a escassez de bolsas, agravada pelo enfraquecimento de programas de pós-graduação e a concentração de recursos nos estados mais ricos [8].

4. Conclusão e perspectivas

O grande desafio atual, portanto, passa a ser fortalecer infraestrutura e formar uma nova geração de cientistas que saiba aproveitar o poder da computação e da IA sem abrir mão do pensamento crítico, que é a base do método científico. A bioinformática, essencial para o avanço científico

e tecnológico de ponta em setores estratégicos como saúde, agricultura e meio ambiente, ainda enfrenta dificuldades formativas e políticas crônicas. A próxima década, ao que tudo indica, será uma década de transformações, determinante para consolidar a área no Brasil. Para navegarmos as marés da mudança com sucesso, serão necessárias ações urgentes para priorizar a formação de uma nova geração de bioinformatas capazes, com uma formação curricular que integre biologia, ciência de dados e ética em IA, e a criação de carreiras estáveis no país para combater a fuga de cérebros.

Para tal, será fundamental estabelecer parcerias sólidas entre universidades e empresas de biotecnologia, garantindo que a formação tenha um componente prático alinhado às necessidades nacionais. Além disso, um esforço contínuo de política pública científica é necessário para que a bioinformática seja reconhecida como área estratégica e valorizada no âmbito do financiamento nacional. Neste processo, a formação de novos bioinformatas deve mirar na competência técnica integrada e no desenvolvimento de consciência social e científica que lhes permita navegar pelos desafios éticos do futuro com pensamento crítico afiado. A próxima década será decisiva para que o Brasil transforme seu potencial biológico em soberania tecnológica através de investimentos sérios, para que possamos participar desse campo de ponta em força. Um plano de ação para enfrentar os desafios na formação de novos profissionais em bioinformática está disponível na Figura 1.



Figura 1. Plano de ação para enfrentar os desafios na formação de novos profissionais capazes no campo da bioinformática

Agradecimentos. Os autores agradecem às agências de fomento à pesquisa: CAPES, CNPq e Fapemig.

4. Referências

- [1] Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56, 2019.
- [2] Saparov, A. Big data and transformative bioinformatics in genomic diagnostics and beyond, *Parkinsonism & Related Disorders*, Volume 134, 2025, 107311, ISSN 1353-8020.
- [3] Mardinoglu A, Turkez H, Shong M, Srinivasulu VP, Nielsen J, Palsson BO, Hood L, Uhlen M. Longitudinal big biological data in the AI era. *Mol Syst Biol*. 2025 Sep;21(9):1147-1165.
- [4] Huixiao , H. Editorial: Big data and artificial intelligence for genomics and therapeutics – Proceedings of the 19th Annual Meeting of the MidSouth Computational Biology and Bioinformatics Society (MCBIOS). *Frontiers in Bioinformatics*. Volume 4 - 2024
- [5] Jumper J. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596:583–589.
- [6] Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**, 983–987 (2018)
- [7] Lee, H.P. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 1 May–26 April 2025; pp. 1–22.
- [8] Oliveira, M.F. Cuts to postgraduate funding threaten Brazilian science — again. *Nature* 627, 735, 2024