

UNIVERSIDADE FEDERAL DE MINAS GERAIS INSTITUTO DE  
CIÊNCIAS BIOLÓGICAS, DEPARTAMENTO DE GENÉTICA,  
ECOLOGIA E EVOLUÇÃO, PROGRAMA DE PÓS - GRADUAÇÃO  
EM GENÉTICA



DISSERTAÇÃO

Pan genômica e análises de resistência e virulência sob seleção positiva darwiniana de  
*Gardnerella vaginalis*

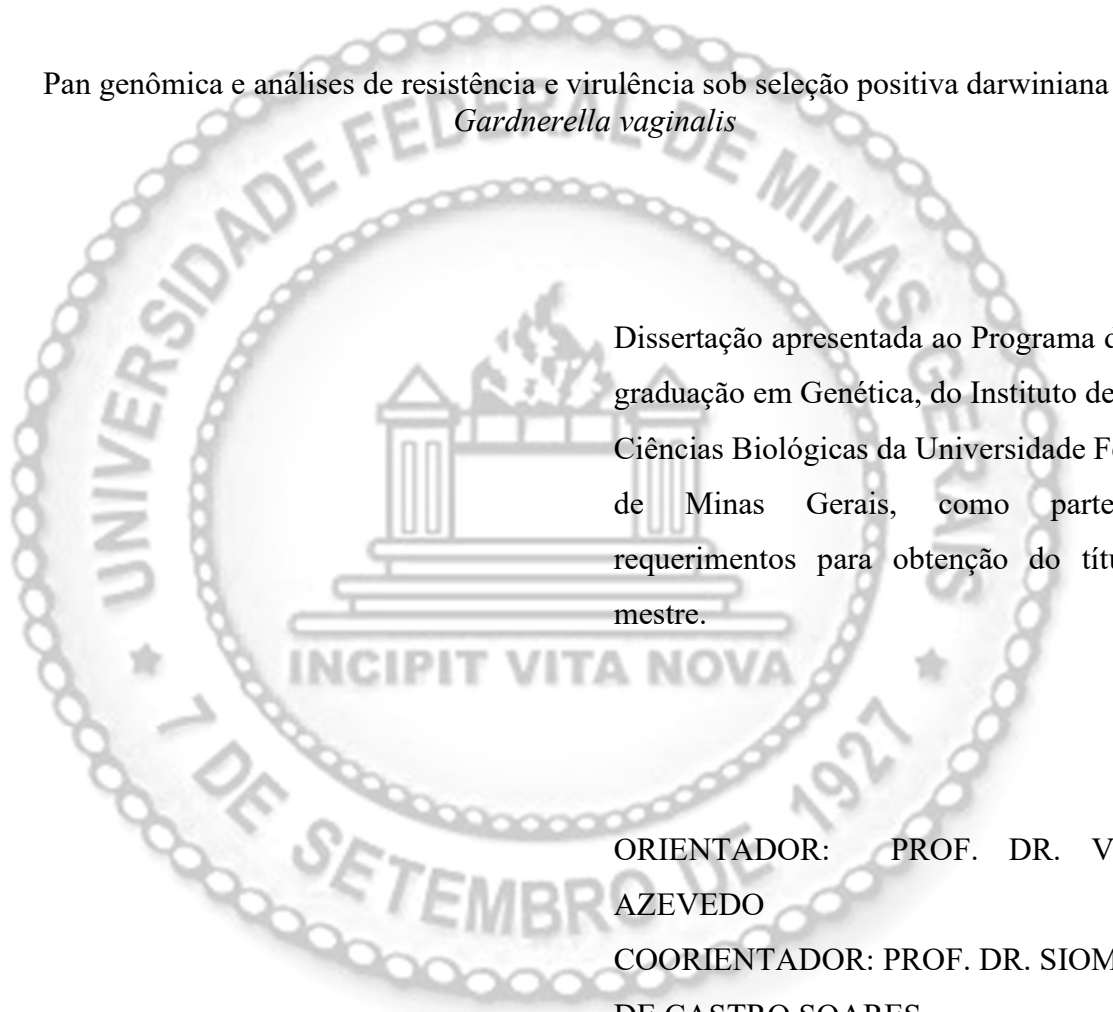
ORIENTADO: EDUARDA GUIMARÃES SOUSA  
ORIENTADOR: PROF. DR. VASCO AZEVEDO  
COORIENTADOR: PROF. DR. SIOMAR DE CASTRO SOARES

BELO HORIZONTE

2024

EDUARDA GUIMARÃES SOUSA

Pan genômica e análises de resistência e virulência sob seleção positiva darwiniana de  
*Gardnerella vaginalis*



Dissertação apresentada ao Programa de Pós-graduação em Genética, do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como parte dos requerimentos para obtenção do título de mestre.

ORIENTADOR: PROF. DR. VASCO AZEVEDO

COORIENTADOR: PROF. DR. SIOMAR DE CASTRO SOARES

BELO HORIZONTE

2024

043

Sousa, Eduarda Guimarães.

Pan genômica e análises de resistência e virulência sob seleção positiva darwiniana de *Gardnerella vaginalis* [manuscrito] / Eduarda Guimarães Sousa. – 2024.

115 f. : il. ; 29,5 cm.

Orientador: Vasco Ariston de Carvalho Azevedo. Coorientador: Siomar de Castro Soares.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Bioinformática. 3. *Gardnerella vaginalis*. 4. Vaginose Bacteriana. 5. Genômica Comparativa. 6. Seleção Genética. I. Azevedo, Vasco Ariston de Carvalho. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 575.1



UNIVERSIDADE FEDERAL DE MINAS GERAIS

**ATA DE DEFESA DE DISSERTAÇÃO / TESE**

Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Genética

**ATA DA DEFESA DE DISSERTAÇÃO**

**Eduarda Guimarães Sousa**

355/2024  
entrada  
2º/2022  
CPF: 092.161.846-80

Às quatorze horas do dia **23 de abril de 2024**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Pan genômica e análises de resistência e virulência sob seleção positiva Darwiana de Gardnerella vaginalis**", requisito para obtenção do grau de Mestre em **Genética**. Abrindo a sessão, o Presidente da Comissão, **Vasco Ariston de Carvalho Azevedo**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Vasco Ariston de Carvalho Azevedo	UFMG	283.171.225-49	APROVADA
Siomar de Castro Soares	UFTM	056.951.826-11	APROVADA
Pedro Milet Meirelles	UFBA	012.449.595-88	APROVADA
Anderson Miyoshi	UFMG	034.357.036-01	APROVADA

Pelas indicações, a candidata foi considerada: **APROVADA**

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

**Belo Horizonte, 23 de abril de 2024.**



Vasco Ariston de Carvalho Azevedo - Orientador (UFMG)

Siomar de Castro Soares (UFTM)

Pedro Milet Meirelles (UFBA)

Anderson Miyoshi (UFMG)

Assinaturas dos membros da banca:



Documento assinado eletronicamente por **Siomar de Castro Soares, Usuário Externo**, em 24/04/2024, às 09:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Anderson Miyoshi, Membro**, em 24/04/2024, às 10:42, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vasco Ariston de Carvalho Azevedo, Professor do Magistério Superior**, em 26/04/2024, às 10:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Pedro Milet Meirelles, Usuário Externo**, em 11/06/2024, às 13:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3195945** e o código CRC **B87B1975**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Genética

**"Pan genômica e análises de resistência e virulência sob seleção positiva  
Darwiana de *Gardnerella vaginalis*"**

**Eduarda Guimarães Sousa**

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Vasco Ariston de Carvalho Azevedo - Orientador  
UFMG

Siomar de Castro Soares  
UFTM

Pedro Milet Meirelles  
UFBA

Anderson Miyoshi  
UFMG

Belo Horizonte, 23 de abril de 2024.



Documento assinado eletronicamente por **Siomar de Castro Soares, Usuário Externo**, em 24/04/2024, às 09:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Anderson Miyoshi, Membro**, em 24/04/2024, às 10:42, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Vasco Ariston de Carvalho Azevedo, Professor do Magistério Superior**, em 26/04/2024, às 10:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Pedro Milet Meirelles, Usuário Externo**, em 11/06/2024, às 13:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3195978** e o código CRC **638FF6AF**.

---

Aos meus pais, que são meu alicerce.

## **AGRADECIMENTOS**

Agradeço primeiramente aos meus orientadores Professor Vasco e Professor Siomar que sempre foram solícitos e me permitiram crescer no meio acadêmico, sempre com bastante confiança e ensinamento. Sem os dois nada seria possível.

Agradeço também a cada um do Laboratório de Bioinformática da Universidade Federal do Triângulo Mineiro e do Laboratório de Genética Celular e Molecular da Universidade Federal de Minas Gerais. Os amigos que eu fiz durante esse processo foram essenciais no meu crescimento e amadurecimento, com muita disposição e sugestões sempre construtivas sobre meu trabalho. Em especial ao Dr. Marcus Canário e Professora Flávia Aburjaile por todo o ensinamento das metodologias e teorias para a confecção do projeto.

Agradeço muito a minha família que desde sempre estiveram presentes na minha vida, mesmo quando tudo estava sendo perdido eles conseguiram sempre me apoiar e me ajudar em todas minhas dificuldades e em todas minhas dúvidas sobre minha capacidade de exercer esse trabalho. Tenho muito carinho por cada um de vocês, José Márcio, Cíntia, Gabriela, Giovanna, Viki, Ana Elisa, Bruna, Marina, Sônia, Alcir, Jandira e Antônio. E ao meu namorado Gustavo, meu parceiro de vida, que sempre me apoiou emocionalmente e profissionalmente, sem você nada seria possível, obrigada por tornar meus dias mais fáceis.

E por fim, gostaria de agradecer meus apoios emocionais que sempre me escutam e me ajudam a conviver numa paz na rotina corrida de um curso de pós-graduação, obrigada Fernanda, Talita, Larissa, Diego, Juan, Gabriela e Júlia. Em especial gostaria de agradecer a Ana Luiza que sempre esteve ao meu lado desde a graduação e na construção do meu conhecimento. Não menos importante gostaria de agradecer aos meus amigos de coração que sempre torceram por mim, Victor Melazo, Victor Barbosa e Matheus Figueredo, e minha família de coração Dias, muito obrigada.

“Nada na vida deve ser temido, somente compreendido.  
Agora é hora de compreender mais para temer menos.”

Marie Curie (\* 07 de novembro de 1867 † 04 de julho de  
1934)

## RESUMO

O gênero *Gardnerella* inclui bacilos Gram-variáveis, anaeróbios, hemolíticos e imóveis. A principal espécie deste gênero é a *G. vaginalis*, responsável pela vaginose bacteriana (VB). Apesar da sua importância, especialmente para as mulheres, e da disponibilidade de genomas de várias estirpes em bases de dados públicas, a análise genômica e os estudos sobre tratamentos eficazes ainda carecem de detalhes. Neste estudo, propusemo-nos a realizar análises bioinformáticas utilizando 97 genomas da espécie, focando principalmente na pan-genômica para investigar o conjunto completo de genes e a seleção positiva darwiniana para analisar quais os genes que foram selecionados positivamente durante a evolução e que têm atividades relacionadas com fatores antimicrobianos e de virulência. A filogenômica, combinada com a epidemiologia molecular, proporcionou uma maior resolução do que o wgMLST na identificação de micro eventos de evolução e complexos clonais da espécie; a análise da plasticidade genômica visou identificar eventos de aquisição e exclusão através da transferência horizontal de genes. Os resultados revelaram uma variabilidade significativa do genoma, indicando um pangenoma aberto altamente diversificado. O baixo número de genes no genoma central e a análise de similaridade do mapa de calor confirmaram esta variabilidade. Foram identificadas três ilhas de patogenicidade, duas ilhas de resistência e nove ilhas genômicas, o que revela eventos de transferência horizontal de genes. Além disso, as análises de genômica comparativa identificaram genes de resistência, como o *Mef(A)*, relacionado com a resistência aos macrolídeos, e outros genes, como o *tet(M)* e o *tet(L)*, relacionados com a resistência às tetraciclina. Estes genes foram associados à seleção positiva darwiniana, indicando a sua importância na adaptação e sobrevivência da espécie. Duas proteínas também foram selecionadas como alvos terapêuticos por meio de Docking molecular, como WP\_004132099.1 (*sigA*), que é um fator sigma da RNA polimerase necessário para o processo de transcrição bacteriana; e WP\_004131683.1 (*murB*) relacionada à formação de peptidoglicano na parede celular procariótica. Estes resultados sublinham a necessidade de sequenciar novos genomas de *G. vaginalis* a uma escala global para compreender melhor esta variabilidade e os padrões de adaptação desta espécie no seu ambiente para desenvolver tratamentos eficazes contra a vaginose bacteriana.

**Palavras-chave:** Bioinformática; *Gardnerella vaginalis*; Vaginose bacteriana; Genômica comparativa; Seleção Positiva Darwiniana.

## ABSTRACT

The *Gardnerella* genus comprises Gram-variable, anaerobic, hemolytic, and non motile bacilli. The main species of this genus is *G. vaginalis*, responsible for bacterial vaginosis (BV). Despite its importance, especially for women, and the availability of several strain genomes in public databases, genomic analysis and studies on effective treatments still lack detail. In this study, we set out to carry out bioinformatic analyses using 97 genomes of the species, mainly focusing on pan-genomics to investigate the complete set of genes and Darwinian positive selection to analyze which genes have been positively selected during evolution and have activities related to antimicrobial and virulence factors. Phylogenomics, combined with molecular epidemiology, provided greater resolution than wgMLST in identifying micro-events of evolution and clonal complexes of the species, genomic plasticity analysis aimed to identify events of acquisition and exclusion through horizontal gene transfer. The results revealed significant genome variability, indicating a highly diverse open pangenome. The low number of genes in the core genome and the heatmap similarity analysis confirmed this variability. Three pathogenicity islands, two resistance islands, and nine genomic islands were identified, showing horizontal gene transfer events. In addition, comparative genomic analyses identified resistance genes, such as *Mef(A)*, related to macrolide resistance, and other genes, such as *tet(M)* and *tet(L)*, related to tetracycline resistance. These genes were associated with Darwinian positive selection, indicating their importance in the adaptation and survival of the species. Two proteins were also selected as therapeutic targets through molecular modeling, such as WP\_004132099.1 (*sigA*), which is a sigma factor of RNA polymerase necessary for the bacterial transcription process; and WP\_004131683.1 (*murB*) is related to peptide-glycan formation in the prokaryotic cell wall. These findings emphasize the need to sequence new *G. vaginalis* genomes on a global scale to better understand this variability and the adaptation patterns of this species in its environment for developing effective treatments against bacterial vaginosis.

**Keywords:** Bioinformatics; *Gardnerella vaginalis*; Bacterial vaginosis; Comparative genomics; Darwinian positive selection.



## LISTA DE ILUSTRAÇÕES

<b>Figura 1:</b> Representação do microbioma vaginal na saúde e na vaginose bacteriana.....	19
<b>Figura 2:</b> Principais componentes importantes na resposta imune contra à vaginose bacteriana.....	21
<b>Figura 3:</b> Representação esquemática da cura e da recorrência da vaginose bacteriana após tratamento com antibióticos.....	23
<b>Figura 4:</b> Mapa de calor com 7 genomas completos de <i>G. vaginalis</i> e 9 genomas representativos de clados.....	39
<b>Figura 5:</b> Árvore filogenética mostra a percentagem de bootstrap com 1000 bootstraps na árvore.....	40
<b>Figura 6:</b> Previsão da plasticidade genômica com ilhas de patogenicidade (PAI), ilhas de resistência (RI) e ilhas metabólicas (MI) de <i>G. vaginalis</i> .....	41
<b>Figura 7:</b> Análise de sintenia gênica.....	42
<b>Figura 8:</b> Desenvolvimento do pangenoma.....	43
<b>Figura 9:</b> Classificação funcional COG do <i>core</i> genoma.....	44
<b>Figura 10:</b> Classificação funcional COG dos genes <i>shared</i> e <i>singletons</i> .....	44
<b>Figura 11:</b> Fluxo de trabalho das metodologias utilizadas para selecionar os candidatos a medicamentos e os genes relacionados com a seleção positiva Darwiniana.....	60
<b>Figura 12:</b> Análise do pan-resistoma e dos genes de resistência utilizando o software PRAP.....	61
<b>Figura 13:</b> Estrutura terciária de WP_004132099.1 e WP_004131683.1 prevista por Alphafold.....	62
<b>Figura 14:</b> Análises moleculares de docking dos novos alvos de medicamentos propostos.....	63

## LISTA DE TABELAS

### CAPÍTULO I

**Tabela 1:** Informações sobre os 16 genomas das estirpes de *G. vaginalis*.....35

### CAPÍTULO II

**Tabela 1:** Informações sobre os 97 genomas de estirpes de *G. vaginalis*.....55

**Tabela 2:** Proteínas selecionadas para alvos de medicamentos com os seus compostos naturais previstos a partir da base de dados ZINC, ligações de hidrogénio e resíduos analisados pelo software Chimera.....64

## LISTA DE ABREVIATURAS

2D	Two dimensions
3D	Three-dimensional
BLASTn	Nucleotide basic local alignment search tool
BLASTp	Protein basic local alignment search tool
DEG	Database of Essential Genes
DNA	Deoxyribonucleic acid
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas
GC	Guanine Cytosine
GIPSy	Genomic Island Prediction Software
ICB	Instituto de Ciências Biológicas
ICBN	Instituto de Ciências Biológicas e Naturais
IgG	Immunoglobulin A
IL	Interleukin
ILE	Isoleucine
NCBI	National Center for Biotechnology Information
PDB	Protein data bank
PDBQT	Protein Data Bank, Partial Charge (Q), Atom Type (T)
PPA	Profile-Profile threading Alignment
UFMG	Universidade Federal de Minas Gerais
UFTM	Universidade Federal do Triângulo Mineiro.
VB	Bacterial vaginosis

## **APRESENTAÇÃO DO TRABALHO DE TESE**

### **Colaboradores**

Este trabalho foi conduzido sob orientação do Prof. Dr. Vasco Azevedo e coorientação do Prof. Dr. Siomar de Castro no Laboratório de Genética Celular e Molecular (LGCM), Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte.

O grupo de pesquisa (LGCM) está ativamente envolvido em abordagens de bioinformática em genômica de bactéria. Pioneiro em pesquisas em bioinformática no Brasil, o grupo se engajou em projetos de pesquisa intensivos abrangendo diversas áreas da biologia como genômica, transcriptômica e o desenvolvimento de vacinas e diagnósticos que tornaram o grupo referência para o estudo de microrganismos.

Este trabalho apresentou apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

## **PREFÁCIO**

O delineamento da dissertação está organizado nas seguintes seções:

**I - INTRODUÇÃO:** Busca trazer informações relevantes sobre o trabalho, introduzindo o gênero *Gardnerella* e a espécie *Gardnerella vaginalis*, que é o principal microrganismos desse gênero responsável pela vaginose bacteriana, a qual foi descrita de acordo com sua epidemiologia, diagnóstico, manifestação clínica e tratamento. Ademais, também traz informações sobre as metodologias que foram utilizadas na confecção desse projeto para elucidar melhor a importância de cada análise realizada.

**II - JUSTIFICATIVA:** Atualmente, há 117 genomas distintos de várias linhagens de *G. vaginalis* arquivados em repositórios online, o que sublinha a relevância dessa espécie. No entanto, apesar de sua importância médica significativa, especialmente para as mulheres, há uma escassez de estudos abordando o perfil genômico dessa espécie, particularmente no âmbito da bioinformática voltada para a genômica comparativa. Essas pesquisas têm o potencial de revelar padrões de distribuição da espécie através de modernas técnicas de epidemiologia molecular de genoma completo, e também de identificar possíveis fatores de virulência e resistência associados a condições patológicas e sujeitos à seleção positiva darwiniana.

**III – OBJETIVO DA TESE:** Realizar análises de Bioinformática, como genômica comparativa, pan-genômica, plasticidade genômica, sintenia gênica e genômica subtrativa, em linhagens de *Gardnerella vaginalis* para identificar alvos de drogas contra esse microrganismo, incluindo a investigação de genes de resistência e virulência sob seleção positiva Darwiniana.

**IV – CAPÍTULO I:** Este artigo de pesquisa descreve a diversidade gênica de linhagens de *Gardnerella vaginalis* utilizando uma abordagem de pan-genômica; intitulado: Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An *In Silico* Approach. Publicado em: Venereology (MDPI).

**V – CAPÍTULO II:** Este artigo de pesquisa descreve sobre genes de resistência e virulência relacionados com a seleção positiva Darwiniana e a seleção de proteínas que sejam potenciais alvos de drogas visto a resistência presente, intitulado: Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for *Gardnerella vaginalis*. Submetido no Journal of Biomedical Science.

**VI – CONCLUSÃO GERAL:** Sumariza os principais achados da Tese.

**VII – PERSPECTIVAS:** Traz as principais ideias a serem exploradas para a continuidade deste trabalho.

**VIII – REFERÊNCIA BIBLIOGRÁFICAS:** Lista as referências que foram utilizadas no desenvolvimento deste trabalho.

**IX – MATERIAL SUPLEMENTAR:** Lista todos os materiais suplementares, como tabelas e figuras que foram feitas para a confecção de cada artigo de pesquisa.

**X – APÊNDICE:** Lista os trabalhos realizados durante esse período de mestrado.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>17</b>
1.1 GÊNERO <i>GARDNERELLA</i> .....	17
1.2 <i>GARDNERELLA VAGINALIS</i> .....	18
1.3 FATORES DE VIRULÊNCIA RELACIONADOS À <i>GARDNERELLA VAGINALIS</i> .....	18
1.4 VAGINOSE BACTERIANA.....	19
1.4.1 Epidemiologia da vaginose bacteriana.....	20
1.4.2 Resposta imune à vaginose bacteriana.....	20
1.4.3 Diagnóstico e tratamento.....	22
1.5 BIOINFORMÁTICA.....	23
1.5.1 Genômica comparativa.....	24
1.5.2 Filogenômica.....	25
1.5.3 Pan genômica.....	26
1.5.4 Plasticidade genômica.....	27
1.5.5 Sintenia gênica.....	27
1.5.6 Genômica subtrativa.....	28
1.5.7 Docking Molecular.....	28
1.6 SELEÇÃO POSITIVA DARWINIANA.....	29
<b>2. JUSTIFICATIVA DA PESQUISA DA TESE</b> .....	<b>30</b>
<b>3. OBJETIVO DA TESE</b> .....	<b>31</b>
3.1 GERAL.....	31
3.2 ESPECÍFICOS.....	31
<b>CAPÍTULO I</b> .....	<b>32</b>
<b>4. ARTIGO CIENTÍFICO 1: Pangenome Analysis Reveals a High Degree of Genetic Diversity in <i>Gardnerella vaginalis</i>: An <i>In Silico</i> Approach</b> .....	<b>33</b>
<b>CONCLUSÃO CAPÍTULO I</b> .....	49
<b>CAPÍTULO II</b> .....	<b>50</b>

<b>5. ARTIGO CIENTÍFICO 2: Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for <i>Gardnerella vaginalis</i></b> .....	<b>51</b>
<b>CONCLUSÃO CAPÍTULO II</b> .....	<b>70</b>
<b>7. CONCLUSÃO GERAL</b> .....	<b>71</b>
<b>8. PERSPECTIVAS</b> .....	<b>72</b>
<b>9. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>73</b>
<b>11. ANEXOS</b> .....	<b>90</b>
11.1 FIGURAS.....	90
11.2 MATERIAL SUPLEMENTAR.....	102
<b>APÊNDICE</b> .....	<b>103</b>



## 1. INTRODUÇÃO

### 1.1 GÊNERO *GARDNERELLA*

Por volta de 1953, Leopold (LEOPOLD, 1953) descreveu um microrganismo com características Gram-negativas, hemolítico em ágar sangue, microaerofílico e que podia ser isolado tanto de homens na urina quanto no colo do útero feminino (Vanechoutte et al., 2019). Cerca de um ano depois, entre 1954 e 1955, Gardner e Dukes (Gardner & Dukes, 1954, 1955) consideraram que essas características eram semelhantes com *Haemophilus*, descrevendo e classificando tal espécie como *Haemophilus vaginalis*. Anos mais tarde, em 1980, após análises adansonianas, métodos de hibridização DNA-DNA, microscopia eletrônica e análises bioquímicas contra organismos conhecidos de outros gêneros, como *Haemophilus*, *Pasteurella* e *Streptococcus*, Greenwood e Pickett (Greenwood & Pickett, 1980) acreditaram que não haviam relações entre *H. vaginalis* com espécies desses gêneros, e então na ausência de gêneros existentes com tais características, propuseram a criação de um novo gênero chamado *Gardnerella*: *G. vaginalis* (Gardner e Dukes).

Nesse novo gênero, *G. vaginalis*, passou a ser altamente estudado por estar relacionado a mais de 50% dos casos de vaginose bacteriana (VB) encontrados em mulheres. Com o tempo, estudos começaram a avaliar a diversidade das espécies isoladas de *G. vaginalis*, através da amplificação do gene 16S do rRNA. Esses estudos (Ingianni et al., 1997; Lopes Dos Santos Santiago et al., 2011; Pleckaityte et al., 2012; Schellenberg et al., 2016) começaram a identificar genótipos distintos, até que Ahmed et al (Ahmed et al., 2012), mostrou através de análises de agrupamentos vizinhos, com dados de genes distribuídos e centrais, que os clados não-recombinantes tinham pools de genes distintos, o que garantiu a descrição de pelo menos quatro espécies, *G. vaginalis* e mais três novas. Essas espécies foram melhor caracterizadas mais tarde por Vanechoutte et al, em 2019 (Vanechoutte et al., 2019), e separadas e denominadas como *Gardnerella vaginalis*, *Gardnerella leopoldii* sp. nov. *Gardnerella piotii* sp. nov., e *Gardnerella swidsinskii* sp. nov..

Todas essas espécies descritas apresentam relação com sintomas do quadro clínico típico de VB, porém algumas tem maior relação com alguns critérios de Nugent/Amsel (Plourd, 1997) que outras. *G. leopoldii* tem associação com esses critérios e abundâncias relativas menores que as outras espécies. Os sintomas de odor anormal e corrimento vaginal estão mais ligados a *G. vaginalis* e *G. swidsinskii*, do que com outras espécies, por exemplo (Hill & Albert, 2019).

## 1.2 *GARDNERELLA VAGINALIS*

Dessas 4 espécies classificadas recentemente dentro do gênero *Gardnerella*, essa dissertação terá enfoque na principal espécie descrita, a *Gardnerella vaginalis*, o mais importante patógeno causador da vaginose bacteriana. Essas bactérias fastidiosas possuem características de um bacilo imóvel, com cerca de 0,5 a 1,5  $\mu\text{m}$ , sendo não encapsulado, e que pode ou não apresentar fímbrias. Além dessas características, ele possui uma fina camada de peptidoglicano em sua parede, que o caracterizaria como um Gram-positivo. Contudo, essa fina camada pode deixá-lo descorar e se apresentar microscopicamente como se fosse um Gram-negativo, sendo por muitas vezes considerado como um Gram-variável. Macroscopicamente, bactérias dessa espécie crescem bem em ágar chocolate, incubados como anaeróbios e enriquecido com  $\text{CO}_2$ , entre 2 e 5 dias, e podem apresentar colônias lisas e esbranquiçadas/acinzentadas com  $\beta$ -galactosidase positiva (Hernández, 1998; Piot, 1991).

São bactérias hemolíticas, as quais apresentam essa atividade apenas em ágar sangue com glóbulos vermelhos de origem humana ou de coelho. São catalase e oxidase negativa. Com essas características morfológicas sua identificação é possível juntamente com a coloração cocobacilares Gram variáveis (Vanechoutte et al., 2019). Outras bactérias oriundas de secreções vaginais como *Actinomyces pyogenes* e *A. Turicensis* são bastante similares morfolologicamente com *Gardnerella vaginalis* e são classificadas como agentes similares a *Gardnerella* (Van Esbroeck et al., 1996).

## 1.3 FATORES DE VIRULÊNCIA RELACIONADOS À *GARDNERELLA VAGINALIS*

Alguns estudos já identificaram bem o potencial patogênico dessa espécie, muitas vezes subestimado. Esse potencial pode ser explicado pela sua capacidade de formação de biofilmes, importante para aumentar sua chance de sobrevivência dentro do organismo humano (Patterson et al., 2010). Essa bactéria também possui atividades de prolidase e sialidase, que são enzimas que vão funcionar como toxinas para aumentar a capacidade de adesão e destruição do tecido humano, e que por vezes já foram associadas a nascimento prematuro ou abortos (Briselden et al., 1992; Gelber et al., 2008; Lopes Dos Santos Santiago et al., 2011). Há ainda a vaginolisina (gene *VLY*), fator de virulência importante que atua causando lise de eritrócitos e distúrbios no sistema imune do hospedeiro (Cauci et al., 1993; Gelber et al., 2008; Pleckaityte et al., 2012).

A toxina citolítica também já foi descrita, em que Rottini, 1990, demonstra uma atividade sob os polimorfonucleares e células epiteliais, também sendo capaz de promover a lise das hemácias humanas (Rottini et al., 1990). Essas bactérias também podem ter a presença

de pili e uma camada de exopolissacarídeo, o que ajudam a aumentar o seu poder de adesão à células epiteliais humanas (Catlin, 1992; De Oliveira Silveira et al., 2010).

#### 1.4 VAGINOSE BACTERIANA

A microbiota vaginal é composta majoritariamente por lactobacilos de Döderlein (*Boletim Epidemiológico Aids - 2003 | Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis*, n.d.). Esses lactobacilos protegem a vagina contra patógenos com a produção de peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>), o que mantém seu pH ácido. Estudos mostram que o decréscimo desses lactobacilos deixa a vagina mais propensa a infestação de *G. vaginalis*, uma das causadoras da vaginose bacteriana (Leite et al., 2010; Schwebke, 2003; Wilson, 2004). A Vaginose Bacteriana (VB) é um distúrbio na microbiota vaginal, em que há predomínio de microrganismos anaeróbios e que, apesar de não causarem reações inflamatórias, causam um quadro clínico de corrimento vaginal de coloração branca, cinza ou amarelada, odor fétido e que pode aumentar com o coito ou menstruação. Se não tratada, pode causar problemas mais sérios como salpingites e endometrites nas mulheres; já nos homens, pode causar coceiras na região peniana e desconforto ao urinar (Fethers & Morris, 2001; Paulo et al., n.d.).

A VB pode ser classificada em tipo I, quando é causada predominantemente por *G. vaginalis*, ou tipo II, quando é causada por outras bactérias anaeróbias, como *Mobiluncus sp*, *Peptostreptococcus*, *Prevotella sp* e *Porphyromonas sp* (Nakamura et al., 2009). Com a ascensão dessas bactérias, principalmente no tipo II, pode surgir outras complicações decorrentes da VB, como endometrites, infecções pós cirúrgicas (pós-cesáreas), doença inflamatória pélvica aguda (DIPA), aminiorrexe, parto prematuro, entre outros (Gravett et al., 1986; Hay, 2000).



**Figura1. Representação do microbioma vaginal na saúde e na vaginose bacteriana.** Um microbioma vaginal em boa forma é caracterizado pela prevalência de tipos de bactérias do gênero *Lactobacillus*, os quais produzem

ácido láctico e peróxido de hidrogênio, criando um ambiente ácido. Na condição de vaginose bacteriana, o microbioma vaginal é amplamente ocupado por variedades de bactérias anaeróbias, que interagem e formam um biofilme polimicrobiano no revestimento vaginal. Os sintomas típicos dessa infecção incluem a presença de secreção vaginal, odor forte, e um aumento no pH vaginal.

Fonte: Dados obtidos de Sousa et al., 2023.

#### **1.4.1 Epidemiologia da vaginose bacteriana**

Kenyon e colaboradores em 2013 fizeram uma revisão de literatura dos dados encontrados em artigos de vaginose ao redor do mundo, entre os anos de 1990 e 2009. As idades avaliadas nos estudos encontrados por eles variaram de 15 a 72 anos, e pode se perceber que a presença de VB foi maior em países que tinham uma população estudada na cor preta, e com média de 30 anos na faixa etária. A África subsaariana foi o país com a maior porcentagem de presença (50%) do grupo estudado com os indicadores para VB (Kenyon et al., 2013).

Outro estudo, norte-americano, ocorrido entre 2001 e 2004, realizou testes com 4646 mulheres de idades entre 14 e 49 anos, e identificou que 29,2% do grupo estudado apresentava vaginose bacteriana, porém apenas ~16% do grupo relatava sintomas. A prevalência foi maior em negros não-hispânicos e proporcional ao aumento de parceiros sexuais ao longo da vida. Todo o processo de triagem das participantes desse estudo seguiu os critérios de Nugent (Koumans et al., 2007).

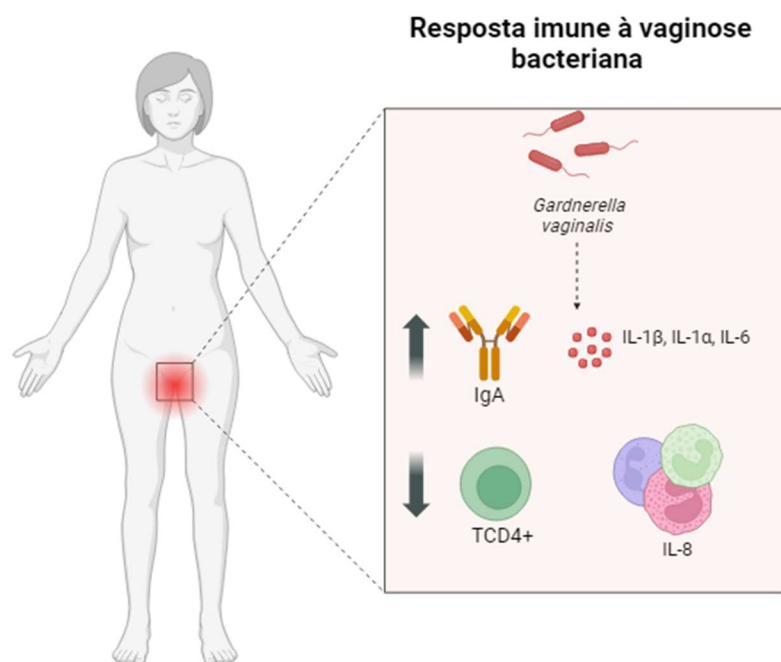
No Brasil um estudo com 277 participantes mulheres residentes do estado do Recife, os quadros clínicos de corrimento vaginal em 74,4% das participantes e odor de peixe na secreção vaginal em 68,6%, foram os mais frequentes. O quadro ocorreu principalmente em mulheres com faixa etária de 25 a 34 anos, sendo este grupo cerca de 50% do total de participantes estudados. Após as citologias e coloração de Gram, *Gardnerella vaginalis* foi encontrada em 96,8% das participantes, sendo seguida por *Mobiluncus*, com presença em 53,1% dos casos (Leite et al., 2010).

#### **1.4.2 Resposta imune à vaginose bacteriana**

Alguns estudos encontraram um tipo de resposta formado por IgA específico para *G. vaginalis* e altos níveis de IL-1 $\alpha$ , normalmente responsável pela formação da inflamação, causando febre e sepse (Di Paolo & Shayakhmetov, 2016), na secreção vaginal de mulheres portadoras (Cauci et al., 1993; Platz-Christensen et al., 1993). Outros estudos encontraram níveis de IL-1 $\beta$ , também responsável pelo início da inflamação, aumentado em até 20 vezes. Além da estimulação da produção de citocinas como IL-1 $\alpha$  e IL-1 $\beta$ , citocinas pró- inflamatórias,

alguns estudos demonstram que a VB também foi associada a um aumento nas células-alvo do HIV-1 na mucosa genital. Algumas células de mulheres quenianas, como T CD4 +, CCR5 +, CD69 +, foram encontradas diminuídas no colo do útero (Rebbapragada et al., 2008). Já em mulheres brasileiras, células T CD4 + estavam diminuídas quando em comparação com mulheres sem infecções vaginais (Giraldo et al., 2012; Mitchell & Marrazzo, 2014).

Um estudo sobre mulheres com genótipos suscetíveis (genes recessivos) a nascimento prematuro mostrou que alguns genes tiveram aumento de expressão, como por exemplo de IL-6, e que as razões de probabilidade variaram de 1,9 a 4 para mulheres que tinham pesquisa de VB positiva (Gómez et al., 2010). Outra citocina, a interleucina 8 (IL-8), principal mediador da resposta imune inata e fator ativador dos leucócitos polimorfonucleares (LPN), normalmente não está aumentada na vaginose bacteriana, e junto com a ação da vaginolisina, pode explicar ausência dos LPN no combate a esta doença (Cauci et al., 1993; Sobel, 2005).



**Figura 2. Principais componentes importantes na resposta imune contra à vaginose bacteriana .** A imagem retrata o complexo de proteínas e citocinas pró-inflamatórias encontrados no processo inflamatório de vaginose bacteriana causada por *Gardenerella vaginalis*.

Fonte: Do autor, 2024.

### 1.4.3 Diagnóstico e tratamento

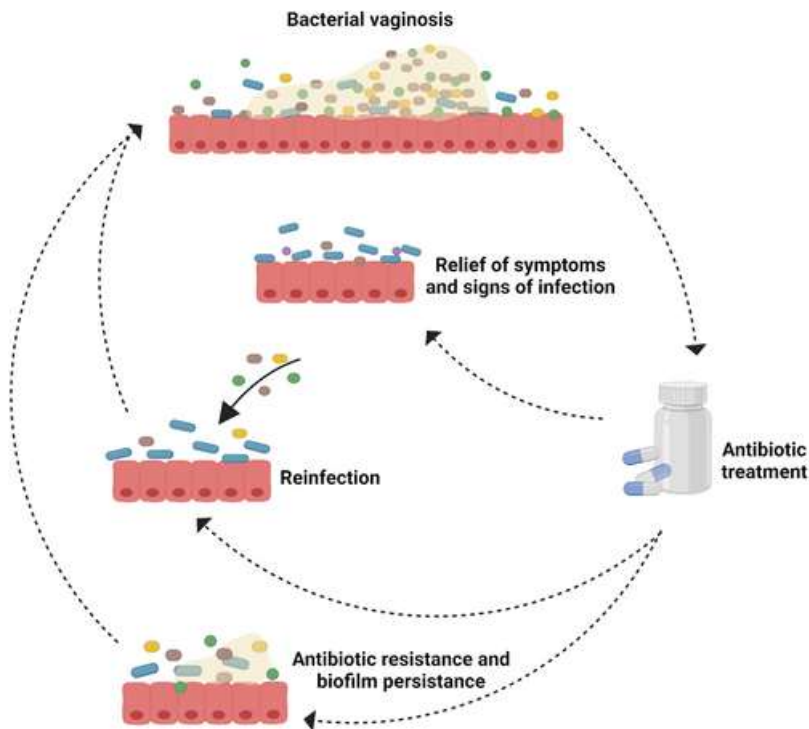
Segundo os critérios de Amsel/Nugent, o diagnóstico de VB é confirmado quando há a presença de pelo menos três dos quatro sinais: I) teste de amina positivo; II) presença de “clue-cells”; III) secreção acinzentada ou amarelada e, IV) pH vaginal superior a 4,7 (Fethers & Morris, 2001; Hawes et al., 1996; Paulo et al., n.d.). Também pode ser usado para confirmatório do diagnóstico os exames de Papanicolau e a cultura qualitativa, porém só esses exames não confirmam diagnóstico (Plourd, 1997).

O tratamento da VB tem o intuito de reestabelecer a microbiota normal da vagina, diminuindo microrganismos anaeróbios e, conseqüentemente, aumento dos lactobacilos. Alguns nitroimidazólicos costumam ser usados no combate da vaginose, sendo o principal representante o metronidazol, que é um antibiótico e antiparasitário usado em dose de 400 mg de 12 em 12 horas, via oral, e tem grande atuação contra Gram positivos e negativos, inclusive *G. vaginalis*. Outros nitroimidazólicos, como o tinidazol e o secnidazol, podem ser usados em dose única de 2g/via oral e têm índices de cura semelhantes ao metronidazol. Estes costumam ser usados em VB não complicada (Paavonen et al., 2000). A clindamicina pode também ser usada em vulvovaginites intensas e/ou recorrentes com dose oral de 300mg a cada 12 horas, com eficácia semelhante aos antibióticos anteriores quando estes estão em tratamento prolongado (Verstraelen et al., 2012).

Por ser uma síndrome que pode ser causada por vários microrganismos, há uma gama de opções terapêuticas para seu tratamento. Alguns estudos buscam o controle fisiológico dessa vaginose por meio de ingestão de peróxido de hidrogênio por duchas vaginais, com dose única (Molteni et al., 2004; Reid, 2002).

Estudos vêm mostrando uma resistência aos principais medicamentos utilizados para o tratamento de vaginose bacteriana causado por *G. vaginalis*, é encontrado principalmente uma resistência à principal linha de tratamento contra a patogênese, o metronidazol (K. Zhang et al., 2022). Um estudo classificou diferentes linhagens de *G. vaginalis* em relação a sua susceptibilidade ou não à metronidazol, eles perceberam que haviam ainda algumas linhagens susceptíveis ao medicamento, o que poderia estar relacionado com a ancestralidade genética compartilhada de linhagens provindas de um mesmo local de isolamento (Rigo & Tasca, 2020). Todavia, há muitos relatos mostrando a forte resistência contra esse medicamento por este patógeno, *G. vaginalis* tem uma grande capacidade de persistir após o tratamento contra a VB, com taxas de recorrência de 76% que ocorrem nos 06 meses seguintes ao tratamento, tanto pela resistência ao antibiótico quanto pela formação de biofilmes (Javed et al., 2019).

Ademais, outros medicamentos que podem ajudar no tratamento contra a VB foram observada resistência também como estudos *in silico* e *in vitro* que demonstraram resistência aos macrólidos eritromicina na identificação dos genes *ermX* e *lsaC*, e os genes de resistência à tetraciclina *tetL* e *tetM* também (K. Zhang et al., 2022).



**Figura 3. Representação esquemática da cura e da recorrência da vaginose bacteriana após tratamento com antibióticos.** Uma recorrência pode acontecer tanto pela reinfecção com patógenos anaeróbios após o tratamento com antibióticos, quanto devido à resistência antimicrobiana das bactérias e à incapacidade dos antibióticos em eliminar o biofilme, que recupera sua capacidade de crescer e causar a infecção.

Fonte: Dados obtidos de Sousa et al., 2023.

## 1.5 BIOINFORMÁTICA

A bioinformática em seus primórdios, começou a se desenvolver numa época em que a estrutura do DNA e seu armazenamento de informação, não eram completamente esclarecidos, assim, nos anos 1950, muitas discussões eram geradas a partir dessa incerteza. As análises de aminoácidos, por sua vez, representaram um ponto de partida para a bioinformática, com a publicação da primeira sequência de uma proteína, a insulina (Avery et al., 1944). Onde surgiu o método de degradação de Edman, o qual permitiu o sequenciamento dessas e daí foram surgindo softwares para automatizar esse processo e poder identificar diversas proteínas no início da década de 1960 (Jaskolski et al., 2014).

Margaret Dayhoff (1925-1983), foi uma físico-química pioneira no campo da bioinformática, onde aplicou diversos métodos computacionais para sua tese de doutorado em eletroquímica, e daí ficou conhecida como “a mãe da bioinformática” (Moody, 2004). Um dos programas foi o *COMPROTEIN*, juntamente com Robert S. Ledley para determinar estrutura primária de proteína usando dados de sequenciamento de peptídeos de Edman (November, 2011; Oakley & Kimball, 1949). Mais tarde, em 1970, os biólogos Paulien Hogeweg e Ben Hesper cunharam o termo Bioinformática, para representar estudos de sistemas biológicos, utilizando a matemática, a biologia molecular e os algoritmos computacionais (Tapprich et al., 2021).

Com isso foram surgindo diversos pesquisadores como Needleman e Wunsch, que desenvolveram o primeiro algoritmo de programação dinâmica para alinhamentos de sequências de proteínas em pares (Needleman & Wunsch, 1970), e Dayhoff, Schwartz e Orcutt (*Atlas of Protein Sequence and Structure. Vol. 5, Supplement 3; [Edited by] Margaret O. Dayhoff. - University of Edinburgh, n.d.*) ao desenvolver o primeiro modelo probabilístico de substituições de aminoácidos.

Em 1976 foi adotado o primeiro método de sequenciamento de DNA por Maxam-Gilbert (Maxam & Gilbert, 1977), e posteriormente por Frederick Sanger em 1977 (SANGER & THOMPSON, 1953), onde até hoje os princípios das técnicas são utilizados com modificações nas tecnologias de Sequenciamento de Nova Geração. Atualmente, muitas ferramentas para analisar genomas, transcriptomas e proteomas são empregadas a fim de compreender e elucidar melhor os parâmetros genéticos e moleculares de cada uma das moléculas (Hert et al., 2008).

Além disso, tecnologias de sequenciamento de Nova Geração permitem analisar com precisão, uma maior quantidade de dados do que em tecnologias anteriores. Assim, com alta reprodutibilidade de resultados computacionais e profundo poder preditivo, ferramentas de bioinformática associadas as novas tecnologias de sequenciamento são fundamentais para analisar de forma precisa e automatizada, o “BigData” biológico. Com implicações profundas no poder preditivo e na reprodutibilidade dos resultados da bioinformática e mineração e gestão de dados, exigindo mais conhecimentos especializados da ciência da computação, para uma melhor análise biológica (Gauthier et al., 2019).

### **1.5.1 Genômica comparativa**

Segundo Sivashankari & Shanmughavel, em 2007, a genômica comparativa é classificada como uma comparação direta entre o conteúdo genético de dois ou mais organismos



afim de entender a biologia da espécie estudada (Sivashankari & Shanmughavel, 2007). Através dessa metodologia é possível identificar regiões codificantes do genoma estudado, além também de conseguir identificar relações filogenéticas e eventos evolutivos (Ogier et al., 2010; Rust et al., 2002).

A genômica comparativa pode e já foi usada de diferentes formas já descritas na literatura, como por exemplo, comparação entre genes de *Drosophila melanogaster* e de seres humanos, em que cerca de 548 genes mostraram ter algum tipo de homologia com genes que podem estar relacionados a doenças em humanos, como doenças cardiovasculares (Reiter et al., 2001). Tal metodologia também pode ser utilizada em estudos comparativos entre bactérias, como no caso de *Bacillus licheniformis*, que foi comparado com outras duas espécies do gênero para completar a montagem do seu genoma, e identificar eventos de rearranjo, que tiveram transferência horizontal de genes (Rey et al., 2004). Outra forma que a genômica comparativa pode ser usada, é para identificar padrões evolutivos, estilos de vida e de tamanho de genoma, como no caso de um estudo que percebeu que patógenos intracelulares têm maior propensão à perda gênica via evolução reducionista (Merhej et al., 2009). Essa abordagem também vem sendo utilizada em conjunto com outras metodologias da Bioinformática, como filogenômica, pan-genômica e vacinologia reversa, salientando sua importância (Benevides et al., 2017; Figueiredo et al., 2015; Merhej et al., 2009; Muzzi et al., 2007; Tettelin et al., 2005).

### **1.5.2 Filogenômica**

A filogenômica foi aperfeiçoada com o uso de marcadores nucleotídicos através de técnicas moleculares, para evitar os antigos problemas que eram causados pelas análises baseadas somente nos pontos de divergências entre as espécies, as quais geravam dados complicados devido as diferentes evoluções nessas regiões convergentes e divergentes (Delsuc et al., 2005). Essa metodologia pode ser definida como a junção da filogenética e genômica para reconstruir árvores de espécies confiáveis, analisar a distribuição e espalhamento bacteriano e predizer genes ortólogos e parálogos (Chan & Ragan, 2013; Ocaña & Dávila, 2011).

Essa nova era da filogenômica trouxe muitas mudanças nessas inferências devido a melhora dos dados encontrados, como conteúdo e ordem gênica; ortologia; e, estrutura da fita de DNA ou assinatura do DNA (Delsuc et al., 2005; Kumar et al., 2012). O software Gegenees (Ågren et al., 2012) é um dos programas que fazem inferências das divergências entre as espécies e também é usado para reconstrução de árvores filogenômicas. Este programa quebra

os genomas em sequências de tamanhos pré-definidos e através do algoritmo BLAST identifica os genes compartilhados e monta a matriz baseada tanto na similaridade de genes quanto nos conteúdos variáveis de cada linhagem. A abordagem gene a gene ou wgMLST, também é uma estratégia interessante, já que identifica sítios polimórficos em uma tipagem da sequência multilocus do genoma inteiro (Cody et al., 2014; Sheppard et al., 2012). O serviço online PGAdb- builder realiza essa análise, que também pode criar uma árvore filogenômica baseada na similaridade e diferença dessas espécies, gerando um dendograma com maior resolução que os métodos tradicionais (Benevides et al., 2017).

### 1.5.3 Pan genômica

A ideia introduzida por Tettelin e colaboradores em 2005 (Tettelin et al., 2005) era criar uma metodologia que pudesse dar uma visão mais ampla, tanto do gênero quanto da espécie. Outros trabalhos já utilizam essa metodologia desde então e com diferentes espécies, como por exemplo *Streptococcus pneumoniae* (Tettelin et al., 2005), *Methanobrevibacter smithi* (Hansen et al., 2011), *Corynebacterium diphtheriae* (Soares et al., 2013), *Corynebacterium pseudotuberculosis* (Soares et al., 2013), *Pantoea ananatis* (De Maayer et al., 2014), *Bartonella henselae* (*The Pan-Genome of the Zoonotic Neglected Pathogen Bartonella Henselae Reveals Two Groups with Different Patterns of Adaptation to Hosts*, n.d.), o gênero *Faecalibacterium* (Benevides et al., 2017), entre outros.

Neste contexto, o objetivo principal da pan-genômica é a comparação genômica de diferentes linhagens de uma dada espécie ou até mesmo gênero (Alcaraz et al., 2010; Snipen et al., 2009). O pan-genoma é definido como o repertório completo e não-redundante de genes de uma dada espécie ou gênero e é composto de três subgrupos: o *core* genoma, que agrupa todos os genes compartilhados do conjunto de dados estudado, normalmente envolvidos em processos celulares essenciais; o genoma acessório (também chamado de core estendido ou genoma dispensável), que consiste de genes que são compartilhados por duas ou mais linhagens, mas não estão presentes em todas as linhagens; e, os singletons, que são genes específicos para cada linhagem, ou seja, estão presentes em somente uma linhagem do conjunto estudado (Tettelin et al., 2005, 2008). Alguns estudos sugerem que normalmente os genes presentes no core estendido e/ou no singletons estão relacionados com processos de adaptação, tanto ao ambiente quanto ao hospedeiro (Medini et al., 2005).

#### **1.5.4 Plasticidade genômica**

Através da plasticidade genômica pode se observar o quanto organismos como procariotos possuem maior facilidade que eucariotos para realizar incorporações de DNA recebido de organismos via transferência horizontal de genes (Horizontal Gene Transfer ou HGT) (Boto, 2010).

Existem diversos mecanismos diferentes envolvidos em eventos HGT, exercendo um papel importante na evolução, devido a modificações causadas, através da incorporação de: plasmídeos, bacteriófagos, transposons, elementos de inserção e, até mesmo ilhas genômicas. Devido a características específicas da fonte e a mecanismos utilizados na incorporação genômica, regiões adquiridas horizontalmente possuem em comum: um desvio na assinatura genômica (conteúdo G+C e uso de códon), que reflete a assinatura genômica do organismo doador; a presença de sequências de inserção (IS) e/ou tRNAs flanqueadores, que, por sua vez, podem apresentar uma IS específica em suas regiões 3'-terminais; e, a presença de transposases (Azevedo et al., 2014). Além disso, ilhas genômicas podem estar ausentes em organismos do mesmo gênero ou espécie correlata e ainda possuem altas concentrações de genes com funções correlatas e específicas, as quais podem ser classificadas em: ilhas de resistência (RI), ilhas simbióticas (SI), ilhas metabólicas (MI) e ilhas de patogenicidade (PAI).

As RIs, albergam altas concentrações de genes de resistência a antibióticos (Krizova & Nemeč, 2010); SIs podem estar correlacionadas com a associação de bactérias a plantas hospedeiras (Barcellos et al., 2007); As MIs, possuem genes associados com a biossíntese de metabólitos secundários (Tumapa et al., 2008); e as PAIs possuem altas concentrações de fatores de virulência, e estão presentes em bactérias patogênicas e ausentes em espécies não-patogênicas do mesmo gênero e/ou espécie correlata, e que podem estar envolvidas na re-emergência de vários organismos patogênicos devido à inserção de novos determinantes de virulência (Dobrindt et al., 2000). As ilhas genômicas são muito interessantes para análises de genômica comparativa, principalmente para pan-genômica, já que causam incorporações de blocos de genes, por vezes instáveis, cheios de informações dessas linhagens doadoras.

#### **1.5.5 Sintenia gênica**

A sintenia gênica pode ser caracterizada pela presença de genes ortólogos em cromossomos correspondentes em diferentes espécies em análises comparativas (Tang et al., 2008; Xu et al., 2016). Dentro dessa área, também encontramos a colinearidade, que pode ser caracterizada por ser um tipo de sintenia em que os genes são conservados na mesma ordem

(Tang et al., 2008). Essa metodologia se tornou muito frequente em áreas da genômica comparativa por ajudar no aumento da precisão da identificação da homologia e dos rearranjos cromossômicos dentro e/ou entre os genomas (Ma et al., 2006). Diversos programas são úteis para essas análises, como por exemplo, ADHoRe (Vandepoele et al., 2002), MicroSyn (Cai et al., 2011), CYNTENATOR (Rödelsperger & Dieterich, 2010), OrthoCluster (Vergara & Chen, 2009), entre outros. O software Mauve (Darling et al., 2010), também importante ferramenta para essas análises, fragmenta os genomas em tamanhos pré- definidos e cria blocos colineares baseados nos eventos de rearranjos de bases dentro desses genomas.

### **1.5.6 Genômica subtrativa**

A genômica subtrativa é uma abordagem moderna utilizada na identificação de potenciais candidatos à alvos de drogas (Hossain et al., 2017). Basicamente, dois ou mais genomas são comparados e um conjunto de dados central é identificado, que cujos genes podem possuir informações fenotípicas e genéticas de cada linhagem (Jaiswal et al., 2021). Esses genes são direcionados para análises de homologia contra hospedeiro, para evitar reações adversas se forem usados como alvos de drogas. Esses genes não-homólogos são testados quanto à sua localização subcelular e os que foram identificados como proteínas citoplasmáticas, serão direcionadas à análise de essencialidade, ou seja, se estão ou não envolvidos em processos vitais ao organismo. Após isso, os alvos podem ser testados com fármacos para ver sua capacidade de ação *in vivo* através da avaliação da estrutura tridimensional pelas análises de docking molecular.

### **1.5.7 Docking Molecular**

O docking molecular pode ser entendido como o mecanismo de reconhecimento molecular entre moléculas baseado em estrutura de um alvo específico *in silico*. Vai ser estudo como os compostos químicos interagem com os seus alvos moleculares. Há uma grande aplicação do docking na identificação de novos compostos de interesse terapêutico, prevendo interações ligante-alvo em nível molecular (Pinzi & Rastelli, 2019). As chances de encontrar os candidatos a medicamentos foi melhorado diante de estudos de docking, visto de ser uma triagem virtual de milhões de compostos num tempo acessível, reduzindo assim os custos iniciais de identificação de resultados (Meng et al., 2011).

Um dos primeiros estudos envolvendo a descoberta de novos alvos de medicamento foi a de Kuntz et al. no início da década de 1980, onde descreveram um método computacional que permite a exploração de alinhamentos ligantes-receptores viáveis para as estruturas conhecidas

de heme-mioglobina, o qual empregaram o docking para prever conformações potenciais de complexos moleculares (Kuntz et al., 1982).

Ao longo dos anos foram desenvolvidos diversos programas baseados em diferentes algoritmos para estudos de docking molecular. Sendo cada vez mais importantes em pesquisas na procura de alvos de drogas. Esse processo é dividido em duas etapas: na previsão da conformação do ligante, bem como sua “pose”, que seria a posição e orientação dentro desses locais no sítio de ligação da proteína alvo, seguido da avaliação da afinidade de ligação (Wang et al., 2016).

## 1.6 SELEÇÃO POSITIVA DARWINIANA

A seleção natural proposta por Darwin no livro “A origem das espécies”, continua sendo a principal explicação para a variabilidade do fenótipo em organismos, sendo a principal teoria de evolução amplamente utilizada. Um dos pontos-chave dessa teoria é a seleção positiva Darwiniana, a qual parte do princípio dos fenótipos que foram mantidos durante a evolução (J. Zhang, n.d.). Utilizando a genômica é possível identificar por busca computacional grupos de genes homólogos que evoluem sob seleção positiva Darwiniana, e que geralmente têm uma alta taxa de substituição não-sinônima, para uma melhor compreensão no nível molecular (Hongo et al., 2015).

Uma análise convencional na biologia evolutiva é compreender a arquitetura fenotípica e a sua evolução em uma população. Estudos dessa natureza estão utilizando a riqueza de dados genômicos gerados pela genômica comparativa e sequenciamento de nova geração para extrair informações biológicas dos padrões de conservação entre elementos genômicos compartilhados entre os genomas. A utilização da seleção positiva Darwiniana em escala genômica tem auxiliado muito na análise de genes que estão envolvidos na especiação de cada população. A seleção positiva pode atuar em códons específicos em momentos específicos da história evolutiva, sendo que ao nível de códon é possível discriminar a relação entre substituição não sinônima e sinônima baseado na estimação de  $d_N/d_S$ . Para estudar esse processo, os cientistas desenvolveram modelos estatísticos de ramificação (Viana et al., 2022). Além disso, muitas pesquisas foram realizadas durante a “corrida armamentista” na detecção de genes envolvidos na relação co-evolutiva entre hospedeiro-patógeno de diversos microrganismos patogênicos como *Escherichia coli* (Chen et al., 2006), *Salmonella* (Soyer et al., 2009), *Staphylococcus* (Suzuki et al., 2012), *Streptococcus* (Lefébure & Stanhope, 2007) e *Trypanosoma brucei* (Emes & Yang, 2008).

## 2. JUSTIFICATIVA DA PESQUISA DA TESE

Atualmente 117 genomas de diferentes linhagens de *G. vaginalis* estão depositados em bancos de dados online, salientando a importância da espécie. Contudo, apesar de sua grande importância médica, especialmente para o sexo feminino, existem poucos estudos sobre o perfil genômico dessa espécie, principalmente trabalhos de Bioinformática voltados para genômica comparativa.

Há necessidade de mais estudos genômicos, com a utilização da bioinformática, para ajudar na investigação futura sobre esta espécie. Tais trabalhos podem ajudar a identificar os padrões de distribuição da espécie utilizando técnicas recentes de epidemiologia molecular de genoma completo e identificar potenciais fatores de virulência e resistência que possam estar relacionados com fatores patológicos e sob seleção positiva Darwiniana. Neste trabalho, são utilizadas várias metodologias para identificar mecanismos e a diversidade dos organismos para ajudar em futuras investigações com a espécie *G. vaginalis*. Além disso, a geração de virulência e de resistência a antibióticos exige a identificação de novos alvos de medicamentos, pelo que foram efetuadas, em simultâneo, análises de modelagem molecular e de acoplamento com compostos naturais.

### 3. OBJETIVO DA TESE

#### 3.1 GERAL

Realizar análises de Bioinformática, como genômica comparativa, pan-genômica, plasticidade genômica, sintenia gênica e genômica subtrativa, em linhagens de *Gardnerella vaginalis* para identificar alvos de drogas contra esse microrganismo, incluindo a investigação de genes de resistência e virulência sob seleção positiva Darwiniana.

#### 3.2 ESPECÍFICOS

- Identificação e homogeneização da anotação genômica das linhagens de *G. vaginalis*;
- Realização de análises filogenômicas;
- Predição e identificação de ilhas genômicas;
- Predição de blocos localmente colineares para sintenia gênica;
- Caracterização de genes ortólogos e classificação dos *clusters* desses genes nos subconjuntos pan-genoma, *core genoma* e *singletons*;
- Classificação dos subconjuntos de genes das linhagens nas categorias funcionais do COG. Identificação de genes ortólogos e seleção dos clusters de genes pertencentes ao *core genoma*, *shared* e *singleton*;
- Seleção dos orthogroups para análise estatística de Seleção Positiva Darwiniana;
- Análise funcional dos genes selecionados positivamente;
- Predição de alvos de drogas utilizando genômica subtrativa;
- Análise de essencialidade e de homologia à microbiota intestinal dos alvos.

# CAPÍTULO I



#### 4. ARTIGO CIENTÍFICO 1: Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An *In Silico* Approach

**Publicado em: Venereology (MDPI); Fator de impacto (atual): 4.7.**

- **DOI:** <https://doi.org/10.3390/venereology2040012>

O gênero *Gardnerella* foi proposto na década de 1980 por Greenwood e Pickett após análises abrangentes contra outros microrganismos. *Gardnerella vaginalis*, a espécie mais estudada, é Gram-variável, com potencial patogênico significativo, formando biofilmes e produzindo enzimas que podem causar danos ao tecido humano.

Além de *G. vaginalis*, foram descritas três novas espécies relacionadas à vaginose bacteriana (VB). A VB é causada pelo desequilíbrio na microbiota vaginal, onde há uma redução de lactobacilos e aumento de bactérias anaeróbicas como *G. vaginalis*. Essa condição pode causar sintomas desconfortáveis, como corrimento anormal e odor fétido.

A VB é uma infecção comum entre mulheres sexualmente ativas, com fatores de risco incluindo múltiplos parceiros sexuais e baixa escolaridade. O diagnóstico é feito através de exames como a citologia oncótica e requer a presença de sinais específicos, como teste de amina positivo e pH vaginal elevado.

O tratamento da VB visa restaurar o equilíbrio da microbiota vaginal, com opções terapêuticas incluindo o uso de antibióticos e duchas vaginais com peróxido de hidrogênio. No entanto, apesar dos tratamentos disponíveis, o número de casos de VB continua aumentando, destacando a importância de pesquisas adicionais sobre *G. vaginalis* e suas implicações clínicas.

O estudo de *G. vaginalis* é essencial devido à abundância de genomas depositados em bancos de dados, representando uma importância médica significativa, especialmente para as mulheres. Há uma urgência para mais investigações genômicas, incluindo o uso de bioinformática, a fim de aprimorar o entendimento sobre essa espécie. Tais esforços podem elucidar padrões de distribuição das diferentes linhagens utilizando técnicas avançadas de epidemiologia molecular de genomas inteiros, além de identificar possíveis fatores de virulência associados a condições patológicas.

Nesses estudos, diversas metodologias são empregadas para a identificação de mecanismos e a avaliação da diversidade dos organismos, contribuindo para a elaboração de estratégias mais eficazes de diagnóstico, prevenção e tratamento das condições relacionadas à presença de *G. vaginalis* na microbiota vaginal.

## Article

# Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An In Silico Approach

Andrei Giacchetto Felice <sup>1,\*</sup>, Eduarda Guimarães Sousa <sup>2</sup>, Fabiana Vieira Dominici <sup>1</sup>, Vasco Ariston de Carvalho Azevedo <sup>2</sup> and Siomar de Castro Soares <sup>1,\*</sup>

<sup>1</sup> Laboratory of Bioinformatics, Institute of Biological and Natural Sciences, Department of Microbiology, Immunology and Parasitology, Federal University of Triângulo Mineiro, Educational Center Campus, Uberaba 38025-180, MG, Brazil; fabianadominici88@gmail.com

<sup>2</sup> Cellular and Molecular Genetics Laboratory, Institute of Biological Sciences, Department of General Biology, Federal University of Minas Gerais, Pampulha Campus, Belo Horizonte 31270-901, MG, Brazil; eduardaguimaraessousa@gmail.com (E.G.S.); vascoariston@gmail.com (V.A.d.C.A.)

\* Correspondence: andreigf@hotmail.com (A.G.F.); siomar.soares@uftm.edu.br (S.d.C.S.)

**Abstract:** The genus *Gardnerella* comprises Gram-variable, anaerobic, hemolytic, and non-motile bacilli, with four known species, where *Gardnerella vaginalis* is the main species responsible for bacterial vaginosis (BV). However, quantifying this species is challenging due to a lack of data and underreporting. Despite its significance, particularly for women, and the availability of several genomes in online databases, genomic analyses and studies on effective treatments still lack details. This study aimed to conduct bioinformatic analyses focused on pangenomics to investigate the complete gene repertoire of the species. Genomes of the bacterium available in online databases were used for comparative genomics, genomic plasticity, gene synteny, and pangenome prediction analyses. The results revealed considerable genome variability, indicating a highly diverse pangenome. The low number of genes in the core genome and similarity analysis confirmed this variability. Three pathogenicity islands, two resistance islands, and nine genomic islands were identified, suggesting horizontal gene transfer events during evolution. These findings underscore the need for sequencing new *G. vaginalis* genomes to better comprehend its variability and adaptation patterns.

**Keywords:** *Gardnerella vaginalis*; bioinformatics; diversity



**Citation:** Felice, A.G.; Sousa, E.G.; Dominici, F.V.; Azevedo, V.A.d.C.; Soares, S.d.C. Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An In Silico Approach. *Venereology* **2023**, *2*, 132–146. <https://doi.org/10.3390/venereology2040012>

Academic Editors: Jacek C. Szepietowski and Qinxue Hu

Received: 21 August 2023

Revised: 18 September 2023

Accepted: 28 September 2023

Published: 30 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The genus *Gardnerella* was proposed in the 1980s by Greenwood and Pickett after conducting Adansonian analyses, DNA–DNA hybridization methods, electron microscopy, and biochemical analyses against microorganisms of other genera such as *Haemophilus*, *Pasteurella*, and *Streptococcus* [1]. They believed there was no relationship between *Haemophilus vaginalis*, previously proposed as the Gram-negative microorganism isolated in both men and women [1–3], and these bacteria.

The species *Gardnerella vaginalis* has a thin layer of peptidoglycan and is considered Gram-positive. However, this layer can bleach, making it appear Gram-negative and, thus, it is classified as Gram-variable. They are non-motile bacilli and may present fimbriae or not [4–6]. This bacterium has a well-understood but often underestimated pathogenic potential. It can form biofilms, hindering effective treatment and increasing its chance of survival in the host [7]. Additionally, it can produce enzymes with prolidase and sialidase activities, which function as toxins, enabling greater adhesion and destruction of human tissue, and are also related to cases of premature births and abortions [8–10]. Another important virulence factor is vaginolysin, which causes disturbances in the host immune system and lysis of erythrocytes [11–13].

In this new genus, the *G. vaginalis* species accounts for more than 50% of bacterial vaginosis (BV) cases in women and is the most recently studied [14]. Previous studies

allowed the description of three more new species, later characterized by Vaneechoutte et al. in 2019 [6], named *Gardnerella leopoldii* sp. nov., *Gardnerella biotic* sp. nov., and *Gardnerella swidsinskii* sp. nov. This classification was possible by analyzing the isolated *G. vaginalis* species using 16S RNA gene amplification [4,8,11,15,16] and neighboring cluster analysis by Ahmed et al. [17]. Additionally, all these species are related to cases of BV, with some species having a more significant relation to symptoms of the clinical condition than others, according to some Nugent/Amsel criteria [18].

The vaginal microbiota is composed mainly of Döderlein Bacillus, which are microorganisms of the genus *Lactobacillus* responsible for maintaining the acidic pH of the vagina, protecting against pathogens through the production of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>). An imbalance in this vaginal flora causes bacterial vaginosis (BV)-type infections, where there is a decrease in lactobacilli [18–20]. In association, an increase in anaerobic bacteria such as *G. vaginalis* is observed, resulting in a higher vaginal pH and making the patient susceptible to BV.

The main characteristics of BV, besides the increased pH, are the discomfort reported due to a white, gray, or yellowish vaginal discharge and a foul-smelling odor that may increase with coitus or menstruation. If left untreated, it can cause more severe problems such as salpingitis and endometritis in women, and in men, it can cause itching in the penile area and discomfort when urinating [21,22].

BV is a widespread infection reported among sexually active women; it is estimated that between 10% and 30% of women are diagnosed worldwide [23,24]. Additionally, countries with a high frequency of HIV report that the frequency of BV is 50% in women. Studies have shown that the primary pathogens related to vaginosis are *Candida* sp. (56.3%), followed by *G. vaginalis* (35%) [25]. Several risk factors are related to the development of this pathology, including having many sexual partners, using intrauterine devices (IUDs) and vaginal douches, and engaging in oral sex [26–29]. Moreover, a strong relationship exists between the very early initiation of sexual activity and a low educational level [27,30].

Through oncotic cytology, more specifically the Papanicolaou technique, the inflammatory and infectious processes can be diagnosed along with their etiological agents. This exam is usual in women's health routine every year; it is easy to perform, is a low-cost approach, and is an advantage that facilitates the diagnosis of BV and enables faster treatment [31]. However, a confirmatory diagnosis of BV requires the presence of at least three of four signs: (I) positive amine test; (II) presence of clue cells; (III) grayish or yellowish discharge; and (IV) vaginal pH higher than 4.7 [32–34].

BV treatment focuses primarily on restoring the balance of the vaginal flora by decreasing the number of anaerobic microorganisms and increasing the number of lactobacilli. There is a range of therapeutic options, including using antibiotics such as nitroimidazole or even establishing physiological control of the vaginosis through hydrogen peroxide ingestion via vaginal douching [3,34–39]. However, despite all these treatments, the number of BV cases has constantly increased over the years.

The study of *G. vaginalis* is necessary and essential, given the number of different genomes deposited in databases, and it is of great medical importance, especially for females. There is a need for more genomic studies, such as using bioinformatics, to assist in future research on this species. Such works can help identify the distribution patterns of the species using recent whole-genome molecular epidemiology techniques and identify potential virulence factors that may be related to pathological factors. In this work, several methodologies are used to identify mechanisms and the diversity of the organisms to help in future research with the species *G. vaginalis*.

## 2. Materials and Methods

### 2.1. Genome Information

RefSeq and NCBI (National Center for Biotechnology Information) databases (<https://www.ncbi.nlm.nih.gov/>-accessed on 9 August 2022) were used to download the genomes in “.fna” format to perform the genomic analyses of *G. vaginalis*. Sixteen



genomes were used for phylogenomic and phylogenetic analyses, along with seven complete genomes of *G. vaginalis* and nine representative genomes from different clades, according to the tree available at the NCBI (Table 1).

**Table 1.** Information about the 16 genomes of *G. vaginalis* strains.

Organism	Strain	Assembly Level	Size (Mb)	GC%	Genbank Access	Extraction Location
<i>Gardnerella vaginalis</i>	NCTC10287	Complete	1.67	41.4	GCA_900637625.1	-
<i>Gardnerella vaginalis</i>	GV37	Complete	1.75	41.8	GCA_001953155.1	Paris, France
<i>Gardnerella vaginalis</i>	FDAARGOS_568	Complete	1.72	41.3	GCA_003812765.1	-
<i>Gardnerella vaginalis</i>	409-05	Complete	1.62	42	GCA_000025205.1	-
<i>Gardnerella vaginalis</i>	ATCC 14018/ JCM 11026	Complete	1.67	41.4	GCA_001042655.1	-
<i>Gardnerella vaginalis</i>	ATCC 14019	Complete	1.67	41.4	GCA_000159155.2	-
<i>Gardnerella vaginalis</i>	HMP9231	Complete	1.73	41.4	GCA_000213955.1	-
<i>Gardnerella vaginalis</i>	DSM 4944	Chromosome	1.68	41.3	GCA_900105405.1	-
<i>Gardnerella vaginalis</i>	UMB0913	Scaffold	1.51	42.1	GCA_002861145.1	Maywood, United States of America
<i>Gardnerella vaginalis</i>	UMB0736	Scaffold	1.73	41.1	GCA_013315025.1	Maywood, United States of America
<i>Gardnerella vaginalis</i>	UMB0298	Scaffold	1.68	41.2	GCA_002861975.1	Maywood, United States of America
<i>Gardnerella vaginalis</i>	JCP7672	Scaffold	1.69	41.2	GCA_000414645.1	-
<i>Gardnerella vaginalis</i>	JCP8108	Scaffold	1.67	41.1	GCA_000414525.1	-
<i>Gardnerella vaginalis</i>	UMB0833	Scaffold	1.62	42.1	GCA_002861885.1	Maywood, United States of America
<i>Gardnerella vaginalis</i>	JCP7659	Scaffold	1.54	41.9	GCA_000414665.1	-
<i>Gardnerella vaginalis</i>	JCP8070	Scaffold	1.48	42.2	GCA_000414545.1	-

To perform the pangenome analysis, 107 genomes of *G. vaginalis* available in the NCBI (till the first half of 2022) in the “.faa” format were used, which were classified as follows: 7 complete genomes; 1 chromosome; 63 scaffolds; and 36 contigs. The annotation of each of these genomes was performed with the Prokka software (version 1.14.6, Tseemann, 2020) [40] using the default parameters plus the options: ‘addgenes’ and ‘rnammer’. More information about the 107 genomes is described in Supplementary Table S1.

In addition to the genomes of the bacteria under study, the reference genome of the organism *Alloscardovia omnicolens* (DSM21503) was downloaded in the same formats as mentioned above. It was added to all the analyses to function as an outgroup, which helps the analyses to be more parsimonious when analyzing the group of interest [41].

## 2.2. Phylogenomic Analysis and Phylogenetic Reconstruction

The software Gegenees (version 3.1, Segerman, 2019) [42] was used to perform the phylogenomic analyses; its methodology is based on the similarity analysis between the genomes through a strategy followed by genome sequence fragmentation and DNA alignment using a BLASTn algorithm to obtain a distance matrix and the phylogeny analysis.

The genome fragmentation occurs using user-defined lengths and step sizes. Here, we used a size of 500 bp and a step size of 500 bp too. Finally, a heatmap was generated, showing the percentage of similarity between the strains in a range of 0 to 100%.

In addition, an analysis to identify polymorphic genes and whether there is phylogenetic interference in a gene-by-gene analysis was performed using the online software PGAdb-builder [43] with a wgMLST (whole genome multilocus sequence typing) analysis. The resulting PGAdb profile was performed using 90% coverage and 90% identity filters. This profile was exported in the “.newick” format to the MEGA software (version 11, Kumar, 2021) [44] for generating a phylogenetic tree.

### 2.3. Genomic Plasticity via Identification of Genomic Islands

Plasticity analyses were performed with GIPSY (version 1.1.3, Soares, 2016) (Genomic Island Prediction Software) to identify genomic islands, genomic regions acquired via horizontal gene transfer (HGT), which contain information about the microorganism lifestyle. The prediction is based on several features, such as codon usage and G + C content; insertion sequences or flanking tRNAs flanking; presence of transposases; and varied size [45]. The genomic islands were classified into pathogenicity, resistance, and metabolic islands with their respective genes. For this analysis, one non-pathogenic organism was chosen to be compared with the reference organism chosen in this work.

The genomic islands identified with GIPSY were plotted using the BRIG software (version 0.95, Alikhan, 2011) (BLAST Ring Image Generator) [46], which performs a comparative analysis of all genomes against a reference genome. A circular figure was generated where each ring represents a different genome, and the blank areas represent deletion areas. The last ring exhibits the genomic islands.

### 2.4. Gene Synteny

The software Mauve (version snapshot\_2015\_02\_13 build 0, Darling Lab, 2015) [47], along with the progressive Mauve algorithm, was used for gene synteny analysis, where it was possible to identify possible gene rearrangement events. By fragmenting the genomes into pre-defined sizes, the software built multiple genome alignments identifying locally collinear blocks (LCB) plotted in a figure, which is usable for identifying these rearrangements.

### 2.5. Prediction of Orthologous Genes and Pangenome Development

Using the Orthofinder software (version 2.5.4, Emms, 2021) [48], the prediction of orthologous genes was performed, where all genomes were compared against all using the DIAMOND algorithm (version 2.0.14, Buchfink, 2021). This software uses the MCL program (Markov Clustering algorithm) for such prediction [49]. Through in-house script (orthofinder\_pangenome\_splitter.pl), three ortho groups were generated, which are core genome, which is composed of all genes that are commonly shared by all strains, involved mainly in essential cellular processes; shared, which is composed of genes that are present in two or more strains but not all; and singletons that are classified as strain-specific genes, which are present in only one strain. The genes in the latter subset are related to adaptation processes, both to the environment and the host [50,51].

To analyze the pangenome, Heap’s Law was used together with *in-house* scripts (pandev.pl) to estimate the fixed parameters and the least squares fitting of the exponential regression decay (core genome and singletons). For the fixed parameters from the pangenome, Heap’s Law is used. It is an empirical law, and it is represented by the formula  $n = k \cdot N^\alpha$ , where a given number of genes ( $n$ ) is calculated for a given number of genomes ( $N$ ), and  $k$  and  $\alpha$  ( $\alpha = 1 - y$ ) are free parameters. When  $\alpha > 1$ , according to the law, one can consider this pangenome closed, meaning the addition of genomes has no substantial effect in the total number of genes. When  $\alpha < 1$ , the pangenome is considered open, meaning there will be a significant increase in the number of genes for each new genome added to the analysis. The formula used for least-squares fit is represented by  $n = k \times \exp[-x/t] + tg\theta$ , where  $n$  is the number of genes, and  $k$ ,  $t$ , and  $tg\theta$  are free parameters. This law is utilized to



infer how many genes will compose the core genome after stabilized [52,53] and to estimate approximately how many genes are added by each new sequenced genome.

### 2.6. Orthology Assignments and Functional Annotation

The analyses of the core genome, shared genes, and singletons subsets were classified according to the Cluster of Orthologous Genes (COG), where the CDS of the subsets were aligned and classified according to the functional categories of the COG. For this, eggNOG-mapper (version 2.1.11, Cantalapiedra, 2023) was used, a database that performs this annotation according to orthologous genes [54]. Each sequence was mapped using either the hidden Markov model (HMM) or DIAMOND to align with the eggNOG database. The best matching sequence of the target sequence is classified according to its taxonomy and finally categorized and annotated according to gene ontology (GO) [55], KEGG pathways [56], and COG functional categories [57].

The functional categories are divided mainly into (1) information processing and storage; (2) signaling and cellular processes; (3) metabolism; and (4) poorly characterized and are subdivided into 17 subcategories [58].

## 3. Results

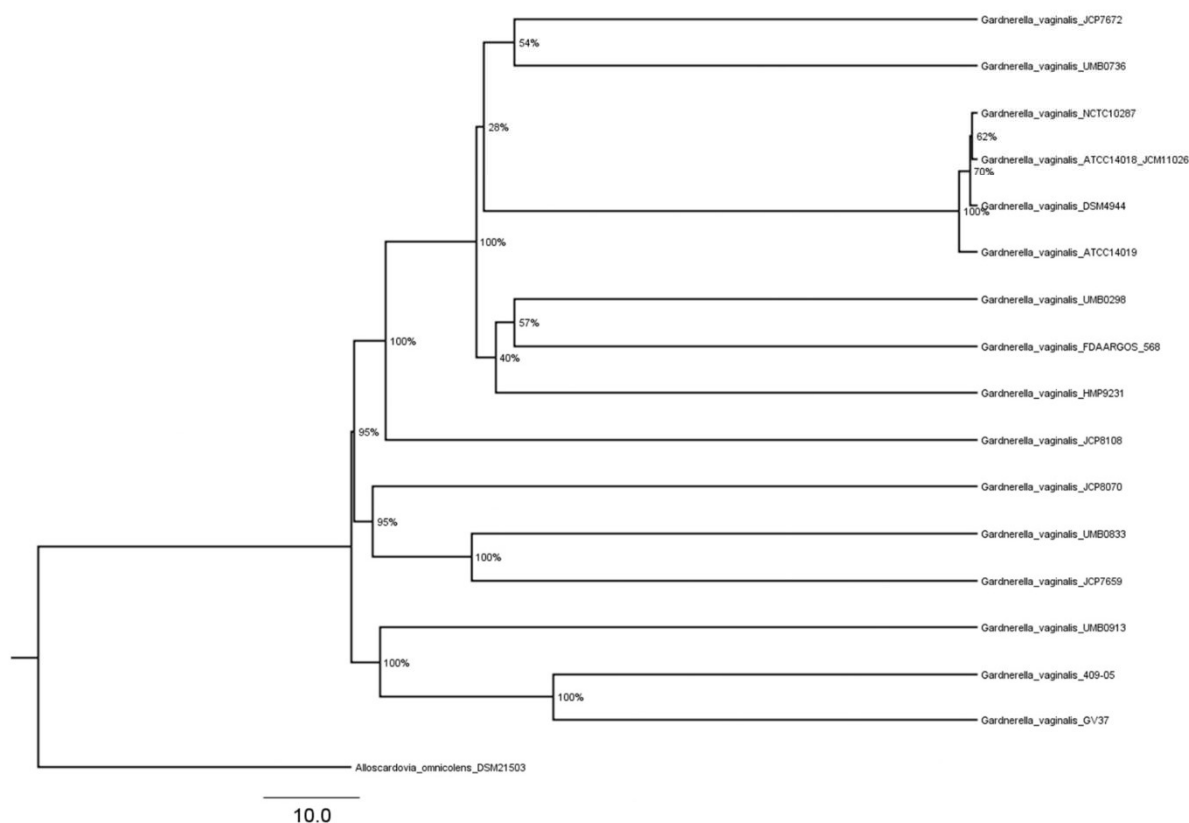
### 3.1. Phylogenetic Analysis

The Gegenees software analyses demonstrated a color variation from green to red showing the degree of similarity between the strains (Figure 1), where almost all the complete genomes are close with similarity ranging from 71 to 100% inside this cluster (cluster 1). Strains such as UMB0833, JCP7659, and JCP8070, scaffold genomes, grouped in cluster 2, with similarity ranging from 72 to 100% inside the cluster, with a medium similarity of 38 to 46% with cluster 1 and a low similarity with cluster 3 (15 to 19%). In addition, *G. vaginalis* GV37, 409-05 (both complete genomes), and UMBB0913, scaffold, grouped in cluster 3, showing a low similarity of approximately 18% with the other two clusters. The others represented in an orange shade had a median similarity and comprised scaffold genomes.

Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1: Alloscardovia_omnicolens_DSM21503	100	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2: Gardnerella_vaginalis_UMB0913	2	100	77	75	17	15	16	17	18	18	18	18	18	17	18	18	18
3: Gardnerella_vaginalis_409-05	1	72	100	88	19	17	17	17	19	20	19	19	18	19	19	19	19
4: Gardnerella_vaginalis_GV37	1	65	82	100	19	18	15	21	23	24	20	18	17	19	19	19	19
5: Gardnerella_vaginalis_UMB0833	1	16	19	19	100	80	72	41	46	46	46	45	43	45	45	45	45
6: Gardnerella_vaginalis_JCP7659	1	15	18	19	84	100	74	42	46	46	45	45	44	45	45	45	45
7: Gardnerella_vaginalis_JCP8070	1	16	18	18	78	77	100	43	46	46	46	46	45	46	46	46	46
8: Gardnerella_vaginalis_JCP8108	1	15	17	19	40	39	38	100	77	75	73	74	71	73	73	73	73
9: Gardnerella_vaginalis_UMB0736	1	16	18	21	43	41	40	74	100	89	87	87	83	86	86	86	86
10: Gardnerella_vaginalis_HMP9231	1	16	19	22	43	41	40	73	90	100	87	87	81	87	87	87	87
11: Gardnerella_vaginalis_FDAARGOS_568	1	16	18	20	44	40	40	72	88	88	100	87	81	87	87	87	87
12: Gardnerella_vaginalis_UMB0298	1	16	18	18	43	41	41	74	90	89	89	100	85	88	88	88	88
13: Gardnerella_vaginalis_JCP7672	1	16	18	18	43	42	42	74	89	87	86	89	100	87	87	87	87
14: Gardnerella_vaginalis_DSM4944	1	17	19	19	44	42	41	74	90	90	90	89	85	100	100	100	100
15: Gardnerella_vaginalis_NCTC10287	1	17	19	19	44	42	42	74	90	90	90	89	85	100	100	100	100
16: Gardnerella_vaginalis_ATCC14019	1	17	19	20	44	42	41	74	90	90	90	89	84	100	100	100	100
17: Gardnerella_vaginalis_ATCC14018_JCM11026	1	17	19	20	44	42	42	74	90	90	90	89	85	100	100	100	100

**Figure 1.** Heatmap with 7 complete genomes of *G. vaginalis* and 9 representative genomes of clades, according to the tree available at the NCBI. In green, a high similarity can be observed ranging from 65% to 100%; in orange, a median similarity ranging from 38% to 46%; and finally, a low similarity in red, ranging from 1% to 22%.

The data generated with the PGAdB-builder software were plotted in the MegaX software to visualize a phylogenetic tree (Figure 2), where the strains form the same clusters found in the heatmap of Gegenees.



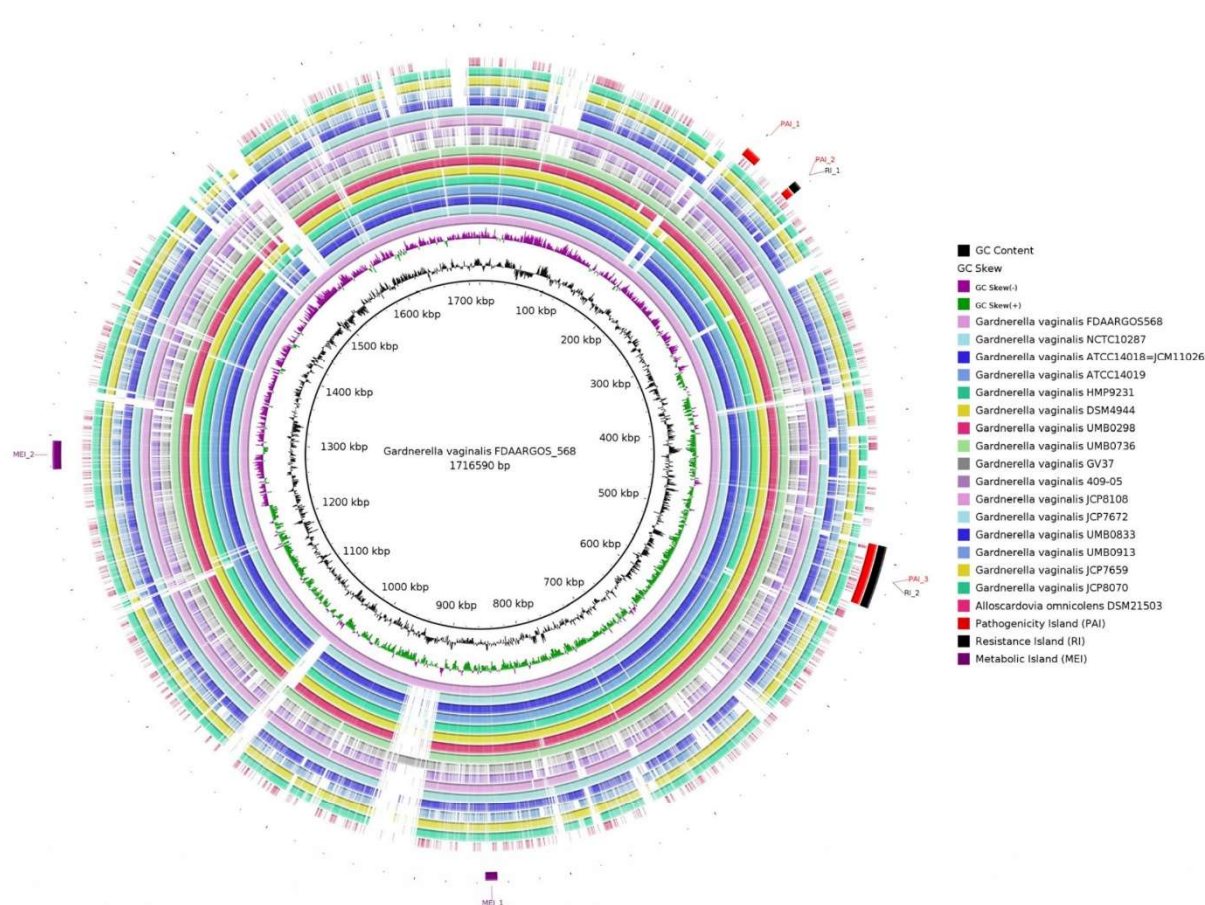
**Figure 2.** The phylogenetic tree shows the bootstrap percentage with 1000 bootstraps on the tree branches, ranging from 28 as the lowest percentage to 100 as the highest one. The same clusters found on the heatmap may be visualized here.

### 3.2. Genomic Plasticity

For the prediction of genomic islands, the genome plasticity analysis with software GIPSY was performed, where we can observe through the BRIG's results the pathogenicity and resistance islands plotted in the outermost circle of the image, with comparative genomic plasticity results. For the comparative genomic plasticity analysis, it was used a reference genome according to the NCBI, which is the *G. vaginalis* strain FDAARGOS 568. In the circle image, each ring represents one strain, and the presence of blank regions in each ring demonstrates a deletion of this genome part concerning the reference genome in the most central ring. It can be seen that there are two large deletion regions in almost all the genomes and the other deletions appear only in preliminary genomes and are present in almost all the complete genomes. The last ring related to the genome of *Alloscardovia omnicolens* DSM21503 is a ring with many deletions compared to the species studied.

Figure 3 also shows the presence of three pathogenicity islands, two resistance islands, and two metabolic islands in the last three outermost rings. Resistance island 1 and pathogenicity island 2 overlap approximately at a position of 200 kpb, and resistance island 2 and pathogenicity island 3 overlap approximately at a position of 500~600 kpb. These regions can present genes related to virulence increase and resistance against antibiotics, and some draft genomes can present some deletions at the beginning of these islands. The other islands also present some deletions when compared with some genomes.





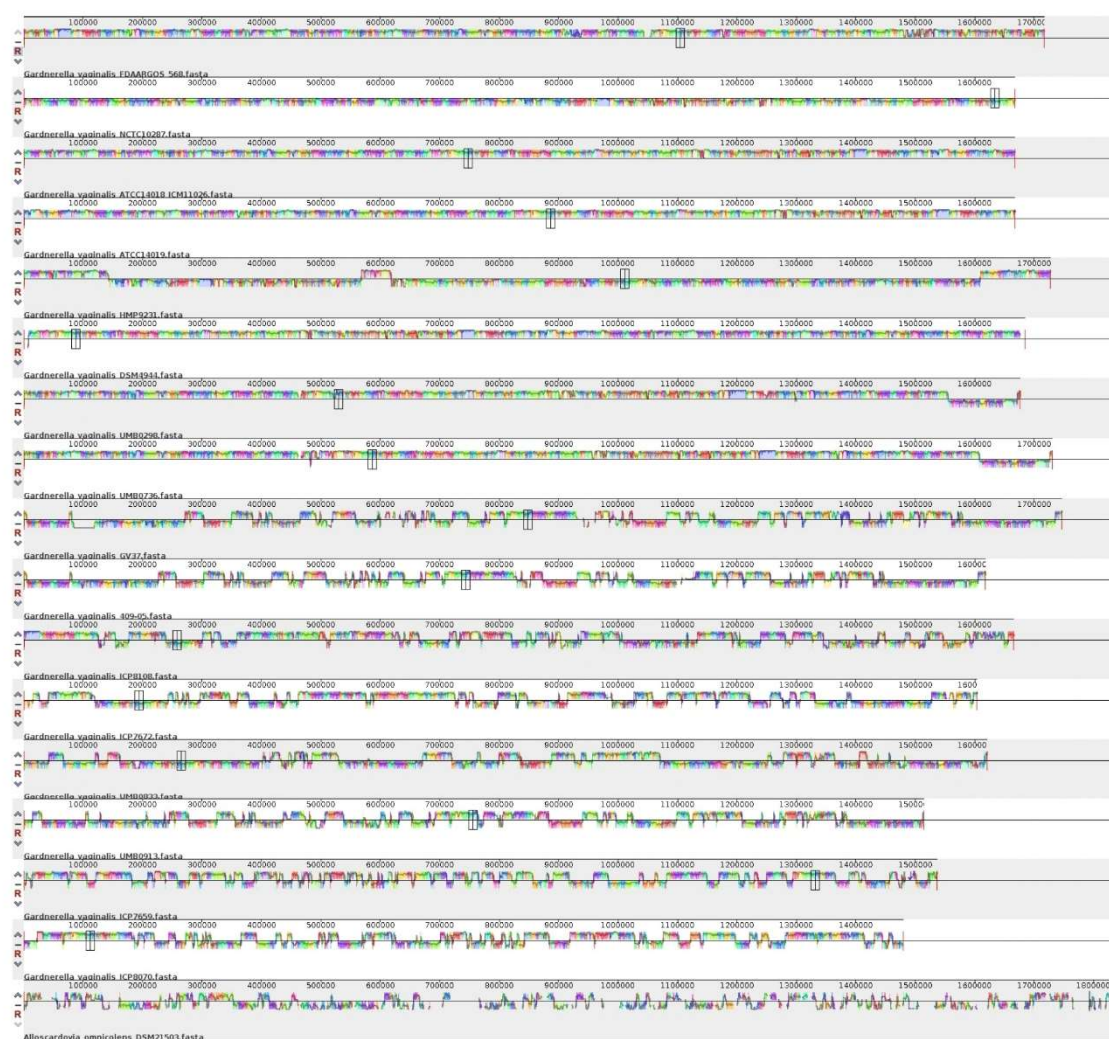
**Figure 3.** Genomic plasticity prediction with pathogenicity islands (PAI), resistance islands (RI), and metabolic islands (MI) of *G. vaginalis*. Each ring represents one strain of the working group, ranging from complete genomes in inner rings to draft genomes in the outermost rings. At the outermost, three rings represent the genomic islands—in red, the pathogenicity islands; in black, the resistance islands; and in purple, the metabolic islands.

Supplementary Figures S1–S3 show a comparative analysis between the clusters according to the similarity predicted in the previous results. When the comparative analysis was conducted, the genomes classified as scaffolds showed a greater, more significant deletion. Furthermore, Supplementary Figure S3 shows a more homogeneous relationship of these genomes forming clusters, while the other figures of clusters 1 and 2 (Supplementary Figures S1 and S2) show the presence of more broken regions.

### 3.3. Gene Synteny

The software Mauve was used to visualize the synteny between the various genomic sequences, where through the homology between them, it was possible to visualize possible rearrangements. Synteny is observed through gene blocks for better identification. Several broken regions may be observed, with many inversions and deletions of many regions, in the comparison between all genomes (Figure 4). A remarkable homogeneity was observed in cluster 1 (Supplementary Figure S4), with highly conserved blocks when compared to clusters 2 and 3 (Supplementary Figure S5 and S6), which showed inversions and deletions.



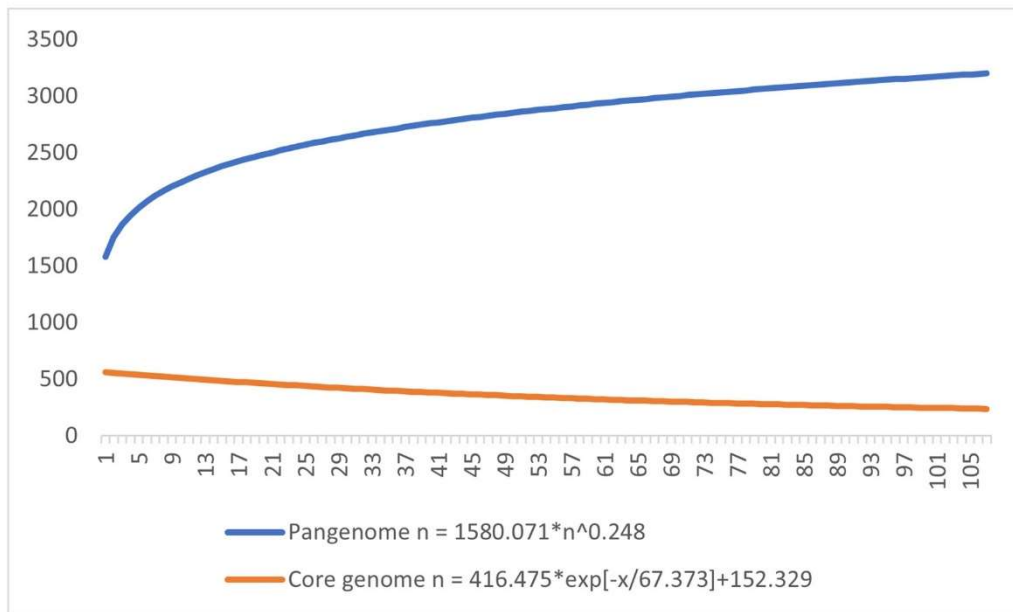


**Figure 4.** Gene synteny analysis. All genomes were broken and presented in continuous blocks. Each color represents one region that has passed through evolution for inversions, translocations, or deletions.

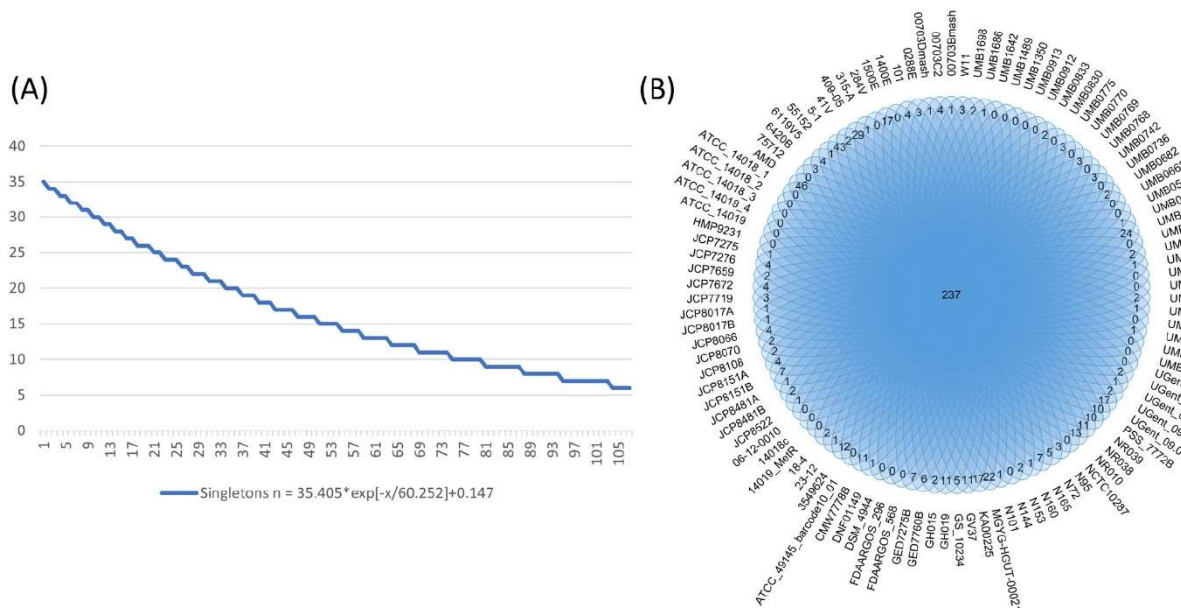
### 3.4. Pangenome Development

OrthoFinder separated these genomes into three different groups, core genes, shared genes, and singletons to predict orthologous genes, where 3563 non-redundant genes were found and classified as pangenome genes, with 208 genes being in the core genome group and 237 genes in the singletons group.

For the analyses with pangenome, the alpha value from Heap's Law when all genomes are analyzed was less than 1, being approximately 0.75 ( $\alpha = 1 - 0.248$ ), inferring an open genome. However, the core genome is close to closing, with 208 genes and stabilizing with 152 genes (Figure 5). Figure 6 shows the number of singletons in each of the genomes used, where the number changes, and some have more singletons than others.



**Figure 5.** The pangenome development. In blue, the pangenome growth curve with Heap’s law correction stabilizes approximately in just over 3000 genes, characterizing an open pangenome. The core genome stabilization curve is in orange, finalizing in approximately 208 genes.

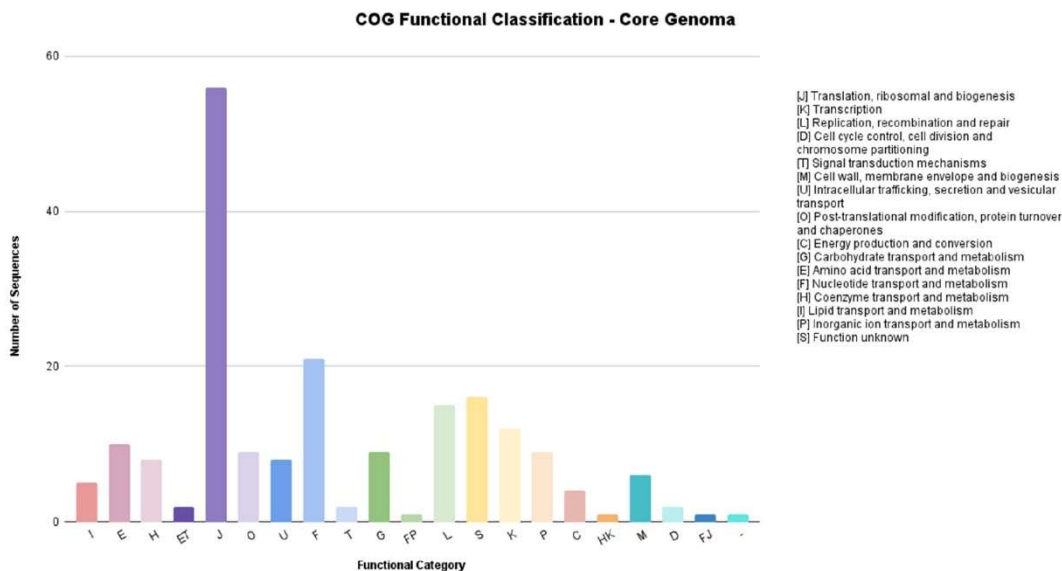


**Figure 6.** A singletons diagram. On side (A), a diagram represents the “singletons” curve generated using the least-squares fits law, with the tangent theta equal to 0.147, i.e., for each newly sequence genome, approximately one gene is added to the pangenome. On side (B), a Venn diagram showing singletons present in each strain. The number of singletons ranged from 0 to 46 genes in the strains.

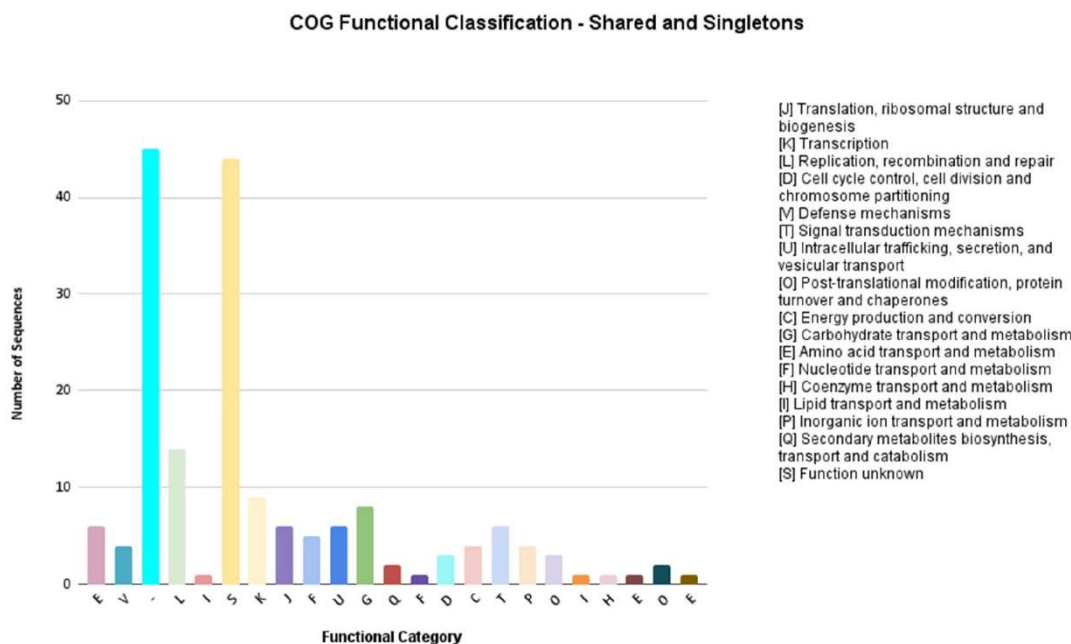
3.5. Orthologous Genes Characterization

The orthologous gene analyses are plotted in Figures 7 and 8, showing the COG classifications of the core genome, shared genome, and singletons, respectively. It is observed that the categories between these two analyses differ regarding the classification,

prevailing in the core, the COG J, which is related to translation, ribosomal, and biogenesis functions, differently from what is seen from the shared and singleton types that prevail but are not yet categorized and have unknown functions.



**Figure 7.** COG functional classification of core genome. Bar graph representing the COG classifications of the core genome. Each bar is related to one classification, COG J, with more genes classified, followed by COG F classification. Note: the column that displays “-” is related to the empty cells in the program’s result, that is, genes that were unable to be classified by the program.



**Figure 8.** COG functional classification of shared and singleton’s genes. Bar graph representing the COG classifications of shared plus singletons group. Each bar is related to one classification, COG S, with more genes classified, followed by COG L classification. Note: the column that displays “-” is related to the empty cells in the program’s result, that is, genes that were unable to be classified by the program.



The second category that is more prevalent in core classifications is COG F, more related to nucleus transport and metabolism, and for shared and singletons classifications, COG L is related to the replication, recombination, and repair of genomic content. Furthermore, there was no prevalence of a common category in these analyses.

#### 4. Discussion

*G. vaginalis* is a species of great medical importance, being one of the primary pathogens responsible for bacterial vaginosis, with high importance in women's health, and there are several reports and a certain problematization in the notification of these diseases as well as in the understanding of the consequences that it brings and how the different strains of this same species can behave differently in the host. Little is known about studies using comparative genomics to analyze the gene diversity of *G. vaginalis*. First, sixteen (16) genomes considered representative of the clusters grouped in the phylogenetic tree in the NCBI databases were used to analyze their distribution at the genomic level, their similarities, and differences throughout the evolution.

In all comparative analyses performed in this work, the strains present the same pattern. There is a formation of three clusters, with the first grouping all of the complete genomes; the second grouping the draft genome of the strains UMB0833, JCP7659, and JCP8070; and finally, the third cluster grouping the strains GV37, 409-05 (these being complete genomes), and UMB0913, considered as draft genome. Cluster 1 presents similarity above 80%, being always grouped in the same way and with a few deletions. Although close, it shows much translocation and a few inversion processes over time, which can be related to adaptation processes. Cluster 2 presents a median similarity, around 45%, with more deleted regions and much translocation and inversion process. Moreover, cluster 3 was more unexpected because it grouped different degrees of complementary types, being less similar with more events of evolution when compared with the other clusters, but inside this cluster, the genomes are so close and present almost the same patterns.

Firstly, in an attempt to explain how these clusters were formed, we researched more information about the epidemiology of these genomes, especially their extraction site. Unfortunately, only five samples had their extraction sites identified in the NCBI, as shown in Table 1 of the materials and methods. Once these sites had been identified, we realized that the clustering of these species was not related to their extraction site, since species as UMB0736 and UMB0298, both extracted in Maywood, USA were allocated in cluster 1; UMB0833, also extracted in Maywood, USA, was allocated in cluster 2; and, finally, GV37, which were extracted in Paris, France and UMB0913, which were extracted in Maywood, USA, were grouped in cluster 3. Then, although most of the species that we are able to identify were extracted from the same place, they do not show high clonality patterns, which corroborates the high plasticity found in the genomic plasticity analyses.

However, this work corroborates another work in diversity analysis inside the *Gardnerella* genus, where they found a polyphyletic organization, sorting this genus into nine distinct genotypes. In agreement with this work, our cluster 3, which grouped GV37, 409-05, and UMB0913 strains, can be considered as part of GGtype4, a group that presented genomes from strains that were considered less virulent organisms, and the other strains belonging to cluster 1 can be considered as part of GGtype9, a group which presents genomes from strains that were considered as more virulent organisms [59]. In addition, this work also suggests that the UMB0833 strain, part of cluster 2, can be considered as GGtype7, which presents characteristics like cluster 1 and may be classified as more virulent organisms. A study by Bohr L. et al. [60] used a different methodology, from DNA extraction to the characterization of the genes present in the pangenome and also corroborates the findings in our study and in the aforementioned study, in which they separated the clusters in the same way. Furthermore, this study demonstrates the significance of lateral gene transfer in the diversity process under discussion. Our findings, as depicted in Figure 3, have successfully identified certain genes associated with pathogenicity and resistance through genomic islands originating from horizontal gene transfer. Our research paves the

way for more in-depth investigations into the genes related to these mechanisms in an effort to ascertain whether HGT may indeed play a pivotal role in influencing diversity. Moreover, it guides future endeavors aimed at exploring these genes and their encoded proteins for potential utilization as diagnostic markers, as well as therapeutic and preventive measures against bacterial vaginosis.

This genomic difference between the strains of *G. vaginalis* in different groups was also observed in several studies, such as in the amplified restriction analysis of ribosomal DNA (ARDRA) identifying two genotypes among 17 samples of *G. vaginalis*, where they considered the existence of a biotype 3 even though it was not identified due to the limiting number of isolates and, also, to the region where they were obtained [11]. Other comparative analysis studies showed the existence of four different biotypes of *G. vaginalis* in the analysis of neighboring clusters between genes and perceived the formation of unexpected clades between them, which corroborates our study [17]. This was later confirmed in PCR analyses of the *cpn 6* gene sequence, confirming the previous result where four different groups were formed within the same species [6].

A pangenome analysis was performed in this work, where the genes were separated into three distinct groups: the core genes present in all strains; shared genes among some but not all strains; and singletons, with genes specific to each strain. The latter two groups are related to the process of evolution and adaptation of each strain, thus helping in better understanding the behavior of this species during its evolution. In the results, the species showed a wide-open pangenome with an alpha value lower than 1 (0.75), a core genome with approximately 208 genes, and the number of singletons added by each sequenced genome to the pangenome was shown to be very low after stabilization, with an estimate of less than one gene added. The core stabilization value is very close (~152 genes), and the wide-open pangenome and the tangent of theta value of singletons indicate the possibility of this pangenome closing, as genes are being added; however, it may take some time for this to happen. This variation is also observed in the Venn diagram of singleton genes (Figure 6B), where some genomes have more genes, being more variable than others. For example, strain AMD with 46 genes, 5-1 with 43 genes, and 409-05 with 29 genes.

Altogether, we can better understand the result found in the functional analysis of COG that represents a functional analysis of the behavior of this bacterium. When analyzing the core genome, proteins classified as COG J were more abundant than the others. COG J harbors proteins associated with translation, ribosomal structure, and biogenesis, which are functions commonly related to the functioning and survival of the bacterium, thus being essential for proliferation, differentiation, and development [61], as well as also helping the pathogen to adapt to changes in external physicochemical parameters; furthermore, another highly abundant functional category in the core genome was COG F, which is related to nucleotide transport and metabolism that are essential to all genomes once it is related to energy production in the organism. From another perspective, when we compare the COG categories between the singleton and shared genes with those of the core genes, in its totality, there are many genes in categories that are not very well described yet, which can be possibly accounted to genes acquired over the years in the process of adaptation of each strain, which would explain this diversity as well. In addition, for this analyzed set, the second category found was COG L, related to replication, recombination, and repair, which have essential roles in the adaptation process and diversity [62,63].

## 5. Conclusions

In conclusion, it can be observed that some genomes exhibit greater similarity among themselves when considering genomes within the same clade, as opposed to comparisons with genomes from different clades. When examining the evolutionary history and, in particular, the arrangement of their genes, some degree of this similarity becomes evident. However, a more pronounced divergence is now apparent when analyzing the gene order across the entirety of the genomes over the course of evolution. This comprehensive understanding sheds light on the genome's diversity. Moreover, regions of horizontal



gene transfer were identified, which may serve as a pivotal factor contributing to this diversity. When compared with the pangenome, it becomes apparent that new genes can be added with the discovery of new genomes, further amplifying this aspect. Analysis of the functions associated with these genes was anticipated when examining the core genome, with genes fulfilling crucial metabolic roles for survival. Nonetheless, the other two subsets contain genes with unknown functions, which may potentially generate new genes and contribute to an expanded diversity.

These findings hold significant relevance for future studies focusing on the search for resistance and virulence genes as targets for diagnostics or therapeutics. However, they should be corroborated through *in vitro* and *in vivo* experimentation.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/venereology2040012/s1>, Figure S1: Cluster 1 analysis with BRIG software; Figure S2: Cluster 2 analysis with BRIG software; Figure S3: Cluster 1 analysis with BRIG software; Figure S4: Cluster 1 analysis with Mauve software; Figure S5: Cluster 2 analysis with Mauve software; Figure S6: Cluster 3 analysis with Mauve software; Table S1: Genomes informations.

**Author Contributions:** A.G.F. and E.G.S. performed the download, all data processing and analysis, participated in the study design, and wrote this manuscript; F.V.D. helped in the data processing and in the writing of this manuscript; V.A.d.C.A. coordinated the study; S.d.C.S. designed, coordinated the study, helped with analysis, and wrote this manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Genome sequences are available in the NCBI database.

**Acknowledgments:** We thank the agencies of the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) and the Minas Gerais Research Funding Foundation (FAPEMIG). We give our thanks to the post-graduate program in Genetics at the Federal University of Minas Gerais and the post-graduate program in Medicina Tropical e Infectologia at the Federal University of Triângulo Mineiro.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Greenwood, J.R.; Pickett, M.J. Transfer of *Haemophilus vaginalis* Gardner and Dukes to a new genus, *Gardnerella*: *G. vaginalis* (Gardner and Dukes) comb. nov. *Int. J. Syst. Bacteriol.* **1980**, *30*, 170–178. [[CrossRef](#)]
- Leopold, S. Heretofore undescribed organism isolated from the genitourinary system. *U. S. Armed Forces Med. J.* **1953**, *4*, 263–266. [[PubMed](#)]
- Gardner, H.L.; Dukes, C.D. *Haemophilus vaginalis* vaginitis: A newly defined specific infection previously classified “nonspecific” vaginitis. *Am. J. Obstet. Gynecol.* **1955**, *69*, 962–976. [[CrossRef](#)] [[PubMed](#)]
- Piot, P.; van Dyck, E.; Peeters, M.; Hale, J.; Totten, P.A.; Holmes, K.K. Biotypes of *Gardnerella vaginalis*. *J. Clin. Microbiol.* **1984**, *20*, 677–679. [[CrossRef](#)] [[PubMed](#)]
- Hernández, F. *Gardnerella vaginalis* y *mobiluncus* en la etiología de la vaginosis bacteriana. *Rev. Costarric Cienc Méd.* **1998**, *19*, 107–116.
- Vanechoutte, M.; Guschin, A.; Van Simaey, L.; Gansemans, Y.; Van Nieuwerburgh, F.; Cools, P. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 679–687. [[CrossRef](#)]
- Patterson, J.L.; Stull-Lane, A.; Girerd, P.H.; Jefferson, K.K. Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of *Gardnerella vaginalis* relative to other bacterial-vaginosis-associated anaerobes. *Microbiology* **2010**, *156 Pt 2*, 392–399. [[CrossRef](#)]
- Lopes Dos Santos Santiago, G.; Deschaght, P.; El Aila, N.; Kiama, T.N.; Verstraelen, H.; Jefferson, K.K.; Temmerman, M.; Vanechoutte, M. *Gardnerella vaginalis* comprises three distinct genotypes of which only two produce sialidase. *Am. J. Obstet. Gynecol.* **2011**, *204*, 450.e1–450.e7. [[CrossRef](#)]
- Briselden, A.M.; Moncla, B.J.; Stevens, C.E.; Hillier, S.L. Sialidases (neuraminidases) in bacterial vaginosis and bacterial vaginosis-associated microflora. *J. Clin. Microbiol.* **1992**, *30*, 663–666. [[CrossRef](#)]

10. Cauci, S.; Thorsen, P.; Schendel, D.E.; Bremmelgaard, A.; Quadrifoglio, F.; Guaschino, S. Determination of Immunoglobulin A against *Gardnerella vaginalis* Hemolysin, Sialidase, and Prolidase Activities in Vaginal Fluid: Implications for Adverse Pregnancy Outcomes. *J. Clin. Microbiol.* **2003**, *41*, 435. [[CrossRef](#)]
11. Pleckaityte, M.; Janulaitiene, M.; Lasickiene, R.; Zvirbliene, A. Genetic and biochemical diversity of *Gardnerella vaginalis* strains isolated from women with bacterial vaginosis. *FEMS Immunol. Med. Microbiol.* **2012**, *65*, 69–77. [[CrossRef](#)] [[PubMed](#)]
12. Gelber, S.E.; Aguilar, J.L.; Lewis, K.L.T.; Ratner, A.J. Functional and phylogenetic characterization of vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis*. *J. Bacteriol.* **2008**, *190*, 3896–3903. [[CrossRef](#)] [[PubMed](#)]
13. Cauci, S.; Monte, R.; Ropele, M.; Missero, C.; Not, T.; Quadrifoglio, F.; Menestrina, G. Pore-forming and haemolytic properties of the *Gardnerella vaginalis* cytolysin. *Mol. Microbiol.* **1993**, *9*, 1143–1155. [[CrossRef](#)] [[PubMed](#)]
14. Bautista, C.T.; Wurapa, E.; Sateren, W.B.; Morris, S.; Hollingsworth, B.; Sanchez, J.L. Bacterial vaginosis: A synthesis of the literature on etiology, prevalence, risk factors, and relationship with chlamydia and gonorrhoea infections. *Mil Med Res.* **2016**, *13*, 4. [[CrossRef](#)]
15. Ingianni, A.; Petruzzelli, S.; Morandotti, G.; Pompei, R. Genotypic differentiation of *Gardnerella vaginalis* by amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Immunol. Med. Microbiol.* **1997**, *18*, 61–66. [[CrossRef](#)]
16. Schellenberg, J.J.; Jayaprakash, T.P.; Gamage, N.W.; Patterson, M.H.; Vanechoutte, M.; Hill, J.E. *Gardnerella vaginalis* Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. *PLoS ONE* **2016**, *11*, e0146510. [[CrossRef](#)]
17. Ahmed, A.; Earl, J.; Retchless, A.; Hillier, S.L.; Rabe, L.K.; Cherpes, T.L.; Powell, E.; Janto, B.; Eutsey, R.; Hiller, N.L.; et al. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. *J. Bacteriol.* **2012**, *194*, 3922–3937. [[CrossRef](#)]
18. Leite, S.R.R.D.F.; Amorim, M.M.R.D.; Calábria, W.B.; Leite, T.N.D.F.; Oliveira, V.S.D.; Ferreira Júnior, J.A.A.; Ximenes, R.A.D. A Perfil clínico e microbiológico de mulheres com vaginose bacteriana. *Rev. Bras. De Ginecol. E Obs.* **2010**, *32*, 82–87. [[CrossRef](#)]
19. Wilson, J. Managing recurrent bacterial vaginosis. *Sex. Transm. Infect.* **2004**, *80*, 8–11. [[CrossRef](#)]
20. Schwabke, J.R. Gynecologic consequences of bacterial vaginosis. *Obstet. Gynecol. Clin. N. Am.* **2003**, *30*, 685–694. [[CrossRef](#)]
21. Fethers, K.; Morris, M. Is bacterial vaginosis a sexually transmitted infection. *Sex. Transm. Infect.* **2001**, *77*, 390. [[CrossRef](#)]
22. Tanaka, V.; Fagundes, L.J.; Catapan, A.; Gotlieb, S.L.D.; Belda, W., Jr.; Arnone, M.; Soreano, R.; Moraes, F.R. Epidemiological profile of women with bacterial vaginosis treated at a clinic for sexually transmitted diseases in the city of Sao Paulo, SP. *An. Bras. Dermatol.* **2007**, *82*, 41–46. [[CrossRef](#)]
23. Formana, D.; de Martel, C.; Lacey, C.J.; Soerjomatarama, I.; Lortet-Tieulent, J.; Bruni, L.; Vignat, J.; Ferlay, J.; Bray, F.; Plummer, M.; et al. Global Burden of Human Papillomavirus and Related Diseases. *Vaccine* **2012**, *30* (Suppl. S5), F12–F23. [[CrossRef](#)]
24. Workowski, K.A.; Bachmann, L.H. Centers for Disease Control and Prevention's Sexually Transmitted Diseases Infection Guidelines. *Clin. Infect. Dis.* **2022**, *74* (Suppl. S2), S89–S94. [[CrossRef](#)] [[PubMed](#)]
25. Lukic, A.; Napoli, A.; Santino, I.; Bianchi, P.; Nobili, F.; Ciampittiello, G.; Nardone, M.R.; Santomauro, M.; Di Properzio, M.; Caserta, D. Cervicovaginal bacteria and fungi in pregnant diabetic and non-diabetic women: A multicenter observational cohort study. *Eur. Rev. Med. Pharmacol. Sci.* **2017**, *21*, 2303–2315. [[PubMed](#)]
26. Bradshaw, C.S.; Morton, A.N.; Hocking, J.; Garland, S.M.; Morris, M.B.; Moss, L.M.; Horvath, L.B.; Kuzevska, I.; Fairley, C.K. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J. Infect. Dis.* **2006**, *193*, 1478–1486. [[CrossRef](#)] [[PubMed](#)]
27. Marion, L.L.; Meeks, G.R. Ectopic pregnancy: History, incidence, epidemiology, and risk factors. *Clin. Obstet. Gynecol.* **2012**, *55*, 376–386. [[CrossRef](#)]
28. Holzman, C.; Leventhal, J.M.; Qiu, H.; Jones, N.M.; Wang, J. Factors Linked to Bacterial Vaginosis in Nonpregnant Women. *Am. J. Public Health* **2011**, *91*, 1664–1670. [[CrossRef](#)]
29. Marrazzo, J.M.; Wiesenfeld, H.C.; Murray, P.J.; Busse, B.; Meyn, L.; Krohn, M.; Hillier, S.L. Risk Factors for Cervicitis among Women with Bacterial Vaginosis. *J. Infect. Dis.* **2006**, *193*, 617–624. [[CrossRef](#)]
30. Small, S.A.; Luster, T. Adolescent Sexual Activity: An Ecological, Risk-Factor Approach. *J. Marriage Fam.* **1994**, *56*, 181. [[CrossRef](#)]
31. Anand, K.V.; Pimple, S.A.; Mishra, G.A.; Sahare, R.V.; Pathuthara, S.; Deodhar, K.K.; Shastri, S.S. Reliability of conventional Papanicolaou smear in diagnosing bacterial vaginosis among women with clinical genital infection. *South Asian J. Cancer* **2020**, *9*, 13–16. [[CrossRef](#)] [[PubMed](#)]
32. Li, X.D.; Wang, C.C.; Zhang, X.J.; Gao, G.P.; Tong, F.; Li, X.; Hou, S.; Sun, L.; Sun, Y.-H. Risk factors for bacterial vaginosis: Results from a cross-sectional study having a sample of 53,652 women. *Eur. J. Clin. Microbiol. Infect. Dis.* **2014**, *33*, 1525–1532. [[CrossRef](#)] [[PubMed](#)]
33. Denney, J.M.; Culhane, J.F. Bacterial vaginosis: A problematic infection from both a perinatal and neonatal perspective. *Semin. Fetal Neonatal Med.* **2009**, *14*, 200–203. [[CrossRef](#)] [[PubMed](#)]
34. Adinkra, P.; Lamont, R.F. Adverse obstetric sequelae of bacterial vaginosis. *Hosp. Med.* **2000**, *61*, 475–477. [[CrossRef](#)]
35. Molteni, B.; D'Antuono, A.; Bandini, P.; Sintini, G.; Barcellona, E.; Agnello, A.; Milani, M. Efficacy and tolerability of a new chlorhexidine-based vaginal gel in vaginal infections. *Curr. Med. Res. Opin.* **2004**, *20*, 849–853. [[CrossRef](#)]
36. Miko, E.; Barakonyi, A. The Role of Hydrogen-Peroxide (H<sub>2</sub>O<sub>2</sub>) Produced by Vaginal Microbiota in Female Reproductive Health. *Antioxidants* **2023**, *12*, 1055. [[CrossRef](#)]



37. Verstraelen, H.; Verhelst, R.; Roelens, K.; Temmerman, M. Antiseptics and disinfectants for the treatment of bacterial vaginosis: A systematic review. *BMC Infect. Dis.* **2012**, *12*, 148. [[CrossRef](#)]
38. Turovskiy, Y.; Sutyak Noll, K.; Chikindas, M.L. The aetiology of bacterial vaginosis. *J. Appl. Microbiol.* **2011**, *110*, 1105–1128. [[CrossRef](#)]
39. Petrova, M.I.; Lievens, E.; Malik, S.; Imholz, N.; Lebeer, S. *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. *Front. Physiol.* **2015**, *6*, 81. [[CrossRef](#)]
40. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
41. Maddison, W.P.; Donoghue, M.J.; Maddison, D.R. Outgroup Analysis and Parsimony. *Syst. Biol.* **1984**, *33*, 83–103. [[CrossRef](#)]
42. Ågren, J.; Sundström, A.; Håfström, T.; Segerman, B. Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLoS ONE* **2012**, *7*, e39107. [[CrossRef](#)] [[PubMed](#)]
43. Liu, Y.Y.; Chiou, C.S.; Chen, C.C. PGADB-builder: A web service tool for creating pan-genome allele database for molecular fine typing. *Sci. Rep.* **2016**, *6*, 36213. [[CrossRef](#)]
44. Rodrigues, T.C.V.; Jaiswal, A.K.; De Sarom, A.; Oliveira, L.D.C.; Oliveira, C.J.F.; Ghosh, P.; Tiwari, S.; Miranda, F.M.; de Jesus Benevides, L.; de Carvalho Azevedo, V.A.; et al. Reverse vaccinology and subtractive genomics reveal new therapeutic targets against *Mycoplasma pneumoniae*: A causative agent of pneumonia. *R. Soc. Open Sci.* **2019**, *6*, 1909070. [[CrossRef](#)]
45. Soares, S.C.; Geyik, H.; Ramos, R.T.J.; de Sá, P.H.C.G.; Barbosa, E.G.V.; Baumbach, J.; Figueiredo, H.C.P.; Miyoshi, A.; Tauch, A.; Silva, A.; et al. GIPSY: Genomic island prediction software. *J. Biotechnol.* **2016**, *232*, 2–11. [[CrossRef](#)] [[PubMed](#)]
46. Alikhan, N.F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [[CrossRef](#)]
47. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)]
48. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
49. Enright, A.J.; Van Dongen, S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584. [[CrossRef](#)]
50. Medini, D.; Donati, C.; Tettelin, H.; Massignani, V.; Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589–594. [[CrossRef](#)]
51. Boto, L. Horizontal gene transfer in evolution: Facts and challenges. *Proc. R. Soc. B Biol. Sci.* **2010**, *277*, 819. [[CrossRef](#)] [[PubMed](#)]
52. Soares, S.C.; Silva, A.; Trost, E.; Blom, J.; Ramos, R.; Carneiro, A.; Ali, A.; Santos, A.R.; Pinto, A.C.; Diniz, C.; et al. The Pan-Genome of the Animal Pathogen *Corynebacterium pseudotuberculosis* Reveals Differences in Genome Plasticity between the Biovar *ovis* and *equi* Strains. *PLoS ONE* **2013**, *8*, e53818. [[CrossRef](#)] [[PubMed](#)]
53. Carlos Guimaraes, L.; Benevides de Jesus, L.; Vinicius Canario Viana, M.; Silva, A.; Thiago Juca Ramos, R.; de Castro Soares, S.; Azevedo, V. Inside the Pan-genome—Methods and Software Overview. *Curr. Genom.* **2015**, *16*, 245. [[CrossRef](#)]
54. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)]
55. Gene Ontology Consortium. Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [[CrossRef](#)] [[PubMed](#)]
56. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)] [[PubMed](#)]
57. Galperin, M.Y.; Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **2015**, *43*, D261–D269. [[CrossRef](#)]
58. Tatusov, R.L.; Koonin, E.V.; Lipman, D.J. A genomic perspective on protein families. *Science* **1997**, *278*, 631–637. [[CrossRef](#)]
59. Tarracchini, C.; Lugli, G.A.; Mancabelli, L.; Milani, C.; Turrone, F.; Ventura, M. Assessing the Genomic Variability of *Gardnerella vaginalis* through Comparative Genomic Analyses: Evolutionary and Ecological Implications. *Appl. Environ. Microbiol.* **2020**, *87*, 1–16. [[CrossRef](#)]
60. Bohr, L.L.; Mortimer, T.D.; Pepperell, C.S. Lateral Gene Transfer Shapes Diversity of *Gardnerella* spp. *Front. Cell Infect. Microbiol.* **2020**, *10*, 293. [[CrossRef](#)]
61. Jiao, L.; Liu, Y.; Yu, X.Y.; Pan, X.; Zhang, Y.; Tu, J.; Song, Y.-H. Ribosome biogenesis in disease: New players and therapeutic targets. *Signal Transduct. Target. Ther.* **2023**, *8*, 15. [[CrossRef](#)] [[PubMed](#)]
62. Epum, E.A.; Haber, J.E. DNA replication: The recombination connection. *Trends Cell Biol.* **2022**, *32*, 45–57. [[CrossRef](#)] [[PubMed](#)]
63. Xu, Y.; Xu, D. Repair pathway choice for double-strand breaks. *Essays Biochem.* **2020**, *64*, 765–777. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.



## CONCLUSÃO CAPÍTULO I

Os resultados deste estudo destacam os padrões de similaridade e divergência genômica da história evolutiva de *Gardnerella vaginalis*. A análise dos arranjos genômicos revelou uma maior coesão entre os genomas dentro do mesmo clado, enquanto as comparações interclados evidenciaram uma divergência mais acentuada, especialmente na ordem dos genes ao longo do tempo evolutivo. O que pode ser explicado pelas regiões de transferência horizontal de genes que destaca um mecanismo fundamental na promoção da diversidade genômica. Essas transferências podem introduzir novos genes nos genomas, ampliando as capacidades adaptativas das espécies. A constatação de que novos genes continuam sendo descobertos com a análise de novos genomas enfatiza a importância da abordagem do pangenoma para uma compreensão completa da diversidade genética.

A análise funcional dos genes identificados revelou uma gama de funções, desde aquelas essenciais para processos metabólicos básicos até aquelas com funções ainda desconhecidas. Essa diversidade funcional sugere um vasto potencial para a descoberta de novos genes e suas implicações nas adaptações evolutivas. Esses achados são de grande relevância para estudos futuros, especialmente aqueles voltados para a identificação de genes de resistência e virulência como alvos para diagnóstico e terapêutica e além disso para compreender melhor o comportamento da espécie ao longo do tempo, a fim de elucidar melhor quais as características evolutivas foram mantidas dentro de *Gardnerella vaginalis*.

## **CAPÍTULO II**

## **5. ARTIGO CIENTÍFICO 2: Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for *Gardnerella vaginalis***

**Manuscrito submetido de Análise de seleção positiva de genes de resistência e virulência de *Gardnerella vaginalis* em língua inglesa formato de acordo com as normas do periódico Genes & Genomics (Springer), Fator de Impacto (atual): 2.1.**

Uma questão recorrente quando se estuda genômica comparativa é tentar encontrar a base genética que está relacionada com o processo de diversidade entre um grupo taxonômico. Esse artigo visa entender quais seriam as características fenotípicas específica da espécie e quais genes possibilitam visualizar essa diversidade dentro de *Gardnerella vaginalis*, como poderia também ser explicado o comportamento de resistência e virulência dentre essas linhagens dessa bactéria patogênica. A seleção positiva Darwiniana pode ajudar a compreender melhor esse comportamento que foi mantido durante a evolução da espécie na busca de descrever os fenômenos de que variantes genéticas que sejam benéficas para fixar em uma população e que ao longo do tempo aumenta sua aptidão.

*Gardnerella vaginalis* exibe um potencial patogênico maior em comparação com outros organismos que causam esta condição, formando biofilmes e aderindo mais intensamente às células epiteliais vaginais. A resistência a antibióticos também foi observada em cepas de *G. vaginalis*, atribuída à presença de vários genes de resistência. Estudos genômicos comparativos revelaram diversidade genética entre as espécies e diferentes genótipos, possivelmente relacionados à virulência e resistência. O estudo visa identificar fenótipos de resistência e virulência, bem como os genes relacionados à seleção positiva de Darwin, e identificar novos alvos terapêuticos por meio de docking molecular e análises de acoplamento com compostos naturais.

**Genes & Genomics**  
**Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection**  
**for Novel Drug Discovery for Gardnerella vaginalis**  
 --Manuscript Draft--

<b>Manuscript Number:</b>	GENG-D-24-00308
<b>Full Title:</b>	Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for Gardnerella vaginalis
<b>Article Type:</b>	Original Article
<b>Corresponding Author:</b>	Eduarda Guimarães Sousa, MD Universidade Federal de Minas Gerais Belo Horizonte, Minas Gerais BRAZIL
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	Universidade Federal de Minas Gerais
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Eduarda Guimarães Sousa, MD
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	Eduarda Guimarães Sousa, MD Andrei Giacchetto Felice Fabiana Vieira Dominici Arun Kumar Jaiswal Mariana Letícia Costa Pedrosa Luiza Pereira Reis Lucas Gabriel Rodrigues Gomes Vasco Ariston de Carvalho Azevedo Siomar de Castro Soares
<b>Order of Authors Secondary Information:</b>	
<b>Funding Information:</b>	
<b>Abstract:</b>	Gardnerella vaginalis is a Gram-variable bacillus, characterized as a non-motile and non-capsule producer, capable of causing bacterial vaginosis, a condition prevalent in reproductive age women, with this bacterium present in nearly 100% of cases. This organism exhibits high pathogenicity linked to virulence and resistance genes acquired throughout evolution, showcasing elevated resistance to a broad spectrum of drug classes. This study conducted comparative genomic analyses to identify these genes and correlate their presence with positive Darwinian selection. Additionally, new drug targets were selected through docking and molecular modeling, guided by the heightened antimicrobial resistance exhibited by this microbial species. The available genomes of G. vaginalis were analyzed, the orthologous genes were delineated, and the positively selected genes, where 29 groups were found. Of these genes, one of great importance was predicted, Mef(A), which is related to resistance to the macrolide group of antibiotics, which are the first line of choice for treating bacterial vaginosis. The selection covered genes ubiquitously present in all the strains studied; other resistance genes also predicted are those related to resistance to tetracycline, tet(M), and tet(L). In addition, phenotypic analyses of virulence were carried out, and one of the findings was the sigA (RpoV) gene, which is closely related to resistance to various classes of antibiotics and is also a promising drug target. Additionally, two potential protein candidates were selected as drug targets, implicated in essential processes for bacterial cell function. Subsequently, these two proteins underwent modeling, and their structures were predicted, followed by docking simulations with natural compounds.

# Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for *Gardnerella vaginalis*

Eduarda Guimarães Sousa<sup>1\*</sup>, Andrei Giacchetto Felice<sup>2\*</sup>, Fabiana Vieira Dominici<sup>2</sup>, Arun Kumar Jaiswal<sup>1</sup>, Mariana Leticia Costa Pedrosa<sup>1</sup>, Luiza Pereira Reis<sup>1</sup>, Lucas Gabriel Rodrigues Gomes<sup>1</sup>, Vasco Ariston de Carvalho Azevedo<sup>1</sup>, Siomar de Castro Soares<sup>2\*</sup>

<sup>1</sup> Cellular and Molecular Genetics Laboratory, Institute of Biological Sciences, Department of General Biology, Federal University of Minas Gerais, Pampulha Campus, Belo Horizonte 31270-901, MG, Brazil

<sup>2</sup> Laboratory of Immunology and Bioinformatics, Institute of Biological and Natural Sciences, Department of Microbiology, Immunology and Parasitology, Federal University of Triângulo Mineiro, Educational Center Campus, Uberaba 38025-180, MG, Brazil

\* Correspondence: [eduardaguimaraessousa@gmail.com](mailto:eduardaguimaraessousa@gmail.com); [siomars@gmail.com](mailto:siomars@gmail.com)

**Abstract:** *Gardnerella vaginalis* is a Gram-variable bacillus, characterized as a non-motile and non-capsule producer, capable of causing bacterial vaginosis, a condition prevalent in reproductive age women, with this bacterium present in nearly 100% of cases. This organism exhibits high pathogenicity linked to virulence and resistance genes acquired throughout evolution, showcasing elevated resistance to a broad spectrum of drug classes. This study conducted comparative genomic analyses to identify these genes and correlate their presence with positive Darwinian selection. Additionally, new drug targets were selected through docking and molecular modeling, guided by the heightened antimicrobial resistance exhibited by this microbial species. The available genomes of *G. vaginalis* were analyzed, the orthologous genes were delineated, and the positively selected genes, where 29 groups were found. Of these genes, one of great importance was predicted, *Mef(A)*, which is related to resistance to the macrolide group of antibiotics, which are the first line of choice for treating bacterial vaginosis. The selection covered genes ubiquitously present in all the strains studied; other resistance genes also predicted are those related to resistance to tetracycline, *tet(M)*, and *tet(L)*. In addition, phenotypic analyses of virulence were carried out, and one of the findings was the *sigA (RpoV)* gene, which is closely related to resistance to various classes of antibiotics and is also a promising drug target. Additionally, two potential protein candidates were selected as drug targets, implicated in essential processes for bacterial cell function. Subsequently, these two proteins underwent modeling, and their structures were predicted, followed by docking simulations with natural compounds. These proteins were linked with a natural compound each and are considered good potential drug targets. The analyses contribute to analyzing the evolution of the species and how resistance genes are related to their permanence as a potential pathogen.

**Keywords:** *Gardnerella vaginalis*; Resistance; Positive-selection; Drug Targets.

## 1. Introduction

*Gardnerella vaginalis* is a bacillus-shaped bacterium, commonly described as Gram-variable due to its unpredictable response to Gram staining, influenced by the age of cultures and the physiological state of the bacterium. In a specific analysis of its cell wall, a skinny layer of peptidoglycan was observed as it ages, along with the absence of lipopolysaccharide and its components, such as heptose and hydroxylated fatty acids. This characteristic has led many authors to characterize it as Gram-positive. This organism lacks flagella or any other means of locomotion, does not produce spores, and lacks a capsule (Ping WONG et al.).

1 This pathogen can potentially cause bacterial vaginosis (BV), a common dysbiosis in the lower  
2 genital tract of women of reproductive age. This condition is characterized by a reduction in *Lactobacillus*  
3 species in the vaginal region, followed by an increase in facultative anaerobic bacteria. BV has been  
4 associated with various women's health issues, including an increased risk of pregnancy and childbirth  
5 complications, and contracting sexually transmitted diseases. Despite being considered a dysbiosis caused  
6 by different bacterial genera, *G. vaginalis* is the species that predominantly causes BV, detected in samples  
7 from affected women in up to 95% of cases (Ping WONG et al.).  
8

9 This bacillus exhibits a higher pathogenic potential compared to other organisms causing this  
10 condition, adhering more intensely to vaginal epithelial cells and forming biofilms ([CSL STYLE ERROR:  
11 reference with no printed form.]), a capability acquired through the presence of the sialidase enzyme.  
12 Additionally, it possesses the *VLY* gene, which encodes vaginolysin, a toxic compound that assists in lysing  
13 target cells (Ma et al. 2022). The biofilm formed by this bacterium features a matrix that impedes the drug  
14 diffusion to the core, phenotypic changes, bacterial communication through quorum-sensing, upregulation  
15 of genes enabling antimicrobial resistance, presence of enzymes hindering drug penetration into the biofilm,  
16 and adaptation in physiological structure (Rigo and Tasca 2020).  
17

18 Strains of *G. vaginalis* exhibiting intermediate resistance to kanamycin have been identified,  
19 attributed to the presence of a gene belonging to the aminoglycoside phosphotransferase (APH) family.  
20 Additionally, resistance mediated by the *tetM* and *tetL* genes, reducing sensitivity to tetracycline, was  
21 observed. Furthermore, genes associated with superfamilies of ABC transporters (ATP Binding Cassette)  
22 and PBP (penicillin-binding proteins) were found in these strains, strongly linked to drug resistance (Rigo  
23 and Tasca 2020; Zhang et al. 2022). A study utilized strains from China to assess antibiotic resistance,  
24 reporting the involvement of antibiotic resistance genes *ermX* and *lsaC* in macrolide resistance. Macrolides  
25 are commonly used to treat bacterial vaginosis (Schuyler et al. 2015). The study also identified a resistance  
26 phenotype for nitroimidazoles and aminoglycosides (Zhang et al. 2022).  
27

28 These resistance genes may be involved in processes of positive Darwinian selection (Bochner et  
29 al. 1980), in addition to those related to the host-pathogen relationship, immunity and virulence (Hongo  
30 et al. 2015) as reported in several important pathogenic taxa, *Escherichia coli* (Petersen et al. 2007),  
31 *Campylobacter* (Lefébure and Stanhope 2009) and *Salmonella* (Soyer et al. 2009). Generally, studies of  
32 this nature are involved in understanding evolution and phenotypic characteristics that have been positively  
33 selected.  
34

35 Comparative genomic studies of *G. vaginalis* have shown a certain gene diversity among the  
36 species (Giacchetto Felice et al. 2023), as well as including nine distinct genotypes (GGtype1 to GGtype9),  
37 and this may be related to the virulence and resistance potential of this microbial taxon (Tarracchini et al.  
38 2020) that can be associated with genome-scale positive selection detection (GSPSD).  
39

40 In this comparative genomics study, we aim to identify the resistance and virulence phenotypes  
41 present in the core and which of these genes are related to Darwinian positive selection. In addition,  
42 generating virulence and antibiotic resistance requires identifying new drug targets, so molecular modeling  
43 and docking analyses with natural compounds will be carried out concurrently.  
44

## 45 2. Materials and Methods

46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59



## 2.1 Genome Information

We used 97 *Gardnerella vaginalis* Refseq and annotated genomes available from the National Center for Biotechnology Information Datasets (NCBI) (<https://www.ncbi.nlm.nih.gov/datasets/> - accessed August 2023) for the comparative genomics analyses using “.fna”, “.faa” and “.gbk” formats to perform the studies (Table 1).

**Table 1:** Information about the 97 genomes of *G. vaginalis* strains.

Assembly Accession	Organism Name	Assembly BioSample Accession
GCF_002861965.1	<i>Gardnerella vaginalis</i> UMB0386	SAMN08193674
GCF_001049785.1	<i>Gardnerella vaginalis</i> 3549624	SAMN03801593
GCF_001278345.1	<i>Gardnerella vaginalis</i> 14019_MetR	SAMN04014465
GCF_001546455.1	<i>Gardnerella vaginalis</i> GED7760B	SAMN03851015
GCF_001546485.1	<i>Gardnerella vaginalis</i> PSS_7772B	SAMN03851016
GCF_001563665.1	<i>Gardnerella vaginalis</i> CMW7778B	SAMN03851013
GCF_001660735.1	<i>Gardnerella vaginalis</i> 23-12	SAMN04625558
GCF_001660755.1	<i>Gardnerella vaginalis</i> 18-4	SAMN04625602
GCF_001913835.1	<i>Gardnerella vaginalis</i> ATCC 49145	SAMN05757759
GCF_002206225.1	<i>Gardnerella vaginalis</i> FDAARGOS_296	SAMN06173309
GCF_002861165.1	<i>Gardnerella vaginalis</i> UMB0061	SAMN08193668
GCF_002861925.1	<i>Gardnerella vaginalis</i> UMB0775	SAMN08193678
GCF_002861945.1	<i>Gardnerella vaginalis</i> UMB0770	SAMN08193677
GCF_002861975.1	<i>Gardnerella vaginalis</i> UMB0298	SAMN08193676
GCF_002862005.1	<i>Gardnerella vaginalis</i> UMB0032B	SAMN08193675
GCF_002862015.1	<i>Gardnerella vaginalis</i> UMB0032A	SAMN08193673
GCF_002862045.1	<i>Gardnerella vaginalis</i> UMB0233	SAMN08193672
GCF_002884835.1	<i>Gardnerella vaginalis</i> UMB0768	SAMN07511408
GCF_002894105.1	<i>Gardnerella vaginalis</i> DNF01149	SAMN05578253
GCF_002896555.1	<i>Gardnerella vaginalis</i> KA00225	SAMN05578087
GCF_003034925.1	<i>Gardnerella vaginalis</i> ATCC 49145	SAMN08644262
GCF_003369875.1	<i>Gardnerella vaginalis</i> KA00225	SAMN03145604
GCF_003369895.1	<i>Gardnerella vaginalis</i> NI01	SAMN03145579
GCF_003369935.1	<i>Gardnerella vaginalis</i> NI53	SAMN03145603
GCF_003369965.1	<i>Gardnerella vaginalis</i> N95	SAMN03145504
GCF_003397605.1	<i>Gardnerella vaginalis</i> UGent 25.49	SAMN09373179
GCF_003397665.1	<i>Gardnerella vaginalis</i> UGent 09.07	SAMN09373175
GCF_003408745.1	<i>Gardnerella vaginalis</i> GH015	SAMN04446401
GCF_003408775.1	<i>Gardnerella vaginalis</i> NI60	SAMN04446403
GCF_003408785.1	<i>Gardnerella vaginalis</i> NI65	SAMN04446402
GCF_003408835.1	<i>Gardnerella vaginalis</i> NI44	SAMN04446400

	GCF_003408845.1	<i>Gardnerella vaginalis</i> NR010	SAMN04446404
1	GCF_003585655.1	<i>Gardnerella vaginalis</i> NR038	SAMN07490630
2	GCF_003585755.1	<i>Gardnerella vaginalis</i> NR039	SAMN07490631
3	GCF_003812765.1	<i>Gardnerella vaginalis</i> FDAARGOS_568	SAMN10163192
4	GCF_004336715.1	<i>Gardnerella vaginalis</i> 14018c	SAMN11037839
5	GCF_013315005.1	<i>Gardnerella vaginalis</i> UMB0143	SAMN15064064
6	GCF_013315025.1	<i>Gardnerella vaginalis</i> UMB0736	SAMN15064063
7	GCF_013315045.1	<i>Gardnerella vaginalis</i> UMB0540	SAMN15064062
8	GCF_013315075.1	<i>Gardnerella vaginalis</i> UMB0202	SAMN15064060
9	GCF_013315085.1	<i>Gardnerella vaginalis</i> UMB0358	SAMN15064061
10	GCF_013315115.1	<i>Gardnerella vaginalis</i> UMB0558	SAMN15064059
11	GCF_014857145.1	<i>Gardnerella vaginalis</i> 06-12-0010	SAMN16294983
12	GCF_023016185.1	<i>Gardnerella vaginalis</i> KC2	SAMN23424279
13	GCF_023016205.1	<i>Gardnerella vaginalis</i> KC1	SAMN23424278
14	GCF_023016225.1	<i>Gardnerella vaginalis</i> KC4	SAMN23424281
15	GCF_023016245.1	<i>Gardnerella vaginalis</i> KC3	SAMN23424280
16	GCF_023277565.1	<i>Gardnerella vaginalis</i> JNFY17	SAMN21246408
17	GCF_023277605.1	<i>Gardnerella vaginalis</i> JNFY14	SAMN21246406
18	GCF_023277625.1	<i>Gardnerella vaginalis</i> JNFY13	SAMN21246405
19	GCF_023277645.1	<i>Gardnerella vaginalis</i> JNFY11	SAMN21246404
20	GCF_023277665.1	<i>Gardnerella vaginalis</i> JNFY9	SAMN21246403
21	GCF_023277685.1	<i>Gardnerella vaginalis</i> JNFY4	SAMN21246402
22	GCF_023277725.1	<i>Gardnerella vaginalis</i> JNFY1	SAMN21246400
23	GCF_030213965.1	<i>Gardnerella vaginalis</i> UMB9230	SAMN34996711
24	GCF_030215405.1	<i>Gardnerella vaginalis</i> UMB6972	SAMN34996565
25	GCF_030216615.1	<i>Gardnerella vaginalis</i> UMB6789	SAMN34996560
26	GCF_030217865.1	<i>Gardnerella vaginalis</i> UMB1190A	SAMN34996494
27	GCF_030218185.1	<i>Gardnerella vaginalis</i> UMB1019	SAMN34996474
28	GCF_030228365.1	<i>Gardnerella vaginalis</i> UMB1218B	SAMN35153957
29	GCF_030228445.1	<i>Gardnerella vaginalis</i> UMB1190B	SAMN35153955
30	GCF_030233905.1	<i>Gardnerella vaginalis</i> UMB10121	SAMN35153918
31	GCF_900105405.1	<i>Gardnerella vaginalis</i> DSM 4944	SAMN04488545
32	GCF_900637625.1	<i>Gardnerella vaginalis</i> NCTC10287	SAMEA4535760
33	GCF_000263555.1	<i>Gardnerella vaginalis</i> 0288E	SAMN02393775
34	GCF_000263495.1	<i>Gardnerella vaginalis</i> 1400E	SAMN02393779
35	GCF_000263595.1	<i>Gardnerella vaginalis</i> 1500E	SAMN02393780
36	GCF_000263435.1	<i>Gardnerella vaginalis</i> 284V	SAMN02393773
37	GCF_000214315.1	<i>Gardnerella vaginalis</i> 315-A	SAMN00138210
38	GCF_000165635.1	<i>Gardnerella vaginalis</i> 41V	SAMN02472074
39	GCF_000263475.1	<i>Gardnerella vaginalis</i> 55152	SAMN02393778
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			



1	GCF_000263655.1	<i>Gardnerella vaginalis</i> 6119V5	SAMN02393784
2	GCF_000263535.1	<i>Gardnerella vaginalis</i> 75712	SAMN02393774
3	GCF_000178355.1	<i>Gardnerella vaginalis</i> ATCC 14018	SAMN02471014
4	GCF_001042655.1	<i>Gardnerella vaginalis</i> ATCC 14018	SAMD00061047
5	GCF_003397685.1	<i>Gardnerella vaginalis</i> ATCC 14018	SAMN09373172
6	GCF_004336685.1	<i>Gardnerella vaginalis</i> ATCC 14018	SAMN11037755
7	GCF_000159155.2	<i>Gardnerella vaginalis</i> ATCC 14019	SAMN00001462
8	GCF_000213955.1	<i>Gardnerella vaginalis</i> HMP9231	SAMN00100736
9	GCF_000414705.1	<i>Gardnerella vaginalis</i> JCP7275	SAMN02436832
10	GCF_000414685.1	<i>Gardnerella vaginalis</i> JCP7276	SAMN02436904
11	GCF_000414645.1	<i>Gardnerella vaginalis</i> JCP7672	SAMN02436831
12	GCF_000414525.1	<i>Gardnerella vaginalis</i> JCP8108	SAMN02436830
13	GCF_000414465.1	<i>Gardnerella vaginalis</i> JCP8481A	SAMN02436910
14	GCF_000414445.1	<i>Gardnerella vaginalis</i> JCP8481B	SAMN02436829
15	GCF_000263615.1	<i>Gardnerella vaginalis</i> 00703Bmash	SAMN02393781
16	GCF_000263515.1	<i>Gardnerella vaginalis</i> 00703Cmash	SAMN02393782
17	GCF_000263635.1	<i>Gardnerella vaginalis</i> 00703Dmash	SAMN02393783
18	GCF_000414665.1	<i>Gardnerella vaginalis</i> JCP7659	SAMN02436712
19	GCF_000414625.1	<i>Gardnerella vaginalis</i> JCP7719	SAMN02436711
20	GCF_000414605.1	<i>Gardnerella vaginalis</i> JCP8017A	SAMN02436912
21	GCF_000414585.1	<i>Gardnerella vaginalis</i> JCP8017B	SAMN02436773
22	GCF_001546445.1	<i>Gardnerella vaginalis</i> GED7275B	SAMN03851014
23	GCF_002861905.1	<i>Gardnerella vaginalis</i> UMB0830	SAMN08193679
24	GCF_002861885.1	<i>Gardnerella vaginalis</i> UMB0833	SAMN08193680
25	GCF_002884775.1	<i>Gardnerella vaginalis</i> UMB1686	SAMN07511412
26	GCF_000165615.1	<i>Gardnerella vaginalis</i> 101	SAMN02472073

## 2.2. Identification of Orthologues

For the prediction of orthologous genes, the Orthofinder v2.5.4 software was employed (<https://github.com/davidemms/OrthoFinder>), performing a clustering calculation based on the Markov Clustering Algorithm (MCL). This algorithm compared all genomes against each other to define clusters with high levels of similarity within the analyzed data (Emms and Kelly 2015). Associating these results with script *in-house* (`ortho_pangenome_splitter.pl`) that classifies genes, three sets were obtained: the core genome, which is present in all analyzed lineages and essential for the microorganism's survival; shared, which gathers genes present in two or more lineages but not in all; and singletons, which are specific to only one lineage (Kim et al. 2020).

## 2.3 Identification of Positively Selected Genes

POsitive selecTION (POTION) v1.2 (<https://www.lmb.cnptia.embrapa.br//share/POTION/>) is a massively parallel program that identifies positive Darwinian selection in genomic analyses of groups of homologous genes through phylogenetic comparisons of protein-coding genes. An end-to-end pipeline, open source, to select groups that are related with non-synonymous substitution, where occurs a mutation which will cause the change of amino acid in the protein that are targets of natural selection and thus know which phenotypic patterns have evolved at the molecular level (Hongo et al. 2015). To measure positive selection is at the codon level discriminating  $\omega$ , the ratio of nonsynonymous to synonymous substitution rates (Gao et al. 2019). Orthofinder's orthogroups.tsv output was used as an input for signaling the homologous groups, and an *in-house* script (ortho2mcl.pl) was used to select the count of each orthogroup and how many times it appeared in each of the lineages.

These orthogroups were used for functional analysis and identification of which genes and proteins are mainly related to the resistance phenotype of *G. vaginalis*, using the Basic Local Alignment Search Tool (BLASTp) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

#### 2.4 Subtractive Genomics and Protein Subcellular Localization Prediction

It was necessary to use the BLASTp algorithm to compare the core genome with the human genome, given the classification of orthologous genes previously performed by Orthofinder, to perform subtractive genomics. This analysis was performed to analyze which protein coding genes are inserted in the genomes and do not have homology with the host, in this case, the human being, to avoid adverse effects. To continue the analysis of possible targets for drugs (Soares et al. 2013), an *in house* script (core-non-host.pl) was used to align the amino acid sequences (faa) in analysis against the human genome.

An analysis was performed to predict the subcellular localization of the proteins using SurfG+ software (Pacheco et al. 2011) to define which ones would be analyzed for drug targets. Those proteins related to vital metabolic processes and bacterial survival, which are usually cytoplasmic, were selected for drug target analysis (Rodrigues et al. 2019).

#### 2.5 Genomic Resistance and Virulence Analysis

Pan Resistome Analysis Pipeline (PRAP) (<https://github.com/syyrjx-hyc/PRAP>); was used to predict Antibiotic resistance genes (ARGs). This platform-independent Python3 tool predict genes related to resistance using whole genomes based on the Comprehensive Antibiotic Resistance Database (CARD) or ResFinder databases; furthermore, the annotations are used for the characterization of pan-resistomes that can be used to characterize the distribution of ARGs among the input genomes (He et al. 2020). Only the CARD database was used for Pan-resistome analyses.

Pan Virulence and resisTance Analysis (PanViTa) (<https://github.com/dlnrodrigues/panvita>) is a tool that not only predicts ARGs and Pan-resistomes like PRAP, but it also analyzes virulence and antibacterial biocide and metal resistance, using the CARD, Antibacterial Biocide and Metal Resistance Genes Database (BACMET) and Virulence Factor Database (VFDB). For the analyses of the *G. vaginalis* genomes, only virulence analyses were carried out using minimum identity to infer presence (“=70”) (Rodrigues et al. 2023).

## 2.6 Selection of Possible Drug Target Candidates

The Essential Gene Database (DEG) was used to assess whether these candidate proteins for drug targets are translated from genes that are essential for the permanence of the studied bacteria and are indispensable for the biological processes of it (Zhang et al. 2004).

Subsequently, the DrugBank platform ([www.drugbank.ca](http://www.drugbank.ca)) was employed to predict proteins with significant interactions and drugability. This database offers comprehensive information on drugs, encompassing their drug interactions, bindings, associations, and mechanisms of action. Within this context, we exclusively selected drug targets whose drugability has been assessed as high (Wishart et al. 2018).

The proteins identified by the VFDB database with the virulence phenotype were selected as drug targets since they are essential for the survival and spread of the pathogen to the host (Passarelli-Araujo et al. 2019).

## 2.7 Protein Tertiary Structure Prediction

Through the sequence of the proteins, the prediction of their 3D structure was performed using AlphaFold v2.3.2 (<https://github.com/google-deepmind/alphafold/tree/main>) (Jumper et al. 2021a). This AI system developed by DeepMind predicts the secondary and later the tertiary structure of the protein in a reliable and reproducible way, where an accurate structure prediction is represented according to multiple sets of sequence alignments (MSAs) and procedures based on evolutionary, geometric and physical constraints, where all the 3D coordinates of all the protein atoms are predicted from the aligned primary amino acid sequence, according to their distance and interactions comparing with the PDB.

The network of this software comprises two main steps; the first is the processing of an MSA comparing the number of sequences with the number of residues, a neural network block called Evoformer to make a matrix  $N_{seq} \times N_{res}$  ( $N_{seq}$ , number of sequences;  $N_{res}$ , number of residues), there will be defined the spatial and evolutionary relationships. Afterward, it is evaluated according to the rotation and translation of the 3D structure. The model with the highest predicted Local Distance Difference Test (LDDT) score (pLDDT) will be selected, which measures the local distance differences of all the atoms in a model (Tunyasuvunakool et al. 2021; Jumper et al. 2021b).

To improve the structure of the selected proteins, GalaxyRefine (Heo et al. 2013) was used, which refines the side chains by molecular dynamics simulation to improve the quality of the structure. The Ramachandran Diagram predicted using the PROCHECK v.3.5 software confirmed the protein structure by evaluating the stereochemical quality of protein structures (Laskowski et al. 1993).

## 2.8 Molecular Modeling Analysis

The Autodock tool (ADT) of the MGLTool package v1.5.7 (Morris et al. 2009) was used for 3D structure analysis of the final drug target candidates, where a grid box for each target was made to cover the region of the protein active site predicted by the DoGSiteScorer software of the Protein plus server (Volkamer et al. 2012; Fährrolfes et al. 2017). In parallel, 5008 natural compounds were selected as ligands from the ZINC database (Irwin et al. 2012). The Natural Compound ligand library with 5008 molecules was prepared according to the criteria stated by Lipinski's rule, such as - hydrogen bond donors not greater than 5, hydrogen bond acceptors not greater than 10, molecular weight not greater than 500 Da, and octanol-



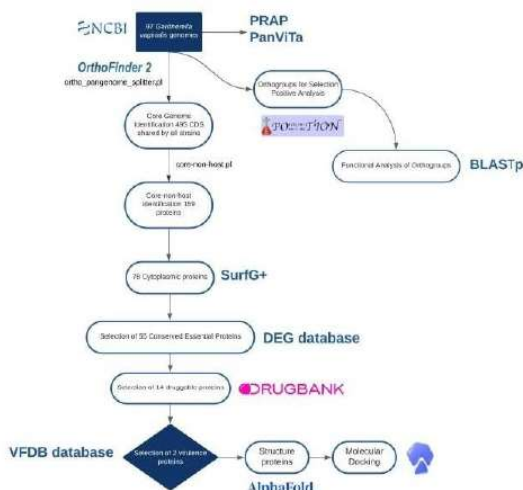
water partition coefficient (log P) not greater than 5 for their drug-like properties. The compound molecules were prepared according to the above-mentioned rule and downloaded from the ZINC database in 2017 and the library was first used in our published work by Alissa de Sarom et al., in 2018 (Gao et al. 2019) against *Haemophilus ducreyi*. Af. AutoDock Vina software was used to perform the modeling analysis itself (Trott and Olson 2010). The top 10 molecules were extracted for each target sequentially through an *in-house* Python script and analyzed according to their binding affinity and hydrogen bridges. For visualization of this binding and extraction of the 3D image of the target the Chimera software was used (Pettersen et al. 2004).

BIOVIA Discovery Studio 21.1.0.0 (Sharma et al. 2021) was used to create 2D interaction photos of the complex between the select proteins and their respective ligands predict by Docking in Autodock toll.

### 3. Results

All the steps that were carried out are included in a flowchart. 97 *Gardnerella vaginalis* genomes were compared using the methodology inserted in Figure 1, which summarizes the proteins selected for molecular docking and the orthogroups selected for positive selection analysis.

**Figure 1:** Workflow of methodologies used to select the drug candidates and genes related to Darwinian positive selection.



Source: Authors, 2024.

Note: In this figure, a workflow with all the materials and methods used in this work can be observed, with the methods inside each of the oval balloons and the program symbol next to it.

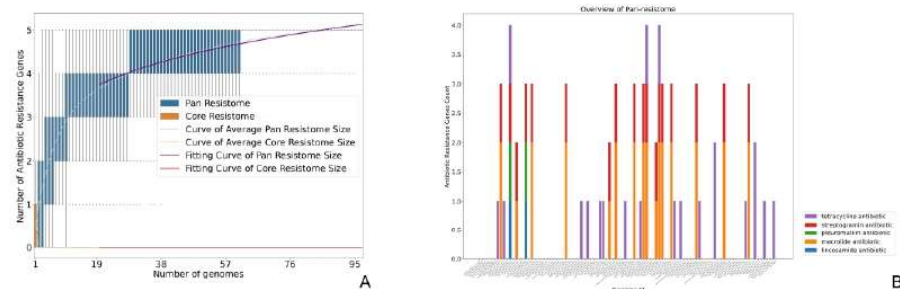
#### 3.1 Analysis of Positively Selected Orthogroups

Orthofinder selected 2568 orthogroups to predict positive Darwinian selection by Potion. After filtering, 231 groups were selected as valid. 29 groups were classified in Model 8, chosen for having the best natural log-likelihood value (lnL value) of the observed data sequence given the model parameters. A cut-off point of 5% significance was used for the program's statistical analyses.

To perform a functional analysis of the orthogroups that were positively selected (Supplementary Table S2), we performed blast analyses (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) with the blastp option. Of the 29 groups, 19 were hypothetical groups with no known functions. Among the other 10 groups, several are involved in ion, amino acid, and carbohydrate transport processes and in cell wall biogenesis and transcriptional movements. In addition, one group is engaged with antibiotic efflux, for example, gene *Mef(A)*, which is related to antibiotic macrolide.

### 3.2 Resistance and Virulence Analysis

**Figure 2:** Analysis of pan resistome and resistance genes using PRAP software.



Source: Authors, 2024.

Note: Figure A shows a growth curve of pan resistome (in blue) and core resistome (in orange). In addition, in dotted lines, a curve of average and the fitting curve for pan and core resistome are shown. In figure B, can be observed the number of antibiotic resistance genes for each antibiotic class in different colors, such as tetracycline in purple, streptogramin in red, pleuromutilin in green, macrolide in orange, and lincosamide in blue.

For pan resistome analysis (Figure 2A), no genes belonging to the core resistome were found, i.e., those found in all strains. However, the power law model performed the pan-resistome, generating a pan-resistome size  $P = 2.078 * x^{0.197}$  ( $R^2 = 0.803$ ). This R-value of less than 1 suggests an open pan-resistome, revealing that *G. vaginalis* can acquire resistance genes from other organisms throughout its evolution.

Through the CARD database, the PRAP software also found gene *IsaC* in two lineages that provides resistance to lincosamide, pleuromutilin, and streptogramin antibiotic. The gene 'mel' was found in seventeen lineages, possibly providing resistance to macrolide and streptogramin antibiotics. The gene *mef(A)* was found in thirteen, providing resistance to macrolide antibiotic. Finally, the genes *tet(M)* and *tet(L)* were found in nineteen and two lineages, respectively, and both provide resistance against tetracycline antibiotics (Figure 2B).

Using PanViTa we found *sigA/rpoV* and *msbB* alleles related to virulence, which were predicted by using the VFDB database (Supplementary Table S3).

### 3.3 Selection of Drug Targets

Initially, proteins were selected based on the orthogroup they belonged to, using Orthofinder and *in-house* scripts (*ortho\_pangenome\_splitter.pl*). The analysis encompassed all proteins identified in the

1 core, resulting in 495 proteins across all genomes. From this pool, only 159 proteins were chosen (core-  
 2 non-host.pl) due to their non-homology with the host, a precaution taken to mitigate potential adverse  
 3 effects. In the search for proteins with cytoplasmic localization for use as drug targets, SurfG+ identified  
 4 and selected only 78 proteins.  
 5

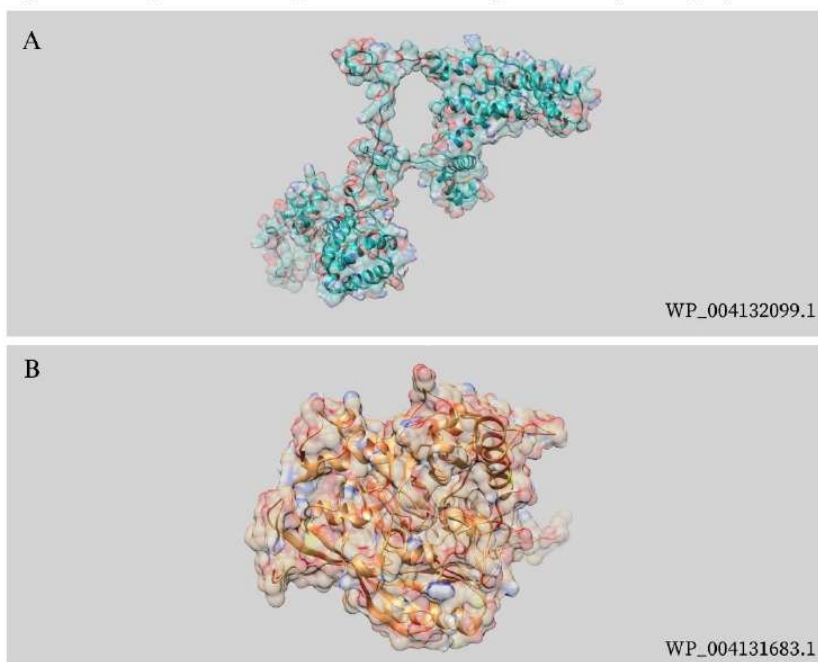
6 The remaining proteins underwent scrutiny based on essentiality (n = 55) for the species,  
 7 druggability (n = 14), and virulence, employing DEG, VFDB, and DrugBank programs, respectively.  
 8  
 9 Ultimately, two proteins, WP\_004132099.1 (RNA polymerase sigma factor sigA) (gene *sigA*) and  
 10 WP\_004131683.1 (UDP-N-acetylenolpyruvoylglucosamine reductase) (gene *murB*), were chosen as  
 11 potential drug targets for subsequent structural and molecular docking analyses.  
 12  
 13

### 14 3.4 Prediction of the Tertiary Structure of the Selected Proteins

15 AlphaFold was used to predict the tertiary structure of two previously selected proteins,  
 16 WP\_004132099.1 and WP\_004131683.1 (Figure 3). For each, 24 models were predicted after reordering  
 17 by model confidence, containing the prediction with the (i + 1)-th highest confidence, predicted from the  
 18 pLDDT in which "ranked\_0.pdb" has the highest confidence.  
 19  
 20 The predicted Ramachandran diagrams showed a high quality of the structure of the predicted  
 21 proteins and are present in the supplementary material (Figure S1 and Figure S2).  
 22  
 23

24 The predicted Ramachandran diagrams showed a high quality of the structure of the predicted  
 25 proteins and are present in the supplementary material (Figure S1 and Figure S2).  
 26  
 27

28 **Figure 3:** Tertiary structure of WP\_004132099.1 and WP\_004131683.1 predict by AlphaFold.



54 Source: Authors, 2024.

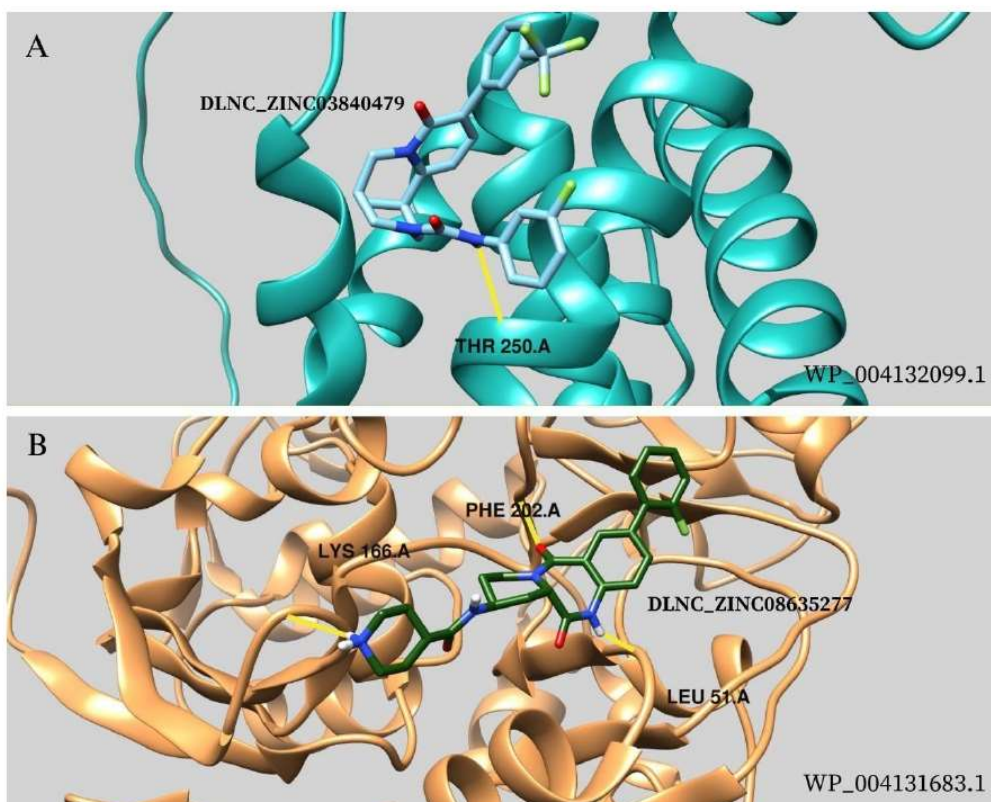
55  
 56 Note: Figure A shows WP\_004132099.1 tertiary structure predicts by AlphaFold (in blue). Figure B shows  
 57 WP\_004131683.1 tertiary structure predicts by AlphaFold (in orange).  
 58  
 59



### 3.5 Molecular Docking of Selected Proteins with Natural Compounds

The previously selected proteins WP\_004132099.1 and WP\_004131683.1 were used for molecular docking analysis with 5008 natural compounds. The best ligands were respectively DLNC\_ZINC08635277 with a free energy score of -9.68 and a hydrogen bond on Tyrosine 250 belonging to the active site of protein WP\_004132099.1 and DLNC\_ZINC03840479 with a score of -9.689 with three hydrogen bonds on residues Lysine 166, Phenylalanine 202 and Leucine 51 of the active site of protein WP\_004131683.1 (Figure 4 and Table 2). This interactions in 2D are illustrated in supplementary material (Figure S3 and Figure S4).

**Figure 4:** Docking molecular analyses of the proposed new drug targets.



Source: Authors, 2024.

Note: In figure A, can be observed the protein WP\_004132099.1 in blue color and its best compound ligand DLNC\_ZINC08635277 in shades of blue color. The link between the protein and its ligand by the hydrogen bridge with the amino acid Tyrosine 250 (THR 250) is stained yellow. In figure B, the protein WP\_004131683.1 in orange color and its best compound ligand DLNC\_ZINC03840479 in shades of green color can be observed. The link between the protein and its ligand by the hydrogen bridge with the amino acids Lysine 166 (LYS 166), Phenylalanine 202 (PHE 202) and Leucine 51 (LEU 51) is stained yellow.

**Table 2:** Proteins selected for drug targets with their predicted natural compounds from the ZINC database, hydrogen bonding and residue analyzed by Chimera software.

Protein	ZINC Compounds	Auto Dock Vina	Hbonds number	Residue	Ångström (Å)
WP_004132099.1	DLNC_ZINC08635277	-9.68	1	THR 250 LYS 166	2.291 Å 2.123 Å
WP_004131683.1	DLNC_ZINC03840479	-9.689	3	PHE 202 LEU 51	2.327 Å 2.227 Å

#### 4. Discussion

*G. vaginalis* is a species of great medical importance, especially for women's health. Recent studies have shown its great gene diversity, antibiotic resistance, and virulence phenotype (Schellenberg et al. 2017; Chen et al. 2021). Comparative analyses were carried out to evaluate the positive Darwinian selection of all the annotated genomes deposited in the NCBI database of *G. vaginalis*.

Positive Darwinian selection is understood as characteristics that have been selected over time to maintain phenotypes that are considered advantageous for the survival of the species. What occurs are mutations that lead to amino acid changes in the protein that are targets of natural selection, i.e., non-synonymous substitution (Farhat et al. 2013). Antibiotic resistance may be one of these mechanisms that are evolutionarily selected. Through phylogenetic comparison of protein-coding genes, Darwinian positive selection analyses showed relevant orthogroups that were selected, which shows that *G. vaginalis* is a species with high genetic diversity and that genes have been added through evolution (Mu et al. 2010). Fundamental groups that were analyzed are the *Mef(A)* gene, which is related to resistance to macrolide class antibiotics (Zhang et al. 2016), and *tet(M)* and *tet(L)* related to tetracycline resistance (Ammor et al. 2008). This relationship shows how much this pathogenic species has evolved towards the resistance phenotype.

The resistance and virulence of these strains were then predicted to correlate with the orthogroup positively selected by PRAP and PanVita, respectively. These analyses also showed resistance to the macrolide group class. Several studies have shown strains of *G. vaginalis* resistant to macrolides, which are the first line of treatment for bacterial vaginosis, and can be associated with treatment failure and a high recurrence of the condition (Schuyler et al. 2015). Another comparative genomics study demonstrated the presence of two genes related to resistance to this class of antibiotic as well as one major facilitator superfamily (MFS) transporter, as well as four unknown multidrug efflux systems (Tarracchini et al. 2020).

In addition, when analyzing the pan-resistome using the power law model, the R-value was lower than 1, suggesting an open pan-resistome; this can be understood as a greater capacity to acquire resistance genes, as shown by the analysis of which genes were positively selected. Furthermore, many singletons show the relationship of resistance genes that evolve with the species, guaranteeing this phenotype for *G. vaginalis* (Decano et al. 2021).

Based on this result and the possibility that *G. vaginalis* may continue to be a microorganism that will acquire resistance genes throughout its evolution, it already has resistance genes; reverse vaccinology and molecular docking analyses were carried out. Firstly, within subtractive genomics, the genes present in all the strains, i.e., those that belonged to the core genome, were compared in terms of their presence in their host to avoid possible adverse reactions. The proteins encoded by these genes, which were only present in the bacteria, were tested for their subcellular localization. Proteins considered to be cytoplasmic are



usually used as drug targets because they are more likely to be involved in the microorganism's survival mechanisms (Peraman et al. 2021), so 78 proteins went on to be analyzed for essentiality.

After the essentiality analysis, these proteins were tested against databases to assess whether they could bind to antibiotic compounds and were already considered virulence mechanisms among the most diverse organisms. Finally, two proteins passed all the quality filters and have undergone molecular docking analysis to bind to any natural compounds belonging to an antibiotic.

The first protein, WP\_004132099.1, is a RNA polymerase sigma factor sigA. This protein has the vital function of directing the RNA polymerase enzyme to its promoter, which allows the bacterial transcription process to begin (Ganapathy et al. 2021; Switzer et al. 2022). Drugs that involve this protein as a drug target are already widely studied, for example, acting as an inhibitor of infection caused by *Staphylococcus aureus* (Haupenthal et al. 2020) or even tuberculosis (Mukinda et al. 2012; Caputo et al. 2022). In addition, this gene is related to the putative transcriptional regulator *WhiB7*, which is crucial for resistance to various classes of antibiotics; this has been demonstrated in *Mycobacterium smegmatis*, and so inhibiting this WhiB7 binding could make ineffective antibiotics effective again (Burian et al. 2013). Finally, a detail that is striking about *G. vaginalis* is that some species present the *sigA* (*RpoV*) gene as a virulence factor, predicted by the PanVita software in five strains, also called sigma A, and involved in this transcription initiation process. For these reasons, this drug target is essential, especially if it is to be used against the diseases caused by these bacteria.

The second protein, WP\_004131683.1, is an UDP-N-acetylenolpyruvylglucosamine reductase. This protein catalyzes the reaction of UDP-N-acetylmuramic acid, an essential part of peptidoglycan. Peptidoglycan is responsible for maintaining the structure of the cell wall of prokaryotes ([CSL STYLE ERROR: reference with no printed form.]; Amera et al. 2020). Because it is a molecule that plays an essential role in forming the cell wall and maintaining the structure and, consequently, the presence of the microorganism in the host, it is often used in studies as a drug target. *In silico* work using *Acinetobacter baumannii*, *Escherichia coli*, *Pseudomonas aeruginosa*, and even *in vitro* studies against *Corynebacterium glutamicum*, demonstrated that the *murB* gene, which encodes this protein, had important effects on maintaining the survival of microorganisms and could be used as drug targets (Shi et al. 2019; Amiranashvili et al. 2020; Amera et al. 2020).

## 5. Conclusions

In conclusion, the *G. vaginalis* genomes showed the selection of several positively selected genes, especially those related to resistance and virulence. In addition, the species' pan-resistome is open, demonstrating that in addition to the genes already found, it may be able to acquire new genes throughout evolution. This shows that this phenotype is essential for maintaining the species in its pathogenic trait during its evolution.

Coupled with this, two proteins with the potential to promote virulence were predicted, selected by subtractive genomics. Given the resistance observed, molecular docking analyses were carried out on these proteins with natural compounds to choose them *in silico* as potential drug targets. These proteins were linked with two natural compounds with good binding energy; in addition to that, these proteins reveal themselves to be excellent targets due to their functions in critical survival processes because they are also

1 linked to the resistance and virulence genes found here and have already been tested against other  
 2 microorganisms. These targets should also be tested *in vitro* and *in vivo* to find more promising results  
 3 against *G. vaginalis*.  
 4

5  
 6 **Acknowledgments:** We thank the agencies of the Brazilian Federal Agency for Support and Evaluation of Graduate Education  
 7 (CAPES) and the Minas Gerais Research Funding Foundation (FAPEMIG). We are thankful for the post-graduate program in Genetics  
 8 at the Federal University of Minas Gerais and the Medicina Tropicale Infectologia at the Federal University of Triângulo Mineiro.  
 9

## 10 11 12 **References**

- 13  
 14 Amera GM, Khan RJ, Pathak A, et al (2020) Structure based drug designing and discovery of promising  
 15 lead molecules against UDP-N-acetylenolpyruvoylglucosamine reductase (MurB): A potential drug  
 16 target in multi-drug resistant *Acinetobacter baumannii*. *J Mol Graph Model* 100:107675.  
 17 <https://doi.org/10.1016/J.JMGM.2020.107675>  
 18  
 19 Amiranashvili L, Nadaraia N, Merlani M, et al (2020) Antimicrobial Activity of Nitrogen-Containing 5- $\alpha$ -  
 20 Androstane Derivatives: In Silico and Experimental Studies. *Antibiot* 2020, Vol 9, Page 224 9:224.  
 21 <https://doi.org/10.3390/ANTIBIOTICS9050224>  
 22  
 23 Ammor MS, Gueimonde M, Danielsen M, et al (2008) Two different tetracycline resistance mechanisms,  
 24 plasmid-carried tet(L) and chromosomally located transposon-associated tet(M), coexist in  
 25 *Lactobacillus sakei* rits 9. *Appl Environ Microbiol* 74:1394–1401.  
 26 [https://doi.org/10.1128/AEM.01463-07/ASSET/A61F9864-0484-4C7B-B3A6-  
 27 61B17ACD0122/ASSETS/GRAPHIC/ZAM0050886410004.JPEG](https://doi.org/10.1128/AEM.01463-07/ASSET/A61F9864-0484-4C7B-B3A6-61B17ACD0122/ASSETS/GRAPHIC/ZAM0050886410004.JPEG)  
 28  
 29 Bochner BR, Huang HC, Schieven GL, Ames BN (1980) Positive selection for loss of tetracycline  
 30 resistance. *J Bacteriol* 143:926–933. <https://doi.org/10.1128/JB.143.2.926-933.1980>  
 31  
 32 Burian J, Yim G, Hsing M, et al (2013) The mycobacterial antibiotic resistance determinant WhiB7 acts as  
 33 a transcriptional activator by binding the primary sigma factor SigA (RpoV). *Nucleic Acids Res*  
 34 41:10062–10076. <https://doi.org/10.1093/NAR/GKT751>  
 35  
 36 Caputo A, Sartini S, Levati E, et al (2022) An Optimized Workflow for the Discovery of New Antimicrobial  
 37 Compounds Targeting Bacterial RNA Polymerase Complex Formation. *Antibiotics* 11:1449.  
 38 <https://doi.org/10.3390/ANTIBIOTICS11101449/S1>  
 39  
 40 Chen X, Lu Y, Chen T, Li R (2021) The Female Vaginal Microbiome in Health and Bacterial Vaginosis.  
 41 *Front Cell Infect Microbiol* 11:631972. <https://doi.org/10.3389/FCIMB.2021.631972/BIBTEX>  
 42  
 43 Decano AG, Pettigrew K, Sabiiti W, et al (2021) Pan-resistome characterization of uropathogenic  
 44 *Escherichia coli* and *Klebsiella pneumoniae* strains circulating in Uganda and Kenya, isolated from  
 45 2017–2018. *Antibiotics* 10:1547. <https://doi.org/10.3390/ANTIBIOTICS10121547/S1>  
 46  
 47 Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons  
 48 dramatically improves orthogroup inference accuracy. *Genome Biol* 16:1–14.  
 49 <https://doi.org/10.1186/S13059-015-0721-2/FIGURES/7>  
 50  
 51 Fährrolfes R, Bietz S, Flachsenberg F, et al (2017) ProteinsPlus: a web portal for structure analysis of  
 52 macromolecules. *Nucleic Acids Res* 45:W337–W343. <https://doi.org/10.1093/NAR/GKX333>  
 53  
 54 Farhat MR, Shapiro BJ, Kieser KJ, et al (2013) Genomic analysis identifies targets of convergent positive  
 55  
 56  
 57  
 58  
 59

- selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* 2013 45:10 45:1183–1189.  
<https://doi.org/10.1038/ng.2747>
- 1  
2  
3 Ganapathy S, Wiegard JC, Hartmann RK (2021) Rapid preparation of 6S RNA-free *B. subtilis*  $\sigma$ A-RNA  
4 polymerase and  $\sigma$ A. *J Microbiol Methods* 190:106324.  
5  
6 <https://doi.org/10.1016/J.MIMET.2021.106324>
- 7 Gao F, Chen C, Arab DA, et al (2019) EasyCodeML: A visual tool for analysis of selection using CodeML.  
8  
9 *Ecol Evol* 9:3891. <https://doi.org/10.1002/ECE3.5015>
- 10 Giacchetto Felice A, Guimarães Sousa E, Vieira Dominici F, et al (2023) Pangenome Analysis Reveals a  
11  
12 High Degree of Genetic Diversity in *Gardnerella vaginalis*: An In Silico Approach. *Venereol* 2023,  
13 Vol 2, Pages 132-146 2:132–146. <https://doi.org/10.3390/VENEREOLGY2040012>
- 14  
15 Hauptenthal J, Kautz Y, Elgaher WAM, et al (2020) Evaluation of Bacterial RNA Polymerase Inhibitors in  
16 a *Staphylococcus aureus*-Based Wound Infection Model in SKH1 Mice. *ACS Infect Dis* 6:2573–  
17  
18 2581. [https://doi.org/10.1021/ACSINFECDIS.0C00034/SUPPL\\_FILE/ID0C00034\\_SI\\_001.PDF](https://doi.org/10.1021/ACSINFECDIS.0C00034/SUPPL_FILE/ID0C00034_SI_001.PDF)
- 19 He Y, Zhou X, Chen Z, et al (2020) PRAP: Pan Resistome analysis pipeline. *BMC Bioinformatics* 21:1–8.  
20  
21 <https://doi.org/10.1186/S12859-019-3335-Y/FIGURES/5>
- 22 Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking.  
23  
24 *Nucleic Acids Res* 41:W384. <https://doi.org/10.1093/NAR/GKT458>
- 25 Hongo JA, de Castro GM, Cintra LC, et al (2015) POTION: An end-to-end pipeline for positive Darwinian  
26  
27 selection detection in genome-scale data through phylogenetic comparison of protein-coding genes.  
28 *BMC Genomics* 16:1–16. <https://doi.org/10.1186/S12864-015-1765-0/FIGURES/3>
- 29  
30 Irwin JJ, Sterling T, Mysinger MM, et al (2012) ZINC: A free tool to discover chemistry for biology. *J*  
31 *Chem Inf Model* 52:1757–1768.  
32  
33 [https://doi.org/10.1021/CI3001277/SUPPL\\_FILE/CI3001277\\_SI\\_001.PDF](https://doi.org/10.1021/CI3001277/SUPPL_FILE/CI3001277_SI_001.PDF)
- 34  
35 Jumper J, Evans R, Pritzel A, et al (2021a) Highly accurate protein structure prediction with AlphaFold.  
36 *Nat* 2021 5967873 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- 37  
38 Jumper J, Evans R, Pritzel A, et al (2021b) Applying and improving AlphaFold at CASP14. *Proteins Struct*  
39 *Funct Bioinforma* 89:1711–1721. <https://doi.org/10.1002/PROT.26257>
- 40  
41 Kim Y, Gu C, Kim HU, Lee SY (2020) Current status of pan-genome analysis for pathogenic bacteria. *Curr*  
42 *Opin Biotechnol* 63:54–62. <https://doi.org/10.1016/J.COPBIO.2019.12.001>
- 43  
44 Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the  
45 stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.  
46  
47 <https://doi.org/10.1107/S0021889892009944>
- 48 Lefebvre T, Stanhope MJ (2009) Pervasive, genome-wide positive selection leading to functional  
49  
50 divergence in the bacterial genus *Campylobacter*. *Genome Res* 19:1224–1232.  
51 <https://doi.org/10.1101/GR.089250.108>
- 52  
53 Ma X, Wang X, Ye S, et al (2022) Biofilm and pathogenic factor analysis of *Gardnerella vaginalis*  
54 associated with bacterial vaginosis in Northeast China. *Front Microbiol* 13:1033040.  
55  
56 <https://doi.org/10.3389/FMICB.2022.1033040/BIBTEX>
- 57  
58 Morris GM, Ruth H, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking  
59 with selective receptor flexibility. *J Comput Chem* 30:2785–2791.

<https://doi.org/10.1002/JCC.21256>

- 1 Mu J, Myers RA, Jiang H, et al (2010) Plasmodium falciparum genome-wide scans for positive selection,  
2 recombination hot spots and resistance to antimalarial drugs. *Nat Genet* 2010 42:268–271.  
3 <https://doi.org/10.1038/ng.528>  
4
- 5 Mukinda FK, Theron D, Van Der Spuy GD, et al (2012) Rise in rifampicin-mono-resistant tuberculosis in  
6 Western Cape, South Africa. *Int J Tuberc Lung Dis* 16:196–202.  
7 <https://doi.org/10.5588/IJTL.11.0116>  
8
- 9 Pacheco LG, Slade SE, Seyffert N, et al (2011) A combined approach for comparative exoproteome  
10 analysis of *Corynebacterium pseudotuberculosis*. *BMC Microbiol* 11:1–14.  
11 <https://doi.org/10.1186/1471-2180-11-12/FIGURES/5>  
12
- 13 Passarelli-Araujo H, Palmeiro JK, Moharana KC, et al (2019) Genomic analysis unveils important aspects  
14 of population structure, virulence, and antimicrobial resistance in *Klebsiella aerogenes*. *FEBS J*  
15 286:3797–3810. <https://doi.org/10.1111/FEBS.15005>  
16
- 17 Peraman R, Sure SK, Dusthacker VNA, et al (2021) Insights on recent approaches in drug discovery  
18 strategies and untapped drug targets against drug resistance. *Futur J Pharm Sci* 2021 71 7:1–25.  
19 <https://doi.org/10.1186/S43094-021-00196-5>  
20
- 21 Petersen L, Bollback JP, Dimmic M, et al (2007) Genes under positive selection in *Escherichia coli*.  
22 *Genome Res* 17:1336–1343. <https://doi.org/10.1101/GR.6254707>  
23
- 24 Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera—A visualization system for exploratory  
25 research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/JCC.20084>  
26
- 27 Ping WONG Y, Chin TAN G, Ken WONG K, et al *Gardnerella vaginalis* in perinatology: An overview of  
28 the clinicopathological correlation  
29
- 30 Rigo GV, Tasca T (2020) Vaginitis: Review on Drug Resistance. *Curr Drug Targets* 21:1672–1686.  
31 <https://doi.org/10.2174/1389450121666200804112340>  
32
- 33 Rodrigues DLN, Ariute JC, Rodrigues da Costa FM, et al (2023) PanViTa: Pan Virulence and resisTance  
34 analysis. *Front Bioinforma* 3:1070406. <https://doi.org/10.3389/FBINF.2023.1070406>  
35
- 36 Rodrigues TCV, Jaiswal AK, De Sarom A, et al (2019) Reverse vaccinology and subtractive genomics  
37 reveal new therapeutic targets against *Mycoplasma pneumoniae*: a causative agent of pneumonia. *R*  
38 *Soc Open Sci* 6:. <https://doi.org/10.1098/RSOS.190907>  
39
- 40 Schellenberg JJ, Patterson MH, Hill JE (2017) *Gardnerella vaginalis* diversity and ecology in relation to  
41 vaginal symptoms. *Res Microbiol* 168:837–844. <https://doi.org/10.1016/J.RESMIC.2017.02.011>  
42
- 43 Schuyler JA, Chadwick SG, Mordechai E, et al (2015) Draft Genome Sequence of a Metronidazole-  
44 Resistant *Gardnerella vaginalis* Isolate. *Genome Announc* 3:.  
45 <https://doi.org/10.1128/GENOMEA.00992-15>  
46
- 47 Sharma S, Shama A, Gupta U (2021) Molecular Docking Studies on the Anti-Fungal Activity of *Allium*  
48 *Sativum* (Garlic) Against *Mucormycosis* (Black Fungus) by BIOVIA Discovery Studio Visualizer  
49 21.1.0.0. <https://doi.org/10.21203/RS.3.RS-888192/V1>  
50
- 51 Shi T, Ma Q, Liu X, et al (2019) Double deletion of *murA* and *murB* induced temperature sensitivity in  
52 *Corynebacterium glutamicum*. *Bioengineered* 10:548–560.  
53 <https://doi.org/10.1080/21655979.2019.1685058>  
54
- 55  
56  
57  
58  
59  
60

- 1 Soares SC, Silva A, Trost E, et al (2013) The Pan-Genome of the Animal Pathogen *Corynebacterium*  
2 pseudotuberculosis Reveals Differences in Genome Plasticity between the Biovar ovis and equi  
3 Strains. *PLoS One* 8:e53818. <https://doi.org/10.1371/JOURNAL.PONE.0053818>
- 4 Soyer Y, Orsi RH, Rodriguez-Rivera LD, et al (2009) Genome wide evolutionary analyses reveal serotype  
5 specific patterns of positive selection in selected *Salmonella* serotypes. *BMC Evol Biol* 9:1–18.  
6 <https://doi.org/10.1186/1471-2148-9-264/TABLES/6>
- 7  
8  
9 Switzer A, Burchell L, Mitsidis P, et al (2022) A Role for the RNA Polymerase Gene Specificity Factor  
10  $\sigma 54$  in the Uniform Colony Growth of Uropathogenic *Escherichia coli*. *J Bacteriol* 204:.  
11 <https://doi.org/10.1128/JB.00031-22>
- 12  
13 Tarracchini C, Lugli GA, Mancabelli L, et al (2020) Assessing the Genomic Variability of *Gardnerella*  
14 vaginalis through Comparative Genomic Analyses: Evolutionary and Ecological Implications. *Appl*  
15 *Environ Microbiol* 87:1–16. [https://doi.org/10.1128/AEM.02188-20/SUPPL\\_FILE/AEM.02188-20-](https://doi.org/10.1128/AEM.02188-20/SUPPL_FILE/AEM.02188-20-)  
16 [SD002.XLSX](https://doi.org/10.1128/AEM.02188-20/SUPPL_FILE/AEM.02188-20-SD002.XLSX)
- 17  
18  
19 Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring  
20 function, efficient optimization and multithreading. *J Comput Chem* 31:455.  
21 <https://doi.org/10.1002/JCC.21334>
- 22  
23  
24 Tunyasuvunakool K, Adler J, Wu Z, et al (2021) Highly accurate protein structure prediction for the human  
25 proteome. *Nat* 2021 5967873 596:590–596. <https://doi.org/10.1038/s41586-021-03828-1>
- 26  
27 Volkamer A, Kuhn D, Grombacher T, et al (2012) Combining global and local measures for structure-based  
28 druggability predictions. *J Chem Inf Model* 52:360–372.  
29 [https://doi.org/10.1021/CI200454V/ASSET/IMAGES/MEDIUM/CI-2011-00454V\\_0015.GIF](https://doi.org/10.1021/CI200454V/ASSET/IMAGES/MEDIUM/CI-2011-00454V_0015.GIF)
- 30  
31  
32 Wishart DS, Feunang YD, Guo AC, et al (2018) DrugBank 5.0: a major update to the DrugBank database  
33 for 2018. *Nucleic Acids Res* 46:D1074–D1082. <https://doi.org/10.1093/NAR/GKX1037>
- 34  
35  
36 Zhang K, Lu M, Zhu X, et al (2022) Antibiotic resistance and pathogenicity assessment of various  
37 *Gardnerella* sp. strains in local China. *Front Microbiol* 13:1009798.  
38 <https://doi.org/10.3389/FMICB.2022.1009798/BIBTEX>
- 39  
40  
41 Zhang R, Ou HY, Zhang CT (2004) DEG: a database of essential genes. *Nucleic Acids Res* 32:.  
42 <https://doi.org/10.1093/NAR/GKH024>
- 43  
44  
45 Zhang Y, Tatsuno I, Okada R, et al (2016) Predominant role of *msr(D)* over *mef(A)* in macrolide resistance  
46 in *Streptococcus pyogenes*. *Microbiol (United Kingdom)* 162:46–52.  
47 <https://doi.org/10.1099/MIC.0.000206/CITE/REFWORKS>
- 48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

## CONCLUSÃO CAPÍTULO II

Neste trabalho foram encontrados diversos genes que possibilitaram uma melhor compreensão do comportamento evolutivo de *Gardnerella vaginalis*, o qual foi possível observar a presença de genes com características de virulência e resistência antimicrobiana. Esses achados corroboram com o fenótipo patogênico e resistente dessa espécie que foi mantido ao longo do tempo, a qual pode ser explicada pela grande permanência e recorrência da vaginose bacteriana mesmo após tratamento, tanto pela resistência aos Macrolídeos, a principal linha de tratamento como pelos fatores de virulência, como a formação de biofilme a presença de vaginolisina.

Visto essas resistências a antimicrobianos, foram propostas duas proteínas com funções essenciais para a sobrevivência da bactéria como alvos de drogas utilizando a metodologia de docking molecular, prevendo interações ligante-alvo em nível molecular na utilização de compostos naturais para inibição do sítio ativo desses alvos, melhorando as chances de encontrar candidatos a medicamento contra vaginose bacteriana que é um problema crucial na saúde e bem-estar da mulher.

## 7. CONCLUSÃO GERAL

É perceptível que há uma maior semelhança entre os genomas dentro do mesmo clado, em comparação com genomas de clados diferentes. Ao examinar a história evolutiva e, especialmente, a organização dos seus genes, torna-se evidente um certo grau dessa similaridade. No entanto, uma divergência mais significativa é agora evidente ao analisar a ordem dos genes em todos os genomas ao longo da evolução. Essa compreensão abrangente lança luz sobre a diversidade do genoma.

Além disso, os genomas da espécie demonstraram a seleção positiva de vários genes, especialmente aqueles relacionados à resistência e virulência. Além disso, o pan-resistoma da espécie é expansível, sugerindo que, além dos genes já identificados, ela pode adquirir novos genes ao longo da evolução, o que é crucial para manter a característica patogênica da espécie. Além disso, foram identificadas duas proteínas com potencial para promover a virulência, selecionadas por meio de genômica subtrativa. Dada a resistência observada, foram efetuadas análises de *docking molecular* destas proteínas com compostos naturais para as escolher *in silico* como potenciais alvos de medicamentos, esses dados todos podem contribuir para pesquisas futuras por buscas de terapias vacinais e medicamentosas para o tratamento da doença.

## 8. PERSPECTIVAS

Os capítulos desse trabalho, trazem análises complementares de diversidade gênica seguida de seleção positiva Darwiniana e por fim na identificação de proteínas como potenciais alvos de droga. Será realizada em análises posteriores a dinâmica molecular dessas proteínas, bem como Docking molecular das suas estruturas secundárias para validação da estrutura proteica de cada.

Ademais, será realizado a busca de novos isolados de linhagens de *Gardnerella vaginalis* para um estudo com outros patógenos envolvidos na vaginose bacteriana para seleção de alvos que possam ser utilizados na confecção de um fármaco.



## 9. REFERÊNCIAS BIBLIOGRÁFICAS

- Ågren, J., Sundström, A., Håfström, T., & Segerman, B. (2012). Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLOS ONE*, 7(6), e39107. <https://doi.org/10.1371/JOURNAL.PONE.0039107>
- Ahmed, A., Earl, J., Retchless, A., Hillier, S. L., Rabe, L. K., Cherpes, T. L., Powell, E., Janto, B., Eutsey, R., Luisa Hiller, N., Boissy, R., Dahlgren, M. E., Hall, B. G., William Costerton, J., Christopher Post, J., Hu, F. Z., & Ehrlich, G. D. (2012). Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. *Journal of Bacteriology*, 194(15), 3922–3937. [https://doi.org/10.1128/JB.00056-12/SUPPL\\_FILE/ZJB999091690SO1.PDF](https://doi.org/10.1128/JB.00056-12/SUPPL_FILE/ZJB999091690SO1.PDF)
- Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L., & Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics*, 11(1), 1–17. <https://doi.org/10.1186/1471-2164-11-332/FIGURES/4>
- Ariston de Carvalho Azevedo, V., Silva Almeida, S., Santos, A. R., & Azevedo, V. (2014). *Whole Genome Annotation: In Silico Analysis*. <https://doi.org/10.5772/23724>
- Atlas of protein sequence and structure. Vol. 5, Supplement 3; [edited by] Margaret O. Dayhoff. - University of Edinburgh.* (n.d.). Retrieved February 28, 2024, from [https://discovered.ed.ac.uk/discovery/fulldisplay?vid=44UOE\\_INST:44UOE\\_VU2&mode=advanced&tab=Everything&docid=alma991768313502466&query=sub,exact,Genetics,AND&context=L&lang=en](https://discovered.ed.ac.uk/discovery/fulldisplay?vid=44UOE_INST:44UOE_VU2&mode=advanced&tab=Everything&docid=alma991768313502466&query=sub,exact,Genetics,AND&context=L&lang=en)
- Avery, O. T., Macleod, C. M., & McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF

PNEUMOCOCCAL TYPES INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *Journal of Experimental Medicine*, 79(2), 137–158.  
<https://doi.org/10.1084/JEM.79.2.137>

Barcellos, F. G., Menna, P., Batista, J. S. D. S., & Hungria, M. (2007). Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium (Ensifer) fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Applied and Environmental Microbiology*, 73(8), 2635–2643.  
<https://doi.org/10.1128/AEM.01823-06>

Benevides, L., Burman, S., Martin, R., Robert, V., Thomas, M., Miquel, S., Chain, F., Sokol, H., Bermudez-Humaran, L. G., Morrison, M., Langella, P., Azevedo, V. A., Chatel, J. M., & Soares, S. (2017). New insights into the diversity of the genus *Faecalibacterium*. *Frontiers in Microbiology*, 8(SEP), 300416.  
<https://doi.org/10.3389/FMICB.2017.01790/BIBTEX>

*Boletim Epidemiológico Aids - 2003 | Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis*. (n.d.). Retrieved February 25, 2024, from <https://antigo.aids.gov.br/pt-br/pub/2003/boletim-epidemiologico-aids-2003>

Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences*, 277(1683), 819.  
<https://doi.org/10.1098/RSPB.2009.1679>

Briselden, A. M., Moncla, B. J., Stevens, C. E., & Hillier, S. L. (1992). Sialidases (neuraminidases) in bacterial vaginosis and bacterial vaginosis-associated microflora. *Journal of Clinical Microbiology*, 30(3), 663–666. <https://doi.org/10.1128/JCM.30.3.663-666.1992>

Cai, B., Yang, X., Tuskan, G. A., & Cheng, Z. M. (2011). MicroSyn: A user friendly tool for

- detection of microsynteny in a gene family. *BMC Bioinformatics*, 12(1), 1–12.  
<https://doi.org/10.1186/1471-2105-12-79/FIGURES/7>
- Catlin, B. W. (1992). *Gardnerella vaginalis*: characteristics, clinical considerations, and controversies. *Clinical Microbiology Reviews*, 5(3), 213–237.  
<https://doi.org/10.1128/CMR.5.3.213>
- Cauci, S., Monte, R., Ropele, M., ero, C., Not, T., Quadrifoglio, F., & Menestrina, G. (1993). Pore-forming and haemolytic properties of the *Gardnerella vaginalis* cytolysin. *Molecular Microbiology*, 9(6), 1143–1155. <https://doi.org/10.1111/J.1365-2958.1993.TB01244.X>
- Chan, C. X., & Ragan, M. A. (2013). Next-generation phylogenomics. *Biology Direct*, 8(1), 1–6. <https://doi.org/10.1186/1745-6150-8-3/TABLES/1>
- Chen, S. L., Hung, C. S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R. R., Ozersky, P., Armstrong, J. R., Fulton, R. S., Latreille, J. P., Spieth, J., Hooton, T. M., Mardis, E. R., Hultgren, S. J., & Gordon, J. I. (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15), 5977–5982.  
<https://doi.org/10.1073/PNAS.0600938103>
- Cody, A. J., Bennett, J. S., & Maiden, M. C. J. (2014). Multi-locus sequence typing and the gene-by-gene approach to bacterial classification and analysis of population variation. *Methods in Microbiology*, 41, 201–219. <https://doi.org/10.1016/BS.MIM.2014.06.001>
- Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLOS ONE*, 5(6), e11147.  
<https://doi.org/10.1371/JOURNAL.PONE.0011147>
- De Maayer, P., Chan, W. Y., Rubagotti, E., Venter, S. N., Toth, I. K., Birch, P. R. J., & Coutinho, T. A. (2014). Analysis of the *Pantoea ananatis* pan-genome reveals factors

- underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics*, *15*(1), 1–14. <https://doi.org/10.1186/1471-2164-15-404/FIGURES/5>
- De Oliveira Silveira, A. C., De Mello De Souza, H. A. P. H., & Albini, C. A. (2010). A *Gardnerella vaginalis* e as infecções do trato urinário. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, *46*(4), 295–300. <https://doi.org/10.1590/S1676-24442010000400006>
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics*, *6*(5), 361–375. <https://doi.org/10.1038/NRG1603>
- Di Paolo, N. C., & Shayakhmetov, D. M. (2016). Interleukin 1 $\alpha$  and the inflammatory process. *Nature Immunology*, *17*(8), 906–913. <https://doi.org/10.1038/NI.3503>
- Dobrindt, U., Janke, B., Piechaczek, K., Nagy, G., Ziebuhr, W., Fischer, G., Schierhorn, A., Hecker, M., Blum-Oehler, G., & Hacker, J. (2000). Toxin genes on pathogenicity islands: impact for microbial evolution. *International Journal of Medical Microbiology: IJMM*, *290*(4–5), 307–311. [https://doi.org/10.1016/S1438-4221\(00\)80028-4](https://doi.org/10.1016/S1438-4221(00)80028-4)
- Emes, R. D., & Yang, Z. (2008). Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. *PloS One*, *3*(5). <https://doi.org/10.1371/JOURNAL.PONE.0002295>
- Fethers, K., & Morris, M. (2001). Is bacterial vaginosis a sexually transmitted infection. *Sexually Transmitted Infections*, *77*(5), 390. <https://doi.org/10.1136/STI.77.5.390>
- Figueiredo, H. C. P., Soares, S. C., Pereira, F. L., Dorella, F. A., Carvalho, A. F., Teixeira, J. P., Azevedo, V. A. C., & Leal, C. A. G. (2015). Comparative genome analysis of *Weissella ceti*, an emerging pathogen of farm-raised rainbow trout. *BMC Genomics*, *16*(1), 1–18. <https://doi.org/10.1186/S12864-015-2324-4/TABLES/3>
- Gardner, H. L., & Dukes, C. D. (1954). New etiologic agent in nonspecific bacterial vaginitis. *Science (New York, N.Y.)*, *120*(3125), 853.

<https://doi.org/10.1126/SCIENCE.120.3125.853>

- Gardner, H. L., & Dukes, C. D. (1955). Haemophilus vaginalis vaginitis: A newly defined specific infection previously classified “nonspecific” vaginitis. *American Journal of Obstetrics and Gynecology*, 69(5), 962–976. [https://doi.org/10.1016/0002-9378\(55\)90095-8](https://doi.org/10.1016/0002-9378(55)90095-8)
- Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in Bioinformatics*, 20(6), 1981–1996. <https://doi.org/10.1093/BIB/BBY063>
- Gelber, S. E., Aguilar, J. L., Lewis, K. L. T., & Ratner, A. J. (2008). Functional and phylogenetic characterization of vaginolysin, the human-specific cytolysin from Gardnerella vaginalis. *Journal of Bacteriology*, 190(11), 3896–3903. [https://doi.org/10.1128/JB.01965-07/SUPPL\\_FILE/JB1965-07SUPPLCDCALIGN.DOC](https://doi.org/10.1128/JB.01965-07/SUPPL_FILE/JB1965-07SUPPLCDCALIGN.DOC)
- Giraldo, P. C., de Carvalho, J. B. J., do Amaral, R. L. G., da Silveira Gonçalves, A. K., Eleutério, J., & Guimarães, F. (2012). Identification of immune cells by flow cytometry in vaginal lavages from women with vulvovaginitis and normal microflora. *American Journal of Reproductive Immunology (New York, N.Y. : 1989)*, 67(3), 198–205. <https://doi.org/10.1111/J.1600-0897.2011.01093.X>
- Gómez, L. M., Sammel, M. D., Appleby, D. H., Elovitz, M. A., Baldwin, D. A., Jeffcoat, M. K., Macones, G. A., & Parry, S. (2010). Evidence of a gene-environment interaction that predisposes to spontaneous preterm birth: a role for asymptomatic bacterial vaginosis and DNA variants in genes that control the inflammatory response. *American Journal of Obstetrics and Gynecology*, 202(4), 386.e1-386.e6. <https://doi.org/10.1016/J.AJOG.2010.01.042>
- Gravett, M. G., Nelson, H. P., Derouen, T., Critchlow, C., Eschenbach, D. A., & Holmes, K. K. (1986). Independent Associations of Bacterial Vaginosis and Chlamydia trachomatis

- Infection With Adverse Pregnancy Outcome. *JAMA*, 256(14), 1899–1903.  
<https://doi.org/10.1001/JAMA.1986.03380140069024>
- Greenwood, J. R., & Pickett, M. J. (1980). Transfer of *Haemophilus vaginalis* Gardner and Dukes to a new genus, *Gardnerella*: *G. vaginalis* (Gardner and Dukes) comb. nov. *International Journal of Systematic Bacteriology*, 30(1), 170–178.  
<https://doi.org/10.1099/00207713-30-1-170/CITE/REFWORKS>
- Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., Goodfellow, J., Zaneveld, J. R., McDonald, D. T., Goodrich, J. A., Heath, A. C., Knight, R., & Gordon, J. I. (2011). Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proceedings of the National Academy of Sciences of the United States of America*, 108(SUPPL. 1), 4599–4606.  
<https://doi.org/10.1073/PNAS.1000071108/-/DCSUPPLEMENTAL/SD01.XLSX>
- Hawes, S. E., Hillier, S. L., Benedetti, J., Stevens, C. E., Koutsky, L. A., Wølner-Hanssen, P., & Holmes, K. K. (1996). Hydrogen peroxide-producing lactobacilli and acquisition of vaginal infections. *The Journal of Infectious Diseases*, 174(5), 1058–1063.  
<https://doi.org/10.1093/INFDIS/174.5.1058>
- Hay, P. (2000). Recurrent Bacterial Vaginosis. *Current Infectious Disease Reports*, 2(6), 506–512. <https://doi.org/10.1007/S11908-000-0053-5>
- Hernández, F. (1998). *Gardnerella vaginalis* y *mobiluncus* en la etiología de la vaginosis bacteriana. *Rev. Costarric. Cienc. Méd.*, 107–116.  
[http://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S0253-29481998000100010&lng=pt&nrm=iso&tlng=es](http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0253-29481998000100010&lng=pt&nrm=iso&tlng=es)
- Hert, D. G., Fredlake, C. P., & Barron, A. E. (2008). Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *ELECTROPHORESIS*, 29(23), 4618–4626.



<https://doi.org/10.1002/ELPS.200800456>

- Hill, J. E., & Albert, A. Y. K. (2019). Resolution and Cooccurrence Patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. piovii*, and *G. vaginalis* within the Vaginal Microbiome. *Infection and Immunity*, *87*(12). <https://doi.org/10.1128/IAI.00532-19>
- Hongo, J. A., de Castro, G. M., Cintra, L. C., Zerlotini, A., & Lobo, F. P. (2015). POTION: An end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. *BMC Genomics*, *16*(1), 1–16. <https://doi.org/10.1186/S12864-015-1765-0/FIGURES/3>
- Hossain, T., Kamruzzaman, M., Choudhury, T. Z., Mahmood, H. N., Nabi, A. H. M. N., & Hosen, M. I. (2017). Application of the Subtractive Genomics and Molecular Docking Analysis for the Identification of Novel Putative Drug Targets against *Salmonella enterica* subsp. *enterica* serovar Poona. *BioMed Research International*, *2017*. <https://doi.org/10.1155/2017/3783714>
- Ingianni, A., Petruzzelli, S., Morandotti, G., & Pompei, R. (1997). Genotypic differentiation of *Gardnerella vaginalis* by amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Immunology and Medical Microbiology*, *18*(1), 61–66. <https://doi.org/10.1111/J.1574-695X.1997.TB01028.X>
- Jaiswal, A. K., Tiwari, S., Jamal, S. B., de Castro Oliveira, L., Sales-Campos, H., Andrade-Silva, L. E., Oliveira, C. J. F., Ghosh, P., Barh, D., Azevedo, V., Soares, S. C., Junior, V. R., & da Silva, M. V. (2021). Reverse vaccinology and subtractive genomics approaches for identifying common therapeutics against *Mycobacterium leprae* and *Mycobacterium lepromatosis*. *Journal of Venomous Animals and Toxins Including Tropical Diseases*, *27*. <https://doi.org/10.1590/1678-9199-JVATITD-2020-0027>
- Jaskolski, M., Dauter, Z., & Wlodawer, A. (2014). A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *The FEBS Journal*,

- 281(18), 3985–4009. <https://doi.org/10.1111/FEBS.12796>
- Javed, A., Parvaiz, F., & Manzoor, S. (2019). Bacterial vaginosis: An insight into the prevalence, alternative treatments regimen and it's associated resistance patterns. *Microbial Pathogenesis*, *127*, 21–30. <https://doi.org/10.1016/J.MICPATH.2018.11.046>
- Kenyon, C., Colebunders, R., & Crucitti, T. (2013). The global epidemiology of bacterial vaginosis: a systematic review. *American Journal of Obstetrics and Gynecology*, *209*(6), 505–523. <https://doi.org/10.1016/J.AJOG.2013.05.006>
- Koumans, E. H., Sternberg, M., Bruce, C., McQuillan, G., Kendrick, J., Sutton, M., & Markowitz, L. E. (2007). The prevalence of bacterial vaginosis in the United States, 2001–2004; associations with symptoms, sexual behaviors, and reproductive health. *Sexually Transmitted Diseases*, *34*(11), 864–869. <https://doi.org/10.1097/OLQ.0B013E318074E565>
- Krizova, L., & Nemeč, A. (2010). A 63 kb genomic resistance island found in a multidrug-resistant *Acinetobacter baumannii* isolate of European clone I from 1977. *The Journal of Antimicrobial Chemotherapy*, *65*(9), 1915–1918. <https://doi.org/10.1093/JAC/DKQ223>
- Kumar, S., Filipiński, A. J., Battistuzzi, F. U., Kosakovsky Pond, S. L., & Tamura, K. (2012). Statistics and truth in phylogenomics. *Molecular Biology and Evolution*, *29*(2), 457–472. <https://doi.org/10.1093/MOLBEV/MSR202>
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X)
- Lefébure, T., & Stanhope, M. J. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology*, *8*(5). <https://doi.org/10.1186/GB-2007-8-5-R71>
- Leite, S. R. R. de F., de Amorim, M. M. R., Calábria, W. B., Leite, T. N. de F., de Oliveira, V.

- S., Ferreira Júnior, J. A. A., & Alencarximenes, R. A. (2010). Perfil clínico e microbiológico de mulheres com vaginose bacteriana. *Revista Brasileira de Ginecologia e Obstetrícia*, 32(2), 82–87. <https://doi.org/10.1590/S0100-72032010000200006>
- LEOPOLD, S. (1953). Heretofore undescribed organism isolated from the genitourinary system. *United States Armed Forces Medical Journal*, 4(2), 263–266. <https://pubmed.ncbi.nlm.nih.gov/13015741/>
- Lopes Dos Santos Santiago, G., Deschaght, P., El Aila, N., Kiama, T. N., Verstraelen, H., Jefferson, K. K., Temmerman, M., & Vaneechoutte, M. (2011). Gardnerella vaginalis comprises three distinct genotypes of which only two produce sialidase. *American Journal of Obstetrics and Gynecology*, 204(5), 450.e1-450.e7. <https://doi.org/10.1016/J.AJOG.2010.12.061>
- Ma, J., Zhang, L., Suh, B. B., Raney, B. J., Burhans, R. C., Kent, W. J., Blanchette, M., Haussler, D., & Miller, W. (2006). Reconstructing contiguous regions of an ancestral genome. *Genome Research*, 16(12), 1557. <https://doi.org/10.1101/GR.5383506>
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560–564. <https://doi.org/10.1073/PNAS.74.2.560>
- Medini, D., Donati, C., Tettelin, H., Massignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6), 589–594. <https://doi.org/10.1016/J.GDE.2005.09.006>
- Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular Docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, 7(2), 146. <https://doi.org/10.2174/157340911795677602>
- Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biology Direct*, 4(1), 1–25. <https://doi.org/10.1186/1745-6150-4-13/FIGURES/6>

- Mitchell, C., & Marrazzo, J. (2014). Bacterial vaginosis and the cervicovaginal immune response. *American Journal of Reproductive Immunology (New York, N.Y. : 1989)*, 71(6), 555–563. <https://doi.org/10.1111/AJI.12264>
- Molteni, B., D'Antuono, A., Bandini, P., Sintini, G., Barcellona, E., Agnello, A., & Milani, M. (2004). Efficacy and tolerability of a new chlorhexidine-based vaginal gel in vaginal infections. *Current Medical Research and Opinion*, 20(6), 849–853. <https://doi.org/10.1185/030079904125003692>
- Moody, G. (2004). *Digital code of life: how bioinformatics is revolutionizing science, medicine, and business*. 389. [https://books.google.com/books/about/Digital\\_Code\\_of\\_Life.html?hl=pt-BR&id=Q960CIDzRuIC](https://books.google.com/books/about/Digital_Code_of_Life.html?hl=pt-BR&id=Q960CIDzRuIC)
- Muzzi, A., Masignani, V., & Rappuoli, R. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discovery Today*, 12(11–12), 429–439. <https://doi.org/10.1016/J.DRUDIS.2007.04.008>
- Nakamura, A., Sheppard, K., Yamane, J., Yao, M., Söll, D., & Tanaka, I. (2009). Two distinct regions in *Staphylococcus aureus* GatCAB guarantee accurate tRNA recognition. *Nucleic Acids Research*, 38(2), 672–682. <https://doi.org/10.1093/NAR/GKP955>
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453. [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4)
- November, J. A. (2011). Early biomedical computing and the roots of evidence-based medicine. *IEEE Annals of the History of Computing*, 33(2), 9–23. <https://doi.org/10.1109/MAHC.2011.35>
- Oakley, M. B., & Kimball, G. E. (1949). Punched Card Calculation of Resonance Energies. *The Journal of Chemical Physics*, 17(8), 706–717. <https://doi.org/10.1063/1.1747374>

- Ocaña, K. A. C. S., & Dávila, A. M. R. (2011). Phylogenomics-based reconstruction of protozoan species tree. *Evolutionary Bioinformatics Online*, 7(7), 107–121. <https://doi.org/10.4137/EBO.S6861>
- Ogier, J. C., Calteau, A., Forst, S., Goodrich-Blair, H., Roche, D., Rouy, Z., Suen, G., Zumbihl, R., Givaudan, A., Tailliez, P., Médigue, C., & Gaudriault, S. (2010). Units of plasticity in bacterial genomes: New insight from the comparative genomics of two bacteria interacting with invertebrates, *Photorhabdus* and *Xenorhabdus*. *BMC Genomics*, 11(1), 1–21. <https://doi.org/10.1186/1471-2164-11-568/FIGURES/8>
- Paavonen, J., Mangioni, C., Martin, M. A., & Wajszczuk, C. P. (2000). Vaginal clindamycin and oral metronidazole for bacterial vaginosis: a randomized trial. *Obstetrics and Gynecology*, 96(2), 256–260. [https://doi.org/10.1016/S0029-7844\(00\)00902-9](https://doi.org/10.1016/S0029-7844(00)00902-9)
- Patterson, J. L., Stull-Lane, A., Girerd, P. H., & Jefferson, K. K. (2010). Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of *Gardnerella vaginalis* relative to other bacterial-vaginosis-associated anaerobes. *Microbiology (Reading, England)*, 156(Pt 2), 392–399. <https://doi.org/10.1099/MIC.0.034280-0>
- Paulo, S., Tanaka, A., Jorge Fagundes, L., Catapan, A., Léa Davidson Gotlieb, S., Belda Jr, W., Arnone, M., Soreano, R., & Regina Moraes, F. B. (n.d.). *Epidemiological profile of women with bacterial vaginosis treated at a clinic for sexually transmitted diseases in the city of Sao Paulo, SP* \*.
- Pinzi, L., & Rastelli, G. (2019). Molecular Docking: Shifting Paradigms in Drug Discovery. *International Journal of Molecular Sciences*, 20(18). <https://doi.org/10.3390/IJMS20184331>
- Piot, P. (1991). *Gardnerella, Streptobacillus, Spirillum and Calymmatobacterium* (pp. 483–487). American Society for Microbiology (ASM). <https://research.itg.be/en/publications/gardnerella-streptobacillus-spirillum-and->

calymmatobacterium

- Platz-Christensen, J. J., Mattsby-Baltzer, I., Thomsen, P., & Wiqvist, N. (1993). Endotoxin and interleukin-1 alpha in the cervical mucus and vaginal fluid of pregnant women with bacterial vaginosis. *American Journal of Obstetrics and Gynecology*, *169*(5), 1161–1166. [https://doi.org/10.1016/0002-9378\(93\)90274-M](https://doi.org/10.1016/0002-9378(93)90274-M)
- Pleckaityte, M., Janulaitiene, M., Lasickiene, R., & Zvirbliene, A. (2012). Genetic and biochemical diversity of *Gardnerella vaginalis* strains isolated from women with bacterial vaginosis. *FEMS Immunology & Medical Microbiology*, *65*(1), 69–77. <https://doi.org/10.1111/J.1574-695X.2012.00940.X>
- Plourd, D. M. (1997). Practical Guide to Diagnosing and Treating Vaginitis. *Medscape Women's Health*, *2*(2), 2.
- Rebbapragada, A., Howe, K., Wachih, C., Pettengell, C., Sunderji, S., Huibner, S., Ball, T. B., Plummer, F. A., Jaoko, W., & Kaul, R. (2008). Bacterial vaginosis in HIV-infected women induces reversible alterations in the cervical immune environment. *Journal of Acquired Immune Deficiency Syndromes* (1999), *49*(5), 520–522. <https://doi.org/10.1097/QAI.0B013E318189A7CA>
- Reid, G. (2002). Probiotics for urogenital health. *Nutrition in Clinical Care: An Official Publication of Tufts University*, *5*(1), 3–8. <https://doi.org/10.1046/J.1523-5408.2002.00512.X>
- Reiter, L. T., Potocki, L., Chien, S., Gribskov, M., & Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research*, *11*(6), 1114–1125. <https://doi.org/10.1101/GR.169101>
- Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., de Leon, A., Xiang, H., Gusti, V., Clausen, I. G., Olsen, P. B., Rasmussen, M. D., Andersen, J. T., Jørgensen, P. L., Larsen, T. S., Sorokin, A., Bolotin, A., Lapidus, A., Galleron, N.,



- ... Berka, R. M. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biology*, 5(10), r77. <https://doi.org/10.1186/GB-2004-5-10-R77>
- Rigo, G. V., & Tasca, T. (2020). Vaginitis: Review on Drug Resistance. *Current Drug Targets*, 21(16), 1672–1686. <https://doi.org/10.2174/1389450121666200804112340>
- Rödelsperger, C., & Dieterich, C. (2010). CYNTENATOR: Progressive Gene Order Alignment of 17 Vertebrate Genomes. *PLOS ONE*, 5(1), e8861. <https://doi.org/10.1371/JOURNAL.PONE.0008861>
- Rottini, G., Dobrina, A., Forgiarini, O., Nardon, E., Amirante, G. A., & Patriarca, P. (1990). Identification and partial characterization of a cytolytic toxin produced by *Gardnerella vaginalis*. *Infection and Immunity*, 58(11), 3751. <https://doi.org/10.1128/IAI.58.11.3751-3758.1990>
- Rust, A. G., Mongin, E., & Birney, E. (2002). Genome annotation techniques: New approaches and challenges. *Drug Discovery Today*, 7(11). [https://doi.org/10.1016/S1359-6446\(02\)02289-4](https://doi.org/10.1016/S1359-6446(02)02289-4)
- SANGER, F., & THOMPSON, E. O. (1953). The amino-acid sequence in the glycol chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical Journal*, 53(3), 353–366. <https://doi.org/10.1042/BJ0530353>
- Schellenberg, J. J., Jayaprakash, T. P., Gamage, N. W., Patterson, M. H., Vaneechoutte, M., & Hill, J. E. (2016). *Gardnerella vaginalis* Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. *PloS One*, 11(1). <https://doi.org/10.1371/JOURNAL.PONE.0146510>
- Schwebke, J. R. (2003). Gynecologic consequences of bacterial vaginosis. *Obstetrics and Gynecology Clinics of North America*, 30(4), 685–694. [https://doi.org/10.1016/S0889-8545\(03\)00086-X](https://doi.org/10.1016/S0889-8545(03)00086-X)

- Sheppard, S. K., Jolley, K. A., & Maiden, M. C. J. (2012). A Gene-By-Gene Approach to Bacterial Population Genomics: Whole Genome MLST of *Campylobacter*. *Genes*, *3*(2), 261. <https://doi.org/10.3390/GENES3020261>
- Sivashankari, S., & Shanmughavel, P. (2007). Comparative genomics - A perspective. *Bioinformatics*, *1*(9), 376. <https://doi.org/10.6026/97320630001376>
- Snipen, L., Almøy, T., & Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics*, *10*(1), 1–8. <https://doi.org/10.1186/1471-2164-10-385/TABLES/1>
- Soares, S. C., Silva, A., Trost, E., Blom, J., Ramos, R., Carneiro, A., Ali, A., Santos, A. R., Pinto, A. C., Diniz, C., Barbosa, E. G. V., Dorella, F. A., Aburjaile, F., Rocha, F. S., Nascimento, K. K. F., Guimarães, L. C., Almeida, S., Hassan, S. S., Bakhtiar, S. M., ... Azevedo, V. (2013). The Pan-Genome of the Animal Pathogen *Corynebacterium pseudotuberculosis* Reveals Differences in Genome Plasticity between the Biovar *ovis* and *equi* Strains. *PLOS ONE*, *8*(1), e53818. <https://doi.org/10.1371/JOURNAL.PONE.0053818>
- Sobel, J. D. (2005). What's new in bacterial vaginosis and trichomoniasis? *Infectious Disease Clinics of North America*, *19*(2), 387–406. <https://doi.org/10.1016/J.IDC.2005.03.001>
- Sousa, L. G. V., Pereira, S. A., & Cerca, N. (2023). Fighting polymicrobial biofilms in bacterial vaginosis. *Microbial Biotechnology*, *16*(7), 1423–1437. <https://doi.org/10.1111/1751-7915.14261>
- Soyer, Y., Orsi, R. H., Rodriguez-Rivera, L. D., Sun, Q., & Wiedmann, M. (2009). Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. *BMC Evolutionary Biology*, *9*(1). <https://doi.org/10.1186/1471-2148-9-264>
- Suzuki, H., Lefébure, T., Bitar, P. P., & Stanhope, M. J. (2012). Comparative genomic analysis

- of the genus *Staphylococcus* including *Staphylococcus aureus* and its newly described sister species *Staphylococcus simiae*. *BMC Genomics*, 13(1).  
<https://doi.org/10.1186/1471-2164-13-38>
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science (New York, N.Y.)*, 320(5875), 486–488.  
<https://doi.org/10.1126/SCIENCE.1153917>
- Tapprich, W. E., Reichart, L., Simon, D. M., Duncan, G., McClung, W., Grandgenett, N., & Pauley, M. A. (2021). An instructional definition and assessment rubric for bioinformatics instruction. *Biochemistry and Molecular Biology Education : A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 49(1), 38–45.  
<https://doi.org/10.1002/BMB.21361>
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome.” *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13950.  
<https://doi.org/10.1073/PNAS.0506758102>
- Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11(5), 472–477.  
<https://doi.org/10.1016/J.MIB.2008.09.006>
- The Pan-Genome of the Zoonotic Neglected Pathogen Bartonella henselae Reveals Two Groups with Different Patterns of Adaptation to Hosts.* (n.d.). Retrieved February 28, 2024, from <https://austinpublishinggroup.com/bacteriology/fulltext/bacteriology-v7-id1133.php>

- Tumapa, S., Holden, M. T. G., Vesaratchavest, M., Wuthiekanun, V., Limmathurotsakul, D., Chierakul, W., Feil, E. J., Currie, B. J., Day, N. P. J., Nierman, W. C., & Peacock, S. J. (2008). Burkholderia pseudomallei genome plasticity associated with genomic island variation. *BMC Genomics*, *9*. <https://doi.org/10.1186/1471-2164-9-190>
- Van Esbroeck, M., Vandamme, P., Falsen, E., Vancanneyt, M., Moore, E., Pot, B., Gavini, F., Kersters, K., & Goossens, H. (1996). Polyphasic approach to the classification and identification of Gardnerella vaginalis and unidentified Gardnerella vaginalis-like coryneforms present in bacterial vaginosis. *International Journal of Systematic Bacteriology*, *46*(3), 675–682. <https://doi.org/10.1099/00207713-46-3-675/CITE/REFWORKS>
- Vandepoele, K., Saeys, Y., Simillion, C., Raes, J., & Van de Peer, Y. (2002). The Automatic Detection of Homologous Regions (ADHoRe) and Its Application to Microcolinearity Between Arabidopsis and Rice. *Genome Research*, *12*(11), 1792. <https://doi.org/10.1101/GR.400202>
- Vaneechoutte, M., Guschin, A., Van Simaey, L., Gansemans, Y., Van Nieuwerburgh, F., & Cools, P. (2019). Emended description of Gardnerella vaginalis and description of gardnerella leopoldii sp. Nov., gardnerella piotii sp. nov. and Gardnerella swidsinskii sp. nov., with delineation of 13 genomic species within the genus Gardnerella. *International Journal of Systematic and Evolutionary Microbiology*, *69*(3), 679–687. <https://doi.org/10.1099/IJSEM.0.003200/CITE/REFWORKS>
- Vergara, I. A., & Chen, N. (2009). Using OrthoCluster for the detection of synteny blocks among multiple genomes. *Current Protocols in Bioinformatics*, Chapter 6(SUPPL. 27). <https://doi.org/10.1002/0471250953.BI0610S27>
- Verstraelen, H., Verhelst, R., Roelens, K., & Temmerman, M. (2012). Antiseptics and disinfectants for the treatment of bacterial vaginosis: A systematic review. *BMC Infectious*

*Diseases*, 12. <https://doi.org/10.1186/1471-2334-12-148>

- Viana, M. V. C., Profeta, R., Cerqueira, J. C., Wattam, A. R., Barh, D., Silva, A., & Azevedo, V. (2022). Evidence of episodic positive selection in *Corynebacterium diphtheriae* complex of species and its implementations in identification of drug and vaccine targets. *PeerJ*, 10, e12662. <https://doi.org/10.7717/PEERJ.12662/SUPP-18>
- Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., & Hou, T. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics : PCCP*, 18(18), 12964–12975. <https://doi.org/10.1039/C6CP01555G>
- Wilson, J. (2004). Managing recurrent bacterial vaginosis. *Sexually Transmitted Infections*, 80(1), 8–11. <https://doi.org/10.1136/STI.2002.002733>
- Xu, Y., Bi, C., Wu, G., Wei, S., Dai, X., Yin, T., & Ye, N. (2016). VGSC: A Web-Based Vector Graph Toolkit of Genome Synteny and Collinearity. *BioMed Research International*, 2016. <https://doi.org/10.1155/2016/7823429>
- Zhang, J. (n.d.). *Positive Darwinian Selection in Gene Evolution*.
- Zhang, K., Lu, M., Zhu, X., Wang, K., Jie, X., Li, T., Dong, H., Li, R., Zhang, F., & Gu, L. (2022). Antibiotic resistance and pathogenicity assessment of various *Gardnerella* sp. strains in local China. *Frontiers in Microbiology*, 13, 1009798. <https://doi.org/10.3389/FMICB.2022.1009798/BIBTEX>

## 11. ANEXOS

### 11.1 FIGURAS

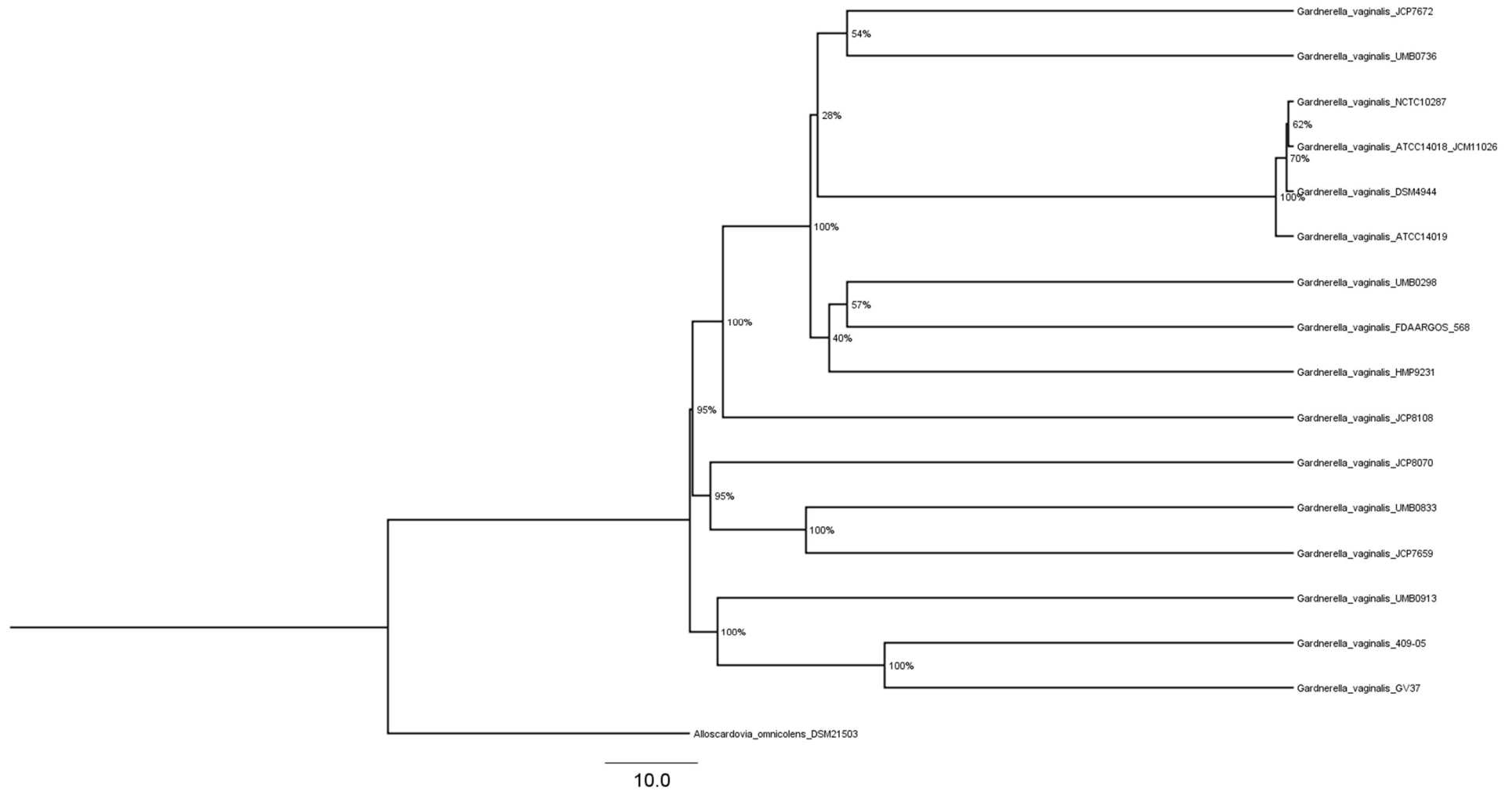
#### CAPÍTULO I

Figuras do artigo de pesquisa intitulado: Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An *In Silico* Approach.

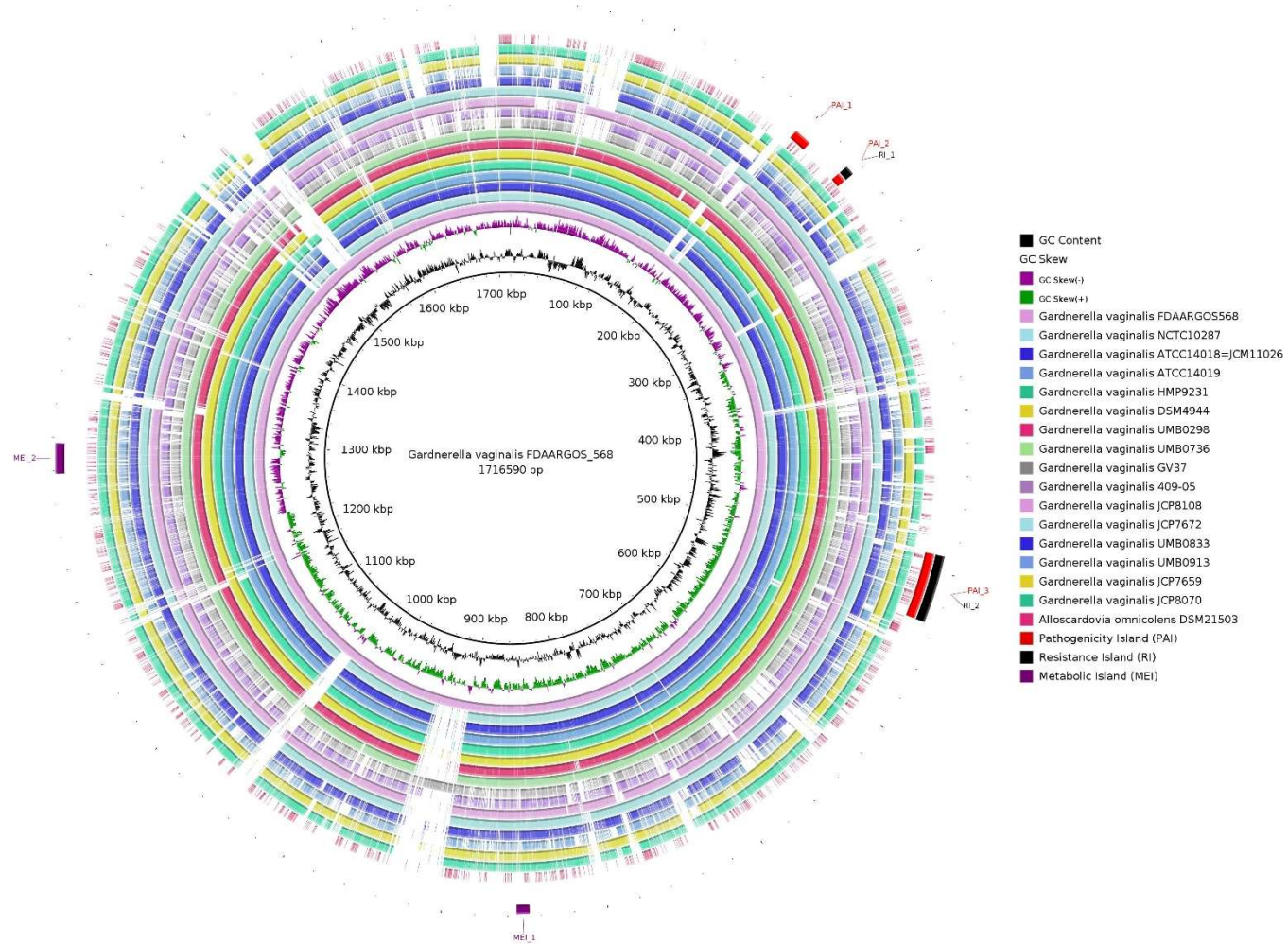
Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1: <i>Alloscardovia_omnicolens_DSM21503</i>	100	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2: <i>Gardnerella_vaginalis_UMB0913</i>	2	100	77	75	17	15	16	17	18	18	18	18	17	18	18	18	18
3: <i>Gardnerella_vaginalis_409-05</i>	1	72	100	88	19	17	17	17	19	20	19	19	18	19	19	19	19
4: <i>Gardnerella_vaginalis_GV37</i>	1	65	82	100	19	18	15	21	23	24	20	18	17	19	19	19	19
5: <i>Gardnerella_vaginalis_UMB0833</i>	1	16	19	19	100	80	72	41	46	46	46	45	43	45	45	45	45
6: <i>Gardnerella_vaginalis_JCP7659</i>	1	15	18	19	84	100	74	42	46	46	45	45	44	45	45	45	45
7: <i>Gardnerella_vaginalis_JCP8070</i>	1	16	18	18	78	77	100	43	46	46	46	46	45	46	46	46	46
8: <i>Gardnerella_vaginalis_JCP8108</i>	1	15	17	19	40	39	38	100	77	75	73	74	71	73	73	73	73
9: <i>Gardnerella_vaginalis_UMB0736</i>	1	16	18	21	43	41	40	74	100	89	87	87	83	86	86	86	86
10: <i>Gardnerella_vaginalis_HMP9231</i>	1	16	19	22	43	41	40	73	90	100	87	87	81	87	87	87	87
11: <i>Gardnerella_vaginalis_FDAARGOS_568</i>	1	16	18	20	44	40	40	72	88	88	100	87	81	87	87	87	87
12: <i>Gardnerella_vaginalis_UMB0298</i>	1	16	18	18	43	41	41	74	90	89	89	100	85	88	88	88	88
13: <i>Gardnerella_vaginalis_JCP7672</i>	1	16	18	18	43	42	42	74	89	87	86	89	100	87	87	87	87
14: <i>Gardnerella_vaginalis_DSM4944</i>	1	17	19	19	44	42	41	74	90	90	90	89	85	100	100	100	100
15: <i>Gardnerella_vaginalis_NCTC10287</i>	1	17	19	19	44	42	42	74	90	90	90	89	85	100	100	100	100
16: <i>Gardnerella_vaginalis_ATCC14019</i>	1	17	19	20	44	42	41	74	90	90	90	89	84	100	100	100	100
17: <i>Gardnerella_vaginalis_ATCC14018_JCM11026</i>	1	17	19	20	44	42	42	74	90	90	90	89	85	100	100	100	100

**Figura 1.** Mapa de calor com 7 genomas completos de *G. vaginalis* e 9 genomas representativos de clados

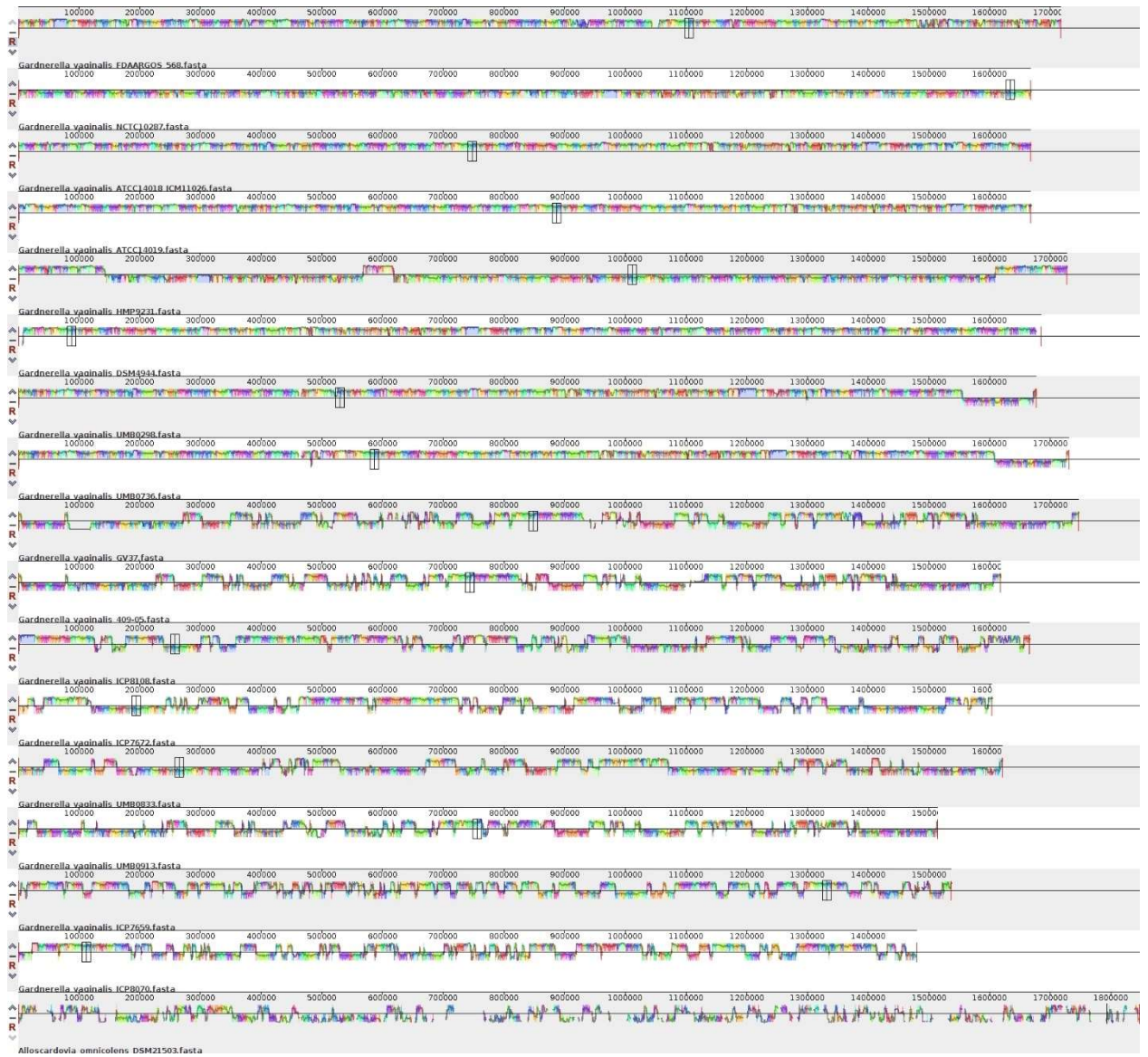




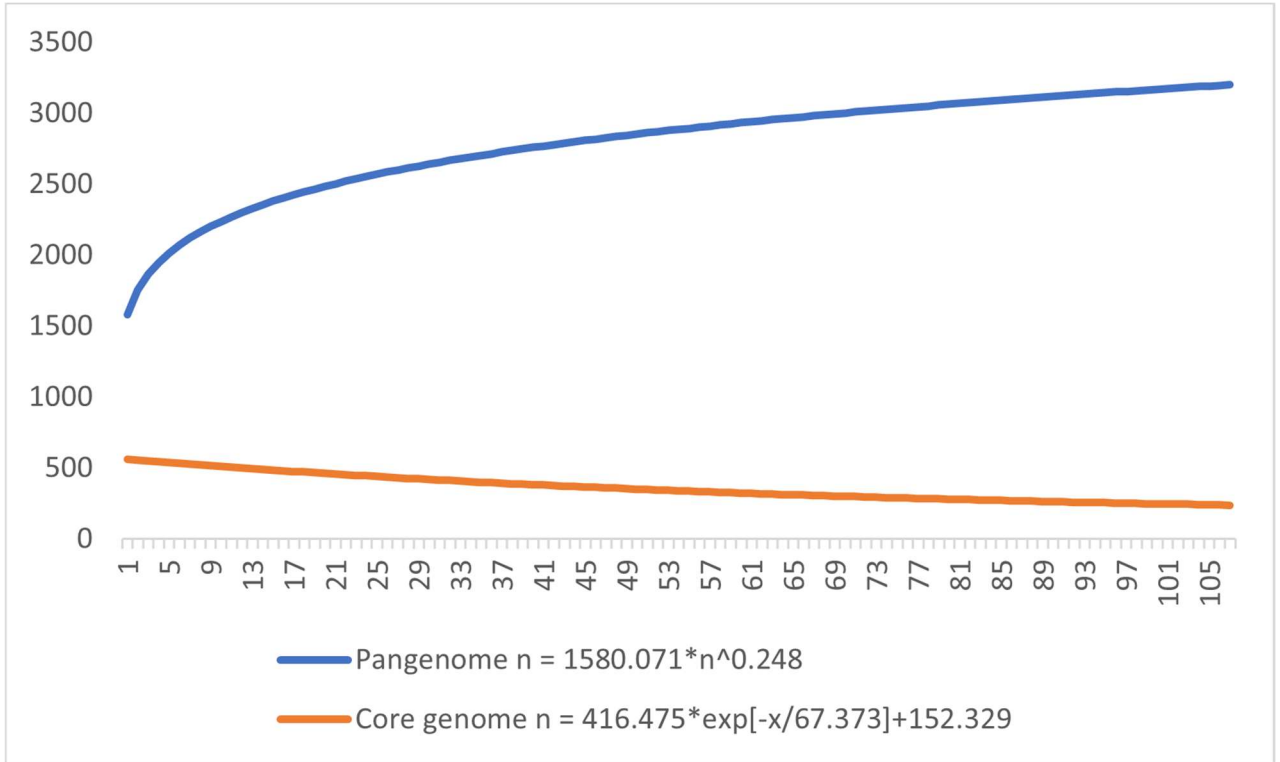
**Figura 2.** Árvore filogenética mostra a percentagem de bootstrap com 1000 bootstraps na árvore.



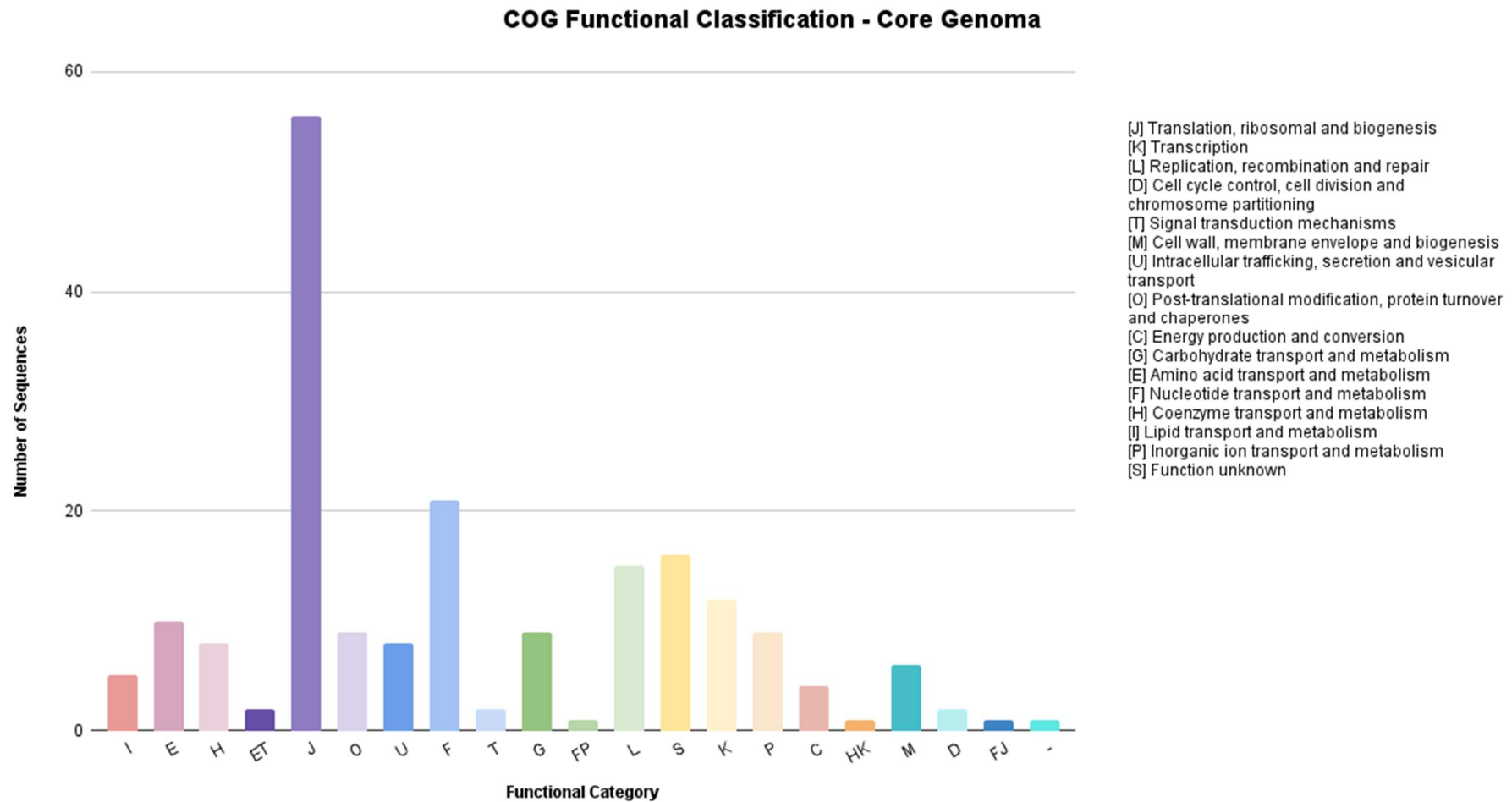
**Figura 3.** Previsão da plasticidade genômica com ilhas de patogenicidade (PAI), ilhas de resistência (RI) e ilhas metabólicas (MI) de *G. vaginalis*



**Figura 4.** Análise de sintenia gênica.

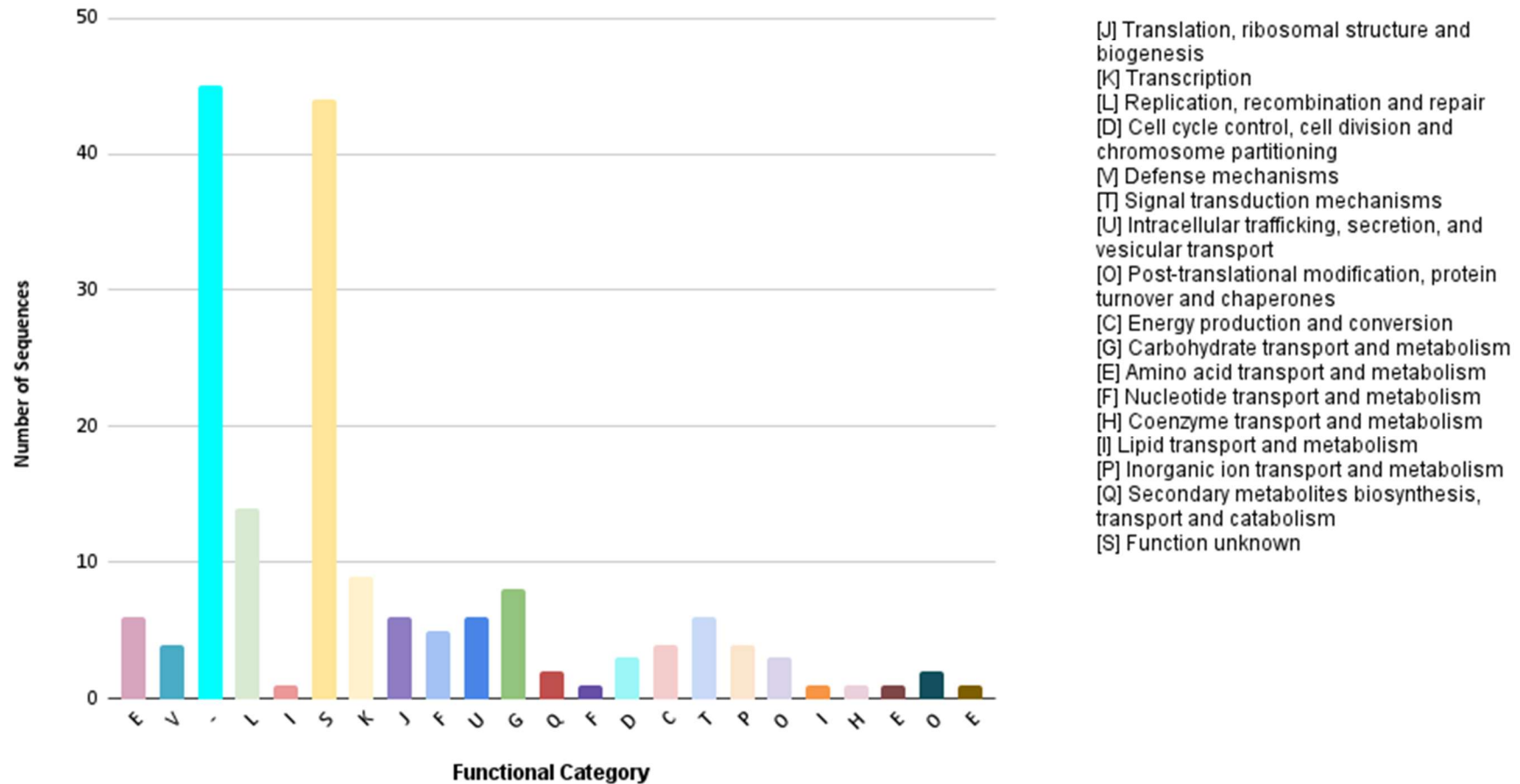


**Figura 5.** Desenvolvimento do pangenoma.



**Figura 6.** Classificação funcional COG do *core* genoma.

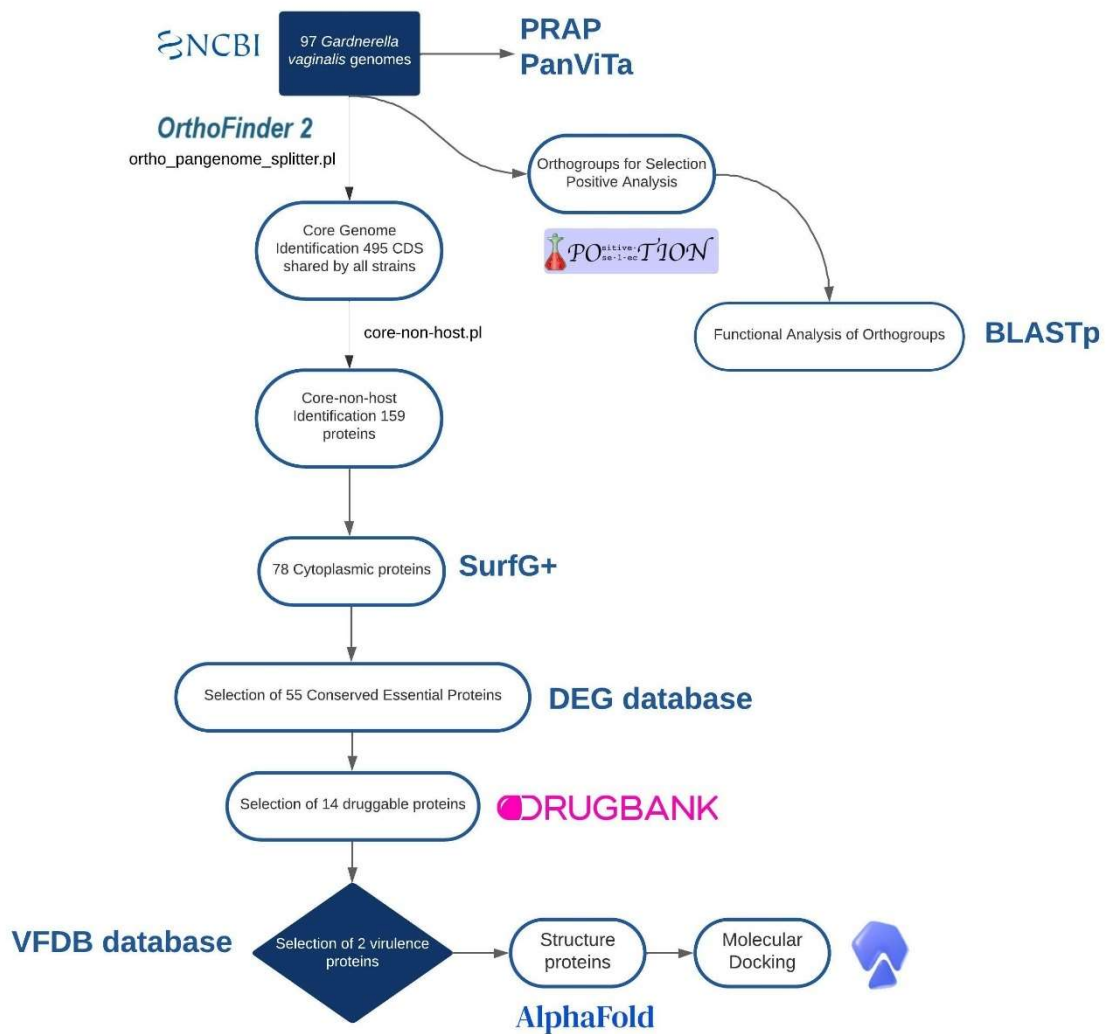
### COG Functional Classification - Shared and Singletons



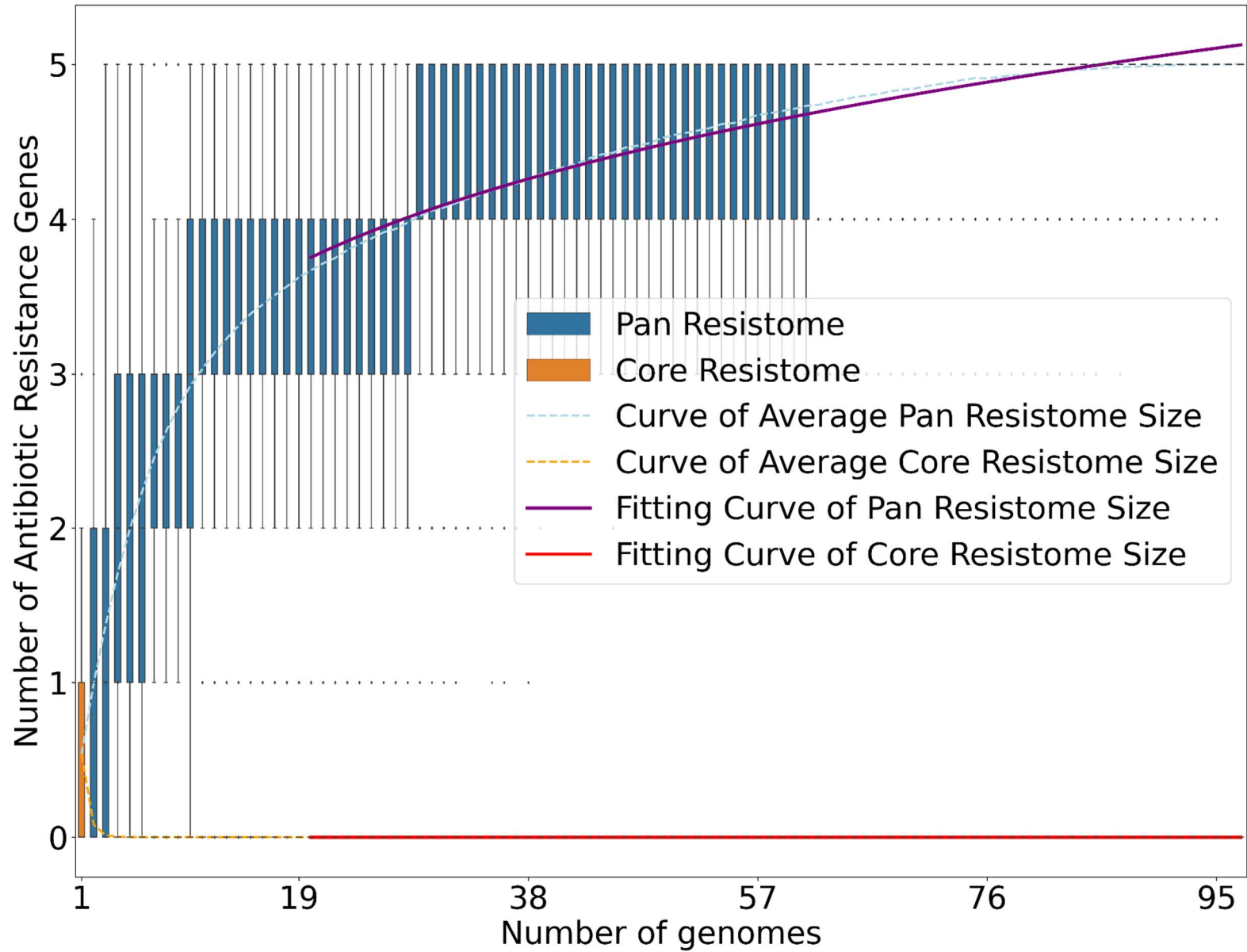
**Figura 7.** Classificação funcional COG dos genes *shared* e *singletons*.



## CAPÍTULO II



**Figura 1.** Fluxo de trabalho das metodologias utilizadas para selecionar os candidatos a medicamentos e os genes relacionados com a seleção positiva Darwiniana.



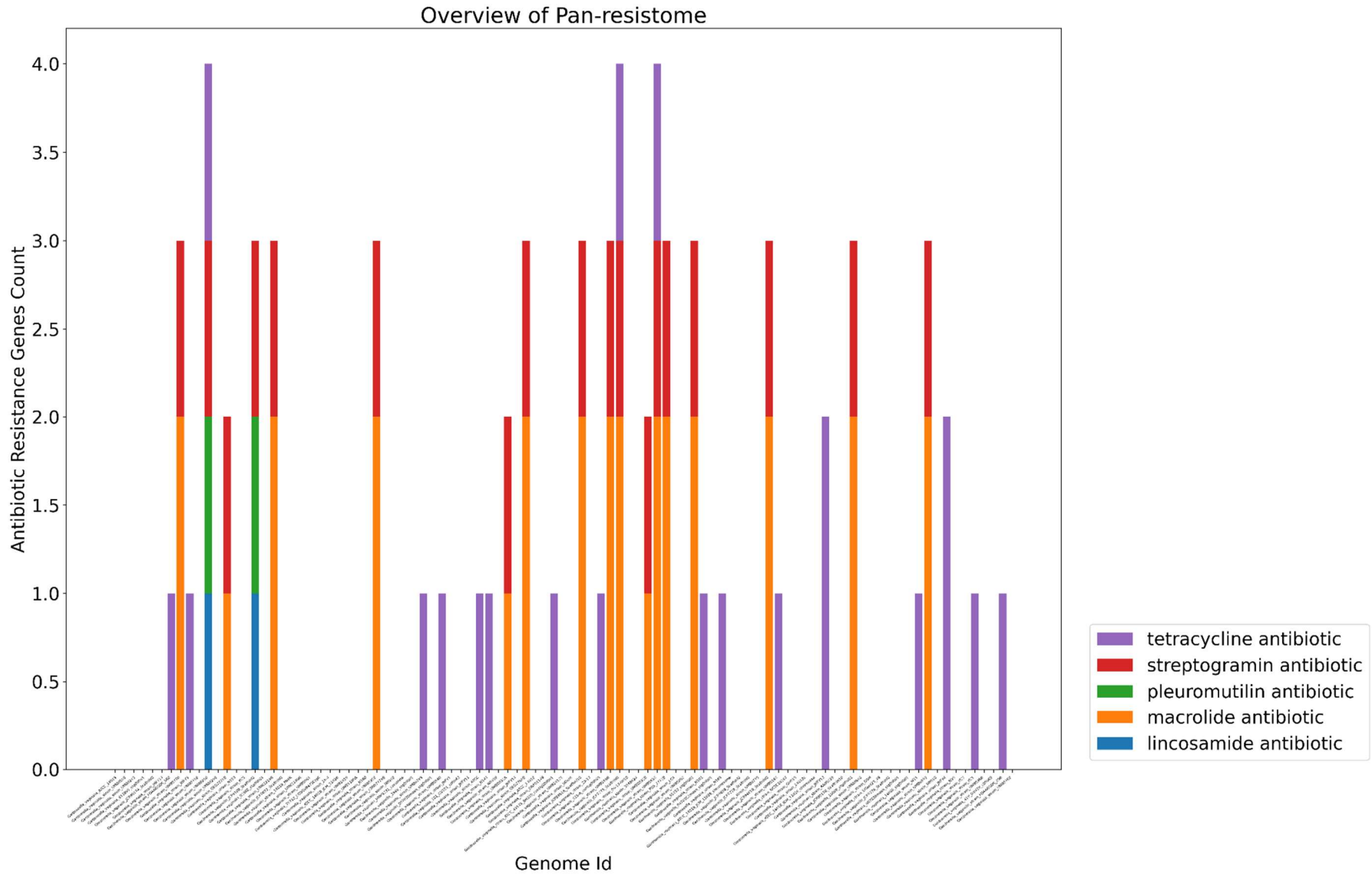
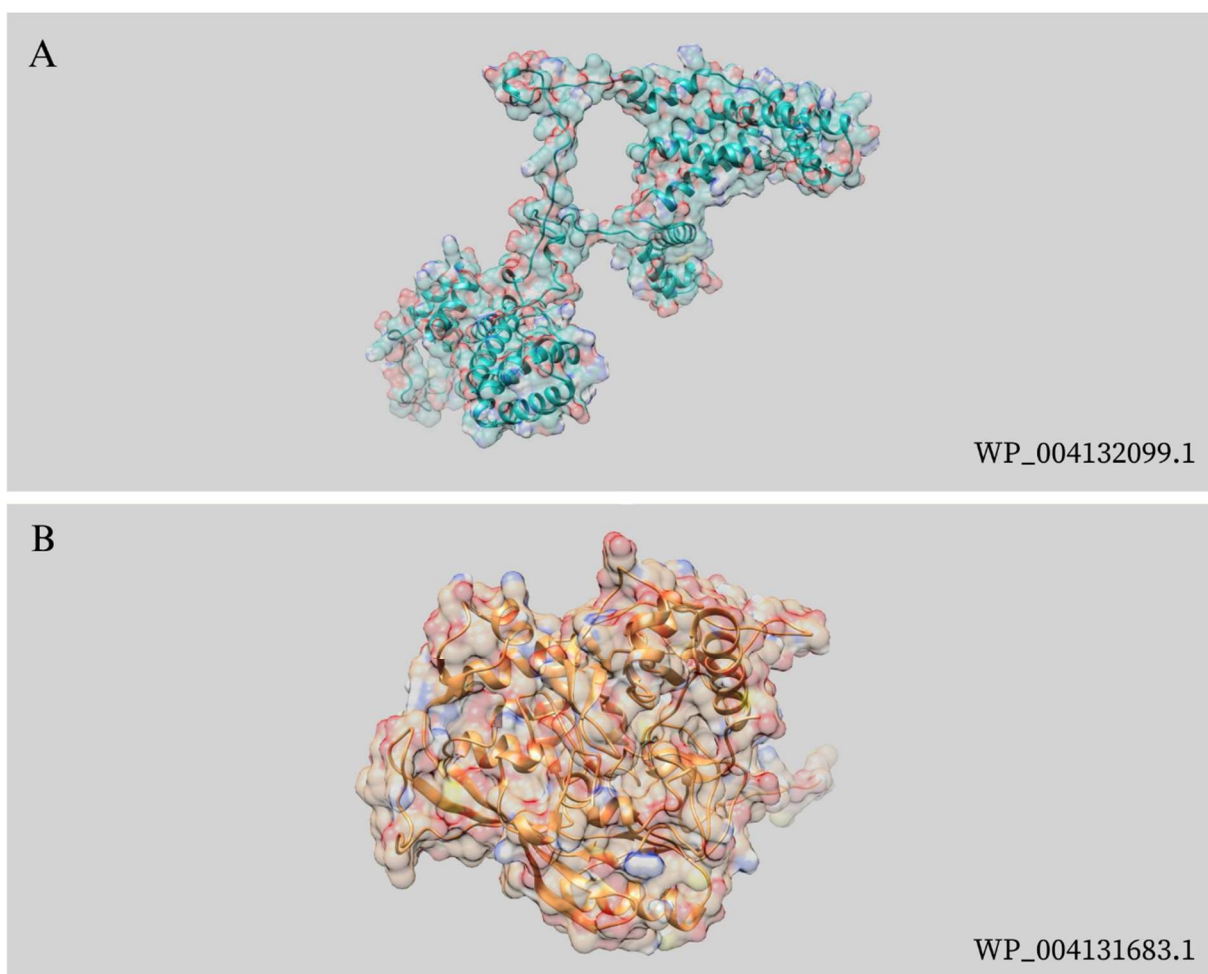
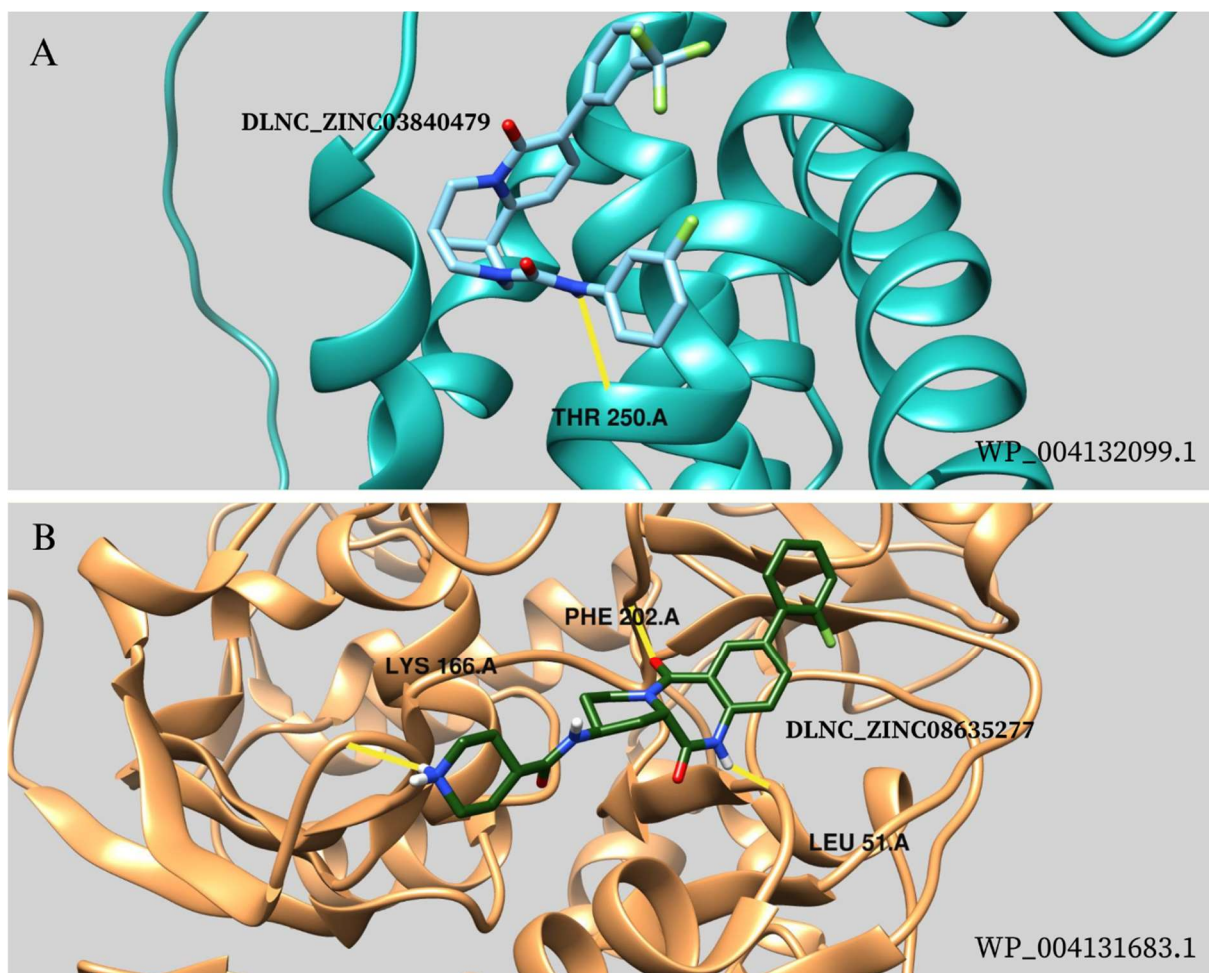


Figura 2. Análise do pan-resistoma e dos genes de resistência utilizando o software PRAP.



**Figura 3.** Estrutura terciária de WP\_004132099.1 e WP\_004131683.1 prevista por Alphafold.



**Figura 4.** Análises moleculares de docking dos novos alvos de medicamentos propostos.

## 11.2 MATERIAL SUPPLEMENTAR

### **CAPÍTULO I**

Material suplementar do artigo de pesquisa intitulado: Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An *In Silico* Approach.

[https://drive.google.com/file/d/1PpRjcAGkvDT8h3Gi2dhXSswL9QZEwotC0/view?usp=drive\\_link](https://drive.google.com/file/d/1PpRjcAGkvDT8h3Gi2dhXSswL9QZEwotC0/view?usp=drive_link)

### **CAPÍTULO II**

Material suplementar do artigo de pesquisa intitulado: Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for *Gardnerella vaginalis*.

[https://drive.google.com/file/d/1J\\_nG21BRRKTQc-HzRooI\\_5zAKV-hddSS/view?usp=sharing](https://drive.google.com/file/d/1J_nG21BRRKTQc-HzRooI_5zAKV-hddSS/view?usp=sharing)



## APÊNDICE

### Produção científica


Produção científica referente às atividades de pesquisa durante o período de 2021 a 2024. Dois artigos submetidos como primeira autora, um publicado com primeira autoria compartilhada e 2 como coautora submetidos. Além disso, foram publicados 3 capítulos de livros como co-autor e 1 submetido como primeira autora, conforme pode ser observado abaixo no currículo lattes resumido e comprovantes.

### Curriculum Lattes (Resumido)

**Eduarda Guimarães Sousa**

### Curriculum Vitae

#### Formação acadêmica/titulação

<b>2022</b>	Mestrado em Programa de Pós-graduação em Genética. Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil  Orientador: Vasco Ariston de Carvalho Azevedo  Co-orientador: Siomar de Castro Soares Bolsista do(a): Fundação de Amparo à Pesquisa do Estado de Minas Gerais
<b>2017 - 2022</b>	Graduação em Biomedicina. Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil Título: VACINOLOGIA REVERSA E GENOMICA SUBTRATIVA PARA PREDIÇÃO DE ALVOS VACINAIS E DE DROGRAS CONTRA STAPHYLOCOCCUS EPIDERMIDIS Orientador: Siomar de Castro Soares

#### Formação complementar

<b>2023 - 2023</b>	Curso de curta duração em Curso de Inverno de Bioinformática. (Carga horária: 40h). Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
<b>2023 - 2023</b>	Curso de curta duração em Introdução à Lógica da Programação e Algoritmos. (Carga horária: 4h). Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
<b>2023 - 2023</b>	Curso de curta duração em Tá tudo sobre controle:

- Versionamento de código com Git e Github. (Carga horária: 4h).  
Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
- 2023 - 2023** Curso de curta duração em Montagem e Anotação de Genomas. (Carga horária: 4h).  
Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
- 2023 - 2023** Curso de curta duração em Workshop on Phenomics. (Carga horária: 15h).  
Instituto de Ciências Agrárias - UFMG, UFMG, Montes Claros, Brasil
- 2022 - 2022** Curso de curta duração em IX CVIP - Curso de Verão em Imunoparasitologia.  
(Carga horária: 12h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2021 - 2021** Curso de curta duração em VIII Curso de Verão em Imunoparasitologia.  
(Carga horária: 12h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2021 - 2021** Curso de curta duração em Aula Aberta Epigenética e Sociedade.  
(Carga horária: 2h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2021 - 2021** Curso de curta duração em Curso de Excel. (Carga horária: 4h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2021 - 2021** Curso de curta duração em Powerpoint e Word. (Carga horária: 4h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2020 - 2020** Curso de curta duração em Fundamentals of Neuroscience,  
Part 1: The electrical properties of the Neur.  
Harvard University, HARVARD, Cambridge, Estados Unidos
- 2019 - 2019** Curso de curta duração em XII Curso de Inverno em Imunologia.  
(Carga horária: 30h).  
Universidade de São Paulo, USP, Sao Paulo, Brasil
- 2019 - 2019** Curso de curta duração em Rotina e desafios de um biomédico perfusionista.  
(Carga horária: 8h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2018 - 2018** Curso de curta duração em I Curso de Anatomia Radiológica Aplicada a Clínica.  
(Carga horária: 45h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil
- 2018 - 2018** Curso de curta duração em Anatomia Clínica e Funcional do Sistema Circulatório.  
(Carga horária: 30h).  
Universidade Federal do Triângulo Mineiro, UFTM, Uberaba, Brasil




## Prêmios e títulos

- 2023** Artigo destacado na página principal da revista Venereology,






- 2022** Menção Honrosa pela apresentação do trabalho VACINOLOGIA REVERSA E GENÔMICA SUBTRATIVA PARA PREDIÇÃO DE ALVOS VACINAIS E DE DROGAS CONTRA STAPHYLOCOCCUS EPIDERMIDIS, AGENTE CAUSADOR DE ENDOCARDITE, UFTM
- 2022** Menção Honrosa pela apresentação e desempenho no XVII Encontro Mineiro de Biomedicina, ocupando o 2º lugar na modalidade Apresentação de Pôster - Graduação, UFTM

## Produção


### Artigos completos publicados em periódicos

1.   FELICE, ANDREI GIACCHETTO; **SOUSA, EDUARDA GUIMARÃES**;  DOMINICI, FABIANA VIEIRA; AZEVEDO, VASCO ARISTON DE CARVALHO; SOARES, SIOMAR DE CASTRO  
Pangenome Analysis Reveals a High Degree of Genetic Diversity in Gardnerella vaginalis: An In Silico Approach. Venereology. , v.2, p.132 - 146, 2023.

### Artigos aceitos para publicação

1. **SOUSA, E. G.**; ARAUJO, M.; PRATES, F.; FONSECA, P.; VIANA, M.; SOARES, S. C.; AZEVEDO, V. A. C.  
Genomic associations analyze for identify novel genetic determinants potentially associated with AMR in C. striatum. Frontiers in Antibiotics. , 2024.
2. ARAUJO, M.; PRATES, F.; RAMOS, J. N.; **SOUSA, E. G.**; BOKERMANN, S.; SACCHI, C. T.; MATTOS-GUARALDI, A. L.; CAMPOS, K. R.; SOUSA, M. A. B.; VIEIRA, V. V.; SANTOS, M. B. N.; CAMARGO, C. H.; SANTANNA, L. O.; SANTOS, L. S.; AZEVEDO, VASCO ARISTON DE CARVALHO  
Infection by multidrug-resistant Corynebacterium diphtheriae strain: whole-genome sequencing, prediction of antimicrobial resistance genes, virulence factors, and CRISPR-Cas system analysis. JOURNAL OF MEDICAL MICROBIOLOGY. , 2024.
3.  **Eduarda Guimarães Sousa**; PRADO, L. C.; FELICE, ANDREI GIACCHETTO; MARQUES, P. H.; JAIWAL, A. K.; OLIVEIRA, C. J. F.; CARNEIRO, A. C. D. M.; SILVA, M. V.; TIWARI, S.; AZEVEDO, V. A. C.; SOARES, S. C.  
Reverse vaccinology and subtractive genomics for vaccine and drug targets prediction against Staphylococcus epidermidis.  
Journal of Biomolecular Structure and Dynamics. , 2024.
4.  **SOUSA, E. G.**; FELICE, A. G.; DOMINICI, FABIANA VIEIRA; JAIWAL, A. K.; PEDROSA, M. L. C.; REIS, L. P.; AZEVEDO, VASCO ARISTON DE CARVALHO; SOARES, S. C.  
Unveiling Resistance and Virulence Mechanisms Under Darwinian Positive Selection for Novel Drug Discovery for Gardnerella vaginalis.  
JOURNAL OF BIOMEDICAL SCIENCE. , 2024.

### Capítulos de livros publicados

1. **SOUSA, E. G.**; GOMES, L. G. R.; PRATES, F.; **JAISWAL, A. K.**; VIANA, M.; GOMES, T. P.; REIS, L. P.; PEDROSA, M. L. C.; VINHAL, A. L. O.; GOMES, G. C.; MENDONCA, B. B. A.; FERREIRA, A.; **SOARES, S. C.**; **AZEVEDO, V. A. C.**  
State of The Art of Immunoinformatics In: State of The Art of Immunoinformatics.1 ed.Portugal: IGI Global, 2024, p. 33-.
2. Vilela Rodrigues, Thaís Cristina; Jaiswal, Arun Kumar; Aburjaile, Flávia Figueira; Almeida, Carlos Augusto; Dias de Oliveira Carvalho, Rodrigo; de Paula, Janaíne Aparecida; Santos, Rodrigo Profeta Silveira; DOMINICI, FABIANA VIEIRA; FELICE, ANDREI GIACCHETTO; Marques, Pedro Henrique; Wu, Michele Min San; Mascarenhas, Ynggrid Victória Cassiano; Junior, Alexandre Claudino; de Jesus Sousa, Thiago; Ceballos, Victor Augusto Sallum; **SOUSA, EDUARDA GUIMARÃES**; Barh, Debmalya; AZEVEDO, VASCO ARISTON DE CARVALHO; Tiwari, Sandeep; SOARES, SIOMAR DE CASTRO  
An Overview of Genomic Islands; Main Features and Computational Prediction: The CMNR Group of Bacteria As a Case Study In: Microbial Genomic Islands in Adaptation and Pathogenicity.1 ed.: Springer Nature Singapore, 2023, p. 33-62.
3. Lemes, Marcela Rezende; FELICE, ANDREI GIACCHETTO; **SOUSA, EDUARDA GUIMARÃES**; Zen, Felipe Lucas; Costa-Madeira, Juliana; Prado, Ligia Carolina Da Silva; Marques, Pedro Henrique; Tiveron, Rafael Destro Rosa; Rodrigues, Thaís Cristina Vilela; Nogueira, Wylerson Guimarães; Sales-Campos, Helioswilton; Silva, Marcos Vinícius; AZEVEDO, VASCO ARISTON DE CARVALHO; SOARES, SIOMAR DE CASTRO; Tiwari, Sandeep  
ABORDAGEM EM IMUNOINFORMÁTICA PARA CARACTERIZAÇÃO E IDENTIFICAÇÃO DE EPÍTOPOS EM BUSCA DE VACINAS In: O Estado da Arte nas Pesquisas Em Vacinologia. 1 ed.: Editora Creative, 2021, p. 17-40.
4.  Lemes, Marcela Rezende; FELICE, ANDREI GIACCHETTO; **SOUSA, EDUARDA GUIMARÃES**; Zen, Felipe Lucas; Prado, Ligia Carolina Da Silva; Marques, Pedro Henrique; Tiveron, Rafael Destro Rosa; Rodrigues, Thaís Cristina Vilela; Nogueira, Wylerson Guimarães; Silva, Marcos Vinícius; AZEVEDO, VASCO ARISTON DE CARVALHO; SOARES, SIOMAR DE CASTRO; Tiwari, Sandeep  
VACINAS MULTIEPÍTOPO USANDO IMUNOINFORMÁTICA EM BACTÉRIAS, VÍRUS, PROTOZOÁRIOS E PARASITOS PATOGÊNICOS  
In: O Estado da Arte nas Pesquisas Em Vacinologia.1 ed.: Editora Creative, 2021, v.1, p. 121-136.

### Trabalhos publicados em anais de eventos (resumo)

1. **Eduarda Guimarães Sousa**; VIANA, M.; DANTAS, C. W. D.; RODRIGUES, D. L. N.; FONSECA, B. B.; ABURJAILE, F. F.; **SOARES, S. C.**; **AZEVEDO, V. A. C.**  
Taxogenomic and probiogenomic analyses for the identification and characterization of species from Bacillus genus In: Brazilian Congress of Genetics, 2023, Ouro Preto.  
**Paleogenomics Sequencing ancient DNA.**, 2023.

## 2. SOUSA, E. G.; OLIVEIRA, C. J. F.; SANTOS, M. M.

Reatividade Humoral de Pacientes com Pênfigo Foliaceus contra proteínas presentes na saliva de carrapatos *Amblyomma cajennense* In: I Congresso Multidisciplinar de Análises Clínicas e Toxicológicas, 2021

**I Congresso Multidisciplinar de Análises Clínicas e Toxicológicas.**

Brazilian Journal of Health and Pharmacy, 2021. v.3.

**Demais produções bibliográficas**1. SOUSA, E. G. **Reatividade Humoral de Pacientes com Pênfigo Foliaceus contra proteínas presentes na saliva de carrapatos**

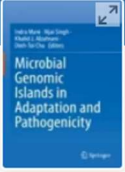
*Amblyomma cajennense*. Congresso. Dvinópolis: Brazilian Journal of Health and Pharmacy (BJHP), 2021. (Outra produção bibliográfica)

**Totais de Produção**

<b>Produção bibliográfica</b>	
Artigos completos publicados em periódico	<b>1</b>
Artigos aceitos para publicação	<b>4</b>
Capítulos de livros publicados	<b>4</b>
Trabalhos publicados em anais de eventos	<b>2</b>
Apresentações de trabalhos (Comunicação)	<b>1</b>
Apresentações de trabalhos (Congresso)	<b>4</b>
Demais produções bibliográficas	<b>1</b>

<b>Eventos</b>	
<b>Participações em eventos (congresso)</b>	<b>7</b>
<b>Participações em eventos (seminário)</b>	<b>2</b>
<b>Participações em eventos (simpósio)</b>	<b>6</b>
<b>Participações em eventos (encontro)</b>	<b>1</b>
<b>Participações em eventos (outra)</b>	<b>8</b>
<b>Organização de evento (congresso)</b>	<b>1</b>
<b>Organização de evento (outro)</b>	<b>2</b>

## Capítulos de Livros



**Microbial Genomic Islands in Adaptation and Pathogenicity** pp 33–62 | [Cite as](#)


Home > [Microbial Genomic Islands in Adaptation and Pathogenicity](#) > Chapter

### An Overview of Genomic Islands' Main Features and Computational Prediction: The CMNR Group of Bacteria As a Case Study

[Thaís Cristina Vilela Rodrigues](#), [Arun Kumar Jaiswal](#), [Flávia Figueira Aburjaile](#), [Carlos Augusto Almeida](#), [Rodrigo Dias de Oliveira Carvalho](#), [Janaine Aparecida de Paula](#), [Rodrigo Profeta Silveira Santos](#), [Fabiana Vieira Dominici](#), [Andrei Giacchetto Felice](#), [Pedro Henrique Marques](#), [Michele Min San Wu](#), [Yngrid Victória Cassiano Mascarenhas](#), [Alexandre Claudino Junior](#), [Thiago de Jesus Sousa](#), [Victor Augusto Sallum Ceballos](#), [Eduarda Guimarães Sousa](#), [Debmalya Barh](#), [Vasco Ariston de Carvalho Azevedo](#), [Sandeep Tiwari](#) & [Siomar de Castro Soares](#)

Chapter | [First Online: 01 April 2023](#)

**Contribuição:** Levantamento bibliográfico, escrita e revisão do manuscrito



**O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA**

ISBN: 978-65-995536-3-9  
<https://doi.org/10.53924/vac1>

**CAPÍTULO 02 - ABORDAGEM EM IMUNOINFORMÁTICA PARA CARACTERIZAÇÃO E IDENTIFICAÇÃO DE EPÍTOPOS EM BUSCA DE VACINAS**

Autores: [Marcela Rezende Lemes](#), [Andrei Giacchetto Felice](#), [Eduarda Guimarães Sousa](#), [Felipe Lucas Zen Juliana Costa-Madeira](#), [Ligia Carolina Da Silva Prado](#), [Pedro Henrique Marques](#), [Rafael Destro Rosa Tiveron](#), [Thaís Cristina Vilela Rodrigues](#), [Wylerson Guimarães Nogueira](#), [Helioswilton Sales-Campos](#), [Marcos Vinícius Da Silva](#), [Vasco Ariston De Carvalho Azevedo](#), [Siomar De Castro Soares](#), [Sandeep Tiwari](#)

<https://doi.org/10.53924/vac1.02> [DOWNLOAD](#) [Fale conosco](#)

**Contribuição:** Levantamento bibliográfico, escrita e revisão do manuscrito

## CAPÍTULO 07

**Vacinas Multiépítipo Usando Imunoinformática Em  
Bactérias, Vírus, Protozoários E Parasitos Patogênicos**

Marcela Rezende Lemes<sup>1</sup>; Andrei Giacchetto Felice<sup>2</sup>; Eduarda Guimarães Sousa<sup>3</sup>  
 Felipe Lucas Zen<sup>3</sup>; Ligia Carolina da Silva Prado<sup>4</sup>; Pedro Henrique Marques<sup>3</sup>  
 Rafael Destro Rosa Tiveron<sup>2</sup>; Thaís Cristina Vilela Rodrigues<sup>5</sup>;  
 Wylerson Guimarães Nogueira<sup>1</sup>; Marcos Vinícius da Silva<sup>6</sup>;  
 Vasco Ariston de Carvalho Azevedo<sup>7</sup>; Siomar de Castro Soares<sup>6</sup>; Sandeep Tiwari<sup>4</sup>

<sup>1</sup> Doutoranda em Bioinformática. Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais - UFMG

<sup>2</sup> Mestrando em Ciências, ênfase em Parasitologia e Imunologia Aplicadas. Programa de Pós-graduação em Medicina Tropical e Infectologia - UFTM

<sup>3</sup> Graduanda do curso de Biomedicina. Universidade Federal do Triângulo Mineiro - UFTM.

<sup>4</sup> Programa Interunidades de Pós-Graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG.

<sup>5</sup> Doutoranda no PPG em Genética, Depto de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais - UFMG

<sup>6</sup> Professor Adjunto do Departamento de Microbiologia, Imunologia e Parasitologia. Universidade Federal do Triângulo Mineiro - UFTM.

<sup>7</sup> Professor Titular do Departamento de Genética, Ecologia e Evolução - Universidade Federal de Minas Gerais - UFMG

**RESUMO**

A imunoinformática, braço da bioinformática, permite a predição e construção de novas vacinas *in silico*, que podem ser produzidas e chegarem à população em geral. A partir do genoma dos organismos de interesse são feitas diferentes conversões e análises até se chegar à proteínas imunogênicas e que são compartilhadas entre diferentes espécies ou linhagens. Essas proteínas são então utilizadas para a produção de vacinas. Modelos vacinais *in silico* contra bactérias, vírus e parasitos já estão disponíveis, demonstrando o grande potencial da imunoinformática. No capítulo são abordados esses organismos com alguns dos avanços já alcançados.

**Palavras-chave:** Imunoinformática. Parasitos. Vírus. Bactérias. Vacinas.





Dear Eduarda Guimarães Sousa,

This email is to confirm that we received your chapter submission, "State of The Art of Immunoinformatics." Your chapter will be evaluated by the book's editor(s) and you will be advised as soon as possible through email of its status, as well as any revisions that may be necessary. Your chapter can be managed from the Chapters page: <https://www.igi-global.com/submission/chapters/>. Please bookmark this page for easy access.

Please kindly note that as an author of a submission under consideration for this publication, we do ask that as a professional courtesy, you participate in the peer review process of other authors' contributions submitted to this publication, just as they will perform a peer review of your work. We appreciate this level of support in ensuring the highest level of quality and integrity for the publication. It is likely that you will receive a chapter manuscript for review within the next 1-2 weeks.

If you have any questions, please contact the book editor, Paulo Fazendeiro, at [fazendeiro@ubi.pt](mailto:fazendeiro@ubi.pt).

**Contribuição:** Levantamento bibliográfico, escrita e revisão do manuscrito

## Artigos científicos

IK

Order Article Reprints



Open Access Article

### **Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An In Silico Approach**

by Andrei Giacchetto Felice <sup>1,\*</sup> , Eduarda Guimarães Sousa <sup>2</sup> ,  
 Fabiana Vieira Dominici <sup>1</sup> , Vasco Ariston de Carvalho Azevedo <sup>2</sup> and  
 Siomar de Castro Soares <sup>1,\*</sup>

<sup>1</sup> Laboratory of Bioinformatics, Institute of Biological and Natural Sciences, Department of Microbiology, Immunology and Parasitology, Federal University of Triângulo Mineiro, Educational Center Campus, Uberaba 38025-180, MG, Brazil

<sup>2</sup> Cellular and Molecular Genetics Laboratory, Institute of Biological Sciences, Department of General Biology, Federal University of Minas Gerais, Pampulha Campus, Belo Horizonte 31270-901, MG, Brazil

\* Authors to whom correspondence should be addressed.

*Venereology* **2023**, 2(4), 132-146; <https://doi.org/10.3390/venereology2040012>

Submission received: 21 August 2023 / Revised: 18 September 2023 / Accepted: 28 September 2023 /  
Published: 30 September 2023

**Acesso:** <https://doi.org/10.3390/venereology2040012>

**Primeira autoria compartilhada.**



Dear Eduarda Guimarães Sousa,






Many thanks for contributing to the Research Topic "[Predicting antimicrobial resistance \(AMR\) profiles from whole-genome sequences of pathogenic bacteria: current strategies, challenges, and future outlook](#)" with your manuscript summary "[Genomic associations analyze for identify novel genetic determinants potentially associated with AMR in \*C. striatum\*](#)":




To move forward, please submit your manuscript no later than 25 March 2024 via the following link [Predicting antimicrobial resistance \(AMR\) profiles from whole-genome sequences of pathogenic bacteria: current strategies, challenges, and future outlook](#).

When submitted, your manuscript will undergo Frontiers collaborative peer-review process, through which authors can directly interact with editors and reviewers in a fully-online, dedicated review forum. If accepted, your article will be published immediately and become freely accessible to all readers. Your paper will be indexed in relevant repositories, and as an author in Frontiers, you retain the copyright to your own papers and figures. You will also have access to advanced article-level metrics to track your impact.


All articles published incur a publishing fee, allowing us to make articles free to read and download. [Find more information on article fees](#).

**Contribuição:** Auxílio na realização do experimento, análise dos dados, escrita e revisão do manuscrito

15:40   99 99 •   4G  29%

←   

Your submission JMM-D-24-00010 has been assigned to an Editor - [EMID:340228996c9192db]

 em.jmm.0.88a546.4447f6fe@edito... 11 de jan.  
 Em nome de JMM  
 Max Roberto Batista de Araújo

Manuscript number: JMM-D-24-00010  
 Title: Infection by multidrug-resistant *Corynebacterium diphtheriae* strain: whole-genome sequencing, prediction of antimicrobial resistance genes, virulence factors, and CRISPR-Cas system analysis  
 Authors: Max Roberto Batista de Araújo; Fernanda Diniz Prates; Juliana Nunes Ramos; Eduarda Guimarães Sousa; Sérgio Bokermann; Cláudio Tavares Sacchi; Ana Luiza Mattos-Guaraldi; Karoline Rodrigues Campos; Mireille Ângela Bernardes Sousa; Verônica Viana Vieira; Marlon Benedito Nascimento Santos; Carlos Henrique Camargo; Lincoln Oliveira Sant'anna; Louisy Sanches Santos; Vasco Azevedo

Dear Dr Batista de Araújo,

Thank you for submitting your manuscript to *Journal of Medical Microbiology*.

Your manuscript has now been assigned to Arnoud H. M. van Vliet, PhD who will assess your article for *Journal of Medical Microbiology*.

← ∨ Responder

**Contribuição:** Escrita e análises do manuscrito.



Dear Eduarda Sousa,

Thank you for your submission.

Submission ID            231401535  
Manuscript Title        **Reverse vaccinology and subtractive genomics for vaccine and drug targets prediction against Staphylococcus epidermidis**  
**Journal**                    **Journal of Biomolecular Structure and Dynamics**

If you made the submission, you can check its progress and make any requested revisions on the [Author Portal](#)

Thank you for submitting your work to our [journal](#).

If you have any queries, please get in touch with [TBSD-peerreview@journals.tandf.co.uk](mailto:TBSD-peerreview@journals.tandf.co.uk).

Kind Regards,

[Journal of Biomolecular Structure and Dynamics](#) Editorial Office

**Primeira autoria**