UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

JANAÍNA CANÁRIO CERQUEIRA

PREDIÇÃO DE ALVOS QUIMIOTERÁPICOS E *DESIGN* DE VACINA MULTI-EPÍTOPO EM *Corynebacterium ulceranS* E *C. silvaticum*

BELO HORIZONTE
2022

JANAÍNA CANÁRIO CERQUEIRA

PREDIÇÃO DE ALVOS QUIMIOTERÁPICOS E DESIGN DE VACINA MULTI-EPÍTOPO EM *Corynebacterium ulcerans* E *C. silvaticum*

Versão final

Dissertação apresentada ao Programa de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Bioinformática.

Orientador: Prof. Dr. Vasco Ariston de Carvalho Azevedo

BELO HORIZONTE
2022

043 Cerqueira, Janaína Canário.
Predição de alvos quimioterápicos e design de vacina multi epítopo em Corynebacterium ulcerans e C. silvaticu [manuscrito] / Janaína Canário Cerqueira. – 2022.
90 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Vasco Ariston de Carvalho Azevedo. Coorientador: Prof. Dr. Sandeep Tiwari.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2 Corynebacterium. 3. Zoonoses. 4. Noxas. 5. Vacinologia. I. Azevedo, Vasco Ariston de Carvalho. II. Tiwari, Sandeep. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

PARECER Nº **22/2022**
PROCESSO Nº 23072.247486/2022-13

FOLHA DE APROVAÇÃO

"Predição de alvos quimioterápicos e design de vacina multi-epítopo para Corynebacterium ulcerans e C. silvaticumPredição de alvos quimioterápicos e design de vacina multi-epítopo para Corynebacterium ulcerans e C. silvaticum"

Janaína Canário Cerqueira

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. Dr. Vasco Ariston de Carvalho Azevedo - Orientador
Universidade Federal de Minas Gerais

Prof. Dr. Sandeep Tiwari - Coorientador
Universidade Federal de Minas Gerais

Prof. Dr. Bruno Silva Andrade
Universidade Estadual do Sudoeste da Bahia

Profa. Dra. Liza Figueiredo Felicori Vilela

Universidade Federal de Minas Gerais

Belo Horizonte, 16 de agosto de 2022.

Documento assinado eletronicamente por **Sandeep Tiwari**, **Usuário Externo**, em 16/08/2022, às 11:36, 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 **Bruno Silva Andrade**, **Usuário Externo**, em 16/08/2022, às 13:21, 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 17/08/2022, às 14:34, 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 **Liza Figueiredo Felicori Vilela**, **Professora do Magistério Superior**, em 21/08/2022, às 20:18, 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, informando o código verificador **1665069** e o código CRC **10AD7FF3**.

Referência: Processo nº 23072.247486/2022-13 SEI nº 1665069

AGRADECIMENTOS

Gostaria de agradecer a todos os que colaboraram com o meu mestrado:

- Ao Prof. Dr. Vasco Ariston de Carvalho Azevedo, pela orientação.
- Ao programa de Programa de Pós-graduação em Bioinformática da UFMG, pelas disciplinas e suporte.
- Ao Laboratório de Genética Celular e Molecular, especialmente Arun, Marcus e Sandeep pelo apoio no desenvolvimento do trabalho.
- As organizações de fomento, especialmente a CAPES pelo suporte financeiro.
- Aos meus pais, e, especialmente ao Marcus que é meu irmão e colega da ciência e de laboratório, que me influenciou, incentivou, apoiou e me ajudou nos momentos de dificuldade e aprendizado.
- Aos amigos/amigas e namorado, pelo incentivo e apoio emocional.

RESUMO

Corynebacterium é um gênero de bactérias gram-positivas que contém espécies de importância médica, veterinária e biotecnológica. Entre as bactérias patogênicas existe um clado contendo seis espécies que podem ser lisogenizadas por corynefagos *tox*$^{+}$ e produzir a toxina diftérica (DT), denominado complexo da *C. diphtheriae*. Entre as espécies do complexo,

C. ulcerans é uma bactéria zoonótica que infecta diversos mamíferos e pode causar diferentes doenças. Atualmente linhagens produtoras de DT são a maior causa de difteria, que vem aumentando em países subdesenvolvidos devido a vacinação negligenciada. Pessoas vacinadas podem ser infectadas e transmitir o patógeno. Além disso, a vacina não protege contra linhagens não produtoras de DT, que são capazes de causar outras doenças. *C. silvaticum* é uma espécie recentemente descrita de importância veterinária que causa linfadenite caseosa em porcos e corças, e possui potencial zoonótico. Pouco se sabe sobre sua distribuição geográfica e espectro de hospedeiros. Neste contexto, o desenvolvimento de uma nova vacina capaz de proteger contra linhagens produtoras e não produtoras de DT das duas espécies teria relevância média eveterinária. Além disso, para os casos de tratamento da infecção, a identificação de novos quimioterápicos seria uma alternativa em caso de resistência aos atuais. Neste trabalho utilizamos estratégias *in silico* para identificar, para as duas espécies, novas proteínas alvo de quimioterápicos e desenvolver novas vacinas. Foram analisados 72 genomas de *C. ulcerans* e 36 de *C. silvaticum*. Por meio da genômica subtrativa foi possível identificar quatro proteínas alvo de quimioterápicos. Utilizando-se *docking* molecular e uma biblioteca de 5.008 compostos naturais, foram identificados três possíveis quimioterápicos para cada uma das proteínas alvo. Por meio da vacinologia reversa foram identificadas nove prováveis proteínas alvo. Uma análise imunoinformática foi realizada para desenhar vacinas multi-epitopos a partir de 25 epítopos em comum entre oito das nove proteínas identificadas anteriormente. Quatro vacinas multi-epítopos foram desenhadas com diferentes adjuvantes, capazes de induzir respostas imunes inatas e adaptativas. As informações geradas poderão reduzir o custo e o tempo para a descoberta de novos antibióticos e para o desenvolvimento de uma vacina para as duas espécies, capaz de proteger contra linhagens não produtoras de DT.

Palavras-chave: alvos de drogas, *Corynebacterium*, doença zoonótica, imunoinformática, patógeno, vacinologia reversa.

ABSTRACT

Corynebacterium is a genus of gram-positive bacteria that contains species of medical, veterinary and biotechnological importance. Among the pathogenic bacteria there is a clade containing six species that can be lysed by corynephages tox+ and produce diphtheria toxin (DT), called the *C. diphtheriae* complex. Among the species of the complex, *C. ulcerans* is a zoonotic bacterium that infects several mammals and can cause different diseases. Currently DT-producing strains are the major cause of diphtheria, which is increasing in underdeveloped countries due to neglected vaccination. Vaccinated people can become infected and transmit the pathogen. In addition, the vaccine does not protect against non-DT-producing strains, which are capable of causing other diseases. *C. silvaticum* is a recently described species of veterinary importance that causes caseous lymphadenitis in pigs and roe deer, and has zoonotic potential. Little is known about its geographic distribution and host spectrum. In this context, the development of a new vaccine capable of protecting against DT-producing and non-DT-producing strains of the two species would have medium and veterinary relevance. In addition, for cases of infection treatment, the identification of new chemotherapeutics would be an alternative in case of resistance to the current ones. In this work we use *in silico* strategies to identify, for the two species, new chemotherapeutic target proteins and to develop new vaccines. The 72 *C. ulcerans* and 36 *C. silvaticum* genomes were analyzed. By means of subtractive genomics it was possible to identify four target proteins of chemotherapeutic agents. Using molecular docking and a library of 5,008 natural compounds, three possible chemotherapeutics were identified for each of the target proteins. Through reverse vaccinology, nine probable target proteins were identified. An immunoinformatic analysis was performed to design multi-epitope vaccines from 25 epitopes in common among eight of the nine previously identified proteins. Four multi-epitope vaccines were designed with different adjuvants, capable of inducing innate and adaptive immune responses. The information generated could reduce the cost and time for the discovery of new antibiotics and for the development of a vaccine for bothspecies, capable of protecting against non-DT-producing strains.

Keywords: drug targets, *Corynebacterium*, zoonotic disease, immunoinformatics, pathogen, reverse vaccinology.

LISTA DE FIGURAS

Capítulo I

Capítulo II

LISTA DE TABELAS

Capítulo I

Capítulo II

LISTA DE ABREVIAÇÕES

ANI *Average Nucleotide Identity*

BLAST *Basic Local Alignment Search tool*

CAMP Christie–Atkins–Munch-Peterson

CDS *Coding Sequence*

CWSS *Cell Wall Sorting Signal*

GO *Gene Ontology*

DT Toxina diftérica

DTPa Difteria, Tétano e Pertussis

EBS Ensaio Bactericida de Soro

GBDP *Genome Blast Distance Phylogeny*

KEGG *Kyoto Encyclopedia of Genes and Genomes*

LGCM Laboratório de Genética Celular e MolecularLCA Linfadeninte caseosa

MALDI-TOF *Matrix-Assisted Laser Desorption Ionization-Time of Flight*

MHC Major Histocompatibility Complex

MLST *Multilocus Sequence Typing*

MS *Mass Spectrometry*

NTTB *Non-Toxigenic but Tox gene-Bearing*

PATRIC *Pathosystems Resource Integration Center*

PBIT *Pipeline Builder for Identification of Targets*

PLD *Phospholipase D* (Fosfolipase D)

RASTtk *Rapid Annotation using Subsystems Technology tool kit*

SNP *Single Nucleotide Polymorphism*

ST *Sequence Type*

TLR Receptor *Toll-Like*

TYGS *Type (Strain) Genome Server*

VFDB *Virulence Factors Database*

VR Vacinologia Reversa

WGS *Whole Genome Sequencing* (Sequenciamento de Genoma Completo)

SUMÁRIO

I. INTRODUÇÃO

I.1 Gênero *Corynebacterium*

O gênero *Corynebacterium* se caracteriza por bactérias aeróbicas bastonetes Gram-positivos e corineformes (claviforme), pertencente a classe Actinobacteria. As corinebactérias possuem alto conteúdo de guanina e citosina (G+C) em seus genomas, a parede celular contém arabinose e galactose e a maioria das espécies contém ácido micólico. *Corynebacterium* possui bactérias fermentadoras e não fermentadoras, as espécies de importância clínica são catalase positivas e imóveis (BERNARD; FUNKE, 2015). A espécie tipo do gênero é *Corynebacterium diphtheriae* (Kruse 1886) Lehmann e Neumann 1896, 350 ("*Bacillus diphtheria*" Kruse in Flügge 1886, 225). O gênero possui 136 espécies descritas e publicadas de acordo com o *List of Prokaryotic names with Standing in Nomenclature* (https://lpsn.dsmz.de/genus/corynebacterium, acesso em junho de 2022).

Cinquenta e quatro espécies possuem relevância médica, e 14 destas são patógenos zoonóticos transmitidos de diversos mamíferos de produção e domésticos (BERNARD, 2012) Seis espécies patogênicas de interesse médico e veterinário fazem parte de um clado que pode ser lisogenizado por corinefagos *tox*$^{+}$ e produzir a toxina diftérica (DT): *C. diphtheriae*, *C. belfantii*, *C. rouxii*, *C. pseudotuberculosis*, *C. ulcerans* e *C. silvaticum* (BADELL et al., 2020; BERNARD; FUNKE, 2015; DANGEL et al., 2020). Este clado foi denominado de complexo *C. diphtheriae* (BADELL et al., 2020).

C. diphtheriae pode causar difteria em humanos devido a capacidade de algumas linhagens de produzir a DT, sendo o primeiro agente etiológico descrito para esta doença. Além disso, pode causar outras doenças que acometem os sistemas respiratório, nervoso e outros tecidos do sistema circulatório (miocárdio), linfático e urinário (rins). A espécie pode infectar equinos, bovinos e gatos domésticos (GUARALDI; HIRATA; AZEVEDO, 2014) . *C. belfantii* (BADELL et al., 2020) e *C. rouxii* (DAZAS et al., 2018) são reclassificações de linhagens previamente consideradas *C. diphtheriae* biovar belfanti. *C. pseudotuberculosis* é o patógeno causador da linfadenite caseosa em caprinos e ovinos, porém infecta outros mamíferos, incluindo humanos, causando diferentes manifestações (GUARALDI; HIRATA; AZEVEDO, 2014). *C. ulcerans* e *C. silvaticum* serão discutidas em seguida.

I.2 Corynebacterium ulcerans

C. ulcerans foi isolada de humano com doença respiratória similar à difteria, e descrita previamente por Gilbert e Stewart em 1927 (RIEGEL et al., 1995). As linhagens de *C. ulcerans* caracterizam-se por serem bastonetes Gram-positivos, pleomórficos, catalase positivos, não móveis e sem esporos, podem apresentar arranjos coccobacilares sem grânulos de fosfato, ter em seu DNA alto conteúdo de G+C. As principais moléculas componentes de sua parede celular são os açúcares galactose e arabinose, e ácidos micólicos. As cepas são capazes de fermentar maltose e glicose, produz urease e DNase, cistinase positivo em ágar de Tinsdale e negativo para o teste de PYZ (pirazinamidase). Reações reversas de Christie–Atkins–Munch-Peterson (CAMP) são observadas em ágar sangue quando inoculadas transversalmente com uma cepa de *Staphylococcus aureus* (BERNARD; FUNKE, 2015). Além da DT, *C. ulcerans* pode produzir a exotoxina Fosfolipase D (PLD), que hidrodroliza a esfingomielina das membranas celulares de mamíferos. Esta toxina auxilia a disseminação das bactérias do sítio inicial da infecção e causa a morte de macrófagos (TAUCH; BURKOVSKI, 2015). A linhagem tipo desta espécie é a NCTC 7910^{T} (RIEGEL et al., 1995).

C. ulcerans infecta humanos e diversos animais. Seu principal reservatório são diversos mamíferos de grande e pequeno porte, incluindo roedores, suínos e primatas. Linhagens que não produzem DT podem causar diversas doenças no sistema respiratório como pneumonia, sinusite, amigdalite, faringite, peritonite e outras relacionadas mais comuns. Zoonoses podem ser transmitidas principalmente de animais de produção, mas também por animais domésticos e selvagens (HACKER et al., 2016) .

Casos de transmissões e infecções por *C. ulcerans* vêm aumentando, principalmente em países subdesenvolvidos, mesmo em populações vacinadas com a DTPa (difteria, tétano e pertussis). Essa vacina é baseada no toxoide da DT, mas não está sendo suficiente para o controle da difteria e não protege contra outras doenças causadas por linhagens não produtoras da toxina diftérica (HACKER et al., 2016).

Corynebacterium ulcerans fez-se, atualmente, um patógeno emergente em humanos tanto em países industrializados quanto em desenvolvimento. Nesses últimos, tem-se considerado um baixo número de notificações de infecções, em consequência de escassa especialização de diagnóstico. Com isso, faz-se necessário aprimoramento do monitoramento global dessas infecções que envolvem testes em humanos e animais previamente contatados com a bactéria para que possa ser identificada e compreendida a forma como a doença é transmitida. Além do exposto, é requerido maior informações sobre fatores de virulência,

mecanismos de colonização em hospedeiros por esse organismo, já que ainda são insuficientes. Em vista disso, estudos básicos e adicionais sobre essas demandas são fundamentais para o entendimento da patogenicidade e consequente infecção dessa espécie (HACKER et al., 2016).

I.3 Corynebacterium silvaticum

Corynebacterium silvaticum foi isolada de abcessos de Linfadenite Caseosa (LCA) de javalis selvagens (*Sus scrofa*), corça (*Capreolus capreolus*), e porcos domésticos (*Sus scrofa domesticus*), e identificada previamente como *C. pseudotuberculosis* ou *C. ulcerans* (CONTZEN et al., 2011; EISENBERG et al., 2014; OLIVEIRA et al., 2014). *C. silvaticum* tem crescimento lento em condições aeróbicas ou microaeróbicas, as colônias são pequenas, cerosas a brancas cremosas, com superfície lisa e discreta β-hemólise após 48 h de incubação em Ensaio Bactericida de Soro (EBS) e crescimeto bom em pH 7 a 8. As linhagens são *tox* positivas mas não expressam a DT, portanto são linhagens portadoras de *tox* mas não toxigênicas (*Non-toxigenic tox gene bearing* ou NTTB) (DANGEL et al., 2020). A espécie pode ser identificada pela filogenia dos genes 16S e *rpoB*, filogenômica, Tipagem de Sequência Multilocus (MLST), conteúdo G+C (54.5%), espectrometria de massa *Matrix-Assisted Laser Desorption Ionization-Time of Flight* (MALDI-TOF MS). A linhagem tipo desta espécie é a KL0182^{T} (DANGEL et al., 2020).

A variedade de hospedeiros da espécie ainda é desconhecida, o que se sabe é que javalis selvagens, porcos domésticos e corças são seus reservatórios (DANGEL et al., 2020). A distribuição geográfica inclui Alemanha (CONTZEN et al., 2011; DANGEL et al., 2019; EISENBERG et al., 2014), Áustria (BUSCH et al., 2019) , Portugal (MÖLLER et al., 2020; OLIVEIRA et al., 2014; VIANA et al., 2020) e Suíça (https://www.ncbi.nlm.nih.gov/biosample/SAMN16955830). *C. silvaticum* pode ser um patógeno humano devido ao seu potencial zoonótico por análise *in silico* (VIANA et al., 2020) e a confirmação de sua toxicidade para células epiteliais humanas *in vitro* (MÖLLER et al., 2021). Essa espécie já é considerada de importância veterinária, pois causa linfadenite em mamíferos.

I.4 Aplicação de análises genômicas para o controle de patógenos

Como discutido anteriormente, *C. ulcerans* e *C. silvaticum* possuem como reservatórios animais domesticados e selvagens, além de relevância médica e veterinária (HACKER et al.,

2016; MÖLLER et al., 2021). A vacina DTPa foi desenhada para proteger contra linhagens produtoras de DT, não oferecendo proteção contra linhagens não produtoras e que são capazes de causar outras doenças (HACKER et al., 2016). Uma vacina mais eficiente poderia utilizar proteínas alvo compartilhadas entre o maior número de linhagens possível, além de oferecer proteção contra as duas espécies. Neste contexto, a análise de genomas contribui para o entendimento da biologia dos patógenos e para propor estratégias de controle (HARDISON, 2003). Com o auxílio de ferramentas computacionais, pode-se realizar predições e simulações experimentais (*in silico*) qualitativa e quantitativamente (NOGUEIRA et al., 2021). As próximas seções abordarão métodos e aplicações da análise de genomas para o controle de patógenos.

I.5 Genômica comparativa

A genômica comparativa implica na comparação de genomas de diferentes organismos por meio de suas sequências de nucleotídeos para inferir função, conservação, filogenia, evolução, níveis de expressão gênica, entre outros (HARDISON, 2003). A comparação de genomas bacterianos de uma mesma espécie revelou uma variação no conteúdo gênico e deu origem ao conceito de pangenoma, caracterizado por um genoma central composto de genes compartilhados por todos os genomas e um genoma dispensável composto de genes ausentes em alguns genomas ou linhagem-específicos (MEDINI et al., 2005). O genoma central contém genes que codificam funções básicas da biologia celular, enquanto o genoma dispensável, ou acessório, pode conter genes associados a adaptação à diferentes ambientes. Um pangenoma é considerado aberto quando o sequenciamento de novos genomas aumenta significativamente o número de novos genes descobertos (TETTELIN et al., 2008).

A genômica comparativa é amplamente usada em estudos de doenças infecciosas, contribuindo para a epidemiologia (origem e propagação) molecular. É vantajosa também, para a identificação de genes de resistencia à antibióticos e identificação de fatores de virulência, consequentemente podendo estipular a possibilidade de um organismo causar doenças (MOUSTAFA et al., 2020).

Neste contexto, a equipe do LGCM tem publicado nos últimos anos estudos que utilizam análises comparativas para elucidar diversidade genômica, filogenômica, mecanismos de patogenicidade e predição de alvos vacinais e terapêuticos para diferentes espécies do gênero *Corynebacterium* (VIANA et al., 2022; GOMES et al., 2022; JAISWAL et al., 2021; VIANA

et al., 2017, 2020; VILELA RODRIGUES et al., 2022).

I.6.1 Genômica Subtrativa

A Genômica Subtrativa é uma abordagem comumente utilizada na identificação de alvos terapêuticos e no desenvolvimento *in silico* de vacinas. Fundamenta-se na filtragem de sequências proteômicas do patógeno baseada nos principais critérios: (i) fator de virulência, (ii) essencialidade na sobrevivência do organismo e (iii) não homologia com o hospedeiro. No fim, resultam os alvos selecionados de vacinas ou drogas, que posteriormente podem ser submetidos a outras análises (BARH et al., 2011). Em pesquisas com bactérias e vírus, o pangenoma é utilizado na abordagem de genômica subtrativa, pois estes organismos são divididos em linhagens, biovares e serovares. Nesto caso, utiliza-se o genoma central no intuito de selecionar alvos em comum em todas a linhagens analisadas (PRADO et al., 2021).

I.6.2 Imunoinformática e Vacinologia Reversa

Imunoinformática é um campo da Bioinformática, cuja interface está entre métodos computacionais e a imunologia experimental. Utiliza-se de ferramentas computacionais para o entendimento de respostas imunes e testar hipóteses experimentais *in silico*. A principal aplicação da Imunoinformática é no *design* de vacinas e modelagem do sistema imunológico (OLI et al., 2020; TOMAR; DE, 2010).

A vacina multi-epítopo é constituída por uma proteína quimérica, resultante da junção de vários epítopos selecionados de proteínas do patógenos, pequenos *linkers* e adjuvantes que são sequências peptídicas para o aumento da imunogenicidade da vacina. Epítopos são pequenos peptídeos presentes em antígenos, com a capacidade de induzir resposta imune pela ligação de anticorpos específicos de seu sítio ativo. A vacina multi-epítopo é testada *in silico* em análises de interação molecular com um Receptor *Toll-Like* (TLR), a fim de checar a adequação da vacina candidata (VILELA RODRIGUES et al., 2022).

As áreas de pesquisa desse campo incluem: predição de epítopos de células B e T, identificação de genes/proteínas virulentos, interação patógeno-hospedeiro, predição de alergia, vacinação *in silico*, *design* de banco de dados imunológicos, *design* e engenharia de terapias e diagnósticos imunológicos, estudo de genes e suas expressões em estados saudáveis e doentes, compreensão de genes relacionados ao sistema imunológico, modelagem de resposta do sistema imunológico e simulação de experimentos de laboratório. Existem pelo menos 14 bancos de

dados imunológicos que permitem as pesquisas nas diversas áreas e aplicações citadas anteriormente (OLI et al., 2020; TOMAR; DE, 2010).

No desenvolvimento de vacinas, a vacinologia convencional embasa-se no isolamento e cultivo do patógeno e posterior identificação dos elementos essenciais para a imunidade. Existem dois tipos principais de vacinas: as que contêm vírus ou bactérias vivos enfraquecidos (atenuados) e as que apresentam esses organismos mortos (inativados). O tipo inativado pode conter o organismo inteiro ou ser uma vacina de subunidades, que são construídas por componentes antigênicos ou que induzem resposta imunológica como proteínas, polissacarídeos e ácidos ribonucleicos (BAGNOLI et al., 2011).

Com o surgimento e o desenvolvimento da genômica, foi possível o surgimento da abordagem de vacinologia reversa (VR). Esse termo foi mencionado pela primeira vez em 2001, o qual consiste na utilização de dados genômicos e análises computacionais para a identificação dos componentes necessários para a geração de imunidade no hospedeiro (BAGNOLI et al., 2011). Basicamente, através dos genomas pode-se utilizar *pipelines* para predizer proteínas dos micro-organismos de interesse, identificar suas propriedades imunogênicas e essenciais para o metabolismo do patógeno, além de analisar e simular a interação de moléculas terapêuticas (existentes em bancos de dados) com essas proteínas.

O primeiro imunizante feito através da vacinologia reversa, foi a vacina meningocócica *Bexsero*^{R} contra a bactéria *Neisseria meningitidis* (serogrupo B) (SERRUTO et al., 2012). A partir de então, a abordagem foi utilizada para muitas outras bactérias patogênicas, além de vírus e fungos. Desde o primeiro protocolo de VR até os dias atuais, avanços em Sequenciamento de Genomas Completos (WGS) contribuiu para ampliação da aplicabilidade de VR na genômica, imunologia, entre outras áreas afins. Além de identificar alvos quimioterápicos para a produção de drogas e vacinas, pode-se também avaliar a diversidade e evolução de antígenos, juntamente com a interação patógeno-hospedeiro. Concomitantemente, vários programas e ferramentas foram desenvolvidos para patógenos e vêm sendo refinados para a identificação e a classificação de moléculas antigênicas e, também, com dados genômicos e imunológicos sobre os hospedeiros (DALSASS et al., 2019; MOXON et al., 2020).

II. OBJETIVOS

II.1 Objetivo geral

O objetivo deste trabalho foi a predição *in silico* de proteínas alvo candidatas para tratamento com quimioterápicos e a predição de vacina multi-epítopo para o controle de *Corynebacterium ulcerans* e *C. silvaticum*.

II.2 Objetivos específicos

Os objetivos específicos foram:

a) Montar e anotar os genomas de *C. ulcerans* e *C. silvaticum* os quais apenas os dados brutos estavam disponíveis publicamente
b) Identificar proteínas alvo e moléculas candidatas a quimioterápicos
c) Identificar proteínas antigênicas candidatas pata o desenvolvimento de vacina
d) Predizer vacinas multi-epítopo para *C. ulcerans* e *C. Silvaticum*

III. CAPÍTULO I

ARTIGO DE PESQUISA

Neste capítulo foi feita uma pesquisa *in silico* sobre as bactérias *Corynebacterium ulcerans* e *C. silvaticum*, através das abordagens de genômica subtrativa e vacinologia reversa, para a identificação de proteínas alvos de quimioterápicos e proteínas antigênicas que possam ser utilizadas no desenvolvimento de vacinas. Além disso, foram propostos compostos naturais como prováveis quimioterápicos para as proteínas drogáveis. Essa é uma nova investigação feita para ambas as bactérias desse gênero, no intuito de propor estratégias para o controle de infecções causadas por elas.

Candidate common targets (vaccine and drug) and potential drug likemolecules against *Corynebacterium ulcerans* and *C. Silvaticum*

Janaína Canário Cerqueira[a], Marcus Vinicius Canário Viana[b], Arun Kumar Jaiswal[a], Siomar Castro Soares[c], Sandeep Tiwari[a], Alice Rebecca Wattam[d], Artur Silva[b], Khalid J.Alzahrani[e,] Murtaza M. Tambuwala[f], Debmalya Barh[ag], **Vasco Azevedo**[a*]

O artigo de pesquisa foi submetido na revista **Journal of Bioscienses** em 07/06/2022.

Candidate common targets (vaccine and drug) and potential drug like molecules against *Corynebacterium ulcerans* and *C. silvaticum*

Janaína Canário Cerqueira[a], Marcus Vinicius Canário Viana[b], Arun Kumar Jaiswal[a],Siomar Castro Soares[c], Sandeep Tiwari[a], Alice Rebecca Wattam[d], Artur Silva[b], Khalid J. Alzahrani[e,] Murtaza M. Tambuwala[f], Debmalya Barh[ag], Vasco Azevedo[a*]

[a]Laboratory of Cellular and Molecular Genetics, Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil;
[b]Laboratory of Bioinformatics, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil;
[c]Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Minas Gerais, Brazil;
[d]Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, USA;
[e]Department of Clinical Laboratories Sciences, College of Applied Medical Sciences, Taif University, Taif 21944,Saudi Arabia;
[f]School of Pharmacy & Pharmaceutical Sciences, Ulster University, Coleraine, County Londonderry, BT52 1SA,Northern Ireland, United Kingdom;
[g]Institute of Integrative Omics and Applied Bio-technology (IIOAB), Nonakuri, Purba Medinipur, West Bengal,India.

Running title: Targets against *C. silvaticum* and *C. ulcerans.*

Corresponding author: Vasco Azevedo, e-mail: vascoariston@gmail.com.

Abstract:

The availability of complete genomes of pathogens enables *in silico* analyses that can be used to develop new control methods, reducing the time, cost, and necessity of pathogen cultivation. In this study, we used bioinformatics' approach to predict common drug and vaccine candidates against the diphtheria toxin producing *Corynebacterium ulcerans* and the recently characterized *C. silvaticum species*, considered as a zoonotic potential. We found four common, non-host homologous, virulent, essential cytoplasmic druggable proteins that belong to metabolic pathways and are involved in the regulation in other essential genes. In addition, the docking analysis showed drug like natural compounds as possible drug candidate against the target proteins. We also identified nine vaccine candidates involved in transport and regulation of permeability of substances important to the cell. We hypothesize that these identified vaccine candidates, drug targets and drug molecules could be considered for prophylaxis and hence, should be subjected to further experimental validations.

Keywords: Drug target, homology modeling, reverse vaccinology, subtractive genomics, zoonotic.

Candidate common targets (vaccine and drug) and potential drug like molecules against *Corynebacterium ulcerans* and *C. silvaticum*

Janaína Canário Cerqueira[a], Marcus Vinicius Canário Viana[b], Arun Kumar Jaiswal[a],Siomar Castro Soares[c], Sandeep Tiwari[a], Alice Rebecca Wattam[d], Artur Silva[b], Khalid J. Alzahrani[e,] Murtaza M. Tambuwala[f], Debmalya Barh[ag], Vasco Azevedo[a*]

[a]Laboratory of Cellular and Molecular Genetics, Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil;

[b]Laboratory of Bioinformatics, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil; [c]Institute of Biological Sciences and Natural Sciences, Federal University ofTriângulo Mineiro, Uberaba, Minas Gerais, Brazil;

[d]Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, USA;

[e]Department of Clinical Laboratories Sciences, College of Applied Medical Sciences, Taif University, Taif 21944, Saudi Arabia;

[f]School of Pharmacy & Pharmaceutical Sciences, Ulster University, Coleraine, County Londonderry, BT52 1SA, Northern Ireland, United Kingdom;

[g]Institute of Integrative Omics and Applied Bio-technology (IIOAB), Nonakuri, Purba Medinipur, West Bengal, India.

1. Introduction

The genus *Corynebacterium* comprise Gram positive, non-motile and sporeless rods, with strains capable of producing diphtheria toxin (DT) or carrying the gene. *C. ulcerans* is a pathogenic bacterium belonging to the same clade as *C. diphtheria* and *C. tuberculosis*. It is a species of economic and social importance, as it is the main current etiological agent of diphtheria in addition to infecting several mammals. Some strains produce DT, others do not produce the toxin yet cause other human diseases (endocarditis, septic arthritis, osteomyelitis and sepsis) (Hacker et al. 2016). *C. silvaticum* was recently described and is closely related to *C. ulcerans*. This bacterium has pig and deer as reservoirs, but its potential host range is still unknown. It is known so far that it is cytotoxic in humans (Möller et al. 2021). Therefore, the bacterium has a zoonotic potential that leads to a potential economic loss in animal production

(Dangel et al. 2020; Viana et al. 2020).

C. ulcerans was selected based on its emergence as a causative agent of diphtheria and the factthat it has a wide range of reservoir species (Tiwari et al. 2008; Hacker et al. 2016), which increases its threat potential. *C. silvaticum* was selected based on its ability to infect production animals and has a, as yet, unknown potential to cause zoonotic infections (Dangel et al. 2020; Viana et al. 2020). The current vaccine for diphtheria uses the toxoid from *C. diphtheriae* (Rappuoli and Malito 2014) and the lack of sequence diversity could lower the efficiency of the

current vaccine and antitoxin against *C. ulcerans tox*$^{+}$ strains (Otsuji et al. 2019) and could play arole in diphtheria cases in vaccinated individuals (Vandentorren et al. 2014). Other factors that impede the control of this disease is that immunization does not last for a lifetime and decreases with time, and has the possibility of immunized people being colonized and transmitting the disease (Truelove et al. 2020). Additionally, non-toxigenic strains of *C. ulcerans* can cause severeinfections such as septic arthritis, endocarditis, osteomyelitis and sepsis (Zasada 2013; Grosse- Kock et al. 2017).

Genomic data can directly assist in the development of control methods. Reverse vaccinology is the prediction of target proteins for vaccine development based on genomic sequencing data, decreasing the time and cost required and the need for cultivation of the pathogen (Rappuoli 2000; Rappuoli et al. 2016). Examples of products developed from this approach are the vaccines*Bexsero* for *Neisseria meningitidis* (Christodoulides and Heckels 2017) and *Engerix-B* for Hepatitis B (Keating and Noble 2003). Recently, vaccine targets have been predicted *in silico* andtested *in vitro* or *in vivo* for *N. meningitidis* (Masignani et al. 2019), *Shigella* (Hajialibeigi et al. 2021; de Alwis et al. 2021), *Leishmania mexicana* (Burgos-Reyes et al. 2021), and *Rhipicephalusbursa* (Couto et al. 2021). Drug targets and their respective drugs can also be predicted by a *in silico* approach (Das et al. 2021).

In this work, we applied a standard computational approach to identify common vaccine and drug targets and targeting drug like molecules against *Corynebacterium ulcerans* and *C. silvaticum*.

2. Materials and method

2.1. Samples

Genome data of 108 strains of were retrieved from GenBank, 72 of *C. ulcerans* and 36 of *C. silvaticum* (Supplementary Table S1). Of the 108 genomes, 76 were available as sequencing

reads from the Sequence Read Archive (Supplementary Table S1, SRR accession only). They were assembled in PATRIC (Davis et al. 2020) using Unicycler (Wick et al. 2017). All the genomes were annotated in PATRIC using RASTtk (Brettin et al. 2015). The taxonomy of the samples had previously been verified as *C. ulcerans* or *C. silvaticum* (Viana et al. 2020) using FastANI (Jain et al. 2018), and were additionally submitted to the Type Strain Genome Server (Meier-Kolthoff and Göker 2019) to confirm their taxonomy.

2.2. Identification of new targets

2.2.1. Identification of core, non-homologous to human host, essential genes, and cell localization

The core proteome was obtained using OrthoFinder v2.3.11 (Emms and Kelly 2015). Of these,proteins not homologous to the host were identified by BLASTp (Camacho et al. 2009) against the human proteome (GenBank accession GRCh38.p13). Furthermore, the dataset of core and non-host homologous proteins were used for essential protein identification by the PipelineBuilder for Identification of Targets (PBIT) (Shende et al. 2017). Finally, the cellular location of the proteins was identified using the SurfG+ v1.2.1 (Barinov et al. 2009). Putative Surface Exposed (PSE), secreted and membrane proteins were screened for probable vaccine targets, while cytoplasmatic proteins were screened for probable drug targets.

2.2.2. Characterization of vaccine targets

For secreted PSE, and membrane proteins, the Vaxign2 web-based server (He et al. 2010) was used to identify likely vaccines targets using the following criteria: i) epitope binding to the Major Histocompatibility Complex I and II (MHC I and MHC II) and; ii) probability of being an adhesin greater than 0.51. For virulence prediction, the server Virulence Factor Database (VFDB) was used (CHEN *et al.*, 2005). These genes were reannotated for function using eggNOG-mapper(Huerta-Cepas et al. 2017). Genes in genomic islands (GI), specifically in pathogenic islands (PAIs), were predicted using GIPSy (Soares et al. 2016) with *C. glutamicum* 13032 (CP025533.1) as non-pathogenic reference.

2.2.3. Identification of drug targets and docking analysis

For the 84 cytoplasmic proteins, 3D structures were predicted through the MHOLline pipeline (http://www.mholline2.lncc.br/). Genes with predicted structures classified as G2 (e-

value ≤ $10e^{-5}$ and Identity ≥ 0.25 and Length Variation Index ≤ 0.7) were selected for the following steps. The Ramachandran plot of the structures was generated PROCHECK (Laskowski et al. 1993) implemented in SAVES server v6.0 (https://saves.mbi.ucla.edu/). Furthermore, the target proteinswere prioritized according to function and metabolic pathway using eggNOG-mapper, molecular weight using Uniprot (Wasmuth et al. 2017), virulence using (VFDB) (CHEN *et al.*, 2005), and genomic islands using GIPSy.

The residues with high drug scores were identified using DoGSiteScorer (Volkamer et al. 2012). InterProScan online software was used to identify the domains of each of these proteins (JONES *et al.*, 2014). Moreover, the DoGSiteScorer pocket with the highest drug score and the residues of the protein domain were used for the preparation of targets for docking using AutoDockTools (Morris et al. 2009). To identify drugs for these target proteins, a library of 5,008 drug-like natural compounds were obtained from the ZINC database (http://zinc.docking.org/). The docking analysis was made using AutoDock Vina (Trott and Olson 2010). The top 10 compounds were analysed for interactions between their residues interaction and the targetproteins using Chimera (Pettersen et al. 2004).

2.3. Workflow

The methods used here and the total number of proteins in each step are briefed in Figure 1.

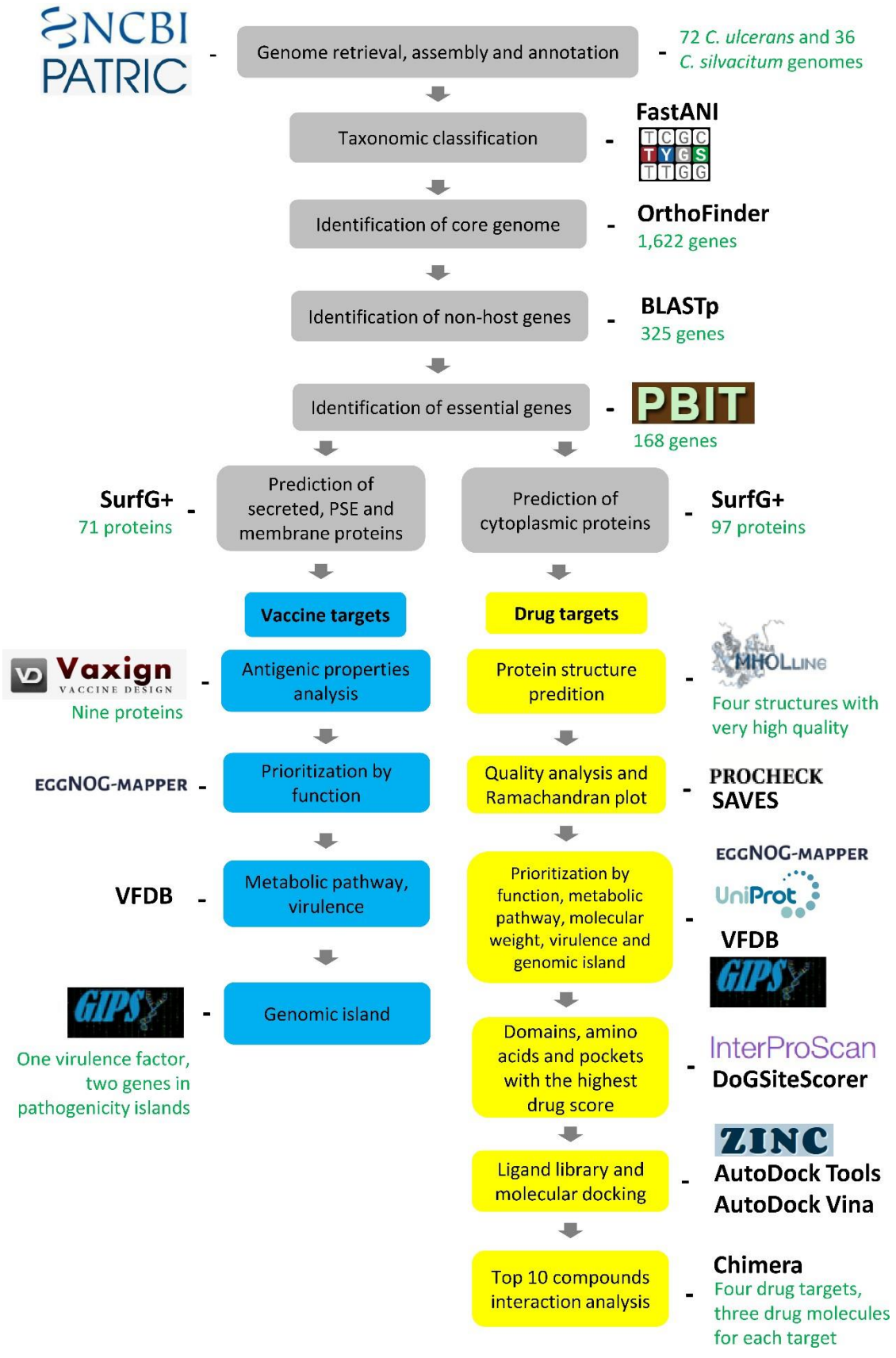


Figure 1. Workflow of the methodologies and the number of genes identified and selected ateach step.

3. Results

3.1. Nine candidate *vaccine targets are common to C. silvaticum and C. ulcerans*

1,622 proteins were identified as the core proteome, and of these, 325 were characterized as non-host proteins, 168 as essential, 71 as PSE or secreted and 97 as cytoplasmatic proteins. Within the PSE, Vaxign2 predicted nine vaccine targets with adhesin probability > 0.51 and binding epitopes for MHC I and II (Figure 1, Table 1). Two of these were in pathogenicity islands. An ABC transporter, substrate-binding protein (*mntA*) was found in PAI 3, and a phosphate ABC transporter, substrate-binding protein PstS (TC 3.A.1.7.1) (*pstS*) was found in PAI 6 (Figure 2, Supplementary Table S2).

Table 1. Vaccine targets candidates for *Corynebacterium ulcerans* and *C. silvaticum.*

Product (Gene)	Adhesin probability	MHC class	Virulence factor (GI)[a]	PATRIC ID
ABC transporter, substrate- binding protein (cluster 14, Mn/Zn)	0.630	I and II	No	996634.5.peg.167
ABC transporter, substrate- binding protein (cluster 14, Mn/Zn) (*mntA*)	0.602	I and II	Yes (PAI 3)[b]	996634.5.peg.623
Glutamate ABC transporter, permease protein 1 GluC (*gluC*)	0.555	I and II	No	996634.5.peg.835
Integral membrane protein	0.566	I and II	No	996634.5.peg.316
Hypothetical protein	0.512	I and II	No	996634.5.peg.652
Hypothetical protein	0.586	I and II	No	996634.5.peg.823
Hypothetical protein	0.661	I and II	No	996634.5.peg.1276
Phosphate ABC transporter, substrate-binding protein PstS (TC 3.A.1.7.1) (*pstS*)	0.696	I and II	No (PAI 6)	996634.5.peg.1997
Phospholipase / thioesterase	0.571	I and II	No	996634.5.peg.1106

[a] *genomic island*
[b] *pathogenicity island*

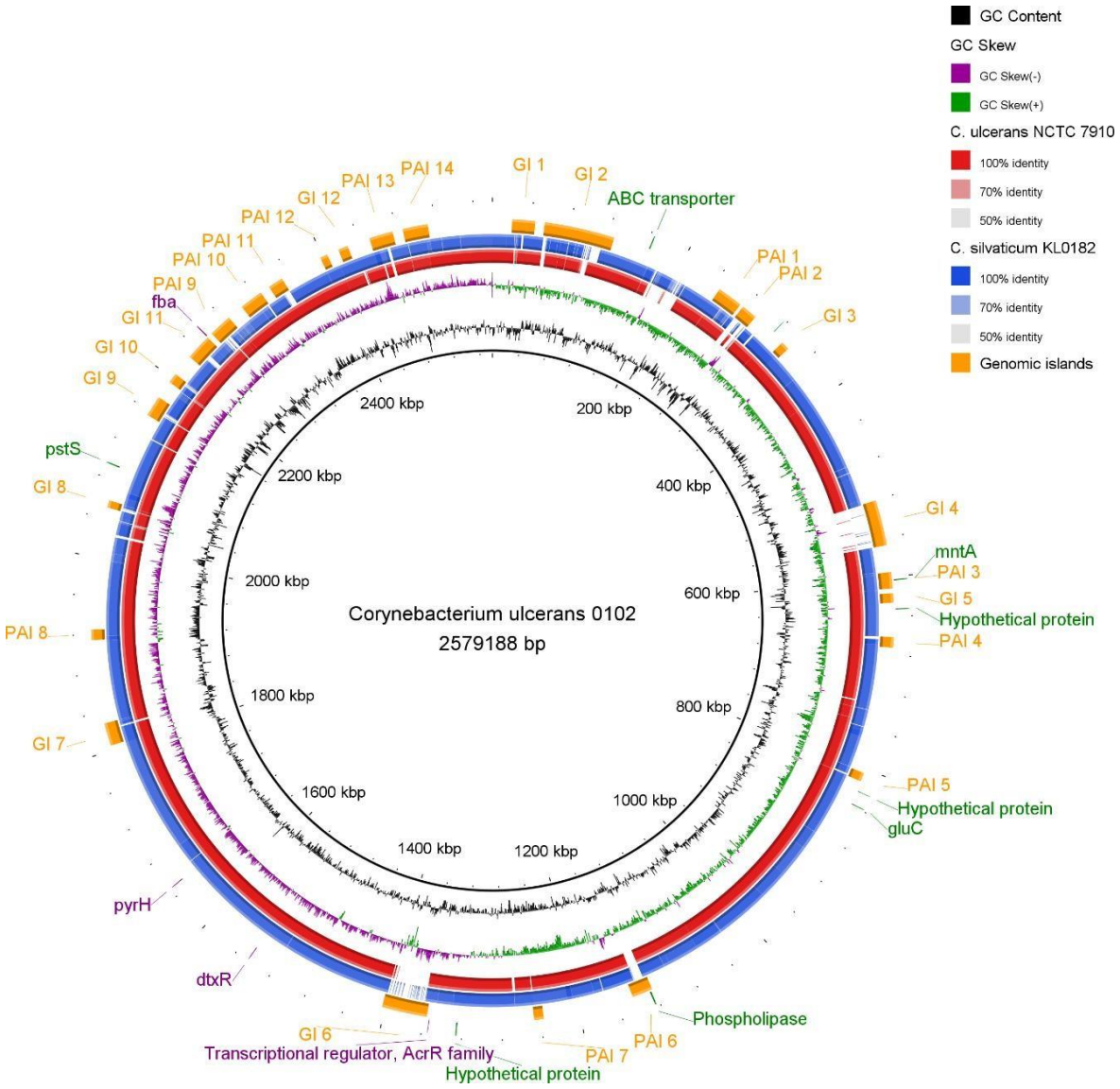


Figure 2. Circular map of *Corynebacterium ulcerans* 0102 showing genomic islands and the position of vaccine (green) and drug targets (purple). GI – genomic island, PAI – pathogenicity island.

3. Four common putative *drug targets in C. silvaticum and C. ulcerans*

For 84 proteins, the sequence alignment with a reference protein had the structure quality classified as G2 (e-value ≤ 10-5, identity ≥ 0.25% and Length Variation Index ≤ 0.1). From these,only the four proteins with alignment quality classified as Very High (identity ≥ 0.75% and Length Variation Index ≤ 0.1) were selected for further analysis (Table 2). The Ramachandran plot for 3D structure validation of these final 4 targets is shown in Supplementary Figure S1, and their respective actives residues in Table S3. These four potential targets are essential and only DtxR is a virulence factor according to the PBIT database prediction (Supplementary Table S2). None of the drug targets were in a PAI or GI (Figure 2).

Table 2. Drug target candidates for *Corynebacterium ulcerans* and *C. silvaticum.*

Product (Gene)	**Molecular Function**	**Biological Process**	**Mass (Da)**	**Metabolic Pathways**	**Virulence Factor**	**Ramachandran residues in most favoral region**	**PATRIC ID**
Iron-dependent repressor IdeR/DtxR (*dtxR*)	DNA-binding transcription factor activity, protein dimerization activity, transition metal ion binding	Transcription, Transcription regulation	25,203	-	Yes	92.2%	996634.5.peg.1487
Uridine monophosphate kinase UMPK (EC 2.7.4.22) (*pyrH*)	ATP binding, UMP kinase activity	'*de novo*' CTP biosynthetic process	26,168	CTP biosynthesis via de novo pathway and in Pyrimidine metabolism	No	94.2%	996634.5.peg.1586
Fructose-Biphosphate aldolase classII (EC 4.1.2.13) (*fba*)	Fructose-bisphosphate aldolase activity, zinc ion binding	Glycolytic process	37,144	Synthesizes D-glyceraldehyde 3-phosphate and glycerone phosphate from D-glucose	No	94.9%	996634.5.peg.2154
Transcriptional regulator, AcrR family	DNA binding, magnesium ion binding	Regulation of transcription	21,831	-	No	95.5%	996634.5.peg.1300

3.3 Top 12 candidate drug like molecules to target four common targets

We have used DoGSiteScorer to select the druggable pockets with score greater than

0.79 (Supplementary Table S3). For the docking analysis, we selected three natural compounds (for each target) based on lower binding energy values and the largest number of hydrogen bonds.The output was the 10 top ranked ligands, based on the best interactions with the proteins actives residues (domain), according to the requisites mentioned. We selected the final 3 out of 10 best compounds for each target protein. The hydrogen bonds, binding energy values, ligands IDs and proteins information are shown in Table 3. For the graphic representation of the docking analysis,we selected the best one compound for each protein (Figures 3-6).

AcrR family protein is a transcriptional regulator that normally acts as a repressor and had better interaction with the compound ZINC04258896 (Figure 3). Fructose-biphosphate aldolaseis an enzyme that participates in the glycolytic pathway and showed the best docking with the compound ZINC04235626 (Figure 4). The IdeR/DtxR target is an iron-dependent transcriptional regulator, classified as a virulence factor that showed the best interaction with the ligand ZINC03840461 (Figure 5). Finally, Uridine monophosphate kinase is a converting enzyme in the metabolism of pyrimidines, It had the best docking with the ligand ZINC08300249 (Figure 6).

Table 3. Docking results of candidate proteins targets and its ligands showing number ofhydrogen bonds with its respective residues and binding energy values.

Protein (PATRICID)	**Ligand ZINC ID**	**IUPAC name**	**Binding energy value (Kcal/mol)**	**H bond / Residues**
Transcriptional regulator, AcrR family (996634.5.peg.1300)	ZINC04258896	1-(1,3-benzodioxol-5-yl)-N-methyl-2-[(E)-3-phenylprop-2-enoyl]-1,3,4,9-tetrahydropyrido[3,4-b]indole-3-carboxamide	-7.9	3/Arg21, Arg92
	ZINC04277685	N-[(1R,9S)-11-(naphthalene-2-carbonyl)-6-oxo-7,11-diazatricyclo[7.3.1.0^{2,7}]trideca-2,4-dien-5-yl]cyclohexanecarboxamide	-8	2/Tyr29, Asp54
	ZINC08300353	N-[(1R,9S)-11-(4-fluorobenzoyl)-6-oxo-7,11-diazatricyclo[7.3.1.0^{2,7}]trideca-2,4-dien-3-yl]-2H-1,3-benzodioxole-5-carboxamide	-8.2	2/Arg22, Asp94

Fructose- biphosphate aldolaseclass II (996634.5.peg.2154)	ZINC04235626	(3R,3aS,6S,6aR)-6-(naphthalene-2-sulfonamido)-hexahydrofuro[3,2-b]furan-3-ylN-(4-methoxyphenyl)carbamate	-9.1	3/Thr29, Ser53, Asp276
	ZINC04235829	(3R,3aS,6S,6aR)-6-(naphthalene-2-sulfonamido)-hexahydrofuro[3,2-b]furan-3-ylN-(4-acetylphenyl)carbamate	-9.3	3/Thr29, Ser53, His212
	ZINC08300421	N-[(1R,9S)-11-(1-acetylpiperidin-4-yl)-6-oxo- 7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-dien- 3-yl]-2H-1,3-benzodioxole-5-carboxamide	-9.6	2/Gly213, Thr277
Iron-dependent repressor IdeR/DtxR (996634.5.peg.1487)	ZINC03840461	N-[(11S,13S)-5-(4-chlorophenyl)-2,10-dioxo-1,9-diazatricyclo[9.4.0.0³,⁸]pentadeca- 3(8),4,6-trien-13-yl]pyridine-3-carboxamide	-9.8	2/Ser70, Thr159
	ZINC08300249	(1R,9S)-11-(2H-1,3-benzodioxole-5-carbonyl)-5-[3-(trifluoromethyl) phenyl]-7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-dien-6-one	-10.2	2/Thr124, Leu135
	ZINC04259070	(1R,9R)-5-{[(3-fluorophenyl)carbamoyl]amino}-N-(4-methoxyphenyl)-6-oxo-7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-diene-11-carboxamide	-10.4	2/Thr24, Thr191
Uridine monophosphate kinase (996634.5.peg.1586)	ZINC08300249	(1R,9S)-11-(2H-1,3-benzodioxole-5-carbonyl)-5-[3-(trifluoromethyl)phenyl]-7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-dien-6-one	-10.3	2/Gly59, Arg64
	ZINC08300441	3-[(1R,9S)-11-(1-acetylpiperidin-4-yl)-6-oxo-7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-dien-3-yl]-1-(3-cyanophenyl)urea	-8.6	3/Glu22, Phe145, Thr147
	ZINC08300421	N-[(1R,9S)-11-(1-acetylpiperidin-4-yl)-6-oxo- 7,11-diazatricyclo[7.3.1.0²,⁷]trideca-2,4-dien- 3-yl]-2H-1,3-benzodioxole-5-carboxamide	-9.9	2/Gly59, Arg64

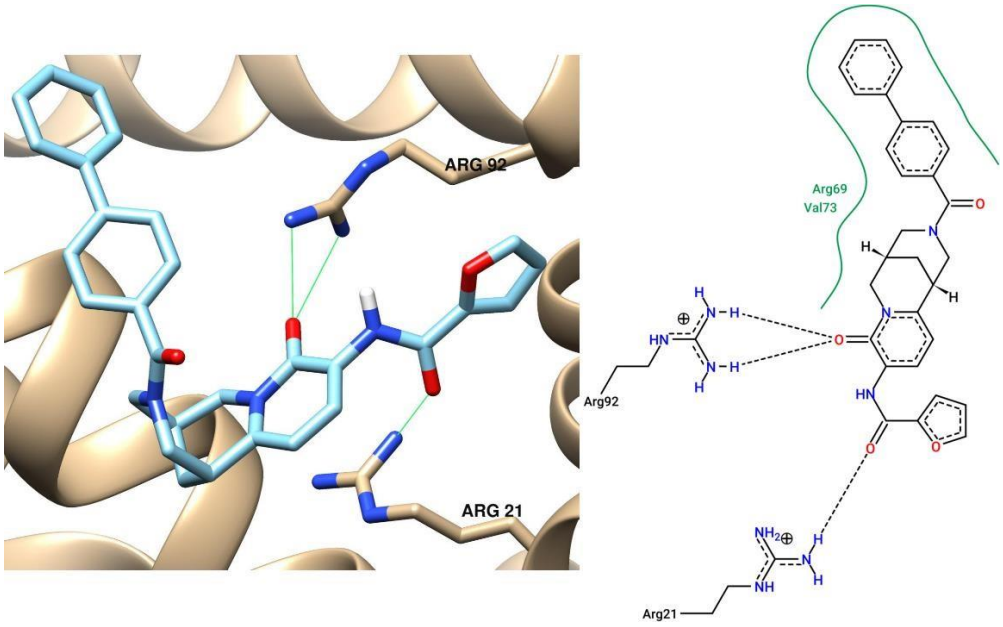


Figure 3. 3D and 2D graphic representation of the docking analysis for the most druggableprotein cavity of drug target AcrR family protein with ZINC04258896.

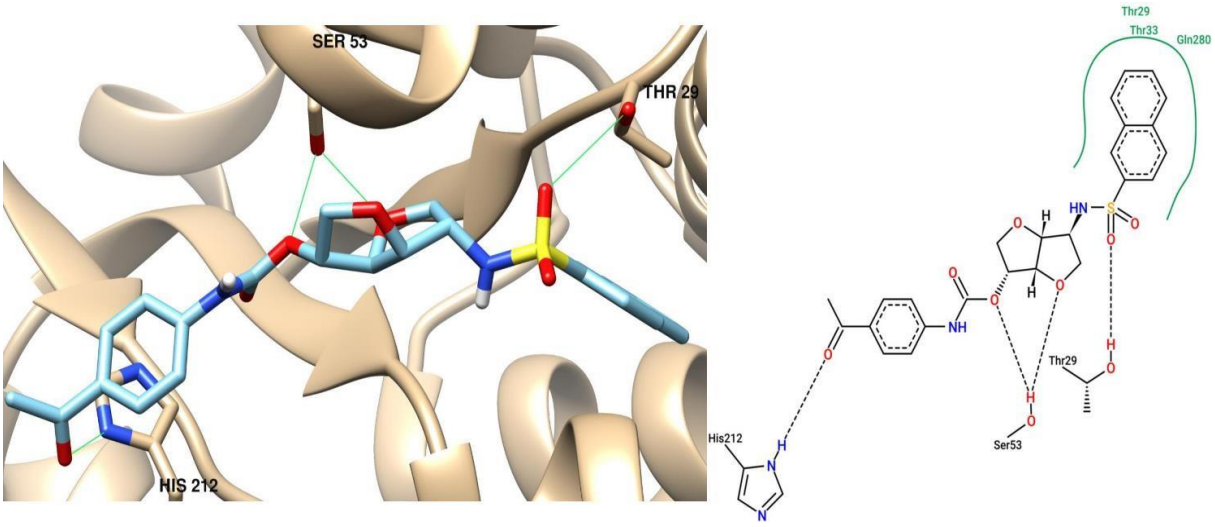


Figure 4. 3D and 2D graphic representation of the docking analysis for the most druggableprotein cavity of drug target Fructose-bisphosphate aldolase class II with ZINC04235829.

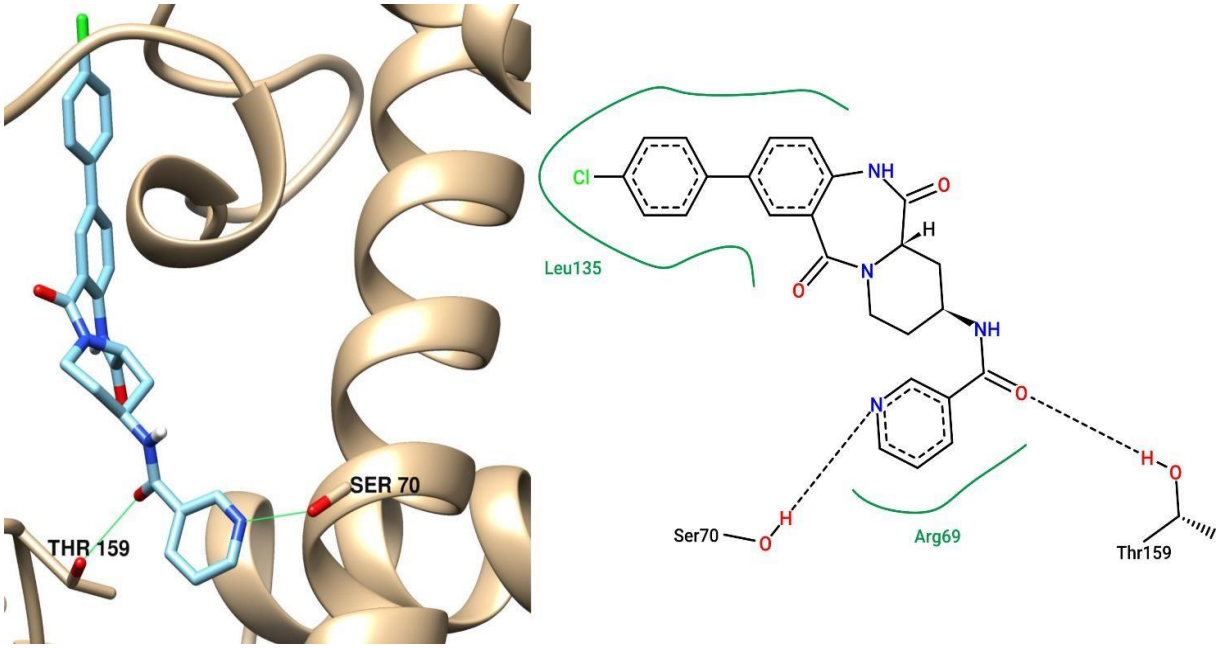


Figure 5. 3D and 2D graphic representation of the docking analysis for the most druggableprotein cavity of drug target Iron-dependent repressor IdeR/DtxR with ZINC03840461.

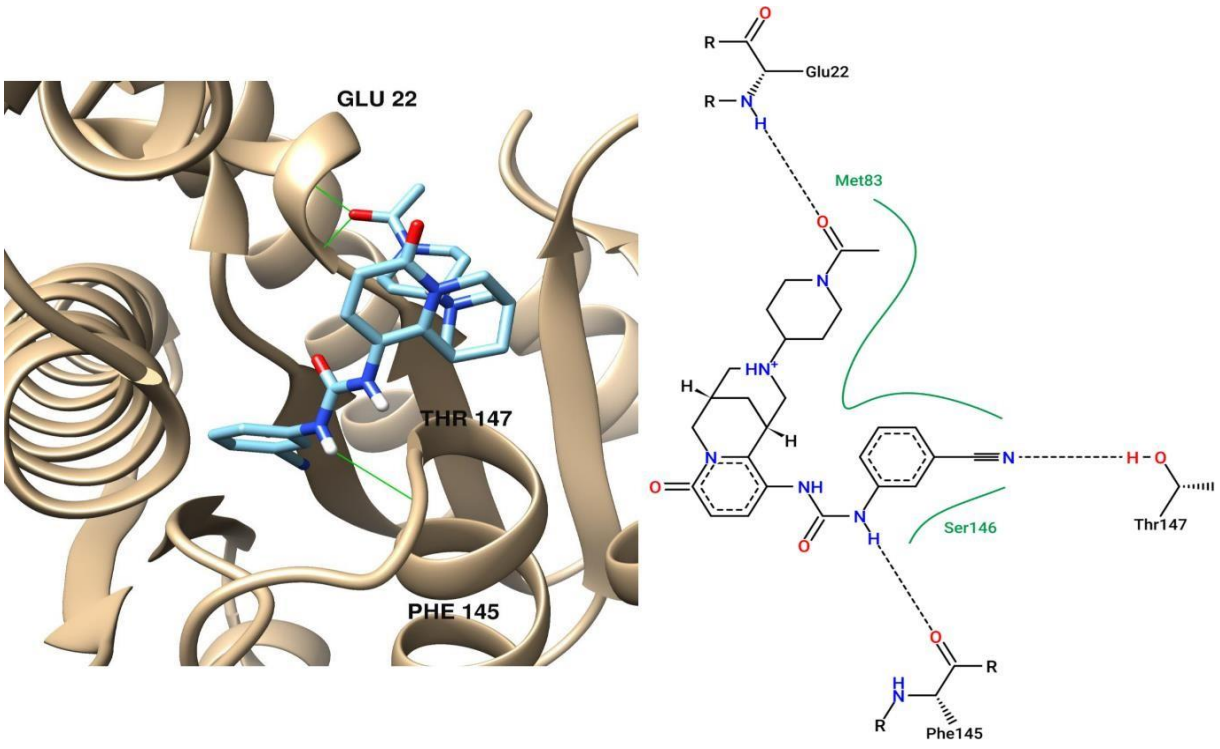


Figure 6. 3D and 2D graphic representation of the docking analysis for the most druggableprotein cavity of drug target Uridine monophosphate kinase UMPK with ZINC08300441.

4. Discussion

C. ulcerans can infect humans and a variety of wild and domestic animals that act as a reservoir. Manifestations can be asymptomatic, or cause damage to the respiratory tract, mastitis and gangrenous dermatitis (Tiwari et al. 2008; Hacker et al. 2016). Non-toxigenic strains can be classified as tox-negative or non-toxigenic but *tox* gene-bearing (NTTB) (Fuursted et al. 2015; Dangel et al. 2019). *C. silvaticum* is a recently described NTTB species close to *C. ulcerans* that causes caseous lymphadenitis in wild animals such as wild boar and roe deer, which are reservoirs, and domestic pigs (Dangel et al. 2020). Its potential host range is unknown, and It could potentially cause zoonotic infections (Viana et al. 2020).

Subtractive genomics, as the name suggests, is an approach based on several steps of filtering proteomic sequences, in order to select and identify final proteins that are indispensable to microorganisms and that are not homologous to the host. This methodology is widely used in reverse vaccinology, a method which includes subtractive genomics, modeling, docking andother in silico techniques in order to predict vaccine and drug targets, without financial expense and in a short time (Hassan et al. 2018; Masignani et al. 2019). The Bexsero vaccine for Neisseria meningitidis (Christodoulides and Heckels 2017) and Engerix-B vaccine for Hepatitis B (Keatingand Noble 2003) were developed using these approaches.

We predicted nine vaccine targets, four drug targets and four probable dug molecules for *C. ulcerans* and *C. silvaticum* using reverse vaccinology and *in silico* drug targeting (Tables 1 and2). Among nine predicted vaccine targets, four were annotated as subunits of ABC transporters. Two are substrate-binding protein for manganese (Mn) and zinc (Zn), one for phosphate, and one is a permease for glutamate. Both metals are vital for several organisms, required by transcriptional transporters and regulators and oxidative stress response (Mn), and several ribosomal proteins and tRNA synthetases (Zn). The acquisition of both metals is disputed withthe host and those ABC transporters are required for colonization (Nies and Grass 2009; Juttukonda and Skaar 2015). One of the transporter's subunits (*mntA*) is in PAI 3 (Figure 2). Glutamate is an essential metabolite that plays an relevant role in the metabolism of nitrogen and carbon. The ABC glutamate transporter has already been characterized in other bacterial genera and is a known vaccine target, considering its indispensable role in the survival of the bacterium (Bhatia et al. 2014). Phosphate is essential for energy metabolism and is a component of nucleic acids, phospholipids, and other cell molecules. The Phosphate ABC transporter, substrate-bindingprotein PstS is part of the complex PstSACB involved in phosphate import (Santos-Beneit 2015). Inactivation of the *pstS* gene increased the susceptibility to penicillin in

Streptococcus pneumoniae (Soualhine et al. 2005). The transport of these essential nutrients makes thoseproteins promising vaccine candidates.

Among the non-transporter vaccine target candidates is a bifunctional protein from the Phospholipase/Thioesterase family, with the gene located in pathogenicity island PAI 6 (Figure 2). This enzyme is involved in the non-ribosomal synthesis of peptides, including antimicrobial peptides (Schneider and Marahiel 1998; Santucci et al. 2018). The Integral membrane protein belong to the "Putative Actinobacterial Holin-X, holin superfamily III". Holins are produced by double-stranded DNA bacteriophages that utilize an endolysin-holin tactics to lysogenize their hosts (InterPro entry IPR009937). The other three candidates are hypothetical proteins classified by InterProScan as "Papain-like cysteine peptidase superfamily" (InterPro entry IPR038765)(PATRIC ID 996634.5.peg.652), "Calcium-dependent phosphotriesterase" (SSF63829) (996634.5.peg.1276) and a transmembrane protein (996634.5.peg.823).

Within the drug targets, DtxR was the only predicted virulence factor. It is an iron-dependent repressor in prokaryotes that controls the expression of genes involved in iron homeostasis and virulence factors, such as diphtheria toxin (*tox*) (Wagner et al. 2014). DtxR regulates genes according to the concentration of Fe_2^+ ions in the environment, allowing to adjust iron uptake. Itis considered a virulence factor due to its importance for host colonization (Merchant andSpatafora 2014). Furthermore, DxtR has been considered an attractive drug target due to its rolein regulating genes involved in iron homeostasis in bacteria, and such genes are generallyvirulence factors (Cheng et al. 2018; Parise et al. 2021). Residue SER 70 is hydrogen bonded to the nitrogen of the aromatic ring, THR 159 is hydrogen linked to the third oxygen of the compound and residues ARG 69 and LEU 135 have hydrophobic interactions with the three aromatic rings.

Uridine Monophosphate Kinase is an important enzyme that acts on the metabolism of pyrimidines in bacteria, specifically on uridine monophosphate (UMP). It converts UMP to UDP that is later used in the nucleoside biosynthesis. UMPK plays an essential role in the productionof nucleotides that are constituents of nucleic acids. This enzyme is of interest, as It has been recognized as a potential drug target for tuberculosis (ARVIND et al. 2013) as It also plays a role in regulating the balance of pyrimidine and purine nucleosides. In addition, It is structurally different from the eukaryotic UMP kinases (Rostirolla et al. 2011), which makes It a more attractive candidate. ASP 79 makes two hydrogen bonds with NH (NH-O), GLY 21 makes two hydrogen bonds with the first oxygen of the molecule. The 2D image shows hydrogen bonding ofMET 140 to the last nitrogen. MET 83 shows hydrophobic interaction with the last aromatic ring.

The Fructose-1,6-bisphosphate aldolase (FBA) is a glycolytic protein that participates in the glucose metabolic pathway and in the production of organic acid when oxygen is scarce (Altenhoff and Dessimoz 2009; Teramoto et al. 2010). In some genera, It has been shown that reducing FBA leads to cell death. This enzyme is present in the membrane and exposed on the bacterium's cell surface, and its association to virulence is already know as It binds to human plasminogen and mammalian cells (Shams et al. 2014). FBA and other proteins are expressed under conditions of alkaline stress, which indicates that this enzyme is also involved in the physiological breakdown of the bacterium during environmental changes. For these reasons, this protein has previously been recognized as a potential candidate for drug and vaccine targets in various pathogens such as bacteria, fungi and parasites of animals and humans (Pirovich et al. 2021)(Shams et al. 2014). SER 53 is hydrogen bonded to two oxygens (one from the pentose), HIS 212 is hydrogen bonded to an oxygen at the ligand end, and THR 29 is hydrogen bonded to one oxygen (O=S=O). THR 29, THR 33 and GLN 280 have a hydrophobic interaction with naphthalene.

The AcrR family of transcriptional regulators belongs to the one-component system and is associated with several essential cellular mechanisms of bacteria and archaea, such as cell signaling, carbon, nitrogen and lipid metabolism, amino acid metabolism and cofactor metabolism, production of antibiotics, among other physiological factors. They normally act as repressors, having different regulatory mechanisms in different species of bacteria, which can be positive or negative. In addition, these regulators play a key role in antibiotic resistance (Cuthbertson and Nodwell 2013). For these reasons, proteins from this family have been identified as a broad-spectrum drug target (Deng et al. 2013) and for novel treatments (Gonzáles et al. 2018). ARG 21 has two hydrogen bonds with the oxygen that is attached to the aromatic ring, ARG 92 has a hydrogen bond with an oxygen of the ligand. The two aromatic rings show hydrophobic interactions with VAL 73 and ARG 69.

In this work we presented a first-time approach applied to these bacteria, in which we use common drugs and vaccines for both species. We predicted nine vaccines and four drug target candidates, as well as the respective drug molecules for *C. ulcerans* and *C. silvaticum*. This type of analysis could reduce the cost and time for the development of vaccine and identification of antimicrobial compounds and should be subjected to further experimental validation, as these species can be cultivated. This pipeline is not restricted to *Corynebacterium* but could be used on any prokaryotic or eukaryotic pathogen.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

We acknowledge the collaboration and assistance of all team members, the Programa Interunidades de Pós-graduação em Bioinformática da UFMG and the Brazilian funding agenciesCAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação deAmparo à Pesquisa de Minas Gerais). KKJA is supported by the Taif University Researchers Supporting Program (project number: TURSP-2020/128), Taif University, Saudi Arabia; MMT acknowledges the grant from the Dowager Countess Eleanor Peel Trust (Number # 295). VA is supported by a CNPq research fellowship.

SUPPLEMENTARY MATERIAL

Supplementary Table S1. List of 108 genome samples, 72 of *Corynebacterium ulcerans* and 36 of *C. silvaticum*.

Supplementary Table S2. Functional annotation of the drug and vaccine targets shared by *Corynebacterium ulcerans* and *C. silvaticum*

Supplementary Table S3. Druggable pockets in drug targets with score greater than 0.79 predicted using DoGSiteScorer.

Supplementary Figure S1. Ramachandran plot for 3D structure validation of the final 4 drugtargets.

REFERENCES

Altenhoff AM, Dessimoz C 2009 Phylogenetic and functional assessment of orthologs inferenceprojects and methods. *PLoS Comput Biol* **5**: e1000262. https://doi.org/10.1371/journal.pcbi.1000262

Arvind A, Jain V, Saravanan P, Mohan Cg 2013 Uridine Monophosphate Kinase as Potential Target for Tuberculosis: From Target to Lead Identification. *Interdiscip Sci Comput Life Sci*296–311. https://doi.org/10.1007/s12539-013-0180-y

Barinov A, Loux V, Hammani A, *et al* 2009 Prediction of surface exposed proteins in Streptococcus pyogenes, with a potential application to other Gram-positive bacteria. *Proteomics* **9**:61–73. https://doi.org/10.1002/pmic.200800195

Bhatia B, Ponia SS, Solanki AK, *et al* 2014 Identification of glutamate ABC-Transporter component in Clostridium perfringens as a putative drug target. *Bioinformation* **10**:401–405.https://doi.org/10.6026/97320630010401

Brettin T, Davis JJ, Disz T, *et al* 2015 RASTtk: A modular and extensible implementation of theRAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports* **5**:8365. https://doi.org/10.1038/srep08365

Burgos-Reyes MA, Baylón-Pacheco L, Espíritu-Gordillo P, *et al* 2021 Effect of Prophylactic Vaccination with the Membrane-Bound Acid Phosphatase Gene of Leishmania mexicana inthe Murine Model of Localized Cutaneous Leishmaniasis. *J Immunol Res* **2021**:6624246. https://doi.org/10.1155/2021/6624246

Camacho C, Coulouris G, Avagyan V, *et al* 2009 BLAST+: architecture and applications. BMC *Bioinformatics* **10**:421. https://doi.org/10.1186/1471-2105-10-421

Cheng Y, Yang R, Lyu M, *et al* 2018 IdeR, a DtxR Family Iron Response Regulator, Controls Iron Homeostasis, Morphological Differentiation, Secondary Metabolism, and the OxidativeStress Response in Streptomyces avermitilis. *Applied and Environmental Microbiology* **84**: https://doi.org/https://doi.org/10.1128/ AEM.01503-18

Christodoulides M, Heckels J 2017 Novel approaches to Neisseria meningitidis vaccine design. *Pathogens and Disease* **75**:1–16. https://doi.org/10.1093/femspd/ftx033

Couto J, Seixas G, Stutzer C, *et al* 2021 Probing the rhipicephalus bursa sialomes in potential anti-tick vaccine candidates: A reverse vaccinology approach. *Biomedicines* **9**:. https://doi.org/10.3390/biomedicines9040363

Cuthbertson L, Nodwell JR 2013 The TetR Family of Regulators. Microbiology and Molecular *Biology Reviews* **77**:440–475. https://doi.org/10.1128/mmbr.00018-13

Dangel A, Berger A, Konrad R, Sing A 2019 NGS-based phylogeny of diphtheria-related pathogenicity factors in different *Corynebacterium* spp. implies species-specific virulencetransmission. *BMC Microbiology* **19**:1–16. https://doi.org/10.1186/s12866-019-1402-1

Dangel A, Berger A, Rau J, *et al* 2020 *Corynebacterium* silvaticum sp. Nov., a unique group ofNTTB corynebacteria in wild boar and roe deer. International Journal of Systematic and *Evolutionary Microbiology* **70**:3614–3624. https://doi.org/10.1099/ijsem.0.004195

Das P, Sercu T, Wadhawan K, *et al* 2021 Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nat Biomed Eng*. https://doi.org/10.1038/s41551-021-00689-x

Davis JJ, Wattam AR, Aziz RK, *et al* 2020 The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Research* **48**:D606–D612. https://doi.org/10.1093/nar/gkz943

de Alwis R, Liang L, Taghavian O, *et al* 2021 The identification of novel immunogenic antigensas potential Shigella vaccine components. *Genome Medicine* **13**:8. https://doi.org/10.1186/s13073-020-00824-4

Deng W, Li C, Xie J 2013 The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cellular Signalling* **25**:1608–1613. https://doi.org/10.1016/j.cellsig.2013.04.003

Emms DM, Kelly S 2015 OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**:1–14. https://doi.org/10.1186/s13059-015-0721-2

Fuursted K, Søes LM, Crewe BT, *et al* 2015 Non-toxigenic tox gene-bearing *Corynebacterium ulcerans* in a traumatic ulcer from a human case and his asymptomatic dog. *Microbes and Infection* **17**:717–719. https://doi.org/10.1016/j.micinf.2015.07.004

Gonzáles A, Fillat MF, Lanas Á 2018 Transcriptional regulators: valuable targets for novel antibacterial strategies. *Future Medicinal Chemistry* **10**:541–560. https://doi.org/10.4155/fmc-2017-0181

Grosse-Kock S, Kolodkina V, Schwalbe EC, *et al* 2017 Genomic analysis of endemic clones oftoxigenic and non-toxigenic *Corynebacterium diphtheriae* in Belarus during and after

the major epidemic in 1990s. *BMC Genomics* **18**:1–10. https://doi.org/10.1186/s12864-017- 4276-3

Hacker E, Antunes CA, Mattos-Guaraldi AL, *et al* 2016 *Corynebacterium ulcerans*, an emerginghuman pathogen. *Future Microbiology* **11**:1191–1208. https://doi.org/10.2217/fmb-2016- 0085

Hajialibeigi A, Amani J, Gargari SLM 2021 Identification and evaluation of novel vaccine candidates against Shigella flexneri through reverse vaccinology approach. Applied *Microbiology and Biotechnology* **105**:1159–1173. https://doi.org/10.1007/s00253-020-11054-4

Hassan SS, Jamal SB, Radusky LG, *et al* 2018 The druggable pocketome of *Corynebacterium diphtheriae*: A new approach for in silico putative druggable targets. *Frontiers in Genetics***9**:1–9. https://doi.org/10.3389/fgene.2018.00044

He Y, Xiang Z, Mobley HLT 2010 Vaxign: The First Web-Based Vaccine Design Program for Reverse Vaccinology and Applications for Vaccine Development. *Journal of Biomedicineand Biotechnology* **2010**:1–15. https://doi.org/10.1155/2010/297505

Huerta-Cepas J, Forslund K, Coelho LP, *et al* 2017 Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution* **34**:2115–2122. https://doi.org/10.1093/molbev/msx148

Jain C, Rodriguez-R LM, Phillippy AM, *et al* 2018 High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**:5114. https://doi.org/10.1038/s41467-018-07641-9

Juttukonda LJ, Skaar EP 2015 Manganese homeostasis and utilization in pathogenic bacteria. *Molecular Microbiology* **97**:216–228. https://doi.org/10.1111/mmi.13034

Keating GM, Noble S 2003 Recombinant Hepatitis B Vaccine Engerix-B. *Drugs* **63**:1021–1051.https://doi.org/10.2165/00003495-200363100-00006

Laskowski RA, MacArthur MW, Moss DS, Thornton JM 1993 PROCHECK: a program to checkthe stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**:283–291. https://doi.org/10.1107/S0021889892009944

Masignani V, Pizza M, Moxon ER 2019 The development of a vaccine against MeningococcusB using reverse vaccinology. *Frontiers in Immunology* **10**:1–14. https://doi.org/10.3389/fimmu.2019.00751

Meier-Kolthoff JP, Göker M 2019 TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications* **10**:2182. https://doi.org/10.1038/s41467-019-10210-3

Merchant AT, Spatafora GA 2014 A role for the DtxR family of metalloregulators in gram-positive pathogenesis. *Molecular Oral Microbiology* **29**:1–10. https://doi.org/10.1111/omi.12039

Möller J, Busch A, Berens C, *et al* 2021 Newly Isolated Animal Pathogen *Corynebacterium silvaticum* Is Cytotoxic to Human Epithelial Cells

Morris GM, Huey R, Lindstrom W, *et al* 2009 AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**:2785–2791. https://doi.org/10.1002/jcc.21256

Nies DH, Grass G 2009 Transition Metal Homeostasis. *EcoSal Plus* **3**: ecosalplus.5.4.4.3. https://doi.org/10.1128/ecosalplus.5.4.4.3

Otsuji K, Fukuda K, Ogawa M, Saito M 2019 Mutation and diversity of diphtheria toxin in *Corynebacterium ulcerans*. *Emerging Infectious Diseases* **25**:2122–2123. https://doi.org/10.3201/eid2511.181455

Parise D, Teixeira M, Parise D, *et al* 2021 The Transcriptional Regulatory Network of *Corynebacterium pseudotuberculosis*. *Frontiers in Microbiology* **12**:656435

Pettersen EF, Goddard TD, Huang CC, *et al* 2004 UCSF Chimera - A visualization system for

exploratory research and analysis. *Journal of Computational Chemistry* **25**:1605–1612. https://doi.org/10.1002/jcc.20084

Pirovich DB, Da AA, Skelly PJ 2021 Multifunctional Fructose 1, 6-Bisphosphate Aldolase as a*Therapeutic Target*. **8**:1–19. https://doi.org/10.3389/fmolb.2021.719678 Rappuoli R 2000 Reverse vaccinology. *Current Opinion in Microbiology* **3**:445–450. https://doi.org/10.1016/S1369-5274 00 00119-3

Rappuoli R, Bottomley MJ, D'Oro U, *et al* 2016 Reverse vaccinology 2.0: Human immunologyinstructs vaccine antigen design. *Journal of Experimental Medicine* **213**:469–481. https://doi.org/10.1084/JEM.20151960

Rappuoli R, Malito E 2014 History of Diphtheria Vaccine Development. In: Burkovski A ed *Corynebacterium diphtheriae* and Related Toxigenic Species. *Springer Netherlands*, Dordrecht, pp 225–238

Rostirolla DC, Breda A, Rosado LA, *et al* 2011 UMP kinase from Mycobacterium tuberculosis:Mode of action and allosteric interactions, and their likely role in pyrimidine metabolism regulation. *Archives of Biochemistry and Biophysics* **505**:202–212. https://doi.org/10.1016/j.abb.2010.10.019

Santos-Beneit F 2015 The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology* **6**: https://doi.org/10.3389/fmicb.2015.00402

Santucci P, Point V, Poncin I, *et al* 2018 LipG a bifunctional phospholipase/thioesterase involvedin mycobacterial envelope remodeling. *Bioscience Reports* **38**:1–21. https://doi.org/10.1042/BSR20181953

Schneider A, Marahiel MA 1998 Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Archives of Microbiology* **169**:404–410. https://doi.org/10.1007/s002030050590

Shams F, Oldfield NJ, Wooldridge KG, Turner DPJ 2014 Fructose-1,6-bisphosphate aldolase FBA -A conserved glycolytic enzyme with virulence functions in bacteria: "I'll met by moonlight." *Biochemical Society Transactions* **42**:1792–1795. https://doi.org/10.1042/BST20140203

Shende G, Haldankar H, Barai RS, *et al* 2017 PBIT: Pipeline builder for identification of drug targets for infectious diseases. *Bioinformatics* **33**:929–931. https://doi.org/10.1093/bioinformatics/btw760

Soares SC, Geyik H, Ramos RTJ, *et al* 2016 GIPSy: Genomic Island prediction software. *Journalof Biotechnology* **232**:2–11. https://doi.org/10.1016/j.jbiotec.2015.09.008

Soualhine H, Brochu V, Ménard F, *et al* 2005 A proteomic analysis of penicillin resistance in Streptococcus pneumoniae reveals a novel role for PstS, a subunit of the phosphate ABC transporter. *Mol Microbiol* **58**:1430–40. https://doi.org/10.1111/j.1365-2958.2005.04914.x

Teramoto H, Inui M, Yukawa H 2010 Regulation of genes involved in sugar uptake, glycolysis and lactate production in *Corynebacterium glutamicum*. *Future Microbiology* **5**:1475–1481.https://doi.org/10.2217/fmb.10.114

Tiwari TSP, Golaz A, Yu DT, *et al* 2008 Investigations of 2 cases of diphtheria-like illness due totoxigenic *Corynebacterium ulcerans*. *Clin Infect Dis* **46**:395–401. https://doi.org/10.1086/525262

Trott O, Olson AJ 2010 AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**:455–461. https://doi.org/10.1002/jcc.21334

Truelove SA, Keegan LT, Moss WJ, *et al* 2020 Clinical and epidemiological aspects of diphtheria: A systematic review and pooled analysis. *Clinical Infectious Diseases* **71**:89–97

Vandentorren S, Guiso N, Badell E, *et al* 2014 Toxigenic *Corynebacterium ulcerans* in a fatal human case and her feline contacts, France, March 2014. *Euro surveillance: bulletin* Europeen sur les maladies transmissibles = European communicable disease bulletin **19**:20910. https://doi.org/10.2807/1560-7917.es2014.19.38.20910

Viana MVC, Profeta R, da Silva AL, *et al* 2020 Taxonomic classification of strain PO100/5 shows a broader geographic distribution and genetic markers of the recently described *Corynebacterium silvaticum*. *PLOS ONE* **15**: e0244210. https://doi.org/10.1371/journal.pone.0244210

Volkamer A, Kuhn D, Rippmann F, Rarey M 2012 DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**:2074–2075.https://doi.org/10.1093/bioinformatics/bts310

Wagner KS, Zakikhany K, White JM, *et al* 2014 Diphtheria Surveillance. In: *Corynebacteriumdiphtheriae* and Related Toxigenic Species. *Springer Netherlands*, Dordrecht, pp 207–224

Wasmuth E V, Lima CD, Bateman A, *et al* 2017 UniProt: The universal protein knowledgebase.

Nucleic Acids Research **45**: D158–D169. https://doi.org/10.1093/nar/gkw1099 Wick RR, Judd LM, Gorrie CL, Holt KE 2017 Unicycler: Resolving bacterial genome

assemblies from short and long sequencing reads. *PLOS Computational Biology* **13**: e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Zasada AA 2013 Nontoxigenic highly pathogenic clone of *Corynebacterium diphtheriae*, Poland, 2004-2012. *Emerging Infectious Diseases* **19**:1870–1872. https://doi.org/10.3201/eid1911.130297

Suporting Information

Supplementary Table S1. List of 108 genome samples, 72 of *Corynebacterium ulcerans* and 36of *C. silvaticum*.

Species	**Strain**	**Host**	**Country**	**Accession number**	**SRR accession**
C. silvaticum	KL0182	Wild boar	Germany	SDQO02	SRR7825394
C. silvaticum	KL0183	Wild boar	Germany	-	SRR7825401
C. silvaticum	KL0259	Wild boar	Germany	-	SRR7825414
C. silvaticum	KL0260	Wild boar	Germany	-	SRR7825417
C. silvaticum	KL0374	Wild boar	Germany	-	SRR7825405
C. silvaticum	KL0382	Wild boar	Germany	-	SRR7825406
C. silvaticum	KL0386	Wild boar	Germany	-	SRR7825407
C. silvaticum	KL0394	Wild boar	Germany	-	SRR7825389
C. silvaticum	KL0395	Wild boar	Germany	-	SRR7825388
C. silvaticum	KL0396	Wild boar	Germany	-	SRR7825387
C. silvaticum	KL0400	Wild boar	Germany	-	SRR7825386
C. silvaticum	KL0401	Wild boar	Germany	-	SRR7825393
C. silvaticum	KL0581	Wild boar	Germany	-	SRR7825370
C. silvaticum	KL0598	Wild boar	Germany	-	SRR7825373
C. silvaticum	KL0615	Wild boar	Germany	-	SRR7825366
C. silvaticum	KL0707	Wild boar	Germany	-	SRR7825360
C. silvaticum	KL0709	Wild boar	Germany	-	SRR7825361
C. silvaticum	KL0773	Wild boar	Germany	-	SRR7825362
C. silvaticum	KL0774	Wild boar	Germany	-	SRR7825363

C. silvaticum	KL0882	Wild boar	Germany	-	SRR7825436
C. silvaticum	KL0883	Wild boar	Germany	-	SRR7825437
C. silvaticum	KL0884	Wild boar	Germany	-	SRR7825442
C. silvaticum	KL0886	Wild boar	Germany	-	SRR7825443
C. silvaticum	KL0887	Wild boar	Germany	-	SRR7825440
C. silvaticum	KL0938	Wild boar	Germany	-	SRR7825434
C. silvaticum	KL0957	Wild boar	Germany	SDVC01	SRR7825429
C. silvaticum	KL0968	Wild boar	Germany	-	SRR7825428
C. silvaticum	KL1003	Wild boar	Germany	-	SRR7825431
C. silvaticum	KL1006	Wild boar	Germany	-	SRR7825430
C. silvaticum	KL1007	Wild boar	Germany	-	SRR7825433
C. silvaticum	KL1008	Wild boar	Germany	-	SRR7825432
C. silvaticum	KL1009	Wild boar	Germany	-	SRR7825425
C. silvaticum	KL1010	Wild boar	Germany	-	SRR7825424
C. silvaticum	KL1196	Deer	Germany	SDVD01	-
C. silvaticum	PO100/5	-	Portugal	CP021417.1	-
C. silvaticum	W25	Wild boar	Germany	VFEM01	-
C. ulcerans	0102	-	-	AP012284.1	-
C. ulcerans	809	-	Brazil	CP002790.1	-
C. ulcerans	2590	Human	Brazil	MPSS01	-
C. ulcerans	4940	Human	Belarus	LSWN01	-
C. ulcerans	05146	-	-	CP009716.1	-
C. ulcerans	131001	-	-	CP010818.1	-
C. ulcerans	131002	-	-	CP011095.1	-
C. ulcerans	210931	-	-	CP009583.1, CP009584.1	-
C. ulcerans	210932	-	-	CP009500.1	-
C. ulcerans	0211	Human	Japan	AP019662.1	-
C. ulcerans	03-8664	Human	France	LGSY02	-
C. ulcerans	04-3911	Human	France	LGSX01	-
C. ulcerans	04-7514	Dog	France	LJVH01	-
C. ulcerans	BR-AD 2649	Dog	Brazil	MPST01	-
C. ulcerans	BR-AD22	-	Brazil	CP002791.1	-
C. ulcerans	FH20161	Human	Japan	AP019663.1	-
C. ulcerans	FRC11	-	France	CP009622.1	-
C. ulcerans	FRC58	-	France	CP011913.1	-
C. ulcerans	KZN-2016-48390	-	South Africa	MIOS01	-
C. ulcerans	LSPQ-04227	Human	Canada	JZUS01	-
C. ulcerans	LSPQ-04228	Human	Quebec	JZUT01	-
C. ulcerans	MRi49	Horse	UK	CP046863.1	-
C. ulcerans	NCTC 12077	Homo sapiens	UK	AYUJ01	-
C. ulcerans	NCTC7908	-	-	LS483400.1	-
C. ulcerans	NCTC7910	-	UK	LT906443.1	-
C. ulcerans	NCTC8639	-	-	LS483416.1	-
C. ulcerans	NCTC8666	-	-	UFXR01	-
C. ulcerans (NTTB)	KL0195	Human	-	-	SRR7825399
C. ulcerans (NTTB)	KL0251-cb4	Cat	-	-	SRR7825402
C. ulcerans (NTTB)	KL0252-cb5	Cat	-	-	SRR7825415

C. ulcerans (NTTB)	KL0315-cb6	Human	-	-	SRR7825408
C. ulcerans (NTTB)	KL0318-cb7	Dog	-	-	SRR7825409
C. ulcerans (NTTB)	KL0345	Cat	-	-	SRR7825410
C. ulcerans (NTTB)	KL0387-cb8	Human	-	-	SRR7825412
C. ulcerans (NTTB)	KL0392-cb9	Cat	-	-	SRR7825413
C. ulcerans (NTTB)	KL0442	Human	-	-	SRR7825391
C. ulcerans (NTTB)	KL0468	Human	-	-	SRR7825384
C. ulcerans (NTTB)	KL0472	Human	-	-	SRR7825382
C. ulcerans (NTTB)	KL0501	Human	-	-	SRR7825378
C. ulcerans (NTTB)	KL0515	Human	-	-	SRR7825379
C. ulcerans (NTTB)	KL0541	Human	-	-	SRR7825377
C. ulcerans (NTTB)	KL0818	Human	-	-	SRR7825344
C. ulcerans (NTTB)	KL0846	Human	-	-	SRR7825340
C. ulcerans (NTTB)	KL0867	Human	-	-	SRR7825353
C. ulcerans (NTTB)	KL0870	Human	-	-	SRR7825352
C. ulcerans (NTTB)	KL0880	Human	-	-	SRR7825439
C. ulcerans (NTTB)	KL1015	Human	-	-	SRR7825347
C. ulcerans (NTTB)	KL1025	Cat	-	-	SRR7825349
C. ulcerans (*tox*-)	KL0199	Human	-	-	SRR7825398
C. ulcerans (*tox*-)	KL0349	Human	-	-	SRR7825411
C. ulcerans (*tox*-)	KL0451	Human	-	-	SRR7825390
C. ulcerans (*tox*-)	KL0853	Human	-	-	SRR7825339
C. ulcerans (*tox*-)	KL0876	Human	-	-	SRR7825438
C. ulcerans (*tox*-)	KL0941	Human	-	-	SRR7825435
C. ulcerans (*tox*-)	KL1017	Cattle	-	-	SRR7825348
C. ulcerans (*tox*+)	08-1143-CB1	Pig	-	-	SRR7825397
C. ulcerans (*tox*+)	KL0126-cb2	Human	-	-	SRR7825396
C. ulcerans (*tox*+)	KL0194	Human	-	-	SRR7825400
C. ulcerans (*tox*+)	KL0246-cb3	Human	-	-	SRR7825403

C. ulcerans (*tox+*)	KL0350	Human	-	-	SRR7825404
C. ulcerans (*tox+*)	KL0433	Human	-	-	SRR7825392
C. ulcerans (*tox+*)	KL0475	Human	-	-	SRR7825383
C. ulcerans (*tox+*)	KL0483	Human	-	-	SRR7825380
C. ulcerans (*tox+*)	KL0497	Human	-	-	SRR7825381
C. ulcerans (*tox+*)	KL0540	Human	-	-	SRR7825376
C. ulcerans (*tox+*)	KL0547	Human	-	-	SRR7825374
C. ulcerans (*tox+*)	KL0556	Human	-	-	SRR7825375
C. ulcerans (*tox+*)	KL0785	Human	-	-	SRR7825346
C. ulcerans (*tox+*)	KL0796	Human	-	-	SRR7825345
C. ulcerans (*tox+*)	KL0825	Human	-	-	SRR7825343
C. ulcerans (*tox+*)	KL0832	Human	-	-	SRR7825342
C. ulcerans (*tox+*)	KL0840	Human	-	-	SRR7825341

Supplementary Table S2. Functional annotation of the drug and vaccine targets shared by *Corynebacterium ulcerans* and *C. silvaticum*

	GenBank ID			GIPSy	VFDB	PBIT	eggNOG-mapper	
Target	**ID**	**Gene**	**Product**	**Genomic island**	**Virulence Factor**	**Virulence Factor**	**Preferred name**	**Description**
Vaccine	CULC0102_0158	-	ABC transporter, substrate-binding protein(cluster 14, Mn/Zn)	-	-	Yes	*troA*	Belongs to the bacterial solute- binding protein 9family
Vaccine	CULC0102_0597	*mntA*	ABC transporter, substrate-binding protein(cluster 14, Mn/Zn)	PAI 3	-	Yes	*mntA*	Belongs to the bacterial solute- binding protein 9 family
Vaccine	CULC0102_0803	*gluC*	Glutamate ABC transporter, permease protein 1 GluC	-	-	Yes	*gluC*	ABC-type amino acid transport system permease component
Vaccine	CULC0102_0624	-	Hypothetical protein	-	-	Yes		Cell wall-associated hydrolases
Vaccine	CULC0102_0791	-	Hypothetical protein	-	-	No	*yccF*	membrane
Vaccine	CULC0102_1236	-	Hypothetical protein	-	-	No	*lppL*	amine dehydrogenase activity

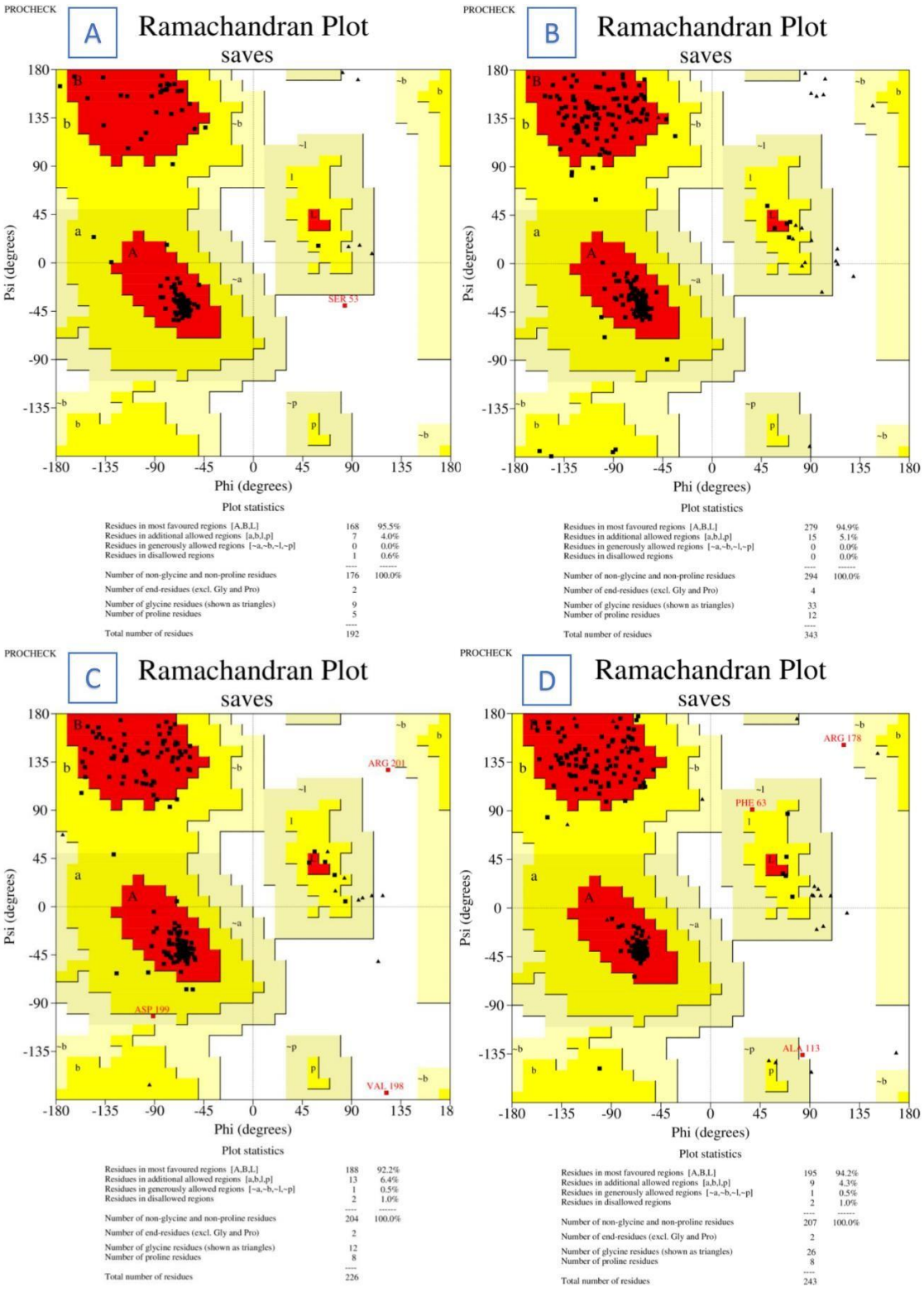
Vaccine	CULC0102_0305	-	Integral membrane protein	-	-	No		Putative Actinobacterial Holin-X, holin superfamily III
Vaccine	CULC0102_1930	*pstS*	Phosphate ABC transporter, substrate-binding protein PstS (TC3.A.1.7.1)	-	-	No	*pstS*	Part of the ABC transporter complex PstSACB involved in phosphate import
Vaccine	CULC0102_1070	-	Phospholipase/thioesterase	PAI 6	-	No	*pla*	Cutinase
Drug	CULC0102_1442	*dtxR*	Iron-dependent repressor IdeR/DtxR	-	Yes	Yes	*ideR*	repressor
Drug	CULC0102_2084	*fba*	Fructose-bisphosphate aldolase class II	-	No	No	*fbaA*	aldolase
Drug	CULC0102_1536	*pyrH*	Uridine monophosphate kinase	-	No	No	*pyrH*	Catalyzes the reversible phosphorylation ofUMP to UDP
Drug	CULC0102_1260	-	Transcriptional regulator, AcrR family	-	No	No	*acnR*	transcriptional

Supplementary Table S3. Druggable pockets in drug targets with score greater than 0.79predicted using DoGSiteScorer.

DoGSiteScorer results						
GenBank ID	**Name**	**Volume (Angstrom3)**	**Surface Area (Angstrom2)**	**Drug Score**	**Active Residues**	**InterProScan Domain Active Residues**
CULC0102_1442	Iron-dependent repressor IdeR/DtxR	979,26	1337,53	0,79	TYR 16, GLU 19, GLU 20,GLU 21, GLY 22, VAL 23,THP 24, PRO 25, MET 64,ARG 69, THR 70, THR 73,ALA 74, ARG 77, ASP 110, PRO 131, PRO 133,GLY 134, GLU 137, LEU138, SER 158, MET 159,PRO 160, ARG 161, LYS162, VAL 193, ILE 195,SER 205, HIS 206, ASN 207, GLY 208, ASP 210.	GLU83, HIS79, HIS98,MET10, CYS102, GLU105, HIS106, LEU4

CULC0102_1536	Uridine monophosphate kinase (EC 2.7.4.22)	823,24	1117,43	0,85	VAL 212A, LEU 213A, GLY 216A, LEU 217A, ARG 218A, VAL 219A, ALA 220A, ALA 222A, PHE 225A, SER 226A, MET 229A, ASP 166B, ALA 169B, ALA 170B, VAL 181B, MET 183B, LYS 185B, VAL 187B, GLY 189B, VAL 190B, PHE 191B, ALA 192B, GLU 193B, ASP 194B, PRO 195B, VAL 207B, VAL 212B, LEU 213B, GLY 216B, LEU 217B, ARG 218B, VAL 219B, ALA 220B, ASP 221B, ALA 222B, ALA 224B, PHE 225B, SER 226B, CYS 228B, MET 229B, ILE 235B, VAL 237B, THR 257B, VAL 259B.	GLU135,HIS145, GLN132, ALA113, GLY131, ARG123, ARG132, LYS130, ARG126, ARG123, PRO118, LEU120, LEU122,PRO121
CULC0102_2084	Fructose-bisphosphate aldolase classII (EC 4.1.2.13)	478,72	877,8	0,86	ALA 23, ASN 27, GLN 51,SER 53, ASP 95, HIS 96,ASN 210, VAL 211, HIS212, GLY 213, VAL 214,TYR 215, LYS 216, PRO217, GLY 218, ASN 219,VAL 220, LEU 222, PHE251, HIS 252, GLY 253,GLY 254, SER 255, GLY256, SER 257, LEU 258,GLU 261, ASN 274, VAL275, ASP 276, THR 277,ZN 352.	HIS96, HIS212, HIS252
CULC0102_1260	Transcriptional regulator, AcrR Family	721,15	705,54	0,84	SER 162A, ARG 163A, THR 166A, GLY 167A, ALA 168A, ALA 6B, ASP66B, ALA 67B, MET 70B,TRP 95B, MET 96B, ARG99B, LEU 100B, LEU106B, ARG 113B, TRP116B, ILE 117B, GLN120B, LEU 123B, ASP 124B, VAL 127B, ARG 128B, GLU 147B, HIS150B, THR 151B, GLU154B, LEU 157B, ASP158B, ILE 161B, SER162B.	LEU153, LEU131, LEU149, VAL184, MET140, ASN134, ARG130, GLU181, ARG185

Supplementary Figure S1. Ramachandran plot for 3D structure validation of the final 4 drugtargets.



IV. CAPÍTULO II
ARTIGO DE PESQUISA

Neste capítulo, foi feito um primeiro estudo de desenvolvimento in silico de vacinas de subunidades para a prevenção de doenças causadas por C. ulcerans e C. silvaticum. O design de 4 vacinas multi-epítopo foi feito por meio da imunoinformática, a partir das 9 proteínas encontradas na pesquisa do Capítulo I. Experimentos in silico de docking molecular, simulação imune, clonagem e expressão em vetor bacteriano foram realizados e mostraram bons resultados que levaram à proposição de 4 vacinas candidatas para a proteção de doenças causadas por ambas as bactérias citadas.

Pan-vaccinomics approach toward a universal multi-epitope based vaccines candidate against *Corynebacterium ulcerans* and *C. silvaticum*

Janaína Canário Cerqueira[a], Lucas Gabriel Rodrigues Gomes[a], Marcus Vinicius Canário Viana[a], Arun Kumar Jaiswal[a], Thaís Cristina Vilela Rodrigues[a], Siomar Castro Soares[b], Debmalya Barh[a], **Vasco Azevedo**[a]* and Sandeep Tiwari[a]*

O artigo de pesquisa foi submetido na revista **Immunoinformatics** em 25/05/2022.

Pan-vaccinomics approach toward a universal multi-epitope based vaccines candidate against *Corynebacterium ulcerans* and *C. silvaticum*

Janaína Canário Cerqueira[a], Lucas Gabriel Rodrigues Gomes[a], Marcus Vinicius Canário Viana[a], Arun Kumar Jaiswal[a], Thaís Cristina Vilela Rodrigues[a], Siomar Castro Soares[b], Debmalya Barh[a], Vasco Azevedo[a*] and Sandeep Tiwari[a*]

[a]Laboratory of Cellular and Molecular Genetics, Department of Genetics, Ecology, and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[b]Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Minas Gerais, Brazil.

*Corresponding authors
E-mail addresses: sandip_sbtbi@yahoo.com (S. Tiwari); vascoariston@gmail.com (V. Azevedo).

Abstract: Zoonotic diseases have been rising in underdeveloped countries with low vaccination rates, posing risks to human and cattle populations. Diphtheria, one such disease, is mainly caused by *C. ulcerans*, which can also cause other diseases of the respiratory system, against which the current diphtheria vaccine (DTPa) does not protect.The recently described species *C. silvaticum*, of the same genus, is cytotoxic in the humanepithelium and has zoonotic potential. The objective of this research was to use immunoinformatics to propose multi-epitope vaccines for both species, which are composed of selected epitopes from proteins essential to both bacteria. The designed multi-epitope vaccine contains 25 common epitopes from 8 proteins from both species with great predicted immunological properties. These constructs' stability, physicochemical properties, allergenicity, antigenicity, and toxicity were stipulated. In addition, docking and molecular dynamics simulations were performed to certify the link between TLR2 and each constructed vaccine. Immune stimulation and cloning in pET- 28a(+) vector in *E. coli* were performed. The results demonstrated that these four vaccineconstructs might induce innate and adaptive immune responses to protect against *C. ulcerans* and *C. silvaticum*. However, experimental validation is still required to prove the *in silico* models predicted here *in vitro* and *in vivo*.

Keywords: *Corynebacterium,* immunoinformatics, multi-epitope vaccine, pan-vaccinomics, zoonotic pathogen.

1. INTRODUCTION

The genus *Corynebacterium* comprises Gram-positive and coryneform bacteria, with more than 90 described species. About 54 species are of medical importance worldwide, 14 of which are zoonotic bacteria infecting reservoirs, usually mammals, ranging from cattle to pets [1].

C. ulcerans is one of these zoonotic pathogens, the leading cause of diphtheria, even more than *C. diphtheriae*, the principal producer of diphtheria toxin (DT). Some C.ulcerans produce DT, which causes diphtheria in humans, while others produce other virulence factors such as the well-known PLD (Phospholipase D), which kills macrophages. *C. ulcerans* causes respiratory system diseases such as pneumonia, tonsillitis, pharyngitis, and other related diseases and is commonly transmitted from farmanimals. Infection and disease cases have been increasing in underdeveloped countries despite vaccination with the DTPa immunizer to prevent diphtheria, tetanus, and pertussis[2].

C. silvaticum is a new species described in 2020 that infects roe deer, wild boar, and pigs [3]. This bacterium is genetically related to *C. ulcerans* and another pathogen, *C. pseudotuberculosis*. Recent genomic and *in vivo* studies have shown the ability of this organism to infect the human epithelium, demonstrating its zoonotic potential andrelevance to human health [4,5].

The circumstances above point to the economic importance of these pathogens inthe meat, milk, and animal reproduction sectors and consequently their importance in public health, demonstrating the need for their prevention and control. Through computational tools and the use of genomics-based approaches, it is possible to analyze the genomes of these species to answer questions and propose strategies to solve problemsrelated to infectious agents.

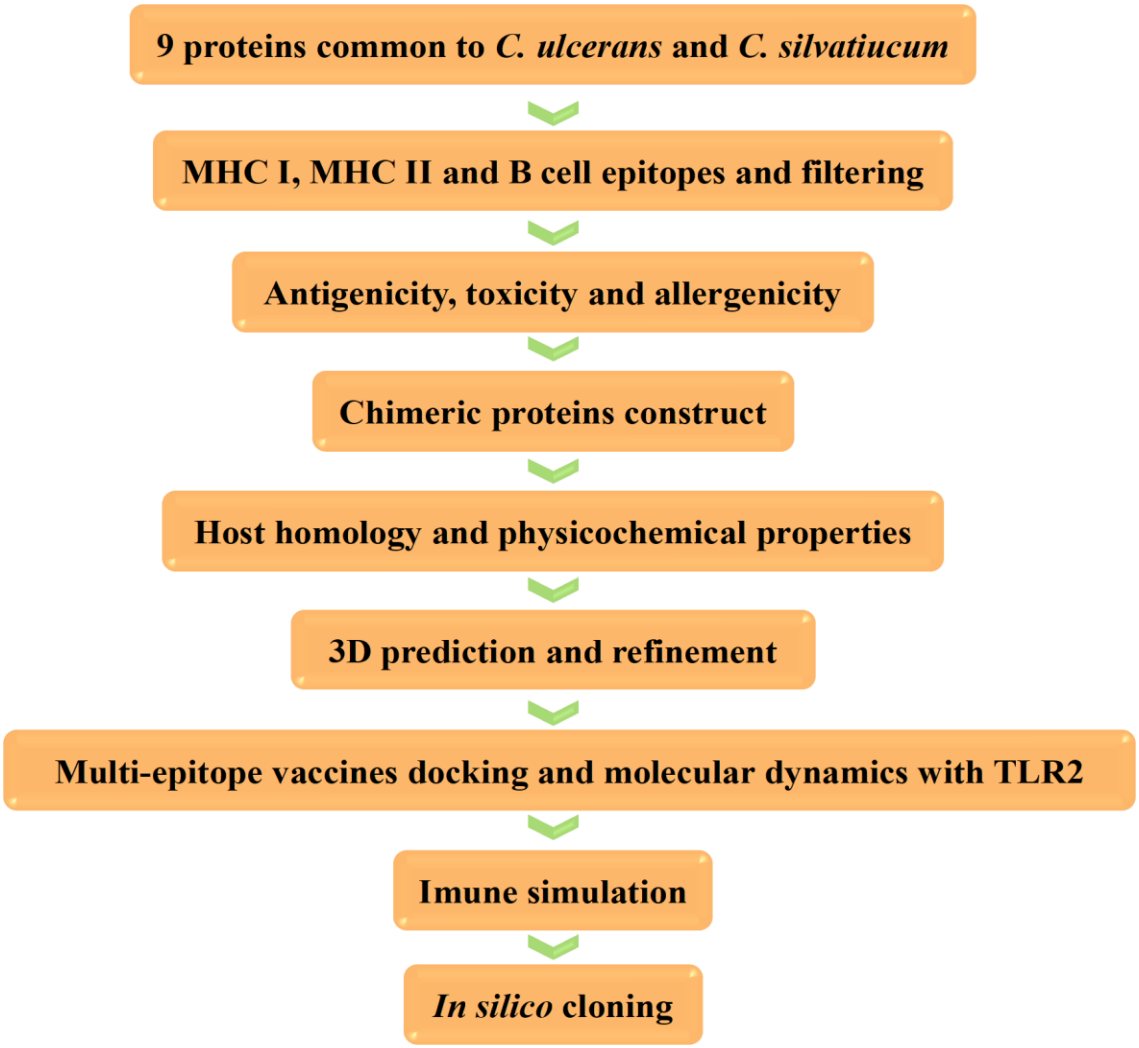
Subtractive genomics and immunoinformatics are *in silico* methods that contribute to vaccinological research and vaccine design. Such methods allow identifyingtarget proteins of drugs and vaccine candidates and constructing chimeric proteins (multi-epitope vaccines) based on these candidates. Reverse vaccinology and immunoinformatics make it possible to propose new vaccines quickly, with the low cost necessary for their development [6,7]. This methodology has been previously used in thedevelopment of control methods for various microorganisms, such as the important causative agents of yellow fever [8], schistosomiasis [9], leprosy [10], pneumonia [11], and syphilis [12].

The objective of this work was to propose four multi-epitope vaccines for the control and prevention of zoonoses caused by *C. ulcerans* and *C. silvaticum.* These vaccines have been

designed with immunogenic epitopes derived through vaccine candidates common to both species and have displayed good immunogenic potential. However, the experimental validation of these constructs *in vivo* and *in vitro* is still necessary to ensure their function.

2. MATERIALS AND METHOD

The overall workflow used in this study is shown in **Figure 1**.



2.1 Selection and retrieval of core antigenic proteins

The nine core genomes, non-host homologous proteins, and antigenic properties derived from *C. ulcerans* and *C. silvaticum* strains were selected from our previous work (Cerqueira et al. 2022, unpublished), in which common drug targets and vaccine candidates from these

species were identified using 108 genomes of both species. The characteristics of these nine vaccine candidates' proteins are shown in Table 1. The FASTA sequences of these proteins were retrieved from the National Center for Biotechnology Information (NCBI) database.

Table 1. Proteins selected for the multi-epitope analysis

Name	**NCBI ID**
ABC transporter, substrate-binding protein (cluster 14, Mn/Zn)	WP_013910564
Integral membrane protein	WP_013910876
ABC transporter, substrate-binding protein (cluster 14, Mn/Zn)	WP_046693528
Hypothetical protein	WP_014836045
Phospholipase/thioesterase	WP_014836279
Hypothetical protein	WP_014836377
Hypothetical protein	WP_014525496
Phosphate ABC transporter PstS	WP_014836748
Glutamate ABC transporter, permease protein 1 GluC	WP_013911078

2.2 Prediction of B cell, MHC class I, and MHC class II epitopes

We used the server ABCpred [13] for the B cell epitope prediction and the Immune Epitope Database and Analysis Resource (IEDB) server [14] for the MHC classI and class II T cell epitope predictions based on mouse alleles for the nine proteins sequences previously selected (**Table 1**). The data were filtered based on Ic50 lower than500nM and percentile rank lower than 1.0 for MHC class I and lower than 5.0 for MHC class II epitopes. Default parameters were used in both predictors.

2.3 Epitope screening

The selected epitopes from IEDB and APCpred were then filtered based on antigenicity, allergenicity, and toxicity by the software VaxiJen v2.0 [15], AllerTOP 2.0server [16], and ToxinPred [17], respectively. Afterward, the screened peptides were subjected to an overlap analysis performed by an *in-house* python script made by our research group. The overlapping was made for MHC-I epitopes against B cell epitopes and MHC-II against B cell epitopes, and redundant epitopes were excluded.

2.4 Construction of the chimeric protein (vaccine)

Four adjuvants (**Table 2**) were selected to increase the immunogenicity of multi-epitope vaccines and added to the chimeric protein sequence. The adjuvants were 50s ribosomal L7/L12 protein, Beta-defensin, HBHA protein (*M. tuberculosis*), and HBHA conserved sequence [18]. We constructed four chimeric proteins (V1, V2, V3, and V4), each with an adjuvant sequence. For the vaccine construct, linker sequences were used for joining the MHC-I/B and MHC-II/B epitopes: AAY and GPGPG, respectively. The EAAAK linker was used to attach the adjuvant to the first epitope of the chimeric protein[11]. We also placed a histidine tag (HHHHHH) at the end of the sequences.

2.5 Prediction of antigenicity, allergenicity, solubility, and host homology of the vaccine construct

The solubility of these four vaccines was calculated using Protein-Sol [19] and SOLpro [20] servers. The AllerTOP 2.0 server [16] was used to predict allergenicity andVaxiJen 2.0 [15] for antigenicity prediction with a threshold of 0.4 (for the bacteria model). Homology analysis with typical host proteomes (*Bos taurus* taxid:9913, *Sus scrofa* taxid:9823, *Ovis aries* taxid:9940, *Equus ferus caballus* taxid:9796, *Capra aegagrus hircus* taxid:9925 and *Homo sapiens* taxid: 9606) was performed by BlastP at the National Center for Biotechnology Information (NCBI) with standard databases and "BLASTp" algorithm.

2.6 Prediction of tertiary structure and refinement

We used AlphaFold v2.0 Colab [21] to predict the tertiary structure of the vaccine.After that, we used the GalaxyRefine available at the GalaxyWEB server [22] to make the refinement in the 3D structure. To analyze the quality of the structure of the chimericprotein, we performed a Ramachandran plot analysis using the PROCHECK program available on the SAVES web server [23].

2.7 Molecular docking between the chimeric protein and the TLR2

We performed molecular docking with TLR2 (Toll-Like Receptor) and the multi-epitope vaccines. The TLR2 is commonly used in *Corynebacterium* research. The TLR structure (**PDB ID: 2Z81**) was retrieved from the PDB database, and the Chimera software [24] was used to

remove water molecules, ligands, and other additional molecules. We performed molecular docking between the TLR2 and each chimeric protein using the ClusPro web server [25]. To determine and visualize the interactions inthe complex, we used LigPlot+ [26].

2.8 Molecular dynamic simulation of the receptor-ligand complex

The iMODS server was used to evaluate the flexibility and stability of vaccine- TLR2 complexes. The program uses Normal Mode Analysis (NMA) to predict dynamic characteristics such as eigenvalue, deformability, and protein variance [27].

2.9 Vaccine Immuno simulation

We used the CImmSim server [28] to evaluate the immune response of the multi-epitope vaccines. To this end, three injections were given over four weeks (each one). Each inoculum carried 1000 vaccine proteins and the periods were 1, 84, and 168 (each period means 8h in real life, and time '1' is the initial inoculation time = 0). The simulationsteps of 1050 were used, keeping the other parameters as default.

2.10 *In silico* cloning in plasmid

We carried out *in silico* cloning to investigate the cloning and expression potentialof the vaccines in a plasmid vector. The peptide sequences of the vaccines were convertedinto their coding DNA sequences, optimizing codon usage for expression in *Escherichiacoli*. The Java Codon Adaptation Tool (JCat) [29] program was used for this adaptation,using default settings and selecting the *E. coli* K12 strain. To ensure the quality of expression in the *E. coli* expression system, the Codon Adaptation Index (CAI) of the sequences should be above 80%, and the GC content rate should be between 30 to 70%. For the *in silico* cloning, the SnapGene software (from Insightful Science, available at snapgene.com) was used, and the *E. coli* pET28a(+) vector was selected as the sequence.Restriction sites for the *SacI* and *BamHI* restriction enzymes were inserted into the coding sequences of the vaccine constructs, and the optimized sequences were inserted into thevector through the SnapGene software to ensure the expression of the vaccine.

3. RESULTS

3.1 Predicted and filtered MHC class I, MHC class II and cell B epitopes

We predicted epitopes in the IEBD (MHC I: 20648 and MHC II: 7581) and ABCpred (B cell: 265) predictors for the 9 proposed proteins. The predictions were basedon mouse alleles, with a size of 9 amino acids for MHC and a size of 16 for B cells. Epitopes with Ic50 values below 500nM were excluded. Afterward, the epitopes with percentile rank values below 1.0 and 5.0 were selected for MHC I and MHC II. 1198 epitopes (B/MHC-I/MHC-II) after these filters remained. These 1198 epitopes were further analyzed to filter based on criteria antigenicity, non-allergenicity, overlapping between the filtered epitopes of MHC II against B cells and MHC II against B cells, followed by removal of redundancy to get the best epitopes.

3.2 Antigenicity, allergenicity, and toxicity analysis

Antigenicity, allergenicity, and toxicity analyzes were performed using standard parameters. Epitopes that did not fit these criteria were excluded, the remaining 120. Overlapping between the filtered epitopes of MHC II against B cells and MHC II againstB cells was performed to remove redundant peptides, decreasing the number of 76 epitopes. Similar peptides remained after the overlapping; they were excluded based on lower percentile rank. The integral membrane protein (WP_013910876) did not pass theselected filter criteria and was excluded from further analysis. As a result, the final immunogenic epitopes of 18 MHC-I and 7 MHC-II epitopes from eight proteins were used for the final candidate's multi-epitope vaccine construction (**Table 2**).

Table 2. Final epitopes sequences for the vaccine construct.

Proteins	MHC I/B epitopes	Percentile rank	MHC II/B epitopes	Percentile rank
ABC transporter, substrate-binding protein (cluster 14, Mn/Zn)	DTYTGALRY	0.57	VPVIFQDASANPQAV	1.62
ABC transporter, substrate-binding protein (cluster 14, Mn/Zn)	PEFTHDPHI	0.45		
Hypothetical protein	CVPLFHAAV	0.45	GTPYQWGGTSTAGFD	0.89
	AALLLPRVF	0.48	NPEPLEPNTPNVNES	1.65
	GTQVTHEQL	0.86		
Phospholipase / thioesterase	GSKTGFDRF	0.52	VGIPAAGAQPAQCPR	2.25

	NSATTTTGY	0.58		
	KVGAPIWWL	0.87		
Hypothetical protein	ARVNNEHTI	0.42	STSGVQRTAIAGLLA	1.81
	SATGDGLQI	0.98		
	LPCPDGVHV	0.60		
	ASDTIGGQL	0.92		
Hypothetical protein	GIPLAIANI	0.91	STSSSNPLAG	2.60
	VIWFLVAGL	0.39		
Phosphate ABC transporter PstS	YDQTTRDQV	0.81	ETGSSSGAGAST	2.75
	EKDAYPLVL	0.86		
	LPFVIGPVA	0.89		
Glutamate ABC transporter, permease protein 1 GluC	RSGINTVDF	0.29		

3.3 Multi-epitope vaccine construct and host homology

Four chimeric proteins were constructed using identical epitope sequences withdifferent adjuvants (**Figure 2**). The four adjuvants used were: 50s ribosomal L7/L12 protein, Beta-defensin, HBHA protein (*M. tuberculosis*), and HBHA conserved sequence.Each adjuvant was followed by the EAAAK linker and then the epitopes. AAY peptideslinked the 18 CTL/B epitopes, GPGPG linked the 7 HTL/B epitopes, and 6 tagged histidine residues were added at the end. The schematic of these four multi-epitope vaccine constructs is shown (V1: 518 amino acids; V2: 509 amino acids; V3: 404 aminoacids, and V4: 489 amino acids) in Figure 2. In the host homology analysis, all four multi-epitope vaccine constructs did not show any homology with the host proteome.

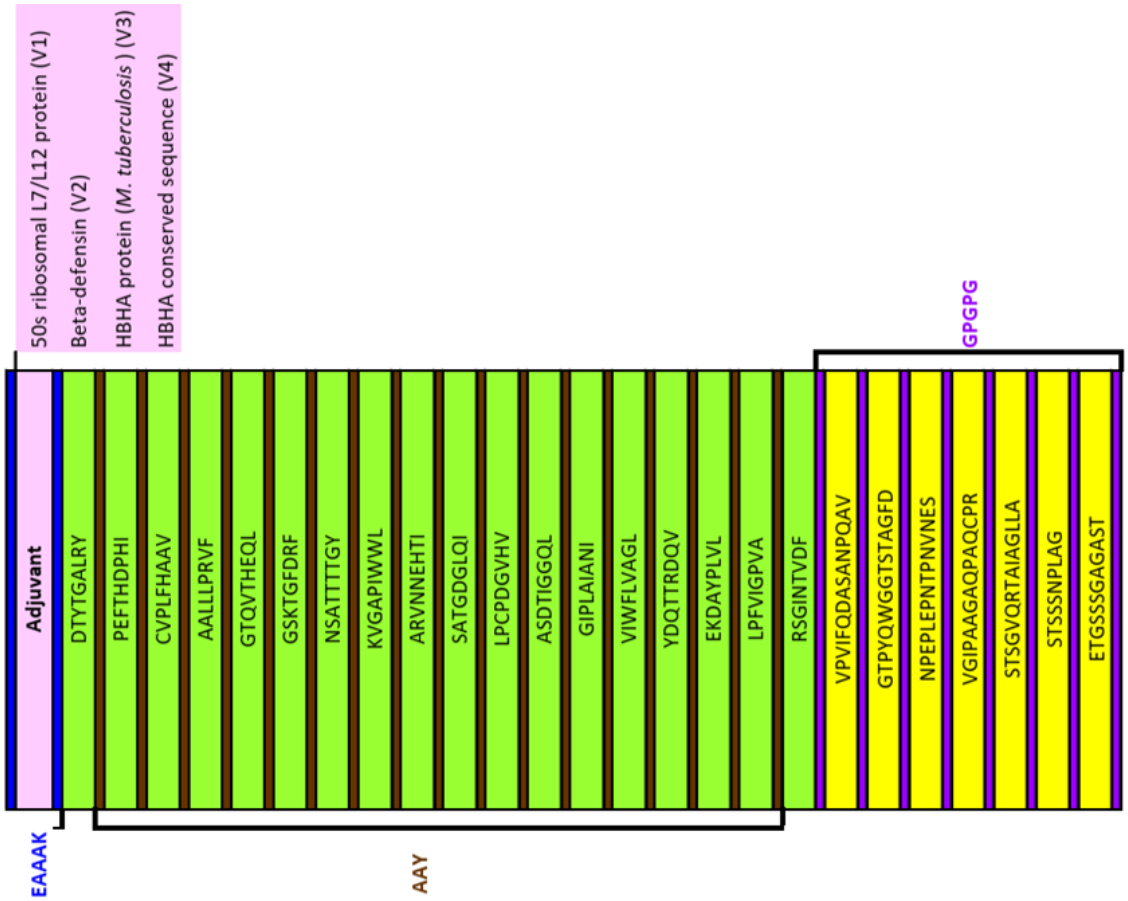


Figure 2. Construction of multi-epitope vaccines. The green boxes represent the MHC-Iepitopes, and the yellow boxes represent the MHC-II epitopes. In the boxes are the epitopes interspersed with the linkers; in the first box, each adjuvant is shown laterally (light pink).

3.4 Physical-chemical and structural properties

The ProtParam server analyzed the physicochemical properties of constructed multi-epitope vaccines. The vaccines showed a molecular weight between 41 to 54 kDa (Table 4). All vaccines are non-allergenic, antigenic, non-toxic, and soluble. The V3 construct had a pI value of 8.08, which differs from the other 3 constructs with similar pIvalues, between 4.98-5.06. V1, V2, and V3 showed negative GRAVY results indicating they are hydrophilic and capable of interacting with water molecules. V4 showed a positive GRAVY value but very close to zero. The instability index was below 40.0 for all vaccines, which indicates stability in biological processes. The aliphatic index (AI) ranged from 72.75 to 81,78, indicating high-temperature stability.

We obtained the best models from AlphaFold and subjected them to refinement. As shown in figure 3, the four vaccines showed good structural quality with expected values - of residues above 90% in more favorable regions (Table 3). After refinement, V1 showed 90.8%

of the residues in the most favorable regions and 1.6% in the disallowed regions, V2 showed 92.8% of the residues in the most favorable regions and 0.5% in thedisallowed regions, V3 showed 92.7% of the residues in the most favorable regions and 0.9% in disallowed regions, and V4 showed 96.5% of residues in more favorable regionsand 0.5% in disallowed regions.

Table 3. Physicochemical properties of multi-epitope vaccines. **MW**: Molecular Weight; **pI**: isoelectric pH; **GRAVY**: Grand Average of Hydrophathicity. NA = non-allergen.

Name	Antigenicity	Solubility	MW	pI	Allergenicity	GRAVY	AI
V1	0.7995	0.350/0.979565	54012.04	5.06	NA	-0.152	79.02
V2	0.8046	0.390/0.989099	52893.81	4.98	NA	-0.127	80.96
V3	0.8747	0.287/0.829559	41544.59	8.08	NA	-0.078	72.75
V4	0.7635	0.511/0.984098	49823.87	5.05	NA	0.048	81.68

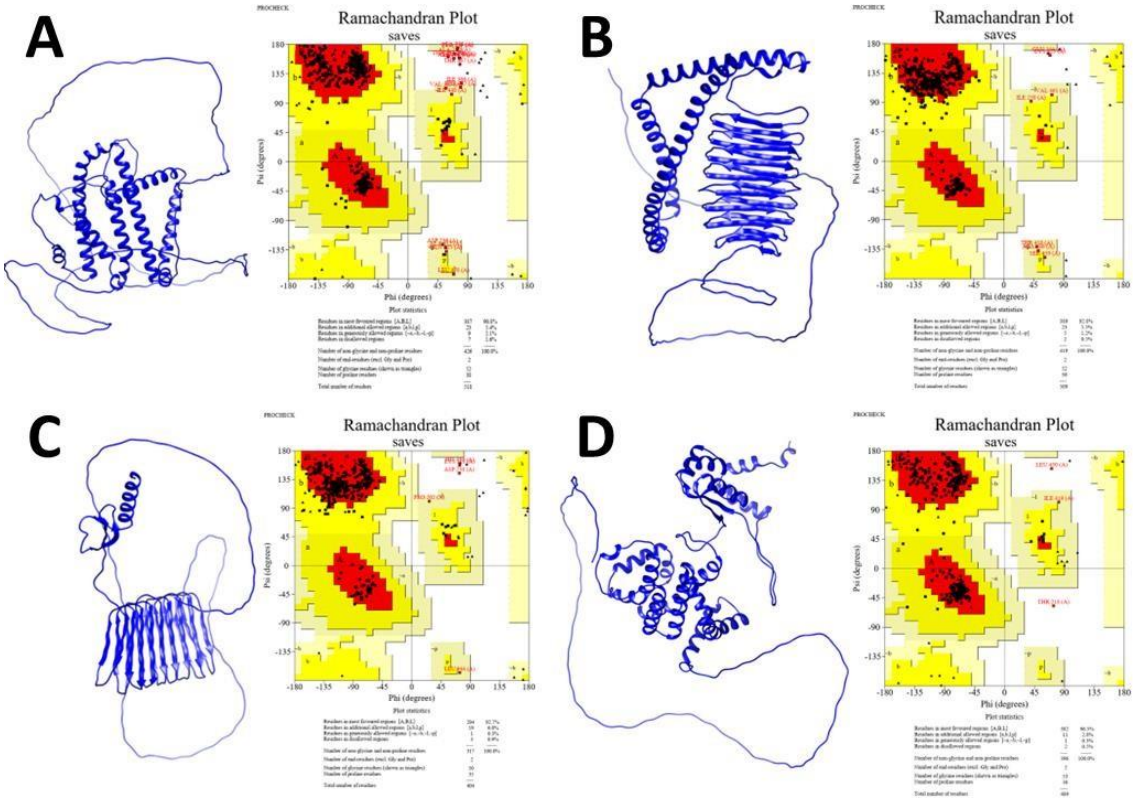


Figure 3. The refined 3D structure and Ramachandran plot of multi-epitope vaccine construct. **A**: The Ramachandran plot for V1 construct after refinement shows 90.8% residues in the most favored regions and 1.6% in disallowed regions.; **B**: The Ramachandran plot for V2 construct after refinement, showing 92.8% residues in most favored regions and 0.5% in disallowed regions.; **C**: The Ramachandran plot for V3 construct after refinement shows 92.7% residues in most favored regions and 0.9% in disallowed regions.; **D**: The Ramachandran plot for V4 construct after refinement shows96.5% residues in most favored regions and 0.5% in disallowed regions.

3.5 Molecular docking with TLR2 and MD simulation

After the docking analysis, the UCSF Chimera software was used to analyze the top 5 democratic docking models for each docked muti-epitope with a TLR2 complex. The best-docked complex was selected based on the lowest binding energy (stable), the highest number of hydrogen bonds, hydrophobic interaction, and other intermolecular interactions. The docking result showed binding energy ranging from -1230.7 to -1410.7 with the highest number of intermolecular interactions, which suggests a good interaction between TLR2 multi-epitope vaccine constructs (**Table 4, Figures 4 and 5**). Furthermore, these best complexes were subjected to the iMOD server for the moleculardynamic simulation analysis. The Normal Mode Analysis (NMA) was used, a computermethod for analyzing the flexibility of the protein. The B-factor graph depicts the complex's link between the Normal Mode Analysis and PDB regions. The eigenvalues ofall docked complexes (**Table 4 and Supplementary Figure 2)**.

Table 4. The docked properties of the best-selected vaccine -TLR2 complexes and iMODeigenvalues.

Vaccine	Energy	Hydrogen Bonds	Hydrophobic BondsTLR2	Hydrophobic Bonds Vaccine	Note	Eigenvalues
V1	-1410.7	18	25	18	2 salt bridge	3.895739 e-09
V2	-1230.7	36	22	20	2 salt bridge	1.892432 e-06
V3	-1284.8	6	25	8	NA	4.953906 e-08
V4	-1372.6	16	32	13	2 salt bridge	3.509202 e-07

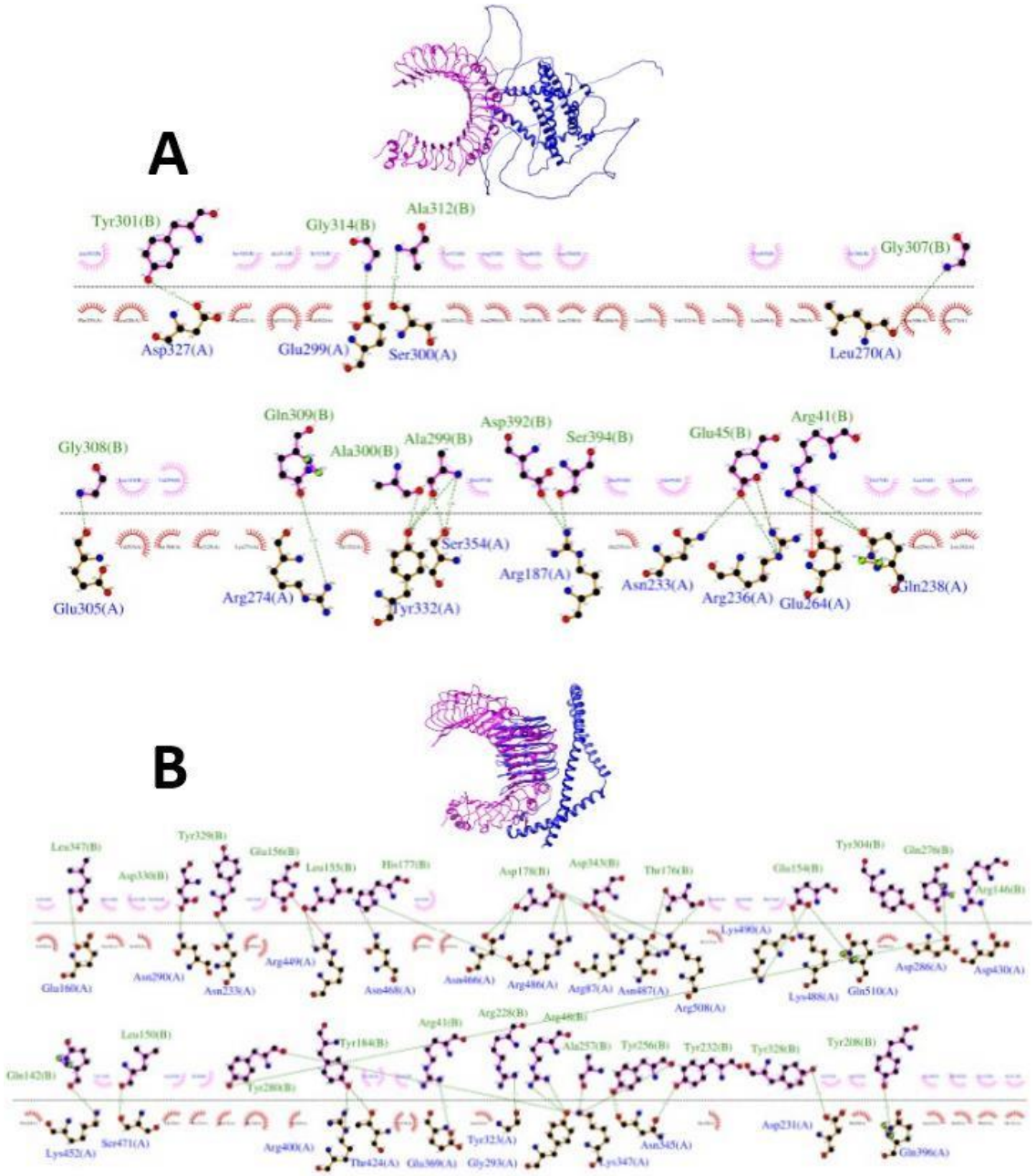


Figure 4. 2D and 3D representation of TL2-vaccine interactions. **A:** V1-TLR2 complex showed 18 hydrogen bonds. Also, 18 residues of V1 and 25 residues of TLR2 are involvedin the hydrophobic interactions, and 2 residues for salt bridge formation. **B**: V2-TLRD complex showed 36 hydrogen bonds. Also, 20 residues of V2 and 22 residues of TLR2 are involved in the hydrophobic interactions, and 2 residues for salt bridge formation.

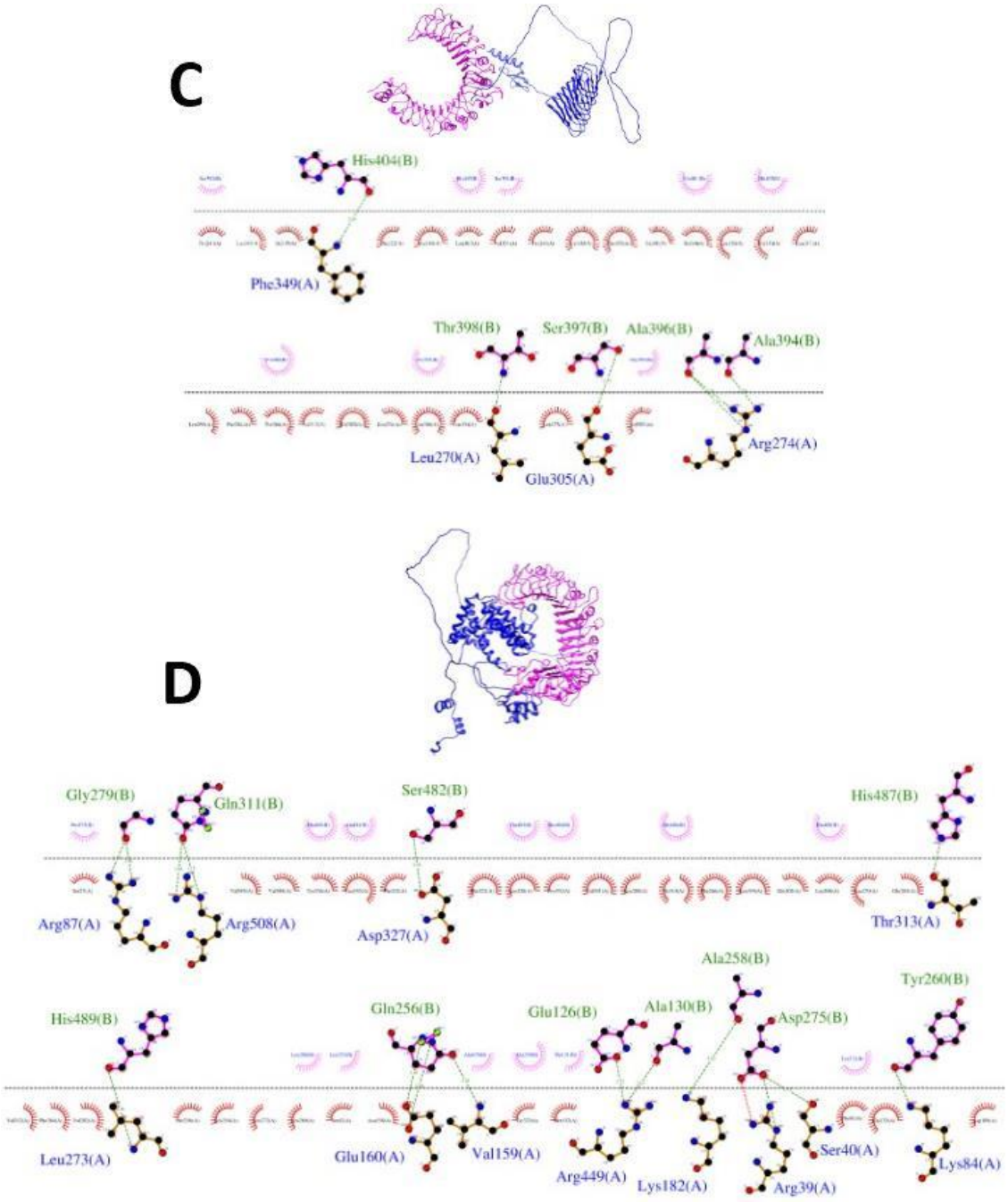


Figure 5. 2D and 3D representation of TL2-vaccine interactions. **C**: V3-TLR2 complex showed 6 hydrogen bonds. Also, 8 residues of V3 and 25 residues of TLR2 are involvedin the hydrophobic interactions. **D**: V4-TLR2 complex showed 16 hydrogen bonds. Also,13 residues of V4 and 32 residues of TLR2 are involved in the hydrophobic interactions,and 2 residues for salt bridge formation.

3.6 Vaccines Immune simulation

The Immune simulation analysis showed the four vaccines to induce immune responses with similar results. The simulation result for the multi-epitope vaccine (V1) isshown in figure 6. The result for other multi-epitope vaccines is shown in **supplementarymaterial figures 3-5**.

The complexes were able to activate and increase B cell and T cell populations after the first injection, showed in graphs A-C. Peak activation and populationgrowth of B and T cells are observed between 50 and 100 days after injections.

The B cells differentiated into isotypes for immunoglobulin expression andmemory cells (total in black line), which supposedly confirms the recognition of immunogens by the defense cells (Figure 6A). In the same figure in B, a significant increase in active T cells cytotoxic (TC) (pink) is observed after the first injection until about day 100, in contrast to the low number of deactivated cells (blue). After this periodof infection, activated cells decrease over time. Helper T cells (TH) are also in high concentrations (total black). Memory TH cells decrease a little over time but maintain a certain blood concentration even after 350 days. TH cells help activate other defense cellssuch as B cells, TC cells, and macrophages. In D, activation and increase of active macrophages are also observed between days 50 and 100 (blue), and their concentration is maintained for an extended period (after 350 days). The total number of macrophages (black) is increased from the first inoculum.

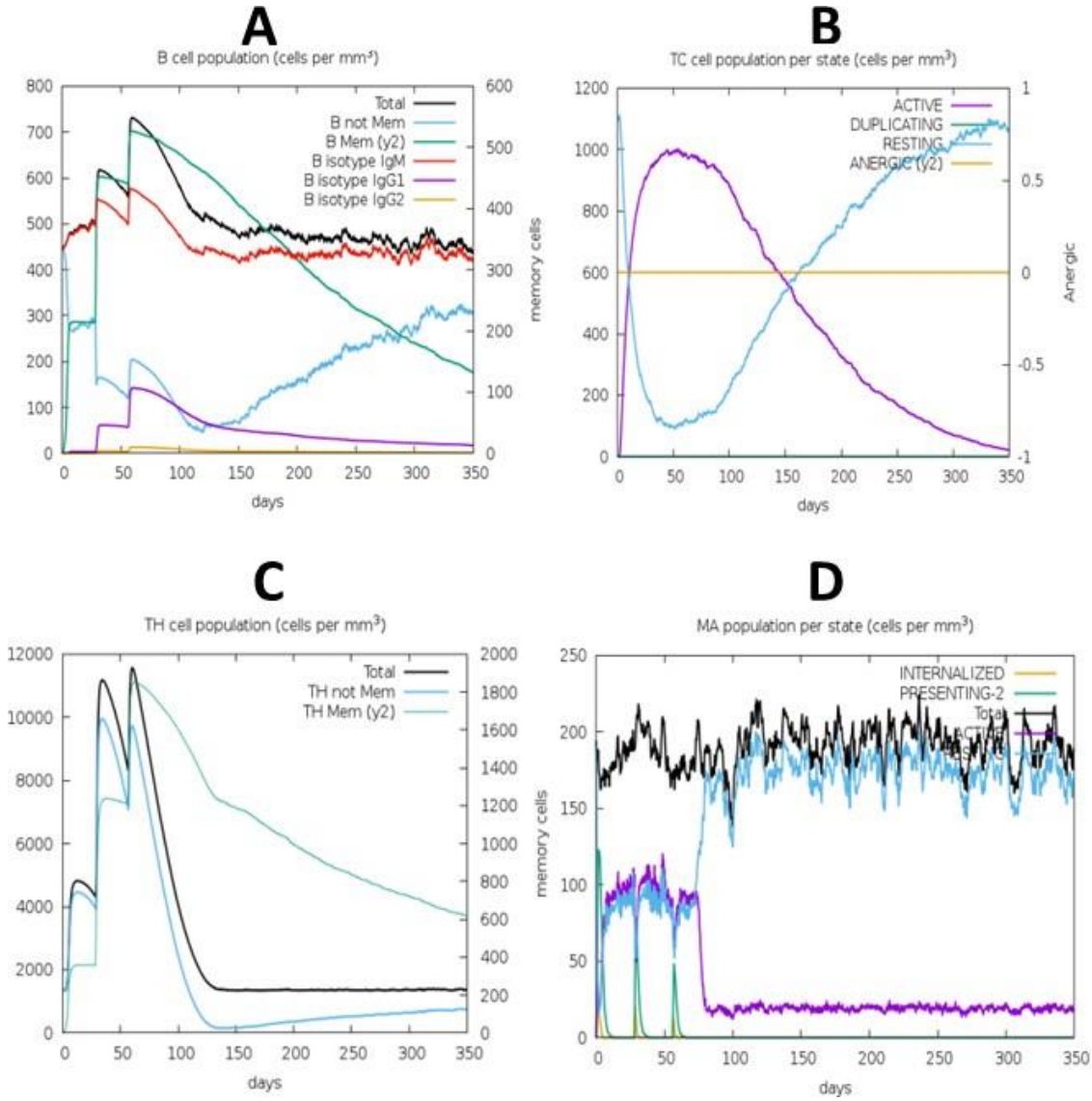


Figure 6. Immune simulation results show (Vaccine construct V1) the amount (cells permm^3) and activation of the leading defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cell population; macrophages population per state.

3.10 *In silico* cloning in plasmid

Codon adaptation for expression in *E. coli* (strain K12) was performed through the JCat program. The CAI score of all vaccine constructs was 1.0. The GC content for Vaccine 1 was 57.66, for Vaccine 2 was 56.97, for Vaccine 3 was 57.84, and for Vaccine4 was 56.03. The optimized sequences' GC content and CAI scores were within the optimal range. Restriction sites for *SacI* and *BamHI* were added to the 5' and 3' ends of the inserts, respectively, and the sequence was cloned into the pET28a(+) vector.

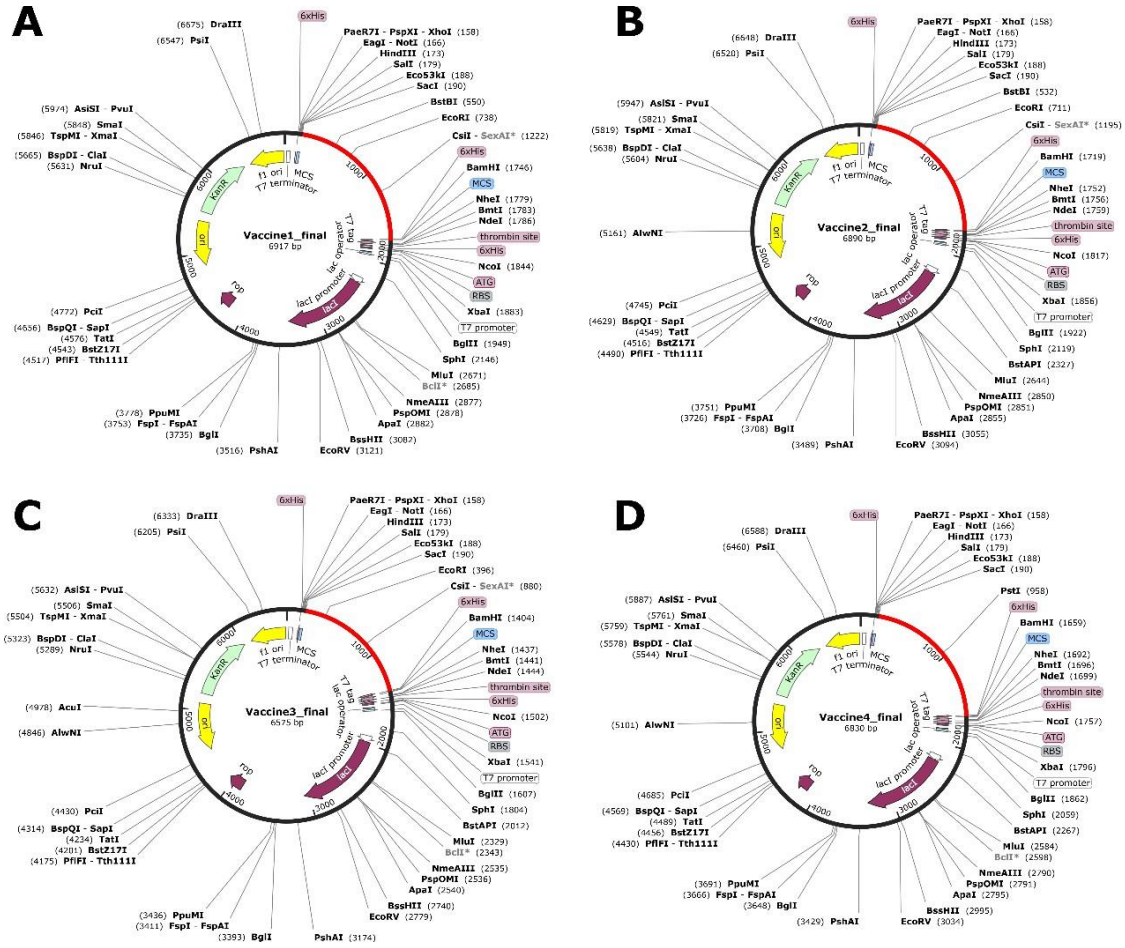


Figure 7. *In silico* cloning circular map. DNA sequences of vaccines (red) in pET28a(+)insertion in a vector of *E. coli* K12. A; B, C, D

4. DISCUSSION

The immunoinformatics approach produces a hetero-valent vaccine capable of targeting multiple pathogens/strains, reducing the risk of the pathogen escaping the immune system [30]. Zoonotic diseases have been rising in underdeveloped countries with low vaccination rates, posing risks to human and cattle populations [31]. *C. silvaticum* has just been described, and a study has already shown that this bacterium caninfect human epithelial cells, in addition to having been previously isolated from infectious lesions of pigs and roe deer. Genomic studies already showed the zoonotic potential of this species, which was inserted in the clade called *C. diphtheriae* complex [4,5]. Some strains of *C. ulcerans* are *tox* positive (producing DT), but many strains are pathogens causing other diseases common in the respiratory system, and it is a zoonotic pathogen of economic, veterinary, and public health importance [2]. These traits

indicate the necessity of developing preventative and diagnostic methods for *C. ulcerans* and *C. silvaticum*. A previous comparative genomics study performed with *C. ulcerans* (72 genomes) and *C. silvaticum* (36 genomes) strains [32] has revealed that the transmembrane and secreted proteins represent suitable vaccine candidates. These proteins are localized in the outer membrane, accessible to the host immune system, are known to be associated with virulence, and are crucial for the survival of bacteria. Here we applied an immunoinformatics strategy to predict the epitopes based on proteins common to all strains of both species. The designed multi-epitope vaccine contains 25 common epitopes from 8 proteins from both species with great predicted immunologicalproperties. Appropriate linkers connected the selected epitopes. The used linkers were experimentally confirmed in antigen presentation assays and have been reported to function as cleavage sites for different sets of immune cells and comprise stereo-chemicalproperties required for protein stability [33].

The final four multi-epitope vaccines proved to be potent immunogenic properties, non-host homologous, non-toxic, and non-allergen with a high aliphatic index,which shows their stability in a wide range of temperatures. In the protein-protein dockingand molecular dynamic simulation analyses, we observed that all four multi-epitope vaccines demonstrated stable intermolecular interactions with TLR 2. TLR2 and TLR4 are widely used in immunological studies of gram-positive and gram-negative bacteria, as they recognize many components of the bacterial cell wall. TLR2 has a major functionin recognition of gram-positive bacteria, is known to sense mycolic acids (glycolipids) from the cell wall of corynebacteria, and is involved in the activation of macrophages (Takeuchi et al. 1999; Schick et al. 2017). in addition, the physicochemical features for these multi-epitope candidates suggest stable, soluble and heterologous expression.

The immune simulation suggests the vaccine constructs may have the potential for the induction of humoral and cellular immune responses. These immune responses are already known in *C. ulcerans* and its closely related species, such as *C. pseudotuberculosis* and *C. jeikeium* [36,37]. Corynebacteria induce innate and adaptive immune responses in humans and animals by recruiting and activating B and T cells andmacrophages. An *in vivo* study in a murine model demonstrated that *C. ulcerans* causes macrophage immune responses. Both toxigenic and non-toxigenic strains of *C. ulcerans*cause infections in the respiratory system, kidneys, spleen, and peripheral blood. As a result of this spread, it causes systemic infection [38]. The *in silico* immune simulation using the C-ImmSim server showed increased immunoglobulins observed, indicating theproduction of memory B-cells.

These immunogens can be used as multi-epitope subunit vaccines to control different diseases caused by C. ulcerans and prevent C. silvaticum infections. They can be used in diagnostic analysis at any stage of disease or in possible asymptomatic cases of both bacteria [39]. Interestingly the level of T-helper lymphocytes and the activation of cytotoxic T lymphocytes significantly support that the vaccine candidates have great potential to elicit humoral and adaptive immune responses.

CONCLUSION

In this study, we built four multi-epitope vaccines based on eight common proteinsof *C. ulcerans* and *C. silvaticum*, with immunogenic properties and the ability to induce innate and adaptive immune responses. *In silico* approaches such as genomics and immunoinformatics allow these experiments to be carried out without financial costs andin a short time. The limitation of computational tools is that they will never imitate nature,which is why experimental validations in vitro and in vivo are always necessary, but whenreaching these final stages, the financial investment and time will be much lower.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

We acknowledge the collaboration and assistance of all team members, thePrograma Interunidades de Pós-graduação em Bioinformática da UFMG and the Brazilian funding agencies CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais).

SUPPLEMENTARY MATERIAL

Supplementary Figure 1. Three-dimensional structure modelling by Alphafold and Ramachandran plots for the initial models (Before refinement).

Supplementary Figure 2. Molecular dynamics simulation, Normal Mode Analysis. The figure reprsents Eigenvalue of vaccine-TLR2 complex (**A**: V1; **B**: V2; **C**: V3 and **D**: V4).

Supplementary Figure 3. Immune simulation results show (Vaccine construct V2) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.

Supplementary Figure 4. Immune simulation results show (Vaccine construct V3) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.

Supplementary Figure 5. Immune simulation results show (Vaccine construct V4) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.

REFERENCES

[1] Bernard K. The genus Corynebacterium and other medically relevantcoryneform-like bacteria. J Clin Microbiol 2012;50:3152–8. https://doi.org/10.1128/JCM.00796-12.

[2] Hacker E, Antunes CA, Mattos-Guaraldi AL, Burkovski A, Tauch A. Corynebacterium ulcerans, an emerging human pathogen. Future Microbiol 2016;11:1191–208. https://doi.org/10.2217/fmb-2016-0085.

[3] Dangel A, Berger A, Rau J, Eisenberg T, Kämpfer P, Margos G, et al. Corynebacterium silvaticum sp. Nov., a unique group of NTTB corynebacteria inwild boar and roe deer. Int J Syst Evol Microbiol 2020;70:3614–24. https://doi.org/10.1099/ijsem.0.004195.

[4] Viana MVC, Profeta R, da Silva AL, Hurtado R, Cerqueira JC, Ribeiro BFS, etal. Taxonomic classification of strain PO100/5 shows a broader geographic distribution and genetic markers of the recently described Corynebacterium silvaticum. PLoS One 2020;15:e0244210. https://doi.org/10.1371/journal.pone.0244210.

[5] Möller J, Busch A, Berens C, Hotzel H, Burkovski A. Newly Isolated AnimalPathogen Corynebacterium silvaticum Is Cytotoxic to Human Epithelial Cells 2021. https://doi.org/10.3390/ijms22073549.

[6] Bagnoli F, Baudner B, Mishra RPN, Bartolini E, Fiaschi L, Mariotti P, et al.Designing

the next generation of vaccines for global public health. Omi A J Integr Biol 2011;15:545–66. https://doi.org/10.1089/omi.2010.0127.

[7] Oli, Angus Nnamdi; Odimegwu DC, Okoyeh JN, Emechebe GO, Adejumo SA,Ibeanu GC. Immunoinformatics and Vaccine Development : An Overview 2020:13–30. https://doi.org/10.2147/ITT.S241064.

[8] Tosta SF de O, Passos MS, Kato R, Salgado Á, Xavier J, Jaiswal AK, et al.Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches. J Biomol Struct Dyn 2021;39:219–35. https://doi.org/10.1080/07391102.2019.1707120.

[9] Sanches RCO, Tiwari S, Ferreira LCG, Oliveira FM, Lopes MD, Passos MJF, etal. Immunoinformatics Design of Multi-Epitope Peptide-Based Vaccine Against Schistosoma mansoni Using Transmembrane Proteins as a Target. Front Immunol 2021;12:1–16. https://doi.org/10.3389/fimmu.2021.621706.

[10] Jaiswal AK, Tiwari S, Jamal SB, de Castro Oliveira L, Sales-Campos H, Andrade-Silva LE, et al. Reverse vaccinology and subtractive genomics approaches for identifying common therapeutics against Mycobacterium leprae and Mycobacterium lepromatosis. J Venom Anim Toxins Incl Trop Dis 2021;27. https://doi.org/10.1590/1678-9199-JVATITD-2020-0027.

[11] Vilela Rodrigues TC, Jaiswal AK, Lemes MR, da Silva MV, Sales-Campos H, Alcântara LCJ, et al. An immunoinformatics-based designed multi-epitope candidate vaccine (mpme-VAC/STV-1) against Mycoplasma pneumoniae. Comput Biol Med 2022;142. https://doi.org/10.1016/j.compbiomed.2021.105194.

[12] Gomes LGR, Rodrigues TCV, Jaiswal AK, Santos RG, Kato RB, Barh D, et al.In Silico Designed Multi-Epitope Immunogen "Tpme-VAC/LGCM-2022" MayInduce Both Cellular and Humoral Immunity against Treponema pallidum Infection. Vaccines 2022;10:1019. https://doi.org/10.3390/vaccines10071019.

[13] Saha, Sudipto; Raghava GPS. Prediction of Continuous B-Cell Epitopes in anAntigen Using Recurrent Neural Network. Wiley Intersci 2006;65:40–8. https://doi.org/10.1002/prot.21078.

[14] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. TheImmune Epitope Database (IEDB): 2018 update. Nucleic Acids Res 2019;47:D339–43. https://doi.org/10.1093/nar/gky1006.

[15] Flower DR, Doytchinova I, Zaharieva N, Dimitrov I. Immunogenicity Predictionby VaxiJen: A Ten Year Overview. J Proteomics Bioinform 2017;10:298–310. https://doi.org/10.4172/jpb.1000454.

[16] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2 - A server for insilico prediction of allergens. J Mol Model 2014;20. https://doi.org/10.1007/s00894-014-2278-5.

[17] Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLoS One 2013;8. https://doi.org/10.1371/journal.pone.0073957.

[18] Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. Sci Rep 2019;9:1–19. https://doi.org/10.1038/s41598-019-41496-4.

[19] Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: A web tool for predicting protein solubility from sequence. Bioinformatics 2017;33:3098–100. https://doi.org/10.1093/bioinformatics/btx345.

[20] Magnan CN, Randall A, Baldi P. SOLpro: Accurate sequence-based prediction of protein solubility. Bioinformatics 2009;25:2200–7. https://doi.org/10.1093/bioinformatics/btp386.

[21] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.

[22] Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven byside-chain repacking. Nucleic Acids Res 2013;41:384–8. https://doi.org/10.1093/nar/gkt458.

[23] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91. https://doi.org/10.1107/S0021889892009944.

[24] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis.J Comput Chem 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.

[25] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusProweb server for protein-protein docking. Nat Protoc 2017;12:255–78. https://doi.org/10.1038/nprot.2016.169.

[26] Roman AL, Mark BS. LigPlot+: Multiple Ligand-Protein Interaction Diagramsfor Drug Discovery. J Chem Inf Model 2011;51:2778–86.

[27] López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: Internal coordinates normal mode analysis server. Nucleic Acids Res 2014;42:271–6. https://doi.org/10.1093/nar/gku339.

[28] Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in thesimulation of the immune system. PLoS One 2010;5. https://doi.org/10.1371/journal.pone.0009862.

[29] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:526–31. https://doi.org/10.1093/nar/gki376.

[30] Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide vaccine:Progress and challenges. Vaccines 2014;2:515–36. https://doi.org/10.3390/vaccines2030515.

[31] Tam JP. Synthetic peptide vaccine design: Synthesis and properties of a high-density multiple antigenic peptide system. Proc Natl Acad Sci U S A 1988;85:5409–13. https://doi.org/10.1073/pnas.85.15.5409.

[32] Cerqueira, Janaína Canário; Viana, Marcus Vinicius Canário; Jaiswal, Arun Kumar; Soares, Siomar Castro; Tiwari, Sandeep; Wattam, Alice Rebecca; Silva, Artur; Alzahrani, Khalid J.; Tambuwala, Murtaza M.; Barh, Debmalya; AzevedoVAC. In silico identification of vaccine and drug targets for Corynebacterium ulcerans and the recently described C . silvaticum. Reasearch Sq 2022:1–14. https://doi.org/10.21203/rs.3.rs-1439819/v1.

[33] Chauhan V, Rungta T, Goyal K, Singh MP. Designing a multi-epitope based vaccine

to combat Kaposi Sarcoma utilizing immunoinformatics approach. SciRep 2019;9:1–15. https://doi.org/10.1038/s41598-019-39299-8.

[34] Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H, Ogawa T, et al. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity 1999;11:443–51. https://doi.org/10.1016/S1074-7613(00)80119-3.

[35] Schick, Judith; Etschel, Philipp; Bailo, Rebeca; Ott, Lisa; Bhatt, Apoorva; Bernd; Kirschning, Carsten; Burkovski, Andreas; Lang R. Toll-Like Receptor 2 and Mincle Cooperatively Sense Corynebacterial Cell Wall Glycolipids. Infect Immun 2017. https://doi.org/10.1128/IAI.00075-17.

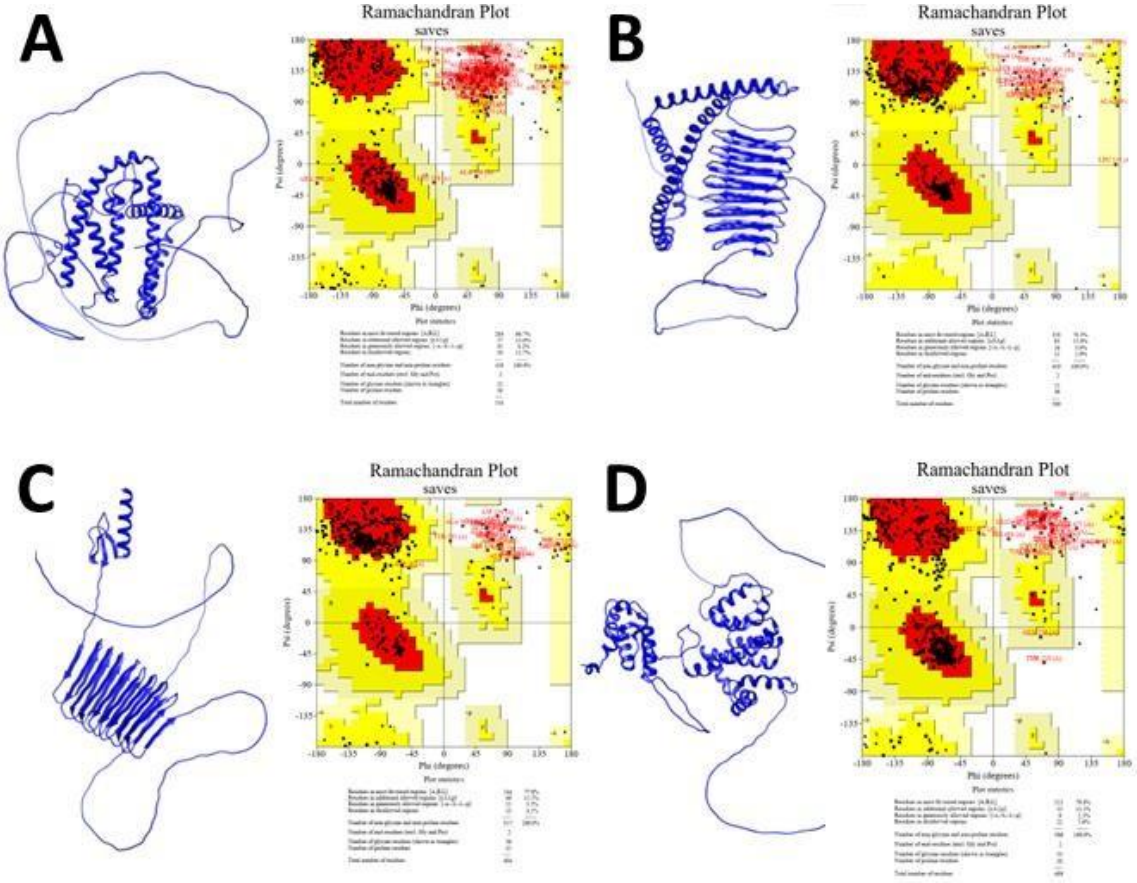
[36] Lopes Bastos B. Corynebacterium pseudotuberculosis: Immunological Responsesin Animal Models and Zoonotic Potential. J Clin Cell Immunol 2012;01. https://doi.org/10.4172/2155-9899.s4-005.

[37] Shahbazi S, Sabzi S, Noori Goodarzi N, Fereshteh S, Bolourchi N, Mirzaie B, etal. Identification of novel putative immunogenic targets and construction of a multi-epitope vaccine against multidrug-resistant Corynebacterium jeikeium using reverse vaccinology approach. Microb Pathog 2022;164:105425. https://doi.org/10.1016/j.micpath.2022.105425.

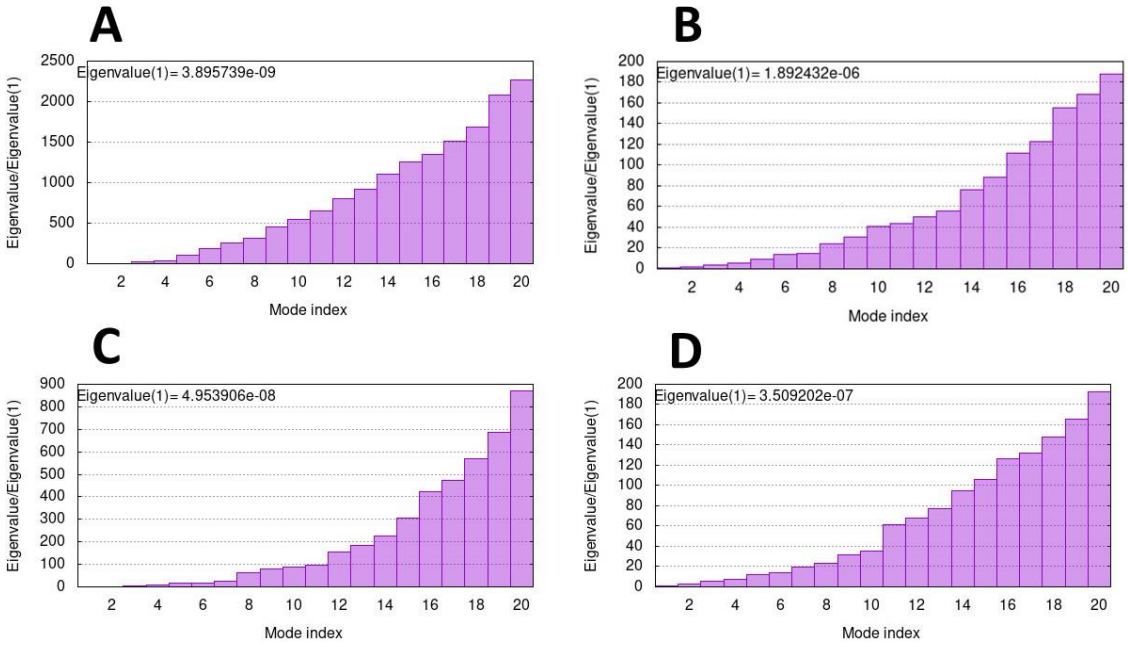
[38] Mochizuki Y, Saeki H, Iwaki M, Takagi H, Shibayama K, Amao H, et al. A novel experimental platform for toxigenic and non-toxigenic Corynebacteriumulcerans infection in mice. Pathog Dis 2016;74:1–7. https://doi.org/10.1093/femspd/ftv109.

[39] Jameie F, Dalimi A, Pirestani M, Mohebali M. Development of a multi-epitope recombinant protein for the diagnosis of human visceral leishmaniasis. Iran J Parasitol 2021;16:1–10. https://doi.org/10.18502/ijpa.v16i1.5506.

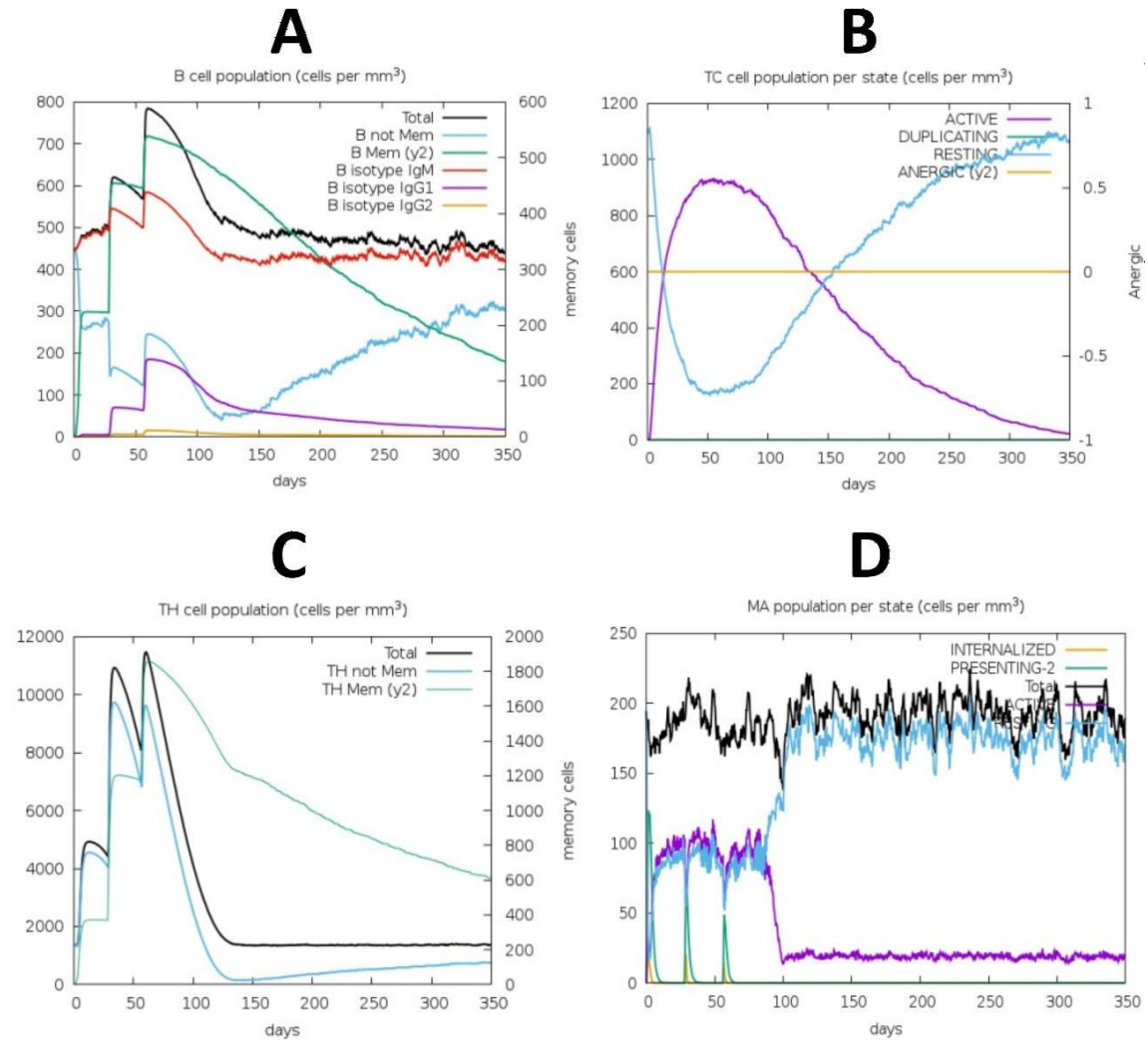
Supporting information



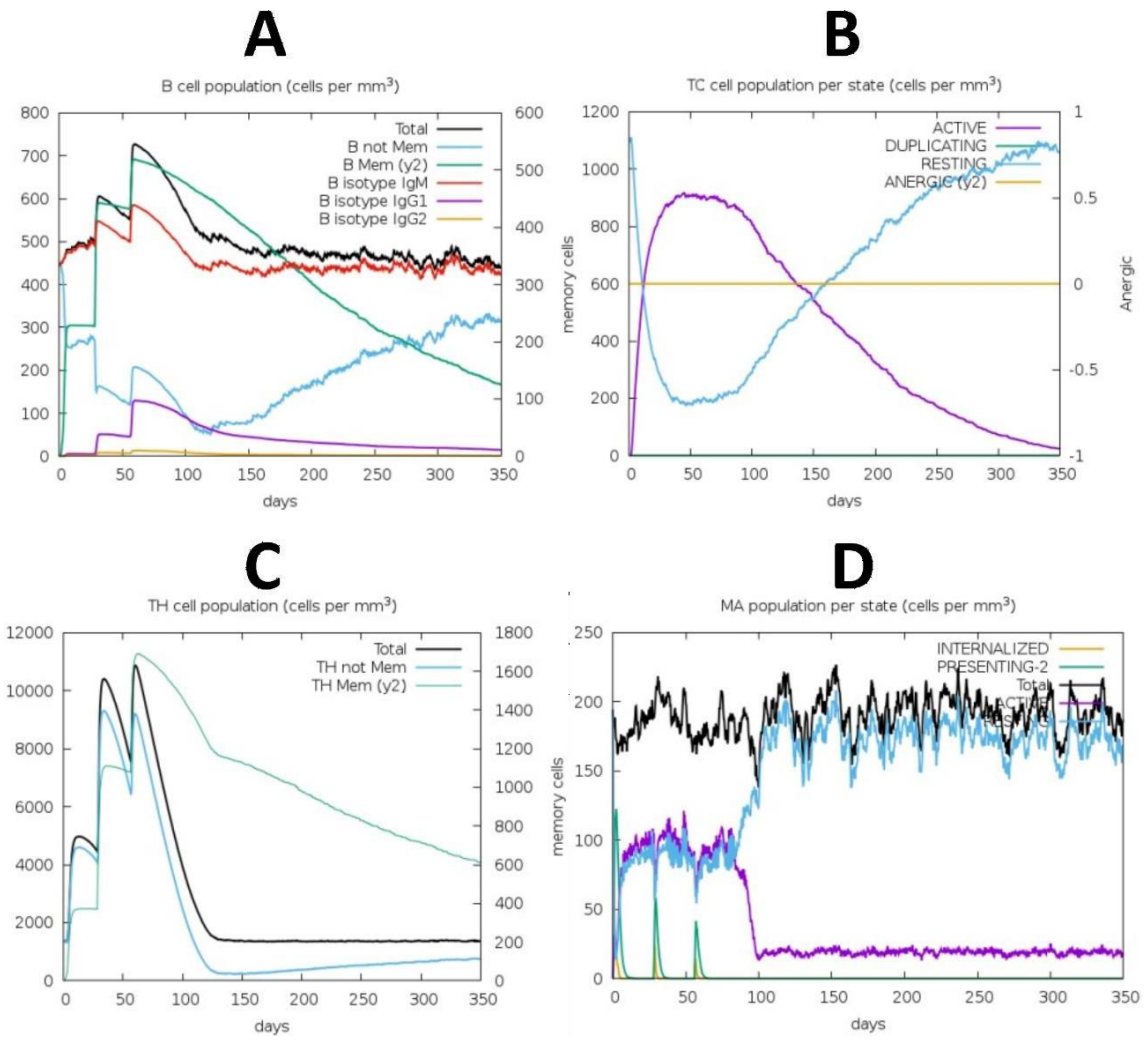
Supplementary Figure 1. Three-dimensional structure modelling by Alphafold and Ramachandran plots for the initial models (Before refinement). **A**: V1 showing 66.7% residues in most favored regions and 11.7% in disallowed regions; **B**: V2 showing 78.3%residues in most favored regions and 2.9% in disallowed regions; **C**: V3 showing 77.0%residues in most favored regions and 4.1% in disallowed regions and **D**: V4 showing 79.0% residues in most favored regions and 5.6% in disallowed regions.



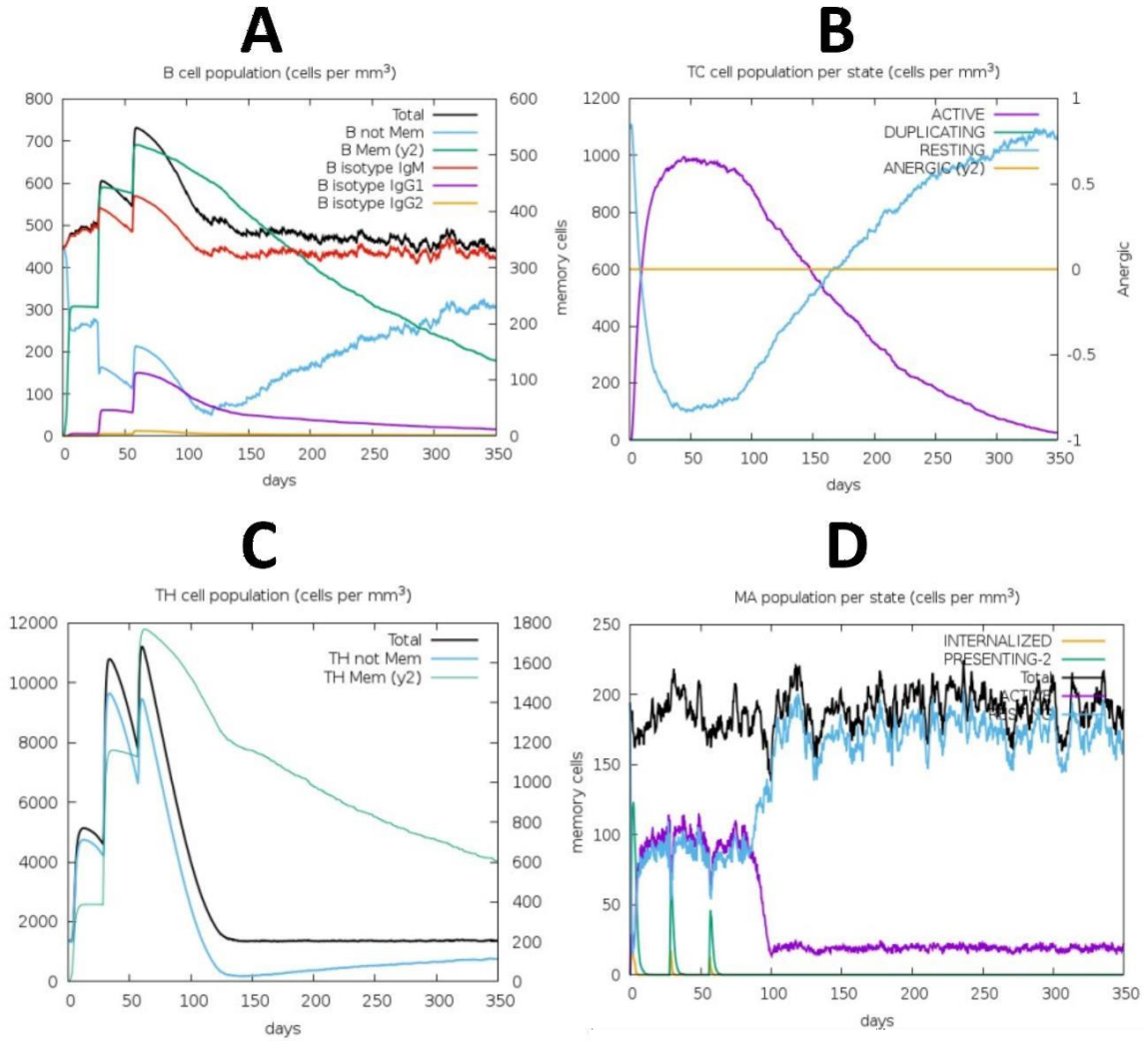
Supplementary Figure 2. Molecular dynamics simulation, Normal Mode Analysis. The figure reprsents Eigenvalue of vaccine-TLR2 complex (**A**: V1; **B**: V2; **C**: V3 and **D**: V4).



Supplementary Figure 3. Immune simulation results show (Vaccine construct V2) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.



Supplementary Figure 4. Immune simulation results show (Vaccine construct V3) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.



Supplementary Figure 5. Immune simulation results show (Vaccine construct V4) the amount (cells per mm^3) and activation of the main defense cells (vaccine 1 representation). **A**: B cell population; **B**: TC cell population per state; **C**: TH cellpopulation; **D**: macrophages population per state.

V. CONCLUSÕES GERAIS

1. A aplicação de métodos *in silico* da genômica como a genômica subtrativa, vacinologia reversa e imunoinformática permitiu a predição de 9 alvos de vacinas e 4 alvos de quimioterápicos para *C. ulcerans* e *C. silvaticum*, respectivamente. Esta informação associada à validação experimental contribuirá para a redução de custo e tempo necessário para o desenvolvimento destes métodos de controle.

2. Foram selecionados os respectivos compostos naturais com melhor capacidade de inibira função das proteínas citoplasmáticas (alvos de drogas) e, assim, interditar a colonização e crescimento desses organismos.

3. Esse trabalho é o primeiro a identificar alvos de drogas e vacinas, e propor imunógenosem comum para duas bactérias patogênicas do gênero *Corynebacterium*.

4. Através da imunoinformática, pôde-se propor 4 vacinas multi-epítopo para a prevençãode doenças causadas por *C. ulcerans* e *C. silvaticum*. As quatro vacinas de subunidades,foram construídas por epítopos preditos de 8 proteínas em comum das duas bactérias, tais vacinas apresentaram características adequadas de imunogenicidade, indução de respostas imunes esperadas por corinebactérias. Além de poderem ser expressas em vetor plasmídico de *Escherichia coli*.

VI. PERSPECTIVAS

1. *C. ulcerans* e *C. silvaticum* podem ser cultivadas e, portanto, é necessário realizar validação experimental por meio de análises *in vitro* e *in vivo*, a fim de testar a capacidade de indução das respostas imunes além da estabilidade estrutural da vacina candidata.

2. Realizar o monitoramento *in silico* da evolução dos genomas através da pangenômica de *C. ulcerans* e *C. silvaticum*, em consequência de futuros depósitos de sequências genômicas.

3. Proceder análises futuras de novos alvos de vacinas, à medida em que os bancos de dados genômicos e de compostos quimioterápicos vão aumentando e se atualizando.

4. Fazer análises de suscetibilidade e resistência antimicrobiana de ambas as espécies.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

BADELL, E. et al. *Corynebacterium rouxii* sp. nov., a novel member of the diphtheriae species complex. **Research in Microbiology**, v. 171, n. 3–4, p. 122–127, abr. 2020.

BAGNOLI, F. et al. Designing the next generation of vaccines for global public health. **OMICS A Journal of Integrative Biology**, v. 15, n. 9, p. 545–566, 2011.

BARH, D. et al. A novel comparative genomics analysis for common drug and vaccine targets in *Corynebacterium pseudotuberculosis* and other CMN group of human pathogens. **Chemical Biology and Drug Design**, v. 78, n. 1, p. 73–84, 2011.

BERNARD, A. L.; FUNKE, G. *Corynebacterium*. In: **Bergey's Manual of Systematic of Archaea and Bacteria (Online)**. [s.l.] John Wiley & Sons, Bergey's Manual Trust, 2015. p. 1–70.

BERNARD, K. The genus *Corynebacterium* and other medically relevant coryneform-like bacteria. **Journal of Clinical Microbiology**, v. 50, n. 10, p. 3152–3158, 2012.

BUSCH, A. et al. Genome sequence of a pathogenic *Corynebacterium ulcerans* strain isolated from a wild boar with necrotizing lymphadenitis. **BMC Research Notes**, v. 12, n. 1, p. 10–12, 2019.

CANÁRIO VIANA, M. V. et al. Evidence of episodic positive selection in *Corynebacterium diphtheriae* complex of species and its implementations in identification of drug and vaccine targets . **PeerJ**, v. 10, p. e12662, 16 fev. 2022.

CONTZEN, M. et al. *Corynebacterium ulcerans* from Diseased Wild Boars. **Zoonoses and Public Health**, v. 58, n. 7, p. 479–488, 2011.

DALSASS, M. et al. Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. **Frontiers in Immunology**, v. 10, n. FEB, p. 1–12, 2019.

DANGEL, A. et al. NGS-based phylogeny of diphtheria-related pathogenicity factors in different *Corynebacterium* spp. implies species-specific virulence transmission. **BMC Microbiology**, v. 19, n. 1, p. 1–16, 2019.

DANGEL, A. et al. *Corynebacterium silvaticum* sp. Nov., a unique group of NTTB corynebacteria in wild boar and roe deer. **International Journal of Systematic and Evolutionary Microbiology**, v. 70, n. 6, p. 3614–3624, 2020.

DAZAS, M. et al. Taxonomic status of *Corynebacterium diphtheriae* biovar Belfanti and proposal of *Corynebacterium belfantii* sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 68, n. 12, p. 3826–3831, dez. 2018.

EISENBERG, T. et al. Nontoxigenic tox-bearing *Corynebacterium ulcerans* Infection among Game Animals, Germany. **Emerging Infectious Diseases**, v. 20, n. 3, p. 448–452, 2014.

GOMES, L. G. R. et al. In Silico Designed Multi-Epitope Immunogen "Tpme-VAC/LGCM-2022" May Induce Both Cellular and Humoral Immunity against *Treponema pallidum* Infection. **Vaccines**, v. 10, n. 7, p. 1019, 2022.

GUARALDI, A. L. DE M.; HIRATA, R.; AZEVEDO, V. A. DE C. *Corynebacterium diphtheriae*, *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis* - General

Aspects. In: BURKOVSKI, A. (Ed.). . ***Corynebacterium diphtheriae*** **and Related Toxigenic Species**. Dordrecht: Springer Netherlands, 2014. p. 15–37.

HACKER, E. et al. *Corynebacterium ulcerans*, an emerging human pathogen. **Future Microbiology**, v. 11, n. 9, p. 1191–1208, set. 2016.

HARDISON, R. C. Comparative genomics. **PLoS Biology**, v. 1, n. 2, p. 156–160, 2003.

JAISWAL, A. K. et al. 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**, v. 27, n. March 2020, 2021.

MEDINI, D. et al. The microbial pan-genome. **Current Opinion in Genetics & Development**, v. 15, n. 6, p. 589–594, dez. 2005.

MÖLLER, J. et al. Phylogenomic characterisation of a novel corynebacterial species pathogenic to animals. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 113, n. 8, p. 1225–1239, 2020.

MÖLLER, J. et al. Newly isolated animal pathogen *Corynebacterium silvaticum* is cytotoxic to human epithelial cells. **International Journal of Molecular Sciences**, v. 22, n. 7, 2021.

MOUSTAFA, A. M.; LAL, A.; PLANET, P. J. Comparative genomics in infectious disease. **Current Opinion in Microbiology**, v. 53, p. 61–70, 2020.

MOXON, C. A. et al. New Insights into Malaria Pathogenesis. **Annual Review of Pathology: Mechanisms of Disease**, v. 15, p. 315–343, 2020.

NOGUEIRA, W. G. et al. Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*. **Genomics**, v. 113, n. 4, p. 2730–2743, 2021.

OLI, ANGUS NNAMDI; ODIMEGWU, D. C. et al. Immunoinformatics and Vaccine Development : An Overview. p. 13–30, 2020.

OLIVEIRA, M. et al. First report of *Corynebacterium pseudotuberculosis* from caseous lymphadenitis lesions in Black Alentejano pig (Sus scrofa domesticus). **BMC Veterinary Research**, v. 10, n. 1, p. 1–5, 2014.

PRADO, L. C. DA S. et al. New putative therapeutic targets against *Serratia marcescens* using reverse vaccinology and subtractive genomics. **Journal of Biomolecular Structureand Dynamics**, v. 0, n. 0, p. 1–16, 2021.

RIEGEL, P. et al. Taxonomy of *Corynebacterium diphtheriae* and related taxa, with recognition of Corynebacterium ulcerans sp. nov. nom. rev. **FEMS Microbiology Letters**, v. 126, n. 3, p. 271–276, 1995.

SERRUTO, D. et al. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: Immunological, functional and structural characterization of the antigens. **Vaccine**, v. 30, n. SUPPL. 2, p. B87–B97, 2012.

TAUCH, A.; BURKOVSKI, A. Molecular armory or niche factors: Virulence determinants of *Corynebacterium* species. **FEMS Microbiology Letters**, v. 362, n. 23, p. 1–6, 2015.

TETTELIN, H. et al. Comparative genomics: the bacterial pan-genome. **Current Opinion in Microbiology**, v. 11, n. 5, p. 472–477, out. 2008.

TOMAR, N.; DE, R. K. Immunoinformatics: An integrated scenario. **Immunology**, v. 131, n. 2, p. 153–168, 2010.

VIANA, M. V. C. et al. Genome Sequences of Two *Brucella suis* Strains Isolated from the Same Patient, 8 Years Apart. **Genome Announcements**, v. 5, n. 9, p. e01687-16, 2 mar. 2017.

VIANA, M. V. C. et al. Taxonomic classification of strain PO100/5 shows a broader geographic distribution and genetic markers of the recently described *Corynebacterium silvaticum*. **PLOS ONE**, v. 15, n. 12, p. e0244210, 21 dez. 2020.

VILELA RODRIGUES, T. C. et al. An immunoinformatics-based designed multi-epitope candidate vaccine (mpme-VAC/STV-1) against *Mycoplasma pneumoniae*. **Computers in Biology and Medicine**, v. 142, n. December 2021, 2022.

VIII. ANEXOS

Nesta seção estão apresentadas as colaborações em artigos publicados em periódicos revisados por pares, além de certificados de participação em cursos e eventos, organização de eventos e apresentação de seminários, bem como certificados de cursos de formação complementar.

A: COLABORAÇÃO EM ARTIGOS PUBLICADOS EM PERIÓDICOS

Evidence of episodic positive selection in *Corynebacterium diphtheriae* complex of species and its implementations in identification of drug and vaccine targets

Marcus Vinicius Canário Viana[1,2], Rodrigo Profeta[1], Janaína Canário Cerqueira[1], Alice Rebecca Wattam[3], Debmalya Barh[1,4], Artur Silva[2] and Vasco Azevedo[1]

[1] Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2] Departamento de Genética, Universidade Federal do Pará, Belém, Pará, Brazil
[3] Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, United States
[4] Institute of Integrative Omics and Applied Biotechnology, Nonakuri, West Bengal, India

ABSTRACT

Background: Within the pathogenic bacterial species *Corynebacterium* genus, six species that can produce diphtheria toxin (*C. belfantii*, *C. diphtheriae*, *C. pseudotuberculosis*, *C. rouxii*, *C. silvaticum* and *C. ulcerans*) form a clade referred to as the *C. diphtheria* complex. These species have been found in humans and other animals, causing diphtheria or other diseases. Here we show the results of a genome scale analysis to identify positive selection in protein-coding genes that may have resulted in the adaptations of these species to their ecological niches and suggest drug and vaccine targets.

Methods: Forty genomes were sampled to represent species, subspecies or biovars of *Corynebacterium*. Ten phylogenetic groups were tested for positive selection using the PosiGene pipeline, including species and biovars from the *C. diphtheria* complex. The detected genes were tested for recombination and had their sequences alignments and homology manually examined. The final genes were investigated for their function and a probable role as vaccine or drug targets.

Results: Nineteen genes were detected in the species *C. diphtheriae* (two), *C. pseudotuberculosis* (10), *C. rouxii* (one), and *C. ulcerans* (six). Those were found to be involved in defense, translation, energy production, and transport and in the metabolism of carbohydrates, amino acids, nucleotides, and coenzymes. Fourteen were identified as essential genes, and six as virulence factors. Thirteen from the 19 genes were identified as potential drug targets and four as potential vaccine candidates. These genes could be important in the prevention and treatment of the diseases caused by these bacteria.

Subjects Bioinformatics, Genomics, Microbiology, Molecular Biology
Keywords Corynebacterium, Positive selection, Drug target, Vaccine target

Submitted 22 July 2021
Accepted 30 November 2021
Published 16 February 2022

Corresponding author
Vasco Azevedo, vasco@icb.ufmg.br

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 13

DOI 10.7717/peerj.12662

Characterization of the first vaginal *Lactobacillus crispatus* genomes isolated in Brazil

Marcelle Oliveira de Almeida[1], Rodrigo Carvalho[1], Flavia Figueira Aburjaile[1], Fabio Malcher Miranda[1], Janaína Canário Cerqueira[1], Bertram Brenig[2], Preetam Ghosh[3], Rommel Ramos[4], Rodrigo Bentes Kato[5], Siomar de Castro Soares[6], Artur Silva[4], Vasco Azevedo[1] and Marcus Vinicius Canário Viana[1]

[1] Department of Genetics, Ecology, and Evolution, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2] Institute of Veterinary Medicine, University of Göttingen, Göttingen, Germany
[3] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
[4] Department of Genetics, Federal University of Pará, Belém, Pará, Brazil
[5] Post-graduation Program in Bioinformatics, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[6] Department of Immunology, Microbiology, and Parasitology, Federal University of Triângulo Mineiro, Uberaba, Minas Gerais, Brazil

Submitted 9 December 2020
Accepted 17 February 2021
Published 10 March 2021

Corresponding author
Vasco Azevedo,
vascoariston@gmail.com

Academic editor
Hector Mora-Montes

Additional Information and Declarations can be found on page 16

DOI **10.7717/peerj.11079**

ABSTRACT

Background: *Lactobacillus crispatus* is the dominant species in the vaginal microbiota associated with health and considered a homeostasis biomarker. Interestingly, some strains are even used as probiotics. However, the genetic mechanisms of *L. crispatus* involved in the control of the vaginal microbiome and protection against bacterial vaginosis (BV) are not entirely known. To further investigate these mechanisms, we sequenced and characterized the first four *L. crispatus* genomes from vaginal samples from Brazilian women and used genome-wide association study (GWAS) and comparative analyses to identify genetic mechanisms involved in healthy or BV conditions and selective pressures acting in the vaginal microbiome.

Methods: The four genomes were sequenced, assembled using ten different strategies and automatically annotated. The functional characterization was performed by bioinformatics tools comparing with known probiotic strains. Moreover, it was selected one representative strain (*L. crispatus* CRI4) for in vitro detection of phages by electron microscopy. Evolutionary analysis, including phylogeny, GWAS and positive selection were performed using 46 public genomes strains representing health and BV conditions.

Results: Genes involved in probiotic effects such as lactic acid production, hydrogen peroxide, bacteriocins, and adhesin were identified. Three hemolysins and putrescine production were predicted, although these features are also present in other probiotic strains. The four genomes presented no plasmids, but 14 known families insertion sequences and several prophages were detected. However, none of the mobile genetic elements contained antimicrobial resistance genes. The genomes harbor a CRISPR-Cas subtype II-A system that is probably inactivated due to fragmentation of the genes *csn2* and *cas9*. No genomic feature was associated with a

GMR

Probiotic potential of novel Brazilian *Lactobacillus crispatus* strains

Vasco Azevedo, Marcus Vinicius Canário Viana, Janaína Canário Cerqueira, Luís Cláudio Lima de Jesus, Tales Fernando da Silva, Rodrigo Carvalho and, Flavia Figueira Aburjaile

Department of Genetics, Ecology, and Evolution, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Corresponding author: V. Azevado
E-mail: mailto:vasco@icb.ufmg.br

Genet. Mol. Res. 20 (4): gmr30376
Received April 27, 2021
Accepted May 11, 2021
Published May 18, 2021

ABSTRACT. *Lactobacilli* are predominant bacterial species colonizing the vaginal surfaces of healthy women, where they play a protective role against opportunistic and polymicrobial infections such as Bacterial Vaginosis (BV). Several *Lactobacillus* species, especially *L. crispatus*, have been prospected for probiotic applications due to their potential antimicrobial and anti-inflammatory capacity. In the last decade, several genomic studies have investigated the genetic factors of *L. crispatus* strains for identifying novel probiotic strains and evaluating their application for improving human or animal health. This mini-review highlights the primary genes associated with *l. crispatus* protective mechanisms identified, in our previous work, on four novel strains isolated from healthy Brazilian women of reproductive age. Among the probiotic features discussed, the role of a pyruvate oxidase-encoding gene, lactate synthesis-related enzymes, bacteriocins genes, and genomic islands is reviewed, and the following steps to confirm their activity are pointed out.

An Acad Bras Cienc (2021) 93(2): e20200945 DOI 10.1590/0001-3765202120200945
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

MICROBIOLOGY

Novel insights in bacterial vaginosis etiology through genomic approaches

MARCELLE O. ALMEIDA, MARCUS VINICIUS C. VIANA, JANAÍNA C. CERQUEIRA, FLAVIA F. ABURJAILE, ANDREY A.Z. JUNIOR, VASCO AZEVEDO & RODRIGO D.O. CARVALHO

Abstract: Bacterial vaginosis (BV) has been considered as dysbiosis state whose etiology is not fully understood. This condition affects a large number of women of reproductive age and its study has been highly relevant due to the growing association of BV with and gynecological and obstetric complications and diseases, in addition to a greater susceptibility to sexually transmitted diseases, including HIV. The vaginal microbiota composition presents high variability among different ethnic groups of women, although, generally, the prevalence of lactobacilli species has been reported. Several studies suggest they may play a protective role, especially *Lactobacillus crispatus* whose population is typically present in low proportions in women with BV. This review article describes the contributions and limitations of genomic approaches in elucidating protective characteristics and mechanisms associated with colonization and persistence of lactobacilli strains. Although some genetic features were associated with resilience of *L. crispatus* during BV, furher studies are required to uncover their functions.

Key words: bacterial vaginosis, genomics, health, *Lactobacillus*.

PLOS ONE

RESEARCH ARTICLE

Taxonomic classification of strain PO100/5 shows a broader geographic distribution and genetic markers of the recently described *Corynebacterium silvaticum*

Marcus Vinicius Canário Viana[1,2], Rodrigo Profeta[1], Alessandra Lima da Silva[1], Raquel Hurtado[1], Janaína Canário Cerqueira[1], Bruna Ferreira Sampaio Ribeiro[1], Marcelle Oliveira Almeida[1], Francielly Morais-Rodrigues[1], Siomar de Castro Soares[3], Manuela Oliveira[4], Luís Tavares[4], Henrique Figueiredo[5], Alice Rebecca Wattam[6], Debmalya Barh[7], Preetam Ghosh[8], Artur Silva[2], Vasco Azevedo[1]*

1 Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **2** Department of Genetics, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil, **3** Department of Immunology, Microbiology and Parasitology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Minas Gerais, Brazil, **4** Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, Lisboa, Portugal, **5** National Reference Laboratory of Aquatic Animal Disease, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **6** Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, United States of America, **7** Institute of Integrative Omics and Applied Biotechnology, Purba Medinipur, West Bengal, India, **8** Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, United States of America

* vasco@icb.ufmg.br

OPEN ACCESS

Citation: Viana MVC, Profeta R, da Silva AL, Hurtado R, Cerqueira JC, Ribeiro BFS, et al. (2020) Taxonomic classification of strain PO100/5 shows a broader geographic distribution and genetic markers of the recently described *Corynebacterium silvaticum*. PLoS ONE 15(12): e0244210. https://doi.org/10.1371/journal.pone.0244210

Editor: Chih-Horng Kuo, Academia Sinica, TAIWAN

Received: April 17, 2020

Accepted: December 4, 2020

Published: December 21, 2020

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico); CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior); Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG); Universidade Federal de Minas Gerais (UFMG); and Pró-Reitoria de Pesquisa da UFMG (PRPQ-UFMG); A. R. Wattam

Abstract

The bacterial strain PO100/5 was isolated from a skin abscess taken from a pig (*Sus scrofa domesticus*) in the Alentejo region of southern Portugal. It was identified as *Corynebacterium pseudotuberculosis* using biochemical tests, multiplex PCR and Pulsed Field Gel Electrophoresis. After genome sequencing and *rpoB* phylogeny, the strain was classified as *C. ulcerans*. To better understand the taxonomy of this strain and improve identification methods, we compared strain PO100/5 to other publicly available genomes from *C. diphtheriae* group. Taxonomic analysis reclassified it and three others strains as the recently described *C. silvaticum*, which have been isolated from wild boar and roe deer in Germany and Austria. The results showed that PO100/5 is the first sequenced genome of a *C. silvaticum* strain from livestock and a different geographical region, has the unique sequence type ST709, and could be could produce the *diphtheriae* toxin, along with strain 05–13. Genomic analysis of PO100/5 showed four prophages, and eight conserved genomic islands in comparison to *C. ulcerans*. Pangenome analysis of 38 *C. silvaticum* and 76 *C. ulcerans* genomes suggested that C. *silvaticum* is a genetically homogeneous species, with 73.6% of its genes conserved and a pangenome near to be closed ($\alpha > 0.952$). There are 172 genes that are unique to *C. silvaticum* in comparison to *C. ulcerans*. Most of these conserved genes are related to nutrient uptake and metabolism, prophages or immunity against them, and could be genetic markers for species identification. Strains PO100/5 (livestock) and KL0182^T (wild

B: PARTICIPAÇÃO EM EVENTOS, SEMINÁRIO E CURSOS DE CURTA DURAÇÃO

PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

BIOINFORMÁTICA UFMG

icbufmg Instituto de Ciências Biológicas

CERTIFICADO

Certificamos que

Janaína Canário Cerqueira

ministrou seminário intitulado "Taxonogenômica, vacinologia reversa e predição de alvos quimioterápicos em Corynebacterium ulcerans e C. silvaticum", para os alunos do Programa de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, no dia 30 de julho de 2021, via plataforma Zoom.

Belo Horizonte, 30 de julho de 2021.

Prof. Dr. Aristóteles Góes Neto
Coordenador

HOST-PATHOGEN INTERACTION Meeting 2021

CERTIFICADO

Certificamos que

Janaína Canário Cerqueira

participou do minicurso

MICROBIOTA INTESTINAL HUMANA NA SAÚDE E NA DOENÇA

com carga horária de 03 horas, realizado durante o pré-evento do **Host-Pathogen Interaction Meeting 2021**, no dia 03 de novembro de 2021, na modalidade remota.

Curitiba, 28 de novembro de 2021

Dr. Alexander Welker Biondo
Coordenador do Programa de Pós-graduação em Biologia Celular e Molecular da UFPR

Dr. Marcel Ivan Ramirez
Coordenador Geral do Host-Pathogen Interaction Meeting 2021

Organização

Apoio

HOST-PATHOGEN INTERACTION Meeting 2021

CERTIFICADO

Certificamos que

JANAÍNA CANÁRIO CERQUEIRA

participou, na qualidade de OUVINTE, do **Host-Pathogen Interaction Meeting 2021**, realizado no período de 08 a 12 de novembro de 2021, na modalidade remota, contabilizando carga horária total de 20 horas.

Curitiba, 28 de novembro de 2021

Dr. Alexander Welker Biondo
Coordenador do Programa de Pós-graduação em Biologia Celular e Molecular da UFPR

Dr. Marcel Ivan Ramirez
Coordenador Geral do Host-Pathogen Interaction Meeting 2021

Organização

Apoio

1st Latin American Environmental DNA Metagenomics Symposium

CERTIFICATE OF ATTENDANCE

This certificate is presented to

Janaína Canário Cerqueira

for attending the
1st Latin American Environmental DNA Metagenomics Symposium
held on the days 20 to 22 October, 2021, with the duration of 13 hours.

Prof. Daniel Cardoso de Carvalho
Organising Committee Chair

Organizer
Laboratório de Genético da Conservação
PUC Minas

Sponsors
illumina

X-meeting eXperience 2020

November 9 - 10, 2020

CERTIFICATE

This certifies that **Janaína Canário Cerqueira**, has participated in the X-meeting eXperience 2020 held online, between 11-09-2020 - 11-11-2020.

November 11, 2020.

Prof. Dr. Ney Lemke
President of AB³C

Prof. Dr. Fabrício Martins Lopes
First Secretary of AB³C

RECOM

CERTIFICADO

Certifico que **Janaína Canário** participou do **1º Encontro da Rede de Ciências Ômicas (RECOM)** realizado no dia **03 de Julho de 2020**, na qualidade de **ouvinte**, com carga horária de **10 horas**.

Belo Horizonte, 10 de Julho de 2020

PROF. DR. VASCO AZEVEDO
COORDENADOR DA RECOM

RECOM