



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA – UFRA

GABRIELA COSTA DUARTE RIBEIRO

**GENÔMICA COMPARATIVA E SUBTRATIVA ALIADA À VACINOLOGIA
REVERSA PARA IDENTIFICAÇÃO DE CANDIDATOS À ALVOS VACINAIS
CONTRA A *Corynebacterium ulcerans***

BELÉM - PA

2021

GABRIELA COSTA DUARTE RIBEIRO

**GENÔMICA COMPARATIVA E SUBTRATIVA ALIADA À VACINOLOGIA
REVERSA PARA IDENTIFICAÇÃO DE CANDIDATOS À ALVOS VACINAIS
CONTRA A *Corynebacterium ulcerans***

Dissertação submetida ao Programa de Pós-Graduação em Biotecnologia Aplicada à Agropecuária da UFRA, como requisito final para obtenção do grau de Mestre em Biotecnologia.

Orientador: Dr. Luís Carlos Guimarães

BELÉM

2021

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas da Universidade Federal Rural da Amazônia
Gerada automaticamente mediante os dados fornecidos pelo(a) autor(a)

R484g Ribeiro, Gabriela Costa Duarte

Genômica comparativa e subtrativa aliada à vacinologia reversa para identificação de candidatos a alvos vacinais contra a *Corynebacterium ulcerans* / Gabriela Costa Duarte Ribeiro. - 2021.

53 f. : il. color.

Dissertação (Mestrado) - Programa de PÓS-GRADUAÇÃO em Biotecnologia Aplicada à Agropecuária (PPGBAA), Campus Universitário de Belém, Universidade Federal Rural Da Amazônia, Belém, 2021.

Orientador: Prof. Dr. Luís Carlos Guimarães

I. Guimarães, Luís Carlos , orient. II. Título

CDD

GABRIELA COSTA DUARTE RIBEIRO

**GENÔMICA COMPARATIVA E SUBTRATIVA ALIADA À VACINOLOGIA
REVERSA PARA IDENTIFICAÇÃO DE CANDIDATOS À ALVOS VACINAIS
CONTRA A *Corynebacterium ulcerans***

Dissertação apresentada à Universidade Federal Rural da Amazônia, como parte das exigências do Curso de Mestrado em Biotecnologia Aplicada à Agropecuária, para obtenção do título de Mestre. Área de Concentração: Biotecnologia.

Aprovado em 12 de abril de 2021

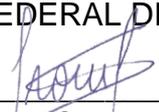
BANCA EXAMINADORA:



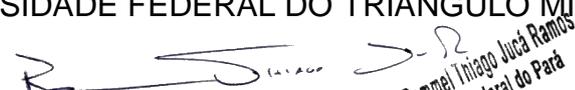
Dr. Luís Carlos Guimarães - Orientador
UNIVERSIDADE FEDERAL DO PARÁ



Dr. Vasco Ariston de Carvalho Azevedo - 1º Examinador
UNIVERSIDADE FEDERAL DE MINAS GERAIS



Dr. Siomar de Castro Soares - 2º Examinador
UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO



Dr. Rommel Thiago Juca Ramus - 3º Examinador
UNIVERSIDADE FEDERAL DO PARÁ



Dra. Adriana Ribeiro Carneiro Folador - 4º Examinador
UNIVERSIDADE FEDERAL DO PARÁ

AGRADECIMENTOS

A toda minha família, especialmente a minha avó, Maria Costa Duarte, por todas as orações. Aos meus pais, Rosimá Ribeiro e Roberto Ribeiro Melo, por todo apoio, incentivo, cuidado e muito amor. E aos meus irmãos, Renato Ribeiro e Lucas Sebastian, por todo carinho.

Ao PPGBAA e aos professores, além da coordenação e secretaria que me proporcionaram as condições necessárias para que eu alcançasse meus objetivos.

Ao Meu orientador, Luís Carlos Guimarães, pela paciência, dedicação e conhecimento fornecido para que esse trabalho pudesse ser realizado.

A equipe do professor Luís, em especial a Sandy Aguiar, por disporem seu tempo e habilidades para ajudar na execução do projeto.

De modo muito especial, agradeço ao James Siqueira, que além de amigo, foi um companheiro de trabalho, estudos e pesquisas, onde juntos aprendemos e ensinamos um ao outro.

À minha tia do coração, Cacilda Pinto e seus filhos, Victor Segtowick e Clarice Costa, por estarem comigo durante esses dois anos, me incentivando e ajudando nos momentos mais difíceis.

Aos meus tios, Renata e Anderson, por todo acolhimento dados durante parte dessa jornada.

Às minhas sobrinhas, Maria Pietra e Maria Cecília, por tornarem essa jornada mais alegre e cheia de amor.

À minha prima e irmã do coração, Roberta Laíne, por toda força, incentivo e ensinamentos e por estar comigo em todas as fases.

A minha família por parte de pai, especialmente ao Eduardo e a Ariana Ribeiro, que me ofereceram hospedagem nos primeiros semestres do curso.

Aos melhores amigos que carrego desde a graduação, Akim Afonso e Lucas Ramon, por estarem comigo em todas as etapas.

Aos melhores amigos que o programa me possibilitou conhecer, James Siqueira, Tamires Natividade e Camille Maia, por todo companheirismo.

Aos meus amigos de infância, Letícia Amoras, Kelson Miranda, Lanhy Mendes, Bruna Magalhães, Franciani Silva, Mateus Silva, Lucas Matheus, Meri Silva, Gabriela Amoras e Mirela Ibiapino, pelo companheirismo durante parte dessa caminhada.

Ao Andrei Giacchetto Felice, por colaborar com algumas análises desse trabalho.

À CAPES, pela bolsa concedida.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

RESUMO

Corynebacterium ulcerans é uma bactéria Gram-positiva, que pode causar infecções respiratória e cutânea em humanos e animais. Casos de infecções tendo como agente causador a *C. ulcerans* vêm crescendo continuamente, visto que algumas cepas dessa espécie podem carregar genes toxigênicos em seus genomas. Por ser um patógeno de interesse médico e veterinário é importante estudar os genes que estão atrelados a sua patogenicidade, bem como predizer alvos vacinais e de drogas. Diante disto, este trabalho realizou uma busca com todos os 29 genomas disponíveis no NCBI (National Center for Biotechnology Information), sendo que as sequências genômicas passaram por uma padronização da predição e anotação gênica através do software Prokka. As similaridades entre as cepas foram calculadas usando o software Gegenes, as ilhas de patogenicidades foram preditas através do Gipsy e a homologia entre elas foi realizada com a ferramenta BLASTn. O software PGAP foi usado para o cálculo do pangenoma e a partir do genoma central foi realizada a busca por genes essenciais disponíveis em bancos de dados online. O BLASTp foi utilizado para identificação de proteínas não homologas ao hospedeiro e a localização subcelular dessas proteínas foi inferida pela ferramenta PSORTb. Dentre as proteínas preditas como exportadas ou ancoradas na membrana foram verificadas as propriedades antigênicas através do software VaxiJen. Dentre as 29 sequências disponíveis no NCBI para *C. ulcerans*, apenas duas cepas (PO100/5, w25) obtiveram baixa porcentagem de similaridade em relação às outras linhagens. Foram preditas 156 ilhas de patogenicidade, dentre as quais 12 ilhas não apresentaram homologia com outras ilhas e uma ilha específica foi predita em 25 genomas. Das vinte e sete cepas analisadas, nove continham a sequência do gene *tox* e dezesseis o gene *pld* em seus genomas. O pangenoma de *C. ulcerans* tem um total de 5.960 genes, onde 1.295 genes fazem parte do genoma central, 1.790 do genoma acessório e 2.205 são genes específicos. A aplicação da genômica subtrativa revelou 2 proteínas putativamente antigênicas que podem ser utilizadas como alvos vacinais e 4 proteínas no desenvolvimento de fármacos.

Palavras-chaves: Toxina da difteria. *Corynebacterium ulcerans*. Pangenômica. Genômica subtrativa. Bioinformática.

ABSTRACT

Corynebacterium ulcerans is a Gram-positive bacteria, which can cause respiratory and cutaneous infections in humans and animals. Reported cases of infections with *C. ulcerans* have continuously been increasing and some of its strains were found to carry toxin genes in their genomes. It is crucial to study about its genes and how they are related to its pathogenicity, as well as to predict vaccine and drug targets. According to this, genomic data were collected from the National Center for Biotechnology Information (NCBI) and genome prediction and annotation were performed using the Prokka software for standardization. Similarity among the strains was calculated using the Gggenes software, the pathogenicity islands were predicted using the Gipsy software and the homology among them was performed using the BLAST tool. The pan-genome was calculated using the Pan-genome Analysis Pipeline (PGAP) software and a search for essential genes available on online databases was carried out according to the core genome. The BLAST tool was used to identify non-host homologous proteins and the subcellular localization of these proteins was predicted using the PSORTb tool. Antigenic properties of secreted or membrane-anchored proteins were verified and predicted using the VaxiJen software. Among the 29 genome sequences from *C. ulcerans* available on the NCBI public database, only two strains (PO100/5, w25) had a low percentage of similarity to the other strains. 156 pathogenicity islands were predicted, among which 12 did not have homology to other pathogenicity islands and 1 specific island was predicted in 25 genomes. Among the 27 strains analyzed, 9 contained the *tox* gene sequence and 16 contained *pld* genes in their genomes. The pan-genome of *C. ulcerans* is composed of 5960 genes, among which 1295 are core genes, 1790 are accessory genes and 2205 are specific genes. The application of substrative genomics revealed 2 putative antigen proteins that can be used as vaccine targets and 4 proteins that can be used for therapeutic drug development.

Keywords: Diphtheria toxin. *Corynebacterium ulcerans*. Pangenomics. Substrative genomics. Bioinformatics.

LISTA DE ILUSTRAÇÕES

Figura 1 – Heatmap: análise de similaridade entre os genomas	30
Figura 2 – Árvore filogenômica. As designações de número de referência da sequência seguidas por denominação das cepas, hospedeiro e país de isolamento, quando essas informações estavam disponíveis	31
Figura 3 – Representação genômica circular de ilhas de patogenicidade nos genomas de <i>C. ulcerans</i> utilizando a cepa 0102 como referência.....	32
Figura 4 – Árvore filogenética demonstrando a similaridade quanto a presença do gene <i>tox</i> entre as linhagens de <i>C. ulcerans</i> e <i>C. diphtheriae</i>	36
Figura 5 – Árvore filogenética demonstrando a similaridade quanto a presença do gene <i>pld</i> entre as linhagens de <i>C. ulcerans</i> e <i>C. pseudotuberculosis</i>	37
Figura 6 – Predição do pangenoma e do genoma do central de vinte e sete genomas de <i>C. ulcerans</i>	38

LISTA DE TABELAS

Tabela 1 – Alguns casos de infecções causados por <i>C. ulcerans</i> em diversos países	16
Tabela 2 – Características de <i>C. ulcerans</i> isolados de animais	18
Tabela 3 – Características e informações dos genomas analisados neste trabalho	28
Tabela 4 – Caracterização das cepas de <i>C. ulcerans</i> quanto a presença de Ilhas de Patogenicidade	33
Tabela 5 – Caracterização da Ilha de Patogenicidade presente em 25 linhagens ...	35
Tabela 6 – Alvos vacinais e de drogas subtraído do genoma central de 27 genomas de <i>C. ulcerans</i>	40

LISTA DE ABREVIATURAS E SIGLAS

TD – toxina da difteria

PAI's – Ilhas de patogenicidade

GC – Guanina e Citosina

DNA – Ácido Desoxirribonucleico

NGS – *Next-Generation Sequencing*

PROKKA – *Rapid Prokaryotic Genome Annotation*

ANI – *Average Nucleotide Identity*

PGAP – *Pan-genomes analysis pipeline*

GF – *Gene family*

tRNA – RNA transportador

DEG – *Database of Essential Genes*

NCBI – *National Center for Biotechnology Information*

BLAST – *Basic Local Alignment Search Tool*

BLASTp – *Basic Local Alignment Search Tool of Proteins*

BRIG – *BLAST Ring Image Generator*

SUMÁRIO

1. CONTEXTUALIZAÇÃO	12
1.1 Justificativa	12
1.1.1 Objetivo Geral	13
1.1.2 Objetivos Específicos	13
1.2 Revisão da Literatura	14
1.2.1 Microbiologia	14
1.2.2 Epidemiologia	14
1.2.3 Apresentação clínica e terapia	18
1.2.4 Bioinformática	19
1.2.4.1 Genômica comparativa	19
1.2.4.2 Plasticidade genômica	20
1.2.4.3 Ilhas de patogenicidade	20
1.2.4.4 Genômica subtrativa	21
2. GENÔMICA COMPARATIVA E SUBTRATIVA ALIADA À VACINOLOGIA REVERSA PARA IDENTIFICAÇÃO DE CANDIDATOS À ALVOS VACINAIS CONTRA A <i>Corynebacterium ulcerans</i>	22
2.1 Introdução	22
2.2 Material e Métodos	23
2.2.1 Aquisição das sequências genômicas	23
2.2.2 Identificação e análise de Ilhas de Patogenicidade	23
2.2.3. Análise filogenômica e similaridade	24
2.2.4 Predição do pangenoma	25
2.2.5 Identificação dos genes essenciais	25
2.2.6 Triagem dos genes não homólogos ao hospedeiro	25
2.2.7 Localização subcelular e previsão de antígenos	26
2.3 Resultados e discussão	26
2.3.1 Dados genômicos de <i>C. ulcerans</i>	26
2.3.2 Análise de similaridade e filogenômica dos genomas baixados	26
2.3.3 Predição e representação circular das Ilhas de patogenicidade em <i>C. ulcerans</i>	31
2.3.4 Análise filogenética dos genes: <i>tox</i> e <i>pld</i>	35
2.3.5 Predição do pangenoma, do genoma central e dos genes linhagem- específicos de <i>C. ulcerans</i>	37
2.3.6 Predição de alvos putativos de vacina e de drogas	39
2.4 Conclusões	42
REFERÊNCIAS BIBLIOGRÁFICAS	44

1. CONTEXTUALIZAÇÃO

1.1 Justificativa

Inicialmente os casos de infecções ocasionados pela *Corynebacterium ulcerans* foram relatados em áreas rurais, a partir de bactérias isoladas, principalmente, de gado leiteiro que apresentavam quadros clínicos de mastite. Entretanto, estudos posteriores demonstraram que a bactéria poderia estar tanto em hospedeiros humanos quanto em animais, classificando-a como um agente causador de infecções zoonóticas (BOSTOCK et al., 1984; HART, 1984; HIGGS et al., 1967; HOMMEZ et al., 1999).

Os relatos de propagação do patógeno em humanos reportados na literatura são em sua maioria de pessoas que contraíram a infecção através do contato com animais infectados, podendo ser de estimação ou até mesmo selvagem (CORTI et al., 2012; FUURSTED et al., 2015; SCHUHEGGER et al., 2009; VANDENTORREN et al., 2014). Os animais também podem apresentar infecções, sendo que os sintomas variam de úlceras de pele ou infecções sistêmicas a doenças semelhantes à difteria (DE ZOYSA et al., 2005; SAEKI et al., 2015; SCHUHEGGER et al., 2009; SYKES et al., 2010). Há casos também em que o portador pode ser assintomático (FUURSTED et al., 2015).

É importante ressaltar que o número de casos relatados de infecção respiratória em humanos por *C. ulcerans* vem crescendo continuamente (DE ZOYSA et al., 2005; HATANAKA et al., 2003; YASUDA et al., 2018). A difteria causada por *C. ulcerans* e *C. diphtheriae* apresentam sintomas muito semelhantes, por isso se torna necessário realizar diagnósticos clínicos laboratoriais para determinar quem é o verdadeiro agente causador desta doença (DE SOUZA DE OLIVEIRA DIAS et al., 2011; SUVAJDZIC et al., 2012; YASUDA et al., 2018).

Igualmente, são necessários estudos que comprovem a eficácia da vacina contra a difteria causada por *C. ulcerans* em humanos, visto que ela foi fabricada com intuito de combater difteria causada por cepas toxigênicas de *C. diphtheriae*. Além disso, também existem estudos controversos que apontam a ineficácia da vacina como medida profilática contra a *C. ulcerans* e até mesmo contra linhagens atoxigênicas de *C. diphtheriae*. (DE SANTIS et al., 2020; DE SOUZA DE OLIVEIRA

DIAS et al., 2011; KONRAD; HÖRMANSDORFER; SING, 2015a; YASUDA et al., 2018).

Outra doença que *C. ulcerans* pode causar é a linfadenite caseosa, que é caracterizada pela formação de abscessos nos linfonodos superficiais, internos e em órgãos de animais (FONTAINE; BAIRD, 2008; TEJEDOR et al., 2000). Essa patologia está atrelada ao gene *pld* (fosfolipase D), que codifica uma potente exotoxina que promove a hidrólise de ligações éster na esfingomiéline em membranas de células de mamíferos (DORELLA et al., 2006). Entretanto, existem poucos estudos de caso associando *C. ulcerans* à essa enfermidade, pois o principal agente causador é a *Corynebacterium pseudotuberculosis*.

Conhecer melhor a genética da *C. ulcerans* é imprescindível para se obter informações como a quantidade de genes envolvidos na virulência que muitas linhagens dessa bactéria carregam e que podem potencialmente causar doenças graves, tais como a difteria em humanos e a linfadenite caseosa em animais (EISENBERG et al., 2014; FUURSTED et al., 2015; SUBEDI et al., 2018). Além disso, através da análise pangenômica, é possível prever proteínas alvo para a fabricação de uma vacina ou de um fármaco (DONKOR, 2013).

1.1.1 Objetivo Geral

Encontrar alvos vacinais através da genômica comparativa e subtrativa em genomas de *Corynebacterium ulcerans*.

1.1.2 Objetivos Específicos

- Obter uma lista genes candidatos a alvos vacinais a partir dos dados do genoma central obtidos através da análise do pangenoma;
- Verificar as relações filogenômicas e de similaridade entre as cepas estudadas;
- Realizar a predição *in silico* das Ilhas de Patogenicidade (PAIs) de cada genoma;
- Utilizar a abordagem de genômica subtrativa para a predição de alvos vacinais e de drogas.

1.2 Revisão da Literatura

1.2.1 Microbiologia

O gênero *Corynebacterium* abrange um grupo diverso de espécies anaeróbicas facultativas, com formas pleomórficas que podem variar desde cocoides a bastonetes, são Gram-positivas e não esporogênicas, com alto conteúdo GC no DNA (FUNKE et al., 1997; JANDA, 1998). Este gênero é notadamente estudado, pois engloba espécies de importância médica, como a *C. diphtheriae* e a *C. ulcerans* que causam difteria em humanos (ZAKIKHANY; EFSTRATIOU, 2012); veterinária, como *C. pseudotuberculosis* que é a principal responsável por causar linfadenite caseosa em animais (DORELLA et al., 2006) e biotecnológica como a *Corynebacterium glutamicum* que é utilizada na fabricação de aminoácidos (KALINOWSKI et al., 2003).

As espécies *C. ulcerans*, *C. pseudotuberculosis* e *C. diphtheriae* formam um clado isolado das demais espécies do gênero, pois esses táxons têm em comum toxinas potentes, como a toxina da difteria (TD) que é transmitida por corinefagos específicos portadores do gene *tox* (FUNKE et al., 1997). Além da produção da TD, a exotoxina codificada pelo gene *pld* é um fator de virulência proeminente produzido por algumas linhagens de *C. ulcerans* e *C. pseudotuberculosis* (SELIM et al., 2016; SIMPSON-LOURÊDO et al., 2019).

1.2.2 Epidemiologia

A primeira vez que a *C. ulcerans* foi isolada foi em 1926 por Gilbert e Stewart a partir de amostras de uma infecção na garganta de um humano (GILBERT, R., STEWART, 1927). Entretanto, a classificação dessa espécie foi baseada nas características bioquímicas e morfológicas, por isso que em 1995 esse patógeno foi filogeneticamente classificado como uma espécie distinta dentro do grupo de organismos *C. diphtheriae*, através da análise molecular do DNA (FUNKE et al., 1997; PASCUAL et al., 1995; RIEGEL et al., 1995).

Baseado no contexto histórico, a difteria toxigênica tem como agente causador a *C. diphtheriae* (MOORE et al., 2015; WAGNER et al., 2012). No entanto, essa doença tem sido cada vez mais relatada em vários trabalhos, onde o responsável pelos sintomas em humanos é a *C. ulcerans* (HATANAKA et al., 2003; TIWARI et al.,

2008). Através desses, é possível notar uma gama de países que apresentam estudos de casos dessa bactéria (TABELA 1).

Com base na tabela 1, observa-se que a principal fonte de contaminação é oriunda de animais, principalmente, gatos e cachorros, embora outros animais podem servir como reservatório para infecção e transmissão dessa bactéria, onde a maioria apresenta sintomas de linfadenite caseosa (TABELA 2). Ainda não se tem confirmação de transmissão de pessoa para pessoa, embora existam trabalhos propondo essa hipótese (KONRAD; HÖRMANSDORFER; SING, 2015b), esse tipo de transferência é um fator crucial para que ocorra surtos da doença.

É importante lembrar que antes da vacina se tornar acessível no mundo todo, a difteria causou epidemias devastadoras. Na década de 1880, as taxas de difteria respiratória atingiram 50% de pessoas contaminadas em algumas áreas da Europa. E durante a primeira e segunda guerra mundial a epidemia causou cerca de 1 milhão de casos e 50.000 mortes em 1943 (WORLD HEALTH ORGANIZATION, 2017). Além disso, houve um surto de difteria com cerca de 140 mil casos entre 1991-1996 nos países da antiga União Soviética (VITEK; WHARTON, 1998).

Tabela 1 – Alguns casos de infecções causados por *C. ulcerans* em diversos países

País de residência	Idade	Vacinação	Quadro Clínico	Produção de toxina	Fonte da Infecção	Desfechos/ Observações	Referencias
Inglaterra	Adulto	ND	Assintomático	ND	Amostras de leite bovino	Recuperado	(BOSTOCK et al., 1984)
EUA	Adulto (54 anos)	ND	Difteria clássica, dispneia e miocardite	Positiva	ND	Recuperado	(“Respiratory Diphtheria Caused by <i>Corynebacterium ulcerans</i> - Terre Haute, Indiana, 1996”, 1997)
Alemanha	Adulto (40 anos)	ND	Ulceração cutânea	Positiva	ND	Recuperado	(SING et al., 2003)
Japão	Adulto (53 anos)	ND	Difteria clássica, dispneia grave e pneumonia	Positiva	Provavelmente um gato de rua que veio a óbito	Recuperado	(HATANAKA et al., 2003)
Brasil	Adulto (80 anos)	ND	Lesão cutânea e pneumonia	Positiva	ND	Óbito / paciente residente em região metropolitana	(MATTOS-GUARALDI et al., 2008)
EUA	Adulto (77 anos)	ND	Difteria clássica	Positiva	ND	Óbito / paciente residente em área rural	(TIWARI et al., 2008)
Alemanha	Adulto (47 anos)	Sim	Difteria clássica	Positiva	Porcos	Recuperado	(SCHUHEGGER et al., 2009)
França	Adulto (56 anos)	Sim	Difteria clássica, dispneia grave	Positiva	Cachorro	Recuperado	(LARTIGUE et al., 2005)
Suíça	Adulto (71 anos)	Sim	Difteria clássica	Positiva	ND	Recuperado	(KAUFMANN; OTT; ZBINDEN, 2002)

Holanda	Adulto (57 anos)	Não	Difteria clássica	Positiva	Provavelmente um gato de estimação	Recuperado	(VISSER et al., 2002)
Bélgica	Adulto (72 anos)	Sim	Ulceração cutânea	Positiva	Provavelmente um gato de estimação	Recuperado	(DETEMMERMAN et al., 2013)
Japão	Adulto (67 anos)	Sim	Difteria clássica e dispneia grave	Positiva	Gato de estimação	Recuperado	(YASUDA et al., 2018)
EUA	Adulto (73 anos)	Não	Conjuntivite	ND	ND	Recuperado	(WEIL et al., 2019)
Japão	Adulto (37 anos)	Sim	Abscesso subcutâneo	Positiva	ND	Recuperado	(URAKAWA et al., 2013)
França	Adulto (80 anos)	Sim	Lesão cutânea, sepse e falência de múltiplos órgãos	Positiva	Gato de rua	Óbito	(VANDENTORREN et al., 2014)

ND: Não Descrito

Tabela 2 – Características de *C. ulcerans* isolados de animais

Espécie de animal	Quant.	País	Quadro clínico	Deteção gene <i>tox</i>	Deteção gene <i>pld</i>	Referência
Dromedário (<i>Camelus dromedarius</i>)	1	ND	Abscesso caseoso	Ausente	Presente	(TEJEDOR et al., 2000)
Lontra (<i>Lutra lutra</i>)	2	Reino Unido	Encontrado morto e com dano pulmonar	Positiva	ND	(FOSTER et al., 2002)
Porco (<i>Sus scrofa domesticus</i>)	1	Alemanha	Assintomático	Positiva	ND	(SCHUHEGGER et al., 2009)
Cervo (<i>Capreolus capreolus</i>)	1	Alemanha	Abscesso caseoso	Ausente	Presente	(RAU et al., 2011)
Javali (<i>Sus scrofa</i>)	2	Alemanha	Abscessos nos linfonodos	Ausente	Presente	(CONTZEN et al., 2011)
Veado (<i>Cervidae</i>)	1	Alemanha	Abscesso caseoso	Ausente	Presente	(EISENBERG et al., 2014)
Furões (<i>Mustela putorius furo</i>)	4	EUA	Infecção no implante cefálico	4 ausente	4 presente	(MARINI et al., 2014)
Raposa vermelha (<i>Vulpes vulpes</i>)	1	Alemanha	ND	Presente	ND	(STING et al., 2015)
Coruja-dos-urais (<i>Strix uralensis</i>)	1	Japão	Assintomático	Presente	ND	(KATSUKAWA et al., 2016)
Urotrichus talpoides (<i>Urotrichus talpoides</i>)	2	Japão	Assintomático	Presente	ND	(KATSUKAWA et al., 2016)
Ouriço (<i>Erinaceus europaeus</i>)	1	Alemanha	Ferida cutânea profunda	Presente	ND	(BERGER et al., 2019)
Ouriço (<i>Erinaceus europaeus</i>)	3	Alemanha	Pneumonia	1 Ausente 2 Presente	ND ND	(BERGER et al., 2019)

ND: Não Descrito

1.2.3 Apresentação clínica e terapia

A difteria respiratória pode originar uma pseudomembrana acinzentada ou esbranquiçada que se instala na faringe causando um edema, podendo evoluir para a obstrução das vias aéreas e, em casos raros, pode gerar uma pneumonia grave ou até mesmo deixar sequelas como miocardite e neuropatia periférica. Por outro lado, a difteria cutânea é caracterizada por uma ulcera dolorosa (CARFORA et al., 2018; MOORE et al., 2015; OTSUJI et al., 2019; YASUDA et al., 2018).

A linfadenite caseosa causa nos linfonodos a formação de abscessos que contém uma massa de coloração amarelada de consistência viscosa a pastosa

disposta em camadas concêntricas, com zonas de necrose, demarcadas por uma cápsula (CONTZEN et al., 2011).

Geralmente as infecções causadas por *C. ulcerans* são tratadas com a antitoxina da difteria e antibióticos, entretanto, estudos têm apontado a existência de linhagens resistentes a alguns tipos de fármacos, como a eritromicina, clindamicina (MOORE et al., 2015; TIWARI et al., 2008), oxacilina (TEJEDOR et al., 2000) e a fosfomicina (VANDENTORREN et al., 2014).

Tendo em vista essa resistência a antimicrobianos e a escassez de vacinas e uma terapia antitoxina específica contra *C. ulcerans*, a grande quantidade de genomas depositados em bancos de dados pode ser utilizada por estratégias de bioinformática na busca por novos alvos vacinais e de drogas para a prevenção e tratamento da doença.

1.2.4 Bioinformática

A bioinformática é definida como um conjunto de técnicas computacionais para o entendimento e análise de problemas sobre questões biológicas (AYYILDIZ; PIAZZA, 2019). Através dela é possível projetar diversos tipos de estudos, como a análise e comparação de similaridade do material genético de espécies (TROST et al., 2012a), pesquisa por genes de resistência a antibióticos (LEE; KIM; CHA, 2021), origem e evolução de patógenos, interação patógeno-hospedeiro (JEAN BELTRAN et al., 2017), análises de vias metabólicas (BARBA et al., 2013), pesquisa por ilhas de patogenicidade e seus fatores de virulência (BELBAHRI et al., 2008), entre outros.

1.2.4.1 Genômica comparativa

A comparação por similaridade é um método imprescindível no estudo do pangenoma de cepas ou de um clado. Desta forma, são utilizados softwares, como o BLAST (Basic Local Alignment Search Tool), que é uma ferramenta heurística muito popular na bioinformática, que além de realizar alinhamentos de sequências, fornece informações estatísticas sobre a significância de cada alinhamento (CAMACHO et al., 2009).

O pangenoma é definido como o repertório de genes de uma determinada espécie ou de um clado. De acordo com a definição, os genes são divididos em três grupos: o primeiro engloba o genoma central que é um grupo de genes que são

compartilhados por todas as linhagens, já o segundo é definido como o genoma acessório e apresenta genes compartilhados por um subconjunto de linhagens e, por fim, os genes específicos de uma cepa (CHEN et al., 2018; XIAO et al., 2015).

O estudo do pangenoma apresenta uma ampla gama de aplicações, principalmente em pesquisas bacterianas, onde uma série de estudos foram desenvolvidos, como na evolução de espécies (FARAG; YOUSSEF; ELSHAHED, 2017; LAWLEY et al., 2017; SYME et al., 2018), investigações de mecanismos patogênicos e de resistência (SUN et al., 2020; YANG et al., 2019), detecção de alvos vacinais e de fármacos (HASSAN et al., 2016; NAZ et al., 2019).

1.2.4.2 Plasticidade genômica

A dinâmica em que o genoma perde, ganha ou rearranja informações genéticas é definida como a plasticidade genômica. Sem isso a evolução adaptativa de bactérias patogênicas não seria possível, já que a aquisição simultânea de muitos genes por transferência horizontal permite a herança de características complexas relacionadas à doença (DOBRINDT et al., 2004).

Estudos filogenéticos e epidemiológicos buscam entender os métodos que induzem a bactéria a sofrer plasticidade, pois há diversos casos que a bactéria patogênica se torna mais virulenta, podendo causar surtos de doenças tanto em humanos quanto em animais (NAVARRO-GARCIA et al., 2019).

Assim, se torna necessário ter precaução com a *C. ulcerans*, pois diferente da *C. diphtheriae*, ela tem a capacidade de ter como hospedeiro várias espécies de animais, isso facilita com que a bactéria sofra variação genética, podendo torná-la mais perigosa para a saúde humana e animal (MOORE et al., 2015; OTSUJI et al., 2019; TIWARI et al., 2008).

1.2.4.3 Ilhas de patogenicidade

Ilhas de patogenicidade, ou PAIs - do inglês Pathogenicity Islands - são um subtipo de ilhas genômicas as quais contém elementos genéticos adquiridos via transmissão horizontal capazes de induzir patogenicidade na bactéria ou mesmo aumentar sua virulência. São frequentemente encontradas no genoma acessório e sua sequência e composição pode variar imensamente entre espécies ou mesmo entre linhagens. Estas incluem genes que conferem resistência a antibióticos e

compostos tóxicos e outras formas de virulência (DOBRINDT et al., 2004; HALLSTROM; MCCORMICK, 2014).

Geralmente as ilhas genômicas podem possuir uma ou mais propriedades em comum, como a habilidade de sofrer recombinação, conteúdo GC que difere do restante do genoma, associação com tRNA e a presença de repetições em ambos os flancos das ilhas genômicas. PAIs são muitas vezes instáveis e podem conter genes de mobilidade que codificam integrases ou transposases (DOBRINDT et al., 2004; HALLSTROM; MCCORMICK, 2014).

1.2.4.4 Genômica subtrativa

A genômica subtrativa é uma análise baseada na genômica comparativa, que utiliza o genoma do hospedeiro e do patógeno para a identificação de genes essenciais e que não tenham nenhuma similaridade com material genético do hospedeiro. Desse modo, a droga ou composto projetados atuam especificamente no(s) alvo(s), sem prejudicar a biologia do hospedeiro. Os alvos identificados podem ser utilizados para planejar e desenvolver medicamentos, vacinas ou alvos de dupla finalidade (BARH et al., 2011a).

A descoberta de drogas normalmente começa com a identificação e validação de um alvo potencial. Alvos de drogas adequados compartilham várias características em comum: devem ser necessários para a sobrevivência, adaptabilidade e crescimento do patógeno; drogável (capaz de se ligar a pequenas moléculas, implicando na presença de um sítio de ligação); caracterizado funcionalmente e estruturalmente; e deve ser distinto de alvos de drogas conhecidos para evitar resistência cruzada. A predição de alvos de drogas depende de múltiplas comparações para identificar genes que se adequam nesses parâmetros (BAKHEET; DOIG, 2009; HOLMAN et al., 2009).

A genômica subtrativa é uma técnica poderosa, rápida e econômica para a identificação de alvos de drogas e vacinais para qualquer patógeno, mas é necessário que os genomas do patógeno e do hospedeiro estejam disponíveis em bancos de dados. Entretanto, torna-se necessário realizar a validação experimental para constatar se os alvos preditos são realmente eficazes (BARH et al., 2011b). Em várias bactérias patogênicas já foram preditos alvos vacinais através da abordagem *in silico*

como a *Mycoplasma pneumoniae* (KUMAR et al., 2010; RODRIGUES et al., 2019), *C. pseudotuberculosis* (ARAÚJO et al., 2019; BARH et al., 2011b; SOARES et al., 2013a), *Staphylococcus aureus* (UDDIN; SAEED, 2014), *Streptococcus pneumoniae* (WADOOD et al., 2018) e *Treponema pallidum* (JAISWAL et al., 2017).

Mesmo que já exista uma gama de antibióticos disponíveis, há inúmeras linhagens emergentes que estão cada vez mais resistentes a estes medicamentos, tornando-as mais difíceis de controlar (ARIAS; MURRAY, 2009). Com base nisso, investimentos contínuos em melhores drogas e vacinas são extremamente necessários, pois as linhagens toxigênicas continuam representando uma ameaça à saúde humana e animal (MATTOS-GUARALDI et al., 2003; OTSUJI et al., 2019).

2. GENÔMICA COMPARATIVA E SUBTRATIVA ALIADA À VACINOLOGIA REVERSA PARA IDENTIFICAÇÃO DE CANDIDATOS À ALVOS VACINAIS CONTRA A *Corynebacterium ulcerans*

2.1 Introdução

Corynebacterium ulcerans é um membro patogênico do gênero *Corynebacterium*, que faz parte da família *Corynebacteriaceae*, da ordem *Actinomycetales* e do filo *Actinobacteria* (TAUCH; SANDBOTE, 2014). É um patógeno tanto de interesse médico quanto veterinário. As linhagens dessa bactéria infectadas por corinefagos específicos portadores do gene *tox* se tornam toxigênicas, representando então um dos agentes causadores de difteria, patologia que pode acometer o sistema respiratório ou causar ulcerações na pele de humanos (MEINEL et al., 2014; SANGAL; HOSKISSON, 2014). Este microrganismo é transportado de forma sintomática ou assintomática por uma variedade de animais, que podem servir como fonte de reservatório e transmissão para possível infecção zoonótica (BERGER et al., 2019; TIWARI et al., 2008).

É crescente o número de casos de infecções causadas por *C. ulcerans*, esse patógeno já foi encontrado em diversos países (SUBEDI et al., 2018). Historicamente, os casos de difteria foram frequentemente causados por *C. diphtheriae*, mas atualmente o número de relatos de difteria causada por *C. ulcerans* têm ganhado mais evidência, pois as linhagens apresentam uma variabilidade genética maior em relação a outra bactéria. (MEINEL et al., 2014; OTSUJI et al., 2019).

Além disto, a *C. ulcerans* pode causar doenças distinguíveis da difteria clássica (FUNKE et al., 1997), pois este microrganismo detém um extenso e variado conjunto de fatores de virulência, tais como a presença de um pili adesivo do tipo SpaDEF, fosfolipase D (*pld*), neuraminidase H (*NanH*), endoglicosidase E (*EndoE*) e o gene *rpb*, que codifica uma proteína com semelhança estrutural da toxina tipo Shiga SLT-1 e SLT-2 de *Escherichia coli* (SANGAL et al., 2014; TROST et al., 2011).

A genômica comparativa consiste em comparar o material genético de espécies ou clado, com diversos tipos de finalidades, como exemplo, estudar a evolução dos genomas, a caracterização do pangenoma e seus subconjuntos, determinantes de virulência, localizar possíveis alvos vacinais contra doenças infecciosas, dentre outras aplicações (BARAÚNA et al., 2017).

Com base em sua importância e por meio da grande quantidade de genomas de *C. ulcerans* depositados no banco de dados do *National Center for Biotechnology Information* (NCBI), é possível promover uma análise comparativa para avaliar características moleculares que contribuem para a virulência e patogenicidade dessa espécie, bem como prever possíveis alvos vacinais e de fármacos.

2.2 Material e Métodos

2.2.1 Aquisição das sequências genômicas

As sequências genômicas no formato fasta da *C. ulcerans* utilizadas neste trabalho foram obtidas a partir do banco de dados do GenBank do NCBI (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/3461/>) depositadas até a data de janeiro de 2020. Os genomas das 29 linhagens foram anotados através do software Prokka v1.12 (SEEMANN, 2014) com o intuito de minimizar o viés da predição realizada por diferentes preditores.

2.2.2 Identificação e análise de Ilhas de Patogenicidade

A predição das Ilhas de Patogenicidade (PAI) em cada genoma, foi realizada através da ferramenta GIPSY: *Genomic Island Prediction Software* v.1.1.2 (SOARES et al., 2016), utilizando-se a *C. glutamicum* 13032 (número de acesso: NZ_CP025533.1) como genoma de referência. Através dessa análise, obteve-se um documento no formato texto (.txt) para cada genoma submetido, indicando as prováveis ilhas preditas bem como a região do genoma na qual estavam localizadas, o tamanho e o grau de

confiabilidade da predição (fraca, normal ou forte). A sequência dos genes de cada PAI, que apresentaram grau forte ou normal, foi recuperada usando o software Artemis v.2.14.0 (CARVER et al., 2012). Para visualizar melhor a localização dessas PAIs, um mapa circular foi construído através do programa BLAST Ring Image Generator (BRIG) (ALIKHAN et al., 2011).

Dando prosseguimento, foi realizado um alinhamento através do algoritmo BLASTn (CAMACHO et al., 2009). Para isso foi criado um arquivo multifasta contendo todas as Ilhas de Patogenicidade preditas, o qual foi formatado como banco de dados, para que então cada um dos genomas dotados de PAI fosse alinhado contra eles mesmos. Foram gerados dois arquivos de saída sendo um no formato tabular e outro no formato não tabular, ambos indicando com quais outras ilhas de patogenicidade aquela ilha em estudo possui similaridade. Foi utilizado o *E-value* dentro dos níveis indicados de significância, ou seja, com o valor de corte de 0.0 e a porcentagem de identidade entre as sequências acima de 80%, pois através desses parâmetros conseguimos assumir que duas PAI eram realmente similares.

2.2.3. Análise filogenômica e similaridade

A construção da árvore filogenômica e a porcentagem de similaridade entre os genomas baixados foi realizada através da ferramenta Gegenees v3.1 (ÅGREN et al., 2012). Para isso, foi utilizado um protocolo de alta precisão, com a fragmentação dos genomas em 200 pb e step size 100 pb. A árvore filogenômica foi exibida e editada utilizando a ferramenta Interactive tree of life (iTOL) (LETUNIC; BORK, 2016). As linhagens que obtiveram uma baixa porcentagem de similaridade foram submetidas a ferramenta ANI (Average Nucleotide Identity), para verificar se elas realmente pertenciam a espécie *C. ulcerans*.

Para verificar em quais linhagens o gene *tox* e o *pld* estavam presentes, foi utilizada a ferramenta BLASTn do NCBI, onde foram utilizadas as sequências com os números de acesso AY141013 e WP_013240889 no GenBank, respectivamente. As sequências de nucleotídeos dos dois genes em análise que estavam presentes nas linhagens de *C. ulcerans* foram recuperadas em formato fasta. Para um estudo comparativo, também foi feito o download da sequência do gene *tox* presente em cinco linhagens de *C. diphtheriae* e do gene *pld* presente em cinco linhagens de *C. pseudotuberculosis*. Separadamente, todas as sequências foram alinhadas no

MUSCLE (EDGAR, 2004), para assim serem comparadas filogeneticamente utilizando o método Neighbor-Joining (SAITOU; NEI, 1987) no software MEGA X (KUMAR et al., 2018).

2.2.4 Predição do pangenoma, core genoma e singletons

Através do PGAP (<http://pgaweb.vlcc.cn/>), foi executada a identificação do pangenoma das sequências genômicas em análise, levando em consideração o método Gene Family (GF) para identificar genes ortólogos. Por se tratarem de genomas da mesma espécie, foram utilizados parâmetros de identidade e cobertura de 80% (CHEN et al., 2018; ZHAO et al., 2012).

A Lei de Heap foi utilizada para saber se o pangenoma da espécie é aberto ou fechado. Essa lei é representada pela fórmula $n = k \cdot N^\alpha$, onde, no contexto genético, a variável n é o número esperado de genes para um determinado número de genomas, N determina o número de genomas e k e α ($\alpha = 1 - \gamma$) são parâmetros livres determinados empiricamente. Segundo essa lei, quando $\alpha > 1$ ($\gamma < 0$) o pangenoma é considerado fechado e não há aumento significativo de genes à medida que novos genomas são adicionados à análise. Por outro lado, quando $\alpha < 1$ ($0 < \gamma < 1$), o pangenoma é considerado aberto, ou seja, haverá aumento significativo de novos genes para cada novo genoma adicionado (SOARES et al., 2013b).

2.2.5 Identificação dos genes essenciais

Com base no resultado gerado pelo PGAP, o genoma central das linhagens de *C. ulcerans* foi subtraído, em formato GenBank, e alinhado contra as sequências proteicas disponíveis no banco de dados de genes essenciais (DGE) (<http://www.essentialgene.org/>), adotando como parâmetro o $e\text{-value} = 1e^{-05}$ (parâmetro padrão), para identificação de genes indispensáveis à sobrevivência de *C. ulcerans*.

2.2.6 Triagem dos genes não homólogos ao hospedeiro

A partir disso, foi gerado um arquivo no formato fasta contendo a sequência das proteínas essenciais ao patógeno, cada uma delas foi submetida à comparação com o proteoma humano, com intuito de triar somente as regiões codificantes que não foram análogas ao genoma do hospedeiro. Esses alinhamentos foram realizados na

ferramenta BLASTp especificamente contra o banco de dados de *homo sapiens* (taxid: 9606), onde os valores de alinhamento padrões foram admitidos.

2.2.7 Localização subcelular e previsão de antígenos

A localização subcelular das proteínas, eleitas na etapa anterior, foi predita utilizando a ferramenta PSORTb versão 3.0 (YU et al., 2010). Posteriormente, as proteínas secretadas e de membrana foram submetidas a predição das propriedades antigênicas através do servidor web VaxiJen (DOYTCHINOVA; FLOWER, 2007), onde foi utilizado um limiar, que frequentemente é utilizado na literatura, de 0,5 para classificar com base nas propriedades físico-químicas, se a proteína realmente tinha um potencial antigênico. A identificação dessas proteínas permite aos pesquisadores direcioná-las para serem manipuladas como uma vacina contra a infecções bacterianas (MUHAMMAD AZAM et al., 2020). As proteínas citoplasmáticas não foram descartadas da análise, uma vez que elas podem ser utilizadas como alvos de fármacos.

2.3 Resultados e discussão

2.3.1 Dados genômicos de *C. ulcerans*

Foi gerada uma tabela para detalhar as informações genotípicas extraídas com base na anotação feita pelo prokka, de cada linhagem analisada. O número de genes presente variou entre 2.233 a 2.594 e o tamanho do genoma variou entre 2,42 a 2,61 Mb (TABELA 3). Também, é possível notar algumas informações biológicas como a variedade de hospedeiro e de países onde essa bactéria foi isolada, todas essas informações foram coletadas no NCBI.

2.3.2 Análise de similaridade e filogenômica dos genomas baixados

Do total de 29 genomas de *C. ulcerans* disponíveis, 17 estavam completos e 12 em contigs (montagem incompleta). Na figura 1 pode ser observado a porcentagem de similaridade entre os genomas de *C. ulcerans*, sendo que as duas linhagens que obtiveram uma baixa porcentagem de similaridade com relação às outras linhagens são a PO100/5 e w25. Com o intuito de confirmar que estas linhagens pertencem a mesma espécie que as demais, calculamos a porcentagem de similaridade no ANI levando em consideração o genoma de referência da *C. ulcerans* NCTC 7910, e obtivemos um valor de 89.82% para a linhagem PO100/5 e 89.79% para a linhagem

w25 (YOON et al., 2017). Para ser considerado da mesma espécie, os valores ANI devem estar acima de 95%, como os valores obtidos estão abaixo desta porcentagem, consideramos que essas duas cepas não pertencem a espécie *C. ulcerans*. Corroborando com os resultados obtidos anteriormente, dois artigos publicados recentemente descreveram que essas duas linhagens fazem parte da espécie *Corynebacterium silvaticum* (GORIS et al., 2007; MÖLLER et al., 2020; VIANA et al., 2021). Através disso, essas linhagens foram removidas das outras análises.

Tabela 3 – Características e informações dos genomas analisados neste trabalho

RefSeq	Linhagem	Tamanho (Mb)	GC%	Número de Genes	Hospedeiro	Data do isolamento	País	Motivo do isolamento	Data de lançamento	Referência
NZ_LT906443.1	NCTC7910	2.45	53.3	2,253	Humano	1948	Reino Unido	Caso suspeito de difteria	15/08/2017	ND
NC_015683.1	BR-AD22	2.61	53.4	2,451	Cachorro	2010	Brasil	Assintomático	02/06/2011	(TROST et al., 2011)
NC_018101.1	0102	2.58	53.4	2,395	Humano	2001	Japão	Difteria Clássica	02/07/2012	(SEKIZUKA et al., 2012)
NZ_AP019663.1	FH2016-1	2.58	53.4	2,391	Humano	2016	Japão	Caso suspeito de difteria	05/09/2019	ND
NZ_AP019662.1	0211	2.58	53.4	2,396	Humano	2002	Japão	Caso suspeito de difteria	05/09/2019	ND
NZ_CP021417.1	PO100/5	2.57	54.4	2,594	Porco	2014	Portugal	Lesões de linfadenite caseosa	31/05/2017	ND
NZ_CP011913.1	FRC58	2.54	53.3	2,377	Humano	ND	França	Bronquite	26/06/2015	(SILVA et al., 2014)
NZ_CP009583.1	210931	2.5	53.3	2,331	Humano	ND	ND	ND	22/10/2014	ND
NC_017317.1	809	2.5	53.3	2,293	Humano	2000	Brasil	Infecção pulmonar fatal	02/06/2011	(TROST et al., 2011)
NZ_CP009500.1	210932	2.48	53.3	2,280	Humano	ND	ND	ND	10/09/2014	(VIANA et al., 2014)
NZ_CP010818.1	131001	2.48	53.3	2,282	Humano	ND	ND	ND	14/09/2015	ND
NZ_CP009716.1	05146	2.47	53.3	2,253	ND	ND	ND	ND	28/10/2014	ND
NZ_LS483416.1	NCTC8639	2.45	53.3	2,247	ND	1900/1951	Reino Unido	ND	17/06/2018	ND
NZ_LS483400.1	NCTC7908	2.45	53.3	2,254	Humano	1946	ND	Assintomático	17/06/2018	ND
NZ_CP009622.1	FRC11	2.44	53.3	2,229	Humano	ND	França	Úlcera cutânea	22/10/2014	(BENEVIDES et al., 2016)
NZ_CP011095.1	131002	2.43	53.4	2,233	Humano	ND	ND	ND	03/04/2015	ND

NZ_CP046863.1	MRI49	2.53	53.2	2,335	Cavalo	ND	Reino Unido: Escócia	Faringite, sinusite e doença extrafaríngea	21/12/2019	ND
NZ_UFXR00000000.1	NCTC8666	2.54	53.4	2,378	ND	1900/1951	ND	ND	30/07/2018	ND
NZ_JZUS00000000.1	LSPQ-04227	2.43	53.4	2,425	Humano	2004	Canadá	Difteria Clássica	05/05/2015	ND
NZ_AYUJ00000000.1	NCTC 12077	2.62	53.4	2,501	Humano	ND	ND	ND	25/11/2013	ND
NZ_MPSS00000000.1	2590	2.5	53.3	2,319	Humano	2014	Brasil	Faringite	08/01/2018	(SUBEDI et al., 2018)
NZ_VFEM00000000.1	W25	2.55	54.4	2,576	Javali	2009	Alemanha	ND	17/06/2019	ND
NZ_MPST00000000.1	BR-AD 2649	2.54	53.3	2,369	Cachorro	2015	Brasil	Assintomático	08/01/2018	(SUBEDI et al., 2018)
NZ_JZUT00000000.1	LSPQ-04228	2.44	53.4	2,506	Humano	2013	Canadá	Úlcera cutânea	05/05/2015	ND
NZ_LJVH00000000.1	04-7514	2.5	53.5	2,414	Cachorro	2004	França	ND	28/09/2015	(GUIMARÃES et al., 2016a)
NZ_MIOS00000000.1	KZN-2016-48390	2.54	53.4	2,365	Humano	2016	África do Sul	Difteria Clássica	10/11/2016	(DU PLESSIS et al., 2017)
NZ_LSWN00000000.1	4940	2.42	53.3	2,233	Humano	2009	Bielorrússia	Assintomático	10/01/2018	(SUBEDI et al., 2018)
NZ_LGSX00000000.1	04-3911	2.49	53.3	2,320	Humano	2004	França	ND	24/09/2015	(GUIMARÃES et al., 2016b)
NZ_LGSY00000000.2	03-8664	2.43	53.5	2,385	Humano	2003	França	Difteria Clássica	24/09/2015	(LARTIGUE et al., 2005)

ND: Não Descrito

Fonte: NCBI

Figura 1 – Heatmap com a análise de similaridade entre os genomas de *C. ulcerans*

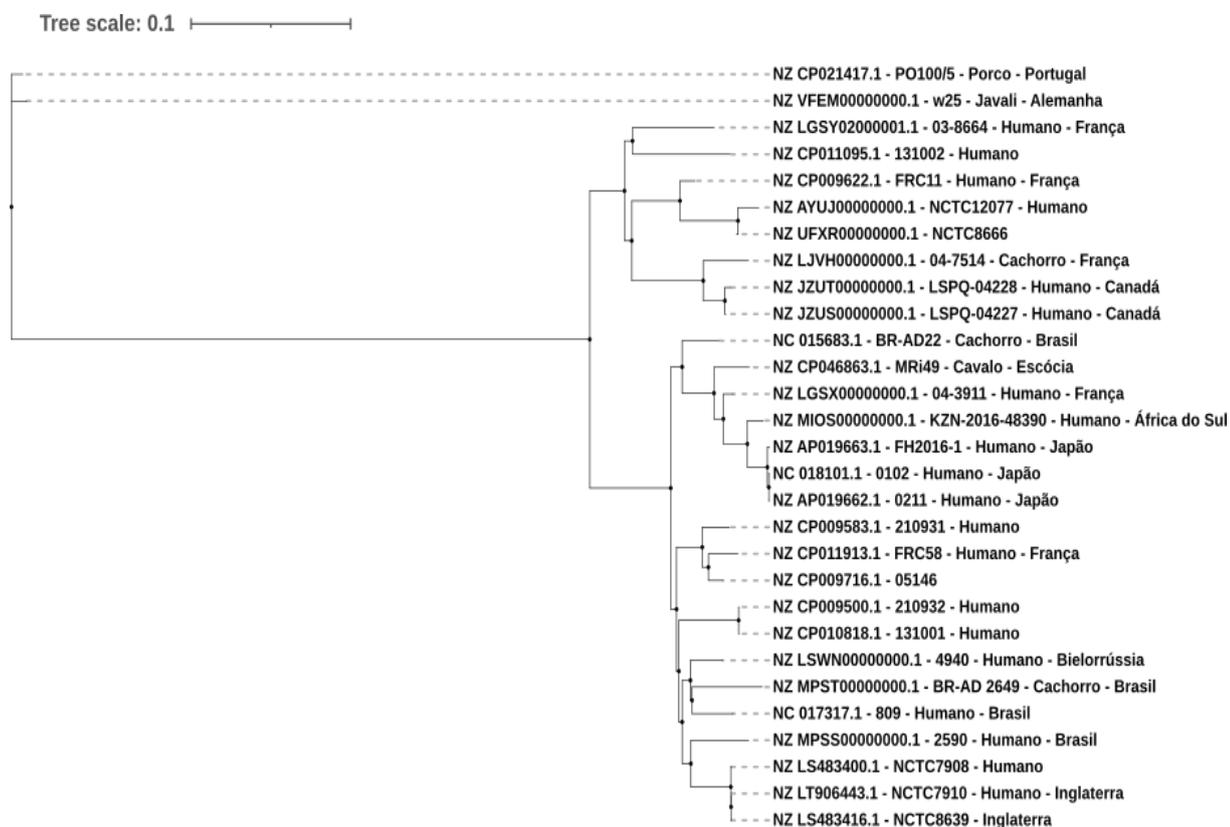
Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1: NC_015683.1	100	89	89	89	90	90	89	89	89	88	88	88	88	90	90	92	93	93	93	79	77	76	77	77	78	82	80	53	53
2: NZ_MPST00000000.1	90	100	88	89	89	91	89	89	89	89	88	88	88	88	88	88	90	90	90	80	77	77	78	78	77	82	80	54	54
3: NZ_MPSS00000000.1	92	90	100	91	92	91	90	90	91	91	91	91	91	91	91	91	91	91	81	80	79	80	80	79	82	82	55	55	
4: NZ_CP011913.1	91	90	90	100	94	91	91	91	94	90	90	90	90	93	93	93	93	93	81	80	79	79	79	79	83	82	55	55	
5: NZ_CP009583.1	93	91	92	95	100	93	92	92	96	92	92	92	92	92	92	92	92	92	82	81	80	81	81	80	84	84	55	55	
6: NC_017317.1	94	93	92	92	93	100	92	92	92	93	92	92	92	92	92	92	93	93	83	80	80	81	81	80	84	83	55	55	
7: NZ_CP009500.1	93	91	92	93	92	93	100	100	92	92	92	92	92	92	92	92	94	94	83	81	81	81	81	80	84	83	56	56	
8: NZ_CP010818.1	93	91	92	93	92	93	100	100	92	92	92	92	92	92	92	92	94	94	83	81	81	81	81	80	84	83	56	56	
9: NZ_CP009716.1	94	92	93	97	98	94	93	93	100	93	93	93	93	95	95	95	95	95	84	82	82	82	82	82	84	84	56	56	
10: NZ_LSWN00000000.1	94	93	94	94	95	95	93	93	94	100	94	94	94	94	94	94	94	94	85	82	82	83	83	82	85	85	57	57	
11: NZ_LT906443.1	93	92	94	94	94	94	93	93	94	94	100	100	100	93	93	93	93	93	84	82	82	82	82	81	84	84	56	56	
12: NZ_LS483416.1	93	92	94	94	94	94	93	93	94	94	100	100	100	93	93	93	93	93	84	82	82	82	82	81	84	84	56	56	
13: NZ_LS483400.1	93	92	94	94	94	94	93	93	94	94	100	100	100	93	93	93	93	93	84	82	82	82	82	81	84	84	56	56	
14: NZ_LGSX00000000.1	94	90	92	94	93	92	92	92	94	92	91	91	91	100	98	100	99	99	83	81	80	81	81	82	83	83	56	56	
15: NZ_CP046863.1	93	89	90	93	91	91	91	91	93	91	91	91	91	96	100	97	96	96	81	79	79	80	80	79	82	83	55	55	
16: NZ_MIOS00000000.1	93	88	89	92	91	90	89	89	91	89	89	89	89	97	95	100	98	98	80	79	78	79	79	81	83	81	54	54	
17: NZ_AP019663.1	94	89	89	91	90	90	90	90	88	89	89	89	89	96	94	98	100	100	80	79	77	78	78	81	83	80	54	54	
18: NC_018101.1	94	89	89	91	90	91	90	90	88	89	89	89	89	96	94	98	100	100	80	79	77	78	78	81	83	80	54	54	
19: NZ_AP019662.1	94	89	89	91	90	90	90	90	88	89	89	89	89	96	94	98	100	100	80	79	77	78	78	81	83	80	54	54	
20: NZ_CP009622.1	84	84	84	85	84	85	84	84	85	84	85	85	85	84	84	84	84	84	100	90	89	91	91	90	100	100	57	57	
21: NZ_LGSY02000001.1	83	82	83	84	84	83	83	83	83	83	83	83	83	84	83	84	84	84	84	90	100	89	89	89	89	91	91	58	58
22: NZ_CP011095.1	82	81	82	82	82	82	82	82	83	82	82	82	82	82	82	82	82	82	82	89	89	100	89	89	88	89	90	58	58
23: NZ_JZUT00000000.1	82	81	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	90	87	88	100	99	97	90	90	57	57	
24: NZ_JZUS00000000.1	82	81	82	83	83	82	82	82	83	82	83	83	83	82	82	82	82	82	91	88	89	100	100	98	91	91	57	57	
25: NZ_LJVH00000000.1	81	79	80	80	81	80	80	80	80	80	80	80	80	82	80	83	83	83	88	86	86	96	96	100	90	88	56	56	
26: NZ_AYUJ00000000.1	82	80	80	81	80	81	80	80	79	79	79	79	79	79	79	81	82	83	94	85	83	85	85	86	100	96	53	53	
27: NZ_UFXR00000000.1	82	81	82	82	83	81	81	81	81	81	81	81	81	81	82	82	82	82	96	87	86	87	87	87	99	100	55	55	
28: NZ_CP021417.1	54	54	54	54	54	54	54	54	54	54	54	53	54	54	54	54	54	54	54	54	54	54	55	55	55	54	54	100	98
29: NZ_VFEM00000000.1	54	54	53	54	54	53	54	54	54	54	53	53	53	54	54	54	54	54	54	54	54	54	54	54	54	54	54	98	100

Fonte: o autor

Por meio da árvore filogenômica (FIGURA 2) é possível visualizar que as cepas PO100/5 e a W25 são polifiléticas e evolutivamente mais distantes em relação as demais linhagens. De modo geral, nota-se que a árvore ramificou em dois grandes galhos, em que as linhagens 03-8664, 131002, FRC11, NCTC 12077, NCTC 8666, 04-7514, LSPQ- 04228 e LSPQ-04227 descenderam de um mesmo ancestral, e o outro grupo de linhagens (BR-AD22, MRi49, 04-3911, KZN-2016-48390, FH2016-1, 0102, 0211, 210931, FRC58, 05146, 210932, 131001, 4940, BR-AD 2649, 809, 2590, NCTC 7908, NCTC 7910, NCTC 8939) também formaram um clado.

As linhagens 0102, 0211 e FH2016-1 formaram um clado similar entre si, sendo que todas elas foram isoladas do mesmo hospedeiro e no mesmo país. Em contraponto, as linhagens oriundas da França (03-8664, FRC11, 04-7514, 04-3911, FRC58), onde a maioria foram isoladas da mesma espécie de hospedeiro, apresentaram-se em ramos diferente da árvore, indicando que a espécie é pouco clonal. Esses resultados corroboraram com a árvore criada por Subedi et al. (2019), onde foram utilizadas 19 das 29 linhagens presentes nesse estudo.

Figura 2 – Árvore filogenômica mostrando as relações evolutivas dentro da espécie *C. ulcerans*. As designações de número de referência da sequência seguidas por denominação das cepas, hospedeiro e país de isolamento, quando essas informações estavam disponíveis

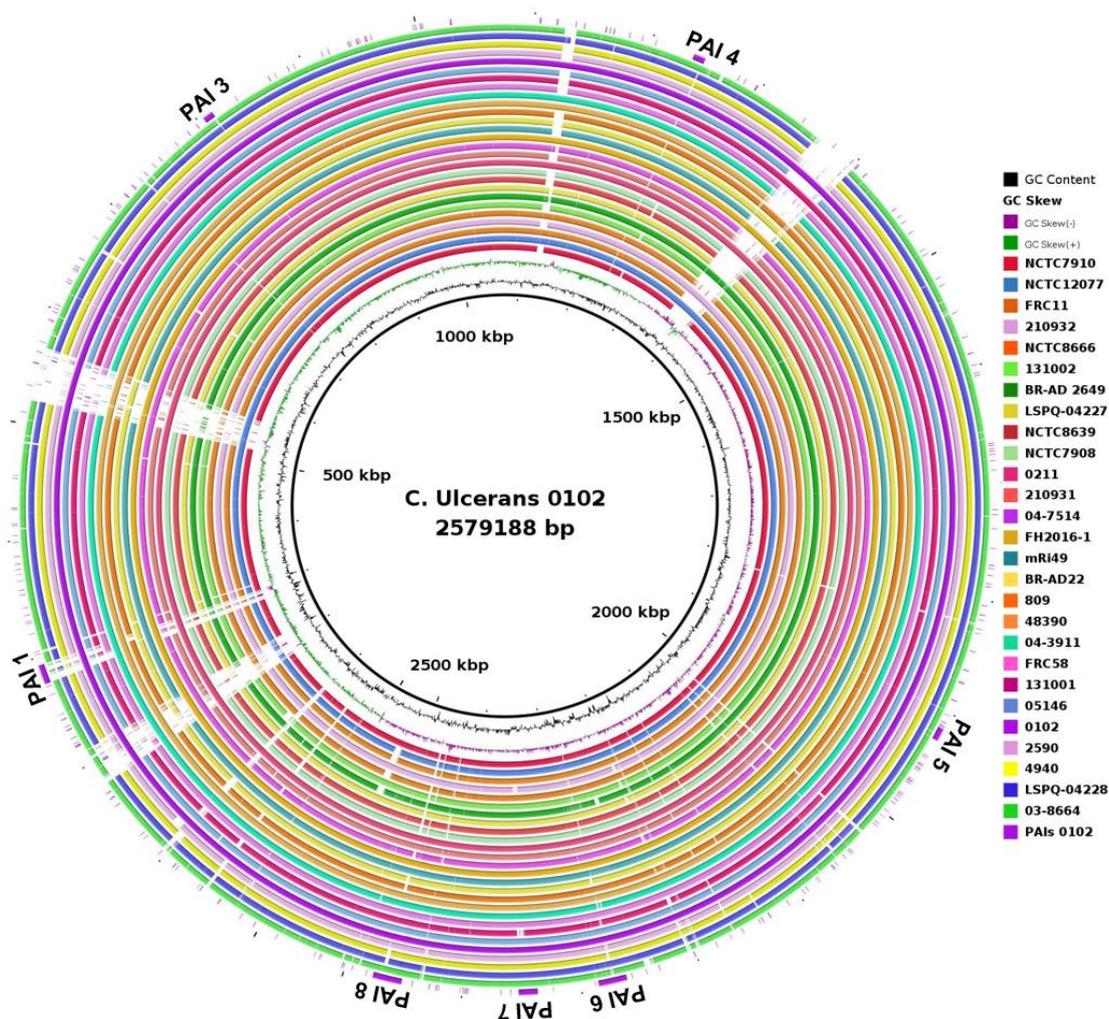


Fonte: o autor

2.3.3 Predição e representação circular das Ilhas de patogenicidade em *C. ulcerans*

Com base no tamanho do genoma, por não apresentar PAIs exclusivas e por apresentar ilhas que tiveram um maior número de homologias com as PAIs dos outros genomas, foi utilizada a linhagem 0102 como referência, e suas respectivas PAIs, para criar o mapa circular. É importante ressaltar que a PAI 2 foi removida da análise, pois era caracterizada como fraca (FIGURA 3). No mapa circular, é possível identificar a posição das ilhas em cada genoma, e que as PAIs 1, 4, 6 e 8 da 0102 são homólogas apenas a algumas linhagens. Em uma comparação geral, os gaps identificados no mapa circular são decorrentes da plasticidade do genoma da cepa 0102 em relação as demais.

Figura 3 – Representação genômica circular de ilhas de patogenicidade nos genomas de *C. ulcerans* utilizando a cepa 0102 como referência



Fonte: o autor

Do total de 27 genomas submetidos ao GIPSy, verificou-se que todos eles possuíam predições significativas de ilhas de patogenicidade com uma média de 6 PAIs por genoma, totalizando 156 ilhas preditas (TABELA 4). No trabalho de (BARAÚNA et al., 2017), com *C. pseudotuberculosis* biovar equi, foram identificados uma média de 18 PAIs por genoma sequenciado, usando o mesmo software. A partir disto, nota-se que esta espécie possui mais PAIs agrupadas em relação a *C. ulcerans*.

Tabela 4 – Caracterização das cepas de *C. ulcerans* quanto a presença de Ilhas de Patogenicidade

RefSeq	Cepa	Nível	Total de PAI	Total de proteínas da menor PAI	Total de proteínas da maior PAI (pb)	Total de Homologias	Nome do organismo
NZ_LT906443.1	NCTC7910	completo	5	7	16	93	<i>Corynebacterium ulcerans</i>
NC_015683.1	BR-AD22	completo	5	7	16	77	<i>Corynebacterium ulcerans</i> BR-AD22
NC_018101.1	0102	completo	7	7	20	113	<i>Corynebacterium ulcerans</i> 0102
NZ_AP019663.1	FH2016-1	completo	5	7	20	82	<i>Corynebacterium ulcerans</i>
NZ_AP019662.1	0211	completo	6	7	20	95	<i>Corynebacterium ulcerans</i>
NZ_CP011913.1	FRC58	completo	7	8	54	101	<i>Corynebacterium ulcerans</i>
NZ_CP009583.1	210931	completo	6	8	24	106	<i>Corynebacterium ulcerans</i>
NC_017317.1	809	completo	7	7	53	107	<i>Corynebacterium ulcerans</i> 809
NZ_CP009500.1	210932	completo	8	6	21	110	<i>Corynebacterium ulcerans</i>
NZ_CP010818.1	131001	completo	8	6	21	109	<i>Corynebacterium ulcerans</i>
NZ_CP009716.1	5146	completo	6	10	36	81	<i>Corynebacterium ulcerans</i>
NZ_LS483416.1	NCTC8639	completo	4	10	70	67	<i>Corynebacterium ulcerans</i>
NZ_LS483400.1	NCTC7908	completo	5	10	70	79	<i>Corynebacterium ulcerans</i>
NZ_CP009622.1	FRC11	completo	6	7	15	103	<i>Corynebacterium ulcerans</i> FRC11
NZ_CP011095.1	131002	completo	5	7	18	86	<i>Corynebacterium ulcerans</i>

NZ_CP046863.1	MRi49	completo	6	7	19	103	<i>Corynebacterium ulcerans</i>
NZ_UFXR00000000.1	NCTC8666	incompleto	2	15	70	25	<i>Corynebacterium ulcerans</i>
NZ_JZUS00000000.1	LSPQ-04227	incompleto	6	7	21	72	<i>Corynebacterium ulcerans</i>
NZ_AYUJ00000000.1	NCTC 12077	incompleto	6	7	53	78	<i>Corynebacterium ulcerans</i> NCTC 12077
NZ_MPSS00000000.1	2590	incompleto	7	7	21	98	<i>Corynebacterium ulcerans</i>
NZ_MPST00000000.1	BR-AD 2649	incompleto	6	6	54	48	<i>Corynebacterium ulcerans</i>
NZ_JZUT00000000.1	LSPQ-04228	incompleto	5	7	53	80	<i>Corynebacterium ulcerans</i>
NZ_LJVH00000000.1	04-7514	incompleto	6	7	25	89	<i>Corynebacterium ulcerans</i>
NZ_MIOS00000000.1	KZN-2016-48390	incompleto	6	7	18	100	<i>Corynebacterium ulcerans</i>
NZ_LSWN00000000.1	4940	incompleto	6	7	16	106	<i>Corynebacterium ulcerans</i>
NZ_LGSX00000000.1	04-3911	incompleto	6	7	16	117	<i>Corynebacterium ulcerans</i>
NZ_LGSY00000000.2	03-8664	incompleto	4	7	15	18	<i>Corynebacterium ulcerans</i>

Fonte: o autor

De posse do alinhamento resultante do BLASTn, entre as Ilhas de Patogenicidade e o banco de dados multifasta, realizou-se manualmente a análise de homologia aplicando-se os parâmetros descritos anteriormente. Assim, foi gerada uma tabela indicando quais PAI são homólogas entre si, onde a cepa 04-3911 teve mais ilhas compartilhadas com os outros genomas, em que cada PAI alinhou com pelo menos 20 cepas, além disso 12 PAIs não alinharam com nenhuma outra ilha (TABELA 4). O maior número de homologia para uma mesma ilha foi de 25, ou seja, cerca de 25 linhagens compartilharam uma mesma ilha de patogenicidade. Com base nisso criamos uma tabela para visualizar o conteúdo gênico desta ilha (TABELA 5). Estes

dados podem ser utilizados para um estudo proteômico, visando uma caracterização básica da expressão e localização dessas proteínas. É importante frisar que as PAIs detêm um grande número de genes de virulência agrupados em blocos que desempenham papéis importantes na patogenicidade do microrganismo (BARAÚNA et al., 2017).

Tabela 5 – Caracterização da Ilha de Patogenicidade presente em 25 linhagens

ID de proteína	Coordenadas PAIs	Gene	Produto
SQH01998.1	1242999..1246238	<i>lysX</i>	lisil-tRNA sintetase 2
SNV08640.1	1246201..1246875	<i>locustag</i>	proteína hipotética
SNV08649.1	1246835..1247500	<i>locustag</i>	proteína hipotética
SNV08655.1	1247510..1247695	<i>locustag</i>	proteína hipotética
SNV08664.1	1247727..1249034	<i>rkpK</i>	UDP-glucose 6-dehydrogenase
SNV08668.1	1249039..1249605	<i>dcd</i>	Desoxicitidina trifosfato desaminase
SQH03068.1	1250134..1251525	<i>gabD1_1</i>	succinato-semialdeído desidrogenase
SNV08680.1	1251645..1252262	<i>locustag</i>	proteína hipotética
SNV08688.1	1252267..1253994	<i>irtB</i>	proteína permease IrtB
SNV08698.1	1253987..1255687	<i>irtA</i>	proteína permease IrtA
SNV08706.1	1255677..1257149	<i>ykoD_1</i>	proteína de importação de tiamina YkoD
SNV08712.1	1257146..1257838	<i>locustag</i>	proteína hipotética
SNV08719.1	1257835..1258425	<i>locustag</i>	proteína hipotética

Fonte: o autor

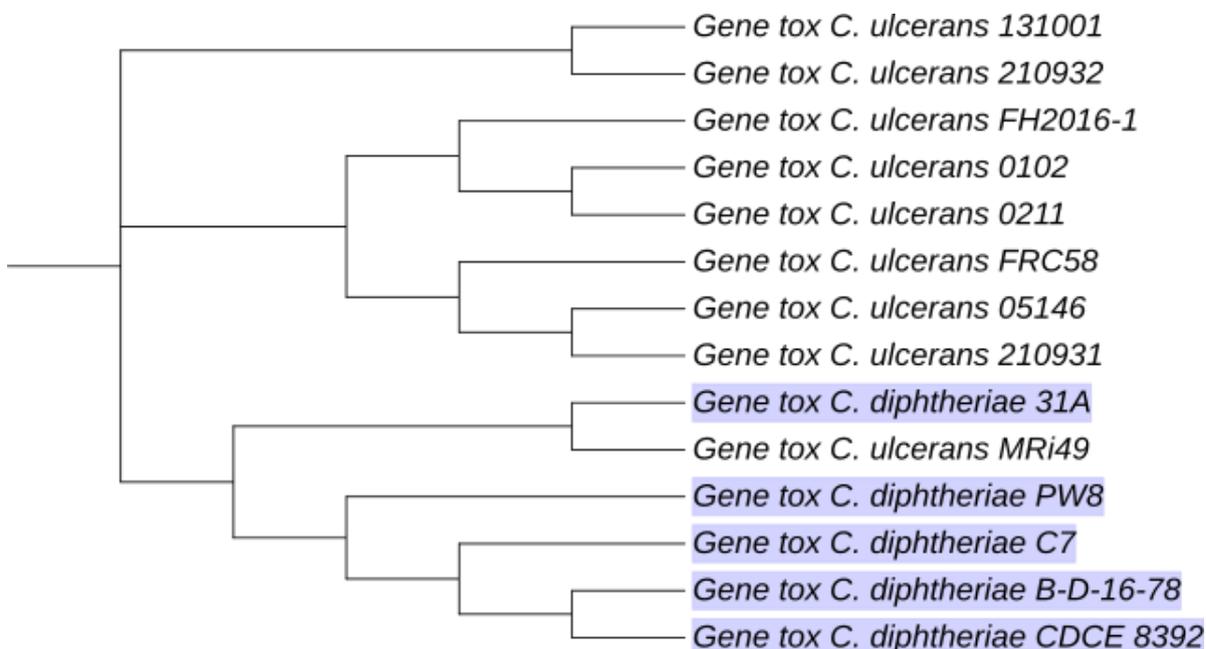
2.3.4 Análise filogenética dos genes: *tox* e *pld*

De vinte e sete, apenas nove linhagens tiveram a sequência do gene *tox* em seus genomas. De modo geral, nota-se que o gene presente nos genomas das linhagens de *C. diphtheriae* estiveram todos presentes em um mesmo ramo da árvore, ao contrário de *C. ulcerans*, em que o gene apresentou pequenas variações em sua sequência de linhagem para linhagem (FIGURA 4). Já foram comprovados em outros trabalhos que a sequência da toxina da difteria isolada de linhagens de *C. ulcerans* são menos conservadas em relação a *C. diphtheriae* (OTSUJI et al., 2019; SING et al., 2003).

Visto que a produção da vacina contra difteria utiliza em sua fórmula a toxina da difteria inativada derivada da cepa *C. diphtheriae* PW8, seria interessante estudar o efeito da vacina aliada a heterogeneidade do gene *tox* em linhagens de *C. ulcerans* (PARK; WILLIAMS, 1896; SANGAL; BURKOVSKI, 2020). Otsuji et al. (2019)

suspeitam que a variação da toxina diftérica pode diminuir a eficácia da vacina diftérica na prevenção e tratamento de doenças causadas por este patógeno.

Figura 4 – Árvore filogenética demonstrando a similaridade quanto a presença do gene *tox* entre as linhagens de *C. ulcerans* e *C. diphtheriae*

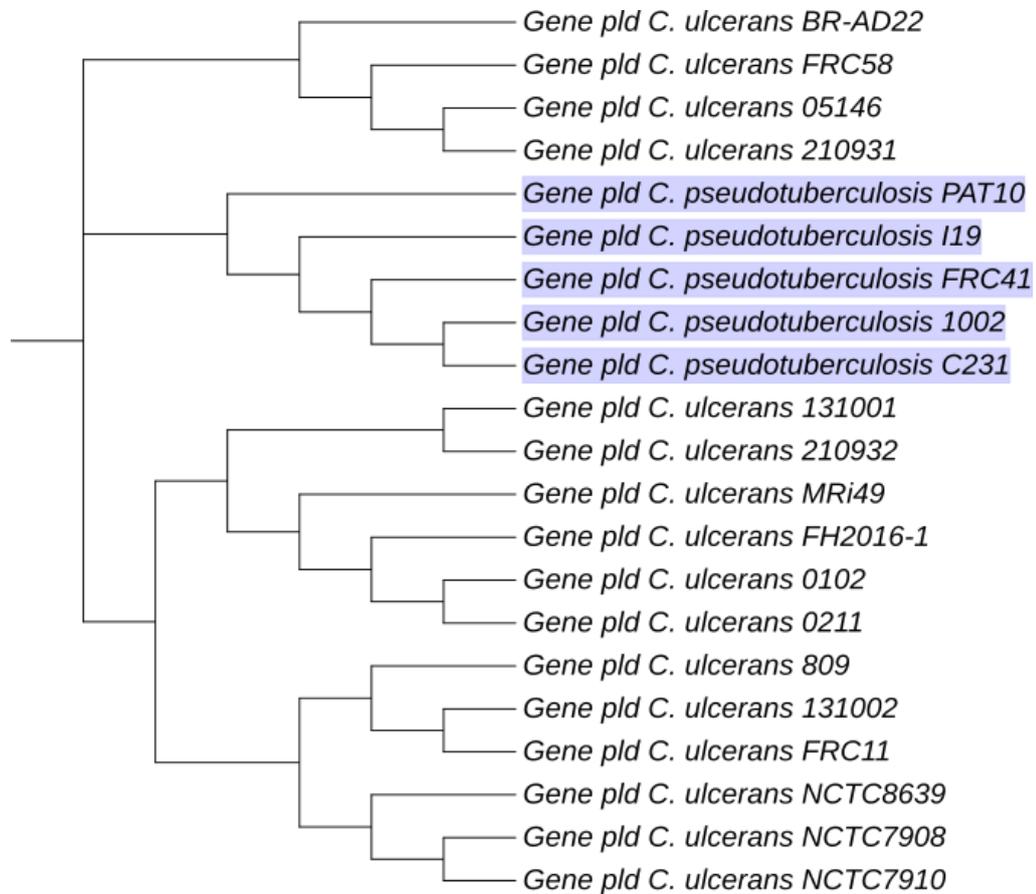


Fonte: o autor

Assim como o gene *tox* em *C. diphtheriae*, o gene *pld* em *C. pseudotuberculosis* se apresentou conservado nas cinco linhagens em análise, onde todos se encontram em um mesmo clado (FIGURA 5). Em contraponto, o gene *pld* presente nas 16 linhagens de *C. ulcerans* tiveram pequenas variações em suas sequências.

Conforme descrito anteriormente, o gene *pld* codifica um importante fator de virulência em *C. pseudotuberculosis* e *C. ulcerans*, pois está atrelado ao aumento da permeabilidade vascular e contribui para a persistência e disseminação da bactéria no hospedeiro, por isso, está intimamente ligado aos sintomas de linfadenite caseosa (FONTAINE; BAIRD, 2008; SOARES et al., 2013b). A variação da sequência genética do gene *pld* em *C. ulcerans* pode estar associada com a diminuição da sua patogenicidade, uma vez que há poucos relatos de infecções semelhantes ao da linfadenite caseosa em alguns animais e até mesmo em humanos. (MÖLLER et al., 2020; TERADA et al., 2015; URAKAWA et al., 2013).

Figura 5 – Árvore filogenética demonstrando a similaridade quanto a presença do gene *pld* entre as linhagens de *C. ulcerans* e *C. pseudotuberculosis*



Fonte: o autor

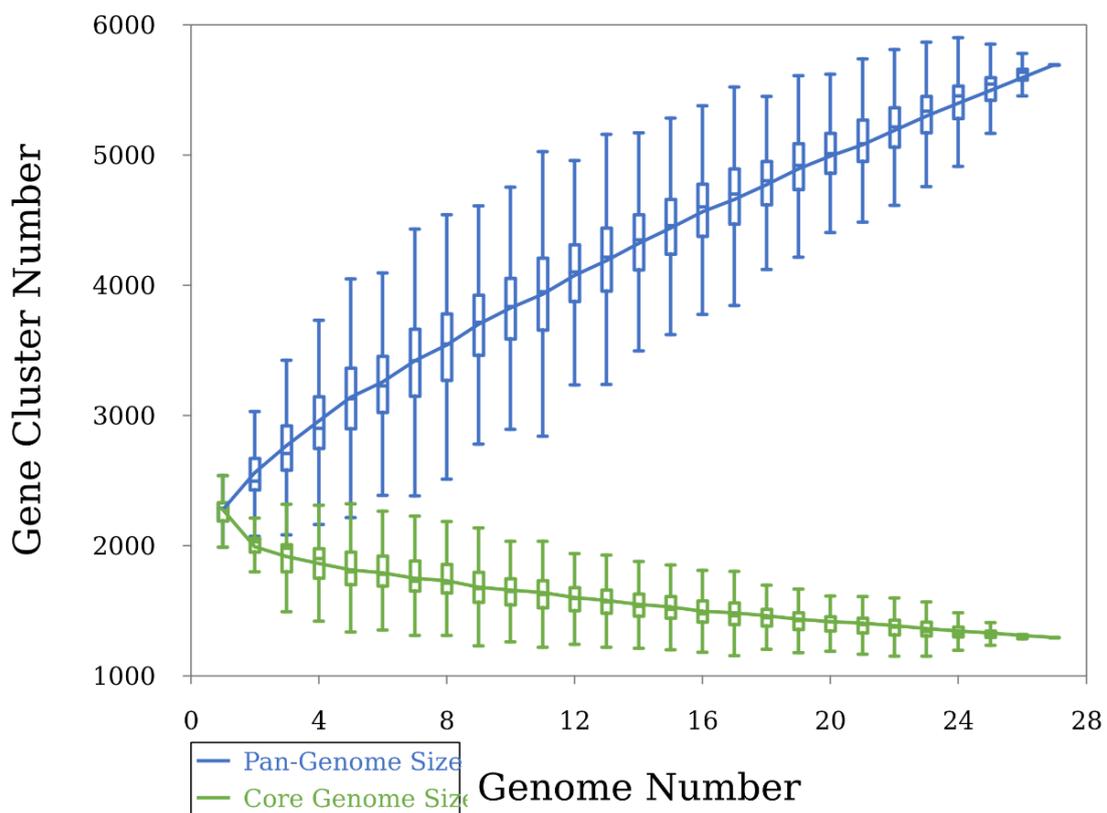
2.3.5 Predição do pangenoma, do genoma central e dos genes linhagem-específicos de *C. ulcerans*

O pangenoma da *C. ulcerans* é composto por 5.690 regiões codificadoras de proteínas, onde 1.790 são genes compartilhados por pelo menos duas cepas e, portanto, fazem parte do genoma acessório (FIGURA 6). Além disso, a quantidade de genes específicos foi de 2.205, esse resultado potencializa a alta variabilidade genética desta espécie. Dessa forma, o genoma central da *C. ulcerans* corresponde a 22.8% do pangenoma, porcentagem baixa em relação ao genoma central posteriormente calculado das espécies *C. pseudotuberculosis* biovar equi e *C. diphtheriae* que obtiveram, respectivamente, aproximadamente 42,5% e 34% (BARAÚNA et al., 2017; TROST et al., 2012a).

O valor do pangenoma encontrado nesse trabalho foi maior em relação ao encontrado por Subedi et al. (2018), que analisando o genoma de 19 linhagens de *C. ulcerans*, chegou ao resultado de que o pangenoma foi composto por 4120 genes, incluindo 1405 genes centrais e 2715 genes acessórios. Comparando esse resultado, nota-se que houve uma grande quantidade de genes adquiridos por essa bactéria, conforme houve adição de novas linhagens para essa análise.

Confirmando essa hipótese, calculamos o valor α da Lei de Heap e constatamos que o número de genes anotados nos genomas sequenciados de *C. ulcerans* cresce com $\alpha = 0.649$, e uma vez que $\alpha < 1$, o pangenoma é considerado aberto. Além disso, o valor de α indica que o pangenoma de *C. ulcerans* tem maior taxa de crescimento em relação à observada para *C. diphtheriae* ($\alpha = 0.69$), *C. pseudotuberculosis* biovar equi ($\alpha = 0.89$) e biovar ovis ($\alpha = 0.94$) (SOARES et al., 2013b; TROST et al., 2012b).

Figura 6 – Predição do pangenoma e do genoma do central de vinte e sete genomas de *C. ulcerans*



Fonte: o autor

2.3.6 Predição de alvos putativos de vacina e de drogas

A análise gerada pelo PGAweb, das regiões codificadoras de proteínas compartilhadas pelas 27 linhagens, serviu como base para a identificação dos genes essenciais para a sobrevivência da *C. ulcerans*, onde as sequências de proteínas foram submetidas ao DGE que, por sua vez, revelou que somente 6 genes codificaram proteínas que não tiveram homologia nenhuma com o proteoma humano, e apenas 2 proteínas estavam localizados na membrana plasmática e possuíam propriedades antigênicas (TABELA 6).

Proteínas exportadas e ancoradas na membrana são consideradas potenciais alvos vacinais, uma vez que essas proteínas podem interagir diretamente com as moléculas do hospedeiro para adesão celular, multiplicação, invasão, evasão da resposta imune, geração de danos ao hospedeiro e sobreviver às defesas da célula hospedeira (KRACHLER; ORTH, 2013; SANTANA-JORGE et al., 2016)

Em *Corynebacterium urealyticum*, a aplicação do método de vacinologia reversa revelou 19 proteínas antigênicas (GUIMARÃES et al., 2015). Através de análises *in silico* realizada em *C. pseudotuberculosis* FRC41, demonstrou que as proteínas codificadas pelos genes SpaC, PknG e NanH apresentam bom potencial como alvos para o desenvolvimento de vacinas (SANTANA-JORGE et al., 2016). Através da abordagem RNA-seq e da vacinologia reversa, Araújo e colaboradores, identificaram seis - copC, ykuE, ndh, mtrB, ftsI, senX3 - potenciais alvos de vacinais usando o genoma central de um conjunto de genes compartilhados entre 65 linhagens de *C. pseudotuberculosis* (ARAÚJO et al., 2019). Um artigo baseado na abordagem da proteômica subtrativa em *C. diphtheriae*, constatou que 3 proteínas (MtrA, MprA e RegX3) são prováveis alvos de drogas (KHALID et al., 2018). Nenhum dos alvos vacinais e de drogas encontrado nesses artigos foi similar com os seis alvos descritos nesse trabalho, como demonstrado na tabela 6.

Tabela 6 – Alvos vacinais e de drogas subtraído do genoma central de 27 genomas de *C. ulcerans*

ID da proteína	Processo biológico	Gene	Pontuação de Antigenicidade	Localização	Comprimento da proteína	Alvo
BAM27215.1	Processo biossintético de alfa-glucana	<i>glgE</i>	0.4361	Citoplasma	693	Drogas
QGZ25337.1	Regulamentação da fidelidade transducional	<i>gatC</i>	0.1691	Citoplasma	99	Drogas
SNV00767.1	Atividade de transporte transmembrana	<i>emrB</i>	0.6124	Membrana citoplasmática	498	Vacinal
SNV00953.1	A proteína reguladora da transcrição de exportação de lisina LysR	<i>lysG</i>	0.4121	Citoplasma	304	Drogas
SNV00991.1"	Processo biossintético da isoleucina e da valina	<i>ilvD</i>	0.4251	Citoplasma	633	Drogas
SNV01009.1	Atividade de transporte transmembrana	<i>mscS_1</i>	0.6569	Membrana citoplasmática	582	Vacinal

Fonte: o autor

O gene *glgE* codifica a enzima glucanase que está envolvida na biossíntese de glicogênio, que é essencial para o crescimento da bactéria (SYSON et al., 2021). Koliwer-Brandl e colaboradores, constataram que o glicogênio derivado de *glgE* participa da formação de uma capsula em *Mycobacterium tuberculosis*, e conseqüentemente, esse fator contribui no aumento da virulência dessa bactéria (KOLIWER-BRANDL et al., 2016). Outros dois estudos apontaram que esse gene é visto como um potencial alvos de drogas, pois quando ocorre a substituição do aminoácido H349Y da proteína GlgE, α -maltose-1-fosfato acumula-se a níveis tóxicos devido a uma perda parcial da atividade da polimerase, particularmente em uma temperatura elevada, ocorre uma perda de estabilidade da proteína (KALSCHEUER; JACOBS, 2010; SYSON et al., 2021).

A proteína expressa pelo gene *gatC* atua na formação do RNA transportador de glutaminil. A reação ocorre na presença de glutamina e ATP através de um fosfo-Asp-tRNA (Asn) ou fosfo-Glu-tRNA (Gln) ativado (DIMMER et al., 2012; FRIEDERICH et al., 2018). O *gatC* é caracterizado por um loop estendido não estruturado flanqueado por duas hélices que envolvem a região de interface realizando interações com *gatA* e *gatB* (NAKAMURA et al., 2006).

O *lysG* faz parte da família de genes que atuam na regulação transcricional do tipo *lysR* no reino procariótico, tais genes podem estar envolvidos em funções como a virulência, motilidade, metabolismo, entre outros (MADDOCKS; OYSTON, 2008).

Um artigo constatou que o *lysG* presente na *Mycobacterium tuberculosis* ativa o gene *lysE* na presença de lisina e histidina e desregula sua própria transcrição, e também exerce um efeito indireto e positivo nas quantidades transcritas dos genes *ppsBCD* (SCHNEEFELD et al., 2017).

O *ilvD* é um gene que codifica a enzima di-hidroxi-ácido desidratase, e faz parte do operon *ilvBHC-leuABCD* (*ilv-leu*), que inclui proteínas que catalisam a formação dos precursores da valina e da isoleucina (MÄDER et al., 2004). A proteína codificada pelo *ilvD* está envolvida na etapa 3 da sub-via que sintetiza L-isoleucina a partir de 2-oxobutanoato, e na etapa 3 da sub-via que sintetiza L-valina a partir do dihidroxi-3-metilbutanoato (MAGRANE; CONSORTIUM, 2011). O mesmo processo ocorre com *C. glutamicum* (ZHANG et al., 2018).

Ainda não existem trabalhos publicados descrevendo que as proteínas codificadas pelos genes *gatC*, *lysG* e *ilvD*, como potenciais alvos de drogas em organismos homólogos a *C. ulcerans*. Por isso, torna-se necessário entender as interações biomoleculares para o desenho e descoberta de fármacos, por meio da bioinformática. E, posteriormente, estudos experimentais para comprovar sua eficácia.

As proteínas codificadas pelos genes *emrB* e *mscS_1*, foram as únicas preditas no Vaxigen, com respectivos valores de antigenicidade de 0.6124 e 0.6569, como capazes de gerar a formação de anticorpos, e ambas estão situadas na membrana plasmática. O *mscS_1* faz parte do canal mecanossensível, que participa do sistema de efluxo de solutos para o exterior da célula, evitando a lise celular (SUKHAREV, 2002). O gene *emrB* traduz uma proteína de membrana integral, altamente hidrofóbica que está envolvida no bombeamento de drogas através da membrana citoplasmática, por isso ela está associada à resistência a múltiplas drogas, como a ampicilina em *Staphylococcus aureus* (ZHANG et al., 2016), o ácido nalidíxico e a tulatromicina em *Escherichia coli* (LOMOVSKAYA; LEWIS, 1992; LOMOVSKAYA; LEWIS; MATIN, 1995), a tetraciclina em *Burkholderia thailandensis* (SABRIN et al., 2019) e a colistina em *Acinetobacter baumannii* (LIN; LIN; LAN, 2017).

Cientificamente, já foi provado que a *C. ulcerans* apresenta resistência a alguns tipos de antibióticos como a eritromicina, clindamicina, oxacilina e a fosfomicina (MOORE et al., 2015; TEJEDOR et al., 2000; TIWARI et al., 2008; VANDENTORREN et al., 2014). Entretanto, nenhum deles demonstram qual gene está sendo expresso

em resposta a substância antibactericida. Seria interessante avaliar os níveis de expressão do gene *emrB* em resposta a esses fármacos.

2.4 Conclusões

Por meio dos resultados apresentados, a análise das sequências completas dos 29 genomas coletados no NCBI descritos como *C. ulcerans* forneceu conhecimentos detalhados sobre a arquitetura do genoma e o repertório de genes candidatos que contribuem para a virulência dessas cepas. Além disto, devido a porcentagem de ANI abaixo de 95%, as linhagens PO100/5 e w25 foram retiradas das demais análises uma vez que estas linhagens pertencem a outra espécie do gênero *Corynebacterium*. Portanto, após uma busca na literatura, foram encontrados dois artigos recém publicados que apontam que estas duas linhagens foram erroneamente descritas como *C. ulcerans*, corroborando com nossos resultados (MÖLLER et al., 2020; VIANA et al., 2021).

Apenas nove cepas apresentaram o gene *tox* em seus genomas, sendo capazes de sintetizar a toxina que causa difteria. Embora, a patogenicidade da *C. ulcerans* não dependa necessariamente da produção da TD, alguns estudos de casos mostraram a capacidade da bactéria de causar infecções graves sem apresentar a toxina (DESSAU et al., 1995; GUBLER et al., 1990; JENSSEN et al., 2017; KIMURA et al., 2011; SING et al., 2003).

O gene *pld*, que codifica a exotoxina que está envolvida na linfadenite caseosa, esteve presente em 16 linhagens de *C. ulcerans*, onde a sequência do gene variou de linhagem para linhagem. Em decorrência disto, seria importante estudar a expressão dessa proteína em linhagens de *C. ulcerans*, com intuito de verificar se essa variação pode estar afetando em sua função, visto que há poucos estudos de casos relacionando essa bactéria como agente causal da linfadenite caseosa em animais.

Apesar do mapa de calor evidenciar uma boa porcentagem de similaridade entre as 27 linhagens, a árvore filogenômica demonstrou que alguns genomas isolados de um mesmo hospedeiro e no mesmo país estiveram presente em ramos diferente da árvore. Supostamente, essa variação indica que as linhagens de *C. ulcerans* podem ser pouco clonais.

Corroborando com a suposição anterior, os valores do pangenoma de *C. ulcerans* foi alto em relação aos trabalhos citados anteriormente, e isso está associado a grande quantidade de genes específicos detectados. Consequentemente, segundo as regras estabelecidas pela lei de Heap, o pangenoma é aberto. Além de que, foi estimado que o pangenoma da espécie estudada cresce com taxas de α superiores aquelas encontradas para *C. diphtheriae* e *C. pseudotuberculosis*.

Encontramos duas proteínas candidatas como alvos vacinais e quatro como alvos de drogas. A produção dessas proteínas em grandes quantidades representa um passo importante para estudos futuros sobre estrutura 3-D, patogenicidade, virulência e desenvolvimento de vacinas e drogas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ÅGREN, J. et al. Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. **PLoS ONE**, v. 7, n. 6, 2012.
- ALIKHAN, N. F. et al. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. **BMC Genomics**, v. 12, 2011.
- ARAÚJO, C. L. et al. Prediction of new vaccine targets in the core genome of *Corynebacterium pseudotuberculosis* through omics approaches and reverse vaccinology. **Gene**, v. 702, p. 36–45, 20 jun. 2019.
- ARIAS, C. A.; MURRAY, B. E. Antibiotic-resistant bugs in the 21st century - A clinical super-challenge. **New England Journal of Medicine**, v. 360, n. 5, 2009.
- AYYILDIZ, D.; PIAZZA, S. Introduction to Bioinformatics. In: **Methods in Molecular Biology**. [s.l.] Humana Press Inc., 2019. v. 1986p. 1–15.
- BAKHEET, T. M.; DOIG, A. J. Properties and identification of human protein drug targets. **Bioinformatics**, v. 25, n. 4, 2009.
- BARAÚNA, R. A. et al. Assessing the genotypic differences between strains of *Corynebacterium pseudotuberculosis* biovar equi through comparative genomics. **PLoS ONE**, v. 12, n. 1, 2017.
- BARBA, M. et al. Identifying reaction modules in metabolic pathways: Bioinformatic deduction and experimental validation of a new putative route in purine catabolism. **BMC Systems Biology**, v. 7, 2013.
- BARH, D. et al. In silico subtractive genomics for target identification in human bacterial pathogens. **Drug Development Research**, v. 72, n. 2, 2011a.
- 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, 2011b.
- BELBAHRI, L. et al. Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. **Gene**, v. 408, n. 1–2, p. 1–8, 31 jan. 2008.
- BENEVIDES, L. DE J. et al. Genome sequence of *Corynebacterium ulcerans* strain FRC11. **Genome Announcements**, v. 3, n. 2, 2016.
- BERGER, A. et al. Tox-positive *Corynebacterium ulcerans* in hedgehogs, Germany*. **Emerging Microbes and Infections**, v. 8, n. 1, 2019.
- BOSTOCK, A. D. et al. *Corynebacterium ulcerans* infection associated with untreated milk. **Journal of Infection**, v. 9, n. 3, 1984.
- CAMACHO, C. et al. BLAST+: Architecture and applications. **BMC Bioinformatics**, v. 10, 15 dez. 2009.
- CARFORA, V. et al. Non-toxigenic *Corynebacterium ulcerans* sequence types 325 and 339 isolated from two dogs with ulcerative lesions in Italy. **Journal of Veterinary**

Diagnostic Investigation, v. 30, n. 3, p. 447–450, 2018.

CARVER, T. et al. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. **Bioinformatics**, v. 28, n. 4, 2012.

CHEN, X. et al. PGAweb: A web server for bacterial pan-genome analysis. **Frontiers in Microbiology**, v. 9, n. JUL, 2018.

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

CORTI, M. A. M. et al. Rare human skin infection with *Corynebacterium ulcerans*: Transmission by a domestic cat. **Infection**, v. 40, n. 5, 2012.

DE SANTIS, A. et al. Non-toxicogenic *Corynebacterium diphtheriae* infective endocarditis with embolic events: a case report. **BMC Infectious Diseases**, v. 20, n. 1, 2020.

DE SOUZA DE OLIVEIRA DIAS, A. A. et al. ***Corynebacterium ulcerans* diphtheria: An emerging zoonosis in Brazil and worldwide** *Revista de Saude Publica*, 2011.

DE ZOYSA, A. et al. Characterization of toxigenic *Corynebacterium ulcerans* strains isolated from humans and domestic cats in the United Kingdom. **Journal of Clinical Microbiology**, v. 43, n. 9, 2005.

DESSAU, R. B. et al. Pulmonary nodules due to *Corynebacterium ulcerans*. **European Respiratory Journal**, v. 8, n. 4, 1995.

DETEMMERMAN, L. et al. Toxigenic *Corynebacterium ulcerans* in human and non-toxicogenic *Corynebacterium diphtheriae* in cat. **New Microbes and New Infections**, v. 1, n. 1, 2013.

DIMMER, E. C. et al. The UniProt-GO Annotation database in 2011. **Nucleic Acids Research**, v. 40, n. D1, 2012.

DOBRINDT, U. et al. Genomic islands in pathogenic and environmental microorganisms. **Nature Reviews Microbiology**, v. 2, n. 5, p. 414–424, 2004.

DONKOR, E. S. **Sequencing of bacterial genomes: Principles and insights into pathogenesis and development of antibiotics** *Genes*, 2013.

DORELLA, F. A. et al. ***Corynebacterium pseudotuberculosis*: Microbiology, biochemical properties, pathogenesis and molecular studies of virulence** *Veterinary Research*, 2006.

DOYTCHINOVA, I. A.; FLOWER, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. **BMC Bioinformatics**, v. 8, 2007.

DU PLESSIS, M. et al. Molecular characterization of *corynebacterium diphtheriae* outbreak Isolates, South Africa, March–June 2015. **Emerging Infectious Diseases**, v. 23, n. 8, 2017.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, 2004.

EISENBERG, T. et al. Nontoxicogenic tox-bearing *Corynebacterium ulcerans* Infection

among Game Animals, Germany. **Emerging Infectious Diseases**, v. 20, n. 3, p. 448–452, mar. 2014.

FARAG, I. F.; YOUSSEF, N. H.; ELSHAHED, M. S. Global distribution patterns and pangenomic diversity of the candidate phylum “Latescibacteria” (WS3). **Applied and Environmental Microbiology**, v. 83, n. 10, 2017.

FONTAINE, M. C.; BAIRD, G. J. Caseous lymphadenitis. **Small Ruminant Research**, v. 76, n. 1–2, p. 42–48, abr. 2008.

FOSTER, G. et al. **Corynebacterium ulcerans in free-ranging otters [1]****Veterinary Record**, 2002.

FRIEDERICH, M. W. et al. Pathogenic variants in glutamyl-tRNA^{Gln} amidotransferase subunits cause a lethal mitochondrial cardiomyopathy disorder. **Nature Communications**, v. 9, n. 1, 2018.

FUNKE, G. et al. **Clinical microbiology of Coryneform bacteria**. [s.l: s.n.]. v. 10

FUURSTED, K. et al. Non-toxigenic tox gene-bearing *Corynebacterium ulcerans* in a traumatic ulcer from a human case and his asymptomatic dog. **Microbes and Infection**, v. 17, n. 10, 2015.

GILBERT, R., STEWART, F. C. *Corynebacterium ulcerans*; a pathogenic microorganism resembling *C. diphtheriae*. **J. Lab. Clin. Med. Volume**, v. 12, p. 756–761, 1927.

GORIS, J. et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. **International Journal of Systematic and Evolutionary Microbiology**, v. 57, n. 1, 2007.

GUBLER, J. G. et al. [Classical pseudomembranous diphtheria caused by *Corynebacterium ulcerans*]. **Schweiz Med Wochenschr**, v. 120, n. 48, 1990.

GUIMARÃES, L. C. et al. Genome informatics and vaccine targets in *Corynebacterium urealyticum* using two whole genomes, comparative genomics, and reverse vaccinology. **BMC Genomics**, v. 16, n. 5, 2015.

GUIMARÃES, L. C. et al. Draft genome sequence of toxigenic *Corynebacterium ulcerans* strain 04-7514, isolated from a dog in France. **Genome Announcements**, v. 4, n. 2, 2016a.

GUIMARÃES, L. C. et al. Draft genome sequence of *Corynebacterium ulcerans* strain 04-3911, isolated from humans. **Genome Announcements**, v. 4, n. 2, 2016b.

HALLSTROM, K. N.; MCCORMICK, B. A. Pathogenicity Islands: Origins, Structure, and Roles in Bacterial Pathogenesis. In: **Molecular Medical Microbiology: Second Edition**. [s.l: s.n.]. v. 1–3.

HART, R. J. C. *Corynebacterium ulcerans* in humans and cattle in North Devon. **Journal of Hygiene**, v. 92, n. 2, 1984.

HASSAN, A. et al. Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. **BMC Genomics**, v. 17, n. 1, 2016.

HATANAKA, A. et al. **Corynebacterium ulcerans diphtheria in Japan [3]****Emerging**

Infectious Diseases Centers for Disease Control and Prevention (CDC), , 1 jun. 2003.

HIGGS, T. M. et al. Corynebacterium ulcerans infections in a dairy herd. **The Veterinary record**, v. 81, n. 2, 1967.

HOLMAN, A. G. et al. Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of Brugia malayi. **BMC Microbiology**, v. 9, 2009.

HOMMEZ, J. et al. Identification of nonlipophilic corynebacteria isolated from dairy cows with mastitis. **Journal of Clinical Microbiology**, v. 37, n. 4, 1999.

JAISWAL, A. K. et al. An in silico identification of common putative vaccine candidates against treponema pallidum: A reverse vaccinology and subtractive genomics based approach. **International Journal of Molecular Sciences**, v. 18, n. 2, 2017.

JANDA, W. M. Corynebacterium species and the coryneform bacteria part I: New and emerging species in the genus Corynebacterium. **Clinical Microbiology Newsletter**, v. 20, n. 6, 1998.

JEAN BELTRAN, P. M. et al. Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. **Molecular Systems Biology**, v. 13, n. 3, p. 922, mar. 2017.

JENSSEN, C. et al. Skin and Soft Tissue Infections Due to Corynebacterium ulcerans - Case Reports. **Deutsche Medizinische Wochenschrift**, v. 142, n. 21, p. 1609–1612, 2017.

KALINOWSKI, J. et al. The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. **Journal of Biotechnology**, v. 104, n. 1–3, 2003.

KALSCHUEER, R.; JACOBS, W. R. **The significance of GlgE as a new target for tuberculosis** *Drug News and Perspectives*, 2010.

KATSUKAWA, C. et al. Toxigenic Corynebacterium ulcerans isolated from a wild bird (ural owl) and its feed (shrew-moles): Comparison of molecular types with human isolates. **BMC Research Notes**, v. 9, n. 1, 2016.

KAUFMANN, D.; OTT, P.; ZBINDEN, R. Laryngopharyngitis by Corynebacterium ulcerans. **Infection**, v. 30, n. 3, 2002.

KHALID, Z. et al. Subtractive proteomics revealed plausible drug candidates in the proteome of multi-drug resistant Corynebacterium diphtheriae. **Meta Gene**, v. 17, p. 34–42, 1 set. 2018.

KIMURA, Y. et al. Acute peritonitis due to Corynebacterium ulcerans in a patient receiving continuous ambulatory peritoneal dialysis: A case report and literature review. **Clinical and Experimental Nephrology**, v. 15, n. 1, 2011.

KOLIWER-BRANDL, H. et al. Metabolic Network for the Biosynthesis of Intra- and Extracellular α -Glucans Required for Virulence of Mycobacterium tuberculosis. **PLoS Pathogens**, v. 12, n. 8, 2016.

- KONRAD, R.; HÖRMANSDORFER, S.; SING, A. Possible human-to-human transmission of toxigenic *Corynebacterium ulcerans*. **Clinical Microbiology and Infection**, v. 21, n. 8, 2015a.
- KONRAD, R.; HÖRMANSDORFER, S.; SING, A. Possible human-to-human transmission of toxigenic *Corynebacterium ulcerans*. **Clinical Microbiology and Infection**, v. 21, n. 8, p. 768–771, 1 ago. 2015b.
- KRACHLER, A. M.; ORTH, K. Targeting the bacteria–host interface. **Virulence**, v. 4, n. 4, 2013.
- KUMAR, G. S. et al. Definition of potential targets in *Mycoplasma pneumoniae* through subtractive genome analysis. **Journal of Antivirals and Antiretrovirals**, v. 2, n. 2, 2010.
- KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, 2018.
- LARTIGUE, M. F. et al. *Corynebacterium ulcerans* in an immunocompromised patient with diphtheria and her dog. **Journal of Clinical Microbiology**, v. 43, n. 2, 2005.
- LAWLEY, B. et al. Differentiation of *Bifidobacterium longum* subspecies *longum* and *infantis* by quantitative PCR using functional gene targets. **PeerJ**, v. 2017, n. 5, 2017.
- LEE, K.; KIM, D. W.; CHA, C. J. **Overview of bioinformatic methods for analysis of antibiotic resistome from genome and metagenome data** *Journal of Microbiology* The Korean Society for Microbiology / The Korean Society of Virology, , 1 mar. 2021.
- LETUNIC, I.; BORK, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. **Nucleic acids research**, v. 44, n. W1, 2016.
- LIN, M. F.; LIN, Y. Y.; LAN, C. Y. Contribution of EmrAB efflux pumps to colistin resistance in *Acinetobacter baumannii*. **Journal of Microbiology**, v. 55, n. 2, p. 130–136, 1 fev. 2017.
- LOMOVSKAYA, O.; LEWIS, K. *emr*, an *Escherichia coli* locus for multidrug resistance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 19, p. 8938–8942, 1 out. 1992.
- LOMOVSKAYA, O.; LEWIS, K.; MATIN, A. EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump *emrAB*. **Journal of Bacteriology**, v. 177, n. 9, p. 2328–2334, 1995.
- MADDOCKS, S. E.; OYSTON, P. C. F. **Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins** *Microbiology*, 2008.
- MÄDER, U. et al. Transcriptional Organization and Posttranscriptional Regulation of the *Bacillus subtilis* Branched-Chain Amino Acid Biosynthesis Genes. **Journal of Bacteriology**, v. 186, n. 8, p. 2240–2252, abr. 2004.
- MAGRANE, M.; CONSORTIUM, U. P. UniProt Knowledgebase: A hub of integrated protein data. **Database**, v. 2011, 2011.
- MARINI, R. P. et al. ***Corynebacterium ulcerans* in ferrets** *Emerging Infectious*

Diseases, 2014.

MATTOS-GUARALDI, A. L. et al. **Diphtheria Remains a Threat to Health in the Developing World - An Overview** *Memorias do Instituto Oswaldo Cruz*, 2003.

MATTOS-GUARALDI, A. L. et al. First detection of *Corynebacterium ulcerans* producing a diphtheria-like toxin in a case of human with pulmonary infection in the Rio de Janeiro metropolitan area, Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 103, n. 4, 2008.

MEINEL, D. M. et al. Next generation sequencing analysis of nine *Corynebacterium ulcerans* isolates reveals zoonotic transmission and a novel putative diphtheria toxin-encoding pathogenicity island. **Genome Medicine**, v. 6, n. 11, p. 1–13, 2014.

MÖLLER, J. et al. Cellular and extracellular proteome of the animal pathogen *Corynebacterium silvaticum*, a close relative of zoonotic *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*. **Proteomes**, v. 8, n. 3, 2020.

MOORE, L. S. P. et al. *Corynebacterium ulcerans* cutaneous diphtheria. **The Lancet Infectious Diseases**, v. 15, n. 9, p. 1100–1107, 2015.

MUHAMMAD AZAM, F. et al. Antigenic outer membrane proteins prediction of *Pasteurella multocida* serotype B:2. **Asia Pacific Journal of Molecular Biology and Biotechnology**, p. 102–116, 28 nov. 2020.

NAKAMURA, A. et al. Ammonia channel couples glutaminase with transamidase reactions in GatCAB. **Science**, v. 312, n. 5782, p. 1954–1958, 30 jun. 2006.

NAVARRO-GARCIA, F. et al. **Type VI secretion system in pathogenic escherichia coli: Structure, role in virulence, and acquisition** *Frontiers in Microbiology* Frontiers Media S.A., , 2019.

NAZ, K. et al. PanRV: Pangenome-reverse vaccinology approach for identifications of potential vaccine candidates in microbial pangenome. **BMC Bioinformatics**, v. 20, n. 1, 2019.

OTSUJI, K. et al. Mutation and Diversity of Diphtheria Toxin in *Corynebacterium ulcerans*. **Emerging Infectious Diseases**, v. 25, n. 11, p. 2122–2123, 2019.

PARK, W. H.; WILLIAMS, A. W. The production of diphtheria toxin. **Journal of Experimental Medicine**, v. 1, n. 1, 1896.

PASCUAL, C. et al. Phylogenetic analysis of the genus *Corynebacterium* based on 16s rRNA gene sequences. **International Journal of Systematic Bacteriology**, v. 45, n. 4, 1995.

RAU, J. et al. *Corynebacterium ulcerans*-infektion bei einem Reh (*Capreolus capreolus*). **Berliner und Munchener Tierärztliche Wochenschrift**, v. 125, n. 3–4, 2011.

Respiratory Diphtheria Caused by *Corynebacterium ulcerans* - Terre Haute, Indiana, 1996. **Morbidity and Mortality Weekly Report**, v. 46, n. 15, 1997.

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, 1995.

- RODRIGUES, T. C. V. et al. Reverse vaccinology and subtractive genomics reveal new therapeutic targets against *Mycoplasma pneumoniae*: A causative agent of pneumonia. **Royal Society Open Science**, v. 6, n. 7, 2019.
- SABRIN, A. et al. An EmrB multidrug efflux pump in *Burkholderia thailandensis* with unexpected roles in antibiotic resistance. **Journal of Biological Chemistry**, v. 294, n. 6, p. 1891–1903, 8 fev. 2019.
- SAEKI, J. et al. The detection of toxigenic *Corynebacterium ulcerans* from cats with nasal inflammation in Japan. **Epidemiology and Infection**, v. 143, n. 12, 2015.
- SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, v. 4, n. 4, 1987.
- SANGAL, V. et al. **Diphtheria-like disease caused by toxigenic corynebacterium ulcerans strain** **Emerging Infectious Diseases**, 2014.
- SANGAL, V.; BURKOVSKI, A. Insights into old and new foes: Pan-genomics of *Corynebacterium diphtheriae* and *Corynebacterium ulcerans*. In: **Pan-genomics: Applications, Challenges, and Future Prospects**. [s.l: s.n.].
- SANGAL, V.; HOSKISSON, P. A. Corynephages: Infections of the infectors. In: **Corynebacterium diphtheriae and Related Toxigenic Species: Genomics, Pathogenicity and Applications**. [s.l: s.n.]. v. 9789400776241.
- SANTANA-JORGE, K. T. O. et al. Putative virulence factors of *Corynebacterium pseudotuberculosis* FRC41: Vaccine potential and protein expression. **Microbial Cell Factories**, v. 15, n. 1, 2016.
- SCHNEEFELD, M. et al. The transcriptional regulator LysG (Rv1985c) of *Mycobacterium tuberculosis* activates lysE (Rv1986) in a lysine-dependent manner. **PLoS ONE**, v. 12, n. 10, 1 out. 2017.
- SCHUHEGGER, R. et al. **Pigs as source for toxigenic Corynebacterium ulcerans** **Emerging Infectious Diseases**, 2009.
- SEEMANN, T. Prokka: Rapid prokaryotic genome annotation. **Bioinformatics**, v. 30, n. 14, 2014.
- SEKIZUKA, T. et al. *Corynebacterium ulcerans* 0102 carries the gene encoding diphtheria toxin on a prophage different from the *C. diphtheriae* NCTC 13129 prophage. **BMC Microbiology**, v. 12, 2012.
- SELIM, S. A. et al. Immunological characterization of diphtheria toxin recovered from *Corynebacterium pseudotuberculosis*. **Saudi Journal of Biological Sciences**, v. 23, n. 2, 2016.
- SILVA, A. DO S. DE S. et al. Draft genome sequence of *Corynebacterium ulcerans* FRC58, isolated from the bronchitic aspiration of a patient in France. **Genome Announcements**, v. 2, n. 1, 2014.
- SIMPSON-LOURÉDO, L. et al. Detection and virulence potential of a phospholipase D-negative *Corynebacterium ulcerans* from a concurrent diphtheria and infectious mononucleosis case. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 112, n. 7, 2019.

- SING, A. et al. Detection of differences in the nucleotide and amino acid sequences of diphtheria toxin from *Corynebacterium diphtheriae* and *Corynebacterium ulcerans* causing extrapharyngeal infections. **Journal of Clinical Microbiology**, v. 41, n. 10, 2003.
- SOARES, S. C. et al. Genome sequence of *Corynebacterium pseudotuberculosis* biovar equi strain 258 and prediction of antigenic targets to improve biotechnological vaccine production. **Journal of Biotechnology**, v. 167, n. 2, 2013a.
- SOARES, S. C. et al. The Pan-Genome of the Animal Pathogen *Corynebacterium pseudotuberculosis* Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. **PLoS ONE**, v. 8, n. 1, p. e53818, jan. 2013b.
- SOARES, S. C. et al. GIPSy: Genomic island prediction software. **Journal of Biotechnology**, v. 232, p. 2–11, 2016.
- STING, R. et al. Toxigenic *Corynebacterium ulcerans* isolated from a free-roaming red fox (*Vulpes vulpes*). **Berliner und Münchener tierärztliche Wochenschrift**, v. 128, n. 5–6, 2015.
- SUBEDI, R. et al. Genomic analyses reveal two distinct lineages of *Corynebacterium ulcerans* strains. **New Microbes and New Infections**, v. 25, p. 7–13, 2018.
- SUKHAREV, S. Purification of the small mechanosensitive channel of *Escherichia coli* (MscS): The subunit structure, conduction, and gating characteristics in liposomes. **Biophysical Journal**, v. 83, n. 1, 2002.
- SUN, Z. et al. Determining the Genetic Characteristics of Resistance and Virulence of the “Epidermidis Cluster Group” Through Pan-Genome Analysis. **Frontiers in Cellular and Infection Microbiology**, v. 10, 2020.
- SUVAJDZIC, L. et al. A proposal of a diagnostic protocol for isolation of *Corynebacterium ulcerans* from cow’s milk. **Acta Scientiae Veterinariae**, v. 40, n. 2, 2012.
- SYKES, J. E. et al. *Corynebacterium ulcerans* bronchopneumonia in a dog. **Journal of Veterinary Internal Medicine**, v. 24, n. 4, 2010.
- SYME, R. A. et al. Pan-parastagonospora comparative genome analysis-effector prediction and genome evolution. **Genome Biology and Evolution**, v. 10, n. 9, p. 2443–2457, 2018.
- SYSON, K. et al. A temperature-sensitive *Mycobacterium smegmatis* glgE mutation leads to a loss of GlgE enzyme activity and thermostability and the accumulation of α -maltose-1-phosphate. **Biochimica et Biophysica Acta - General Subjects**, v. 1865, n. 2, 1 fev. 2021.
- TAUCH, A.; SANDBOTE, J. The family *Corynebacteriaceae*. In: **The Prokaryotes: Actinobacteria**. [s.l: s.n.].
- TEJEDOR, M. T. et al. Caseous lymphadenitis caused by *Corynebacterium ulcerans* in the dromedary camel. **Canadian Veterinary Journal**, v. 41, n. 2, 2000.
- TERADA, T. et al. A paediatric case of lymphadenitis by toxigenic *Corynebacterium ulcerans*. **JMM Case Reports**, v. 2, n. 5, 2015.

TIWARI, T. S. P. et al. Investigations of 2 Cases of Diphtheria-Like Illness Due to Toxigenic *Corynebacterium ulcerans*. **Clinical Infectious Diseases**, v. 46, n. 3, 2008.

TROST, E. et al. Comparative analysis of two complete *Corynebacterium ulcerans* genomes and detection of candidate virulence factors. **BMC Genomics**, v. 12, 2011.

TROST, E. et al. Pangenomic study of *Corynebacterium diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. **Journal of Bacteriology**, v. 194, n. 12, 2012a.

TROST, E. et al. Pangenomic Study of *Corynebacterium diphtheriae* That Provides Insights into the Genomic Diversity of Pathogenic Isolates from Cases of Classical Diphtheria, Endocarditis, and Pneumonia. **Journal of Bacteriology**, v. 194, n. 12, p. 3199–3215, jun. 2012b.

UDDIN, R.; SAEED, K. Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*. **Computational Biology and Chemistry**, v. 48, 2014.

URAKAWA, T. et al. Subcutaneous abscess formation in the upper extremity caused by toxigenic *Corynebacterium ulcerans*. **Journal of Medical Microbiology**, v. 62, n. PART3, p. 489–493, mar. 2013.

VANDENTORREN, S. et al. Toxigenic *Corynebacterium ulcerans* in a fatal human case and her feline contacts, France, March 2014. **Eurosurveillance**, v. 19, n. 38, 2014.

VIANA, M. V. C. et al. Genome sequence of *Corynebacterium ulcerans* strain 210932. **Genome Announcements**, v. 2, n. 6, 2014.

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 December, 2021.

VISSER, L. E. et al. Nasopharyngeal *Corynebacterium ulcerans* diphtheria in the Netherlands. **Weekly releases (1997–2007)**, v. 6, n. 7, 2002.

VITEK, C. R.; WHARTON, M. **Diphtheria in the former Soviet Union: Reemergence of a pandemic disease** **Emerging Infectious Diseases**, 1998.

WADOOD, A. et al. Subtractive genome analysis for in silico identification and characterization of novel drug targets in *Streptococcus pneumoniae* strain JJA. **Microbial Pathogenesis**, v. 115, 2018.

WAGNER, K. S. et al. Diphtheria in the postepidemic period, Europe, 2000-2009. **Emerging Infectious Diseases**, v. 18, n. 2, 2012.

WEIL, L. M. et al. Notes from the Field: Conjunctivitis Caused by Toxigenic *Corynebacterium ulcerans* — Missouri, 2018. **MMWR. Morbidity and Mortality Weekly Report**, v. 68, n. 27, 2019.

WORLD HEALTH ORGANIZATION. Weekly Epidemiology Record. **Releve epidemiologique hebdomadaire**, v. 92, n. 31, p. 417–435, 2017.

XIAO, J. et al. **A brief review of software tools for pangenomics** *Genomics, Proteomics and Bioinformatics*, 2015.

YANG, Y. et al. Genomic diversity, virulence, and antimicrobial resistance of *Klebsiella pneumoniae* strains from cows and humans. **Applied and Environmental Microbiology**, v. 85, n. 6, 2019.

YASUDA, I. et al. Severe pneumonia caused by toxigenic *Corynebacterium ulcerans* infection, Japan. **Emerging Infectious Diseases**, v. 24, n. 3, p. 588–591, 2018.

YOON, S. H. et al. A large-scale evaluation of algorithms to calculate average nucleotide identity. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 110, n. 10, 2017.

YU, N. Y. et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **Bioinformatics**, v. 26, n. 13, p. 1608–1615, 2010.

ZAKIKHANY, K.; EFSTRATIOU, A. **Diphtheria in Europe: Current problems and new challenges** *Future Microbiology*, 2012.

ZHANG, H. et al. Multidrug resistance operon *emrAB* contributes for chromate and ampicillin co-resistance in a *Staphylococcus* strain isolated from refinery polluted river bank. **SpringerPlus**, v. 5, n. 1, 1 dez. 2016.

ZHANG, H. et al. Understanding the high L-valine production in *Corynebacterium glutamicum* VWB-1 using transcriptomics and proteomics. **Scientific Reports**, v. 8, n. 1, 2018.

ZHAO, Y. et al. PGAP: Pan-genomes analysis pipeline. **Bioinformatics**, v. 28, n. 3, 2012.