UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA

INSTITUTO DE QUÍMICA

GABRIELLE TOMÉ CORDEIRO

GENÔMICA COMPARATIVA ENVOLVENDO O SISTEMA DE SECREÇÃO DO TIPO III EM LINHAGENS DE *BURKHOLDERIA CENOCEPACIA* DE ORIGEM AMBIENTAL OU CLÍNICA

RIO DE JANEIRO

2021

GABRIELLE TOMÉ CORDEIRO

GENÔMICA COMPARATIVA ENVOLVENDO O SISTEMA DE SECREÇÃO DO TIPO III EM LINHAGENS DE *BURKHOLDERIA CENOCEPACIA* DE ORIGEM AMBIENTAL OU CLÍNICA

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Rio de Janeiro para obtenção do título de Mestra em Ciências.

Orientadora: Prof.^a Dr.^a Bianca Cruz Neves

Rio de Janeiro 2021

Gabrielle Tomé Cordeiro

Genômica comparativa envolvendo o Sistema de Secreção do Tipo III em linhagens de *Burkholderia cenocepacia* de origem ambiental ou clínica

Dissertação de Mestrado submetida ao Programa de Pós Graduação em Bioquímica, Departamento de Bioquímica do Instituto de Química da Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários para a obtenção do título de Mestre em Ciências.

Aprovada por: Prof. Bianca Cruz Neves (IQ-UFRJ)

Prof. Eidy de Oliveira Santos (UBio-UEZO)

Prof. Leticia Miranda Lery Santos (IOC/Fiocruz)

Prof. Márcia Regina Soares da Silva (IQ-UFRJ)

Rio de Janeiro 2021 Cordeiro, Gabrielle Tomé

GENÔMICA COMPARATIVA ENVOLVENDO O SISTEMA DE SECREÇÃO DO TIPO III EM LINHAGENS DE *BURKHOLDERIA CENOCEPACIA* DE ORIGEM AMBIENTAL OU CLÍNICA/Gabrielle Tomé Cordeiro – Rio de Janeiro; UFRJ/IQ.

Dissertação de Mestrado – Universidade Federal do Rio de Janeiro, Centro de Ciências da Saúde, Instituto de Química / IQ, 2021

Orientadora: Bianca Cruz Neves

Burkholderia cenocepacia 2. sistema de secreção tipo III 3. genômica comparativa
I. Neves, Bianca Cruz II Universidade Federal do Rio de Janeiro

AGRADECIMENTOS

"O sucesso não teria sido atingido se não tivéssemos pessoas capazes construindo nossos projetos" (Moreno, 2021). Essa pequena frase resume o sentimento de gratidão que tenho pelas pessoas que sempre estiveram e estão ao meu lado em momentos tão importantes quanto este.

Acredito que sem a minha fé e o meu Deus não estaria vivenciando essa glória, pois foi com eles que pude enfrentar os obstáculos dessa caminhada surpreendente e um tanto complicada.

Agradeço aos meus pais, Antonio e Eva, que a todo instante me deram carinho nas noites turbulentas e me ofereceram o suporte necessário para que eu pudesse continuar firme na minha jornada. Sou muito grata por tê-los na minha vida e por eles viverem esse sonho, que eu não imaginava que seria possível, junto comigo.

A minha orientadora, Prof.^a Dr.^a Bianca, que esteve junto a mim nos melhores e piores momentos da minha carreira profissional. Anseio um dia ser uma profissional tão amável e admirável quanto ela.

Ao meu melhor amigo e amor da minha vida, Pedro André, que sempre esteve ao meu lado, me apoiando a cada segundo, vibrando comigo as vitórias conquistadas nesta trajetória e deixando os meus dias mais leves.

A toda equipe do Laboratório de Microbiologia Molecular e Proteínas, em especial a Camila, Danielly, Gabriela Breda, Lurian, Michele, Luiza e Rosane, por sempre me ajudar nos trabalhos na bancada e por torcer por mim a todo momento.

As minhas fiéis escudeiras e amigas, Graciela e Hadassa, por me auxiliarem em todo trabalho de bioinformática e por terem sido a minha rocha durante o percurso do mestrado. Sem elas não teria chegado ao fim dessa etapa da minha vida.

As minhas ICs, Yasmin e Ana Gabriela, por me darem a honra de aprender o significado da palavra ensinar. Tenho muito apreço, orgulho e gratidão por acompanhar o crescimento de cada uma.

As minhas amigas Beatriz, Juliana, Núbia, Renata Chapot, Renata Trabach e Gabrielly Ferreira, por disponibilizarem um tempinho da vida delas para me ouvirem e compartilhar a loucura que é ser filha da minerva nesse tempo difícil, chamado pandemia.

Aos meus professores e funcionários do Programa de Pós-Graduação em Bioquímica (PPGBq), que foram essenciais para a minha formação acadêmica e pessoal. Todos os ensinamentos levarei comigo pro resto da vida.

Aos meus amigos e eterna chapa da respresentação discente, Ana Collaço, Luís Felipe e Taissa, por terem dividido comigo o conhecimento que tinham por ser veteranos da casa PPGBq e também por toda força que me davam mesmo quando eu não acreditava no meu potencial.

A minha filha de quatro patas Pretinha, por sempre me fazer companhia, inclusive enquanto escrevia esta dissertação. Solidão é uma palavra que não existe no meu vocabulário por tê-la em minha vida.

Por fim agradeço a cada pessoa que se fez presente na minha vida pessoal e acadêmica, se pudesse colocaria o nome de cada um aqui porém o espaço é restrito. Encerro dizendo o quanto sou feliz por compartilhar esta vida extraordinária com pessoas tão maravilhosas, a vocês o meu singelo e sincero, muito obrigada!

LISTA DE ABREVIATURAS

AAI	Identidade média de aminoácidos
ANI	Identidade média de nucleotídeos
Bcc	Complexo Burkholderia cepacia
bscN	Proteína N de secreção em Burkholderia
cAMP	Adenosina monofosfato cíclico
CFTR	Regulador da condutância transmembranar fibrose cística
dDDH	Hibridização digital DNA-DNA
DGC	Doença granulomatosa crônica
FC	Fibrose cística
HSPs	Pares de segmento de alta pontuação
NADPH	Nicotinamida adenina dinucleotídeo fosfato
NBD	Domínios de ligação a nucleotídeos citoplasmáticos
RND	Resistência - Nodulação - Divisão
Sct	Secreção e translocação celular
T3SS	Sistema de secreção do tipo III
ТМ	Túnel transmebranar
TMD	Domínio transmembranar

LISTA DE SIGLAS

NIH National Institutes of Health

Índice

1. RESUMO	12			
2. ABSTRACT	14			
3. INTRODUÇÃO	16			
3.1 BURKHOLDERIA CENOCEPACIA	16			
3.2 FIBROSE CÍSTICA	17			
3.3 IMPACTO NA SAÚDE HUMANA E NA ECONOMIA	19			
3.4 SISTEMAS DE SECREÇÃO	21			
3.4.1 Sistema de Secreção do Tipo III	22			
3.5 T3SS E A BURKHOLDEIRA CENOCEPACIA	29			
3.6 TAXONOMIA GENÔMICA	30			
4. OBJETIVOS	34			
4.1 OBJETIVO GERAL	34			
4.2 OBJETIVOS ESPECÍFICOS	34			
5. MATERIAIS E MÉTODOS	35			
5.1 SELEÇÃO E ANOTAÇÃO DOS GENOMAS BACTERIANOS	35			
5.2 ANÁLISE COMPARATIVA	35			
5.3 ANÁLISE FILOGENÉTICA E TAXONÔMICA	36			
5.3.1 Taxonomia genômica	36			
5.3.1.1 Genes rRNA 16S, <i>recA</i> e método MLSA	36			
5.3.1.2 Métodos dDDH, ANI e AAI	37			
5.3.2 Alinhamento múltiplo	38			
5.3.3 Árvores filogenéticas e taxonômicas				
5.3.4 Rstudio	39			
6. RESULTADOS E DISCUSSÃO	39			
6.1 CEPAS BACTERIANAS E ANOTAÇÃO GENÔMICA	39			
6.2 ORGANIZAÇÃO E COMPARAÇÃO GENÔMICA	40			
6.3 ANÁLISE FILOGENÉTICA	47			
6.4 TAXONOMIA GENÔMICA	49			
7. CONCLUSÃO E PERSPECTIVAS				
8. REFERÊNCIA BIBLIOGRÁFICA	61			
MATERIAL SUPLEMENTAR	72			

Índice de figuras

Figura 1: Estrutura molecular do CFTR humano na forma desfosforilada e livre de ATP.
Figura 2: Principais sistemas de secreção bacterianos
Figura 3: Composição das estruturas resolvidas dos componentes individuais dos
T3SSs de diferentes espécies bacteriana
Figura 4: Componentes estruturais chave do anel citoplasmático T3SS e da plataforma
de classificação
Figura 5: Componentes que compõem o filamento da agulha
Figura 6: Tamanho dos genomas (Mb) dos isolados de origem ambiental e clínica41
Figura 7: Número de sequências codificantes (CDS) dos isolados de origem ambiental
e clínica41
Figura 8: Percentual do conteúdo de Guanina-Citosina (GC%) dos genomas de origem
ambiental e clínica
Figura 9: Clusters gênicos do Sistema de Secreção do tipo III (T3SS) de B.
cenocepacia
Figura 10: Árvore filogenética com base no gene <i>sctN</i> dos genomas analisados48
Figura 11: Árvore taxonômica baseada no gene rRNA 16S50
Figura 12: Árvore taxonômica baseada no gene <i>recA</i>
Figura 13: Árvore taxonômica baseada nos sete genes housekeeping (recA, gyrB, atpD,
gltB, lepA, phaC e trpB)
Figura 14: Heatmaps dos valores de dDDH, ANI e AAI, a partir de comparações de
genoma-genoma e identidade média de nucleotídeos e aminoácidos

Índice de tabelas

Tabela 1: Principais componentes estruturais dos T3SSs e seus homólogos flagelare	ólogos flagelares.		
	24		
Tabela 2: Número de clusters do T3SS encontrados nos genomas selecionados de	В.		
cenocepacia	45		
Tabela 3: Domínios conservados presentes na proteína conservada.	46		

Cordeiro, GT. Genômica comparativa envolvendo o Sistema de Secreção do Tipo III em linhagens de *Burkholderia cenocepacia* de origem ambiental ou clínica. Rio de Janeiro: Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, 2021.

1. RESUMO

O sistema de secreção do tipo III (T3SS) é descrito em muitas bactérias Gramnegativas, incluindo o patógeno oportunista humano Burkholderia cenocepacia. Além de mediar processos patogênicos em humanos, animais e plantas, o T3SS também pode promover interações mutualísticas benéficas. Algumas espécies associadas a plantas requerem T3SS para colonização completa, desempenhando um papel protetor contra fitopatógenos e ajudando no crescimento do hospedeiro. Na B. cenocepacia J2315, o Bsc-T3SS é composto por 12 genes que codificam proteínas conservadas, responsáveis pela formação do aparelho de secreção. Assim, o presente projeto tem como objetivo identificar genes do T3SS em uma gama de cepas de B. cenocepacia de diferentes nichos, clínicos ou ambientais, a fim de melhor compreender o papel deste sistema no seu ciclo de vida e na associação com hospedeiros. Análises comparativas e filogenéticas foram utilizadas para investigar a distribuição, evolução, variabilidade intraespecífica e estrutura dos operons do T3SS nos genomas sequenciados. Foram utilizados 50 genomas de origem ambiental ou clínica, disponíveis no banco de dados do GenBank. A seleção de genomas foi baseada na origem da cepa e na qualidade das sequências. Foram empregados apenas genomas obtidos com sequenciamento Illumina / PacBio e que foram anotados através da a ferramenta Prokka (v. 1.14.6). Para a identificação dos genes do T3SS, foi realizado um mapeamento contra o banco de dados T3Enc utilizando o programa BLAST (ncbi-blast + (2.9.0-2)), no qual as proteínas foram consideradas homólogas se compartilhassem uma identidade $\geq 40\%$ e com e-value 0,001. Dos 50 genomas selecionados, apenas 1 genoma não apresentou qualquer cluster do T3SS, sendo este um isolado do escarro de um paciente fibrocístico. Análises filogenéticas e taxonômicas foram realizadas pela pelo software MEGA (v. X) utilizando o método estatístico Maximum Likelihood, com valor de bootstrap igual a 1500 e as árvores foram visualizadas pela plataforma online iTOL (v. 5). Os resultados taxonômicos indicam uma distinção filogenética entre as cepas de B.

cenocepacia. Com essas análises, espera-se obter *insights* que possibilitem o conhecimento de fatores relacionados à virulência de *B. cenocepacia* e/ou sua adaptação ao meio ambiente. Essas evidências e dados experimentais obtidos em nosso grupo sugerem que *B. cenocepacia* pode interagir ativamente com plantas como hospedeiros intermediários em seu ciclo de vida, por meio de uma interface funcional dependente de T3SS. Trabalhos futuros serão direcionados para a compreensão dos papéis que esses T3SSs desempenham na biologia de *B. cenocepacia*.

PALAVRA-CHAVE: *Burkholderia cenocepacia*, sistema de secreção tipo III, genômica comparativa

Cordeiro, GT. Comparative genomics involving the Type III Secretion System in *Burkholderia cenocepacia* strains of environmental or clinical origin. Rio de Janeiro: Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, 2021.

2. ABSTRACT

Type III secretion system (T3SS) has been described in many Gram-negative bacteria, including human opportunistic pathogen Burkholderia cenocepacia. In addition to T3SS being associated with diseases in humans, animals and plants, it can also be associated with beneficial mutualistic interactions. Some species associated with plants require T3SS for complete colonization, playing a protective role against phytopathogens and helping host growth. In B. cenocepacia J2315, T3SS is composed of 12 genes that encode conserved proteins, responsible for forming the secretion apparatus. Thus, the current project aimed to identify T3SS genes in a range of B. cenocepacia strains from different niches, clinical or environmental, in order to better understand the role of this system in its life cycle and host association. Comparative and phylogenetic analyses were used to investigate the distribution, evolution, intrastrain variability and operon structure of the T3SS in the sequenced genomes. 50 genomes of environmental or clinical origin were used, which are available in the GenBank database. The selection of genomes was based on the strain's origin and quality of the sequences. Genomes obtained with Illumina / PacBio sequencing and scaffolds above the value of 200 were used and annotated using the Prokka tool (v. 1.14.6). For the identification of T3SS genes in the selected genomes, a mapping was carried out against the T3Enc database using the BLAST program (ncbi-blast + (2.9.0-2)), in which proteins were considered homologous if they shared an identity $\geq 40\%$ and with an e-value of 0.001. Of the 50 selected genomes, only 1 genome did not show any cluster of the T3SS, which is an isolated sputum from a FC patient. Phylogenetic and taxonomic analyzes were performed by the MEGA software (v. X) using the Maximum Likelihood statistical method, with a bootstrap value equal to 1500 and as trees were visualized by the online platform iTOL (v. 5). The taxonomic results indicate

a phylogenetic distinction between *B. cenocepacia* strains. With these analyses, it is expected to obtain insights that enable the knowledge of factors related to the virulence of *B. cenocepacia* and/or its adaptation to the environment. These insights indicate that the two types of T3SSs of *B. cenocepacia* appear to have been independently acquired and may play different roles relating to pathogenicity, host range determination and/or niche adaptation. This evidence and experimental data obtained in our group suggests that *B. cenocepacia* can actively interact with plants as intermediate hosts in their life cycle, through a functional interface dependent on T3SS. Current work is focused on understanding the roles of these T3SSs in the biology of *B. cenocepacia*.

KEYWORD: *Burkholderia cenocepacia*, type III secretion system, comparative genomics

3. INTRODUÇÃO

3.1 BURKHOLDERIA CENOCEPACIA

O gênero *Bukholderia* abrange espécies patogênicas para plantas e animais, além de espécies simbióticas e saprófitas, presentes em uma ampla gama de habitats ambientais e clínicos (ESTRADA-DE LOS SANTOS et al., 2018). Espécies do gênero *Burkholderia*, inicialmente foram classificadas como espécies do gênero *Pseudomonas* (WALTER H. BURKHOLDER, 1950), somente em 1992 o gênero *Burkholderia* foi proposto com a finalidade de destinar essas espécies, antes identificadas como *Pseudomonas* e pertencentes ao grupo II de espécies de acordo com os resultados de homologia da hibridização rRNA-DNA, para um novo gênero a partir de dados das sequências do rRNA 16S e características fenotípicas (YABUUCHI et al., 1992).

Os gêneros *Bukholderia sensu stricto, Caballeronia, Trinickia, Paraburkholderia, Mycetohabitans* e *Robbsia* compõem atualmente o grande gênero *Burkholderia* (DOBRITSA; SAMADPOUR, 2016; ESTRADA-DE LOS SANTOS et al., 2018) A *Burkholderia sensu stricto* integra cerca de 31 espécies distintas e ao menos 22 espécies pertencem ao complexo *Burkholderia cepacia* (Bcc), incluindo a *Burkholderia cenocepacia*. O Bcc reúne espécies que são patógenos humanos oportunistas e comumente isoladas em ambientes úmidos, solo e água, onde podem sobreviver durante um longo período de tempo (SFEIR, 2018).

O desenvolvimento do campo genômico possibilitou a divisão da espécie *Burkholderia cepacia* em cinco genomovares geneticamente distintos, embora sejam similares fenotipicamente (WALLNER et al., 2019). Subsequentemente, os genomovares aumentaram quanto ao seu número e foram gradualmente categorizados em nove táxons distintos, baseando-se a sua caracterização no gene *recA* (TURTON et al., 2007). A partir do polimorfismo deste gene mostrou-se que a *B. cepacia* é geneticamente heterogênea, sendo composta por pelo menos quatro subgrupos filogenéticos (IIIA, IIIB, IIIC e IIID), tendo a sua maior recorrência para os subgrupos IIIA e IIIB. A cepa epidêmica ET12 se encontra no subgrupo IIIA e geralmente é associada à síndrome de cepacia (SAJJAN et al., 2007). Com essas informações e

através de estudos de hibridização de DNA-DNA foi possível diferenciar a *B. cenocepacia* da *B. cepacia* (CHIARINI et al., 2004; VANDAMME et al., 2003). Alguns dos subgrupos de *B. cenocepacia* podem ser frequentemente isolados de ambientes naturais e em associação com plantas (BALDWIN et al., 2007; DALMASTRI et al., 2007) incluindo o subgrupo IIIA.

B. cenocepacia é uma bactéria Gram-negativa, originalmente resistente a distintas classes de antibióticos utilizados em práticas clínicas e sua característica patogênica é ocasionada por diversos determinantes de virulência (SCOFFONE et al., 2017). *B. cenocepacia*, como outras espécies bacterianas do Bcc, é frequentemente isolada do escarro de pacientes com FC, está associada ao declínio acelerado da função pulmonar e ao aumento da mortalidade devido à resistência antimicrobiana (SFEIR, 2018). Além dessas associações, a capacidade de adaptação às mudanças ambientais tornam o tratamento das infecções por *B. cenocepacia* mais desafiador. Esta espécie bacteriana é considerada uma das espécies do Bcc que acarreta mais problemas em pacientes imunocomprometidos, além disso estudos recentes relatam que a *B. cenocepacia* possui forte capacidade de associação com plantas, podendo viver livremente na rizosfera da planta (EBERL; VANDAMME, 2016; YOU et al., 2020).

A *B. cenocepacia* J2315, é a cepa mais comumente estudada no âmbito da taxonomia genômica, fornecendo *insights* sobre a sua adaptação no ambiente clínico. O genoma completo desta cepa compõe três cromossomos circulares de 3.870.082, 3.217.062 e 875.977 pb e um plasmídeo de 92.661 pb (HOLDEN et al., 2009). Esses quatro replicons codificam 3.537, 2.849, 776 e 99 sequências codificantes (CDSs) previstas, respectivamente, dos quais 126 são pseudogenes ou genes parciais (HOLDEN et al., 2009).

3.2 FIBROSE CÍSTICA

Fibrose cística é conhecida como uma doença congênita hereditária e altamente perigosa, causada por mutações no gene regulador da condutância transmembranar da fibrose cística (*CFTR*), o qual codifica a proteína CFTR (CUTTING, 2005). CFTR compõe a superfamília de cassetes de ligação de ATP (ABC), sendo a única entre as

proteínas ABC que funciona como um canal iônico, conduzindo ânions por seu gradiente eletroquímico (LIU et al., 2017a). A estrutura da proteína CFTR (**Figura 1**) consiste em dois domínios transmembranares (TMD1 e TMD2) que constituem a via de translocação, dois domínios de ligação a nucleotídeos citoplasmáticos (NBD1 e NBD2) que hidrolisam o ATP, um domínio regulatório (R) que permite a abertura do canal ao ser fosforilado (CHENG et al., 1991). Assim a fosforilação abre o canal CFTR e a hidrólise (desfosforilação) o fecha. Outra estrutura presente e importante na proteína CFTR é o túnel transmembranar (TM8), uma estrutura de 3 hélices curtas que reveste os poros, desempenhando dessa forma um papel crítico na condução e passagem de íons. A alteração na flexibilidade de TM8, a partir de uma mutação dentro da quebra helicoidal (L927P), causa a doença fibrose cística (LIU et al., 2017a).



Figura 1: Estrutura molecular do CFTR humano na forma desfosforilada e livre de ATP. As estruturas helicoidais na cor verde apontam os domínios TMD2 e NBD2. As estruturas helioidais na cor azul identificam os domínios TMD1 e NBD1. A densidade eletrônica mostrada em vermelho corresponde às regiões não estruturadas, inserção R, dentro de NBD1 ou do domínio R. As regiões não resolvidas na estrutura são identificadas com as linhas tracejadas, elas não são correspondem conectividade ou localização das regiões ausentes, são mostradas para fins de visualização (Fonte: LIU et al., 2017).

A proteína CFTR é encontrada em diversos órgãos do corpo, incluindo o trato respiratório, o trato gastrointestinal, o fígado, o pâncreas, assim como o trato reprodutor masculino. Nas vias aéreas, o mal funcionamento dessa proteína leva ao aumento da

espessura do muco, que não é eliminado pelo sistema mucociliar (MCDERMOTT; REECE; RENWICK, 2019). Isso, por sua vez, predispõe o indivíduo à infecção crônica do trato respiratório, seguida de processos inflamatórios graves (COHEN; PRINCE, 2012). As infecções das vias aéreas estão associadas ao declínio progressivo da função pulmonar e, em último caso, à insuficiência respiratória, sendo esta a principal causa de mortalidade em pacientes fibrocísticos (GIBSON; BURNS; RAMSEY, 2003; KONSTAN et al., 2007; LIOU et al., 2001).

Indivíduos com FC desenvolvem infecções ao longo da sua vida e algumas espécies do complexo *Burkholderia cepacia* (Bcc) podem ser identificadas nesses pacientes. As infecções pulmonares por cepas do Bcc podem evoluir para um quadro sistêmico denominado "síndrome de cepacia", com aumento da mortalidade. Esta síndrome é caracterizada por uma rápida deterioração clínica devido à pneumonia necrosante ou sepse, resultando em morte prematura (BODILIS et al., 2018). Algumas cepas desse complexo aparentam ser mais virulentas do que outras, pois a grande maioria das infecções é causada por *Burkholderia cenocepacia* ou *Burkholderia multivorans* (LIPUMA, 2005). Em algumas regiões da Europa, assim como no Canadá, a *B. cenocepacia* é responsável por mais de 80% das infecções bacterianas em pacientes com FC (WALLNER et al., 2019).

3.3 IMPACTO NA SAÚDE HUMANA E NA ECONOMIA

B. cenocepacia é uma da espécies mais encontradas em pacientes portadores da doença de FC, acarretando a piora do quadro clínico do paciente. Os pacientes que possuem a doença granulomatosa crônica (DGC), sendo esta uma imunodeficiência que sucede na implicação do complexo enzimático nicotinamida adenida dinucleotídeo fosfato (NADPH)-oxidase dos fagócitos, impossibilitando-os de promover uma eficiente resposta oxidativa em combate aos patógenos, também podem ser acometidos de uma piora no seu quadro clínico quando infectados por *B. cenocepacia* (JOHNSTON, 2001).

Estudos no campo da genômica comparativa, relatam a existência de marcadores genéticos que estão intimamente relacionados à patogenicidade em *B*.

cenocepacia. O gene *cblA* (*cable pilus*) e o *adhA* (adesina 22-kDa) são considerados importantes para a adesão à célula hospedeira, inclusive a presença do *adhA* é essencial para o potencial oportunista de cepas da *B. cenocepacia* (WALLNER et al., 2019). O megaplasmídeo pC3 também desempenha um papel importante na patogênese de várias cepas de *B. cenocepacia* e em diversos modelos de infecção (AGNOLI et al., 2012). Wallner *et al* (2019) relatam outros dados genômicos e filogenéticos, que indicam que *B. cenocepacia* pode ter evoluído de um estilo de vida associado a plantas para um estado oportunista em humanos. Neste mesmo trabalho, demonstra-se que a *B. cenocepacia* representa um grupo genômico heterogêneo, o que se reflete nos seus diferentes perfis metabólicos e de virulência, com genes específicos detectados nas análises filogenéticas (WALLNER et al., 2019).

Além disso, a distribuição ubíqua do T3SS nas bactérias Gram-negativas, intimamente associado às relações simbióticas ou patogênicas nas plantas, representa um problema para a saúde humana, pois hospedeiros vegetais podem assim desempenhar um papel de reservatório, assintomático ou não, de *B. cenocepacia*, uma vez que a bactéria é capaz de estabelecer uma associação endofítica.

Com base nos dados da literatura, cepas de *B. cenocepacia* que estão incluídas nos subgrupos IIIA e IIIB, foram identificadas como causadores de doenças em cebola na região semi-árida do Nordeste do Brasil (OLIVEIRA et al., 2017b). Cepas do Bcc também foram encontradas, com base no sequenciamento rRNA 16S, em outras plantas. Dessa forma demonstra-se o potencial, principalmente das linhagens de *B. cenocepacia*, como agente patogênico em plantas nesta região do Brasil (BAIA et al., 2017).

O entendimento dos mecanismos moleculares envolvidos na interface bactériahospedeiro é fundamental para o desenvolvimento de estratégias de controle desta e outras espécies patogênicas. Portanto, a investigação genômica da abundância e características do T3SS em um grupo amplo e diversificado de isolados de *B. cenocepacia*, de diferentes origens e hospedeiros, é a estratégia adotada nesse trabalho. Essa espécie representa não somente uma ameaça à saúde humana, mas também à economia e agricultura, devido ao seu potencial patogênico para algumas espécies vegetais.

3.4 SISTEMAS DE SECREÇÃO

Muitas espécies bacterianas, incluindo a Burkholderia cenocepacia, utilizam mecanismos diversificados para estabelecer a interface molecular para a interação bactéria-hospedeiro e/ou adaptação ao meio ambiente (ABBÊS et al., 2014; KOOGER et al., 2018). Os sistemas de secreção bacterianos estão na linha de frente das interfaces moleculares, mediando a conversa cruzada com os hospedeiros eucarióticos e desempenhando um papel fundamental no resultado de relações patogênicas ou simbióticas. Os principais sistemas são classificados do I ao VI (Figura 2) e alguns deles, incluindo os sistemas de secreção do tipo III (T3SS), do tipo IV (T4SS) e do tipo VI (T6SS), foram extensivamente estudados em algumas interações bactériahospedeiro (CHRISTIE, 2019). O T6SS, sistema amplamente estudado em competição entre espécies bacterianas, atua como um dispositivo injetando proteínas antibacterianas tóxicas em células bacterianas rivais, modulando assim as comunidades polimicrobianas (MONJARÁS FERIA; VALVANO, 2020). Diversas espécies de bactérias utilizam seus T6SS para funções específicas em relação ao hospedeiro, nicho ou estratégia de sobrevivência do organismo. Dessa forma demonstra-se o papel do T6SS exclusivamente na patogênese bacteriana.



Figura 2: Principais sistemas de secreção bacterianos. Bomba de RND e Sistemas de secreção do tipo I, II, II, IV e VI são representados na figura, levando em consideração a nomenclatura utilizada em *Salmonella* spp. (Fonte: (COSTA et al., 2015).

Entre os sistemas de secreção bacterianos, o T3SS é um dos mais sofisticados e bem estudados e tem atraído muita atenção desde a sua descoberta devido ao seu papel central na virulência e patogenicidade bacteriana (PORTALIOU et al., 2016). Os estudos concentrados nos mecanismos de virulência, assim como a caracterização estrutural e funcional do T3SS, abrem portas para o desenvolvimento de drogas antimicrobianas e vacinas (DENG et al., 2017). A partir dessas informações relatadas iremos nos aprofundar sobre o T3SS no próximo capítulo.

3.4.1 Sistema de Secreção do Tipo III

Os T3SSs são nanomáquinas transpassadas em membrana dupla e encontradas em diversos gêneros de bactérias Gram-negativas patogênicas (GALÁN; WOLF-WATZ, 2006). Esse sistema de distribuição proteica proporciona a transferência de proteínas efetoras bacterianas para o citoplasma ou para a membrana plasmática de células eucarióticas alvo. Esses efetores, dentro do hospedeiro, modulam ou corrompem funções específicas da célula alvo, possibilitando a invasão e colonização bacteriana (COSTA et al., 2015). Os T3SSs estão intrinsecamente associados aos flagelos, de maneira evolutiva, e muitas subestruturas e componentes que são responsáveis pela montagem do aparato estrutural são extremamente conservados nessas duas organelas (DENG et al., 2017). A organização e a função dessas estruturas complexas necessitam de mais de 20 proteínas, que ao longo dos anos foram nomeadas

de modo exclusivo para cada espécie de bactéria em que eram caracterizadas, dificultando comparações entre as espécies (BÜTTNER, 2012; NOTTI; STEBBINS, 2016; PORTALIOU et al., 2016). A partir disso, percebeu-se a necessidade de unificar a nomenclatura dos componentes conservados no T3SS, dessa forma foi sugerido o prefixo de secreção e translocação celular (Sct) (HUECK, 1998). A nomenclatura Sct (**Tabela 1**) foi extensamente adotada e será utilizada ao longo deste trabalho.

Nomenclatura Unificada	Função proposta	Yersinia spp.	Salmonella spp.		EPEC e EHEC	Shigella spp.	Chlamydia spp.	P. aeruginosa	P. syringae	Rhizobium spp.	Aparato Flagelar
			SP-1	SP-2							
Corpo Basal											
SctC	Secretina, anel de membrana externa	YscC	InvG	SsaC	EscC	MxiD	CdsC	PscC	HrcC	RhcC1–RhcC2	-
-	Secretina pilotina	YscW	InvH	-	-	MxiM	-	ExsB	HrpT	-	-
SctD	Anel de membrana interna	YscD	PrgH	SsaD	EscD	MxiG	CdsD	PscD	HrpQ	Y4yQ	FliG
SctJ	Anel de membrana interna	YscJ	PrgK	SsaJ	EscJ	MxiJ	CdsJ	PscJ	HrcJ	NolT	FliF
SctF	Agulha	YscF	PrgI	SsaG	EscF	MxiH	CdsF	PscF	HrpA	NopA–NopB	FlgE
SctI	Haste interna	YscI	PrgJ	SsaI	EscI	MxiI	-	PscI	HrpB	NolU	-
-	Transglicosilase lítica	-	IagB	-	EtgA	IpgF	-	-	HrpH e HopP1	l –	FlgJ
Aparelho de exportação											
SctU	Autoprotease	YscU	SpaS	SsaU	EscU	Spa40	CdsU	PscU	HrcU	RhcU	FlhB
SctV	Portão de exportação	YscV	InvA	SsaV	EscV	MxiA	CdsV	PcrD	HrcV	Y4yR	FlhA
SctR	Componente de membrana interna	YscR	SpaP	SsaR	EscR	Spa24	CdsR	PscR	HrcR	RhcR	FliP
SctS	Componente de membrana interna	YscS	SpaQ	SsaS	EscS	Spa9	CdsS	PscS	HrcS	RhcS	FliQ
SctT	Componente de membrana interna	YscT	SpaR	SsaT	EscT	Spa29	CdsT	PscT	HrcT	RhcT	FliR
Anel citoplasmático											
SctQ	Anel citoplasmático	YscQ	SpaO	SsaQ	SepQ	Spa33	CdsQ	PscQ	HrcQ	RhcQ	FliM–FliN
Complexo ATPase											
SctN	ATPase	YscN	InvC	SsaN	EscN	Spa47	CdsN	PscN	HrcN	RhcN	FliI
SctL	Estator	YscL	OrgB	SsaK	EscL	MxiN	CdsL	PscL	HrpE	NoIV	FliH
SctO	Talo/Haste	YscO	Invl	SsaO	EscO	Spa13	CdsO	PscO	HrpO	Y4yJ	FliJ
SctK	Cofator ATPase	YscK	OrgA	-	EscK	MxiK	-	PscK	HrpD	-	-
Reguladores	Dogulador da										
SctP	comprimento de agulha	YscP	InvJ	SsaP	EscP	Spa32	CdsP	PscP	HrpP	-	FliK
SctW	Regulador de liga/desliga	YopN–TyeA	InvE	SsaL	SepL	MxiC	CopN	PopN	HrpJ	-	-
-	Regulador de liga/desliga	-	-	SpiC	SepD	-	-	-	-	-	-
Translocadores											
SctB	Poro de translocação	YopD	SipC	SseD	EspB	IpaC	CopD	PopD	-	-	-
SctE	Poro de translocação	YopB	SipB	SseC	EspD	IpaB	CopB	PopB	HrpK	NopX	
SctA	Ponta da agulha ou filamento	LcrV	SipD	SseB	EspA	IpaD	CT584	PcrV	-	-	FliC

Tabela 1: Principais componentes estruturais dos T3SSs e seus homólogos flagelares.

Na tabela é apresentado os componentes das espécies *Yersinia, Salmonella, Escherichia coli* enteropatogênica (EPEC) e hemorrágica (EHEC), *Shigella, Chlamydia, Pseudomonas aeruginosa, Pseudomonas syringae, Rhizobium.* Além dessas espécies são listados os componentes do aparato flagelar, similar ao T3SS, e a sua nomenclatura universal. (Adaptado de DENG et al., 2017)

A estrutura do T3SS (**Figura 3**) é identificada no formato de seringa e conhecida como complexo de agulha ou injetissoma. Esta arquitetura é constituída por duas subestruturas principais, corpo basal e filamento em forma de agulha (COSTA et al., 2015). O corpo basal é composto por dois anéis de membrana interna e externa e para emplilhá-los é necessária a via secretora geral (via Sec), visto que os componentes do anel apresentam sinais de secreção amino-terminais cliváveis e Sec-dependentes. Os principais componentes do corpo basal são o SctC, SctD e SctJ, que contêm uma dobra em forma de cunha e possui uma sequência primária diversificada, contudo estruturalmente conservada (DENG et al., 2017). O anel de membrana externa é composto pelos domínios N-terminais de SctC e atinge profundamente o periplasma, entrando em contato direto com o anel de membrana interna de SctD e estruturando uma ponte entre os anéis de membrana interna e externa (WORRALL et al., 2016). Os componentes da membrana interna SctJ e SctD apresentam simetrias distintas, assim ambos formam dois anéis concêntricos de membrana interna, com SctD contornando o SctJ (DENG et al., 2017).



Figura 3: Composição das estruturas resolvidas dos componentes individuais dos T3SSs de diferentes espécies bacteriana. Os componentes proteicos e a sua localização no T3SS são demonstrados através da sua estrutura molecular, exceto os componentes SctR, SctS, SctT e SctK (Adaptado de (DENG et al., 2017).

O aparelho de exportação inclui os seguintes componentes: SctR, SctS, SctT, SctU e SctV. Este aparato se localiza na base do corpo basal e se encontra embutido na membrana interna, funcionando como um canal de entrada para os substratos (WAGNER et al., 2010). SctR, SctS e SctT compõe uma subestrutura identificada como um pentâmero central e o SctV se incorpora na base injetável pré-montada, formada por SctRST (DIETSCHE et al., 2016). A autoprotease SctU possui duas conformações diferentes, a pré-clivada e a pós-clivada de seu domínio C-terminal, podendo interagir de maneira distinta com outros componentes do T3SS, incluindo o SctV. Essa interação gera uma regulação na troca de secreção de substratos iniciais (como SctI e SctF) para substratos tardios (como translocadores e efetores) (DENG et al., 2017). O anel C, também conhecido como SctQ, é uma estrutura dinâmica que circula entre um pool citosólico de proteínas e a base do injetissomo, eventualmente definindo a carga e descarga do substrato durante a secreção (DIEPOLD et al., 2015). Este componente se encontra tanto na forma de comprimento total quanto na forma C-terminal mais curta, contendo domínios de apresentação de superfície de antígenos, que medeiam a oligomerização de proteínas (NOTTI et al., 2015).

A ATPase SctN, componente mais conservado do aparato do T3SS, se localiza no centro abaixo do SctV (portão de exportação). As proteínas SctQ e SctK, localizadas ao final de SctL, fornecem um local de encaixe para o SctN (DENG et al., 2017). SctL e SctO funcionam como um estator e talo, respectivamente, ancorando a ATPase ao SctQ regulando a atividade do aparato de exportação (ZARIVACH et al., 2007). As afinidades dos complexos efetor-chaperona (chaperona de classe I e II) pela ATPase SctN na plataforma de classificação definem a hierarquia de secreção (LARA-TEJERO et al., 2011). A plataforma de classificação é carregada inicialmente com as proteínas translocadoras e suas chaperonas de classe II. Uma vez que os translocadores são secretados, formam um poro funcional na membrana do hospedeiro, auxiliando na secreção das proteínas efetoras, que estão acompanhadas das chaperonas de classe I.

Os componentes proteicos, SctR, SctS, SctT e SctK não são demonstrados na figura, pois não há informações estruturais para esses componentes. A **Figura 4a** demonstra a localização do complexo que o SctRST compõe no aparato de exportação, acoplado aos componentes SctU e SctV, assim como na **Figura 4b** onde se encontra o SctK no complexo ATPase, a partir das informações relatadas na literatura.



Figura 4: Principais componentes estruturais do anel citoplasmático do T3SS e da plataforma de classificação. a | O aparelho de exportação do sistema de secreção tipo III (T3SS) representa a porta de entrada do canal T3SS para substratos. É composto por SctQRSTUV na membrana interna e está localizado abaixo dos anéis da membrana interna que são formados por SctD e SctJ. A proteína de porta de exportação SctV (vermelha) forma um anel oco com a ATPase SctN (azul-verde). O complexo da plataforma de classificação ATPase é envolvido pelo anel citoplasmático oco (anel-C) que é formado por SctQ (rosa). O complexo ATPase e o anel C formam a plataforma de classificação, que funciona como centro de recrutamento para substratos. b | O anel citoplasmático e a plataforma de classificação vistos de baixo. Em *Shigella* spp. e *Salmonella* spp., os complexos citoplasmáticos dos T3SSs foram recentemente propostos, consistindo de um eixo central (a ATPase SctN) e seis raios (o estator SctL), cada um com uma estrutura semelhante a uma cápsula terminal (SctQ, que forma o anel C). Acredita-se que esta estrutura possa ser altamente dinâmica e, portanto, menos estática do que esta figura sugere (Adaptado de (BURKINSHAW; STRYNADKA, 2014; DENG et al., 2017).

O complexo do filamento da agulha, SctF, se conecta na extremidade distal do complexo SctRST por cerca de seis cópias, montado em formato helicoidal, da proteína de haste interna SctI (DIETSCHE et al., 2016), as extremidades das subunidades do complexo SctRST ficam voltadas para fora do complexo, enquanto que os terminais da proteína do filamento da agulha SctF se situam voltados para dentro, explicando que SctI é essencial para conectar firmemente o aparelho de exportação e o filamento da agulha (WAGNER et al., 2018). Essa conformação SctI-SctF é apresentado na **Figura 5**. Em alguns patógenos animais, o filamento da agulha é coberta na extremidade

extracelular por um "complexo de ponta", formado por SctA. Este complexo pode interromper o alongamento da agulha, prevenindo a secreção efetora precocemente e auxiliando na montagem do poro de translocação nas membranas da célula hospedeira (DENG et al., 2017). Os filamentos de agulha do T3SS de patógenos vegetais são consideravelmente mais longos do que os filamentos dos T3SSs de patógenos animais, supostamente para que possam penetrar na espessa matriz de celulose da parede celular vegetal (JIN; HE, 2001). Essas mudanças na estrutura do filamento de agulha são provavelmente adaptações evolutivas para infectar diferentes hospedeiros.



Figura 5: Componentes que compõem o filamento da agulha. Estrutura, baseada em homologia da haste interna de *Salmonella* SPI-1 T3SS, SctI em verde, e o filamento de agulha SctF. O SctI, aqui apresentado, representa apenas a primeira volta helicoidal do filamento. Figura a esquerda, vista lateral, e direita vista superior (Adaptado de (TORRES-VARGAS et al., 2019).

3.5 T3SS E A BURKHOLDEIRA CENOCEPACIA

A *B. cenocepacia*, assim como outras espécies do Bcc, parecem utilizar diferentes mecanismos para fins de interação com o hospedeiro, seja vegetal ou animal. O T3SS tem se mostrado essencial para os processos de colonização, adaptação ao meio e infecção ao hospedeiro (DENNIS, 2017). A maioria dos patógenos bacterianos, incluindo a *B. cenocepacia*, entregam os fatores de virulência no ambiente extracelular por meio do T3SS (MALEŠEVIĆ et al., 2017). Estudos direcionados para a caracterização dos *clusters* gênicos do T3SS são necessários para que possamos

compreender envolvimento do T3SS nas interações mutualísticas benéficas entre bactéria-hospedeiro, adaptação bacteriana ao meio e a patogênese (animais e plantas) (AUSSEL; BEUZÓN; CASCALES, 2016; HUECK, 1998; PIROMYOU et al., 2019). Visando investigar o papel do T3SS na patogênese animal, Tomich et al (2003) realizaram a construção de um mutante nulo para o gene bscN ($\Delta bscN$), o qual codifica a proteína SctN responsável pela geração de energia que impulsiona a secreção de proteínas efetoras, e o empregaram em modelo animal (camundongo). Dessa forma o trabalho forneceu indícios de que esse sistema tenha extrema importância para a patogênese, embora não forneça dados conclusivos, visto que o modelo animal era capaz de se livrar rapidamente da cepa mutante (TOMICH et al., 2003). Em um trabalho anterior do nosso grupo foi demonstrado que quando a mesma cepa mutante é inoculada em Oryza. sativa L. (arroz), sendo este um modelo in planta, a bactéria deficiente pro T3SS não consegue realizar o processo de colonização endofítica, diferentemente da selvagem. Na Medicago sativa L. (alfafa), outro modelo in planta, a cepa $\Delta bscN$ não causa prejuízo ao desenvolvimento da planta, enquanto a cepa selvagem causa grave comprometimento ao desenvolvimento da planta. Esses achados revelam de forma nítida a atenuação na capacidade interativa entre B. cenocepacia e as plantas hospedeiras quando esse sistema não é funcional (mutante bscN), indicando a importância do T3SS como um elemento essencial da interação da bactéria com hospedeiros vegetais (ABBÊS et al., 2014).

3.6 TAXONOMIA GENÔMICA

A taxonomia microbiana é uma abordagem que engloba a criação de novos táxons, identificação de isolados em espécies conhecidas e a nomenclatura (THOMPSON et al., 2013). Inicialmente Woese et al. (1990) sugeriu um sistema biológico de classificação para *Archaea, Bacteria* e *Eukarya,* baseando-se nos dados de comparações de sequências de genes da pequena subunidade ribossômica (GARRITY, 2016; WOESE; KANDLER; WHEELIS, 1990). Ao longo dos anos subsequentes, esse procedimento se apresentou como um método padrão para a classificação de procariotos (GARRITY, 2016). Atualmente, esse procedimento se alia a dados

fenotípicos e genotípicos (polifásicos) complementares, assim como informações quimiotaxonômicas para classificar espécies bacterianas (ANDRADE, 2016).

Dentre os métodos amplamente utilizados em diversas pesquisas de taxonomia genômica nos dias atuais, destacam-se a análise de similaridade do gene ribossomal 16S, do gene *recA* (recombinase A) e análise das sequências multilocus (MLSA), bem como os métodos de hibridização digital DNA-DNA (dDDH), identidade média de nucleotídeos e aminoácidos (ANI e AAI, respectivamente), os quais são definidos como métodos de delimitação das espécies microbianas (ANDRADE, 2016; DEPOORTER et al., 2016, 2020; JIN et al., 2020; PEETERS et al., 2013; ROMERO-GUTIÉRREZ et al., 2020).

A classificação de espécies bacterianas baseada na comparação do gene rRNA 16S, de maneira eficiente e acelerada, foi possibilitada a partir do desenvolvimento da tecnologia do DNA recombinante e dos métodos de sequencimento (KITAHARA; MIYAZAKI, 2013; WOESE; KANDLER; WHEELIS, 1990). Os principais pontos a serem destacados, que levaram esta análise (com base no gene rRNA 16S) se tornar o padrão de excelência para as análises taxonômicas são: as sequências dos genes rRNA 16S serem altamente conservadas, onipresentes em procariotos e possuirem probabilidade baixa de terem sofrido transferência horizontal de genes entre as espécies (KITAHARA; MIYAZAKI, 2013). Sabendo dessas características, variados estudos de décadas distintas, no campo da taxonomia e direcionados ao gênero *Burkholderia*, se baseiam no gene rRNA 16S com a finalidade de distinguir os diferentes membros que compõem este gênero (SAWANA; ADEOLU; GUPTA, 2014; YABUUCHI et al., 1992).

Apesar da análise do gene rRNA 16S fazer parte dos principais métodos utilizados nas análises taxonômicas, em diversos gêneros bacterianos, foi identificado que para membros do Bcc esta análise se torna limitada, portanto o gene *recA* se apresenta como um método mais eficaz para a classificação das espécies do Bcc (PAYNE et al., 2005). O gene *recA* está envolvido com a síntese da enzima recombinase A, sendo esta a estrutura proteica encarregada pelo sistema de recombinação homóloga de DNA, além de participar da transferência horizontal de

genes entre cepas distintas (AGRONOMIA; BARAÚNA, 2017; SHENG et al., 2020). Assim como o *recA* outros genes estão relacionados com a manutenção das funções vitais das células bacterianas, genes *housekeeping*.

A técnica de MLSA utiliza os genes *housekeeping*, de maneira concatenada. Estes variados fragmentos de genes que codificam proteínas, tem capacidade de fornecer uma topologia de árvore mais robusta e uma compreensão melhorada dos acontecimentos de especiação em comparação com uma árvore que se baseia nas sequências do gene rRNA 16S (LÓPEZ-HERMOSO et al., 2017; SAWABE et al., 2013). Este termo (MLSA) foi estabelecido por Gevers et al. (2005) e o método está sendo empregado progressivamente com a finalidade de obter um maior poder de resolução entre as espécies dentro de um determinado gênero (GEVERS et al., 2005; GLAESER; KÄMPFER, 2015).

Ao decorrer dos anos, houve um aumento no número de sequências dos genomas disponíveis para espécies bacterianas, incluindo membros do Bcc e assim foram surgindo novos métodos como o dDDH, ANI e AAI (JIN et al., 2020; SAWANA; ADEOLU; GUPTA, 2014).

O método de dDDH permite obter uma estimativa da similaridade entre os genomas de duas espécies microbianas e consequentemente inferir distâncias intergenômicas, através de análises *in silico* (JIN et al., 2020). Para calcular os valores de dDDH, dependendo da fórmula utilizada, a disponibilidade de sequências do genoma completo pode se tornar um aspecto limitante do método, sendo este um fator que não é exigido no DDH tradicional (ORATA, 2017). Simultaneamente a disponibilidade da sequência do genoma traz uma vantagem, visto que as sequências podem ser utilizadas para análises posteriores para constatar a identidade, assim como estudos futuros para descrições das espécies (MEIER-KOLTHOFF et al., 2013).

O método ANI, utilizado para delimitar espécies, apresenta a identidade média de nucleotídeos de genes conservados que se encontram presentes em duas cepas distintas sequenciadas, caracterizando uma medida robusta da distância genética e evolutiva entre elas (GORIS et al., 2007). Este método aplicado por Goris et al. (2007)

se baseia em um corte artificial (genoma 1) em fragmentos de aproximadamente 1.020 pb e e utilizado para buscar a sequencia de outro genoma (genoma 2) alvo, através do algoritmo Blastn (ALTSCHUL et al., 1990) e o método é então revertido (GORIS et al., 2007; ORATA, 2017). O conceito de AAI é similar ao método ANI, sua técnica se baseia na determinação da identidade média de aminoácidos dos genomas de cepas diferentes. Assim como ANI, AAI é utilizado como um meio de delimitação de espécies. Com a implementação de tais métodos genômicos no âmbito da taxonomia procariótica evidenciou diversos erros na classificação de espécies bacterianas (COLSTON et al., 2014; JESUS, 2020).

4. OBJETIVOS

4.1 OBJETIVO GERAL

Realizar a identificação e análise comparativa dos genes do T3SS em uma gama de linhagens de *B. cenocepacia* disponíveis nos bancos de dados, de variadas origens, sendo elas clínicas ou ambientais, com a finalidade de compreender melhor o papel desse sistema na interação bactéria-hospedeiro.

4.2 OBJETIVOS ESPECÍFICOS

- 1 Selecionar genomas de *B. cenocepacia*, quanto à sua origem e qualidade das sequências;
- 2 Realizar a anotação funcional dos genomas selecionados pelo *software* Prokka (v. 1.14.6);
- 3 Analisar comparativamente, utilizando o programa BLAST (ncbi-blast+ (2.9.0-2)), as sequências proteicas, resultantes da anotação dos genomas, contra o banco de dados T3Enc;
- 4 Comparar a organização gênica e similaridade do T3SS de todas os genomas selecionados;
- Analisar filogeneticamente o gene conservado da plataforma de classificação do T3SS, *sctN*, empregando-se o MEGA X;
- 6 Analisar o grupo *B. cenocepacia* através da taxonomia genômica, utilizando os genes rRNA 16S, *recA* e a partir da análise das sequências multilocus (MLSA);
- 7 Investigar a diversidade taxonômica das cepas clínicas e ambientais, através da hibridização digital DNA-DNA (dDDH), identidade média de nucleotídeos e proteínas (ANI, AAI).

5. MATERIAIS E MÉTODOS

5.1 SELEÇÃO E ANOTAÇÃO DOS GENOMAS BACTERIANOS

Os 50 genomas de *B. cenocepacia*, acessíveis publicamente no momento do estudo, foram selecionados de acordo a sua origem, clínica e ambiental, excluindo os genomas de origem desconhecida e o banco de dados utilizado foi o GenBank (BENSON et al., 2013), sendo este um banco de dados de sequência genética do National Institutes of Health (NIH). Para a seleção pela qualidade foram priorizados os genomas completos e contigs ou *scaffolds* abaixo de 200, além disso os métodos de sequenciamento por Illumina e/ou PacBio também foram priorizados. A anotação funcional dos genomas selecionados foi efetuado pela ferramenta Prokka (v. 1.14.6) (SEEMANN, 2014).

5.2 ANÁLISE COMPARATIVA

A análise comparativa foi realizada através do mapeamento contra o banco de dados T3Enc (HU et al., 2017). A categorização e reanotação dos genes do T3SS, para a construção do T3Enc, foram realizados com base em uma estratégia combinatória da análise de sequências filogenéticas e organização do Bloco de Microssintenia (MSB). MSB são pequenas regiões de sintenia identificadas entre duas ou mais regiões genômicas, comparando um baixo número de genes ortólogos (JOSHI; CHAUHAN; DAS, 2018). Os autores deste banco de dados produziram para os 36 genes do T3SSs bem conhecidos, o alinhamento das sequências de proteínas individuais Sct utilizando o ClustalW (LARKIN et al., 2007), Neighbour-joining e árvores de máxima parcimônia, empregando o MEGA 6 (TAMURA et al., 2013), com 1000 réplicas de bootstrap. Para identificar os genes do T3SS nos genomas selecionados, foi utilizado a ferramenta BLASTp (ALTSCHUL et al., 1990) usando os parâmetros de 40% de similaridade e evalue 0.001. Os melhores hits foram filtrados e tabulados para análises posteriores. A ferramenta de visualização Easyfig (v. 2.2.2) (SULLIVAN; PETTY; BEATSON, 2011) foi utilizada para visualizar os *clusters* gênicos do T3SS presentes nos genomas. Os arquivos com a extensão .gbk obtidos da anotação genômica, os quais possuem

informações essenciais como os nomes e as sequências proteicas dos genes, foram empregados como *input* na ferramenta.

5.3 ANÁLISE FILOGENÉTICA E TAXONÔMICA

5.3.1 Taxonomia genômica

5.3.1.1 Genes rRNA 16S, recA e método MLSA

Com base na homologia da sequência rRNA 16S, viabiliza-se a identificação de espécies a partir da ferramenta comparativa BLAST (JIN et al., 2020; RAJENDHRAN; GUNASEKARAN, 2011), assim como na delimitação de cepas dentro de uma mesma espécie. Essa delimitação se dá pela presença de múltiplas cópias do operon rRNA e a heterogeneidade intra-genômica dos genes rRNA 16S (RAJENDHRAN; GUNASEKARAN, 2011). A sequência do comprimento total do gene rRNA 16S, de cada genoma analisado, foi gerada utilizando a ferramenta barrnap (termo em inglês, Basic Rapid Ribosomal RNA Predictor, https://github.com/tseemann/barrnap) (v. 0.9) (SEEMANN, 2013). As sequências do gene rRNA 16S utilizadas continham um tamanho médio de 1529 bp.

O gene *recombinase A* (*recA*), foi sequenciado amplamente em espécies bacterianas (ROCA; COX, 1997) e através deste gene *B. cenocepacia* se mostrou ser geneticamente heterogênea, como já foi descrito no capítulo 3.1 (página 16). Com isso, este gene foi selecionado e as sequências de nucleotídeos encontradas no arquivo de extensão .ffn, gerado a partir da anotação pelo *software* Prokka (v. 1.14.6), foram alinhadas e direcionadas para construir a árvore taxonômica com base no *recA*.

Estudos no campo da análise taxonômica utilizam a análise das sequências multilocus (MLSA), com a finalidade de discriminar de maneira eficiente espécies bacterianas pertencentes ao Bcc, (ANSARI et al., 2019; BAIA et al., 2021; ESTRADA-DE LOS SANTOS et al., 2013; GAUTAM et al., 2016; JIN et al., 2020). Baseando-se nesses estudos citados anteriormente, para o atual trabalho, foram selecionados os seguintes genes *housekeeping*:

1. atpD - *ATP* synthase subunit beta 1;
- 2. gltB ferredoxin-dependent glutamate synthase 1;
- 3. gyrB DNA gyrase subunit B;
- 4. recA recombinase A;
- 5. *lepA elongation factor 4;*
- 6. *phaC poly (3-hydroxyalkanoate) polymerase subunit PhaC;*
- 7. *trpB tryptophan synthase beta chain*.

Os sete genes *housekeeping* foram concatenados para realizar o alinhamento múltiplo e posteriormente gerar a árvore taxonômica baseada no MLSA.

5.3.1.2 Métodos dDDH, ANI e AAI

A hibridização digital DNA-DNA (dDDH), é uma abordagem que possibilita determinar, de maneira computacional, a distância entre dois genomas e estimar se ambos são da mesma espécie ou de espécies distintas (AUCH et al., 2010). Para mensurar os valores de dDDH dos genomas analisados foi utilizado o *web service* GGDC (termo em inglês, *Genome-to-Genome Distance Calculator*; <u>http://ggdc.dsmz.de/ggdc.php#</u>) (MEIER-KOLTHOFF et al., 2013) usando a fórmula 2 e o BLAST+ como ferramenta de alinhamento local.

A fórmula 2 se baseia no alinhamento de um genoma contra o outro e mutuamente, além de produzir pares de HSPs (Pares de Segmento de Alta Pontuação), assim a distância entre os genomas é definida dividindo a soma de todas as identidades encontradas em HSPs pelo comprimento total de HSPs (YAMASSAKI PEREIRA, 2017). Dessa forma, foi possível obter estimativas dos valores de dDDH dos genomas completos e incompletos, visto que o cálculo da distância não depende do comprimento do genoma. Os valores <70% foram considerados como uma indicação que os organismos testados são de espécies diferentes (AUCH et al., 2010).

Além do dDDH, nas análises taxonômicas foram utilizados os métodos de ANI e AAI, que se referem a identidade média de nucleotídeos e a identidade média de aminoácidos respectivamente, esses dois parâmetros resultam através da comparação de genomas calculando a média da identidades das sequências de genes ortólogos. Dessa forma possibilitou direcionar uma análise baseada nos genomas e suas relações evolutivas. O programa GenTaxo foi utilizado para calcular ambas as identidades, ANI e AAI, um programa escrito em Perl, o qual é disponível gratuitamente em <u>sourceforge.net/projects/gentaxo/</u>. Nesta ferramenta foi aplicado como *input* as sequências de nucleotídeos e aminoácidos em formato FASTA de todas as cepas aqui analisadas. As cepas que compartilhavam mais de 95% de identidade (ANI e AAI), foram consideradas como da mesma espécie (JAIN et al., 2018; MEDLAR; TÖRÖNEN; HOLM, 2018).

5.3.2 Alinhamento múltiplo

Os alinhamentos múltiplos dos genes analisados neste estudo foram realizados pelo *software online* Clustal Omega (MADEIRA et al., 2019; SIEVERS et al., 2011). O Clustal Omega permite que alinhamentos de até 4000 sequências ou um tamanho máximo de arquivo de 4 MB, sendo este um dos métodos de alinhamento múltiplo mais rápido e preciso. Na plataforma do Clustal Omega foi utilizado ClustalW como formato de *output*.

5.3.3 Árvores filogenéticas e taxonômicas

Após a etapa de alinhamento, as análises filogenéticas foram conduzidas pelo *software* MEGA (v. X) (KUMAR et al., 2018), sendo este um programa com um conjunto de ferramentas para inferir árvores evolutivas, estimar distância e diversidade genéticas, inferir sequências ancestrais, entre outros métodos. Este programa possibilitou construir as árvores filogenéticas deste trabalho, através do método de *Maximum Likelihood* com *bootstrap* de valor 1500, que nos permitiu observar as semelhanças e divergências entre os genomas e suas origens. A escolha deste método se deu por possibilitar calcular estimativas de menor variância, logo com menos erros de amostragem (CALDART et al., 2018). Os modelos de substituição de nucleotídeos utilizados foram o *Tamura-Nei* (rRNA 16S e *recA*), *Tamura 3-parameter (sctN*) e *General Time Reversible* (MLSA). Para visualizar as árvores taxonômicas e

filogenética, foi utilizada a plataforma *online* iTOL (v. 5) (LETUNIC; BORK, 2021), aplicando os arquivos em formato Newick para o *upload*.

5.3.4 Rstudio

Para as análises taxonômicas ANI, AAI e dDDH foi gerado heatmaps dos seus respectivos valores, utilizando o *software* RStudio (v. 1.4.1717) (<u>http://www.rstudio.com/</u>) que se baseia na programação em R. Este *software* apresenta variados pacotes e para o atual trabalho os pacotes ComplexHeatmap (v. 2.9.1) e circlize (v. 0.4.13) foram utilizados.

O pacote ComplexHeatmap é eficiente para visualizar associações entre distintas fontes de conjunto de dados e revelar os potenciais padrões (GU; EILS; SCHLESNER, 2016). O pacote circlize fornece aos usuários uma maior liberdade para projetar figuras para melhor compreender os padrões genômicos por trás (GU et al., 2014), como por exemplo editar a escala de cores de cada heatmap, permitindo que os gráficos sejam mais atrativos e melhor compreendidos.

Os métodos de distância e de clusterização utilizados foram *euclidean* e *average*, respectivamente. Com o método de distância *euclidean* foi possível minimizar a distância entre pontos dentro de *clusters* e maximizar a distância para pontos de diferentes *clusters*. O método de clusterização *average* foi aplicado com o objetivo de agrupar as cepas de acordo com a distância média entre cada um de seus membros.

6. RESULTADOS E DISCUSSÃO

6.1 CEPAS BACTERIANAS E ANOTAÇÃO GENÔMICA

No período de busca (agosto/2020) dos genomas bacterianos para serem analisados, haviam 329 genomas de *B. cenocepacia*, incluindo *drafts* e genomas completos. Dentre todos os genomas listados, foi realizado um filtro a fim de buscar cepas isoladas de origem clínica ou ambiental. As cepas de origem ambiental selecionadas foram isoladas de distintos ambientes – solo, aerossol, água, rizosfera e raiz do milho. As diversas origens clínicas das cepas bacterianas selecionadas englobam: sangue, escarro, sangue do cordão umbilical, fonte nasal e isolado clínico.

Dos 50 genomas selecionados foram encontrados 16 genomas completos, 3 cromossomos (que podem ser considerados genomas completos) e 31 *drafts*. Esses dados são apresentados na **Tabela S1**, do documento em anexo, além de outros dados genômicos.

A partir da etapa de filtragem, foi realizada a anotação de 50 genomas selecionados através do *software* Prokka (v. 1.14.6), possibilitando obter algumas informações como os números de RNA transportador (tRNA), CRISPR e RNA transportador-mensageiro (tmRNA).

Na **Tabela S1** identifica-se que os genomas de origem ambiental e clínica apresentam um valor médio de tRNA de 74,12 e 82,36, respectivamente, indicando uma diferença significativa (Teste-t Student, p = 6.86E-04). A média do percentual de Guanina-Citosina (GC%) dos genomas ambientais é de 66.85% e dos genomas de origem clínica é de 66.99%, corroborando com a correlação positiva estabelecida para procariotos, segundo Satapathy *et al.* (2010). Além disso, neste mesmo trabalho foi relatado que a redundância do gene tRNA está correlacionada com o tamanho do genoma, que será discutido no capítulo 6.2 (SATAPATHY; DUTTA; RAY, 2010).

O genoma da cepa bacteriana DDS 22E-1 é o único, dentre os genomas analisados, que possui um sistema CRISPR, porém esta cepa não expressa um tamanho distinto das cepas analisadas. Segundo Gao *et al* (2019), os genomas que possuem CRISPR, apesar de por vezes terem incorporações de sequências espaçadoras nas matrizes CRISPR, não têm um alto impacto no tamanho total do genoma (GAO et al., 2019), corroborando com o resultado encontrado.

6.2 ORGANIZAÇÃO E COMPARAÇÃO GENÔMICA

Com a finalidade de analisar a variabilidade da organização genômica entre os genomas de diferentes origens, foi comparada as médias do tamanho do genoma inteiro (Mb), das sequências codificantes (CDS), e do conteúdo percentual de GC (GC%) para as cepas de origem ambiental e clínica e a média. Na **Figura 6** podemos observar que as cepas de origem ambiental apresentaram um tamanho médio de 7.65 Mb, enquanto as cepas clínicas exibiram um tamanho médio de 7.77 Mb, demonstrando que não há

uma diferença significativa (Teste-t Student, p = 3.55E-01). O número médio de CDS foi de 6.803 e 6.981, para as cepas de origem ambiental e clínica respectivamente, indicando que não existe uma diferença significativa (Teste-t Student, p = 1.68E-01) entre os genomas de diferentes origens, esse resultado é retratado na **Figura 7**. A **Figura 8** apresenta a média do conteúdo de GC% para os genomas de origem ambiental (66.85%) e para os de origem clínica (66.99%), não exibindo uma diferença significativa (Teste-t Student, p = 1.14E-01).



Figura 6: Tamanho dos genomas (Mb) dos isolados de origem ambiental e clínica. O tamanho médio dos genomas para as cepas isoladas de fonte ambiental é 7.65 Mb e para as cepas de fonte clínica é 7.77 Mb. Não apresentam diferença significativa, visto que exibem um valor de de p = 3.55E-01.



Figura 7: Número de sequências codificantes (CDS) dos isolados de origem ambiental e clínica. O número médio de CDS dos genomas isolados de fonte ambiental é 6.803 e para os genomas de fonte clínica é 6.981. O valor de p = 1.68E-01 indica que o número médio de CDS não apresentam diferença significativa.



Figura 8: Percentual do conteúdo de Guanina-Citosina (GC%) dos genomas de origem ambiental e clínica. A média do GC% para os genomas de ambas as origens, ambiental ou clínica, são 66.85% e 66.99%, respectivamente. A partir do valor p = 1.14E-01 conclui-se que não há diferença significativa entre as origens.

A patogenicidade e a adaptação bacteriana estão consistentemente associadas ao tamanho do genoma, a partir de uma proporção direta para a adaptação e inversa para a patogenicidade (MURRAY et al., 2021; OLIVEIRA et al., 2017a). O comprimento médio do tamanho do genoma também se relaciona com o conteúdo médio de GC em genomas procarióticos: quanto maior o tamanho, maior será a taxa percentual de GC em determinado genoma (ALMPANIS et al., 2018). E o conteúdo de GC se relaciona com o número de CDS (OLIVER; MARÍN, 1996), além de ser altamente variado nos organismos celulares (13% a 75%) (MAHAJAN; AGASHE, 2021). A composição genômica se torna característica importante que contribui para a aptidão dos organismos (LASSALLE et al., 2015). Diante dessas informações, os resultados denotam que as cepas analisadas, de ambas as origens, são capazes de se adaptarem ao meio.

Arquivos de extensão .gff (*General Feature Format* – GFF) e .gbk (GenBank *Format*) foram originados a partir da anotação genômica realizada no *software* Prokka (v. 1.14.6). O arquivo .gff é um arquivo de texto simples delimitado por tabulação que descreve características genômicas, enquanto que o .gbk é um arquivo de texto que armazena as informações do genoma descritas .gff . Ambos os arquivos foram

utilizados para contruir as figuras dos *clusters* genômicos e portanto possibilitar a comparação genômica.

Com a plataforma EasyFig (v.2.2.2) foi possível identificar e elaborar 9 *clusters* distintos dos genomas analisados, que são apresentados na **Figura 9**. Nos *clusters* é constatada a presença dos genes que codificam as proteínas que são respónsáveis pela estrutura do aparato do T3SS (nomenclatura *sct*), como também a existência do gene *asnB*.



Figura 9: *Clusters* gênicos do Sistema de Secreção do tipo III (T3SS) de *B. cenocepacia*. Os 9 *clusters* e os genes que os constui são ilustrados. Os genes *sctQLK*, que codificam as proteínas que constituem a plataforma de classificação, são representados em amarelo. A maquinaria da membrana interna é composta pelas proteínas codificadas pelo genes *sctRSTUV*, que são correspondidos pela cor verde. O gene *sctN*, que codifica a ATPase do T3SS, é retratado na cor azul. Em violeta, os genes representados são o *sctCDJ*, que compõe o corpo basal. O gene *asnB*, que codifica a Asparagina sintetase, corresponde à cor cinza enquanto os genes que codificam proteínas identificadas como hipotéticas são representadas pela cor vermelha.

A **Tabela 2** permite identificar a quantidade de cepas que possui apenas 1 *cluster*, 2 *clusters* ou nenhum *cluster* do T3SS. As cepas que apresentaram somente um *cluster* do T3SS são de ambas as origens, ambiental ou clínica; as que exibiram 2 *clusters* são apenas cepas de origem ambiental; a cepa que não apresentou nenhum *cluster* (*B. cenocepacia* VC9789) é de origem clínica. Este resultado possibilita inferir que o T3SS não é essencial para mediar com o hospedeiro humano, mas sim com o hospedeiro vegetal.

Quantidade de clusters	Quantidade de cepas	Origem
Nenhum cluster	1	Clínica
1 cluster	39	Ambiental e Clínica
2 clusters	10	Ambiental

Tabela 2: Número de clusters do T3SS encontrados nos genomas selecionados de B. cenocepacia.

O gene *asnB*, codifica a Asparagina sintetase está envolvido com a metabolização da alanina, aspartato e glutamato, e é definido na literatura como um gene relacionado à virulência regulado pelo fator de sinal difusível (DSF) e pelo regulador global Clp (Receptor proteína-*like* cAMP) (QIAN et al., 2013; RYAN et al., 2015). Qian *et al* (2013) demonstraram que o *asnB* é altamente conservado o gênero bacteriano *Xanthomonas*, que engloba espécies fitopatogênicas, e que cepas bacterianas mutantes ($\Delta asnB$) não apresentavam a mesma capacidade de virulência e taxa de crescimento no hospedeiro quando comparados às cepas selvagens, além de prejudicar a resistência do patógeno ao estresse oxidativo (QIAN et al., 2013). Com isso, entende-se que a presença em desde gene reitera o potencial de virulência e de colonização das cepas bacterianas de *B. cenocepacia*.

Outro gene que requer devida atenção é uma CDS que codifica uma proteína hipotética (*hypothetical protein*), que possui o maior tamanho (cerca de 6.637 pares de bases) dentre todos os genes identificado nos *clusters* 1, 2, 3 e 4 dos T3SS, representada em vermelho, na **Figura 9**. Na **Tabela S2** é descrito o *locus tag* de cada cepa que possui o gene *hp* (*hypothetical protein*). Foi realizado um BLASTp na plataforma do NCBI a fim de encontrar domínios conservados desta proteína hipotética

e investigar sua determinada função. Os domínios encontrados são relatados na **Tabela 3** abaixo.

Nome	e-value	Intervalo
HopBF1	3.19E-34	1851-2015
HopBF1 super família	1.43E-15	1406-1568
DUF4781 super família	8.00E-10	752-908
Domínio de exclusão de superfície SEC10 / PgrA	2.90E-09	101-207

Tabela 3: Domínios conservados presentes na proteína conservada.

HopBF1, HopBF1 super família, DUF4781 super família e Domínio de exclusão de superfície SEC10 / PgrA são os domínios conservados encontrados na proteína analisada. O *e-value* para cada domínio também é relatado.

Os três principais domínios HopBF1, DUF4781 e Domínio de exclusão de superficie SEC10 / PgrA, apresentam e-value de 3.19E-34, 8.00E-10 e 2.90E-09 respectivamente, indicando uma alta confiabilidade nos resultados apresentados. O domínio HopBF1 é identificado em Pseudomonas syringae (patógeno de plantas) (CHEN; LIU; FU, 2019) e seus homólogos são encontrados em cepas dos gêneros Enterobacter, Klebsiella, Ewingella, Serratia, Burkholderia e Ralstonia, englobando cepas que são patógenos ou simbiontes animais e vegetais, assim como de vida livre (LOPEZ et al., 2019). Este domínio é responsável pela inativação de uma chaperona essencial (Heat Shock Protein - HSP90) necessária para a imunidade inata do hospedeiro, induzindo necrose e colapso no tecido das folhas das plantas (LOPEZ et al., 2019). O domínio DUF4781 apesar de ser extensamente encontrado em eucariotos e bactérias, este domínio não é caracterizado funcionalmente (BATEMAN; COGGILL; FINN, 2010), não havendo evidências suficientes para compreender a função da proteína conservada. O Domínio de exclusão de superfície SEC10 / PgrA é descrito como um domínio conservado, encontrado em proteínas de superfície de diversos membros do filo Firmicutes e está envolvido na inibição específica da capacidade das células bacterianas de receber plasmídeos homólogos, sendo este fenômeno conhecido como exclusão de superfície (KAO et al., 1991).

A proteína hipotética conservada é identificada, em *B. cenocepacia* J2315, como CDS BCAM2043 – *peptidoglycan-binding protein*, pseudogene – na literatura (WINSOR et al., 2008) (Imagem do gene *hp* em *B. cenocepacia* J2315 - *Burkholderia* *Genome Database*), além de possuir uma mutação *frameshift (HOLDEN et al., 2009; TOLMAN, 2011)*, porém sua função atualmente é desconhecida. Entretanto com esses achados, relacionado aos domínios conservados, podemos inferir que a proteína conservada poderia estar relacionada com a patogênese, empregando o T3SS para secretar os efetores bioquímicos capazes de modular a hipersensibilidade do sistema imune do hospedeiro.

Na **Figura 9** percebe-se que há *clusters* que não possuem determinados genes. O *cluster* 4, que agrupa 6 cepas *B. cenocepacia* de origem clínica, não apresenta o gene *sctQ*; o *cluster 5*, referente ao *cluster* do T3SS encontrado exclusivamente no genoma da cepa GIMC4560:Bcn122 (origem clínica), não possui o gene *sctJ*; o *cluster* 8, que reúne 9 cepas de origem ambiental, não abriga os genes *sctKL*; o *cluster* 9, encontrado unicamente no genoma da cepa MC0-3 (origem ambiental), tem ausente os genes *sctKLQ*. Sabendo que o *sctJ* é responsável por codificar a proteína SctJ constitui o anel de membrana interna, tendo como função direcionar o transporte proteico (KUHLEN et al., 2018), e que os genes *sctKLQ*, que codificam proteínas que estão envolvidas com o recrutamento hierárquico dos substratos do aparato do T3SS, pode-se concluir que ambas as funções se tornam comprometidas caso esses genes estejam ausentes no *cluster* gênico do T3SS.

6.3 ANÁLISE FILOGENÉTICA

O gene *sctN*, como já foi dito neste trabalho, codifica a ATPase da maquinaria de secreção do T3SS, tornando-se essencial para a função desse aparato secretor. Além disso, o gene *sctN* faz parte do *core* T3SS, portanto amplamente utilizado para análises filogenéticas em diferentes estudos (HU et al., 2017; TEULET et al., 2020; TROISFONTAINES; CORNELIS, 2005). Portanto para a análise filogenética, do atual trabalho, e consequentemente ter uma melhor compreensão sobre o T3SS, o gene *sctN* foi selecionado.

Na Figura 10 é apresentada a árvore filogenética gerada pelo programa iTOL (v. 5), utilizando como *input* o gene *sctN* dos genomas analisados. Nesta árvore é



exibida, além da organização filogenética, a disposição das cepas em relação ao seu respectivo *cluster*.

Figura 10: Árvore filogenética com base no gene *sctN* dos genomas analisados. A construção da árvore foi realizada pelo MEGA (v. X), utilizando o valor de *booststrap* de 1500, através do método *Maximum Likelihood* e do modelo *Tamura 3-parameter*. Os nomes das cepas de *B. cenocepacia* coloridas pela cor verde significa que determinada cepa é de origem ambiental e os nomes das cepas com a cor vermelha expressa que a cepa é de origem clínica. Cada cepa se inclui em pelo menos um *cluster*, as cepas que estão repetidas são as cepas que possuem dois *clusters* do T3SS. As cepas que não estão incluídas em nenhum clado possuem o ramo na cor preta.

Como podemos observar, no primeiro clado (rosa) são identificados os *clusters* 1, 2, 4, 5 e 6, apesar deste clado reunir em grande parte cepas de origem clínica identifica-se 4 cepas de origem ambiental (Bp9134, Bp9158, Bp9139 e MSMB384WGS). Para compreender o porquê das quatro cepas ambientais estar presente neste clado, foi realizado um Blastn da sequência nucleotídica do gene *sctN* dessas quatro cepas, na plataforma do NCBI. O *sctN* das quatro cepas ambientais deste clado pertence ao *cluster* 2 e possui um alto percentual de identidade ao *sctN* (cerca de 95%) de *Burkholderia metallica*, sendo esta uma espécie patogênica oportunista e isolada de pacientes fibrocísticos (VANLAERA et al., 2008), denotando que as quatro cepas ambientais são filogeneticamente idênticas às cepas de origem clínica.

No segundo clado (amarelo) a maioria das cepas são de origem ambiental e contém o *cluster* 2, com exceção de cinco cepas que são de origem clínica e uma delas (PC184 Mulks) que contém *cluster* 3. Essas cinco cepas se agrupam neste clado com as ambientais não somente nesta árvore baseada no gene *sctN*, elas se agupam em um mesmo clado de árvores construídas a partir do gene ribossomal 16S, *recA* e nos genes *housekeeping* (dados discutidos no tópicos seguinte). Esses dados sugerem que as cepas VC7848, AU 1054, FDAARGOS 82, VC12802 e PC184 Mulks, possuem características genômicas que estão majoritariamente presentes em genomas de cepas de origem ambiental.

O terceiro clado (azul) apresenta apenas cepas de origem ambiental e o *cluster* 2. É válido ressaltar que todas as cepas presentes neste clado possuem 2 *clusters* e dessa forma essas mesmas cepas se reúnem no quarto clado (violeta) acompanhadas de duas outras cepas de origem ambiental (DDS 22E-1 e MC0-3). Este último clado pertencem majoriataramente ao *cluster* 8 e 9 (apenas para a MC0-3). O *cluster* 7 que inclui a cepa DDS 22E-1 forma um *outgroup* com a cepa DWS 37 E-2, denotando a distância filogenética perante as outras cepas aqui analisadas.

6.4 TAXONOMIA GENÔMICA

Os genes rRNA 16S, *recA*, bem como sete genes *housekeeping* - MLSA (*atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC* e *trpB*) foram utilizados para as análises taxonômicas.

O rRNA 16S foi selecionado por ser um gene amplamente aceito, em diversos estudos, como uma impressão digital biológica para as espécies bacterianas (ESCOBAR-ZEPEDA et al., 2018). Mas após a árvore taxonômica gerada não apresentar valores de *booststrap* suportados, corroborando com dados prévios da

literatura (MAHENTHIRALINGAM et al., 2000; PAYNE et al., 2005; VANDAMME; DAWYNDT, 2011), como é demonstrado da **Figura 11**.



Figura 11: Árvore taxonômica baseada no gene rRNA 16S. A construção da árvore foi realizada pelo MEGA (v. X), utilizando o valor de *booststrap* de 1500, através do método *Maximum Likelihood* e do modelo *Tamura-Nei*. Cepas de origem ambiental são nomeadas em cor verde e as cepas de origem clínica em cor vermelha. Três clados são formados, o primeiro clado (rosa) agrupa na sua maioria cepas clínicas, o segundo (violeta) possui a mesma proporção de cepas de ambas as origens e o terceiro clado (amarelo) reúne na sua grande parte as cepas de origem ambiental. As cepas que possuem o ramo na cor preta aponta que estas não se inserem dentro de um clado.

Esta limitação se dá principalmente no gênero *Burkholderia*, pois não há diferenças significativas em seus genes rRNA 16S e assim impossibilita distinguir com precisão as espécies (ESCOBAR-ZEPEDA et al., 2018; PAYNE et al., 2005). Destacase a existência de três cepas ambientais (Bp9023, DWS 37E-2 e DDS 22E-1) que não se inserem em nenhum dos três clados (rosa, violeta e amarelo). Estes ramos independentes e distintos sugerem a necessidade de divisão da classificação taxonômica atual de B. cenocepacia (JIN et al., 2020), corroborando com os resultados encontrados em trabalhos na literatura (JIN et al., 2020; WALLNER et al., 2019). O fato dos valores de bootstrap serem muito baixos, incentivou a busca de outro gene para que permitisse discenir as cepas da espécie B. cenocepacia, selecionando-se portanto o gene recA. O gene recA pode ser encontrado em diversas espécies bacterianas, visto que exerce funções importantíssimas nos mecanismos de reparo de dano no DNA, por exemplo (SEPE et al., 2008). Este gene, além de ser extensivamente aplicado na sistemática bacteriana, tem se apresentado extremamente útil para a identificação de determinadas espécies, baseando-se na variação da sequência dentro do gene (LANDETA et al., 2011; SEPE et al., 2008). Foi com base neste gene que possibilitou a criação de quatro subgrupos filogenéticos em B. cepacia, como já foi descrito no capítulo 3.1 (página 16), demonstrando o seu poder discriminatório. Porém, este gene exibe alta similaridade, cerca de 94% a 95% entre as espécies do Bcc e 98% a 99% entre as cepas das espécies desse complexo (BAIA et al., 2021; VANDAMME; DAWYNDT, 2011), além das árvores baseadas neste gene exibirem confusão e discordância em relação ao seus clados (JIN et al., 2020), explicando assim os baixos valores de *bootstrap* apresentados na Figura 12.



Figura 12: Árvore taxonômica baseada no gene *recA*. A construção da árvore foi realizada pelo MEGA (v. X), utilizando o valor de *booststrap* de 1500, através do método *Maximum Likelihood* e do modelo *Tamura-Nei*. As cepas identificadas na cor verde indicam ser de origem ambiental e as cepas na cor vermelha são de origem clínica. Três clados são formados, o primeiro clado (rosa) agrupa na sua maioria cepas clínicas, o segundo clado (violeta) possui quatro cepas de origem clínica e duas de origem ambiental e o terceiro clado (amarelo) reúne na sua grande parte as cepas de origem ambiental. Os dois *outgroups* são identificados com o ramo na cor preta.

Apesar da árvore baseada em *recA* ter semelhanças interessantes com a árvore taxonônomica baseada no gene 16S rRNA, como por exemplo o fato de ambas agruparem as cinco cepas clínicas VC7848, AU 1054, FDAARGOS 82, VC12802 e PC184 Mulks com as cepas de origem ambiental, não é possível diferenciar as cepas umas das outras pelo alto grau de similaridade do gene *recA* entre elas, como já foi citado. Uma alternativa para analisar as cepas de *B. cenocepacia* é utilizar o método MLSA. O método MLSA abrange fragmentos internos de vários genes codificadores de proteínas (genes *housekeeping*), os quais evoluem de maneira constante e em uma taxa

mais lenta do que o gene rRNA 16S (GLAESER; KÄMPFER, 2015), apresentando uma melhor discriminação em nível de gênero ou níveis taxonômicos mais baixos. Acredita-se que as árvores baseadas em MLSA refletem uma relação mais fidedigna dos táxons bacterianos (ANSARI et al., 2019; ESTRADA-DE LOS SANTOS et al., 2013). Sabendo do alto poder discriminatório para a diferenciação de espécies do Bcc (BAIA et al., 2021) e que o método MLSA é empregado amplamente em estudos na literatura (ANSARI et al., 2019; BAIA et al., 2021; ESTRADA-DE LOS SANTOS et al., 2013; GAUTAM et al., 2016; JIN et al., 2020; PEETERS et al., 2013), os genes *housekeeping – recA, gyrB, atpD, gltB, lepA, phaC* e *trpB –* foram utilizados como base para a construção de uma nova árvore taxonômica, a qual é apresentada na **Figura 13**



Figura 13: Árvore taxonômica baseada nos sete genes housekeeping (recA, gyrB, atpD, gltB, lepA, phaC e trpB). A construção da árvore foi realizada pelo MEGA (v. X), utilizando o valor de booststrap de 1500, através do método Maximum Likelihood e do modelo General Time Reversible. As cepas que se apresentam com a cor verde são de origem ambiental e as cepas com a cor vermelha são de origem clínica. Três clados são construídos, o primeiro clado (amarelo) agrupa na sua maioria cepas de origem ambiental contendo cinco cepas de origem clínica, o segundo clado (violeta) reúne somente cepas de origem ambiental e o terceiro clado (rosa) possui na sua maior parte as cepas de origem clínica incluindo apenas quatro cepas de origem ambiental. As cepas que não são encontradas em nenhum dos três clados são identificadas com o ramo na cor preta.

A árvore apresentada na **Figura 13** exibe valores de *bootstrap* mais elevados do que as da **Figura 11** e **12**, indicando uma maior confiança neste resultado e possibilitando diferenciar umas cepas das outras. Nesta árvore taxonômica formaram-se três clados distintos que são apontados nas cores amarela (primeiro clado), violeta (segundo clado) e rosa (terceiro clado).

No primeiro clado foram agrupadas cepas de origem ambiental, exceto cinco cepas (VC7848, AU 1054, FDAARGOS 82, VC12802 e PC184 Mulks), que se agruparam em clados que englobam cepas ambientais nas árvores construídas baseadas

no gene sctN, rRNA 16S e recA. Na literatura, é descrito que as cepas VC7848, AU 1054, VC12802 e PC184 Mulks carregam genes que estão envolvidos na aptidão das bactérias associadas a plantas (WALLNER et al., 2019). Estes genes, segundo Wallner et al. (2019) são: bacteriocina lectina-like 88 (llpA), nitrila hidratase subunidade alfa (nthA), nitrila hidratase subunidade beta (nthB), fenilacetaldoxima desidratase (oxd), feruloil-esterase (faeB), cluster de biossíntese de pirrolnitrina (prn), altronato desidratase (uxaA) e altronato oxidoredutase (uxaB). Neste mesmo trabalho é relatado que a cepa VC7848 possui os genes nthAB, oxd, faeB e uxaAB; a cepa AU 1054 contém os genes *llpA*, *nthAB*, *oxd*, *faeB* e *uxaAB*; a cepa VC12802 apresenta os genes *llpA*, nthAB, oxd, faeB, prn e uxaAB; a cepa PC184 Mulks carrega os genes nthAB e prn. A cepa de origem clínica que não foi analisada no trabalho de Wallner et al (2019), FDAARGOS 82, foi investigada no atual trabalho em relação a presença de genes associados à virulência ou à aptidão bacteriana relacionada a plantas. Dentre os genes relacionados à virulência (cblA, adhA, BCESM (Burkholderia cenocepacia Epidemic Strain Marker), kdgR (Regulador transcricional), baiE (ácido biliar 7-alfa desidratase), tauX (taurina desidrogenase), xsc (sulfoacetaldeído acetiltransferase), telA (cluster de resistência de telurito), lxa (locus ativado com baixo oxigênio) e cluster respiratório de nitrato redutase (narIJHGK) e cluster de sensor e regulação de nitrato (narLX)) foi encontrado apenas o gene kdgR, porém nesta cepa foram identificados cinco genes estabelecidos como genes referentes à aptidão bacteriana na associação a plantas (nthA, nthB, oxd, uxaA e uxaB). Os produtos proteicos codificados por nthAB, oxd e faeB, promovem o crescimento da planta que está associada à bactéria, visto que as proteínas codificadas por esses genes estão envolvidas na conversão de fontes de carbono (ácido galacturônico, xilanas e pectina) ao fitohormônio auxina ácido indol-3-acético (IAA) (ETESAMI; ALIKHANI; HOSSEINI, 2015; HOWDEN et al., 2009; LIU et al., 2017b; SPAEPEN; VANDERLEYDEN; REMANS, 2007; WALLNER et al., 2019). A partir disso a planta realiza o processo de rizodeposição (fornecimento de fontes de carbono) que pode gerar aumento da biomassa microbiana, através das proteínas codificadas por uxaAB (LIMA, 2014; SUVOROVA et al., 2011; WALLNER et al., 2019). Esses achados elucidam a presença dessas cinco cepas clínicas neste clado e a relação que têm com as cepas de origem ambiental.

O segundo clado, composto apenas por cepas de origem ambiental, sugere que há consistência em relação às origens das cepas identificadas.

O terceiro clado é constuído por cepas de origem clínica, na sua maior parte, porém quatro cepas de origem ambiental (Bp9139, MSMB384WGS, Bp9134 e Bp9158) foram detectadas. A cepa ambiental MSMB384WGS carreia genes que estão envolvidos com a virulência (*adhA, kdgR, baiE, tauX, xsc, telA, narLX* e *narIJHGK*) (WALLNER et al., 2019) e as cepas Bp9139, Bp9134 e Bp9158, que são analisadas somente no atual estudo, contêm os genes *kdgR*, ácido biliar 7-alfa desidratase (*baiE*), sulfoacetaldeído acetiltransferase (xsc) e *cluster* respiratório de nitrato redutase (*narIJHGK*), genes que promovem a virulência no hospedeiro. A partir da identificação da presença desses genes nas cepas de origem ambiental, indica a semelhança das características que são encontradas nas cepas clínicas, esclarecendo o motivo destas cepas serem reunidas neste terceiro clado.

Com o objetivo de investigar a identidade entre os genomas e apoiar os resultados encontrados nas árvores taxonômicas e filogenética, foi realizada uma comparação entre os valores dDDH, ANI e AAI, que pode ser observada na **Figura 14**.



Figura 14: Heatmaps dos valores de dDDH, ANI e AAI, a partir de comparações de genoma-genoma e identidade média de nucleotídeos e aminoácidos. Três *clusters* são formados em cada heatmap. Os limiares para distinguir as espécies são >95% e >70% para ANI/AAI e dDDH, respectivamente. <u>Ver imagem.</u>

Nos *heatmaps* taxonômicos apresentados na **Figura 14** identificam-se três clados e um *outgroup* (DDS 22E-1 e DWS 37E-2), que agrupam as mesmas cepas em cada clado, do mesmo modo que a árvore taxonômica baseada nos genes *housekeeping*. Observando os três heatmaps (dDDH, ANI e AAI) verifica-se três principais clados. Os genomas das cepas do primeiro clado apresentam identidade média de nucleotídeos entre 91.8% e 97.7%, a variação da identidade média de aminoácidos é de 86.4% a 99.8%, mas como possuem \geq 89.2% de identidade em relação ao valores de dDDH podemos considerá-las como cepas da mesma espécie. Apesar dos genomas do segundo clado possuírem ANI de 91.4% a 97.6% e AAI entre 88.1% e 99.7%, acreditamos que realmente estejam agrupadas como sendo da mesma espécie por exibirem valores de dDDH \geq 78.6%. No terceiro clado é apresentado os genomas com intervalos de ANI entre 96.2% e 98.8%, AAI de 90.7% e 96.9%, e os valores de dDDH \geq 91.3%. Com esses dados apresentados, aliados aos dados de Wallner *et al.* (2019), entende-se que as cepas agrupadas no primeiro, segundo e terceiro clado (exceto as cepas Bp9145 e Bp9146) pertencem a mesma espécie bacteriana, *B. cenocepacia*.

As cepas Bp9145 e Bp9146 se agrupam no terceiro clado com outras cepas ambientais, porém possuem valores considerados baixos de dDDH, ANI e AAI ($\geq 66,4\%$, $\geq 93\%$ e $\geq 85.3\%$, respectivamente), o que podemos sugerir que essas duas cepas possam pertencer a uma outra espécie do gênero *Burkholderia*. As duas cepas DDS 22E-1 e DWS 37E-2 apresentam os seguintes valores de dDDH, ANI e AAI: $\geq 35.7\%$, $\geq 84.9\%$ e $\geq 73.0\%$; essas mesmas cepas formaram um grupo externo em todas as análises realizadas neste trabalho, corroborando com resultados obtidos por Wallner *et al.* (2019) que essas duas cepas não pertencem a espécie *B. cenocepacia*, mas sim a espécie *B. pseudomultivorans* e *B. latens*, apresentando respectivamente os valores ANI 97.5% – em relação a cepa DDS 22E-1 – e 99.0% – em relação a cepa DWS 37E-2 (WALLNER et al., 2019). Esses dados indicam que as cepas DDS 22E-1 e DWS 37E-2 são distintas de todas as outras cepas analisadas ao longo deste trabalho.

Com os resultados apresentados nas figuras **Figura 13** e **14** demonstra-se a necessidade de se criar uma nova espécie, a qual agruparia as cepas do segundo clado da **Figura 14** e simultaneamente as cepas do primeiro clado da figura 13. No trabalho

de Wallner *et al* (2019), é sugerido o nome *Burkholderia servocepacia* (*servare*, proteger ou guardar; epíteto específico, *cepacia*), o qual acreditamos englobar eficientemente o significado referente as carcterísticas das cepas do segundo *cluster*.

7. CONCLUSÃO E PERSPECTIVAS

Nossas análises indicam que os dois *clusters* de T3SSs de *B. cenocepacia*, os quais são identificados em algumas cepas ambientais, parecem ter sido adquiridos de forma independente e podem desempenhar papéis diferentes em relação à patogenicidade, determinação da gama de hospedeiros e / ou adaptação de nicho. Os resultados das análises taxonômicas e filogenéticas apresentam a distinção entre as cepas de B. cenocepacia, sugerindo a divisão dessa espécie. Essas evidências, aliadas aos dados experimentais obtidos em nosso grupo, sugerem que patógenos humanos, como a *B. cenocepacia*, podem interagir ativamente com plantas, como hospedeiros intermediários em seu ciclo de vida, através de uma interface funcional dependente do T3SS. A partir deste estudo, pretende-se identificar os efetores que são secretados pelo T3SS, com o objetivo de investigar o potencial de virulência e adaptação ao meio das cepas de B. cenocepacia, assim como identificar proteínas secretadas, que sejam imunizantes, podendo ser eventuais candidatos a antígenos vacinais. A busca de antígenos protetores que viabilizem a produção de vacinas contra os patógenos oportunistas tem sido objeto de grande esforço internacional. Pretende-se também pesquisar o grau de identidade das cepas Bp9145 e Bp9146 com outras espécies do gênero Burkholderia.

8. REFERÊNCIA BIBLIOGRÁFICA

ABBÊS, B. B. et al. Caracterização de mecanismos moleculares envolvidos na interação entre Burkholderia cenocepacia e plantas hospedeiras – Oryza sativa L. e Medicago sativa L. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2014.

AGNOLI, K. et al. Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid. **Molecular Microbiology**, v. 83, n. 2, p. 362–378, 2012.

AGRONOMIA, P. D. E. P. E. M.; BARAÚNA, A. C. Ecologia e Taxonomia de Rizóbios de Mimosas do Brasil: Diversidade e Interação com seus Hospedeiros de Diferentes Biomas. [s.l.] Universidade Federal Rural do Rio de Janeiro, 2017.

ALMPANIS, A. et al. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. **Microbial genomics**, v. 4, n. 4, p. 0–7, 2018.

ALTSCHUL, S. F. et al. Basic local alignment search tool. Journal of Molecular Biology, v. 215, n. 3, p. 403–410, 1990.

ANDRADE, B. G. N. Exploração da diversidade genômica, taxonômica e funcional dos gêneros Wolbachia e Klebsiella. Rio de Janeiro: Fundação Oswaldo Cruz, 2016.

ANSARI, M. et al. Pathogenicity and molecular phylogenetic analysis reveal a distinct position of the banana fingertip rot pathogen among the Burkholderia cenocepacia genomovars. **Plant Pathology**, v. 68, n. 4, p. 804–815, 2019.

AUCH, A. F. et al. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. **Standards in Genomic Sciences**, v. 2, n. 1, p. 117–134, 2010.

AUSSEL, L.; BEUZÓN, C. R.; CASCALES, E. Meeting report: Adaptation and communication of bacterial pathogens. **Virulence**, v. 7, n. 4, p. 481–490, 2016.

BAIA, A. D. B. et al. Predominance of Burkholderia cenocepacia lineages causing onion sour skin in the semi-arid region of north-east Brazil. **Plant Pathology**, v. 70, n. 3, p. 521–533, 2021.

BALDWIN, A. et al. Environmental Burkholderia cepacia complex isolates in human infections. **Emerging Infectious Diseases**, v. 13, n. 3, p. 458–461, 2007.

BATEMAN, A.; COGGILL, P.; FINN, R. D. DUFs: Families in search of function. Acta Crystallographica Section F: Structural Biology and Crystallization Communications, v. 66, n. 10, p. 1148–1152, 2010.

BENSON, D. A. et al. GenBank. Nucleic Acids Research, v. 41, n. D1, p. 36–42, 2013.

BODILIS, J. et al. Comparative genomics of environmental and clinical Burkholderia cenocepacia strains closely related to the highly transmissible epidemic ET12 lineage. **Frontiers in Microbiology**, v. 9, n. MAR, p. 1–12, 2018.

BÜTTNER, D. Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria. **Microbiology and Molecular Biology Reviews**, v. 76, n. 2, p. 262–310, 2012.

CALDART, E. T. et al. Phylogenetic Analysis: Basic Concepts and Its Use as a Tool for Virology and Molecular Epidemiology. Acta Scientiae Veterinariae, v. 44, n. 1, p. 20, 19 mar. 2018.

CHEN, H.; LIU, F.; FU, Z. Q. Deceiving the chaperone. **Nature Plants**, v. 5, n. 11, p. 1110–1111, 2019.

CHENG, S. H. et al. Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. **Cell**, v. 66, n. 5, p. 1027–1036, 1991.

CHIARINI, L. et al. Exopolysaccharides produced by Burkholderia cenocepacia recA lineages IIIA and IIIB. **Journal of Cystic Fibrosis**, v. 3, n. 3, p. 165–172, 2004.

CHRISTIE, P. J. The Rich Tapestry of Bacterial Protein Translocation Systems. **The Protein Journal**, v. 38, n. 4, p. 389–408, 15 ago. 2019.

COHEN, T. S.; PRINCE, A. Cystic fibrosis: A mucosal immunodeficiency syndrome. **Nature Medicine**, v. 18, n. 4, p. 509–519, 2012.

COLSTON, S. M. et al. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using aeromonas as a test case. **mBio**, v. 5, n. 6, 2014.

COSTA, T. R. D. et al. Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. **Nature Reviews Microbiology**, v. 13, n. 6, p. 343–359, 2015.

CUTTING, G. R. Modifier genetics: Cystic fibrosis. Annual Review of Genomics and Human Genetics, v. 6, n. 95, p. 237–260, 2005.

DALMASTRI, C. et al. Investigating Burkholderia cepacia complex populations recovered from Italian maize rhizosphere by multilocus sequence typing. **Environmental Microbiology**, v. 9, n. 7, p. 1632–1639, 2007.

DENG, W. et al. Assembly, structure, function and regulation of type III secretion systems. **Nature Reviews Microbiology**, v. 15, n. 6, p. 323–337, 2017.

DENNIS, J. J. Burkholderia cenocepacia virulence microevolution in the CF lung: Variations on a theme. **Virulence**, v. 8, n. 6, p. 618–620, 2017.

DEPOORTER, E. et al. Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. **Applied Microbiology and Biotechnology**, v. 100, n. 12, p. 5215–5229, 2016.

DEPOORTER, E. et al. Burkholderia cepacia Complex Taxon K: Where to Split? **Frontiers in Microbiology**, v. 11, n. July, 2020.

DIEPOLD, A. et al. Composition, Formation, and Regulation of the Cytosolic C-ring, a Dynamic Component of the Type III Secretion Injectisome. **PLoS Biology**, v. 13, n. 1, p. 1–21, 2015.

DIETSCHE, T. et al. Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus. **PLoS Pathogens**, v. 12, n. 12, p. 1–25, 2016.

DOBRITSA, A. P.; SAMADPOUR, M. Transfer of eleven species of the genus Burkholderia to the genus Paraburkholderia and proposal of Caballeronia gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia. **International Journal of Systematic and Evolutionary Microbiology**, v. 66, n. 8, p. 2836–2846, 2016.

EBERL, L.; VANDAMME, P. Members of the genus Burkholderia: Good and bad guys [version 1; referees: 3 approved]. **F1000Research**, v. 5, n. May, 2016.

ESCOBAR-ZEPEDA, A. et al. Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics. **Scientific Reports**, v. 8, n. 1, p. 1–13, 2018.

ESTRADA-DE LOS SANTOS, P. et al. Phylogenetic analysis of burkholderia species by multilocus sequence analysis. **Current Microbiology**, v. 67, n. 1, p. 51–60, 2013.

ESTRADA-DE LOS SANTOS, P. et al. Whole genome analyses suggests that Burkholderia sensu lato contains two additional novel genera (Mycetohabitans gen. nov., and Trinickia gen. nov.): Implications for the evolution of diazotrophy and nodulation in the Burkholderiaceae. **Genes**, v. 9, n. 8, 2018.

ETESAMI, H.; ALIKHANI, H. A.; HOSSEINI, H. M. Indole-3-acetic acid (IAA) production trait, a useful screening to select endophytic and rhizosphere competent bacteria for rice growth promoting agents. **MethodsX**, v. 2, p. 72–78, 2015.

GALÁN, J. E.; WOLF-WATZ, H. Protein delivery into eukaryotic cells by type III secretion machines. **Nature**, v. 444, n. 7119, p. 567–573, 2006.

GAO, N. L. et al. Prokaryotic Genome Expansion Is Facilitated by Phages and Plasmids but Impaired by CRISPR. **Frontiers in Microbiology**, v. 10, n. October, p. 1–8, 2019.

GARRITY, G. M. A New Genomics-Driven Taxonomy of Bacteria and Archaea: Are We There Yet? **Journal of Clinical Microbiology**, v. 54, n. 8, p. 1956–1963, ago. 2016.

GAUTAM, V. et al. Multilocus sequence analysis reveals high genetic diversity in clinical isolates of Burkholderia cepacia complex from India. **Scientific Reports**, v. 6, n. May, 2016.

GEVERS, D. et al. Defining prokaryotic species Reevaluating prokaryotic species. **Microbiology**, v. 3, n. September, p. 733–739, 2005.

GIBSON, R. L.; BURNS, J. L.; RAMSEY, B. W. Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis. American Journal of Respiratory and Critical Care Medicine, v. 168, n. 8, p. 918–951, 2003.

GLAESER, S. P.; KÄMPFER, P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. **Systematic and Applied Microbiology**, v. 38, n. 4, p. 237–245, 2015.

GORIS, J. et al. DNA-DNA hybridization values and their relationship to wholegenome sequence similarities. **International Journal of Systematic and Evolutionary Microbiology**, v. 57, n. 1, p. 81–91, 2007.

GU, Z. et al. Circlize implements and enhances circular visualization in R. **Bioinformatics**, v. 30, n. 19, p. 2811–2812, 2014.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HOLDEN, M. T. G. et al. The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. **Journal of Bacteriology**, v. 91, n. 1, p. 261–277, 2009.

HOWDEN, A. J. M. et al. Pseudomonas syringae pv. syringae B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. **Molecular Plant Pathology**, v. 10, n. 6, p. 857–865, 2009.

HU, Y. et al. A global survey of bacterial type III secretion systems and their effectors. **Environmental Microbiology**, v. 19, n. 10, p. 3879–3895, 2017.

HUECK, C. J. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. **Microbiology and Molecular Biology Reviews**, v. 62, n. 2, p. 379–433, 1998.

JAIN, C. et al. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. **Nature Communications**, v. 9, n. 1, p. 1–8, 2018.

JESUS, S. DE. Chalcone analogues on the control of Meloidogyne incognita and genomic studies of the biocontrol agent Bacillus velezensis. Lavras: Universidade Federal de Lavras, 2020.

JIN, Y. et al. Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status. **Biology Direct**, v. 15, n. 1, p. 1–14, 2020.

JOHNSTON, R. B. Clinical aspects of chronic granulomatous disease. **Current Opinion in Hematology**, v. 8, n. 1, p. 17–22, 2001.

JOSHI, G.; CHAUHAN, C.; DAS, S. Microsynteny analysis to understand evolution and impact of polyploidization on MIR319 family within Brassicaceae. **Development Genes and Evolution**, v. 228, n. 6, p. 227–242, 2018.

KAO, S. M. et al. Molecular and genetic analysis of a region of plasmid pCF10 containing positive control genes and structural genes encoding surface proteins involved in pheromone-inducible conjugation in Enterococcus faecalis. **Journal of Bacteriology**, v. 173, n. 23, p. 7650–7664, 1991.

KITAHARA, K.; MIYAZAKI, K. Revisiting bacterial phylogeny: Natural and experimental evidence for horizontal gene transfer of 16S rRNA. **Mobile genetic elements**, v. 3, n. 1, p. e24210, 2013.

KONSTAN, M. W. et al. Risk Factors For Rate of Decline in Forced Expiratory Volume in One Second in Children and Adolescents with Cystic Fibrosis. **Journal of Pediatrics**, v. 151, n. 2, 2007.

KOOGER, R. et al. CryoEM of bacterial secretion systems. Current Opinion in Structural Biology, v. 52, p. 64–70, 2018.

KUHLEN, L. et al. Structure of the core of the type iii secretion system export apparatus. **Nature Structural and Molecular Biology**, v. 25, n. 7, 2018.

KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 2018.

LANDETA, G. et al. Use of recA gene sequence analysis for the identification of Staphylococcus equorum strains predominant on dry-cured hams. **Food Microbiology**, v. 28, n. 6, p. 1205–1210, 2011.

LARA-TEJERO, M. et al. A Sorting Platform Determines the Order of Protein Secretion in Bacterial Type III Systems. **Science**, v. 331, n. 6021, p. 1188–1191, 4 mar. 2011.

LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947–2948, 2007.

LASSALLE, F. et al. GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. **PLoS Genetics**, v. 11, n. 2, p. 1–20, 2015.

LETUNIC, I.; BORK, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. **Nucleic Acids Research**, v. 49, n. W1, p. W293–W296, 2021.

LIMA, L. DA S. Alteração no perfil de exsudação em plântulas de milho tratadas com ácidos húmicos e bactérias promotoras do crescimento vegetal. Campos dos Goytacazes: Universidade Estadual do Norte Fluminense Darcy Ribeiro, 2014.

LIOU, T. G. et al. Predictive 5-Year Survivorship Model of Cystic Fibrosis. American Journal of Epidemiology, v. 153, n. 4, p. 345–352, 15 fev. 2001.

LIPUMA, J. J. Update on the Burkholderia cepacia complex. Current Opinion in **Pulmonary Medicine**, v. 11, n. 6, p. 528–533, 2005.

LIU, F. et al. Molecular Structure of the Human CFTR Ion Channel. **Cell**, v. 169, n. 1, p. 85-95.e8, 2017a.

LIU, Y. et al. Transcriptional analysis of genes involved in competitive nodulation in Bradyrhizobium diazoefficiens at the presence of soybean root exudates. **Scientific Reports**, v. 7, n. 1, p. 1–11, 2017b.

LÓPEZ-HERMOSO, C. et al. Assessment of multilocus sequence analysis as a valuable tool for the classification of the genus Salinivibrio. **Frontiers in Microbiology**, v. 8, n. JUN, p. 1–14, 2017.

LOPEZ, V. A. et al. A Bacterial Effector Mimics a Host HSP90 Client to Undermine Immunity. **Cell**, v. 179, n. 1, p. 205- 218.e21, set. 2019.

MADEIRA, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. **Nucleic Acids Research**, v. 47, n. W1, p. W636–W641, 2019.

MAHAJAN, S.; AGASHE, D. Evolutionary jumps in bacterial GC content. **bioRxiv**, p. 2021.02.16.431469, 2021.

MAHENTHIRALINGAM, E. et al. DNA-based diagnostic approaches for identification of Burkholderia cepacia complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis, and Burkholderia cepacia genomovars I and III. **Journal of Clinical Microbiology**, v. 38, n. 9, p. 3165–3173, 2000.

MALEŠEVIĆ, M. et al. Virulence traits associated with Burkholderia cenocepacia ST856 epidemic strain isolated from cystic fibrosis patients. **Antimicrobial Resistance and Infection Control**, v. 6, n. 1, p. 1–9, 2017.

MCDERMOTT, G.; REECE, E.; RENWICK, J. Microbiology of the cystic fibrosis airway. **Encyclopedia of Microbiology**, p. 186–198, 2019.

MEDLAR, A. J.; TÖRÖNEN, P.; HOLM, L. AAI-profiler: Fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. **Nucleic Acids Research**, v. 46, n. W1, p. W479–W485, 2018.

MEIER-KOLTHOFF, J. P. et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. **BMC Bioinformatics**, v. 14, 2013.

MONJARÁS FERIA, J.; VALVANO, M. A. An Overview of Anti-Eukaryotic T6SS Effectors. Frontiers in Cellular and Infection Microbiology, v. 10, n. October, 2020.

MURRAY, G. G. R. et al. Genome Reduction Is Associated with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. **Molecular biology and evolution**, v. 38, n. 4, p. 1570–1579, 2021.

NOTTI, R. Q. et al. A common assembly module in injectisome and flagellar type III secretion sorting platforms. **Nature Communications**, v. 6, n. May, 2015.

NOTTI, R. Q.; STEBBINS, C. E. The structure and function of type iii secretion systems. **Virulence Mechanisms of Bacterial Pathogens**, v. 4, n. 1, p. 241–264, 2016.

OLIVEIRA, P. H. et al. The chromosomal organization of horizontal gene transfer in bacteria. **Nature Communications**, v. 8, n. 1, p. 25–28, 2017a.

OLIVEIRA, W. J. et al. First Report of Burkholderia cenocepacia Causing Sour Skin of Onion (Allium cepa) in Brazil. **Plant Disease**, v. 101, n. 11, p. 1950, nov. 2017b.

OLIVEIRA, W. J. et al. Elucidating the etiology of onion bacterial scale rot in the semiarid region of Northeastern Brazil. **Tropical Plant Pathology**, v. 44, n. 6, p. 494–502, 2019.

OLIVER, J. L.; MARÍN, A. A relationship between GC content and coding-sequence length. **Journal of Molecular Evolution**, v. 43, n. 3, p. 216–223, 1996.

ORATA, F. D. **Taxonomy of bacteria in the genomic era**. [s.l.] University of Alberta, 2017.

PAYNE, G. W. et al. Development of a recA gene-based identification approach for the entire Burkholderia genus. **Applied and Environmental Microbiology**, v. 71, n. 7, p. 3917–3927, 2005.

PEETERS, C. et al. Burkholderia pseudomultivorans sp. nov., a novel Burkholderia cepacia complex species from human respiratory samples and the rhizosphere. **Systematic and Applied Microbiology**, v. 36, n. 7, p. 483–489, 2013.

PIROMYOU, P. et al. Mutualistic co-evolution of T3SSs during the establishment of symbiotic relationships between Vigna radiata and Bradyrhizobia. **MicrobiologyOpen**, v. 8, n. 7, p. 1–16, 2019.

PORTALIOU, A. G. et al. Type III Secretion: Building and Operating a Remarkable Nanomachine. **Trends in Biochemical Sciences**, v. 41, n. 2, p. 175–189, fev. 2016.

QIAN, G. et al. AsnB, regulated by diffusible signal factor and global regulator Clp, is involved in aspartate metabolism, resistance to oxidative stress and virulence in Xanthomonas oryzae pv. oryzicola. **Molecular Plant Pathology**, v. 14, n. 2, p. 145–157, 2013.

RAJENDHRAN, J.; GUNASEKARAN, P. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. **Microbiological Research**, v. 166, n. 2, p. 99–110, 2011.

ROCA, A. I.; COX, M. M. **RecA protein: structure, function, and role in recombinational DNA repair.** [s.l.] Elsevier Masson SAS, 1997. v. 56

ROMERO-GUTIÉRREZ, K. J. et al. Phenotypic traits of Burkholderia spp. associated with ecological adaptation and plant-host interaction. **Microbiological Research**, v. 236, n. November 2019, p. 126451, 2020.

RYAN, R. P. et al. The DSF Family of Cell–Cell Signals: An Expanding Class of Bacterial Virulence Regulators. **PLoS Pathogens**, v. 11, n. 7, p. 1–14, 2015.

SAJJAN, U. S. et al. Burkholderia cenocepacia ET12 strain activates TNFR1 signalling in cystic fibrosis airway epithelial cells. **Cellular Microbiology**, v. 23, n. 1, p. 070816152918003-???, 14 ago. 2007.

SATAPATHY, S. S.; DUTTA, M.; RAY, S. K. Variable correlation of genome GC% with transfer RNA number as well as with transfer RNA diversity among bacterial groups: α-Proteobacteria and Tenericutes exhibit strong positive correlation. **Microbiological Research**, v. 165, n. 3, p. 232–242, 2010.

SAWABE, T. et al. Updating the Vibrio clades defined by multilocus sequence phylogeny: Proposal of eight new clades, and the description of Vibrio tritonius sp. nov. **Frontiers in Microbiology**, v. 4, n. DEC, p. 1–14, 2013.

SAWANA, A.; ADEOLU, M.; GUPTA, R. S. Molecular signatures and phylogenomic analysis of the genus burkholderia: Proposal for division of this genus into the emended genus burkholderia containing pathogenic organisms and a new genus paraburkholderia gen. nov. harboring environmental species. **Frontiers in Genetics**, v. 5, n. NOV, p. 1–22, 2014.

SCOFFONE, V. C. et al. Burkholderia cenocepacia infections in cystic fibrosis patients: Drug resistance and therapeutic approaches. **Frontiers in Microbiology**, v. 8, n. AUG, p. 1–13, 2017.

SEEMANN, T. barrnap 0.9: rapid ribosomal RNA prediction. Github.Com, 2013.

SEEMANN, T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics, 2014.

SEPE, A. et al. Evaluation of recA sequencing for the classification of Aeromonas strains at the genotype level. Letters in Applied Microbiology, v. 46, n. 4, p. 439–444, 2008.

SFEIR, M. M. Burkholderia cepacia complex infections: More complex than the bacterium name suggest. **Journal of Infection**, v. 77, n. 3, p. 166–170, 2018.

SHENG, D. H. et al. Functional Division Between the RecA1 and RecA2 Proteins in Myxococcus xanthus. **Frontiers in Microbiology**, v. 11, n. February, p. 1–13, 2020.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 539, 2011.

SPAEPEN, S.; VANDERLEYDEN, J.; REMANS, R. Indole-3-acetic acid in microbial and microorganism-plant signaling. **FEMS Microbiology Reviews**, v. 31, n. 4, p. 425–448, 2007.

SULLIVAN, M. J.; PETTY, N. K.; BEATSON, S. A. Easyfig: A genome comparison visualizer. **Bioinformatics**, v. 27, n. 7, p. 1009–1010, 2011.

SUVOROVA, I. A. et al. Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. **Journal of Bacteriology**, v. 193, n. 15, p. 3956–3963, 2011.

TAMURA, K. et al. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, 2013.

TEULET, A. et al. Phylogenetic distribution and evolutionary dynamics of NOD and T3SS genes in the genus bradyrhizobium. **Microbial Genomics**, v. 6, n. 9, p. 1–18, 2020.

THOMPSON, C. C. et al. Microbial genomic taxonomy. **BMC Genomics**, v. 14, n. 1, 2013.

TOLMAN, J. S. Scholarship @ Western The intracellular behaviour of Burkholderia cenocepacia in murine macrophages. 2011.

TOMICH, M. et al. Attenuated virulence of a Burkholderia cepacia type III secretion mutant in a murine model of infection. **Infection and Immunity**, v. 71, n. 3, p. 1405–1415, 2003.

TORRES-VARGAS, C. E. et al. The inner rod of virulence-associated type III secretion systems constitutes a needle adapter of one helical turn that is deeply integrated into the system's export apparatus. **Molecular Microbiology**, v. 112, n. 3, p. 918–931, 2019.

TROISFONTAINES, P.; CORNELIS, G. R. Type III secretion: More systems than you think. **Physiology**, v. 20, n. 5, p. 326–339, 2005.

TURTON, J. F. et al. Revised approach for identification of isolates within the Burkholderia cepacia complex and description of clinical isolates not assigned to any of the known genomovars. **Journal of Clinical Microbiology**, v. 45, n. 9, p. 3105–3108, 2007.

VANDAMME, P. et al. Burkholderia cenocepacia sp. nov. - A new twist to an old story. **Research in Microbiology**, v. 154, n. 2, p. 91–96, 2003.

VANDAMME, P.; DAWYNDT, P. Classification and identification of the Burkholderia cepacia complex: Past, present and future. **Systematic and Applied Microbiology**, v. 34, n. 2, p. 87–95, 2011.

VANLAERA, E. et al. Burkholderia latens sp. nov., Burkholderia diffusa sp. nov., Burkholderia arboris sp. nov., Burkholderia seminalis sp. nov., and Burkholderia metallica sp. nov., novel species within the Burkholderia cepacia complex. **International Journal of Systematic and Evolutionary Microbiology**, v. 58, n. 7, p. 1580–1590, 2008.

WAGNER, S. et al. Organization and coordinated assembly of the type III secretion export apparatus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 41, p. 17745–17750, 2010.

WAGNER, S. et al. Bacterial type III secretion systems: A complex device for the delivery of bacterial effector proteins into eukaryotic host cells. **FEMS Microbiology Letters**, v. 365, n. 19, p. 1–13, 2018.

WALLNER, A. et al. Genomic analyses of Burkholderia cenocepacia reveal multiple species with differential host-Adaptation to plants and humans. **BMC Genomics**, v. 20, n. 1, p. 1–15, 2019.

WALTER H. BURKHOLDER. Sour skin, a bacterial rot of Onion bulbs. **Phytopathology**, 1950.

WINSOR, G. L. et al. The Burkholderia Genome Database: Facilitating flexible queries and comparative analyses. **Bioinformatics**, v. 24, n. 23, p. 2803–2804, 2008.

WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. **Proceedings of the National Academy of Sciences of the United States of America**, v. 87, n. 12, p. 4576–4579, 1990.

WORRALL, L. J. et al. Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. **Nature**, v. 540, n. 7634, p. 597–601, 2016.

YABUUCHI, E. et al. Proposal of Burkholderia gen. nov. and Transfer of Seven Species of the Genus Pseudomonas Homology Group II to the New Genus, with the Type Species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. **Microbiol. Immunol**, v. 36, n. 12, p. 1251–1275, 1992.

YAMASSAKI PEREIRA, V. M. Reconstrução filogenética de procariotos com base em famílias de genes homólogos. São Paulo: Universidade de São Paulo, 2017.

YOU, M. et al. Isolation and characterization of Burkholderia cenocepacia CR318, a phosphate solubilizing bacterium promoting corn growth. **Microbiological Research**, v. 233, p. 126395, 2020.

ZARIVACH, R. et al. Structural analysis of a prototypical ATPase from the type III secretion system. **Nature Structural and Molecular Biology**, v. 14, n. 2, p. 131–137, 2007.

Material Suplementar
Tabela S1: Cepas de *Burkholderia cenocepacia* utilizadas e suas respectivas informações genômicas. Origem Ambiental

Nome do organismo	Fonte de Isolamento	Сера	BioSample	BioProject	Assembly	Integridade do Genoma	Tamanho do Genomas(Mb)	GC%	Scaffolds	CDS t	RNA tmR	NA	CRISPR
Burkholderia cenocepacia	aerossol	DDS 22E-1	SAMN02796386	PRJNA244014 GCA	A_000755725.1	Completo	8.04525	66.9	3	7038	78	1	1
Burkholderia cenocepacia	água	MSMB384WGS	SAMN03449602	PRJNA279182 GC/	A_001718895.1	Completo	7.78060	67.2	3	6895	81	1	-
Burkholderia cenocepacia	solo	DWS 37E-2	SAMN02796570	PRJNA244015 GCA	A_000764955.1	Completo	6.61242	66.5	3	5691	74	1	-
Burkholderia cenocepacia	rizosfera do milho	MC0-3	SAMN02598404	PRJNA17929 GO	CA_000019505.1	Completo	7.97139	66.5	3	7031	79	1	-
Burkholderia cenocepacia	raiz do milho	CR318	SAMN05756084	PRJNA342374 GCA	A_002007585.1	Completo	7.66489	66.8	3	6755	82	1	-
Burkholderia cenocepacia	solo isolado do campo de cebola	HI2424	SAMN00113327	PRJNA13918 GC	CA_000203955.1	Completo	7.70284	66.8	4	6867	84	1	-
Burkholderia cenocepacia	solo	Bp8985	SAMN09073238	PRJNA451205 GCA	A_003854725.1	Contig	7.39275	67.0	52	6574	74	1	-
Burkholderia cenocepacia	água	Bp9023	SAMN09073273	PRJNA451205 GCA	A_003854055.1	Contig	7.33311	67.1	59	6502	71	1	-
Burkholderia cenocepacia	solo	Bp8979	SAMN09073235	PRJNA451205 GCA	A_003854115.1	Contig	7.55294	66.8	61	6683	71	1	-
Burkholderia cenocepacia	solo	Bp8978	SAMN09073234	PRJNA451205 GCA	A_003854125.1	Contig	7.55269	66.8	74	6677	70	1	-
Burkholderia cenocepacia	solo	Bp9150	SAMN09073390	PRJNA451205 GCA	A_003854485.1	Contig	7.51785	67.0	68	6653	69	1	-
Burkholderia cenocepacia	solo	Bp9134	SAMN09073376	PRJNA451205 GCA	A_003854595.1	Contig	7.71317	67.2	68	6875	71	1	-
Burkholderia cenocepacia	solo	Bp8970	SAMN09073226	PRJNA451205 GCA	A_003854785.1	Contig	7.41251	67.1	51	6601	71	1	-
Burkholderia cenocepacia	solo	Bp9127	SAMN09073369	PRJNA451205 GCA	A_003854625.1	Contig	7.69277	66.6	71	6822	68	1	-
Burkholderia cenocepacia	solo	Bp9158	SAMN09073398	PRJNA451205 GCA	A_003854475.1	Contig	7.78324	67.2	78	6950	71	1	-
Burkholderia cenocepacia	solo	Bp9129	SAMN09073371	PRJNA451205 GCA	A_003854585.1	Contig	7.69050	66.6	70	6823	67	1	-
Burkholderia cenocepacia	solo	Bp9145	SAMN09073385	PRJNA451205 GCA	A_003854495.1	Contig	7.55000	67.0	70	6744	73	1	-
Burkholderia cenocepacia	solo	Bp8974	SAMN09073230	PRJNA451205 GCA	A_003854775.1	Contig	7.76081	66.7	104	6825	70	1	-
Burkholderia cenocepacia	solo	Bp9146	SAMN09073386	PRJNA451205 GCA	A_003854505.1	Contig	7.46291	67.1	68	6635	71	1	-
Burkholderia cenocepacia	solo	Bp9139	SAMN09073380	PRJNA451205 GCA	A_003854545.1	Contig	7.56367	67.4	95	6728	70	1	-
Burkholderia cenocepacia	água	Bp9038	SAMN09073284	PRJNA451205 GCA	A_003858255.1	Contig	7.63750	66.8	108	6706	70	1	-
Burkholderia cenocepacia	solo	Bp9117	SAMN09073359	PRJNA451205 GCA	A_003854065.1	Contig	7.72518	66.8	109	6779	68	1	-
Burkholderia cenocepacia	água	Bp9037	SAMN09073283	PRJNA451205 GCA	A_003858275.1	Contig	7.63914	66.8	111	6709	69	1	-
Burkholderia cenocepacia	solo	Bp9128	SAMN09073370	PRJNA451205 GCA	A_003854605.1	Contig	7.53964	66.9	106	6641	70	1	-
Burkholderia cenocepacia	solo agrícola	CEIB S5-2	SAMN04252993	PRJNA301637 GCA	A_001541445.1	Contig	8.97605	65.7	109	7981	111	1	-

Tabela S1: Cepas de *Burkholderia cenocepacia* utilizadas e suas respectivas informações genômicas. Origem Clínica

Nome do organismo	Fonte de Isolamento	Сера	BioSample	BioProject	Assembly	Integridade do Genoma	Tamanho do Genomas(Mb)	GC%	Scaffolds	CDS	RNA tmR	NA	CRISPR
Burkholderia cenocepacia	escarro	VC7848	SAMN05001925	PRJNA289138	GCA_001999785.1	Completo	7.49946	66.9	1	6631	78	1	-
Burkholderia cenocepacia	sangue do cordão umbilical	895	SAMN04574071	PRJNA316047	GCA_001606135.1	Completo	8.73148	66.7356	3	7784	79	1	-
Burkholderia cenocepacia	escarro	VC12802	SAMN05001933	PRJNA289138	GCA_001999825.1	Completo	7.39491	67.0285	2	6508	80	1	-
Burkholderia cenocepacia	escarro	J2315	SAMEA1705928	PRJNA339	GCA_000009485.1	Completo	8.05578	66.9165	4	7199	87	1	-
Burkholderia cenocepacia	nasal	842	SAMN04570263	PRJNA315790	GCA_001606115.1	Completo	8.14970	66.9768	4	7156	82	1	-
Burkholderia cenocepacia	escarro	H111	SAMEA3138403	PRJNA69823	GCA_000236215.4	Completo	7.71489	67.3149	3	6805	78	1	-
Burkholderia cenocepacia	escarro	ST32	SAMN03323790	PRJNA274219	GCA_001484665.1	Completo	8.09039	67.0132	4	7016	81	1	-
Burkholderia cenocepacia	escarro	VC12308	SAMN05001912	PRJNA289138	GCA_001999885.1	Completo	7.63450	67.1419	4	6775	79	1	-
Burkholderia cenocepacia	escarro	PC184 Mulks	SAMN06842015	PRJNA384579	GCA_003076415.1	Completo	7.06705	66.8313	3	6177	76	1	-
Burkholderia cenocepacia	sangue	AU 1054	SAMN02598326	PRJNA13919	GCA_000014085.1	Completo	7.27912	66.9164	3	6372	82	1	-
Burkholderia cenocepacia	escarro	VC2307	SAMN05001798	PRJNA289138	GCA_001999805.1	Cromossomo	7.91588	66.9889	4	7211	82	1	-
Burkholderia cenocepacia	escarro	VC1254	SAMN05001969	PRJNA289138	GCA_001999925.1	Cromossomo	8.17737	66.9950	4	7187	81	1	-
Burkholderia cenocepacia	escarro	GIMC4560:Bcn122	SAMN06611151	PRJNA379546	GCA_002083015.1	Cromossomo	8.04369	67.1053	3	6943	96	2	-
Burkholderia cenocepacia	isolado clínico	FDAARGOS_518	SAMN10163212	PRJNA231221	GCA_003940705.1	Contig	7.49941	67.4	3	6613	78	1	-
Burkholderia cenocepacia	escarro	VC7604	SAMN05001827	PRJNA289138	GCA_001984275.1	Contig	7.09484	66.9	5	6325	82	1	-
Burkholderia cenocepacia	escarro	FDAARGOS_187	SAMN05004742	PRJNA231221	GCA_002891255.1	Contig	7.46533	67.4	6	6683	78	1	-
Burkholderia cenocepacia	escarro	VC9789	SAMN05001819	PRJNA289138	GCA_001984315.1	Contig	7.80352	66.9	7	6947	86	1	-
Burkholderia cenocepacia	escarro	VC4558	SAMN05001820	PRJNA289138	GCA_001984325.1	Contig	8.22889	67	9	7285	82	1	-
Burkholderia cenocepacia	escarro	FDAARGOS_189	SAMN05004744	PRJNA231221	GCA_002891135.1	Contig	8.09455	66.9	9	7355	87	1	-
Burkholderia cenocepacia	escarro	VC5486	SAMN05001975	PRJNA289138	GCA_001984285.1	Contig	7.73614	66.9	10	6785	91	1	-
Burkholderia cenocepacia	escarro	VC3917	SAMN05001990	PRJNA289138	GCA_001984355.1	Contig	7.34467	66.8	13	6534	80	1	-
Burkholderia cenocepacia	escarro	FDAARGOS_82	SAMN02700076	PRJNA231221	GCA_000783555.2	Contig	7.31971	66.9	16	6435	84	1	-
Burkholderia cenocepacia	escarro	VC1255	SAMN05001970	PRJNA289138	GCA_001984375.1	Contig	8.37363	66.9	17	7329	86	1	-
Burkholderia cenocepacia	escarro	K56-2	SAMN02045905	PRJNA279989	GCA_000981305.1	Contig	7.81315	67	17	6880	83	1	-
Burkholderia cenocepacia	escarro	K56-2Valvano	SAMN00255233	PRJNA62783	GCA_000333155.2	Contig	7.75026	67	19	6953	81	1	-

Cluster	Locus_tag do gene hp
2	Bc_842_03472
2	Bc_895_01826
2	Bc_AU 1054_03571
2	Bc_Bp8970_05777
2	Bc_Bp8974_02537
2	Bc_Bp8978_03107
2	Bc Bp8979 01725
2	Bc Bp8985 00096
2	Bc Bp9023 02601
2	Bc Bp9037 01230
2	Bc Bp9038 01036
2	Bc Bp9117 05876
2	Bc Bp9127 05079
2	Bc Bp9128 03482
2	Bc_Bp9129_04830
2	Bc Bp9134 01733
2	Bc_Bp9139_02161
2	Bc_Bp9145_05082
2	Bc_Bp9146_04739
2	Bc_Bp9150_01963
2	Bc_Bp9158_05627
	2 2 <td< td=""></td<>

Tabela S2: Cepas que possuem o gene que codifica a proteína hipotética (*hypothetical protein – hp*).

Nome da cepa	Cluster	Locus tag do gene hp
Burkholderia cenocepacia CEIB S5-2	2	Bc CEIB S5-2 03413
Burkholderia cenocepacia CR318	2	\overline{Bc} CR $\overline{3}18$ $\overline{05008}$
Burkholderia cenocepacia FDAARGOS 187	2	Bc FDAARGOS 187 03937
Burkholderia cenocepacia FDAARGOS 189	1	Bc_FDAARGOS_189_05137
Burkholderia cenocepacia FDAARGOS 518	2	Bc_FDAARGOS_518_04953
Burkholderia cenocepacia FDAARGOS 82	2	Bc_FDAARGOS_82_04088
Burkholderia cenocepacia HI2424	2	Bc_HI2424_04855
Burkholderia cenocepacia J2315	1	Bc_J2315_05687
Burkholderia cenocepacia K56-2	2	Bc_K56-2_01183
Burkholderia cenocepacia K56-2Valvano	2	Bc_K56-2Valvano_04521
Burkholderia cenocepacia MC0-3	2	Bc_MC0-3_05536
Burkholderia cenocepacia MSMB384WGS	2	Bc_MSMB384WGS_05233
Burkholderia cenocepacia PC184 Mulks	3	Bc_PC184_Mulks_01235
Burkholderia cenocepacia ST32	4	Bc_ST32_05812
Burkholderia cenocepacia VC12308	2	Bc_VC12308_05221
Burkholderia cenocepacia VC1254	4	Bc_VC1254_05506
Burkholderia cenocepacia VC1255	4	Bc_VC1255_06753
Burkholderia cenocepacia VC12802	2	Bc_VC12802_02865
Burkholderia cenocepacia VC2307	4	Bc_VC2307_03821
Burkholderia cenocepacia VC3917	4	Bc_VC3917_01988
Burkholderia cenocepacia VC4558	2	Bc_VC4558_06635
Burkholderia cenocepacia VC5486	4	Bc_VC5486_03068
Burkholderia cenocepacia VC7604	2	Bc_VC7604_04177
Burkholderia cenocepacia VC7848	2	Bc_VC7848_06558

<u>Tabela S2: Cepas que possuem o gene que codifica a proteína hipotética (*hypothetical protein – hp*).</u>