UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM FITOPATOLOGIA

Tese de Doutorado

FILOGENÔMICA DE NOVAS ESPÉCIES DO COMPLEXO *Burkholderia cepacia* E GENÔMICA COMPARATIVA DE ISOLADOS DE *B. cenocepacia* LINHAGENS IIIA E IIIB, CAUSADORAS DE PODRIDÃO DAS ESCAMAS DA CEBOLA

Leandro da Silva Velez

RECIFE-PE

2021

LEANDRO DA SILVA VELEZ

FILOGENÔMICA DE NOVAS ESPÉCIES DO COMPLEXO *Burkholderia cepacia* E GENÔMICA COMPARATIVA DE ISOLADOS DE *B. cenocepacia* LINHAGENS IIIA E IIIB, CAUSADORAS DE PODRIDÃO DAS ESCAMAS DA CEBOLA

Tese apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Fitopatologia.

COMITÊ DE ORIENTAÇÃO:

Orientador: Prof. Dr. Marco Aurélio Siqueira da Gama
Coorientadora: Profª. Dra. Flavia Figueira Aburjaile
Coorientadora: Profª. Dra. Ana Maria Benko-Iseppon

RECIFE-PE

SETEMBRO – 2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

V436f Velez, Leandro da Silva

Filogenômica de novas espécies do complexo Burkholderia cepacia e genômica comparativa de isolados de B. cenocepacia linhagens IIIA e IIIB, causadoras de podridão das escamas da cebola / Leandro da Silva Velez. - 2021.

94 f. : il.

Orientador: Marco Aurelio Siqueira da Gama.

Coorientadora: Flavia Figueira Aburjaile.

Inclui referências.

Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Fitopatologia, Recife, 2021.

1. Burkholderia solum. 2. Burkholderia semiaridus. 3. Podridão das escamas . 4. Genômica comparativa. I. Gama, Marco Aurelio Siqueira da, orient. II. Aburjaile, Flavia Figueira, coorient. III. Título

CDD 632

FILOGENÔMICA DE NOVAS ESPÉCIES DO COMPLEXO *Burkholderia cepacia* E GENÔMICA COMPARATIVA DE ISOLADOS DE *B. cenocepacia* LINHAGENS IIIA E IIIB, CAUSADORAS DE PODRIDÃO DAS ESCAMAS DA CEBOLA

LEANDRO DA SILVA VELEZ

Tese defendida e aprovada pela Banca Examinadora em: 29/09/2021.

ORIENTADOR:

_____
Prof. Dr. Marco Aurélio Siqueira da Gama

EXAMINADORES:

_____
Dr. Adriano Márcio Freire Silva

_____
Profª Dra. Ana Maria Benko Iseppon

_____
Profª Dra. Elineide Barbosa de Sousa

_____
Dra. Flávia Figueira Aburjaile

RECIFE - PE

SETEMBRO - 2021

Aos meus irmãos, ***Alde Velez Júnior****.,* ***Aldilene Velez*** *e* ***Elder Velez,*** *pelo apoio, incentivo, amizade e compreensão. A minha esposa,* ***Barbara Marchesini Malta****, pelo companheirismo, paciência e ajuda nos momentos mais difíceis e, aos meus sogros,* ***José Alvaro Malta*** *e* ***Angelica Marchesini,*** *por nos ensinar como a vida deve ser vivida.*

OFEREÇO

Aos meus pais, ***Gilda Velez e Alde Velez****, por todo o esforço que fizeram para que eu pudesse estar aqui, pelo exemplo de caráter e de pessoas determinadas que são, e a minha querida avó,* ***Neuza Velez,*** *que nos mostrou que com fé chegaremos aonde quisermos*

DEDICO

AGRADECIMENTOS

A Deus, pela dádiva da vida, força e fé para superar os contratempos encontrados durante essa caminhada.

Aos meus pais, Alde e Gilda, por todo amor e apoio durante minha vida; à minha esposa e companheira, Barbara, por sua paciência e compreensão nas horas mais difíceis; e ao meu sogro José Alvaro e a minha sogra Angelica, por toda compreensão e apoio durante esta caminhada;

À Universidade Federal Rural de Pernambuco, pelo apoio institucional, e ao CNPq, pela concessão da bolsa de estudo;

Ao meu orientador, Prof. Dr. Marco Aurélio Siqueira da Gama, pela orientação, apoio e pronto atendimento sempre;

Às minhas coorientadoras, Profª. Dra. Flavia Figueira Aburjaile, que despertou a minha curiosidade pela bioinformática me fazendo trilhar esse caminho com mais facilidade e a Profª. Dra. Ana Maria Benko Iseppon, por todo apoio durante o período que utilizei o Laboratório de Genética e Biologia Vegetal (LGBV) e a Profª. Dra. Elineide Barbosa de Souza, por toda paciência e aprendizado durante toda minha vida acadêmica.

Ao Dr. Adriano Márcio Freire Silva, por sua contagiante busca incessante pelo conhecimento, e a Profª Rosa de Lima Mariano, por iluminar a caminhada de todos do LAFIBAC;

Aos professores Jong Hyun Ham e Lawrence E. Datnoff da Lousiana State University (EUA), e a todos os membros do Laboratório do Dr. Ham, pelo apoio incondicional durante minha estada por lá.

Aos amigos Dr. Roberto Farias e Juan Ariute por todo apoio e ajuda no decorrer das análises, com isso o trabalho se tornou menos árduo;

Aos meus companheiros do Laboratório de Fitobacteriologia pela amizade e apoio nos momentos de indecisão;

A todos que de alguma forma colaboraram para meu sucesso, muito obrigado.

Sumário

RESUMO GERAL

No Brasil, a cebola (*Allium cepa* L.) apresenta elevada importância socioeconômica, destacando-se como a hortaliça mais produzida dentro do gênero *Allium*. Diversas doenças podem acometer essa cultura, tanto durante a produção como na fase de pós-colheita. Dentre as que apresentam elevada importância econômica, destaca-se a podridão em escamas, conhecida popularmente como camisa d'água. Esta doença está associada a várias espécies bacterianas, dentre as quais se sobressaem espécies do complexo *Burkholderia cepacia* (CBC), como *B. cepacia* e *B. cenocepacia*, além de *B. gladioli* pv. *alliicola, Pseudomonas aeruginosa* e *Serratia marcescens*. Isolados do CBC são os mais frequentemente associados aos sintomas da doença, e isolados de *B. cenocepacia* são predominantes na região do Vale do São Francisco. Estudos polifásicos incluindo análises de rep-PCR, perfis bioquímicos e patológicos e análise de sequência multilocus (MLSA) com os isolados demonstraram a predominância de isolados de *B. cenocepacia* linhagens IIIA e IIIB e de isolados pertencendo a duas supostas novas linhagens dessa bactéria, que aparentemente atuam exclusivamente como fitopatógenos. Assim, os objetivos desta tese foram sequenciar, montar e anotar cinco genomas de isolados de duas novas linhagens de *B. cenocepacia* associadas à podridão em escamas da cebola na região Nordeste do Brasil e analisar a posição taxonômica desses isolados por meio de uma abordagem filogenômica, e sequenciar, montar e anotar os genomas de isolados fitopatogênicos de *B. cenocepacia* pertencentes as linhagens IIIA e IIIB, visando a caracterização dos fatores de virulência e patogenicidade envolvidos na interação deste patógeno e a comparação com isolados obtidos de diferentes nichos ecológicos. A análise das sequências do core genoma revelaram que as duas supostas novas linhagens de *B. cenocepacia* pertencem ao CBC, porém não estão relacionados a nenhuma espécie já descrita deste complexo. Análise comparativa do genoma total dos cinco isolados contra os isolados tipo membros do CBC revelou valores médios de identidade de nucleotídeos (ANI) de 86,73 a 94,82%. Dados combinados de ANI, digital hibridização DNA-DNA (dDDH) e analise da sequência multilocus do core genome indicaram que estas supostas novas linhagens são, na verdade, duas novas espécies do CBC, as quais foram classificadas no presente estudo como *B. semiaridus* e *B. solum*. Adicionalmente, 12 isolados de *B. cenocepacia*, sendo cinco da linhagem IIIA e sete da linhagem IIIB isolados de cebola com sintomas de podridão em escamas foram comparados com 21 genomas completos isolados de patógenos humanos e do ambiente disponíveis em bancos de dados públicos. Através dos resultados de ANI, dDDH e cMLSA, foi possível constatar que os isolados fitopatogênicos das linhagens IIIA e IIIB apresentaram poucas variações em relação aos outros grupos estudados, contudo, de acordo com os resultados obtidos, essas duas linhagens se comportam como duas espécies distintas. Isolados fitopatogênicos de *B. cenocepacia* linhagem IIIA e IIIB apresentaram uma maior quantidade de fatores de patogenicidade e virulência a humanos. Até onde sabemos, este é o primeiro relato de um drafft genoma de *B. cenocepacia* causando podridão em escamas da cebola, e nossos dados serão, portanto, um recurso valioso para estudos futuros.

Palavras-Chave: *Burkholderia solum*, *Burkholderia semiaridus*, podridão em escamas, genômica comparativa

GENERAL ABSTRACT

In Brazil, the onion (*Allium cepa* L.) has high socioeconomic importance, standing out as the most produced vegetable within the *Allium* genus. Several diseases can affect onion growing, both during production and in the post-harvest process. Among those with high economic importance, the onion sour skin has great importance. This disease is associated with several bacterial species, among which species of the *Burkholderia cepacia* complex (BCC), such as *B. cepacia* and *B. cenocepacia*, in addition to *B. gladioli* pv. *alliicola*, *Pseudomonas aeruginosa* and *Serratia marcescens*. BCC strains are the most frequently associated with disease symptoms, and *B. cenocepacia* strains are predominant in the São Francisco Valley region. After a polyphasic approach including rep-PCR analyses, biochemical and pathological profiles, and multilocus sequence analysis (MLSA) with the strains, a predominance of *B. cenocepacia* lineages IIIA and IIIB and five strains belonging to two new lineages of this bacterium was observed, which act exclusively as phytopathogens. Thus, the aim of this thesis was to sequence, assemble and annotate five genomes of strains of two new strains of *B. cenocepacia* associated with onion sour skin in Northeastern Brazil and analyze the taxonomic position of these strains through a phylogenomic approach, and sequence, assemble and annotate the genomes of phytopathogenic strains of *B. cenocepacia* belonging to IIIA and IIIB lineages, aiming to characterize the virulence and pathogenicity factors involved in the plant-pathogen interaction and comparison with strains obtained from different ecological niches. The analysis of the core genome sequences revealed that they belong to the CBC, but they are not related to any previously described species of this complex. Comparative analysis of the total genome of the five strains against the CBC member-like strain revealed average nucleotide identity (ANI) values of 86.73 to 94.82%. Combined data from ANI, digital DNA-DNA hybridization (dDDH), and analysis of the core genome indicated that the strains IBSBF 3371^{T} and IBSBF 3372T represent two new species of the genus *Burkholderia*, belonging to the *Burkholderia cepacia* complex. Additionally, 12 strains of *B. cenocepacia*, five from lineage IIIA and seven from lineage IIIB isolated from onion, were compared with 21 complete genomes isolated from human pathogens and environmental groups. Through the ANI, dDDH, and core genome, we verify that the phytopathogenic strains from lineage IIIA and IIIB showed little variation in relation to the other studied groups. However, according to them to the results obtained, they behave as two distinct species. Phytopathogenic strains of *B. cenocepacia* lineage IIIA and IIIB showed a more significant amount of pathogenicity and virulence factors to humans. Our data will therefore be a valuable resource for future studies. In addition, our data contribute to the extensive genomic resources available for Burkholderiaceae species.

Keywords: *Burkholderia solum*, *Burkholderia semiaridus*, onion sour skin, comparative genomics

CAPÍTULO I

Introdução Geral

FILOGENÔMICA DE NOVAS ESPÉCIES DO COMPLEXO *Burkholderia cepacia* E GENÔMICA COMPARATIVA DE ISOLADOS DE *B. cenocepacia* LINHAGENS IIIA E IIIB, CAUSADORAS DE PODRIDÃO DAS ESCAMAS DA CEBOLA

INTRODUÇÃO GERAL

1. A cultura da cebola: aspectos culturais e econômicos

O gênero *Allium* pertence à classe Monocotyledoneae, ordem Asparagales e família Alliaceae (KILL *et al*., 2007). Seu centro de origem até o momento não foi totalmente esclarecido, porém, esta espécie é encontrada naturalmente na Ásia, na região entre o sul da Sibéria, no Golfo Pérsico e no Mar Cáspio. No entanto, através de análises filogenéticas da região ITS, foi possível concluir que a cebola comum (*Allium cepa*) tem como possível progenitor *Allium vavilovii* (GURUSHIDZE *et al*., 2007).

O consumo desta hortaliça pode ser realizado das mais diversas formas, desde *in natura*, em saladas, assim como em condimentos industrializados. No mundo, a cebola apresenta elevada importância socioeconômica, devido a características como sabor, pungência, aroma e propriedades terapêuticas, destacando-se como a cultura mais produzida dentro do gênero *Allium* (QUARTIERO *et al*., 2014). Além do consumo e valor econômico, a cebolicultura é considerada a terceira hortaliça com maior valor agregado do mundo, juntamente com a batata e o tomate (EL BALLA; HAMID; ABDELMAGEED, 2013).

A cebola é uma hortaliça herbácea, apresenta folhas grandes subcilíndricas tubulares (ocas), lisas e cobertas por uma camada cerosa. O caule verdadeiro situa-se na base do bulbo de onde partem as folhas e as raízes, e é composto por um disco achatado (prato), localizado abaixo da superfície do solo (KILL *et al*., 2007). No entanto, as sobreposições das bainhas das folhas formam o pseudocaule. Os bulbos são formados pelas bainhas carnosas das folhas e, nas partes externas, são envoltos por túnicas brilhantes de coloração variável (COSTA *et al*., 2002).

A produção mundial de cebola, em 2019, foi de 99,9 milhões de toneladas. Os principais produtores mundiais foram China (24,9 milhões), Índia (22,8 milhões), Estados

Unidos (3,1 milhões) e Egito (3,0 milhões), os quais produziram cerca de 53 milhões de toneladas (FAO, 2021).

Por sua vez, os estados produtores no Brasil são Piauí, Ceara, Rio Grande do Norte, Paraíba, Pernambuco e Bahia, na região Nordeste, Minas Gerais, Espírito Santo e São Paulo, na região Sudeste, Paraná, Santa Catarina e Rio Grande do Sul, na região Sul e por fim, os estados de Goiás e Distrito Federal na região Centro-Oeste. Sendo assim, o Brasil deteve em 2019 de uma produção de aproximadamente 1.556.885 toneladas de cebola, destacando-se o estado de Santa Catarina, com uma produção de 529.885 toneladas. A região Nordeste foi responsável pela produção de 396.594 toneladas, destacando-se a Bahia, com uma produção de 319.704 toneladas, e Pernambuco, com 54.542 toneladas produzidas (IBGE, 2021).

No Nordeste, a região do Submédio do Vale do São Francisco destaca-se como grande produtora de cebola. Nessa região, os principais municípios pernambucanos produtores desta hortaliça são Belém doSão Francisco, Cabrobó, Orocó, Petrolândia e Santa Maria da BoaVista. Por sua vez, no estado da Bahia, os principais municípios produtores de cebola são Abaré, Casa Nova, Curaçá, Itaguaçú, Juazeiro e Sento Sé (BAIA *et al*., 2021). Fora do entorno do Submédio do Vale do São Francisco, a cebolicultura também é realizada nos municípios baianos de Irecê, João Dourado e Mucugê (OLIVEIRA *et al*.,2019).

Pernambuco, além de sua representativa parcela na produção regional, é também um importante produtor de sementes cebola. O Instituto Agronômico Pernambuco (IPA) desde 1970 implantou um Programa de Melhoramento Genético da Cebola como objetivo desenvolver variedades adaptadas às condições edafoclimáticas e resistentes às principais pragas e doenças regionais, bem como dotadas de características compatíveis com as exigidas pelo mercado consumidor. Este programa vem sendo executado ininterruptamente até o presente momento. Como resultados foram desenvolvidas as variedades ValeOuro IPA 11 (bulbos amarelo) e Franciscana IPA 10 (bulbos roxo) (IPA, 2021).

Apesar dessa elevada produção, o Brasil não é autossuficiente e o alto consumo, associado às menores safras em algumas regiões em determinados períodos do ano, torna essencial a importação de cebola, principalmente da Argentina, Holanda e Espanha (SCHMITT, 2010). Além disso, diversas doenças podem acometer a cebola, tanto durante a produção quanto no processo de pós-colheita. Nesse sentido, dentre as doenças de

grande importância econômica, a podridão em escamas, causada por vários gêneros e espécies bacterianas, ocasionam perdas significativas podendo chegar até 50% durante a comercialização dos bulbos (WORDELL FILHO *et al*., 2006).

2. Podridão em escamas da cebola

A podridão em escamas da cebola tem sido observada em áreas de cultivo em váriospaíses, como Estados Unidos, Japão e Brasil (BURKHOLDER, 1950; OLIVEIRA *et al*.,2017; SOTOKAWA; TAKIKAWA, 2004). No Brasil, foi oficialmente relatada nos estados de São Paulo, Rio de Janeiro, Santa Catarina, Minas Gerais (MALAVOLTA JR. *et al*., 2008), Bahia e Pernambuco (OLIVEIRA *et al*., 2017). Os sintomas da doença caracterizam-se por uma podridão viscosa de coloração amarelada a marrom claro, firme, podendo haver a quebra de uma ou poucas escamas internas do bulbo (MOHAN, 1995). As escamas adjacentes às infectadas podem permanecer firmes e, embora sejam observadas podridões nas escamas individualmente, todo o bulbo é comprometido (DAVIS *et al*., 2014; ROBERTS, 2013).

Figura 1: Sintomas da podridão bacteriana das escamas (*B. cepacia*) (A) e da podridão escorregadia (*B. gladioli* pv. *alliicola*) (B) em bulbos de cebola. Fonte: SILVA *et al,* (2019).

O termo “podridão das escamas” tem sido utilizado como uma nomenclatura generalista atribuída a podridões em bulbos de cebola causadas por bactérias de diferentes gêneros e espécies, cuja sintomatologia e nome da doença podem variar em função da

bactéria envolvida (SILVA *et al*., 2019). As principais bactérias associadas a doença são espécies do complexo *Burkholderia cepacia* (CBC) (Palleroni e Holmes) Yabuuchi *et al*., *B. gladioli* pv. *alliicola* (Burkholder) Young *et al*., *Pseudomonas aeruginosa* (Schroeter) Migula (BAIA *et al.*, 2021; OLIVEIRA *et al*., 2019); WORDELL FILHO *et al*., 2006) e *Serratia marcescens* Bizio (MARQUES *et al*., 1994). Adicionalmente, isolados do CBC são os mais frequentemente associados aos sintomas da doença (JACCOUD FILHO *et al*., 1987), além de predominarem na região semiárida produtora de cebola (BAIA *et al*., 2021)e serem mais agressivos a bulbos de cebola (OLIVEIRA *et al*., 2019).

Dentre as diferentes podridões, destaca-se a podridão bacteriana das escamas, a qual é conhecida popularmente no Brasil como camisa d'água ou capa d'água (WORDELL FILHO *et al*., 2006) e está associada a espécies do CBC, como *B. cepacia* (Palleroni e Holmes) Yabuuchi *et al*. (BURKHOLDER, 1950), *B. cenocepacia* Vandamme *et al*. (BAIA *et al*., 2021; OLIVEIRA *et al*., 2017), *B. ambifaria* e *B. pyrrocinia* (TSUJI; KADOTA, 2020). A primeira ocorrência dessa doença foi realizada em 1950, em Nova York, sendo a causa atribuída à *B. cepacia* (sin. *Pseudomonas cepacia*). Nessa ocasião, a doença foi chamada no inglês de "*sour skin*", devido ao odor avinagrado exalado pelos catafilos mais externos, que se apresentavam apodrecidos (BURKHOLDER, 1950). Atualmente pouco se conhece a respeito da diversidade de isolados do CBC associados a essa doença no Brasil e no mundo. Essa ausência de informações não se deve exclusivamente a carência de estudos, mas também a elevada diversidade dos gêneros e espécies envolvidas com a doença (SILVA *et al*., 2019).

A etiologia da doença na região Nordeste do Brasil foi recentemente estudada por sequenciamento e análise filogenética do gene 16S rRNA e da técnica de rep-PCR, sendo observada a presença de bactérias do CBC, *P. aeruginosa* e *B. gladioli*. No entanto, devido à baixa resolução do marcador utilizado, não foi possível identificar quais espécies do CBC estavam associadas à podridão das escamas nesta região por análise de rep-PCR (OLIVEIRA *et al*., 2019). Posteriormente, após uma abordagem polifásica incluindo análises de rep-PCR, perfis bioquímicos e patológicos e análise de sequência multilocus (MLSA) com os isolados do CBC oriundos da região semiárida do Nordeste brasileiro, foi observada predominância de isolados de *B. cenocepacia* linhagens IIIA e IIIB e de duas novas linhagens dessa bactéria f. No entanto, apesar das evidências encontradas, as análises realizadas não foram suficientes para uma classificação específica dessas duas

novas linhagens. Além disso, curiosamente, nesses estudos nenhum isolado foi identificado como *B. cepacia*, demonstrando que essa espécie provavelmente não está associada à doença nesta região (BAIA *et al*., 2021).

A infecção geralmente ocorre no campo após aformação dos bulbos devido a disseminação das bactérias pelo solo ou água de irrigação,mas também pode ocorrer durante o armazenamento (PARKE; GURIAN-SHERMAN, 2001). Na presença de ferimentos, temperatura ideal (30 e 35°C) e presença de água livre,os danos causados pela podridão das escamas podem ser mais severos (DAVIS, *et al*., 2014). Além disso, é durante o armazenamento que grande parte das perdas acontece (PARKE; GURIAN-SHERMAN, 2001). Nesse contexto, os ferimentos causados durante o manejo da cultura e colheita ou quando as folhas caem durante a maturação dos bulbos, são porta de entrada para as bactérias. A infecção progride para escamas individuais e as escamas superficiais através das folhas infectadas, não sendo possível a locomoção para as escamas adjacentes, a menos que haja lesões (Davis, *et al*., 2014).

Para o manejo da doença é necessário ter conhecimento prévio da etiologia e epidemiologia da doença. Assim, as medidas de controle dependem da identificação adequada do agente causal envolvido, pois, do contrário, essas medidas podem ocasionar perda de tempo e dinheiro, levando a perdas ainda mais significativas. Portanto, o correto diagnóstico da doença e a identificação de seu agente causal são vitais para o sucesso do controle (SILVA, *et al*., 2019).

Recomenda-se o uso de bulbos e sementes sadias, evitarexcesso de umidade no solo, evitar a irrigação aérea perto da época da colheita, evitar a utilização de implementos agrícolas de onde a doença foi constatada, realizar rotação com espécies de outras famílias botânicas, controlar insetos pragas, eliminar plantas com sintomas no campo, após a colheita, armazenar os bulbos de cebola em baixas temperaturas, com baixa umidade do ar e em locais aerados evitará a propagação da podridão em escamas da cebola de bulbo para bulbo (GAVA; TAVARES, 2016; ROMEIRO, 2000). Ainda assim, deve-se ter o máximo cuidado nos tratos culturais durante o cultivo e manuseio dos bulbos desta hortaliça durante a colheita e armazenamento, evitando assim, qualquer choque que possa comprometer a integridade das escamas ou ferir as folhas próximo ao pescoço (WORDELL FILHO *et al*., 2006).

Sabendo que a propagação desta bactéria aumenta em temperaturas mais altas, não permitir que os bulbos maduros permaneçam no campo durante períodos de alta

temperatura associados com a espera da cura dos bulbos no campo por muito tempo. Além disso, é recomendado o uso de práticas que reduzam a chance de contaminação da água de irrigação com essa bactéria e a eliminação dos bulbos infectados deve ser realizada antes do armazenamento, pois a doença pode disseminar de bulbos infectados para bulbos saudáveis (SILVA, *et al*., 2019).

Ainda não existem variedades comprovadamente resistentes à podridão em escamas, o que poderia amenizar consideravelmente as perdas ocasionadas pela doença (GAVA; TAVARES, 2016). No entanto, estudo recente verificou a tolerância de 58 genótipos de cebolas a três isolados de *B. gladioli* pv. *alliicola*, e atestou que 37 genótipos apresentaram níveis de tolerância a esta bactéria, contudo apenas sete apresentaram estabilidade na tolerância aos três diferentes isolados deste patógeno. Diante desses resultados, os autores concluíram que esses sete genótipos se apresentam como fontes promissoras tolerantes à podridão escorregadia da cebola, podendo ser utilizados em programas de melhoramento visando à obtenção de variedades de cebola com resistência a doença (VELEZ *et al*., 2020).

3. Complexo *Burkholderia cepacia*

O gênero *Burkholderia* abriga espécies Gram-negativas, em forma debastonetes, oxidade positiva, aeróbicas, possuem células móveis, medindo entre 1,6 e 3,2μm e 8 a 10 μm, com flagelo monotríquio. Está inserido no reino Bacteria, filo Proteobacteria, classe Betaproteobacteria, ordem Burkholderiales e família Burkholderiaceae. Esta espécie tem um metabolismo diversificado, e está disseminada no meio ambiente, comumente dentro da rizosfera (COENYE, 2013; MAHENTHIRALINGAM; BALDWIN; DOWSON, 2008).

O primeiro registro da bactéria atualmente classificada como *B. cepacia* foi realizado em 1950, por Burkholder (1950), no estado de Nova York, ocasião em que o patógeno foi classificado como *P. cepacia*. Em 1980, esse epíteto foi excluído da *"Approved Lists of Bacterial Names*" (PARKE; GURIAN-SHERMAN, 2001), voltando a ser validado em 1981, após estudos taxonômicos (PALLERONI: HOLMES, 1981). Posteriormente, com baseem análises de hibridação DNA-DNA, o gênero *Burkholderia* foi proposto para acomodarespécies de *Pseudomonas* que compunham o grupo II de

homologia rRNA-DNA (YABUUCHI *et al.*, 1992), ocasião em que *P. cepacia* foi reclassificada como *B. cepacia*.

Devido à baixa sensibilidade das técnicas disponíveis para estudos taxonômicos até meados dos anos 90, isolados de diferentes espécies de *Burkholderia* intimamente relacionados foram erroneamente classificadas como *B. cepacia*. No entanto, a notável diversidade entre os isolados classificados nesta espécie foi observada por Vandamme *et al*., (1997) através de técnicas moleculares, que reportou que isolados de diferentes nichos ecológicos classificados como *B. cepacia* pertenciam a pelo menos cinco genomovares (o termo usado para se referir a espécies genômicas), a saber: *B. cepacia* (genomovar I), *B. multivorans* (genomovar II), *B. cenocepacia* (genomovar III), *B. estabilis* (genomovar IV) e *B. vietnamiensis* (genomovar V). Este grupo de cinco genomovares foi referido coletivamente como o complexo *B. cepacia* (COENYE *et al*., 2001).

Atualmente, o CBC é composto por 22 espécies distintas (*Burkholderia alpina*, *B. ambifaria*, *B. arboris*, *B. anthina*, *B. cepacia*, *B. cenocepacia*, *B. contaminans*, *B. diffusa*, *B. dolosa*, *B. latens*, *B. metallica*, *B. multivorans*, *B. puraquae*, *B. pseudomultivorans*, *B.pyrrocinia*, *B. seminalis*, *B. stagnalis*, *B. territorii*, *B. ubonensis*, *B. lata*, *B. estabilis* e *B.vietnamiensis*) (MARTINA *et al*., 2018) que apresentam altos níveis de similaridade nassequências dos genes 16S rDNA (98-100%) e *recA* (94-95%), níveis moderados de hibridização DNA-DNA (30-50%) (COENYE *et al*., 2001) e valores médios de identidade de nucleotídeos (ANI) do genoma abaixo de 90% (85,04 a 89,92%) (PEETERS *et al*., 2016). Espécies do CBC possuem um grande genoma que consiste emmúltiplos replicons cromossômicos circulares, contendo duas vezes o tamanho do material genético de *E. coli*. As espécies do CBC codificam muitos genes de virulência eresistência a antibióticos (SOUSA *et al*., 2017).

As espécies presentes no CBC, podem ser encontradas nos mais variados ambientes, como em água de rios (VANDAMME *et al*., 1997), solo (MARTINA *et al*., 2018), plantas (endofiticamente ou causando doença) (GILLIS *et al*., 1995), ambiente hospitalar (MARTINA *et al*., 2018) e humanos (COENYE *et al*., 2001), (VANDAMME *et al*., 2003).

4. Burkholderia cenocepacia

Burkholderia cenocepacia é um patógeno oportunista que causa infecções em indivíduos imunocomprometidos e em pacientes com fibrose cística (CHIARINI *et al*., 2006), o qual tem sido relatado como patógeno de cebola (BAIA *et al*., 2021; OLIVEIRA *et al*., 2017; TSUJI; KADOTA, 2020) e banana (LEE; CHAN, 2007). Com base em sequências do gene *recA*, *B. cenocepacia* quatro linhagens filogenéticas têm sido caracterizadas: IIIA, IIIB, IIIC e IIID (VANDAMME *et al*., 2003). As linhagens IIIA e IIIB têm sido encontradas com maior frequência em indivíduos com fibrose cística (GOLINI; CAZZOLA; FONTANA, 2006). Recentemente, estas linhagens também foram detectadas causando podridão das escamas no Nordeste brasileiro (BAIA*et al*., 2021; OLIVEIRA *et al*., 2019).

A linhagem IIIC é comumente encontrada em isolados do ambiente, diferentemente da linhagem IIID que é diretamente relacionada a isolados clínicos (MAHENTHIRALINGAM; BALDWIN; DOWSON, 2008). Outro fato interessante observado por BAIA *et al*. (2021) foi a descoberta de uma nova linhagem entre os isolados causadores de podridão em escamas da cebola no Nordeste brasileiro, além da existência de uma possível sinonímia entre a linhagem IIID e *B. contaminans*. No entanto, os mecanismos envolvidos na adaptação de *B. cenocepacia* à colonização de hospedeiros humanos e vegetais ainda não foram estudados.

Embora para isolados clínicos o conhecimento dos fatores de virulência e seus respectivos reguladores sejam bem compreendidos, os mecanismos envolvidos na adaptação de *B. cenocepacia* à colonização em plantas hospedeiras ainda precisam ser esclarecidos. Muitos fatores de virulência foram identificados em *B. cenocepacia*, incluindo enzimas extracelulares, toxinas, sistemas de secreções, sistemas de aquisição de ferro, comunicação célula-célula (*quorum sensing*), proteínas reguladoras, bem como genes que contribuem à motilidade, adesão, invasão celular, sobrevivência intracelular e proteção bacteriana dos fatores do hospedeiro (LOUTET; VALVANO, 2010). O fato de *B. cenocepacia* ser capaz de produzir uma ampla gama de fatores de virulência, resistir ao estresse oxidativo, persistir nos macrófagos, produção de toxinas, proteases e sistemas de secreção tipo VI, contribui para a relevância deste organismo (BARTELL *et al*., 2014).

Um estudo recente comparou isolados ambientais com isolados clínicos, revelando que existem alguns genes em isolados clínicos, como cablepilus (*cblA*), 22 kDa adhesion (*adhA*), transcriptional regulator (*kdgR*), bile acid 7-alpha, dehydratase (*baiE*), taurine dehydrogenase (*tauX*), sulfoacetaldehyde acetyltransferase (*xsc*), tellurite resistance cluster (*telA*), low oxygen activated locus (*lxa*), respiratory nitrate reductase cluster (*narIJHGK*), nitrate sensor e regulation cluster (*narLX*), que são ausentes em isolados ambientais (isolados do solo, rizosfera de plantas e água), demonstrando que existem diferenças na composição genética de acordo com o nicho ecológico do isolado (WALLNER *et al*., 2019). Também foi possível observar que alguns genes estavam presentes em maioria dos isolados ambientais e ausentes nos isolados clínicos, como é o caso dos genes lectin like bacteriocin 88 (*llpA*), nitrile hydratase cluster (*nthAB*), phenylacetaldoxime dehydratase (*oxd*), feruloyl-esterase (*faeB*), pyrrolnitrin biosynthesis cluster (*prn*), galacturonate metabolism genes (*uxaAB*) (WALLNER *et al*., 2019).

Em isolados fitopatogênicos, tem-se observado que a presença do gene da endopoligalacturonase *pehA* caracteriza-se como um fator de virulência e está relacionado à patogenicidade em cebolas. Por sua vez, a patogenicidade à cebola e a presença de *pehA* não tem sido observada em alguns isolados clínicos oriundos de pacientes com fibrose cística. Em contraste, um grupo relacionado consistindo em alguns isolados obtidos em solos e rizosfera de cebola apresentaram o gene *pehA* e foram patogênicos quando inoculados em cebola. Adicionalmente, a ausência do gene *pehA* e da capacidade de causar doença em cebola podem ser explicadas pela perda ancestral da *pehA* na maioria dos isolados clínicos (SPRINGMAN *et al*., 2009).

Um dos fatores de virulência mais conhecidas e essenciais em *B. cenocepacia* é a formação de biofilme. Biofilmes são comunidades de células inseridas em uma matriz de exopolissacarídeos dinâmica autoproduzida, caracterizada por uma estrutura tridimensional complexa. A matriz mantém as células juntas e acima de uma superfície e, eventualmente, as libera, resultando na colonização de outras superfícies (PELLIZZONI *et al*., 2016). O processo de formação de biofilme é parcialmente controlado pelo *quorum sensing*, um sistema de comunicação dependente da densidade celular que coordena a expressão de vários fatores de virulência (COENYE, 2010).

Nesse contexto, *B. cenocepacia* apresenta dois sistemas baseados em acil-homoserina lactona (AHL) chamados de CepIR e CciIR. O sistema CepIR é encontrado

em todos os isoladosdo CBC e é responsável pela regulação positiva de genes regulados por quorum-sensing. Por sua vez, o CciIR atua principalmente como um regulador negativo (O'GRADY *et al*., 2009). Outro sistema de detecção de quorum-sensing em *B. cenocepacia* usa o ácido cis-2-dodecenóico e é conhecido como "*Burkholderia* diffusible signal factor" (BDSF) como uma molécula de sinalização. O BDSF é detectado pelo *Rpf*R e sintetizado pelo *Rpf*F (SUPPIGER *et al*., 2013). Adicionalmente, há uma interação complexa entre a acil- homoserina lactona e os sistemas de *quorum sensing* baseados em BDSF (FAZLI *et al*., 2014).

Outro estudo recente mostrou que a metilação do DNA é uma função reguladora essencial para a expressão de genes relacionados à formação de biofilme, agregação celular e motilidade. Este estudo confirma que dois genes de DNA-metiltransferases (Mtases) estão envolvidos na regulação da expressão gênica pela estrutura do biofilme e agregação celular, causando também a redução da motilidade, confirmando que a metilação do DNA desempenha um papel essencial na regulação da expressão dos genes envolvidos na formação de biofilme e motilidade (VANDENBUSSCHE *et al*., 2020).

5. Filogenômica do gênero *Burkholderia*

Uma maneira reprodutível, confiável e altamente informativa para inferir relações filogenéticas entre procariotos caracteriza-se pelo uso da genômica através da utilização de ferramentas da bioinformática (ROSSELLÓ-MÓRA; AMANN, 2015).Nesse contexto, plataformas de sequenciamento de nova geração produzem massivamente dados de alto rendimento ao longo das últimas décadas. Estes dados são analisados e interpretados para inferir a classificação, identificação e caracterização genômica (CHUNG *et al*., 2018)

Estudos genômicos revelaram que *B. cenocepacia* tem um alto teor de G + C, em torno de 67%, múltiplos replicons, duplicações de genes, sequências de inserções e componentes móveis. Acredita-se que esses componentes aumentem a versatilidade dos genomas de *Burkholderia* e sua capacidade de obter um amplo escopo de vias metabólicas (LOUTET; VALVANO, 2010)

O sequenciamento genômico completo do isolado *B. cenocepacia* J2315, da linhagem IIIA, revelou que este membro do BCC contém um grande genoma com

múltiplos replicons de 8,06 Mb, consistindo em três cromossomos e um plasmídeo, que codificam um total de 7.261 quadros abertos de leitura (HOLDEN *et al*., 2009). O cromossomo 1 contém uma proporção maior de genes que codificam funcionalidades do núcleo, enquanto os cromossomos 2 e 3 contêm principalmente genes quecodificam para outros funcionalidades (HOLDEN *et al*., 2009). Uma vez que um genomabacteriano pode sofrer redução ou expansão como consequência da adaptação às mudanças nas condições ambientais, este tamanho relativamente grande do genoma levanta dúvidas sobre qual genes são indispensáveis e quais genes são necessários para asobrevivência e persistência em nichos diferentes (AUJOULAT *et al*., 2012)

Experimentos como identidade média de nucleotídeos (ANI) e hibridização DNA-DNA digital (dDDH) têm sido realizados para determinar a relação entre as bactérias. Estas foram algumas das poucas técnicas universalmente aplicáveis disponíveis que poderiam oferecer comparações genuínas de todo o genoma entre organismos. Essas duas abordagens avaliam toda a semelhança do genoma das bactérias, e o dDDH é um substituto rápido e preciso para o DDH convencional realizado em laboratório (COLSTON *et al*., 2014). Outra ferramenta disponível é o TYGS (*Type Strain Genome server*) um servidor de alto rendimento para taxonomia de procariontes baseada no genoma. Esta última técnica se baseia nos valores do dDDH como delimitador de espécies, estabelecido em 70%, além da distância entre os genomas, a qual é obtida pelo método GBDP (*Genome Blast Distance Phylogeny*), que realiza inferências filogenéticas utilizando genomas completos (HENZ, *et al*., 2005). Similarmente, a identidademédia de nucleotídeos (ANI) também apresenta alta resolução para delimitação de espécies de bactérias e arqueias, utilizando um limiar de 95 a 96% para classificação de espécies (RICHTER; ROSSELLÓ-MÓRA, 2009).

Além das técnicas citadas acima, a frequência de fragmentos de quatro nucleotídeos no genoma (TETRA) é sugerida como um parâmetro complementar na identificação de genomas em nível de espécie (ONG *et al*., 2016). Recomenda-se ainda que seja realizada a comparação genômica por um ou mais parâmetros, no caso do TETRA, para ser considerado da mesma espécie os valores precisam ser maiores que 0,998 - 0,999 a fim de garantir a delimitação de espécies que deverão ser verificadas de acordo com uma abordagem polifásica (RICHTER; ROSSELLÓ-MÓRA, 2009). Outra abordagem de alta confiabilidade é a utilização do pangenoma, que inclui análises do

core do genoma, o qual é composto por genes compartilhados por todas as linhagens (VERNIKOS *et al*., 2015; MEDINI *et al*., 2005).

Os avanços obtidos com a genômica comparativa para estudo dos isolados do CBC são evidentes. No entanto, estas técnicas ainda não foram aplicadas aos genomas dos isolados que causam podridão em escamas da cebola, evidenciando a necessidade de mais pesquisas para melhorar a compreensão das características intrínsecas desses organismos, como fatores de patogenicidade e virulência, além das relações taxonômicas e filogenéticas entre isolados de outros nichos ecológicos. Portanto, o presente estudo teve como objetivos: I. sequenciar, montar e anotar cinco genomas de isolados de duas novas linhagens de *B. cenocepacia* associadas à podridão em escamas da cebola na região Nordeste do Brasil e analisar a posição taxonômica desses isolados por meio de uma abordagem filogenômica; e II. sequenciar, montar e anotar os genomas de isolados fitopatogênicos de *B. cenocepacia* CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147 e CCRMBC157 pertencentes à linhagem IIIA, e dos isolados CCRMBC18, CCRMBC23, CCRMBC27, CCRMBC31, CCRMBC60, CCRMBC67 e CCRMBC109 pertencentes a linhagem IIIB, visando à caracterização dos fatores de virulência e e comparação com genomas de isolados de *B. cenocepacia* diferentes nichos ecológicos.

REFERÊNCIAS BIBLIOGRÁFICAS

AUJOULAT, F.; ROGER, F.; BOURDIER, A.; LOTTHÉ, A.; LAMY, B.; MARCHANDIN, H.; JUMAS-BILAK, E. From Environment to Man: Genome Evolution and Adaptation of Human Opportunistic Bacterial Pathogens. **Genes**, Basel, v. 3, n.2, p. 191–232, 2012.

BAIA, A. D. B.; SILVA, A. M. F.; RIBEIRO, B.G.; SOUZA, C.C.; SILVA JÚNIOR,W.; BALBINO, V.Q.; LEAL, C.M.; DE FARIAS, A.R. G.; SOUZA, E.B.; GAMA, M.A.S. Predominance of *Burkholderia cenocepacia* lineages causing onion sour skin inthe semi-arid region of north-east Brazil. **Plant Pathology**, Oxford, v. 70, n. 3, p. 521–533, 1 abr. 2021.

BARTELL, J. A.; YEN, P.; VARGA, J. J.; GOLDBERG, J. B.; PAPIN, J. A. Comparative Metabolic Systems Analysis of Pathogenic *Burkholderia*. **Journal of Bacteriology**, Washington, v. 196, n. 2, p. 210–226, 2014.

BURKHOLDER, W. H. Sour skin, a bacterial rot of Onion bulbs. **Phytopathology**, Saint Paul, v.40, n. 1, 1950.

CHIARINI, L.; BEVIVINO, A.; DALMASTRI, C.; TABACCHIONI, S.; VISCA, P. *Burkholderia cepacia* complex species: health hazards and biotechnological potential. **Trends in Microbiology**, Cambridge, v. 14, n. 6, p. 277–286, 2006.

CHUNG, M.; MUNRO, J. B.; TETTELIN, H.; DUNNING HOTOPP, J. C. Using Core Genome Alignments To Assign Bacterial Species. **mSystems**, Washington, v. 3, n. 6, 30 out. 2018.

COENYE, T. Social interactions in the *Burkholderia cepacia* complex: biofilms and quorum sensing. **Future Microbiology**, London, v. 5, n. 7, p. 1087–1099, 2010.

COENYE, T. The Family Burkholderiaceae. **The Prokaryotes: Alphaproteobacteria and Betaproteobacteria**, London, p. 759–776, 2013.

COENYE, T.; VANDAMME, P.; GOVAN, J. R.; LIPUMA, J. J. Taxonomy and Identification of the *Burkholderia cepacia* Complex. **Journal of Clinical Microbiology**, Washington, v. 39, n. 10, p. 3427–3436, 2001.

COLSTON, S. M.; FULLMER, M. S.; BEKA, L.; LAMY, B.; GOGARTEN, J. P.; GRAF, J. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. **mBio**, Washington, v. 5, n. 6, p. e02136, 2014.

COSTA, N. D.; QUEIROZ, M. D.; ARAÚJO, J. C.; SANTOS, C. A. F.; FARIA, C. D.; HAJI, F. N. P.; TAVARES, S. C. C. H. **A cultura da Cebola** - Coleção Plantar. p. 109, **Agricultural Water Management** 2002.

DAVIS, R. M.; AEGERTER, B. J.; LAEMMLEN, F. F.; VOSS, R. E. Onion and garlic, bacterial soft rot, Pathogens: *Erwinia carotovora* ssp. *carotovora*, *E. chrysanthemi*, *Pseudomonas gladioli*, and *Enterobacter cloacae*. California: University of California –UC. **Pest Management Guidelines**, 2014. Disponível em: <http://www.ipm.ucdavis.edu/PMG/r584100811.html>. Acesso em: 18 nov. 2021.

EL BALLA, M. M. A.; HAMID, A. A.; ABDELMAGEED, A. H. A. Effects of time of water stress on flowering, seed yield and seed quality of common onion (*Allium cepa* L.) under the arid tropical conditions of Sudan. **Agricultural Water Management**, Amsterdam, v.121, p. 149–157, 2013.

FAO (FOOD AND AGRICULTURE ORGANIZATION). FAOSTAT [online]. Rome: Food and Agriculture Organization, 2020. Disponível em: https://www.fao.org/faostat/en/#data/QCL/visualize. Acesso em: 18 nov. 2021.

FAZLI, M.; ALMBLAD, H.; RYBTKE, M. L.; GIVSKOV, M.; EBERL, L.; TOLKER-NIELSEN, T. Regulation of biofilm formation in *Pseudomonas* and *Burkholderia* species. **Environmental Microbiology**, Malden, v. 16, n. 7, p. 1961–1981, 2014.

GAVA, C. A. T.; TAVARES, S. C. C. H. **Cultivo da cebola no Nordeste: doenças**, 2016. Disponivel em: https://www.spo.cnptia.embrapa.br/conteudo?p_p_id=conteudoportlet_WAR_sistemasdeproducaolf6_1ga1ceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&p_r_p_-76293187_sistemaProducaoId=3113&p_r_p_-996514994_topicoId=1843 Accesso em: 18 nov. 2021.

GILLIS, M.; VAN, T. V.; BARDIN, R.; GOOR, M.; HEBBAR, P.; WILLEMS, A.; SEGERS, P.; KERSTERS, K.; HEULIN, T.; FERNANDEZ, M. P. Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N2-fixing isolates from rice in Vietnam. **International Journal of Systematic Bacteriology**, Salamanca, v. 45, n. 2, p. 274-289, 1995.

GOLINI, G.; CAZZOLA, G.; FONTANA, R. Molecular epidemiology and antibiotic susceptibility of *Burkholderia cepacia*-complex isolates from an Italian cystic fibrosis centre. **European Journal of Clinical Microbiology and Infectious Diseases**, v. 25, n.3, p. 175–180, 2006.

GURUSHIDZE, M.; MASHAYEKHI, S.; BLATTNER, F. R.; FRIESEN, N.; FRITSCH, R. M. Phylogenetic relationships of wild and cultivated species of *Allium* section *Cepa* inferred by nuclear rDNA ITS sequence analysis. **Plant Systematics and Evolution**, Viena, v. 269, n. 3, p. 259-269, 2007.

HENZ, S. R.; HUSON, D. H.; AUCH, A. F.; NIESELT-STRUWE, K.; SCHUSTER, S. C. Whole-genome prokaryotic phylogeny. **Bioinformatics**, Oxford, v. 21, n. 10, p. 2329–2335, 2005.

HOLDEN, M. T.; SETH-SMITH, H. M.; CROSSMAN, L. C.; SEBAIHIA, M.; BENTLEY, S. D.; CERDEÑO-TÁRRAGA, A. M.; THOMSON, N.R.; BASON, N.; QUAIL, M. A.; SHARP, S.; CHEREVACH, I.; CHURCHER, C.; GOODHEAD, I.; HAUSER, H.; HOLROYD, N.; MUNGALL, K.; SCOTT, P.; WALKER, D.; WHITE, B.; ROSE, H.; IVERSEN, P.; MIL-HOMENS, D.; ROCHA, E. P. C.; FIALHO, A. M.; BALDWIN, A.; DOWSON, C.; BARRELL, B. G.; GOVAN, J. R.; VANDAMME, P.; HART, C. A.; MAHENTHIRALINGAM, E.; PARKHILL, J. The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients. **Journal of Bacteriology**, Washington, v. 191, n. 1, p. 261–277, 2009.

IBGE – Índice Brasileiro de Geografia E Estatística. Indicadores IBGE. **Levantamento Sistemático da Produção Agrícola**. BRASIL, 2021. Disponível em: https://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_[mensal]/Fasciculo/2017/lspa_201712.pdf. Acesso em: 18 nov. 2021.

IPA – Instituto Agronômico de Pernambuco. **Cultura da cebola**. BRASIL, 2021. Disponível em: http://www.ipa.br/resp25.php. Acesso em: 18 nov. 2021.

KIILL, L.H.P.; RESENDE, G.M.; SOUZA, R.J. **Botânica. Cultivo da Cebola no Nordeste.** Sistemas de Produção, Embrapa. 2007. Disponível em: <http://www.cpatsa.embrapa.br:8080/sistema_producao/spcebola/botanica.htm>. Acesso em: 18 nov. 2021.

PARKE, J. L.; GURIAN-SHERMAN, D. Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. **Annual review of phytopathology**, Palo Alto, v. 39, n. 1, p. 225-258, 2001.

ROBERTS, S. Bacterial storage rots in onion caused by *Burkholderia gladioli* pv. *aliicola*. **Warwick: Plant Health Solution.** 2013. Disponível em: http://www.planthealth.co.uk/downloads/Bga_Onions_Poster_2013_A4.pdf. Acesso em:18 nov. 2021.

ROMEIRO, R. S. 2000. Doenças causadas por bactérias em alho e cebola. em: ZAMBOLIM, L.; VALE, F. X. R.; COSTA, H. (eds). **Controle de doenças de plantas hortaliças,** Viçosa: Editora UFV. p.43-81.

LEE, Y. A.; CHAN, C. W. Molecular Typing and Presence of Genetic Markers Among Strains of Banana Finger-Tip Rot Pathogen, *Burkholderia cenocepacia*, in Taiwan. **Phytopathology**, Saint Paul, v. 97, n. 2, p. 195–201, 2007.

LOUTET, S. A.; VALVANO, M. A. A Decade of *Burkholderia cenocepacia* Virulence Determinant Research. **Infection and Immunity**, Washington, v. 78, n. 10, p. 4088–4100, 2010.

MAHENTHIRALINGAM, E.; BALDWIN, A.; DOWSON, C. G. *Burkholderia cepacian* complex bacteria: opportunistic pathogens with important natural biology. **Journal ofApplied Microbiology**, Malden, v. 104, n. 6, p. 1539–1551, 1 jun. 2008.

MALAVOLTA JR.; V. A.; BERIAM, L. O. S.; ALMEIDA, I. M. G.; RODRIGUES NETO, J.; ROBBS, C. F. Bactérias fitopatogênicas assinaladas no Brasil: uma atualização. **Summa Phytopathologica**, Botucatu, v. 34, p. 9-88. 2008.

MARQUES, A. S. A.; ROBBS, C. F.; BOITEUX, L. S.; PARENTE, P. M. índice de fitobacterioses assinaladas no Brasil. Brasília; **EMBRAPA-SPI**, 1994. 65p.

MARTINA, P., LEGUIZAMON, M., PRIETO, C. I., SOUSA, S. A., MONTANARO, P., DRAGHI, W. O., STÄMMLER, M.; BETTIOL, M.; DE CARVALHO, C. C. C. R.; PALAU, J. FIGOLI, C.; ALVAREZ, F.; BENETTI, S.; LEJONA, S.; VESCINA, C.; FERRERAS, J.; LASCH, P.; LAGARES, A.; ZORREGUIETA, A.; LEITÃO, J. H.; YANTORNO, O.M.; BOSCH, A. *Burkholderia puraquae* sp. nov., a novel species of the *Burkholderia cepacia* complex isolated from hospital settings and agricultural soils. **International Journal of Systematic and Evolutionary Microbiology,** Reading, v. 68, n. 1, p. 14–20, 2018.

MEDINI, D.; DONATI, C.; TETTELIN, H.; MASIGNANI, V.; RAPPUOLI, R. The microbial pan-genome. **Current Opinion in Genetics & Development**, London, v. 15,p. 589-594, 2005.

MOHAN, S. K. Leaf streak and bulb rot. In: H. F.; MOHAN, S. K. (Eds). **Compendium of onion and garlic diseases**, Saint Paul: APS. 1995. v. 2., 31 p.

O'GRADY, E. P.; VITERI, D. F.; MALOTT, R. J.; SOKOL, P. A. Reciprocal regulation by the CepIR and CciIR quorum sensing systems in *Burkholderia cenocepacia*. **BMC Genomics**, London, v. 10, p. 441, 2009.

OLIVEIRA, W. J.; SILVA, W. A.; SILVA, A. M. F.; CANDEIA, J. A.; SOUZA, E. B.; MARIANO, R. L. R.; GAMA, M. A. S. First Report of *Burkholderia cenocepacia* Causing Sour Skin of Onion (*Allium cepa*) in Brazil. **Plant Disease**, Saint Paul, v. 101, n. 11, p.1950, 2017.

OLIVEIRA, W. J.; SOUZA, E. B.; SILVA, A. M. F.; LIMA, N. B.; LEAL, C. M.; CANDEIA, J. A.; GAMA, M. A. S. Elucidating the etiology of onion bacterial scale rot in the semi-arid region of Northeastern Brazil. **Tropical Plant Pathology**, Brasilia, v. 44, n. 6, p.494–502, 2019.

ONG, K. S.; AW, Y. K.; LEE, L. H.; YULE, C. M.; CHEOW, Y. L.; LEE, S. M. *Burkholderia paludis* sp. nov., an antibiotic-siderophore producing novel *Burkholderia cepacia* complex species, isolated from malaysian tropical peat swamp soil. **Frontiers in Microbiology**, Lausanne, v. 7, n. DEC, p. 2046, 2016.

PALLERONI, N. J.; HOLMES, B. *Pseudomonas cepacia* sp. nov., nom. rev.

International journal of systematic bacteriology, Reading, v. 31, n. 4, 479-481, 1981.

PARKE, J. L.; GURIAN-SHERMAN, D. Diversity of the Burkholderia Cepacia Complex and Implications for Risk Assessment of Biological Control Strains. **Annual Review of Phytopathology**, Palo Alto, v. 39, n. 1, p. 225–258, 2001.

PEETERS, C.; MEIER-KOLTHOFF, J. P.; VERHEYDE, B.; DE BRANDT, E.; COOPER, V. S.; VANDAMME, P. Phylogenomic Study of *Burkholderia glathei*-like

Organisms, Proposal of 13 Novel *Burkholderia* Species and Emended Descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis*, and *Burkholderia grimmiae*. **Frontiers in Microbiology**, Lausanne, v. 7, n. -, p. 877, 2016.

PELLIZZONI, E.; RAVALICO, F.; SCAINI, D.; DELNERI, A.; RIZZO, R.; CESCUTTI, P. Biofilms produced by *Burkholderia cenocepacia*: influence of media and solid supports on composition of matrix exopolysaccharides. **Microbiology,** Reading, v. 162,n. 2, p. 283–294, 2016.

QUARTIERO, A.; FARIA, M. V.; RESENDE, J. T. V.; FIGUEIREDO, A. S. T.; CAMARGO, L. K. P.; SANTOS, R. L.; KOBORI, R. F. Desempenho agronômico, heterose e estabilidade fenotípica de genótipos de cebola. **Horticultura Brasileira**, Vitória da Conquista, v.32, n. 3, p. 259–266, 2014.

RICHTER, M.; ROSSELLÓ-MÓRA, R. Shifting the genomic gold standard for the prokaryotic species definition. **Proceedings of the National Academy of Sciences ofthe United States of America**, Washington, v. 106, p. 19126–31, 2009.

ROSSELLÓ-MÓRA, R.; AMANN, R. Past and future species definitions for Bacteriaand Archaea. **Systematic and Applied Microbiology**, Stuttgart v. 38, n. 4, p. 209–216, 1 jun. 2015.

SILVA, A. M. F.; BAIA, A. D. B.; VELEZ, L. S.; OLIVEIRA, W. J.; GAMA, M. A. S. Diversidade taxonômica e patológica de espécies de *Burkholderia* causadoras de podridão das escamas da cebola. **Revisão Anual de Patologia de Plantas**, São Paulo, v. 26, p. 81-95, 2018.

SCHMITT, D. R. **Cebola: produção e mercado nacional.** Síntese Anual da Agricultura de Santa Catarina - 2010-2011. Epagri/Cepa, Santa Catarina, SC. 2010.
SOTOKAWA, N.; TAKIKAWA, Y. Occurrence of bacterial rot of onion bulbs causedby

Burkholderia cepacia in Japan. **Journal of General Plant Pathology**, Tokyo, v.70, n. 6, p. 348–352, dez. 2004.

SOUSA, S. A.; FELICIANO, J. R.; PITA, T.; GUERREIRO, S. I.; LEITÃO, J. H. *Burkholderia cepacia* Complex Regulation of Virulence Gene Expression: A Review. **Genes**, Basel, v. 8, n. 1, p. 43, 2017.

SPRINGMAN, A. C.; JACOBS, J. L.; SOMVANSHI, V. S.; SUNDIN, G. W.; MULKS, M. H.; WHITTAM, T. S.; VISWANATHAN, P.; GRAY, R. L.; LIPUMA, J. J.; CICHE, T. A. Genetic Diversity and Multihost Pathogenicity of Clinical and Environmental Strains of *Burkholderia cenocepacia*. **Applied and Environmental Microbiology,** Washington, v. 75, n. 16, p. 5250–5260, 2009.

SUPPIGER, A.; SCHMID, N.; AGUILAR, C.; PESSI, G.; EBERL, L. Two quorum sensing systems control biofilm formation and virulence in members of the *Burkholderia cepacia* complex. **Virulence**, Austin, v. 4, n. 5, p. 400–409, 2013.

TSUJI, M.; KADOTA, I. Identification and phylogenetic analysis of *Burkholderia cepacia* complex bacteria isolated from rot of onion bulbs in Tohoku region of Japan. **Journal of General Plant Pathology**, Tsushima, v. 86, n. 5, p. 376–386, 25 jun. 2020.

VANDAMME, P.; HOLMES, B.; COENYE, T.; GORIS, J.; MAHENTHIRALINGAM, E.; LIPUMA, J. J.; GOVAN, J. R. *Burkholderia cenocepacia* sp. nov.--a new twist to an old story. **Research in Microbiology**, Paris, v. 154, n.2, p. 91–96, 2003.

VANDAMME, P.; HOLMES, B.; VANCANNEYT, M.; COENYE, T.; HOSTE, B.; COOPMAN, R.; REVETS, H.; LAUWERS, S.; GILLIS, M.; KERSTERS, K.; GOVAN, J. R. Occurrence of Multiple Genomovars of *Burkholderia cepacia* in Cystic Fibrosis Patients and Proposal of Burkholderia multivorans sp. nov. **International**

Journal of Systematic and Evolutionary Microbiology, Reading, v. 47, n. 4, p. 1188–1200, 1997.

VANDENBUSSCHE, I.; SASS, A.; PINTO-CARBÓ, M.; MANNWEILER, O.; EBERL, L.; COENYE, T. DNA methylation epigenetically regulates gene expression in *Burkholderia cenocepacia*. **mSphere**, Washington, v. 5, n. 4, p. e00455-20, 2020.

VELEZ, L. S.; SILVA, A. M. F.; SANTOS, C. A. F.; ASSUNÇÃO, E. F.; SILVA, M. S.; SOUZA, E. B.; GAMA, M. A. S. Evaluation of onion genotypes to slippery skin caused by *Burkholderia gladioli* pv. *alliicola*. **Horticultura Brasileira**, Vitória da Conquista, v. 38, n. 4, p.350–355, 1 out. 2020.

VERNIKOS, G.; MEDINI, D.; RILEY, D. R.; TETTELIN, H. Ten years of pan-genome analyses. **Current Opinion in Microbiology**, London, v. 23, p. 148-154, 2015.

WALLNER, A.; KING, E.; NGONKEU, E. L. M.; MOULIN, L.; BÉNA, G. Genomic analyses of *Burkholderia cenocepacia* reveal multiple species with differential host-adaptation to plants and humans. **BMC Genomics**, London, v. 20, n. 1, p. 803, 2019.

WORDELL FILHO, J. A.; BOFF, P. "Doenças de origem parasitária" In: WORDELL FILHO, J. A.; ROWE, E.; GONÇALVES, P. A. S.; DEBARBA, J. F.; BOFF, P.; THOMAZELLI, L. F. **Manejo fitossanitário da cebola. Florianopolis**: EPAGRI,2006. p. 19 - 126.

YABUUCHI, E.; KOSAKO, Y.; OYAIZU, H.; YANO, I.; HOTTA, H.; HASHIMOTO, Y.; EZAKI, T.; ARAKAWA, M. 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. **Microbiology and Immunology**, New Jersey, v. 36, n. 12, p. 1251–1275, 1992.

CAPÍTULO II

Burkholderia semiaridus sp. nov. and *Burkholderia solum* sp. nov., two novel *B. cepacia* complex species causing onion sour skin

***Burkholderia semiaridus* sp. nov. and *Burkholderia solum* sp. nov., two novel *B. cepacia* complex species causing onion sour skin.**

Leandro da Silva Velez[1]; Flavia Figueira Aburjaile[2]; Antonio Roberto Gomes de Farias[3]; Ana Dulce Botelho Baia[1]; Willams José de Oliveira[4]; Adriano Marcio Freire Silva[4]; Ana Maria Benko-Iseppon[2]; Vasco Ariston de Carvalho Azevedo[5]; Bertram Brenig[6]; Elineide Barbosa Souza[7]; *Marco Aurélio Siqueira da Gama[1]

[1]Departamento de Agronomia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil
[2]Laboratório de Genética e Biotecnologia Vegetal, Universidade Federal de Pernambuco, Recife, 50670-901, Pernambuco, Brazil
[3]Center of Excellence in Fungal Research, School of Science, Mae Fah Luang University, Chiang Rai, 57100, Thailand
[4] Estação Experimental de Cana-de-Açúcar do Carpina, Universidade Federal Rural de Pernambuco/ Carpina, PE, Brasil
[5]Laboratório de Genética Molecular e Celular, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Minas Gerais, Brazil
[6]University of Göttingen, Institute of Veterinary Medicine, Göttingen, Germany
[7]Departamento de Biologia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil

*Correspondence: Marco Aurélio Siqueira da Gama, Departamento de Agronomia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil. Email: marco.gama@ufrpe.com.br

Abstract

Two new *Burkholderia cenocepacia* lineages found in Northeastern Brazil causing scale rot on the central onion-producing regions in Pernambuco and Bahia states were studied using genomic approaches to determine their positioning taxonomic. Five strains, four belonging to one new lineage and one strain belonging to another new lineage, had their whole genome sequenced to carry out a phylogenomic analysis.

Genome sizes varied from 7.5 to 8.1 Mb, with G+C content varying from 66.66 to 67.03%. Comparative whole-genome analysis of the five strains analyzed in this study showed average nucleotide identity (ANI) values ranging from 86.73 to 94.82% concerning the type strains of Bcc. dDDH values ranging from 25.0 to 55.9% were observed for all isolates studied. The phylogenomics performed with the core genome of the five isolates demonstrated two different clades among the isolates CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, and the isolate CCRMBC51. Combined data from ANI, dDDH, and core genome analyses indicated that the strains represent two novel species of the genus *Burkholderia*, belonging to *Burkholderia cepacia* complex (Bcc), which were reclassified as *B. semiaridus* and *B. solum*.

1 **INTRODUCTION**

The bacterium *Burkholderia cepacia* (synonym *Pseudomonas cepacia*) was initially described in 1950 as the causal agent of onion sour skin (Burkholder, 1950). Over years, strains phenotypically similar obtained from diverse ecological niches were classified as *B. cepacia* (Coenye et al., 2001). These strains were posteriorly reclassified at five distinct genomic species, which passed to be referred to as the *B. cepacia* complex (Bcc) (Vandamme et al., 1997). Nowadays, Bcc is composed of a versatile group of 22 closely related species (Martina et al., 2018), which may be satisfactorily identified using multilocus sequence analysis (MLSA) with housekeeping genes (Baia et al., 2021; Peeters et al., 2013; Tsuji & Kadota, 2020) and genomic analyses (Santos et al., 2018; Vanlaere et al., 2009; Wallner et al., 2019).

Although most studies about the diversity of Bcc species have been carried out with strains obtained from clinical environments and environmental samples (Coenye et al., 2001, Peeters et al., 2013, Martina et al., 2018, Spilker et a., 2009, Vandamme et al., 1997, Vanlaere et al., 2009), the knowledge about the diversity of species causing onion sour skin has been improved and other Bcc species beyond *B. cepacia* have also been associated with this disease in the last years (Baia et al., 2021; Oliveira et al., 2017; Oliveira et al. 2019; Tsuji and Kadota, 2020). In this context, strains of *B. cepacia*, *B. cenocepacia*, *B. ambifaria*, and *B. pyrrocinia* have been recently associated with the rot of onion bulbs in Japan (Tsuji & Kadota, 2020), where this disease was reported since

2001 (Sotokawa and Takikawa, 2004). In turn, since 2015, strains of the Bcc have also been identified in the main onion-producing regions of north-east Brazil (Oliveira et al., 2017; Oliveira et al., 2019), and MLSA with housekeeping genes have revealed the predominance of strains of *B. cenocepacia* causing onion sour skin (Baia et al., 2021).

Taking into consideration Bcc species associated with the rot of onion bulbs so far, *B. cenocepacia* stands out because its behaver is a genetically heterogeneous group composed of at least four phylogenetic lineages (IIIA, IIIB, IIIC, and IIID) (Mahenthiralingam et al., 2000; Vandamme et al., 2003). Among them, the lineages IIIA and IIIB have been found both, clinical isolates (Wallner et al. 2019) and onion sour skin (Baia et al., 2021), while lineages IIIC and IIID have been detected exclusively in soil and human clinical settings, respectively (Vandamme et al., 2003).

Based on comparisons between genomes of isolates from lineages IIIA and IIIB, it was recently proposed that isolates from the former lineage should be classified as *B. cenocepacia sensu stricto*. In contrast, isolates from the latter lineage were reclassified as *B. servocepacia* (Wallner et al., 2019). In turn, the high percentage of divergence of the concatenated sequence of seven housekeeping genes (*recA*, *gyrB*, *atpD*, *gltB*, *lepA*, *phaC*, and *trpB*) has indicated that lineages IIIC and IIID may be distinct species. This hypothesis is reinforced because the lineage IIID and *B. contaminans* show low percentage divergence, indicating a synonymy (Baia et al., 2021). In addition, two new *Burkholderia cenocepacia* lineages were recently found causing onion sour skin in the semi-arid region of northeast Brazil (Baia et al., 2021), but their taxonomic status has not yet been elucidated. Therefore, this study aims to investigate the taxonomic positioning of these new lineages using a genomic approach to classify them appropriately within the Bcc.

2 **MATERIALS AND METHODS**

2.1 Bacterial strains and data set preparation

Four (CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171) and one (CCRMBC51) strain belonging to two new *B. cenocepacia* lineages associated with onion sour skin in the semi-arid region of north-east Brazil (Baia et al. 2021) were obtained from Phytopathogenic Bacterial Culture Collection of the Phytobacteriology Laboratory

(LAFIBAC) of the Federal Rural University of Pernambuco (UFRPE) and used for the whole-genome DNA sequencing. Additionally, 26 whole-genome sequences of the type strains of all available Bcc species at the NCBI database (by May 2021) were included in the dataset. We also included whole-genome sequences of *B. gladioli*, *B. glumae*, and *Paraburkholderia fungorum* as outgroup (Table 1). For the analysis, the genomes were filtered using as the criteria the N50 > 20.000 and/or BUSCO (presence of conserved genes) > 90%, which is implemented in the QUAST software v. 5.0.2 (Gurevich et al. 2013).

2.2 Whole-genome DNA sequencing, assembly, and annotation

The strains' DNA was extracted using the MiniPrep kit for bacterial genomic DNA extraction (Axygen Biosciences, Massachusetts, USA) following the manufacturer's recommendations. The quality and quantity of DNA samples were assessed, and DNA concentration was adjusted for 10 ng μl^{-1} using a Biodrop (Biodrop^{TM}, Cambridge, Canada).

The genome sequencing was performed in Illumina HiSeq 2500 platform (Illumina, EUA), and libraries were paired-sequenced (2x 150 bp) at the University of Göttingen, Institute of Veterinary Medicine, Göttingen, Germany. The quality of raw sequencing data was evaluated using FastQC v0.11.8 (Andrews 2010). Adapter screening and quality filtering of reads were performed using Trimmomatic 0.36 (Bolger, Lohse, and Usadel 2014). *De novo* genome assemblies were performed using SPAdes v3.13.0 (Bankevich et al. 2012), selecting the careful option. The quality of assemblies was checked using QUAST v5.0.2 (Gurevich et al. 2013). The assembled contigs were automatically annotated using the PROKKA software v.1.13.3 (Seemann 2014). In addition, the draft genome sequences of the strains CCRMBC16, CCRMBC33, CCRMBC51, CCRMBC74, and CCRMBC171 were deposited as whole-genome shotgun projects in GenBank under the accession numbers SAMN19592385, SAMN19592386, SAMN19592387, SAMN19592388, and SAMN19592389 (Table 1), respectively.

2.3 Taxogenomic analysis

Taxonomic affiliations of the isolated *Burkholderia* spp. were inferred at the genomic level using the Type (Strain) Genome Server (TYGS), a free bioinformatics

platform available under https://tygs.dsmz.de (accessed on May 21, 2021), for a whole genome-based taxonomic analysis (Meier-Kolthoff and Göker 2019). The tree was inferred with FastME 2.1.6.1 (Lefort et al. 2015) from genome BLAST distance phylogeny (GBDP), calculated from genome sequences. The branch lengths were scaled in terms of GBDP distance formula d5.

Average nucleotide identity (ANI) and tetranucleotide frequency correlation coefficient (TETRA) were calculated by pyani v.0.10 Python3 (Pritchard et al. 2016) using whole-genome alignments through mummer. In turn, digital DNA–DNA hybridization (*d*DDH) values among the whole-genomes were computed using the web server genome-to-genome distance calculator v.2.1 http://ggdc.dsmz.de/distcalc2.php for genome alignments, under the recommended Formula 2 (DDH estimates based on identities/HSP - High Scoring Pair length), with the alignment tool BLAST+ (Meier-Kolthoff et al. 2014).

2.4 Phylogenomic Analysis

Pan-genomic analyses were performed without the outgroup, using only the Bcc species aiming at high precision to obtain the core genes using Roary v3.13.0 (Page et al. 2015). At this point, .gff files were used as input for Roary pan-genome analysis. For the genome downloaded from the RefSeq/NCBI, the .gbk files were converted to .gff format. In turn, for the genomes sequenced in this study (CCRMBC16, CCRMBC33, CCRMBC51, CCRMBC74, and CCRMBC171) were used the .gff files outputted by PROKKA. A cut-off of 95% identity was used. The resulting orthologous were classified as core-, softcore-, shell-, and cloud-genes according to the software's default configuration. The core genes were automatically aligned with MAFFT v. 7.310 (Katoh and Standley 2013), using the *-e --mafft* parameter in Roary, creating a multi-FASTA alignment of the nucleotide sequence of all the core genes. This file was then used as input to build a maximum-likelihood (ML) phylogenomics by IQ-TREE software (version 2.0.4) (Nguyen et al. 2015) using the automatic selection of nucleotide substitution model on ModelFinder for selection of the best fitting DNA substitution model (Kalyaanamoorthy et al. 2017). The node support was assessed with ultrafast bootstrap (Minh, Nguyen, and von Haeseler 2013) using 100,000 replicates. The ML tree was visualized using the FigTree software v1.4.3 (Rambaut, 2009) (http://tree.bio.ed.ac.uk/software/figtree/). Additionally, the genetic distances were

estimated using the Kimura 2-parameter matrix and constructed using MEGA v.10.2.6 software (Kumar et al. 2018).

3 **RESULTS**

3.1 Whole-genome DNA sequencing, assembly, and annotation

The *de novo* genome assemblies showed a total of 15, 13, 19, and 16 contigs larger than 500 bp to the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, while 61 contigs were observed for the strain CCRMBC51 (Supplementary Table 1). They also showed N50 values ranging from 1.388.868 to 2.176.560 bp among the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, resulting in genome sizes ranging from 7,523,028 to 7,575.069 bp, while the strain CCRMBC51 showed an N50 value of 515,855 and a genome of 8,114,250 bp. The alignment annotated resulted in coding sequences (CDSs) ranging from 6.734 to 6.788 to the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC17. In turn, the number of CDSs in the CCRMBC51 strain was 7.833. The genomic annotation revealed the presence of 6.852, 6.832, 6.861, and 6.811 total genes in strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, showing several CDSs with 6.646, 6.601, 6.635, and 6.582 proteins, respectively. In turn, the strain CCRMBC51 presented a total number of 7.918 genes and CDSs with 7.318 protein, representing around 6.78% of the annotated genes.

3.2 Taxogenomic analysis

The phylogenomic tree from the Bcc species shared the strain CCRMBC51 and strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 in two different clades, which were not clustered with any known species of the Bcc group (Figure 1). Pairwise comparison based on TYGS using the genomes of these strains showed digital DDH (d4) values lower than 56.6% between strain CCRMBC51 and the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 (Supplementary Table 2). Similarly, digital DDH values among genomes of the strain CCRMBC51 and type strains of the BCC were lower than 57.1%. In turn, digital DDH values among the genomes of the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 were higher than

93.2%, while digital DDH values among the genomes of these strains and type strains of the BCC were lower than 55.5%. Additionally, both CCRMBC51 and strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 showed the highest values of digital DDH concerning *B. cenocepacia* lineage IIIA (J2315), which ranged from 55.2% to 57.1%.

ANIm analysis of the 34 genomes of the *Burkholderia* genus showed an average similarity value of 91,74% among all genomes (Figure 2). A high level of similarity was observed among the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, which showed values ranging from 99.2 to 99.9%. Regarding the strain CCRMBC51, the ANIm values ranged from 94.36 to 94.49 in relation to the other four strains analyzed in this study. The strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 showed values below 94.42% compared to all Bcc type strains analyzed. Similarly, the strain CCRMBC51 showed ANIm values below 94.82% in relation to all Bcc type strains. All strains analyzed in this study showed ANI values below 85.38% in relation to species used as outgroup (*B. gladioli*, *B. glumae*, and *P. fungorum*).

The TETRA coefficients between most species of the BCC were above 0.99, including the strain CCRMBC51 and strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 (Supplementary Table 3). However, when using the genomes of the type strains of *B. dolosa*, *B. multivorans*, *B. paludis*, *B. pseudomultivorans*, *B. paraquae*, *B. stagnalis*, *B. thailandensis*, *B. ubonensis*, and *B. vietnamiensis* in comparisons, TETRA coefficients ranged from 0.94 to 0.99. In turn, TETRA coefficients among genomes of the Bcc strains and genomes of the type strains of *B. gladioli*, *B. glumae*, and *P. fungorum* were always below 0.94.

Regarding dDDH analysis, values below 56,6% were observed between the strain CCRMBC51 and strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, while values below 59.0% were observed between the strain CCRMBC51 and Bcc species (Figure 2). In turn, the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 showed dDDH values ranging from 93.2 to 99.5%, while dDDH values between these strains and Bcc species were below 56.0%. Values below 25.1% were observed between all strains analyzed in this study and species used as outgroup. Considering the cut-off value (>70%), the studied genomes agree with all others analyzes carried out, they presented values below 70%, varying from 25.0 to 55.9%.

3.3 Phylogenomic analysis

Analysis of the core and accessory genomes revealed the presence of 90.015 genes, with 309 core genes (99% in ≥ 100% of strains), 592 soft-core genes (95% in ≥ 99% of the strains), 6.023 shell genes (15% in ≥ 95% of the strains), and 83.091 cloud genes (0% in less than 15% of the strains). The phylogenetic ML tree constructed with the core genes of the strains sequenced in this study and Bcc species provided one true reproduction of the current taxonomic positions of these species, with the strain CCRMBC51 and strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 grouping in two exclusive clades with 79,4% and 100% bootstrap, respectively (Figure 3). These clades did not cluster with any known species of the Bcc and both clades clustered more closely to *B. cenocepacia* lineage IIIB.

4 DISCUSSION

Five strains associated with onion sour skin in the semi-arid region of northeast Brazil and described as new lineages of *B. cenocepacia* (Baia et al. 2021) were better analyzed using genomic approaches to find the best taxonomic positioning within Bcc. Among them, four strains (CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171) were characterized as a new lineage, while only one strain (CCRMBC51) belonged to another new lineage. As for the genomic characteristics found in the five genomes from onion bulbs with symptoms of sour skin, it was possible to verify that the strain CCRMBC51 showed a greater genome size and number of CDSs than the CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 strains. However, the G+C content was lower compared to the four other genomes studied. Based on these parameters, it was possible to verify that these values are like those obtained by other studies, within the range reported for other species of Bcc (66-69%) (Vandamme and Dawyndt 2011; Martina et al. 2018; Wallner et al. 2019).

Previous research has shown that using the dDDH cut-off 70% for species delimitation is generally more stringent than the ANI species cut-off (95-96%) (Chung et al., 2018). In this study, we used 70% dDDH and an upper boundary of 95% to classify the species.

Based on TYGS, the strains CCRMBC16, CCRMBC33, CCRMBC74, CCRMBC171, and the strain CCRMBC51 were satisfactorily classified as two new

species (Figure 1). Using TYGS, we checked that the strains CCRMBC16, CCRMBC33, CCRMBC74, CCRMBC171 grouped into the same clade with 87% bootstrap, while the strain CCRMBC51 formed a separate clade, with 84% bootstrap. When analyzing the dDDH data, it was possible to verify that they agree with those already analyzed by TYGS. dDDH values observed among the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, were above 93.2. In turn, dDDH values below 59% were observed when these five strains were compared concerning Bcc species, indicating that the strains analyzed in this study do not belong to any known species of the Bcc.

The TYGS data contains the pairwise dDDH values between the studied genomes and the selected type strain genome and dDDH values are provided along with their confidence intervals for the GBDP formula d4, which is independent of genome length and has been considered as robust to be used for the analysis of incomplete draft genomes (Boxberger et al., 2020). The values of dDDH found by TYGS indicated that these five strains belong to two new species within Bcc because dDDH values were consistently below 70% concerning Bcc species, which is used as a threshold to species classification (Jin et al., 2020).

ANI values observed among the strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171, were above 99.21%, while ANI values observed among these strains and the strain CCRMBC51 were below 94.36%. In turn, ANI values below 95% were observed when these five strains were compared concerning Bcc species, indicating that the strains analyzed in this study do not belong to any known species of the Bcc (Ritcher and Rosselo-Mora, 2009). In addition, when analyzed against the outgroup (strains outside Bcc), ANI values were below 85.96% for CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 strains, and below 85.89% for CCRMBC51 strain. The values observed in these analyses were, in general, like those observed in the TYGS analysis.

Interestingly, two ANI analyses, one based on BLAST+ (ANIm) and the other on MUmmer (ANIm) were performed, and for the strains used in this study, the comportment was the same. However, the delimitation between the species *B. cenocepacia* lineage IIIA and IIIB showed divergences between these approaches. In ANIb analysis, the value was 0.94, below the cut-off value to consider the same species. However, this value in ANIm analysis was 0.95, which can already be considered the same species. This result was also observed in a recent study by Wallner et al. (2019), where lineage IIIB was reclassified

as a new species, named *B. servocepacia*. Nevertheless, other authors used ANI values to classify Bcc species above 96% (Jin et al., 2020).

Regarding the TETRA correlation coefficients results, this analysis does not have a suitable resolution for species separation within Bcc because it could not differentiate the type Bcc species analyzed in this study. In other words, the cut-off value for species delimitation through this analysis is 0.99 (Richter and Rosselló-Móra 2009). For the isolates analyzed in this study, values below 0.99 were only possible to verify when analyzed against the outgroup and *B. multivorans*, *B. dolosa*, *B. ubonensis*, and *B. vietnamienses*. Nevertheless, TETRA coefficients satisfactorily may be used for species delimitation for the *Xanthomonas* genus (Gama et al., 2018).

A set of 309 concatenated single-copy core gene alignments composed of orthologous genes present in one copy in each analyzed genome was used to estimate the phylogeny of these bacteria by maximum likelihood. The phylogenetic analysis revealed that the closest neighbors were the strains CCRMBC51 and the group formed by strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171. The strains CCRMBC16, CCRMBC33, CCRMBC74, and CCRMBC171 grouped in one cluster within the Bcc, supported by bootstrap values of 100% (Figure 3), while another cluster was formed only by strain CCRMBC51, which was grouped closest with the strain AU1054 from *B. cenocepacia* lineage IIIB.

In addition, based on the TYGS, ANI, dDDH, and core-genome phylogeny performed in this study, we conclude that the five strains found causing sour skin in onion bulbs in the semiarid region of north-east Brazil belong to two new species of the Bcc, which were formerly described below. Using a whole-genome approach for the classification and delimitation of species provides a better approach, as was done in this study. We use core phylogeny based on single-copy orthologous genes and pangenome based on ANI and dDDH. These same approaches were used and recommended to Bcc taxonomic studies by Jin et al., (2020).

For all analyses, genomes of type strains available at the NCBI were used. Among the strains, there are some representing *B. cenocepacia* lineages IIIA (J2315^{T}) and IIIB (AU1054). However, as no genomes are available for lineage IIIC and IIID, the inclusion of these lineages is unfeasible in the present work, as observed by Jin et al. (2020). However, the data indicate that the IIIC and IIID lineages may be distinct species since analyzes with the concatenated sequences of the housekeeping genes (*recA*, *gyrB*, *atpD*,

gltB, *lepA*, *phaC*, and *trpB*) showed a synonymy between the IIID lineage and *B. contaminans* and that lineage IIIC is related with other BCC strains (Baia et al., 2021).

Therefore, we are confident that they are different species based on data obtained through the approach used based on whole-genome analyses, as TYGS, dDDH, and core genome.

DESCRIPTION OF *Burkholderia semiaridus* sp. nov.

Burkholderia semiarius (se.mi.a.ri.dus. L. *semiaridus* of the semiarid) referring to the fact that strains of this species have been obtained, causing onion sour skin in the semiarid region of north-east Brazil.

Cells are Gram-negative, non-sporulating rods. All strains grow on TB-T, and NYDA medium. Growth is observed at 30 C. Yellow-pigment productions on NYDA medium, and on Biolog GEN III MicroPlate System show positive reaction for L-malic acid, D-malic acid, D-fructose, D-raffinose, D-galactose, D-mannose, D-trehalose, gentiobiose, L-fucose, N-acetyl-D-glucosamine, 1% NaCl, pectin, pH 6, sucrose, Tween 40, D-malic acid, formic acid, fusidic acid, γ-amino-butyric acid, D-arabitol, D-mannitol, D-sorbitol, guanidine HCl, myo-inositol, Niaproof 4, Aztreonam and α-D-glucose and negative reaction for citric acid, α-keto-glutaric acid, D-cellobiose, D-aspartic acid, L-galactonic acid lactone, N-acetyl-D-galactosamine, N-acetylneuraminic acid, sodium bromate, sodium butyrate, lithium chlorite, D-serine, D-serine,–α-D-lactose, minocycline, 4% NaCl, 8% NaCl, potassium tellurite, troleandomycin, and β-methyl-D-glucoside. The G+C content is about 67%. Strains have been isolated from symptomatic onion bulbs with onion sour skin in the semiarid region of San Francisco Valley, Brazil and may be satisfactorily different from other BCC species using TYGS web server, ANIm, isDDH, and core genome.

The type strain is IBSBF 3371^{T} (= CCRMBC74^{T}). Phenotypic characteristics of the type strain are the same as those described above for the species. Its G+C content is 66.99%.

DESCRIPTION OF *Burkholderia solum* sp. nov.

Burkholderia solum (so.lum. L. *solum* of the alone) referring to the fact that only one strain of this species has been obtained, causing onion sour skin in the semiarid Brazilian northeastern region.

Cells are Gram-negative, non-sporulating rods. All strains grow on TB-T, and NYDA medium. Growth is observed at 30 C. Yellow-pigment productions on NYDA medium and on Biolog GEN III MicroPlate System show 100% positive reaction for L-malic acid, D-fructose, D-fructose-6-PO4, D-galactose, D-mannose, L-fucose, N-acetyl-D-glucosamine, 1% NaCl, pectin, pH 6, sucrose, Tween 40, D-malic acid, formic acid, fusidic acid, γ-amino-butyric acid, D-arabitol, D-mannitol, D-sorbitol, Citric acid, Bromo-succinic acid, myo-inositol, Niaproof 4, and α-d-glucose and negative reaction for citric acid, α-keto-glutaric acid, D-cellobiose, D-aspartic acid, L-galactonic acid lactone, D-trehalose, Gentiobiose, Guanidine HCl, N-acetyl-D-galactosamine, N-acetylneuraminic acid, sodium bromate, sodium butyrate, lithium chlorite, D-serine, D-serine, α-D-lactose, minocycline, 4% NaCl, 8% NaCl, potassium tellurite, troleandomycin, and β-methyl-D-glucoside. The G+C content is about 66.66%. Strains have been isolated from symptomatic onion bulbs with onion sour skin in the semiarid region of San Francisco Valley, Brazil and may be satisfactorily different from other BCC species using TYGS web server, ANIm, isDDH, and core genome.

The type strain is IBSBF 3372^{T} (= CCRMBC51^{T}). Phenotypic characteristics of the type strain are the same as those described above for the species. Its G+C content is 66.66%.

5 REFERENCES

Andrews, S. 2010. “Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data.” FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available Online at:

Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baia, A. D. B., Silva, A. M. F., Ribeiro, B. G., Souza, C. C., Júnior, W. J. S., Balbino, V. Q., Leal, C. M., Farias, A. R. G. de, Souza, E. B., & Gama, M. A. S. (2021). Predominance of *Burkholderia cenocepacia* lineages causing onion sour skin in the semi-arid region of north-east Brazil. *Plant Pathology*, *70*(3), 521–533. https://doi.org/10.1111/PPA.13311

Baldwin, Adam, Eshwar Mahenthiralingam, Pavel Drevinek, Peter Vandamme, John R. Govan, David J. Waine, John J. LiPuma, et al. 2007. "Environmental *Burkholderia cepacia* Complex Isolates in Human Infections." *Emerging Infectious Diseases* 13 (3): 458–61. https://doi.org/10.3201/eid1303.060403.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, Alexander S Kulikov, Valery M Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boxberger, M., Khedher, M. ben, Magnien, S., Cassir, N., & Scola, B. la. (2020). Draft genome and description of *Chryseobacterium manosquense* strain Marseille-Q2069T sp. nov., a new bacterium isolated from human healthy skin. *New Microbes and New Infections*, *38*. https://doi.org/10.1016/J.NMNI.2020.100805

Chung, Matthew, James B. Munro, Hervé Tettelin, and Julie C. Dunning Hotopp. 2018. "Using Core Genome Alignments To Assign Bacterial Species." *MSystems* 3 (6). https://doi.org/10.1128/MSYSTEMS.00236-18.

Coenye, Tom, and Peter Vandamme. 2003. "Diversity and Significance of *Burkholderi*a Species Occupying Diverse Ecological Niches." *Environmental Microbiology*. Environ Microbiol. https://doi.org/10.1046/j.1462-2920.2003.00471.x.

Coenye, Tom, Peter Vandamme, John R W Govan, and John J LiPuma. 2001. "Taxonomy and Identification of the *Burkholderia cepacia* Complex." *Journal of Clinical Microbiology* 39 (10): 3427–36. https://doi.org/10.1128/JCM.39.10.3427-3436.2001.

Depoorter, Eliza, Evelien de Canck, Charlotte Peeters, Anneleen D Wieme, Margo Cnockaert, James E A Zlosnik, John J LiPuma, Tom Coenye, and Peter Vandamme. 2020. "*Burkholderia cepacia* Complex Taxon K: Where to Split?" *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.01594.

Drevinek, P, and E Mahenthiralingam. 2010. "Burkholderia Cenocepacia in Cystic Fibrosis: Epidemiology and Molecular Mechanisms of Virulence." *Clinical Microbiology and Infection* 16 (7): 821–30. https://doi.org/10.1111/j.1469-0691.2010.03237.x.

Gama, M. A. S. da, Mariano, R. de L. R., Júnior, W. J. da S., Farias, A. R. G. de, Barbosa, M. A. G., Ferreira, M. Á. da S. V., Júnior, C. R. L. C., Santos, L. A., & Souza, E. B. de. (2018). Taxonomic Repositioning of *Xanthomonas campestris* pv. *viticola* (Nayudu 1972) Dye 1978 as *Xanthomonas citri* pv. *viticola* (Nayudu 1972)

Ginther, Jennifer L, Mark Mayo, Stephanie D Warrington, Mirjam Kaestli, Travis Mullins, David M Wagner, Bart J Currie, Apichai Tuanyok, and Paul Keim. 2015. "Identification of *Burkholderia Pseudomallei* Near-Neighbor Species in the Northern Territory of Australia." *PLOS Neglected Tropical Diseases* 9 (6): e0003892. https://doi.org/10.1371/journal.pntd.0003892.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. "QUAST: Quality Assessment Tool for Genome Assemblies." *Bioinformatics* 29 (8): 1072–75. https://doi.org/10.1093/bioinformatics/btt086.

Ho, Ying-Ning, Hsing-Mei Chiang, Chih-Ping Chao, Ching-Chung Su, Hui-Fang Hsu, Chen-tong Guo, Ju-Liang Hsieh, and Chieh-Chen Huang. 2015. “In Planta Biocontrol of Soilborne Fusarium Wilt of Banana through a Plant Endophytic Bacterium, *Burkholderia cenocepacia* 869T2.” *Plant and Soil* 387 (1): 295–306. https://doi.org/10.1007/s11104-014-2297-0.

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N., Ren, H., Liang, L., & Yue, J. Chung, Matthew, James B. Munro, Hervé Tettelin, and Julie C. Dunning Hotopp. 2018. “Using Core Genome Alignments To Assign Bacterial Species.” *MSystems* 3 (6). https://doi.org/10.1128/msystems.00236-18.

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N., Ren, H., Liang, L., & Yue, J. (2020). Genome-based classification of *Burkholderia cepacia* complex provides new insight into its taxonomic status. *Biology Direct*, *15*(1), 6. https://doi.org/10.1186/s13062-020-0258-5

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K.F. Wong, Arndt von Haeseler, and Lars S. Jermiin. 2017. “ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates.” *Nature Methods* 14 (6): 587–89. https://doi.org/10.1038/nmeth.4285.

Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.” *Molecular Biology and Evolution* 35 (6): 1547–49. https://doi.org/10.1093/molbev/msy096.

Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol. 2015;32: 2798–2800. DOI: 10.1093/molbev/msv150

LiPuma, John J., Betty Jo Dulaney, Jennifer D. McMenamin, Paul W. Whitby, Terrence L. Stull, Tom Coenye, and Peter Vandamme. 1999. “Development of RRNA-Based PCR Assays for Identification of *Burkholderia cepacia* Complex Isolates Recovered from Cystic Fibrosis Patients.” *Journal of Clinical Microbiology* 37 (10): 3167–70. https://doi.org/10.1128/jcm.37.10.3167-3170.1999.

Mahenthiralingam, E., J. Bischof, S. K. Byrne, C. Radomski, J. E. Davies, Y. Av-Gay, and P. Vandamme. 2000. “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* 38 (9): 3165–73. https://doi.org/10.1128/jcm.38.9.3165-3173.2000.

Martin, M., B. Christiansen, G. Caspari, M. Hogardt, A. J. von Thomsen, E. Ott, and F. Mattner. 2011. “Hospital-Wide Outbreak of *Burkholderia contaminans* Caused by Prefabricated Moist Washcloths.” *Journal of Hospital Infection* 77 (3): 267–70. https://doi.org/10.1016/j.jhin.2010.10.004.

Martina, Pablo, Mariana Leguizamon, Claudia I Prieto, Silvia A Sousa, Patricia Montanaro, Walter O Draghi, Maren Stämmler, et al. 2018. “Burkholderia Puraquae Sp. Nov., a Novel Species of the *Burkholderia cepacia* Complex Isolated from Hospital Settings and Agricultural Soils.” *International Journal of Systematic and Evolutionary Microbiology* 68 (1): 14–20. https://doi.org/10.1099/ijsem.0.002293.

Meier-Kolthoff, Jan P., and Markus Göker. 2019. “TYGS Is an Automated High-Throughput Platform for State-of-the-Art Genome-Based Taxonomy.” *Nature Communications* 10 (1): 1–10. https://doi.org/10.1038/s41467-019-10210-3.

Meier-Kolthoff, Jan P., Hans Peter Klenk, and Markus Göker. 2014. "Taxonomic Use of DNA G+C Content and DNA-DNA Hybridization in the Genomic Age." *International Journal of Systematic and Evolutionary Microbiology* 64 (PART 2): 352–56. https://doi.org/10.1099/ijs.0.056994-0.

Minh, Bui Quang, Minh Anh Thi Nguyen, and Arndt von Haeseler. 2013. "Ultrafast Approximation for Phylogenetic Bootstrap." *Molecular Biology and Evolution* 30 (5): 1188–95. https://doi.org/10.1093/molbev/mst024.

Nguyen, Lam Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74. https://doi.org/10.1093/molbev/msu300.

Oliveira, W J, W A Silva, A M F Silva, J A Candeia, E B Souza, R L R Mariano, and M a. S Gama. 2017. "First Report of *Burkholderia cenocepacia* Causing Sour Skin of Onion (Allium Cepa) in Brazil." *Plant Disease* 101 (11): 1950. https://doi.org/10.1094/PDIS-05-17-0759-PDN.

Oliveira, Willams J, Elineide B Souza, Adriano M F Silva, Nelson B Lima, Carla M Leal, Jonas A Candeia, and Marco A S Gama. 2019. "Elucidating the Etiology of Onion Bacterial Scale Rot in the Semi-Arid Region of Northeastern Brazil." *Tropical Plant Pathology* 44 (6): 494–502. https://doi.org/10.1007/s40858-019-00310-2.

Ong, Kuan Shion, Yoong Kit Aw, Learn Han Lee, Catherine M. Yule, Yuen Lin Cheow, and Sui Mae Lee. 2016. "*Burkholderia paludis* Sp. Nov., an Antibiotic-Siderophore Producing Novel *Burkholderia cepacia* Complex Species, Isolated from Malaysian Tropical Peat Swamp Soil." *Frontiers in Microbiology* 7 (DEC): 2046. https://doi.org/10.3389/fmicb.2016.02046.

Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T.G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, and Julian Parkhill. 2015. "Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis." *Bioinformatics* 31 (22): 3691–93. https://doi.org/10.1093/bioinformatics/btv421.

Papaleo, Maria Cristiana, Elena Perrin, Isabel Maida, Marco Fondi, Renato Fani, and Peter Vandamme. 2010. "Identification of Species of the *Burkholderia cepacia* Complex by Sequence Analysis of the *HisA* Gene." *Journal of Medical Microbiology* 59 (10): 1163–70. https://doi.org/10.1099/jmm.0.019844-0.

Peeters, Charlotte, James E.A. Zlosnik, Theodore Spilker, Trevor J. Hird, John J. LiPuma, and Peter Vandamme. 2013. "*Burkholderia pseudomultivorans* Sp. Nov., a Novel *Burkholderia cepacia* Complex Species from Human Respiratory Samples and the Rhizosphere." *Systematic and Applied Microbiology* 36 (7): 483–89. https://doi.org/10.1016/j.syapm.2013.06.003.

Pritchard, Leighton, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. 2016. "Genomics and Taxonomy in Diagnostics for Food Security: Soft-Rotting Enterobacterial Plant Pathogens." *Analytical Methods*. Royal Society of Chemistry. https://doi.org/10.1039/c5ay02550h.

Rambaut A, Drummond AJ (2012) FigTree version 1.4. 0. https://github.com/rambaut/figtree. Accessed May 12, 2021.

Richter, Michael, and Ramon Rosselló-Móra. 2009. "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition." *Proceedings of the National Academy of Sciences of the United States of America* 106 (45): 19126–31. https://doi.org/10.1073/pnas.0906412106.

Santos, P. E. los, Palmer, M., Chávez-Ramírez, B., Beukes, C., Steenkamp, E. T., Briscoe, L., Khan, N., Maluk, M., Lafos, M., Humm, E., Arrabit, M., Crook, M., Gross, E., Simon, M. F., Junior, F. B. D. R., Whitman, W. B., Shapiro, N., Poole, P. S., Hirsch, A. M., …

James, E. K. (2018). 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 2018, Vol. 9, Page 389*, *9*(8), 389. https://doi.org/10.3390/GENES9080389

Seconds, Christine, Thierry Heulin, Nicole Marty, and Gerard Chabanon. 1999. "Differentiation of *Burkholderia* Species by PCR-Restriction Fragment Length Polymorphism Analysis of the 16S rRNA Gene and Application to Cystic Fibrosis Isolates." *Journal of Clinical Microbiology* 37 (7): 2201–8. https://doi.org/10.1128/jcm.37.7.2201-2208.1999.

Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics (Oxford, England)* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Spilker, Theodore, Adam Baldwin, Amy Bumford, Chris G. Dowson, Eshwar Mahenthiralingam, and John J. LiPuma. 2009. "Expanded Multilocus Sequence Typing for *Burkholderia* Species." *Journal of Clinical Microbiology* 47 (8): 2607–10. https://doi.org/10.1128/JCM.00770-09.

Tseng, Yuan Hsi, Min Yi Wong, Tsung Yu Huang, Bor Shyh Lin, Chun Wu Tung, and Yao Kuang Huang. 2020. "Molecular Characterization of Clinical Isolates from Vascular Access Infection: A Single-Institution Study." *MicrobiologyOpen* 9 (11): e1126. https://doi.org/10.1002/mbo3.1126.

Tsuji, M., & Kadota, I. (2020). Identification and phylogenetic analysis of *Burkholderia cepacia* complex bacteria isolated from rot of onion bulbs in Tohoku region of Japan. *Journal of General Plant Pathology 2020 86:5*, *86*(5), 376–386. https://doi.org/10.1007/S10327-020-00937-Z

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. J., & Govan, J. R. W. (2003). *Burkholderia cenocepacia* sp. nov.--a new twist to an old story. *Research in Microbiology*, *154*(2), 91–96. https://doi.org/10.1016/S0923-2508(03)00026-3

Vandamme, Peter, and Peter Dawyndt. 2011. "Classification and Identification of the *Burkholderia cepacia* Complex: Past, Present and Future." *Systematic and Applied Microbiology* 34 (2): 87–95. https://doi.org/10.1016/J.SYAPM.2010.10.002.

Vanlaere, Elke, Adam Baldwin, Dirk Gevers, Deborah Henry, Evie de Brandt, John J. LiPuma, Eshwar Mahenthiralingam, David P. Speert, Chris Dowson, and Peter Vandamme. 2009. "Taxon K, a Complex within the *Burkholderia cepacia* Complex, Comprises at Least Two Novel Species, *Burkholderia contaminans* Sp. Nov. and *Burkholderia lata* Sp. Nov." *International Journal of Systematic and Evolutionary Microbiology* 59 (1): 102–11. https://doi.org/10.1099/ijs.0.001123-0.

Velez, Leandro S., Adriano Márcio F. Silva, Carlos A.F. Santos, Emanuel F. Assunção, Matheus S. Silva, Elineide B. Souza, and Marco Aurélio S. Gama. 2020. "Evaluation of Onion Genotypes to Slippery Skin Caused by *Burkholderia gladioli* pv. *alliicola*." *Horticultura Brasileira* 38 (4): 350–55. https://doi.org/10.1590/s0102-053620200402.

Wallner, Adrian, Eoghan King, Eddy L M Ngonkeu, Lionel Moulin, and Gilles Béna. 2019. "Genomic Analyses of *Burkholderia cenocepacia* Reveal Multiple Species with Differential Host-Adaptation to Plants and Humans." *BMC Genomics* 20 (1): 803. https://doi.org/10.1186/s12864-019-6186-z.

Zekic, Filip, Brian Weselowski, and Ze-Chun Yuan. 2017. "Complete Genome Sequence of *Burkholderia cenocepacia* CR318, a Phosphate-Solubilizing Bacterium Isolated from Corn Root." *Genome Announcements* 5 (23). https://doi.org/10.1128/genomeA.00490-17.

Table 1: General features of the strains used in this study

Isolate	Geographic region	BioSample	Contigs	Size (bp)	N50 (bp)	CDS	%GC	BUSCO (%)	References
CCRMBC16	Brazil	SAMN19592385	15	7523028	1841467	6773	67.03	99.9	This study
CCRMBC33	Brazil	SAMN19592386	13	7535064	1388868	6752	67.01	99.7	This study
CCRMBC51	Brazil	SAMN19592387	61	8114250	515855	7833	66.66	99.9	This study
CCRMBC74	Brazil	SAMN19592388	19	7575069	2176560	6788	66.99	99.4	This study
CCRMBC171	Brazil	SAMN19592389	16	7543708	1914232	6734	67.03	99.4	This study
B. aenigmata LMG 13014T	Belgium	SAMEA5795692	158	8897342	98861	8336	65.95	99.7	Depoorter et al., 2020
B. ambifaria AMMDT	USA	SAMN02598309	4	7528567	2646969	6439	66.77	100	Johnson et al., 2015
B. anthina LMG 20980T	USA	SAMEA5780075	73	7608177	201942	6819	66.74	100	Depoorter et al., 2020
B. arboris LMG 24066T	USA	SAMEA5780076	83	8271305	191329	7364	66.84	100	Depoorter et al., 2020
B. catarinensis DNA89T	Brazil	SAMN05521522	892	8118333	18250	6128	66.46	99.2	Bach et al., 2017
B. cenocepacia IIIA J2315T	United Kingdom	SAMEA1705928	4	8055782	3217062	7114	66.90	99.6	Holden et al., 2009
B. cenocepacia IIIB AU1054	USA	SAMN02598326	3	727912	2788459	6381	66.92	99.7	Lipuma et al.. 2002
B. cepacia ATCC 25416T	USA	SAMN04167160	4	8605945	3408190	7490	66.61	99.8	Holden et al., 2009
B. contaminans LMG 23361T	Spain	SAMN02402687	18	9263236	1108159	8038	65.89	99.9	Bloodworth et al., 2015
B. diffusa LMG 24065T	USA	SAMEA5780085	87	7077396	201102	6309	66.26	99.5	Depoorter et al., 2020
B. dolosa LMG 18943T	USA	SAMEA5795690	165	6079078	81572	5248	66.97	99.4	Depoorter et al., 2020
B. lata LMG 22485T	Trinidad and Tobago	SAMN02598262	3	8676277	3587082	7552	66.27	99.7	Depoorter et al., 2020
B. latens LMG 24064T	Italy	SAMEA5795707	115	7075522	104925	6253	66.83	99.7	Depoorter et al., 2020
B. metallica LMG 24068T	USA	SAMEA5795708	118	7532498	133502	6685	67.04	100	Depoorter et al., 2020
B. multivorans ATCC BAA-247T	Belgium	SAMN03140189	3	6322746	3428264	5543	67.24	100	Johnson et al., 2015
B. paludis LMG 30113T	Malaysia	SAMEA5795709	98	8627646	187105	7499	67.15	99.7	Depoorter et al., 2020
B. pseudomultivorans LMG 26883T	USA	SAMEA5795710	150	7396511	93947	6712	66.99	99.3	Depoorter et al., 2020
B. puraquae CAMPA 1040T	Argentina	SAMN06675007	82	8097195	290133	7103	66.59	99.8	Leguizamon et al., 2017
B. pyrrocinia DSM 10685T	Japan	SAMN03651233	4	7961346	3214757	6820	66.46	99.9	Kwak and Shin, 2015
B. seminalis LMG 24067T	USA	SAMEA5795711	197	7928753	86332	7151	67.07	99.5	Depoorter et al., 2020
B. stabilis ATCC BAA-67T	Belgium	SAMN05367054	3	8527947	3318880	7509	66.42	99.9	Bugrysheva et al., 2016
B. stagnalis LMG 28156T	Australia	SAMEA5795712	242	7948011	61121	6980	67.18	99.7	Depoorter et al., 2020
B. territorii LMG 28158T	Australia	SAMEA5795713	101	6907990	129850	6033	66.65	100	Depoorter et al., 2020
B. thailandensis E264T	Thailand	SAMN02604021	2	6723972	3809201	5632	67.63	100	Johnson et al., 2015
B. ubonensis LMG 20358T	Thailand	SAMEA5795714	153	7692957	84900	6745	67.24	99.4	Depoorter et al., 2020
B. vietnamiensis LMG 10929T	Vietnam	SAMN03107475	4	6930496	2262093	5851	66.83	100	Johnson et al., 2015
B. gladioli LMG 2216T	USA	SAMN03010439	5	8899459	4668573	7772	67.61	100	Johnson et al., 2015
B. glumae LMG 2196T	Japan	SAMEA6503528	142	6662149	162494	5937	68.30	99.7	Unpublished
P. fungorum LMG 20227T	France	ASM90283364	33	893876	3508455	7885	61.80	100	Johnson et al., 2015

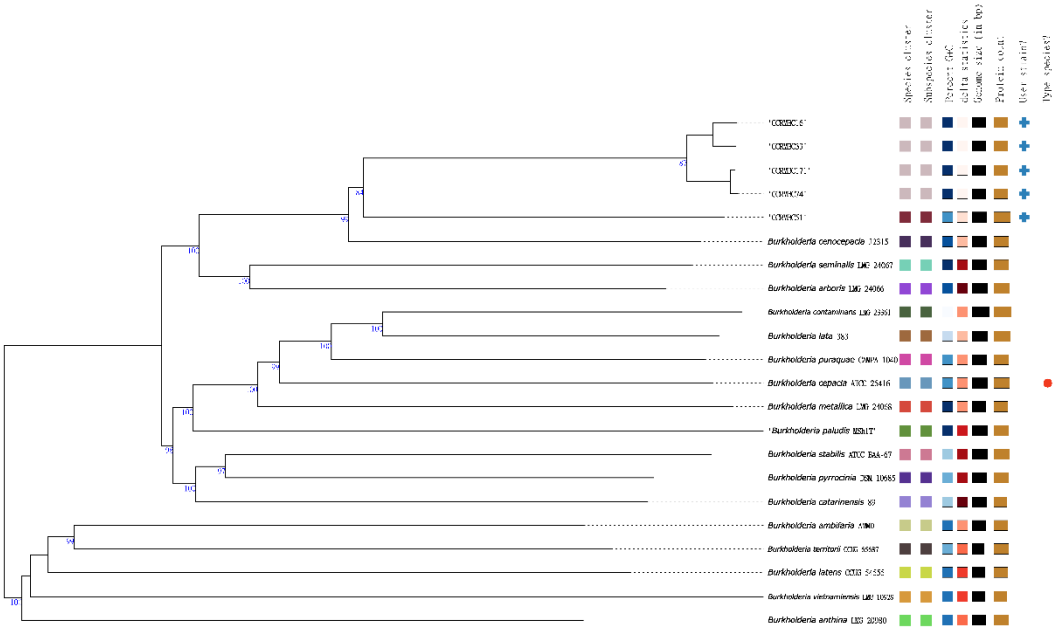

Figure 1: Phylogenomic tree inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. From type strain genome server (TYGS) pairwise comparison of the strains of *Burkholderia cepacia* complex (CCRMBC16, CCRMBC33, CCRMBC51, CCRMBC74, and CCRMBC171) associated with onion sour skin in the semi-arid region of north-east Brazil. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 87.5 %. The symbol + indicates user strains, which represents the strains used in this study and used for this analysis. The tree was rooted at the midpoint.

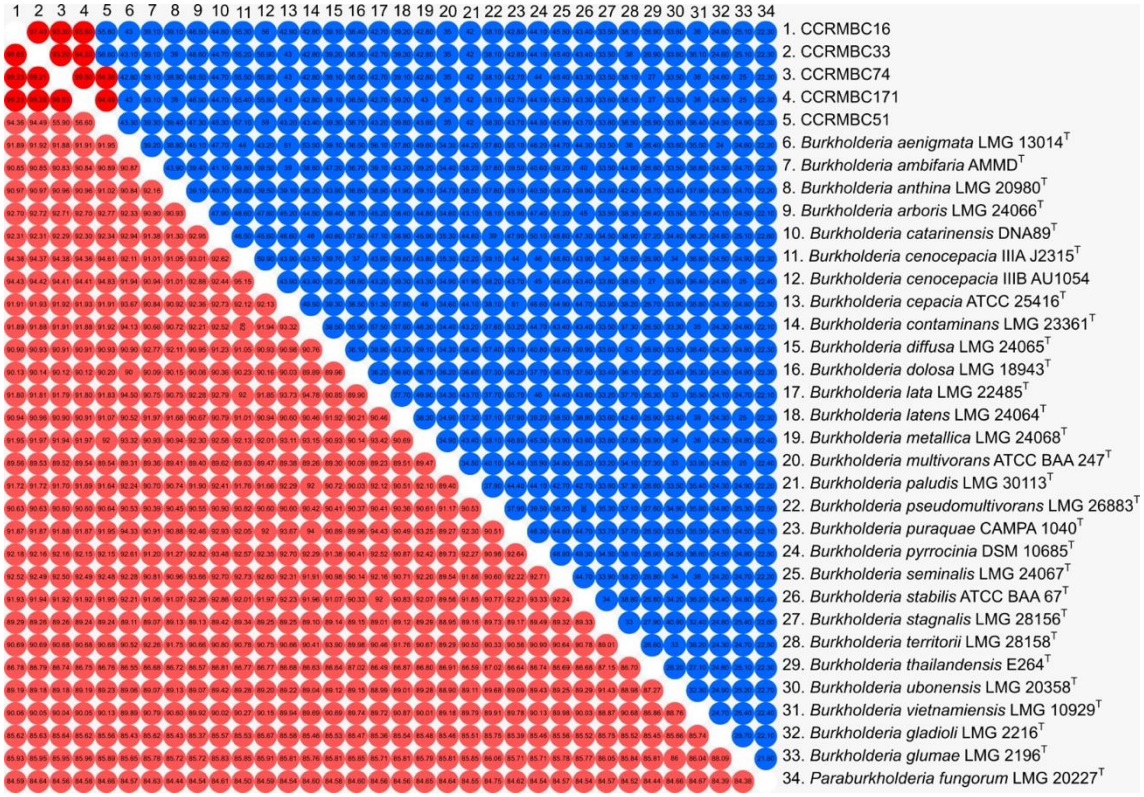

Figure 2: Heatmap of Average Nucleotide Identity (ANIm) and dDDH from pairwise genome comparisons of five strains of *Burkholderia cepacia* complex (BBC) isolated from onion causing onion sour skin in the Northeastern of Brazil, against 26 *Burkholderia* species belonging to BCC available at NCBI database, and an outgroup composed by three strains. The lower triangle displays the ANI values, and the upper triangle displays the dDDH values.

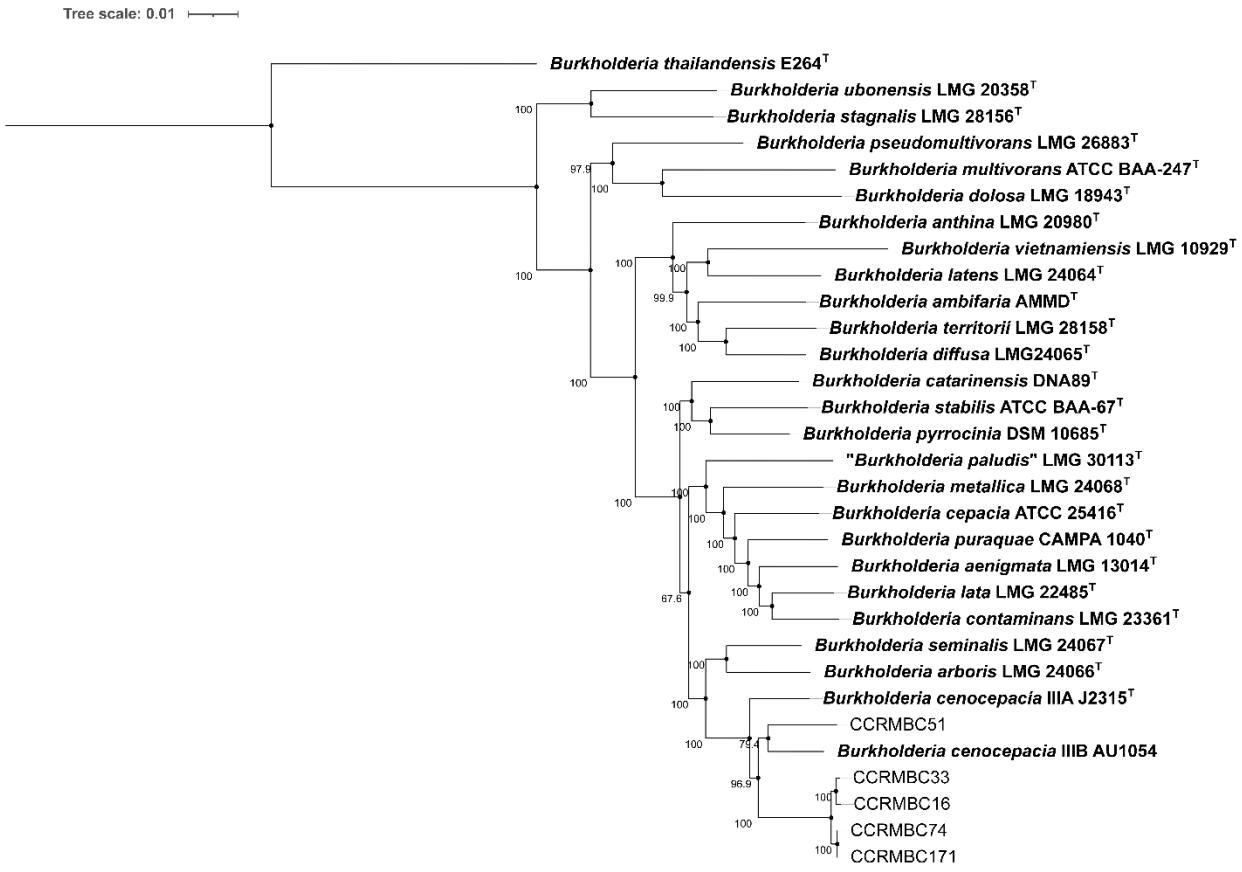

Figure 3: Core gene phylogeny of the *Burkholderia cepacia* complex (BCC) genomes. Core genome phylogeny built using IQ-Tree software from the concatenated nucleotide sequences of 309 single copy orthologous of protein codes from BCC genome strains available in the NCBI database and five new genomes of strains sequenced for this study. The tree structure shows that the new five strains are new species belonging to BCC. Bootstrap values are indicated in the nodes of the tree.

Supplementary Table 1: Metric assembly parameters of the five strains of *Burkholderia cepacia* complex causing sour skin in onion bulbs

Assembly	CCRMBC16	CCRMBC33	CCRMBC51	CCRMBC74	CCRMBC171
# contigs (>= 0 bp)	48	48	307	52	30
# contigs (>= 1000 bp)	13	9	46	13	13
# contigs (>= 5000 bp)	8	6	22	8	8
# contigs (>= 10000 bp)	6	6	21	8	8
# contigs (>= 25000 bp)	5	5	19	8	6
# contigs (>= 50000 bp)	5	5	19	7	6
Total length (>= 0 bp)	7530728	7543403	8186273	7581044	7547108
Total length (>= 1000 bp)	7521954	7532022	8104481	7570980	7541762
Total length (>= 5000 bp)	7514201	7526460	8049476	7562309	7533911
Total length (>= 10000 bp)	7496557	7526460	8042648	7562309	7533911
Total length (>= 25000 bp)	7481693	7512651	8017605	7562309	7501790
Total length (>= 50000 bp)	7481693	7512651	8017605	7524329	7501790
# contigs	15	13	61	19	16
Largest contig	2379354	3418472	1393841	2199546	2061719
Total length	7523028	7535064	8114250	7575069	7543708
GC (%)	67.03	67.01	66.66	66.99	67.03
N50	1841467	1388868	515855	2176560	1914232
N75	1006109	1359968	415158	1341144	1781195
L50	2	2	6	2	2
L75	4	3	11	3	3
# N's per 100 kbp	43.87	46.45	325.27	43.56	29.16

Supplementary table 2: Pairwise comparisons of five strains causing onion sour skin concerning type-strains genomes from *Burkholderia cepacia* complex, performed at Type Strain Genome Server (TYGS).

Strains used in this study

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC74'	CCRMBC171'	98.3	[97.1 - 99.0]	99.5	[99.2 - 99.7]	99.2	[98.6 - 99.5]	0.03
CCRMBC16'	CCRMBC33'	97.8	[96.4 - 98.6]	97.4	[96.3 - 98.1]	98.7	[97.9 - 99.2]	0.02
CCRMBC171'	CCRMBC33'	98.4	[97.3 - 99.1]	94.2	[92.4 - 95.5]	98.9	[98.1 - 99.3]	0.02
CCRMBC16'	CCRMBC171'	96.7	[94.8 - 97.9]	93.6	[91.7 - 95.0]	97.7	[96.5 - 98.5]	0
CCRMBC74'	CCRMBC16'	96.9	[95.2 - 98.0]	93.3	[91.4 - 94.8]	97.8	[96.7 - 98.6]	0.03
CCRMBC74'	CCRMBC33'	95.9	[93.9 - 97.3]	93.2	[91.3 - 94.7]	97.2	[95.8 - 98.1]	0.01
CCRMBC33'	CCRMBC51'	77.2	[73.2 - 80.8]	56.6	[53.8 - 59.3]	75.4	[71.9 - 78.6]	0.37
CCRMBC171'	CCRMBC51'	77	[73.0 - 80.5]	56.6	[53.9 - 59.3]	75.2	[71.7 - 78.4]	0.39
CCRMBC74'	CCRMBC51'	73.7	[69.7 - 77.3]	55.9	[53.1 - 58.6]	72.2	[68.7 - 75.4]	0.35
CCRMBC16'	CCRMBC51'	75.2	[71.2 - 78.8]	55.8	[53.1 - 58.5]	73.5	[70.0 - 76.7]	0.39

CCRMBC16

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC16'	*Burkholderia cenocepacia J2315*	59.8	[56.1 - 63.3]	55.3	[52.6 - 58.0]	60.0	[56.8 - 63.2]	0.12
CCRMBC16'	*Burkholderia arboris LMG 24066*	53.2	[49.7 - 56.7]	46.5	[44.0 - 49.1]	52.2	[49.1 - 55.3]	0.18
CCRMBC16'	*Burkholderia seminalis LMG 24067*	61.2	[57.5 - 64.8]	45.5	[42.9 - 48.0]	58.6	[55.4 - 61.8]	0.06
CCRMBC16'	*Burkholderia catarinensis 89*	53.2	[49.7 - 56.7]	44.8	[42.2 - 47.4]	51.7	[48.6 - 54.8]	0.62
CCRMBC16'	*Burkholderia pyrrocinia DSM 10685*	58.0	[54.4 - 61.5]	44.1	[41.6 - 46.7]	55.5	[52.4 - 58.6]	0.56
CCRMBC16'	*Burkholderia stabilis ATCC BAA-67*	48.0	[44.6 - 51.4]	43.4	[40.8 - 45.9]	46.9	[43.9 - 49.9]	0.6

CCRMBC16'	*Burkholderia cepacia ATCC 25416*	55.5	[51.9 - 59.0]	42.9	[40.4 - 45.5]	53.1	[50.0 - 56.2]	0.41
CCRMBC16'	*Burkholderia puraquae CAMPA 1040*	58.0	[54.4 - 61.5]	42.8	[40.3 - 45.3]	55.1	[51.9 - 58.2]	0.42
CCRMBC16'	*Burkholderia metallica LMG 24068*	60.3	[56.6 - 63.8]	42.8	[40.3 - 45.3]	57.0	[53.8 - 60.1]	0.03
CCRMBC16'	*Burkholderia contaminans LMG 23361*	51.3	[47.8 - 54.7]	42.7	[40.1 - 45.2]	49.5	[46.4 - 52.5]	1.14
CCRMBC16'	*Burkholderia lata 383*	49.2	[45.8 - 52.7]	42.7	[40.2 - 45.2]	47.8	[44.7 - 50.8]	0.75
CCRMBC16'	*Burkholderia paludis MSh1T*	43.5	[40.1 - 46.9]	41.9	[39.4 - 44.5]	42.6	[39.7 - 45.7]	0.1
CCRMBC16'	*Burkholderia territorii CCUG 65687*	51.0	[47.5 - 54.4]	39.1	[36.7 - 41.7]	48.2	[45.2 - 51.3]	0.46

CCRMBC33

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC33'	*Burkholderia cenocepacia J2315*	60.0	[56.3 - 63.6]	55.2	[52.5 - 57.9]	60.2	[57.0 - 63.4]	0.1
CCRMBC33'	*Burkholderia arboris LMG 24066*	53.1	[49.6 - 56.6]	46.6	[44.0 - 49.2]	52.1	[49.0 - 55.1]	0.16
CCRMBC33'	*Burkholderia seminalis LMG 24067*	62.1	[58.4 - 65.7]	45.4	[42.9 - 48.0]	59.4	[56.1 - 62.5]	0.08
CCRMBC33'	*Burkholderia catarinensis 89*	53.0	[49.5 - 56.5]	44.7	[42.2 - 47.3]	51.5	[48.4 - 54.6]	0.6
CCRMBC33'	*Burkholderia pyrrocinia DSM 10685*	58.1	[54.5 - 61.6]	44.1	[41.5 - 46.6]	55.6	[52.4 - 58.7]	0.54
CCRMBC33'	*Burkholderia stabilis ATCC BAA-67*	47.9	[44.5 - 51.3]	43.4	[40.9 - 45.9]	46.8	[43.7 - 49.8]	0.58
CCRMBC33'	*Burkholderia cepacia ATCC 25416*	55.5	[52.0 - 59.0]	43.0	[40.5 - 45.6]	53.1	[50.0 - 56.2]	0.39
CCRMBC33'	*Burkholderia puraquae CAMPA 1040*	57.9	[54.3 - 61.4]	42.8	[40.3 - 45.4]	55.0	[51.9 - 58.1]	0.4
CCRMBC33'	*Burkholderia metallica LMG 24068*	59.9	[56.2 - 63.4]	42.8	[40.3 - 45.4]	56.7	[53.5 - 59.8]	0.05

CCRMBC33'	*Burkholderia lata 383*	49.1	[45.7 - 52.5]	42.7	[40.2 - 45.3]	47.6	[44.6 - 50.7]	0.73
CCRMBC33'	*Burkholderia contaminans LMG 23361*	51.6	[48.1 - 55.0]	42.7	[40.2 - 45.2]	49.7	[46.7 - 52.8]	1.12
CCRMBC33'	*Burkholderia latens CCUG 54555*	57.2	[53.6 - 60.7]	39.3	[36.9 - 41.9]	53.3	[50.2 - 56.4]	0.16
CCRMBC33'	*Burkholderia territorii CCUG 65687*	50.8	[47.4 - 54.3]	39.1	[36.6 - 41.6]	48.1	[45.1 - 51.1]	0.44
CCRMBC33'	*Burkholderia anthina LMG 20980*	48.6	[45.2 - 52.1]	39.0	[36.5 - 41.5]	46.3	[43.3 - 49.3]	0.25
CCRMBC33'	*Burkholderia vietnamiensis LMG 10929*	47.6	[44.2 - 51.0]	36.1	[33.6 - 38.6]	44.6	[41.6 - 47.6]	0.17

CCRMBC51

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC51'	*Burkholderia cenocepacia J2315*	62.1	[58.3 - 65.7]	57.1	[54.3 - 59.8]	62.5	[59.2 - 65.7]	0.26
CCRMBC51'	*Burkholderia arboris LMG 24066*	52.4	[48.9 - 55.9]	47.3	[44.7 - 49.9]	51.6	[48.6 - 54.7]	0.21
CCRMBC51'	*Burkholderia seminalis LMG 24067*	62.6	[58.9 - 66.2]	45.9	[43.3 - 48.4]	59.9	[56.6 - 63.1]	0.44
CCRMBC51'	*Burkholderia catarinensis 89*	52.6	[49.1 - 56.0]	45.3	[42.7 - 47.8]	51.3	[48.2 - 54.4]	0.23
CCRMBC51'	*Burkholderia pyrrocinia DSM 10685*	56.7	[53.1 - 60.2]	44.5	[42.0 - 47.1]	54.6	[51.4 - 57.7]	0.18
CCRMBC51'	*Burkholderia stabilis ATCC BAA-67*	46.9	[43.5 - 50.3]	43.7	[41.2 - 46.3]	46.0	[43.0 - 49.0]	0.21
CCRMBC51'	*Burkholderia puraquae CAMPA 1040*	55.4	[51.8 - 58.9]	43.7	[41.2 - 46.3]	53.2	[50.1 - 56.3]	0.03
CCRMBC51'	*Burkholderia metallica LMG 24068*	55.7	[52.2 - 59.2]	43.6	[41.1 - 46.2]	53.5	[50.4 - 56.6]	0.42
CCRMBC51'	*Burkholderia contaminans LMG 23361*	51.4	[47.9 - 54.8]	43.3	[40.8 - 45.8]	49.7	[46.7 - 52.8]	0.75
CCRMBC51'	*Burkholderia lata 383*	47.9	[44.5 - 51.3]	43.2	[40.7 - 45.7]	46.7	[43.7 - 49.7]	0.36

CCRMBC51'	*Burkholderia cepacia ATCC 25416*	55.6	[52.1 - 59.1]	43.2	[40.7 - 45.8]	53.3	[50.2 - 56.4]	0.03
CCRMBC51'	*Burkholderia paludis MSh1T*	43.1	[39.7 - 46.5]	42.1	[39.6 - 44.7]	42.3	[39.3 - 45.3]	0.48
CCRMBC51'	*Burkholderia territorii CCUG 65687*	48.1	[44.7 - 51.5]	39.4	[36.9 - 41.9]	45.9	[42.9 - 49.0]	0.07

CCRMBC74

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC74'	*Burkholderia cenocepacia J2315*	59.5	[55.8 - 63.0]	55.5	[52.7 - 58.2]	59.8	[56.6 - 63.0]	0.09
CCRMBC74'	*Burkholderia arboris LMG 24066*	53.2	[49.7 - 56.6]	46.5	[43.9 - 49.1]	52.1	[49.0 - 55.2]	0.14
CCRMBC74'	*Burkholderia seminalis LMG 24067*	60.8	[57.1 - 64.3]	45.4	[42.9 - 48.0]	58.2	[55.0 - 61.4]	0.09
CCRMBC74'	*Burkholderia catarinensis 89*	53.2	[49.7 - 56.7]	44.7	[42.1 - 47.2]	51.7	[48.6 - 54.7]	0.58
CCRMBC74'	*Burkholderia pyrrocinia DSM 10685*	57.6	[54.0 - 61.1]	44.0	[41.5 - 46.6]	55.1	[52.0 - 58.2]	0.53
CCRMBC74'	*Burkholderia stabilis ATCC BAA-67*	47.9	[44.5 - 51.4]	43.3	[40.7 - 45.8]	46.8	[43.8 - 49.8]	0.57
CCRMBC74'	*Burkholderia cepacia ATCC 25416*	55.0	[51.5 - 58.5]	43.0	[40.5 - 45.6]	52.7	[49.6 - 55.8]	0.38
CCRMBC74'	*Burkholderia metallica LMG 24068*	59.5	[55.9 - 63.1]	42.8	[40.3 - 45.4]	56.4	[53.2 - 59.5]	0.06
CCRMBC74'	*Burkholderia puraquae CAMPA 1040*	58.0	[54.4 - 61.5]	42.7	[40.2 - 45.3]	55.1	[51.9 - 58.2]	0.39
CCRMBC74'	*Burkholderia lata 383*	49.1	[45.7 - 52.6]	42.7	[40.2 - 45.3]	47.7	[44.7 - 50.7]	0.72
CCRMBC74'	*Burkholderia contaminans LMG 23361*	51.0	[47.6 - 54.5]	42.6	[40.1 - 45.1]	49.3	[46.2 - 52.3]	1.1
CCRMBC74'	*Burkholderia latens CCUG 54555*	57.3	[53.7 - 60.8]	39.2	[36.7 - 41.7]	53.3	[50.2 - 56.4]	0.15
CCRMBC74'	*Burkholderia territorii CCUG 65687*	50.5	[47.1 - 54.0]	39.1	[36.7 - 41.7]	47.9	[44.8 - 50.9]	0.43

CCRMBC74'	*Burkholderia anthina LMG 20980*	48.4	[45.0 - 51.9]	38.9	[36.5 - 41.5]	46.1	[43.1 - 49.1]	0.24
CCRMBC74'	*Burkholderia vietnamiensis LMG 10929*	46.5	[43.1 - 49.9]	36.0	[33.6 - 38.5]	43.6	[40.7 - 46.7]	0.16

CCRMBC171

Query	**Subject**	***d0***	**C.I.d0**	***d4***	**C.I.d4**	***d6***	**C.I.d6**	**Diff. G+C Percent**
CCRMBC171'	*Burkholderia cenocepacia J2315*	60.1	[56.4 - 63.6]	55.4	[52.7 - 58.1]	60.3	[57.1 - 63.5]	0.12
CCRMBC171'	*Burkholderia arboris LMG 24066*	53.3	[49.8 - 56.8]	46.5	[44.0 - 49.1]	52.2	[49.1 - 55.3]	0.18
CCRMBC171'	*Burkholderia seminalis LMG 24067*	61.9	[58.2 - 65.5]	45.5	[42.9 - 48.1]	59.2	[56.0 - 62.3]	0.05
CCRMBC171'	*Burkholderia catarinensis 89*	53.5	[50.0 - 57.0]	44.7	[42.1 - 47.2]	51.9	[48.8 - 55.0]	0.62
CCRMBC171'	*Burkholderia pyrrocinia DSM 10685*	58.0	[54.3 - 61.5]	44.1	[41.5 - 46.6]	55.5	[52.3 - 58.6]	0.56
CCRMBC171'	*Burkholderia stabilis ATCC BAA-67*	48.1	[44.7 - 51.5]	43.3	[40.8 - 45.8]	47.0	[43.9 - 50.0]	0.6
CCRMBC171'	*Burkholderia cepacia ATCC 25416*	55.9	[52.4 - 59.4]	43.0	[40.5 - 45.6]	53.5	[50.4 - 56.6]	0.41
CCRMBC171'	*Burkholderia metallica LMG 24068*	59.9	[56.2 - 63.4]	43.0	[40.4 - 45.5]	56.7	[53.5 - 59.8]	0.03
CCRMBC171'	*Burkholderia puraquae CAMPA 1040*	58.4	[54.8 - 62.0]	42.7	[40.2 - 45.3]	55.4	[52.3 - 58.6]	0.42
CCRMBC171'	*Burkholderia contaminans LMG 23361*	51.5	[48.1 - 55.0]	42.7	[40.2 - 45.2]	49.7	[46.7 - 52.8]	1.14
CCRMBC171'	*Burkholderia lata 383*	49.3	[45.9 - 52.7]	42.7	[40.2 - 45.3]	47.8	[44.8 - 50.9]	0.75
CCRMBC171'	*Burkholderia paludis MSh1T*	43.4	[40.0 - 46.8]	42.0	[39.5 - 44.6]	42.6	[39.6 - 45.6]	0.09
CCRMBC171'	*Burkholderia ambifaria AMMD*	58.1	[54.5 - 61.6]	39.1	[36.6 - 41.6]	53.9	[50.8 - 57.0]	0.25
CCRMBC171'	*Burkholderia territorii CCUG 65687*	50.7	[47.3 - 54.1]	39.1	[36.7 - 41.7]	48.0	[45.0 - 51.0]	0.46

CCRMBC171'	*Burkholderia anthina LMG 20980*	48.8	[45.3 - 52.2]	39.0	[36.5 - 41.5]	46.4	[43.4 - 49.4]	0.27

Supplementary table 3: Tetranucleotide frequency correlation coefficients of 34 *Burkholderia* spp. strains.

	CCRMBC16	CCRMBC33	CCRMBC51	CCRMBC74	CCRMBC171	*B. aenigmata* LMG13014T	*B. ambifaria* AMMDT	*B. anthina* LMG20980T	*B. arboris* LMG24066T	*B. catarinensis* DNA89T	*B. cenocepacia* IIIA J2315T	*B. cenocepacia* IIIB AU1054	*B. cepacia* ATCC 25416T	*B. contaminans* LMG23361T	*B. diffusa* LMG24065T	*B. dolosa* LMG18943T	*B. lata* LMG22485T	*B. latens* LMG24064T	*B. metallica* LMG24068T	*B. multivorans* ATCC BAA 247T	*B. paludis* LMG30113T	*B. pseudomultivorans* LMG26883T	*B. puraquae* CAMPA 1040T	*B. pyrrocinia* DSM 10685T	*B. seminalis* LMG24067T	*B. stabilis* ATCC BAA 67T	*B. stagnalis* LMG28156T	*B. territorii* LMG28158T	*B. thailandensis* E264T	*B. ubonensis* LMG20358T	*B. vietnamiensis* LMG10929T	*B. fungorum* LMG20227T	*B. gladioli* LMG2216T	*B. glumae* LMG2196T
CCRMBC16	1.0000	1.0000	0.9992	0.9999	0.9999	0.9942	0.9977	0.9974	0.9946	0.9969	0.9990	0.9989	0.9960	0.9948	0.9968	0.9834	0.9929	0.9961	0.9981	0.9835	0.9933	0.9931	0.9937	0.9963	0.9973	0.9957	0.9897	0.9962	0.9616	0.9889	0.9883	0.8203	0.9120	0.9284
CCRMBC33	1.0000	1.0000	0.9992	0.9999	1.0000	0.9942	0.9976	0.9974	0.9946	0.9970	0.9989	0.9988	0.9960	0.9949	0.9968	0.9832	0.9929	0.9960	0.9981	0.9834	0.9932	0.9929	0.9937	0.9964	0.9973	0.9957	0.9895	0.9961	0.9618	0.9887	0.9882	0.8206	0.9115	0.9281
CCRMBC51	0.9992	0.9992	1.0000	0.9993	0.9992	0.9949	0.9985	0.9974	0.9950	0.9972	0.9994	0.9991	0.9960	0.9956	0.9972	0.9845	0.9935	0.9970	0.9981	0.9839	0.9930	0.9946	0.9941	0.9967	0.9969	0.9953	0.9904	0.9953	0.9578	0.9898	0.9902	0.8196	0.9108	0.9277
CCRMBC74	0.9999	0.9999	0.9993	1.0000	1.0000	0.9944	0.9977	0.9974	0.9948	0.9971	0.9990	0.9989	0.9961	0.9950	0.9969	0.9832	0.9931	0.9961	0.9981	0.9834	0.9934	0.9930	0.9939	0.9964	0.9974	0.9957	0.9897	0.9960	0.9613	0.9889	0.9882	0.8210	0.9119	0.9287
CCRMBC171	0.9999	1.0000	0.9992	1.0000	1.0000	0.9943	0.9977	0.9974	0.9948	0.9970	0.9989	0.9989	0.9960	0.9949	0.9969	0.9830	0.9931	0.9959	0.9981	0.9833	0.9934	0.9929	0.9938	0.9963	0.9973	0.9957	0.9895	0.9960	0.9613	0.9888	0.9881	0.8215	0.9120	0.9286
B. aenigmata LMG13014T	0.9942	0.9942	0.9949	0.9944	0.9943	1.0000	0.9951	0.9958	0.9979	0.9975	0.9953	0.9954	0.9984	0.9988	0.9968	0.9710	0.9988	0.9900	0.9967	0.9720	0.9947	0.9889	0.9985	0.9975	0.9959	0.9963	0.9878	0.9935	0.9461	0.9851	0.9802	0.8270	0.9160	0.9265
B. ambifaria AMMDT	0.9977	0.9976	0.9985	0.9977	0.9977	0.9951	1.0000	0.9974	0.9945	0.9963	0.9983	0.9982	0.9960	0.9959	0.9980	0.9854	0.9936	0.9975	0.9973	0.9847	0.9923	0.9958	0.9939	0.9961	0.9967	0.9947	0.9925	0.9960	0.9567	0.9914	0.9922	0.8220	0.9182	0.9350
B. anthina LMG20980T	0.9974	0.9974	0.9974	0.9974	0.9974	0.9958	0.9974	1.0000	0.9971	0.9973	0.9978	0.9978	0.9972	0.9961	0.9978	0.9813	0.9956	0.9954	0.9982	0.9800	0.9943	0.9921	0.9962	0.9972	0.9969	0.9975	0.9889	0.9972	0.9564	0.9888	0.9863	0.8317	0.9138	0.9267
B. arboris LMG24066T	0.9946	0.9946	0.9950	0.9948	0.9948	0.9979	0.9945	0.9971	1.0000	0.9979	0.9959	0.9960	0.9980	0.9971	0.9961	0.9725	0.9985	0.9906	0.9973	0.9723	0.9955	0.9888	0.9987	0.9978	0.9967	0.9974	0.9853	0.9947	0.9450	0.9847	0.9792	0.8247	0.9090	0.9202
B. catarinensis DNA89T	0.9969	0.9970	0.9972	0.9971	0.9970	0.9975	0.9963	0.9973	0.9979	1.0000	0.9976	0.9971	0.9984	0.9971	0.9977	0.9761	0.9966	0.9929	0.9979	0.9766	0.9957	0.9927	0.9978	0.9990	0.9981	0.9971	0.9888	0.9953	0.9510	0.9880	0.9820	0.8191	0.9107	0.9246
B. cenocepacia IIIA J2315T	0.9990	0.9989	0.9994	0.9990	0.9989	0.9953	0.9983	0.9978	0.9959	0.9976	1.0000	0.9995	0.9967	0.9957	0.9976	0.9825	0.9943	0.9961	0.9983	0.9817	0.9939	0.9945	0.9944	0.9970	0.9979	0.9962	0.9902	0.9962	0.9546	0.9898	0.9881	0.8183	0.9137	0.9277
B. cenocepacia IIIB AU1054	0.9989	0.9988	0.9991	0.9989	0.9989	0.9954	0.9982	0.9978	0.9960	0.9971	0.9995	1.0000	0.9966	0.9959	0.9977	0.9819	0.9946	0.9957	0.9983	0.9815	0.9935	0.9936	0.9946	0.9968	0.9975	0.9961	0.9891	0.9961	0.9554	0.9892	0.9879	0.8224	0.9127	0.9268
B. cepacia ATCC 25416T	0.9960	0.9960	0.9960	0.9961	0.9960	0.9984	0.9960	0.9972	0.9980	0.9984	0.9967	0.9966	1.0000	0.9980	0.9973	0.9739	0.9979	0.9919	0.9979	0.9736	0.9958	0.9903	0.9983	0.9987	0.9982	0.9977	0.9897	0.9955	0.9507	0.9873	0.9810	0.8193	0.9136	0.9245
B. contaminans LMG23361T	0.9948	0.9949	0.9956	0.9950	0.9949	0.9988	0.9959	0.9961	0.9971	0.9971	0.9957	0.9959	0.9980	1.0000	0.9961	0.9728	0.9980	0.9912	0.9970	0.9725	0.9954	0.9893	0.9976	0.9969	0.9954	0.9965	0.9896	0.9936	0.9484	0.9873	0.9839	0.8330	0.9200	0.9297
B. diffusa LMG24065T	0.9968	0.9968	0.9972	0.9969	0.9969	0.9968	0.9980	0.9978	0.9961	0.9977	0.9976	0.9977	0.9973	0.9961	1.0000	0.9805	0.9955	0.9953	0.9971	0.9809	0.9933	0.9936	0.9961	0.9976	0.9975	0.9961	0.9896	0.9973	0.9540	0.9894	0.9855	0.8255	0.9159	0.9301
B. dolosa LMG18943T	0.9834	0.9832	0.9845	0.9832	0.9830	0.9710	0.9854	0.9813	0.9725	0.9761	0.9825	0.9819	0.9739	0.9728	0.9805	1.0000	0.9687	0.9927	0.9797	0.9939	0.9659	0.9859	0.9716	0.9765	0.9772	0.9737	0.9750	0.9806	0.9570	0.9779	0.9869	0.7955	0.8733	0.9050
B. lata LMG22485T	0.9929	0.9929	0.9935	0.9931	0.9931	0.9988	0.9936	0.9956	0.9985	0.9966	0.9943	0.9946	0.9979	0.9980	0.9955	0.9687	1.0000	0.9884	0.9966	0.9679	0.9944	0.9863	0.9988	0.9970	0.9948	0.9964	0.9853	0.9934	0.9434	0.9831	0.9774	0.8277	0.9117	0.9202
B. latens LMG24064T	0.9961	0.9960	0.9970	0.9961	0.9959	0.9900	0.9975	0.9954	0.9906	0.9929	0.9961	0.9957	0.9919	0.9912	0.9953	0.9927	0.9884	1.0000	0.9947	0.9893	0.9863	0.9944	0.9899	0.9930	0.9931	0.9910	0.9886	0.9941	0.9589	0.9894	0.9921	0.8123	0.9016	0.9236
B. metallica LMG24068T	0.9981	0.9981	0.9981	0.9981	0.9981	0.9967	0.9973	0.9982	0.9973	0.9979	0.9983	0.9983	0.9979	0.9970	0.9971	0.9797	0.9966	0.9947	1.0000	0.9787	0.9943	0.9921	0.9970	0.9979	0.9972	0.9975	0.9901	0.9963	0.9560	0.9885	0.9861	0.8187	0.9097	0.9249
B. multivorans ATCC BAA 247T	0.9835	0.9834	0.9839	0.9834	0.9833	0.9720	0.9847	0.9800	0.9723	0.9766	0.9817	0.9815	0.9736	0.9725	0.9809	0.9939	0.9679	0.9893	0.9787	1.0000	0.9662	0.9875	0.9723	0.9769	0.9781	0.9736	0.9734	0.9807	0.9633	0.9748	0.9853	0.7942	0.8827	0.9150
B. paludis LMG30113T	0.9933	0.9932	0.9930	0.9934	0.9934	0.9947	0.9923	0.9943	0.9955	0.9957	0.9939	0.9935	0.9958	0.9954	0.9933	0.9659	0.9944	0.9863	0.9943	0.9662	1.0000	0.9875	0.9945	0.9942	0.9950	0.9945	0.9899	0.9918	0.9496	0.9891	0.9786	0.8387	0.9255	0.9348
B. pseudomultivorans LMG26883T	0.9931	0.9929	0.9946	0.9930	0.9929	0.9889	0.9958	0.9921	0.9888	0.9927	0.9945	0.9936	0.9903	0.9893	0.9936	0.9859	0.9863	0.9944	0.9921	0.9875	0.9875	1.0000	0.9881	0.9927	0.9929	0.9893	0.9904	0.9906	0.9517	0.9899	0.9911	0.8017	0.9143	0.9327
B. puraquae CAMPA 1040T	0.9937	0.9937	0.9941	0.9939	0.9938	0.9985	0.9939	0.9962	0.9987	0.9978	0.9944	0.9946	0.9983	0.9976	0.9961	0.9716	0.9988	0.9899	0.9970	0.9723	0.9945	0.9881	1.0000	0.9980	0.9955	0.9965	0.9860	0.9942	0.9483	0.9841	0.9775	0.8212	0.9062	0.9184
B. pyrrocinia DSM 10685T	0.9963	0.9964	0.9967	0.9964	0.9963	0.9975	0.9961	0.9972	0.9978	0.9990	0.9970	0.9968	0.9987	0.9969	0.9976	0.9765	0.9970	0.9930	0.9979	0.9769	0.9942	0.9927	0.9980	1.0000	0.9978	0.9976	0.9882	0.9953	0.9513	0.9869	0.9824	0.8157	0.9084	0.9213
B. seminalis LMG24067T	0.9973	0.9973	0.9969	0.9974	0.9973	0.9959	0.9967	0.9969	0.9967	0.9981	0.9979	0.9975	0.9982	0.9954	0.9975	0.9772	0.9948	0.9931	0.9972	0.9781	0.9950	0.9929	0.9955	0.9978	1.0000	0.9966	0.9895	0.9961	0.9523	0.9884	0.9825	0.8171	0.9168	0.9286
B. stabilis ATCC BAA 67T	0.9957	0.9957	0.9953	0.9957	0.9957	0.9963	0.9947	0.9975	0.9974	0.9971	0.9962	0.9961	0.9977	0.9965	0.9961	0.9737	0.9964	0.9910	0.9975	0.9736	0.9945	0.9893	0.9965	0.9976	0.9966	1.0000	0.9871	0.9955	0.9518	0.9865	0.9824	0.8257	0.9157	0.9252
B. stagnalis LMG28156T	0.9897	0.9895	0.9904	0.9897	0.9895	0.9878	0.9925	0.9889	0.9853	0.9888	0.9902	0.9891	0.9897	0.9896	0.9896	0.9750	0.9853	0.9886	0.9901	0.9734	0.9899	0.9904	0.9860	0.9882	0.9895	0.9871	1.0000	0.9861	0.9617	0.9969	0.9871	0.8112	0.9258	0.9469
B. territorii LMG28158T	0.9962	0.9961	0.9953	0.9960	0.9960	0.9935	0.9960	0.9972	0.9947	0.9953	0.9962	0.9961	0.9955	0.9936	0.9973	0.9806	0.9934	0.9941	0.9963	0.9807	0.9918	0.9906	0.9942	0.9953	0.9961	0.9955	0.9861	1.0000	0.9557	0.9860	0.9837	0.8212	0.9126	0.9242
B. thailandensis E264T	0.9616	0.9618	0.9578	0.9613	0.9613	0.9461	0.9567	0.9564	0.9450	0.9510	0.9546	0.9554	0.9507	0.9484	0.9540	0.9570	0.9434	0.9589	0.9560	0.9633	0.9496	0.9517	0.9483	0.9513	0.9523	0.9518	0.9617	0.9557	1.0000	0.9648	0.9578	0.8035	0.8679	0.9074
B. ubonensis LMG20358T	0.9889	0.9887	0.9898	0.9889	0.9888	0.9851	0.9914	0.9888	0.9847	0.9880	0.9898	0.9892	0.9873	0.9873	0.9894	0.9779	0.9831	0.9894	0.9885	0.9748	0.9891	0.9899	0.9841	0.9869	0.9884	0.9865	0.9969	0.9860	0.9648	1.0000	0.9873	0.8201	0.9198	0.9423
B. vietnamiensis LMG10929T	0.9883	0.9882	0.9902	0.9882	0.9881	0.9802	0.9922	0.9863	0.9792	0.9820	0.9881	0.9879	0.9810	0.9839	0.9855	0.9869	0.9774	0.9921	0.9861	0.9853	0.9786	0.9911	0.9775	0.9824	0.9825	0.9824	0.9871	0.9837	0.9578	0.9873	1.0000	0.8201	0.9156	0.9391
B. fungorum LMG20227T	0.8203	0.8206	0.8196	0.8210	0.8215	0.8270	0.8220	0.8317	0.8247	0.8191	0.8183	0.8224	0.8193	0.8330	0.8255	0.7955	0.8277	0.8123	0.8187	0.7942	0.8387	0.8017	0.8212	0.8157	0.8171	0.8257	0.8112	0.8212	0.8035	0.8201	0.8201	1.0000	0.8544	0.8423
B. gladioli LMG2216T	0.9120	0.9115	0.9108	0.9119	0.9120	0.9160	0.9182	0.9138	0.9090	0.9107	0.9137	0.9127	0.9136	0.9200	0.9159	0.8733	0.9117	0.9016	0.9097	0.8827	0.9255	0.9143	0.9062	0.9084	0.9168	0.9157	0.9258	0.9126	0.8679	0.9198	0.9156	0.8544	1.0000	0.9779
B. glumae LMG2196T	0.9284	0.9281	0.9277	0.9287	0.9286	0.9265	0.9350	0.9267	0.9202	0.9246	0.9277	0.9268	0.9245	0.9297	0.9301	0.9050	0.9202	0.9236	0.9249	0.9150	0.9348	0.9327	0.9184	0.9213	0.9286	0.9252	0.9469	0.9242	0.9074	0.9423	0.9391	0.8423	0.9779	1.0000

CAPÍTULO III

Comparative genome analysis of *Burkholderia cenocepacia* lineages IIIA and IIIB causing onion sour skin

Comparative genome analysis of *Burkholderia cenocepacia* lineages IIIA and IIIB causing onion sour skin

Leandro da Silva Velez[1]; Flavia Figueira Aburjaile[2]; Juan Carlos Ariute Oliveira[2]; Antonio Roberto Gomes de Farias[3]; Ana Maria Benko-Iseppon[2]; Bertram Brenig[4]; Vasco Ariston de Carvalho Azevedo[5]; Adriano Marcio Freire Silva[6]; Elineide Barbosa de Souza[7]; *Marco Aurélio Siqueira da Gama[1]

[1]Departamento de Agronomia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil
[2]Laboratório de Genética e Biotecnologia Vegetal, Universidade Federal de Pernambuco, Recife, 50670-901, Pernambuco, Brazil
[3]Center of Excellence in Fungal Research, School of Science, Mae Fah Luang University, Chiang Rai, 57100, Thailand
[4]University of Göttingen, Institute of Veterinary Medicine, Göttingen, Germany
[5]Laboratório de Genética Molecular e Celular, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Minas Gerais, Brazil
[6]Estação Experimental de Cana-de-Açúcar do Carpina, Universidade Federal Rural de Pernambuco/ Carpina, PE, Brasil
[7]Departamento de Biologia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil

*Correspondence: Marco Aurélio Siqueira da Gama, Departamento de Agronomia, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil. Email: marco.gama@ufrpe.com.br

Abstract

Burkholderia cenocepacia is a Gram-negative bacterium that belongs to the *Burkholderia cepacia* complex (Bcc), family Burkholderiaceae, phylum Proteobacteria, and genus *Burkholderia*. This bacterium is a significant opportunistic human pathogen that

causes infections in immunocompromised individuals and patients with cystic fibrosis and has been found in the environment. This bacterium has also been reported to cause diseases in onion bulbs, leading to significant losses in the productivity of this vegetable. Furthermore, many virulence factors have been identified in this species, increasing the potential damage of this microorganism. Here we provide sequence of 12 genomes isolated from onion with symptoms of sour skin. These genomes are compared with 21 complete genomes isolated from human pathogen and environmental groups, available in public databases and Bcc members. *In silico* taxonomic methods and core-genome were carried out the evolutionary relatedness for *B. cenocepacia* strains. Comparative genomic analyses were also performed across three groups, including plant pathogens, human pathogens, and environmental samples. It was possible to determine that our 12 strains belong to the lineages IIIA and IIIB with few variations between the studied groups through the analysis carried out. According to the combined results of the comparative genomic analyses, the values of ANIm and dDDH agreed with the results verified through the core-genome analysis. These lineages behave like two distinct species. We also report the genomic characterization of 12 original strains and identified genes that are likely to contribute to the pathogenicity in humans. To our knowledge, this is the first report of a draft genome sequence for *B. cenocepacia* causing sour onion skin. Therefore, our data will be a valuable resource for future studies and contribute to the comprehensive genomic resources available for Burkholderiaceae.

1 **INTRODUCTION**

Burkholderia cenocepacia is considered an opportunistic pathogen able to cause infections in immunocompromised human individuals and patients with cystic fibrosis (CF) (Chiarini et al., 2006; Wallner et al., 2019) as well as cause onion sour skin (Baia et al., 2020; Oliveira et al., 2017, 2019) and banana finger-tip (Ansari et al., 2019; Lee & Chan, 2007).

This bacterium has also been reported as a plant-beneficial endophytic bacterium (Ho & Huang, 2015), plant growth promoter and biocontrol agent against fungi (Ho et al., 2015), endosymbionts of fungi (Izumi et al., 2007), in the rhizosphere of corn (Zekic et al., 2017b), in samples of water (Ginther et al., 2015a) and soil (Martina et al., 2018). In addition, strains of *B. cenocepacia* have shown potential for degrading pollutants and bioremediation (Martínez-Ocampo et al., 2015). Such a diversity of hosts from three different kingdoms is unusual, turning this group of bacteria one of the most interesting from the point of view of adaptability and opportunism.

Phylogenetic analysis based on the *recA* gene sequences has shown that strains of *B. cenocepacia* may be shared in four lineages: IIIA, IIIB, IIIC, and IIID (Vandamme et al., 2003). Interestingly, *B. cenocepacia* lineages IIIA and IIIB have been found more frequently in clinical environments (Golini et al., 2006), causing also plant disease (Ansari et al., 2019; Baia et al., 2020; Lee & Chan, 2007; Oliveira et al., 2017, Oliveira et al., 2019). In turn, lineage IIIC has been found exclusively in soil (Vandamme, 2003), whereas lineage IIID has also been detected in human clinical settings. Notwithstanding, the whole genomes of strains of the lineages IIIA and IIIB were newly analyzed using average nucleotide identity (ANI) and *in silico* DNA-DNA hybridization. Accordingly, these lineages shall be reclassified as *B. cenocepacia stricto sensu* and *B. servocepacia*, respectively (Wallner et al., 2019). In addition, multilocus sequence analysis (MLSA) has shown that the lineage IIID and *B. contaminans* show a low level of divergence and probably are synonymous (Baia et al., 2021).

Core-genome analyses using strains of *B. cenocepacia* obtained from different ecological niches have shown that the lineage IIIA harbor predominantly clinical strains presenting conservated virulence factors associated with human pathogenesis and lack several genes associated with plant pathogenesis. In contrast, lineage IIIB harbors mainly environmental strains and plant pathogens that show an opposite distribution of these genes

(Wallner et al., 2019). On the other hand, strains of both lineages IIIA and IIIB have been recently found to cause onion sour skin (Baia et al., 2021). However, a characterization of the genetic repertory of these strains was not yet performed. Thus, we performed the genomic sequencing of strains of *B. cenocepacia* lineages IIIA and IIIB associated with onion sour skin to help understand these bacteria's behavior concerning their different hosts and the role of the plants in the evolution of these bacteria. In addition, given the lack of information on virulence factors involved in the colonization and pathogenicity process on onions, our study also characterized virulence and pathogenicity factors involved in the plant-pathogen interaction between this bacterium and the onion bulbs.

2 **MATERIALS AND METHODS**

2.1 Genomes of *B. cenocepacia* lineages IIIA and IIIB

The genomes of five strains of *B. cenocepacia* lineage IIIA (CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147, and CCRMBC157) and seven strains of *B. cenocepacia* lineage IIIB (CCRMBC18, CCRMBC23, CCRMBC27, CCRMBC31, CCRMBC60, CCRMBC67, and CCRMBC109) were sequenced in this study (Table 1). These strains were isolated from onion bulbs showing sour skin symptoms in the semi-arid region of Northeast Brazil and were previously characterized using a polyphasic approach (Baia et al., 2021). The strains are stored at the Phytopathogenic Bacterial Culture Collection of the Phytobacteriology Laboratory (LAFIBAC), Universidade Federal Rural de Pernambuco (UFRPE). Additionally, 21 complete genome of *B. cenocepacia* (J2315, 842, 895, H111, ST32, toggle4, k56-2, VC12308, AU1054, CR318, HI2424, PC184MULKS, VC12802, MC-03, FL-5-3-30-51-D7, PS27, YG3, MSMB384WGS, VC7848, DDS22E1 and DWS37E2) classified in this study as human pathogens (blood, cord blood, nasal inflammation, and CF sputum) and environmental (soil, rhizosphere of maize, endophytic in

poplar, P-solubilizing, and isolated from water) were included in the dataset (Table 1). These genomes were obtained from Genbank/NCBI https://www.ncbi.nlm.nih.gov/genome/?term=burkholderia+cenocepacia in June of 2021.

2.2 DNA Extraction and Sequencing

The DNA of the strains was extracted using the MiniPrep kit for bacterial genomic DNA extraction (Axygen Biosciences, Massachusetts, USA) following the manufactures recommendations. The Biodrop (Biodrop^{TM}, Cambridge, Canada) was used to assess the quality and quantity of DNA samples. Libraries were performed by paired-sequenced (2x 150 bp). The genome sequencing was performed in HiSeq 2500 platform (Illumina, EUA).

2.3 Quality, assembly, and genomic annotation

The quality of raw sequencing data was evaluated using FastQC v0.11.8 (Andrews, 2010). Adapter screening and quality filtering of reads were performed using Trimmomatic 0.36 (Bolger et al., 2014).

The quality of the assemblies was checked using QUAST v5.0.2 (Gurevich et al., 2013). *De novo* genome assemblies from the 12 strains were performed using SPAdes genome assembler v.3.13.0 (Bankevich et al., 2012). The genomes were automatically annotated using Prokka tool v.1.13.3 (Seemann, 2014).

2.4 *recA* gene

The phylogenetics lineages of strains of *B. cenocepacia* were determined from each strain used in this study using sequences of the *recA* gene. The *recA* sequences of the strains J2315^{T}, AU1054, LMG 19230, and LMG 21462 were used as representative sequences of *B. cenocepacia* lineages IIIA, IIIB, IIIC, and IIID, respectively (https://pubmlst.org/bigsdb?db=pubmlst_Bcc_seqdef&page=query&scheme_id=1).

Sequences of the *recA* gene of the type strains of 22 species of Bcc, and the 21 *B. cenocepacia* strains available at GenBank (https://www.ncbi.nlm.nih.gov) downloaded from PubMLST were included in the dataset. In addition, the *recA* sequences of type strains of *B. glalioli*, *B. glumae*, and *P. fungorum* available from PubMLST Bcc and GenBank were also included in the analyses as an outgroup. Multiple sequences alignment (MSA) was performed with the online version of MAFFT 7 (Katoh et al., 2019; Kuraku et al., 2013) using the G-INS-i integrative refinement method 200PAM/k = 2 nucleotide scoring matrix. The alignment used was the result masked that showed a score of 99% of the specific residues. Maximum likelihood (ML) trees were inferred for all 33 strains. The phylogenies were estimated in RAxML – HPC2 (Stamatakis, 2014) implemented on the CIPRES Science Gateway portal (https://www.phylo.org/portal2/login!input.action). The ML tree searches were performed assuming the GTRGAMMA model and bootstrap support were calculated with 1000 pseudoreplicates.

2.5 Genomic comparisons, core-genome calculation, and phylogenomic analyses

Average nucleotide identity (ANIm) and tetranucleotide usage patterns (TETRA) were performed using pyani.py (Pritchard et al., 2015) subsequently used to Pyani tool in ANIm (MUMmer). Digital DNA-DNA hybridization (dDDH) was calculated by GGDC 2.1 (genome-to-genome distance calculator, http://ggdc.dsmz.de/distcalc2.php) under the formula 2 with the alignment tool Blast+ (Meier-Kolthoff et al., 2014). *B. gladioli*, *B. glumae*, and *Paraburkholderia fungorum* were used as an outgroup.

Core-genome analyses were performed without the outgroup strains used previously, aimed at high precision to obtain the core genes. The strains were performed using Roary v3.13.0 (Page et al., 2015).

Two core-genome analyses were performed, the first one with the 12 strains sequenced in this study along with 21 strains of *B. cenocepacia* lineage IIIA and IIIB

available at Genbank/NCBI, and core-genome analysis with all these strains and the 22 type strains of the Bcc. A cutoff of 95% identity was used, and according to the software's default configuration, the resulting orthologous were classified as core-, softcore-, shell-, and cloud-genes. The core genes were automatically aligned with MAFFT v. 7.310 (Katoh & Standley, 2013), using the -e –mafft parameter in Roary, creating a multi-FASTA alignment of the nucleotide sequence of all the core genes. This file was then used as input to build maximum-likelihood phylogenomics by IQ-TREE software v.2.0.4 (Nguyen et al., 2015), in a local computer, using the automatic selection of nucleotide substitution model on ModelFinder for selection of the best fitting DNA substitution model (Kalyaanamoorthy et al., 2017). The node support was assessed with ultrafast bootstrap (Minh et al., 2013) using 100,000 replicates. The maximum likelihood tree was visualized using the FigTree software v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) (Rambaut: FigTree Version 1.4.0).

2.6 Virulence factors in plant pathogenic strains of *B. cenocepacia* lineages IIIA and IIIB

The search for virulence factors in genomes of strains was performed through the Virulence Factor Database – VFDB (http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi) to genes encoding known virulence factors (E-value <1e-15 and sequence identity > 95%) (Liu et al., 2019). The functional annotations were obtained from the categories provided by the VFDB and compared with the annotations in UniProt (https://www.uniprot.org/) and Pfam (https://pfam.xfam.org/) databases.

2.7 Cellular localization, Isoelectric point, and Molecular weight

The bacterial localization prediction PSORTb version 3.0.2 tool (https://www.psort.org/psortb/) was used to predict the cellular localization of the proteins (Yu et al., 2010) from whole-genomes of the strains CCRMBC56, J2315, and

MSMB384WGS from lineage IIIA and CCRMBC56, VC7848, and MC03 from lineage IIIB. Due to these strains having a more significant amount of virulence factors compared to the other strains of each studied group. JvirGel 2.0 software (http://www.jvirgel.de/index.html) was used to predict the theoretical isoelectric points (pI) and molecular weights (MW) of proteins (Hiller et al., 2006). These parameters were verified in the whole-genomes and directly for the virulence factors found by the VFDB database.

3 **RESULTS**

3.1 Sequencing, assembly, and genomic annotation

The genomes of the strains CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147, and CCRMBC157of *B. cenocepacia* lineage IIIA presented a genome size ranging from 7.68 to 7.96 Mb, G + C content from 67.11 to 67.29%, total gene number from 6883 to 7167, and the number of CDSs from 7038 to 7235. In turn, the genomes of the CCRMBC18, CCRMBC23, CCRMBC27, CCRMBC31, CCRMBC60, CCRMBC67, and CCRMBC109 of *B. cenocepacia* lineage IIIB presented a variation ranging from 7.54 to 7.86 Mb in genome size, G + C content from 66.72 to 66.93%, total gene number from 6743 to 8261, and the number of CDSs from 6833 to 8261 (Table 1). The Busco results showed a high level of single-copy ortholog genes retrieval for the genome assemblies of the strains of both *B. cenocepacia* lineages isolated from onions with symptoms of sour skin, with at least a 98% ratio.

3.2 *recA* gene analysis

The Maximum likelihood (ML) trees constructed with the *recA* sequences grouped the strains in the lineages IIIA and IIIB. The strains CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147, and CCRMBC157 grouped along with strain J2315^{T} of the *B. cenocepacia* lineage IIIA, while the strains CCRMBC18, CCRMBC23, CCRMBC27,

CCRMBC31, CCRMBC60, CCRMBC67, and CCRMBC109 were grouped with strains AU1054 of the *B. cenocepacia* lineage IIIB (Figure 1). The strains DDS22E1, DWS37E2, and YG3 of the *B. cenocepacia* did not group in any of the two clades belonging to the lineages IIIA and IIIB.

3.3 Genomic comparisons, core-genome calculation, and phylogenomic analyses

Regardless of whether the strains were obtained causing onion sour skin, in nosocomial infections, or environmental samples, ANIm values above 99.0% were observed among all genomes of the strains of *B. cenocepacia* lineages IIIA. In turn, ANIm values above 97.8% were observed among all genomes of the strains of *B. cenocepacia* lineage IIIB (Figure 2). ANIm values ranging from 95.0 to 95.1% were observed between the genomes of the strains of *B. cenocepacia* lineages IIIA and IIIB. In addition, ANIm values ranging from 86.7 to 93.0% were observed among all genomes of the strains of *B. cenocepacia* lineages IIIA and IIIB concerning genomes of other BCC species.

Tetra nucleotide analyses (TETRA) showed correlations coefficients values above 0.99 among most Bcc species (Supplementary Table 1). TETRA correlations coefficients values below 0.99 were observed only when the comparisons were performed concerning *B. thailandensis* (E264^{T}; TETRA value = 0.95-0.96), *B. dolosa* (LMG18943^{T}; TETRA value = 0.98), and *B. multivorans* (ATCC BAA 247^{T}; TETRA value = 0.98).

dDDH values above 89% were observed among all genomes of the strains of *B. cenocepacia* lineage IIIA, while dDDH values were above 78.2% were observed among all genomes of the strains of *B. cenocepacia* lineage IIIB (Figure 2). dDDH values ranging from 89 to 99.8% were observed between the genomes of the human pathogen and environmental strains of *B. cenocepacia* lineage IIIA and ranging from 78.2 to 100% in lineage IIIB. In addition, dDDH values ranging from 26.7 to 48.6% were observed among all genomes of the

strains of *B. cenocepacia* lineages IIIA and IIIB and genomes of other Bcc species (Figure 2).

The strains DWS37E2, DDS22E1, and YG3 did not present results following the requirements for classification as *B. cenocepacia*. The strain DWS37E2 presented ANI values ranging from 36.9 to 90.23% concerning strains of lineage IIIA and 37 to 90.34% in lineage IIIB. The DD22E1 strain presented ANI values ranging from 37.9 to 90.59% among the lineage IIIA and between 37.4 and 90.61% in lineage IIIB. The YG3 strain presented ANI values between 56.4 and 94.63% for the strains of the IIIA lineage and 58 to 94.98% for the IIIB strains. These same strains showed dDDH values below 59.4% for all *B. cenocepacia* strains studied.

The core-genome analyses performed with the genomes of the 33 strains of *B. cenocepacia* lineages IIIA and IIIB revealed the presence of 30.906 genes, with 1482 core genes (99% in >= 100% of strains), 642 soft-core genes (95% in >= 99% of the strains), 6.856 shell genes (15% in >= 95% of the strains), and 21.926 cloud genes (0% in less than 15% of the strains). Regardless of where the strains were obtained, a phylogenetic tree generated using ML build using core-genes grouped all strains of *B. cenocepacia* lineage IIIA in an exclusive clade. In turn, all strains of *B. cenocepacia* lineage IIIB were grouped in another exclusive clade. The strains DWS37E2, DDS22E1, and YG3 were not grouped along with any strain and formed individual clades (Figure 3).

The supplementary core-genome analysis performed with the genomes of the 33 strains of *B. cenocepacia* lineages IIIA and IIIB and type strains of the Bcc species revealed the presence of 100.452 genes, with 303 core genes (99% in >= 100% of strains), 792 soft-core genes (95% in >= 99% of the strains), 7.111 shell genes (15% in >= 95% of the strains), and 92.246 cloud genes (0% in less than 15% of the strains). Like the first approach, a phylogenetic tree generated using ML build using core-genes grouped all strains of *B. cenocepacia* lineage IIIA in an exclusive clade. In contrast, all strains of *B. cenocepacia*

lineage IIIB were grouped in another exclusive clade (Figure 4). The strains DWS37E2, DDS22E1, and YG3 did not group into any of these two clades belonging to the lineages IIIA and IIIB. In addition, due to ANI, dDDH, and core-genome analysis results, the strains DWS37E2, DDS22E1, and YG3 were removed from the following analyses.

3.4 Virulence Factors

The output of the VFDB database showed a ranging from 70 to 97 genes related to virulence factors to humans among the genomes of the 33 strains of *B. cenocepacia* lineages IIIA and IIIB. Regarding the strains of *B. cenocepacia* lineages IIIA (CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147, and CCRMBC157) and IIIB (CCRMBC18, CCRMBC23, CCRMBC27, CCRMBC31, CCRMBC60, CCRMBC67, and CCRMBC109) causing onion sour skin were observed from 71 to 81 and 94 to 97 genes related to virulence factors, respectively. Regarding the human pathogen strains, the number of virulence factor genes ranged from 71 to 94 for lineage IIIA and from 73 to 77 for lineage IIIB. When analyzing the environmental strains, this variation in the IIIB lineage ranged from 72 to 89 genes of virulence factors, noting that only the environmental strain MSMB384WGS belongs to the IIIA lineage, which showed 70 genes (Table 1).

The main virulence factors found included genes involved in the classes: actin-based motility, adherence, antiphagocytosis, invasion, secretion system, toxin, endotoxin, immune evasion, serum resistance, serum resistance, and immune evasion, cell surface components, anaerobic respiration, colonization, and immune evasion, iron uptake, and stress adaptation. Regarding plant pathogens, human pathogens, and strains from the environment, the strains of *B. cenocepacia* lineage IIIA that presented the highest virulence factors were CCRMBC56, J2315, and MSMB384WGS, reaching 81, 94, and 70 genes. In contrast, the strains of *B. cenocepacia* lineage IIIB that presented the highest virulence factors were CCRMBC23, VC7848, and MC-03, reaching 97, 77, and 89 genes.

3.5 Protein characterization

The molecular weight and isoelectric point for the strains of *B. cenocepacia* lineage IIIA belonging to the groups of plant pathogen (CCRMBC56), human pathogen ($J2315^T$), and environment (MSMB384WGS) showed the molecular weight of the proteins of the genome ranging from 5.35 to 484.79 kDa and the isoelectric point 2.9 to 12.95 pH, 5.35 to 484.57 kDa and isoelectric point ranging from 2.9 to 13.55, and 4.17 to 484.84 kDa and isoelectric point 2.9 to 12.95 pH. These strains have 5178, 5429, and 5215 proteins, respectively.

The IIIB lineage strains CCRMBC23, *B. cenocepacia* VC7848, and *B. cenocepacia* MC-03 show the molecular weight of the proteins of the genome ranging from 8.02 to 477.05, 9.57 to 477,39, and 5.96 to 477.88 kDa, and isoelectric point ranging from 3.00 to 12.9, 3.20 to 12.96, and 2.77 to 12.95 pH. These strains have 5290, 4963, and 5278 proteins. (Figure 5).

When the strains were analyzed only with virulence factors, it was verified for *B. cenocepacia* lineage IIIA a molecular weight ranging from 9.48 to 67.74 and isoelectric point ranging from 3.88 to 10.95 pH for strain CCRMBC56. For the strain J2315, the variation in molecular weight was 9.48 to 288.70 and in the isoelectric point 4.91 to 10.95 and for the environmental strain MSMB384WGS this variation in molecular weight was between 9.48 and 67.72 kDA, and the isoelectric point was 3.88 to 10.95 pH.

For the strains of *B. cenocepacia* lineage IIIB, the variation was from 9.45 to 267.89 in molecular weight and from 4.32 to 11.19 in CCRMBC23 strain, 9.48 to 299.97 in molecular weight and from 4.09 to 11.21 in isoelectric point in VC7848 strain, and finally, the variation was 9.45 to 365.73 in molecular weight and from 4.21 to 11.24 in isoelectric point for strain MC-03 (Figure 6).

In general, the variation between the strains analyzed was minimal. Approximately 43% of proteins are basic, and 57% are acidic in all analyzed strains from *B. cenocepacia* lineages IIIA and IIIB. The predicted subcellular location for all the virulence factors studied was cytoplasmic membrane, extracellular, cytoplasmic, outer membrane, periplasmic, and unknown. All strains showed the same profile.

4. **DISCUSSION**

The whole-genome sequencing of the strains of *B. cenocepacia* lineages IIIA (CCRMBC02, CCRMBC21, CCRMBC56, CCRMBC147, and CCRMBC157) and IIIB (CCRMBC18, CCRMBC23, CCRMBC27, CCRMBC31, CCRMBC60, CCRMBC67, and CCRMBC109) associated with onion sour skin revealed a high percent of G+C content, with average 67.17% in lineage IIIA and 66.83 in lineage IIIB, a large genome with an average of 7.87 Mb in lineage IIIA and 7.70 Mb in lineage IIIB, encoding an average of 7.151 reading frames in lineage IIIA and 7.206 in lineage IIIB (Table 1). These values confirm similar genomic characteristics with the genomes available in public databases and other studies that revealed that *B. cenocepacia* strains have a high percent G+C content, multiple replicons structures, gene duplications, insertions sequences, and mobile components (Loutet & Valvano, 2010). When checked the strain type of these species, the complete genome sequencing of the *B. cenocepacia* strain J2315^{T} revealed that this member of Bcc also contains a large genome with multiple replicons of 8.06 Mb consisting of three chromosomes and a plasmid, which encodes a total of 7.261 reading frames (Holden et al., 2009a).

These characteristics are thought to increase the versatility of *Burkholderia* genomes and their capacity to gain a broad scope of metabolic pathways (Loutet & Valvano, 2010). As verified, with the genomes available in public databases, our 12 genomes have genomic characteristics compatible with the genomes already studied and contain genes in all classes

of virulence factors studied, indicating the need for depth studies to understand this versatility.

The phylogenetic tree built from the *recA* gene grouped the strains of *B. cenocepacia* lineages IIIA and IIIB associated with onion sour skin according to the results presented in the previous study Baia et al. (2020) and Oliveira et al. (2019). Regarding the human pathogen and environmental groups, the strains were grouped as already verified by Wallner 2009. The strains J2315, K56-2, H111, ST32, 895, 842, MSMB384WGS, and VC12308 belonging to lineage IIIA, and strains AU1054, PC184MULKS, MC0-3, HI2424, CR318, VC7848, FL-5-3-30-S1-D7, and VC12802 belong to lineage IIIB are described as a new species named *B. servocepacia*.

Average nucleotide identity (ANI) represents the average nucleotide identity of all orthologous genes shared between two genomes. It offers a complete resolution between strains of the same or closely related species (Jin et al., 2020), as the *B. cenocepacia* lineages IIIA and IIIB. Regardless the source, i.e., if the strains were obtained from onion sour skin, nosocomial infections, or environmental samples, high ANIm values were observed between the isolates within each lineage. However, between the isolates of the two lineages, values from 95.0 to 95.1% were observed, which are within a limit for species division. Theoretically, they should not be considered distinct species. On the other hand, dDDH results clearly showed a division between strains from lineages IIIA and IIIB, with dDDH values above 78% among strains of the same lineage and below 65% among strains of the opposite lineages, agreeing with the proposed value of 70% minimum to dDDH as the recommended standard for delineating species (Depoorter et al., 2020).

TETRA tetranucleotide frequency correlation analysis was used as a complement to ANIm. However, values above 0.99 were observed between all strains, except for three of them. Due to these results, TETRA analysis was not adequate to distinguish the species of the BCC.

Regarding the TETRA correlation coefficients results, this analysis does not have a suitable resolution for species separation within Bcc because it could not differentiate the type Bcc species analyzed in this study. In other words, the cut-off value for species delimitation through this analysis is 0.99 (Richter and Rosselló-Móra 2009), and for the isolates analyzed in this study, values below 0.99 were only possible to verify when analyzed against the outgroup and *B. multivorans*, *B. dolosa*, *B. ubonensis*, *B. thailandensis*, *B. stagnalis,* and *B. vietnamienses*. Nevertheless, for bacteria belonging to the *Xanthomonas* genus, TETRA coefficients satisfactorily may be used for species delimitation (Gama et al., 2018).

Based on the core-genome analysis, Wallner et al. (2019) concluded that the lineage IIIB is phylogenetically distinct from *B. cenocepacia*. They proposed that the *B. cenocepacia* strains (J2315, K56-2, H111, ST32, 895, 842, MSMB384WGS, and VC12308) classified as lineage IIIA, continue to be called *B. cenocepacia*. However, the strains (AU1054, PC184MULKS, MC0-3, HI2424, CR318, VC7848, FL-5-3-30-S1-D7, and VC12802), belonging to lineage IIIB, should be classified as a new species, *Burkholderia servocepacia*. So far, this suggestion has not been accepted yet by the international committee. However, based on this study, it was verified that the strains used in this study behaved in the same way as previously verified. We use six strains for comparative genomic analysis, belonging to the human pathogens, environment, plant-pathogen lineages IIIA and IIIB.

Two core-genome analyzes were performed in this study. The first was performed with our 12 strains, obtained from onion with symptoms of sour skin, together with the 21 *B. cenocepacia* strains available at the NCBI as a complete genome. The second core-genome analysis was performed with the strains used in the first analysis plus the type strain of the members of the Bcc. These results agree with the results verified in the ANI and dDDH analyses and indicate that the lineages IIIA and IIIB are two distinct species, as can also be verified by Wallner et al. (2019), that named the lineage IIIB as *B. servocepacia*. However,

this naming needs confirmation from the List of Prokaryotic names with Standing in Nomenclature (Parte et al. 2020).

Interestingly, regardless of the number of core genes found in the two analyzes that were very contrasting, the structure of the trees was the same. This result shows that some genes are super informative for species delimitation within this genus. Seven housekeeping genes are known to be very informative for this genus. *recA*, *gyrB*, *atpD*, *gltB*, *lepA*, *phaC*, and *trpB* genes have been shown to have high discriminatory power for the differentiation of Bcc species and tested to our strains (Baia et al., 2020).

Human virulence genes were found similarly in the two lineages, regardless of the isolation group. The VFDB database used to determine virulence factors is very effective when discussing virulence factors related to clinical pathogenesis in humans. Because of this, it was necessary to search for genes related to plant pathogenesis through a BLAST using the genes Endo-polygalacturonase (*pehA*) against the genomes used in this study, and it was found that all onion strains have the *pehA* gene, except for the strain CCRMBC109, agreeing with the results already presented demonstrating the pathogenicity of these strains to onion bulbs.

Polygalacturonase is a virulence factor known to be involved in the maceration of plant tissue. This enzyme is encoded by the *pehA* gene and transmitted horizontally through plasmid. According to phytopathogenic strains of *B. cepacia* can produce this enzyme, whereas strains of clinical origin and soil do not. Virulence determinants may be disseminated among the bacteria of the Bcc by horizontal transfer in the onion rhizosphere.

The *pehA* endopolygalacturonase gene presence, a known virulence factor, is related to the pathogenicity in onions (Springman et al., 2009). They detected those only strains consisting of agricultural strains contain the *pehA* gene encoding endopolygalacturonase. The pathogenicity of onion was correlated with the source from which the strains were isolated. Onion pathogenicity and *pehA* were both absent in some strains from cystic fibrosis patients.

In contrast, a related group consisting of some strains from onion field soil or rhizosphere were pathogenic to the onion and contained *pehA*. The characteristic of onion pathogenicity and the presence of the *pehA* gene could be explained by the acquisition of *pehA* by onion strains or the loss of ancestral *pehA* from most clinical isolates.

The ability of *B. cenocepacia* to produce a wide range of virulence factors as cited above, as well as its resistance to oxidative stress, persistence in macrophages, production of toxins, proteases, and type VI secretion systems, contributes to the clinical relevance of this organism (Bartell et al., 2014).

In this context, studies that clarify the pathogenesis mechanisms of *B. cenocepacia* isolates in plants and humans are not yet fully understood. Additional relevant information about which genes are involved in processes such as pathogenicity, virulence, metabolism, and adaptability of *B. cenocepacia* is mandatory to understand the adaptability of this important pathogen.

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Project numbers 870921/1999-1, 433931/2018-3, and 313581/2020-7) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

5 References

Alikhan, N.-F., Petty, N. K., ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics 2011 12:1*, *12*(1), 1–10. https://doi.org/10.1186/1471-2164-12-402

Andrews, S. (2010). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available Online at: Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ansari, M., Taghavi, S. M., Zarei, S., Miri, K., Portier, P., & Osdaghi, E. (2019). Pathogenicity and molecular phylogenetic analysis reveal a distinct position of the banana fingertip rot pathogen among the *Burkholderia cenocepacia* genomovars. *Plant Pathology*, *68*(4), 804–815. https://doi.org/https://doi.org/10.1111/ppa.12976

Baia, A. D. B., Silva, A. M. F., Ribeiro, B. G., Souza, C. C., Júnior, W. J. S., Balbino, V. Q., Leal, C. M., Farias, A. R. G. de, Souza, E. B., & Gama, M. A. S. (2020). Predominance of *Burkholderia cenocepacia* lineages causing onion sour skin in the semi-arid region of north-east Brazil. *Plant Pathology*, *70*(3), 521–533. https://doi.org/https://doi.org/10.1111/ppa.13311

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. v, Sirotkin, A. v, Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bartell, J. A., Yen, P., Varga, J. J., Goldberg, J. B., & Papin, J. A. (2014). Comparative Metabolic Systems Analysis of Pathogenic *Burkholderia*. *Journal of Bacteriology*, *196*(2), 210–226. https://doi.org/10.1128/JB.00997-13

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carlier, A., Agnoli, K., Pessi, G., Suppiger, A., Jenul, C., Schmid, N., Tümmler, B., Pinto-Carbo, M., & Eberl, L. (2014). Genome Sequence of Burkholderia cenocepacia H111, a Cystic Fibrosis Airway Isolate. *Genome Announcements*, *2*(2). https://doi.org/10.1128/GENOMEA.00298-14

Chiarini, L., Bevivino, A., Dalmastri, C., Tabacchioni, S., & Visca, P. (2006). *Burkholderia cepacia* complex species: health hazards and biotechnological potential. *Trends in Microbiology*, *14*(6), 277–286. https://doi.org/10.1016/j.tim.2006.04.006

Daligault, H. E., Davenport, K. W., Minogue, T. D., Bishop-Lilly, K. A., Broomall, S. M., Bruce, D. C., Chain, P. S., Coyne, S. R., Frey, K. G., Gibbons, H. S., Jaissle, J., Koroleva, G. I., Ladner, J. T., Lo, C.-C., Munk, C., Palacios, G. F., Redden, C. L., Rosenzweig, C. N., Scholz, M. B., & Johnson, S. L. (2014). Whole-Genome Assemblies of 56 *Burkholderia* Species. *Genome Announcements*, *2*(6). https://doi.org/10.1128/GENOMEA.01106-14

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–1403. https://doi.org/10.1101/gr.2289704

Darling, P., Chan, M., Cox, A. D., & Sokol, P. A. (1998). Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. *Infection and Immunity*, *66*(2), 874–877. https://doi.org/10.1128/IAI.66.2.874-877.1998

Depoorter, E., de Canck, E., Peeters, C., Wieme, A. D., Cnockaert, M., Zlosnik, J. E. A., LiPuma, J. J., Coenye, T., & Vandamme, P. (2020). *Burkholderia cepacia* Complex Taxon K: Where to Split? *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01594

Fila, L., & Dřevínek, P. (2017). *Burkholderia cepacia* complex in cystic fibrosis in the post-epidemic period: multilocus sequence typing-based approach. *Folia Microbiologica 2017 62:6*, *62*(6), 509–514. https://doi.org/10.1007/S12223-017-0523-X

Gama, Marco Aurélio Siqueira da, Rosa de Lima Ramos Mariano, Wilson José da Silva Júnior, Antônio Roberto Gomes de Farias, Maria Angélica Guimarães Barbosa, Marisa Álvares da Silva Velloso Ferreira, César Raimundo Lima Costa Júnior, Liliana Andréa Santos, and Elineide Barbosa de Souza. 2018. "Taxonomic Repositioning of *Xanthomonas campestris* pv. *viticola* (Nayudu 1972) Dye 1978 as *Xanthomonas citri* pv. *viticola* (Nayudu 1972) Dye 1978 Comb. Nov. and Emendation of the Description of *Xanthomonas citri* pv. *anacardii* to Include Pigmented Isolates Pathogenic to Cashew Plant." *Https://Doi.Org/10.1094/PHYTO-02-18-0037-R* 108 (10): 1143–53. https://doi.org/10.1094/PHYTO-02-18-0037-R.

Ginther, J. L., Mayo, M., Warrington, S. D., Kaestli, M., Mullins, T., Wagner, D. M., Currie, B. J., Tuanyok, A., & Keim, P. (2015a). Identification of *Burkholderia pseudomallei* Near-Neighbor Species in the Northern Territory of Australia. *PLOS Neglected Tropical Diseases*, *9*(6), e0003892. https://doi.org/10.1371/journal.pntd.0003892

Ginther, J. L., Mayo, M., Warrington, S. D., Kaestli, M., Mullins, T., Wagner, D. M., Currie, B. J., Tuanyok, A., & Keim, P. (2015b). Identification of *Burkholderia pseudomallei* Near-Neighbor Species in the Northern Territory of Australia. *PLOS Neglected Tropical Diseases*, *9*(6), e0003892. https://doi.org/10.1371/JOURNAL.PNTD.0003892

Golini, G., Cazzola, G., & Fontana, R. (2006). Molecular epidemiology and antibiotic susceptibility of *Burkholderia cepacia* complex isolates from an Italian cystic fibrosis centre. *European Journal of Clinical Microbiology and Infectious Diseases*, *25*(3), 175–180. https://doi.org/10.1007/s10096-006-0099-x

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hiller, K., Grote, A., Maneck, M., Münch, R., & Jahn, D. (2006). JVirGel 2.0: computational prediction of proteomes separated via two-dimensional gel electrophoresis under consideration of membrane and secreted proteins. *Bioinformatics*, *22*(19), 2441–2443. https://doi.org/10.1093/bioinformatics/btl409

Ho, Y.-N., Chiang, H.-M., Chao, C.-P., Su, C.-C., Hsu, H.-F., Guo, C., Hsieh, J.-L., & Huang, C.-C. (2015). In planta biocontrol of soilborne Fusarium wilt of banana through a plant endophytic bacterium, *Burkholderia cenocepacia* 869T2. *Plant and Soil*, *387*(1), 295–306. https://doi.org/10.1007/s11104-014-2297-0

Ho, Y.-N., & Huang, C.-C. (2015). Draft Genome Sequence of *Burkholderia cenocepacia* Strain 869T2, a Plant-Beneficial Endophytic Bacterium. *Genome Announcements*, *3*(6). https://doi.org/10.1128/genomeA.01327-15

Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C., Sebaihia, M., Bentley, S. D., Cerdeño-Tárraga, A. M., Thomson, N. R., Bason, N., Quail, M. A., Sharp, S., Cherevach, I., Churcher, C., Goodhead, I., Hauser, H., Holroyd, N., Mungall, K., Scott, P., Walker, D., White, B., … Parkhill, J. (2009a). The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients. *Journal of Bacteriology*, *191*(1), 261–277. https://doi.org/10.1128/JB.01230-08

Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C., Sebaihia, M., Bentley, S. D., Cerdeño-Tárraga, A. M., Thomson, N. R., Bason, N., Quail, M. A., Sharp, S., Cherevach, I., Churcher, C., Goodhead, I., Hauser, H., Holroyd, N., Mungall, K., Scott, P., Walker, D., White, B., … Parkhill, J. (2009b). The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *Journal of Bacteriology*, *91*(1), 261–277. https://doi.org/10.1128/JB.01230-08

Izumi, H., Moore, E. R. B., Killham, K., Alexander, I. J., & Anderson, I. C. (2007). Characterisation of endobacterial communities in ectomycorrhizas by DNA- and RNA-based molecular methods. *Soil Biology and Biochemistry*, *39*(4), 891–899. https://doi.org/10.1016/j.soilbio.2006.10.011

Jacobs, J. L., Fasi, A. C., Ramette, A., Smith, J. J., Hammerschmidt, R., & Sundin, G. W. (2008). Identification and onion pathogenicity of *Burkholderia cepacia* complex isolates from the onion rhizosphere and onion field soil. *Applied and Environmental Microbiology*, *74*(10), 3121–3129. https://doi.org/10.1128/AEM.01941-07

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N., Ren, H., Liang, L., & Yue, J. (2020). Genome-based classification of *Burkholderia cepacia* complex provides new insight into its taxonomic status. *Biology Direct*, *15*(1), 6. https://doi.org/10.1186/s13062-020-0258-5

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, *41*(W1), W22–W28. https://doi.org/10.1093/nar/gkt389

Lee, Y.-A., & Chan, C.-W. (2007). Molecular Typing and Presence of Genetic Markers Among Strains of Banana Finger-Tip Rot Pathogen, *Burkholderia cenocepacia*, in Taiwan. *Phytopathology*, *97*(2), 195–201. https://doi.org/10.1094/PHYTO-97-2-0195

LiPuma, J. J., Spilker, T., Coenye, T., & Gonzalez, C. F. (2002). An epidemic *Burkholderia cepacia* complex strain identified in soil. *The Lancet*, *359*(9322), 2002–2003. https://doi.org/10.1016/S0140-6736(02)08836-0

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, *47*(D1), D687–D692. https://doi.org/10.1093/nar/gky1080

Loutet, S. A., & Valvano, M. A. (2010). A Decade of *Burkholderia cenocepacia* Virulence Determinant Research. *Infection and Immunity*, *78*(10), 4088–4100. https://doi.org/10.1128/IAI.00212-10

Martina, P., Leguizamon, M., Prieto, C. I., Sousa, S. A., Montanaro, P., Draghi, W. O., Stämmler, M., Bettiol, M., de Carvalho, C. C. C. R., Palau, J., Figoli, C., Alvarez, F., Benetti, S., Lejona, S., Vescina, C., Ferreras, J., Lasch, P., Lagares, A., Zorreguieta, A., … Bosch, A. (2018). *Burkholderia puraquae* sp. nov., a novel species of the *Burkholderia cepacia* complex isolated from hospital settings and agricultural soils. *International Journal of Systematic and Evolutionary Microbiology*, *68*(1), 14–20. https://doi.org/10.1099/ijsem.0.002293

Martínez-Ocampo, F., Lozano-Aguirre Beltrán, L. F., Hernández-Mendoza, A., Rojas-Espinoza, L. E., Popoca-Ursino, E. C., Ortiz-Hernández, M. L., Sánchez-Salinas, E., Ramos Quintana, F., & Dantán-González, E. (2015). *Burkholderia cenocepacia* Strain CEIB S5-1, a Rhizosphere-Inhabiting Bacterium with Potential in Bioremediation. *Genome Announcements*, *3*(2). https://doi.org/10.1128/genomeA.00056-15

Meier-Kolthoff, J. P., Klenk, H. P., & Göker, M. (2014). Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 2), 352–356. https://doi.org/10.1099/ijs.0.056994-0

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Oliveira, W. J., Silva, W. A., Silva, A. M. F., Candeia, J. A., Souza, E. B., Mariano, R. L. R., & Gama, M. a. S. (2017). First Report of *Burkholderia cenocepacia* Causing Sour Skin of Onion (Allium cepa) in Brazil. *Plant Disease*, *101*(11), 1950. https://doi.org/10.1094/PDIS-05-17-0759-PDN

Oliveira, W. J., Souza, E. B., Silva, A. M. F., Lima, N. B., Leal, C. M., Candeia, J. A., & Gama, M. A. S. (2019). Elucidating the etiology of onion bacterial scale rot in the semi-arid region

of Northeastern Brazil. *Tropical Plant Pathology*, *44*(6), 494–502. https://doi.org/10.1007/s40858-019-00310-2

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Parte, Aidan C., Joaquim Sardà Carbasse, Jan P. Meier-Kolthoff, Lorenz C. Reimer, and Markus Göker. 2020. "List of Prokaryotic Names with Standing in Nomenclature (LPSN) Moves to the DSMZ." *International Journal of Systematic and Evolutionary Microbiology* 70 (11): 5607. https://doi.org/10.1099/IJSEM.0.004332.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2015). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*, *8*(1), 12–24. https://doi.org/10.1039/C5AY02550H

Rambaut: FigTree version 1.4. 0 - Google Scholar. (n.d.). Retrieved May 12, 2021, from https://scholar.google.com/scholar?cluster=10554043299952493487&hl=en&oi=scholarr

Richter, Michael, and Ramon Rosselló-Móra. 2009. "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition." *Proceedings of the National Academy of Sciences of the United States of America* 106 (45): 19126–31. https://doi.org/10.1073/pnas.0906412106.

Sahl, J. W., Vazquez, A. J., Hall, C. M., Busch, J. D., Tuanyok, A., Mayo, M., Schupp, J. M., Lummis, M., Pearson, T., Shippy, K., Colman, R. E., Allender, C. J., Theobald, V., Sarovich, D. S., Price, E. P., Hutcheson, A., Korlach, J., LiPuma, J. J., Ladner, J., … Wagner, D. M. (2016). The effects of signal erosion and core genome reduction on the identification of diagnostic markers. *MBio*, *7*(5). https://doi.org/10.1128/MBIO.00846-16

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Soares, S. C., Geyik, H., Ramos, R. T. J., de Sá, P. H. C. G., Barbosa, E. G. V., Baumbach, J., Figueiredo, H. C. P., Miyoshi, A., Tauch, A., Silva, A., & Azevedo, V. (2016). GIPSy: Genomic island prediction software. *Journal of Biotechnology*, *232*, 2–11. https://doi.org/10.1016/J.JBIOTEC.2015.09.008

Springman, A. C., Jacobs, J. L., Somvanshi, V. S., Sundin, G. W., Mulks, M. H., Whittam, T. S., Viswanathan, P., Gray, R. L., LiPuma, J. J., & Ciche, T. A. (2009). Genetic Diversity and Multihost Pathogenicity of Clinical and Environmental Strains of *Burkholderia cenocepacia*. *Applied and Environmental Microbiology*, *75*(16), 5250–5260. https://doi.org/10.1128/AEM.00877-09

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. J., & Govan, J. R. W. (2003). *Burkholderia cenocepacia* sp. nov.--a new twist to an old story. *Research in Microbiology*, *154*(2), 91–96. https://doi.org/10.1016/S0923-2508(03)00026-3

Wallner, A., King, E., Ngonkeu, E. L. M., Moulin, L., & Béna, G. (2019). Genomic analyses of *Burkholderia cenocepacia* reveal multiple species with differential host-adaptation to plants and humans. *BMC Genomics*, *20*(1), 803. https://doi.org/10.1186/s12864-019-6186-z

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P., & Truper, H. G. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic and Evolutionary Microbiology*, *37*(4), 463–464. https://doi.org/10.1099/00207713-37-4-463

X, W., X, Z., X, L., Z, H., S, N., T, X., J, Z., H, L., T, W., Y, G., M, H., L, C., & Y, Z. (2019). Physiological, biochemical and proteomic insight into integrated strategies of an endophytic bacterium *Burkholderia cenocepacia* strain YG-3 response to cadmium stress. *Metallomics : Integrated Biometal Science*, *11*(7), 1252–1264. https://doi.org/10.1039/C9MT00054B

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(W1), W52–W58. https://doi.org/10.1093/nar/gkz333

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics (Oxford, England)*, *26*(13), 1608–1615. https://doi.org/10.1093/bioinformatics/btq249

Zekic, F., Weselowski, B., & Yuan, Z. C. (2017a). Complete genome sequence of *Burkholderia cenocepacia* CR318, a phosphate-solubilizing bacterium isolated from corn root. *Genome Announcements*, *5*(23). https://doi.org/10.1128/GENOMEA.00490-17

Zekic, F., Weselowski, B., & Yuan, Z.-C. (2017b). Complete Genome Sequence of *Burkholderia cenocepacia* CR318, a Phosphate-Solubilizing Bacterium Isolated from Corn Root. *Genome Announcements*, *5*(23). https://doi.org/10.1128/genomeA.00490-17

Zlosnik, J. E. A., Zhou, G., Brant, R., Henry, D. A., Hird, T. J., Mahenthiralingam, E., Chilvers, M. A., Wilcox, P., & Speert, D. P. (2015). *Burkholderia* Species Infections in Patients with Cystic Fibrosis in British Columbia, Canada. 30 Years' Experience. *Http://Dx.Doi.Org/10.1513/AnnalsATS.201408-395OC*, *12*(1), 70–78. https://doi.org/10.1513/ANNALSATS.201408-395OC

Tabela 2: General features of strains of *Burkholderia cenocepacia* lineages IIIA and IIIB isolated from onion bulbs showing sour skin symptoms (plant pathogens group), nosocomial infections (human pathogens group), and from environmental samples (environmental group)

	Strains	Geographic Region	Assembly accession	Genome size (Mb)	G+C content (%)	CDSs	Number of VF genes	Lineage	Isolation source	Reference
Plant Pathogen	CCRMBC02	Brazil	SAMN19592372	78757	67.11	7035	71	IIIA	Onion bulbs	This study
	CCRMBC21	Brazil	SAMN19592374	79642	67.12	7117	74	IIIA	Onion bulbs	This study
	CCRMBC56	Brazil	SAMN19592378	79341	67.12	7044	81	IIIA	Onion bulbs	This study
	CCRMBC147	Brazil	SAMN19592383	79267	67.23	7167	78	IIIA	Onion bulbs	This study
	CCRMBC157	Brazil	SAMN19592384	76866	67.29	6883	74	IIIA	Onion bulbs	This study
	CCRMBC18	Brazil	SAMN19592373	75486	66.72	6770	94	IIIB	Onion bulbs	This study
	CCRMBC23	Brazil	SAMN19592375	78601	66.79	7092	97	IIIB	Onion bulbs	This study
	CCRMBC27	Brazil	SAMN19592376	77759	66.80	6981	96	IIIB	Onion bulbs	This study
	CCRMBC31	Brazil	SAMN19592377	78598	66.79	7087	96	IIIB	Onion bulbs	This study
	CCRMBC60	Brazil	SAMN19592379	75976	66.93	6743	95	IIIB	Onion bulbs	This study
	CCRMBC67	Brazil	SAMN19592380	77057	66.92	8261	97	IIIB	Onion bulbs	This study
	CCRMBC109	Brazil	SAMN19592382	76148	66.88	6829	96	IIIB	Onion bulbs	This study
Human Pathogen	*B. cenocepacia* J2315*	Scotland	GCF_000009485.1	8,06	66,92	7201	94	IIIA	Human (CF sputum)	(Holden et al., 2009b)
	B. cenocepacia toggle4	Taiwan	GCF_018223805.1	7,95	67,25	6972	71	IIIA	Human (Blood)	Unpublished
	B. cenocepacia ST 32	Czech Republic	GCF_001484665.1	8,09	67,01	7027	72	IIIA	Human (CF sputum)	(Fila & Dřevínek, 2017)
	B. cenocepacia VC12308	Canada	GCF_001999825.1	7,39	67,03	6530	72	IIIA	Human (CF sputum)	(Zlosnik et al., 2015)
	B. cenocepacia H111	Switzerland	GCF_000236215.2	7,71	67,31	6812	72	IIIA	Human (CF sputum)	(Carlier et al., 2014)
	B. cenocepacia 842	Malaysia	GCF_001606115.1	8,15	66,98	7187	76	IIIA	Human (nasal inflammation)	Unpublished

	B. cenocepacia 895	Malaysia	GCF_001606135.1	8,73	66,74	7817	71	IIIA	Human (Cord blood)	Unpublished
	B. cenocepacia K56-2	Canada	GCF_014357995.1	7,74	67,02	6851	71	IIIA	Human (CF sputum)	(P. Darling et al., 1998)
	B. cenocepacia PC 184 MULKS	USA	GCF_003076415.1	7,07	66,83	6206	74	IIIB	Human (CF sputum)	Unpublished
	B. cenocepacia VC7848	Canada	GCF_001999785.1	7,5	66,9	6647	77	IIIB	Human (CF sputum)	(Zlosnik et al., 2015)
	B. cenocepacia VC12802	Canada	GCF_001999885.1	7,63	67,14	6792	76	IIIB	Human (CF sputum)	(Zlosnik et al., 2015)
	B. cenocepacia AU 1054	USA	GCF_000014085.1	7,28	66,92	6381	73	IIIB	Human (Blood)	(LiPuma et al., 2002)
	B. cenocepacia DDS22E1	Australia	GCF_000755725.1	8,05	66,97	7065	59	-	Human (Aerosol)	(Daligault et al., 2014)
Environment	*B. cenocepacia* MSMB 384 WGS	Australia	GCF_001718895.1	7,74	67,24	6918	70	IIIA	Water	(Ginther et al., 2015b)
	B. cenocepacia HI2424	USA	GCF_000203955.1	7,7	66,8	6867	76	IIIB	Soil	(LiPuma et al., 2002)
	B. cenocepacia MC 03	USA	GCF_000019505.1	7,97	66,58	7060	89	IIIB	Soil	(Jacobs et al., 2008)
	B. cenocepacia FL 5 3 30 51 D7	USA	GCF_001718515.1	6,33	67,04	5616	75	IIIB	Soil	(Sahl et al., 2016)
	B. cenocepacia PS27	India	GCF_014211915.1	7,61	66,96	6600	73	IIIB	Soil	Unpublished
	B. cenocepacia CR 318	Canada	GCF_002007585.1	7,66	66,82	6773	72	IIIB	Corn	(Zekic et al., 2017a)
	B. cenocepacia YG3	China	GCF_003966315.1	8,04	66,79	7016	77	-	Populus	(X et al., 2019)
	B. cenocepacia DWS37E2	Australia	GCF_000764955.1	6,61	66,5	5720	69	-	Soil	(Daligault et al., 2014)

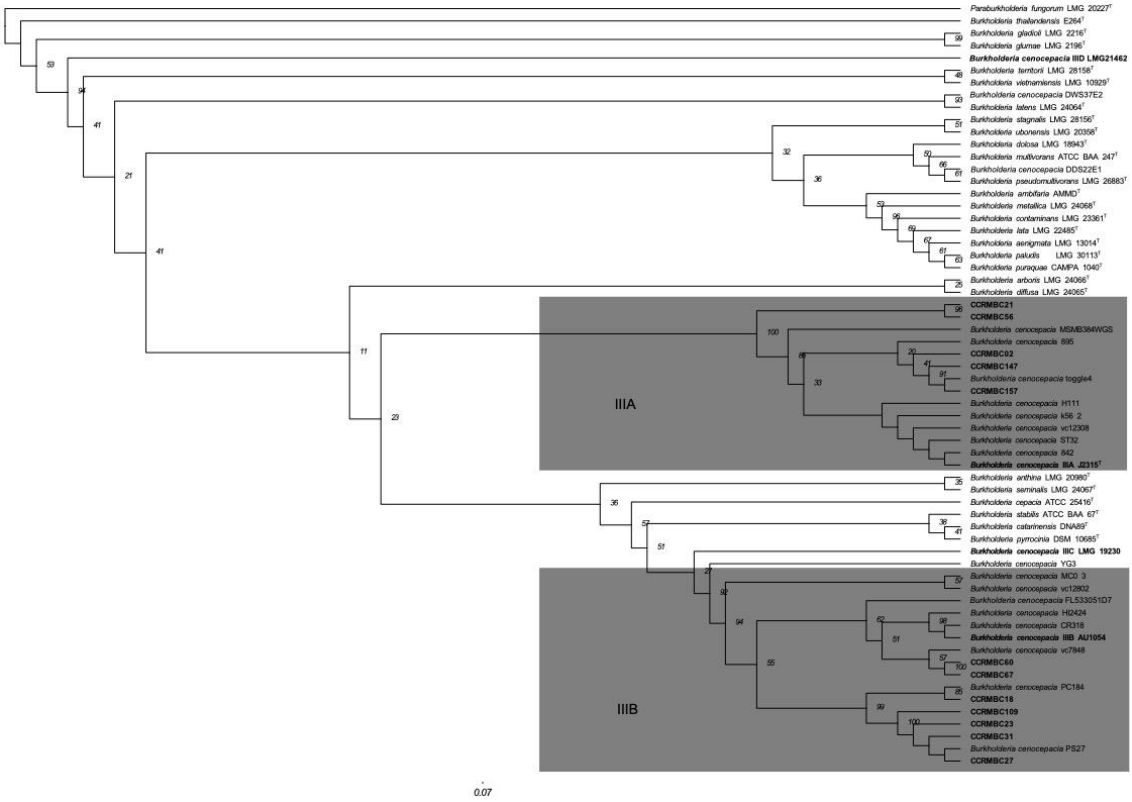

Figure 1: Phylogenetic tree based on the *recA* genes of 12 strains of *Burkholderia cenocepaci*a obtained in this study, against 25 Bcc type strains, represented strains from the lineages (IIIA, IIIB, IIIC, and IIID), 21 *B. cenocepacia* strains from NCBI, and three strains as outgroup. The maximum-likelihood method was used for the construction of the three. Bootstrap scores (1000 replicates) are displayed at each node.

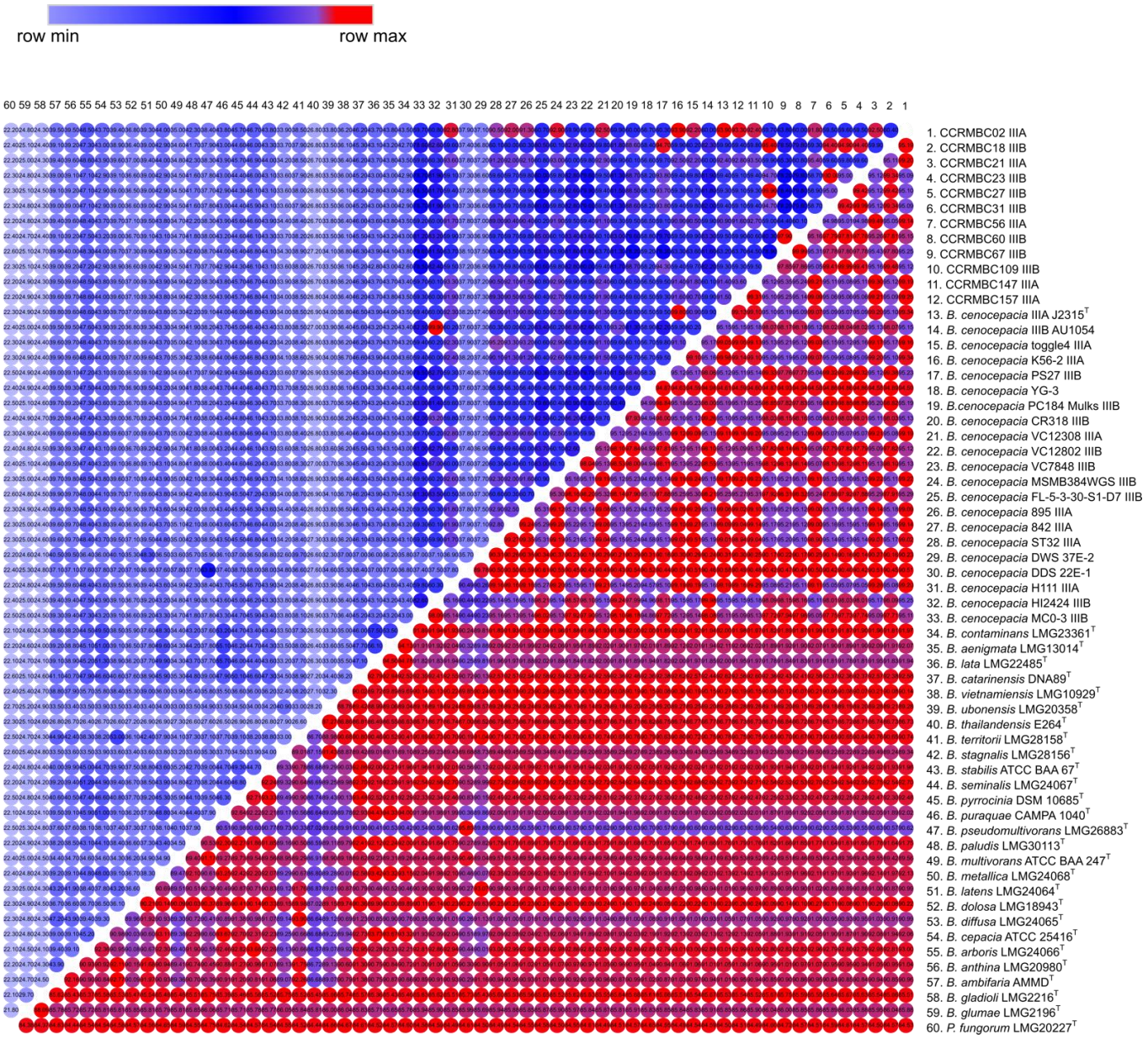

Figure 2: Heatmap of Average Nucleotide Identity (ANIm) and dDDH from pairwise genome comparisons of 12 *Burkholderia cenocepacia* isolated from onion causing onion sour skin in the Northeastern of Brazil, 25 type strains of *B. cepacia* complex, and 21 *B. cenocepacia* strains from NCBI. The outgroup was composed of strains *B. gladioli*, *B.* glumae, and *Paraburkholderia fungorum*. The lower triangle displays the ANI values, and the upper triangle displays the dDDH values.

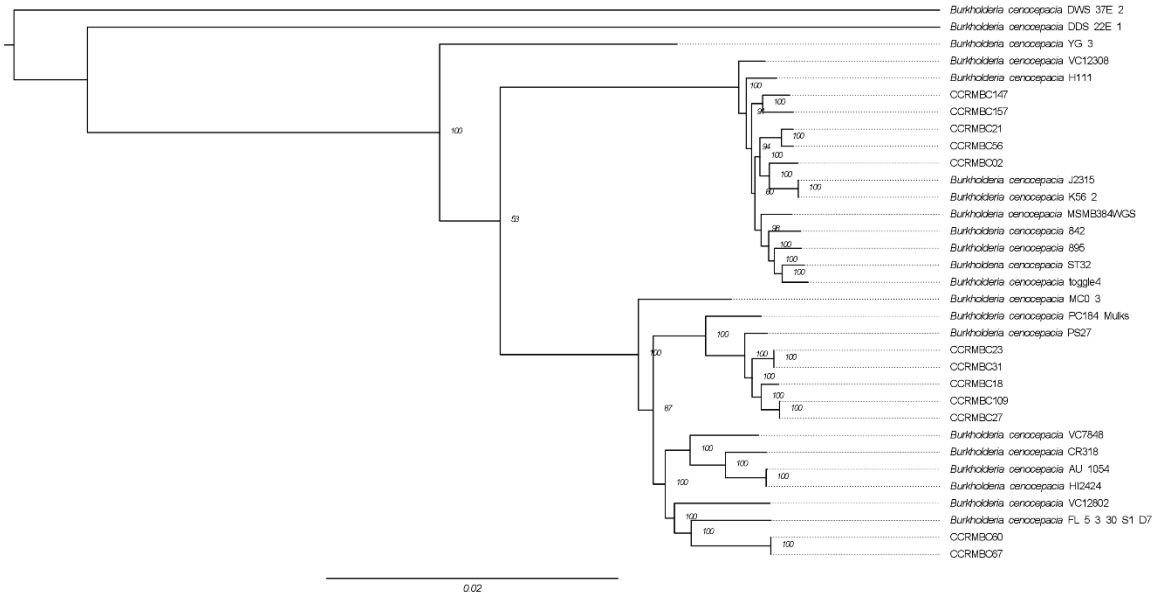

Figure 3: Maximum likelihood tree of 33 core-genome of 12 *Burkholderia cenocepacia* strains from lineages IIIA and IIIB, isolated from onion bulbs symptomatic, and 21 *B. cenocepaci*a strains available at NCBI database as complete genome belonging to human pathogen and environment groups.

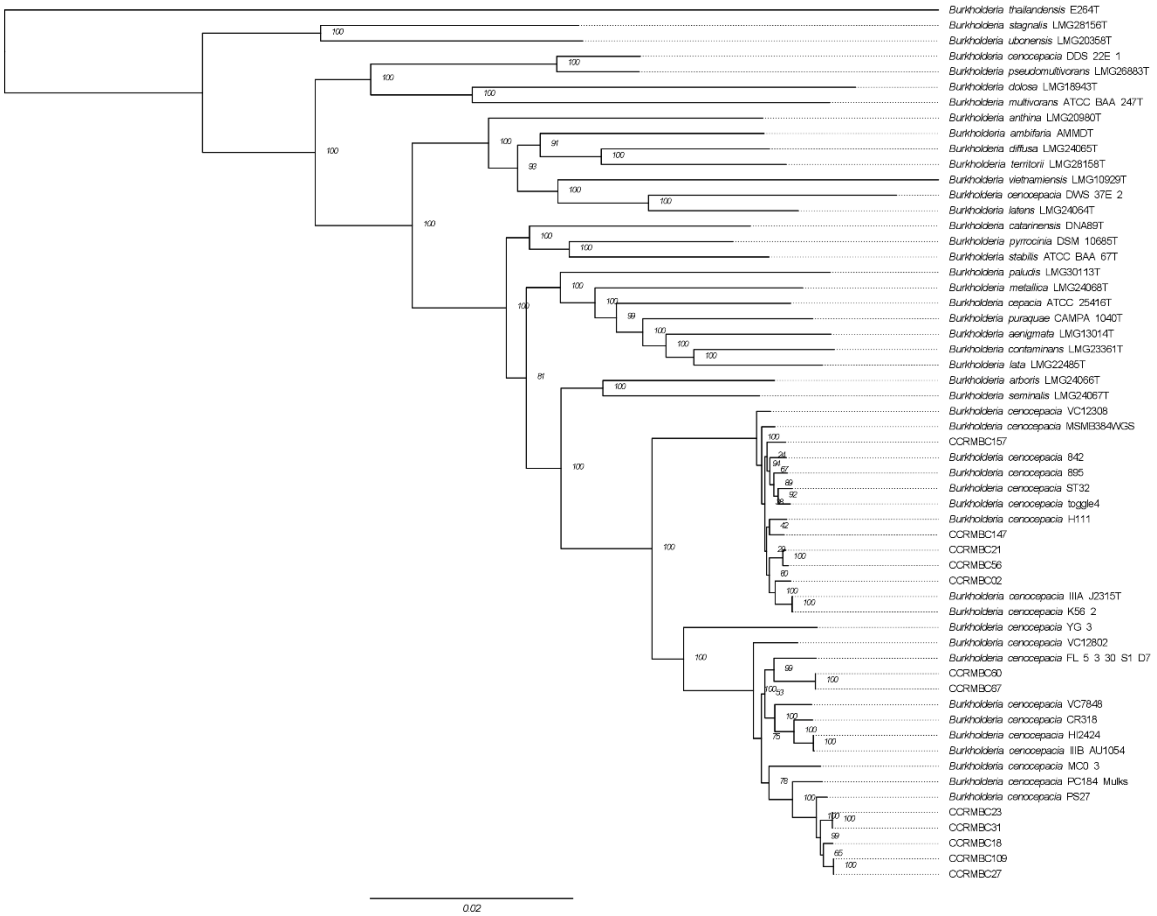

Figure 4: Phylogenomic tree based on the core-genome of 12 *Burkholderia cenocepacia* strains from plant pathology group obtained in this study, 13 strains from human pathogen group, and eight belonging to the environmental group. Besides, type strains from *B. cepacia* complex. The maximum-likelihood method was used for the construction of the three. Bootstrap scores (1000 replicates) are displayed at each node.

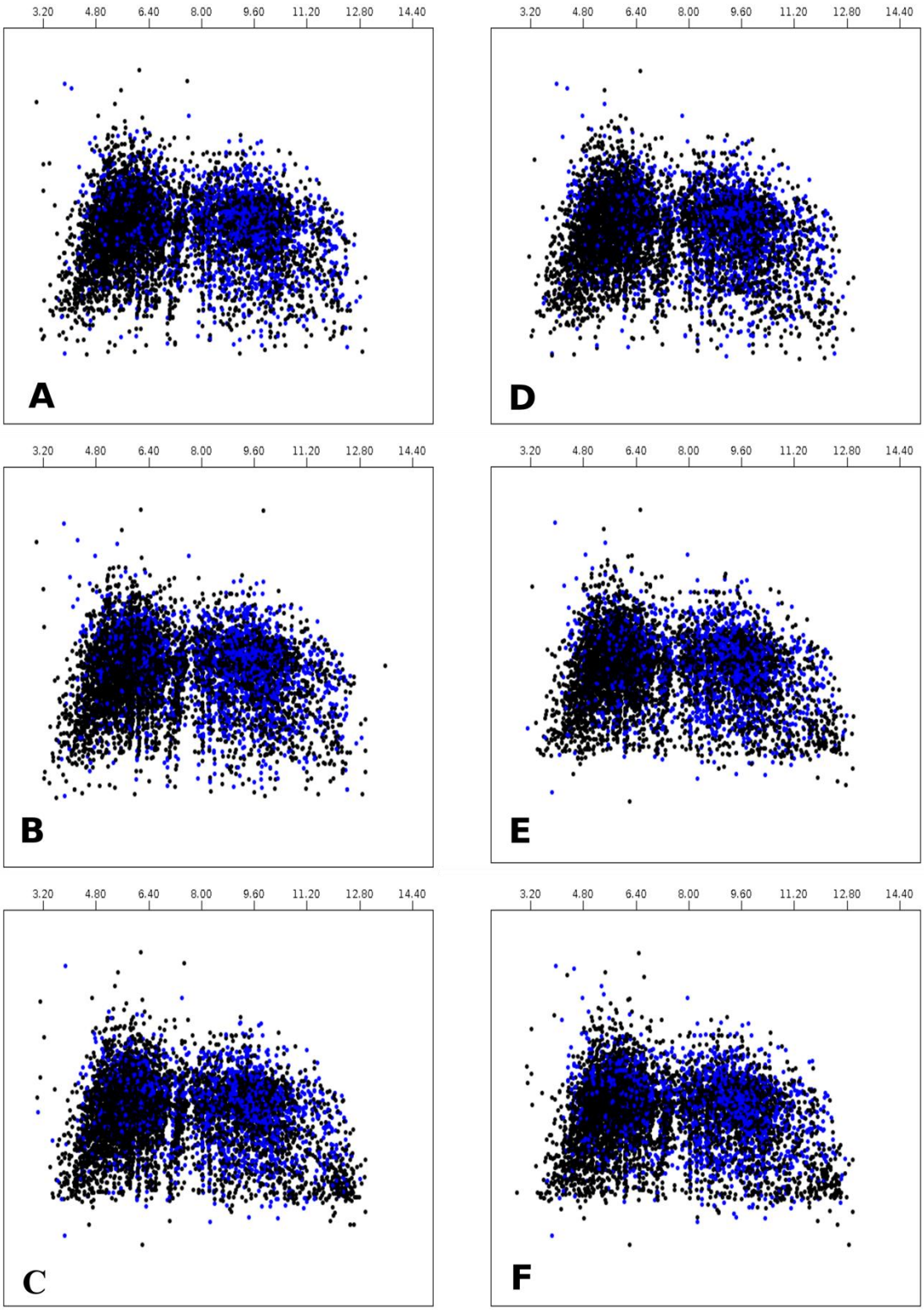

Figure 5: Virtual 2D gel showing the theoretical isoelectric point and molecular weight of genomes belonging to three groups of isolation sources of strains of *Burkholderia cenocepacia*. The gels A, B, and C represent the genomes of the strains from lineage IIIA, and the gels D, E, and F represent the genomes of the strains from lineage IIIB.

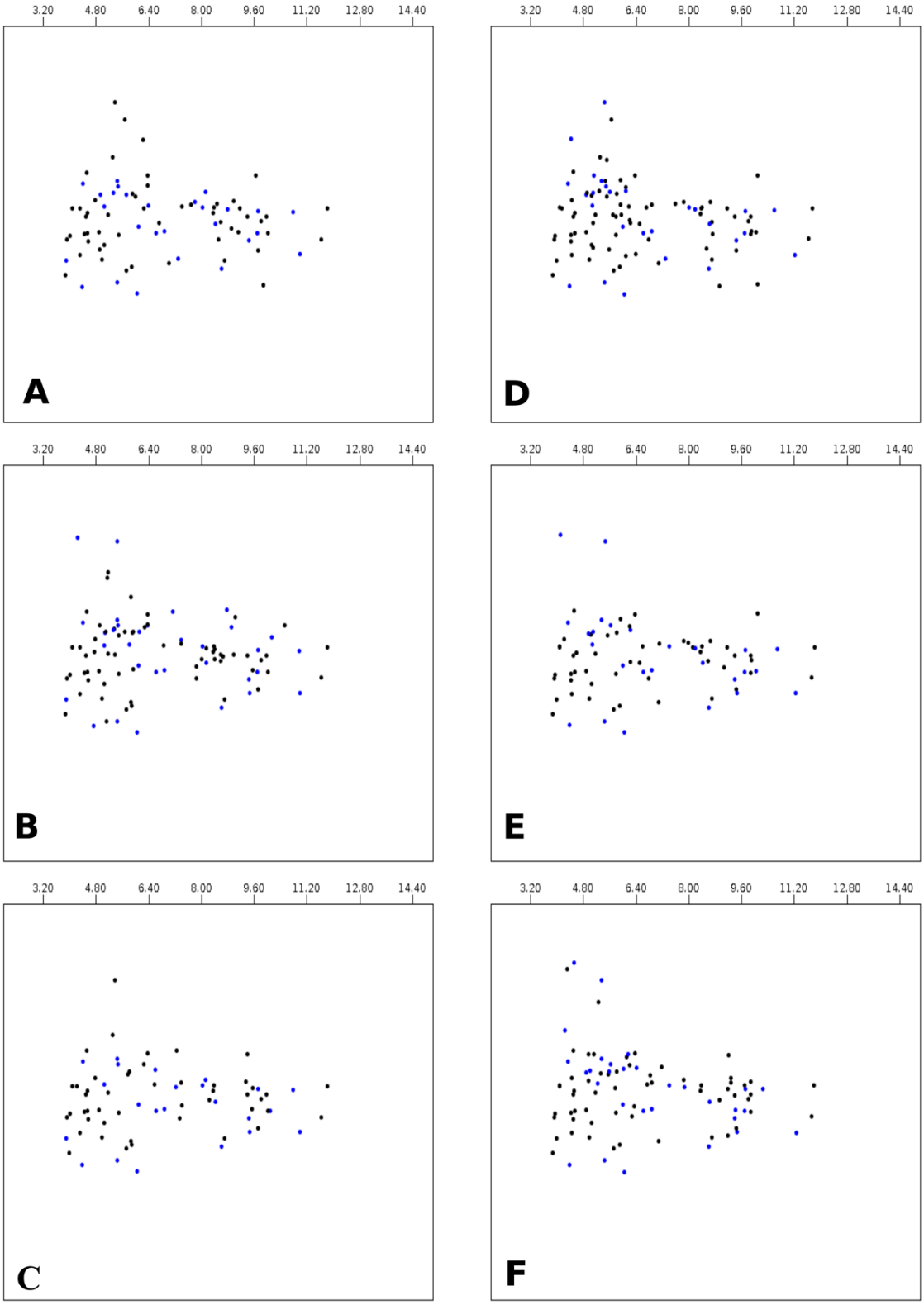

Figure 6: Virtual 2D gel showing the theoretical isoelectric point and molecular weight of the virulence factors outputted from VFDB database belonging to three groups of isolation sources of strains of *Burkholderia cenocepacia*. The gels A, B, and C represent

the corresponding virulence factors of each genome from lineage IIIA, and the gels D, E, and F represent the corresponding virulence factors of each genome from lineage IIIB.

Supporting Information legends

Table 1: General features of *Burkholderia cenocepacia* strains from plant pathogen, human pathogen, and environmental lineages IIIA and IIIB.

Supplementary Table 1: Values of Tetranucleotide from pairwise genomes comparisons of 12 *Burkholderia cenocepacia* strains obtained in this study, against 25 *Bukholderia cepacia* complex type strains, represented strains from the lineages (IIIA and IIIB), 21 *B. cenocepacia* strains from NCBI, and three strains as outgroup.

Figure 1: Phylogenetic tree based on the *recA* genes of 12 *Burkholderia cenocepacia* strains obtained in this study, against 25 *Bukholderia cepacia* complex type strains, represented strains from the lineages (IIIA, IIIB, IIIC, and IIID), 21 *B. cenocepacia* strains from NCBI, and three strains as outgroup. The maximum-likelihood method was used for the construction of the three. Bootstrap scores (1000 replicates) are displayed at each node.

Figure 2: Heatmap of Average Nucleotide Identity (ANIm) and dDDH from pairwise genome comparisons of 12 *Burkholderia cenocepacia* isolated from onion causing onion sour skin in the Northeastern of Brazil, against 25 *Bukholderia cepacia* complex type strains, plus 21 *B. cenocepacia* strains from NCBI, and an outgroup composed of three strains. The lower triangle displays the ANI values, and the upper triangle displays the dDDH values.

Figure 3: Maximum likelihood tree of 33 core-genome of 12 *Burkholderia cenocepacia* strains from lineages IIIA and IIIB, isolated from onion bulbs symptomatic, and 21 *B. cenocepacia* strains available at NCBI database as complete genome belonging to human pathogen and environment groups.

Figure 4: Phylogenomic tree based on the core-genome of 12 *Burkholderia cenocepacia* strains from plant pathology group obtained in this study, 13 strains from human pathogen group, and eight belonging to the environmental group. Besides, type strains from *Burkholderia cepacia* complex. The maximum-likelihood method was used for the construction of the three. Bootstrap scores (1000 replicates) are displayed at each node.

Figure 5: Virtual 2D gel showing the theoretical isoelectric point and molecular weight of genomes belonging to three groups of isolation sources of strains of *Burkholderia cenocepacia*. The gels A, B, and C represent the genomes of the strains from lineage IIIA, and the gels D, E, and F represent the genomes of the strains from lineage IIIB.

Figure 6: Virtual 2D gel showing the theoretical isoelectric point and molecular weight of the virulence factors outputted from VFDB database belonging to three groups of isolation sources of strains of *Burkholderia cenocepacia*. The gels A, B, and C represent

the corresponding virulence factors of each genome from lineage IIIA, and the gels D, E, and F represent the corresponding virulence factors of each genome from lineage IIIB.

CONCLUSÕES GERAIS

CONCLUSÕES GERAIS

- Foram identificadas e descritas duas novas espécies ocasionando podridão das escamas em cebola na região do semiárido, as quais foram classificadas como *Burkholderia semiaridus* e *B. solum*;
- Isolados de *B. cenocepacia* linhagens IIIA e IIIB se comportam e devem ser classificados como duas espécies distintas;
- Isolados fitopatogênicos de *B. cenocepacia* linhagem IIIA e IIIB apresentaram uma maior quantidade de fatores de patogenicidade e virulência a humanos;
- Não há variação no peso molecular e no ponto isoelétrico das proteínas entre as linhagens IIIA e IIIB e as fontes de isolamento;
- Coeficientes de correlação de frequência tetranucleotídica (TETRA) não possui boa resolução para delimitação de espécies dentro do complexo *Burkholderia cepacia*.