GENÔMICA COMPARATIVA DE BACTÉRIAS PROMOTORAS DO CRESCIMENTO VEGETAL E PERTENCENTES A GRUPOS TAXONÔMICOS ASSOCIADOS COM INFECÇÕES OPORTUNISTAS

FRANCISNEI PEDROSA DA SILVA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

Campos dos Goytacazes – RJ
Novembro de 2022

GENÔMICA COMPARATIVA DE BACTÉRIAS PROMOTORAS DO CRESCIMENTO VEGETAL E PERTENCENTES A GRUPOS TAXONÔMICOS ASSOCIADOS COM INFECÇÕES OPORTUNISTAS

FRANCISNEI PEDROSA DA SILVA

Tese apresentada ao Centro de Biociências e Biotecnologia, da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Doutor em Biociências e Biotecnologia.

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

Campos dos Goytacazes – RJ
Novembro de 2022

FICHA CATALOGRÁFICA
UENF - Bibliotecas
Elaborada com os dados fornecidos pelo autor.

S586 Silva, Francisnei Pedrosa da.

Genômica comparativa de bactérias promotoras do crescimento vegetal e pertencentes a grupos taxonômicos associados com infecções oportunistas. / Francisnei Pedrosa da Silva. - Campos dos Goytacazes, RJ, 2023.

96 f. : il.
Inclui bibliografia.

Tese (Doutorado em Biociências e Biotecnologia) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2023.
Orientador: Thiago Motta Venancio.

1. Bioinformática. 2. Genômica. 3. Filogenia. 4. Stenotrophomonas. 5. Pangenoma. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 570

GENÔMICA COMPARATIVA DE BACTÉRIAS PROMOTORAS DO CRESCIMENTO VEGETAL E PERTENCENTES A GRUPOS TAXONÔMICOS ASSOCIADOS COM INFECÇÕES OPORTUNISTAS

FRANCISNEI PEDROSA DA SILVA

Tese apresentada ao Centro de Biociências e Biotecnologia, da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Doutor em Biociências e Biotecnologia.

Aprovada em 29 de novembro de 2022.

Comissão examinadora:

Documento assinado digitalmente
gov.br
DIOGO ANTONIO TSCHOEKE
Data: 24/02/2023 11:14:36-0300
Verifique em https://verificador.iti.br

_____
Dr. Diogo Antonio Tschoeke - UFRJ

Documento assinado digitalmente
gov.br
FABIO LOPES OLIVARES
Data: 24/02/2023 09:43:59-0300
Verifique em https://verificador.iti.br

_____
Dr. Fábio Lopes Olivares - UENF

Documento assinado digitalmente
gov.br
CAIO CEZAR GUEDES CORREA
Data: 24/02/2023 13:32:54-0300
Verifique em https://verificador.iti.br

_____
Dr. Caio Cézar Guedes Corrêa - UENF

Documento assinado digitalmente
gov.br
THIAGO MOTTA VENANCIO
Data: 26/02/2023 09:00:43-0300
Verifique em https://verificador.iti.br

_____
Dr. Thiago Motta Venancio – UENF (orientador)

Campos dos Goytacazes – RJ
Novembro de 2022

DEDICATÓRIA

Aos meus pais Francisco *(in memoriam)* e Sônia, meu amor Isabella, meus irmãos Thiago *(in memoriam)*, Carla, Cláudia, Kátia e Inês, que me apoiaram e me deram forças para enfrentar os momentos mais difíceis desta jornada, mesmo quando pensava não ser possível continuar.

AGRADECIMENTOS

Ao meu orientador e amigo Thiago Venâncio, pelo grande exemplo de profissionalismo, por me apoiar e incentivar nos momentos mais incertos e por acreditar no meu trabalho. Seu apoio foi essencial para a finalização dessa jornada e serei grato eternamente.

Aos meus amigos e pesquisadores do laboratório: Dayana Kelly, Isabella de Oliveira, Kevelin Barbosa, Sarah Henault e Hemanoel Passarelli, pelo companheirismo, força e pelos inúmeros momentos de descontração vividos dentro e fora do laboratório.

Agradeço a todos os meus amigos do NEBBIO e do LQFPP, especialmente Cida Bertoncelli, e aos professores Aline Intorne, Gustavo Lazzaro, Claudete Santa Catarina e Gonçalo Apolinário pelos grandes ensinamentos e pelo apoio nas mais diversas situações.

Ao professor Arnoldo Façanha e todos os membros do programa de pós-graduação, incluindo as secretárias Beatriz Almeida e Marlene Matos, pelo exemplo de competência e profissionalismo.

Ao CNPq e à FAPERJ pelo financiamento durante a pesquisa.

RESUMO

O uso de biofertilizantes à base rizobactérias promotoras do crescimento vegetal tem ganhado cada vez mais destaque devido ao baixo impacto ambiental em comparação com fertilizantes químicos. Entretanto, a ampla adoção das rizobactérias apresenta grandes desafios, dentre os quais está a identificação de cepas que possam causar infecções oportunistas. Diversas rizobactérias têm sido associadas a manifestações clínicas em humanos, incluindo espécies dos gêneros: *Stenotrophomonas*, *Alcaligenes*, e membros da família *Burkholderiaceae*, o que levanta preocupações acerca da segurança destes microrganismos em aplicações biotecnológicas. O objetivo central da presente tese é realizar análises comparativas e sistemáticas dos genomas disponíveis em cada um destes grupos. O primeiro capítulo da tese descreve a análise comparativa com 67 isolados clínicos e ambientais do gênero *Stenotrophomonas,* incluindo o isolado ambiental *Stenotrophomonas maltophilia* UENF-4GII, que nos permitiu identificar grupos gênicos associados exclusivamente com isolados clínicos e ambientais. No segundo capitulo versa sobre a genômica comparativa do gênero *Alcaligenes*, que nos permitiu estabelecer novas reclassificações taxonômicas e prospectar isolados com potencial biotecnológico, especialmente para a promoção de crescimento vegetal e biorremediação. Além disso, identificamos diferenças significativas na composição de genes de resistência antimicrobiana entre isolados clínicos e ambientais. No terceiro capítulo, reportamos uma investigação genômica em larga escala de 1.921 genomas pertencentes a família *Burkholderiaceae.* Através de análises filogenômicas, reclassificamos 61 genomas a nível de gênero. Prospectamos os gêneros com maior potencial para biofertilização e identificamos perfis distintos na composição dos genes de virulência e resistência de gêneros com espécies benéficas e patogênicas. Coletivamente, nossos resultados forneceram informações importantes sobre a diversidade genômica de *Stenotrophomonas*, *Alcaligenes* e *Burkholderiaceae*, que servem tanto como base para a prospecção de isolados candidatos com baixo risco teórico de patogenicidade, quanto para a identificação molecular de potenciais patógenos.

Palavras-chave: Bioinformática, Genômica, Filogenia, *Stenotrophomonas*, *Alcaligenes*, *Burkholderiaceae*

ABSTRACT

The use of rhizobacteria-based biofertilizers that promote plant growth has gained increasing prominence due to the low environmental impact compared to chemical fertilizers. However, the wide adoption of rhizobacteria presents major challenges, among which is the identification of strains that can cause opportunistic infections. Several rhizobacteria have been associated with clinical manifestations in humans, including species from the genera: *Stenotrophomonas*, *Alcaligenes*, and members of the *Burkholderiaceae* family, which raises concerns about the safety of these microorganisms in biotechnological applications. The main objective of this thesis is to carry out comparative and systematic analyzes of the genomes available in each of these groups. The first chapter of the thesis describes the comparative analysis with 67 clinical and environmental isolates of the genus *Stenotrophomonas*, including the environmental isolate *Stenotrophomonas maltophilia* UENF-4GII, which allowed us to identify gene groups associated exclusively with clinical and environmental isolates. The second chapter deals with the comparative genomics of the genus *Alcaligenes*, which allowed us to establish new taxonomic reclassifications and prospect isolates with biotechnological potential, especially for the promotion of plant growth and bioremediation. Furthermore, we identified significant differences in antimicrobial resistance gene composition between clinical and environmental isolates. In the third chapter, we report a large-scale genomic investigation of 1,921 genomes belonging to the *Burkholderiaceae* family. Through phylogenomic analyses, we reclassified 61 genomes at the genus level. We prospected the genera with the greatest potential for biofertilization and identified distinct profiles in the composition of virulence and resistance genes of genera with beneficial and pathogenic species. Overall, our results provided important information on the genomic diversity of *Stenotrophomonas*, *Alcaligenes*, and members of the *Burkholderiaceae* family, which serves as a basis for searching and identifying low pathogenicity risk candidates and potential pathogens.

Keywords: Bioinformatics, Genomics, Phylogeny, *Stenotrophomonas*, *Alcaligenes*, *Burkholderiaceae*

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ACC - 1-aminocyclopropane-1-carboxylate
IAA - Indole-3-acetic acid (ácido indol acétido)
ANI - Average nucleotide identity (identidade média de nucleotídeos)
CARD - Comprehensive antibiotic resistance database (banco de dados de resistência aos antibióticos)
CSM - Complexo *Stenotrophomonas maltophilia*
FBN - Fixação biológica de nitrogênio
GDH - Glucose Dehydrogenase
GI - Genomic Island (Ilhas genômicas)
HGT - Horizontal gene transfer (transferência horizontal de genes)
IAM - Via do indol-3-acetamida
IAN - Via do indol-3-acetonitrila
IPA - Via do indol-3-piruvato
ML - Maximum likelihood (máxima verossimilhança)
MLSA - Multi-locus sequence type (tipo de sequência multilocus)
NRPS -Non-ribosomal peptide synthetases (sintetases peptídicas não-ribossomais)
PGPR - Plant growth promoting rhizobacteria (bactéria promotora do crescimento de plantas)
PGPT - Plant growth promoting traits (características para promoção do crescimento de plantas)
PQQ - Pirroquinolina quinona (Pyrroloquinoline quinone)
RND - Resistência-nodulação-divisão
SMC - *Stenotrophomonas maltophilia complex*
TAM - Via da triptamina
VFDB - Virulence factor database (Base de dados de fatores de virulência)

SUMÁRIO

1. INTRODUÇÃO GERAL

Com a crescente demanda pela produção de alimentos frente ao aumento populacional mundial (United Nations, 2017), a preocupação com a utilização consciente de recursos naturais e com o uso intensivo de fertilizantes e defensivos químicos na agricultura (Powlson *et al.*, 2011), práticas agrícolas mais sustentáveis e ambientalmente corretas têm sido adotadas (Umesha, Singh, & Singh, 2018). Nos sistemas sustentáveis de produção agrícola, há uma busca constante pela redução no uso de insumos químicos e a otimização de processos biológicos que favoreçam o desenvolvimento das culturas (Gupta *et al.*, 2015).

Neste contexto, a utilização de biofertilizantes com microrganismos promotores do crescimento das plantas vem ganhando cada vez mais destaque devido ao baixo impacto ambiental em comparação com fertilizantes e defensivos químicos, sendo uma alternativa estratégica para a agricultura convencional (Suhag, 2016). Entretanto, a ampla adoção destes microrganismos ainda tem grandes desafios, dentre os quais está a identificação de isolados potenciais causadores de infecções oportunistas, particularmente quando filogeneticamente relacionados aos microrganismos promotores do crescimento vegetal (Berg *et al.*, 2013).

1.1. A rizosfera

A interface entre a raiz e solo, conhecida como rizosfera, é considerada o principal reservatório de microrganismos que interagem com plantas, sendo descrito como um dos ecossistemas com maior biodiversidade na terra (Umesha, Singh, & Singh, 2018). Comparada com outras regiões do solo, a rizosfera é uma região abundante em nutrientes, particularmente devido ao acúmulo de uma variedade de exsudatos radiculares ricos em aminoácidos e açúcares que podem ser assimilados por microrganismos (McNear-Jr, 2013). Por consequência, o número de microrganismos na rizosfera é geralmente de 10 a 100 vezes maior do que outras áreas do solo (Beneduzi, Ambrosini, & Passaglia, 2012). Dentre estes microrganismos, bactérias que colonizam a rizosfera, também conhecidas como rizobactérias, constantemente competem entre si para se estabelecerem neste ecossistema e podem ser classificadas em grupos benéficos, deletérios e neutros, com base em seus efeitos sobre o crescimento das plantas (McNear-Jr, 2013). Nas interações negativas, as rizobactérias fitopatogênicas produzem substâncias fitotóxicas (e.g. cianeto de hidrogênio), exibindo influência negativa no crescimento e na fisiologia das plantas. Além dessas bactérias deletérias, existem rizobactérias que promovem o crescimento das plantas. As interações neutras, por outro lado, não beneficiam nem prejudicam diretamente a planta, mas são relevantes para a fertilidade e diversidade microbiológica do solo (Bishnoi, 2015).

1.2. Rizobactérias promotoras do crescimento de plantas

As rizobactérias benéficas são referidas como rizobactérias promotoras de crescimento de plantas (*plant growth promoting rhizobacteria*; PGPR). As PGPR são

um grupo filogeneticamente heterogêneo de microrganismos, incluindo, por exemplo, espécies dos gêneros: *Alcaligenes*, *Azoarcus*, *Azospirillum*, *Burkholderia*, *Bacillus*, *Bradyrhizobium*, *Caballeronia*, *Cupriavidus*, *Gluconacetobacter*, *Herbaspirillum*, *Pandoraea*, *Paraburkholderia*, *Rhizobium e Stenotrophomonas* (Gouda *et al.*, 2017; Kakar *et al.*, 2018).

Estes microrganismos são capazes de promover, direta ou indiretamente, a promoção do crescimento de plantas, enquanto se beneficiam de exsudatos radiculares (Bais *et al.*, 2006; Gopalakrishnan *et al.*, 2015). A promoção direta do crescimento vegetal ocorre quando a bactéria facilita a aquisição de nutrientes essenciais ou modula os níveis de hormônios dentro da planta, o que varia de acordo com a cepa bacteriana e as espécies vegetais (Gouda *et al.*, 2017). Dentre estes mecanismos destacam-se: fixação biológica de nitrogênio atmosférico (Glick, 2012), solubilização de fosfato (Ahemad & Khan, 2012), produção de fitormônios (Duca *et al.*, 2014) e produção da enzima 1-aminociclopropano-1-carboxilato (ACC) desaminase, inibidora do fitormônio etileno (Glick, 2014).

Os mecanismos indiretos de promoção do crescimento vegetal compreendem a mitigação dos efeitos de estresses bióticos causados por fitopatógenos (Beneduzi, Ambrosini, & Passaglia, 2012) e abióticos, como o estresse salino, estresse hídrico e estresse causado por metais pesados (Habib, Kausar, & Saud, 2016; Mishra, Singh, & Arora, 2017). Dentre os mecanismos utilizados para mitigar estresses bióticos, várias PGPR produzem enzimas hidrolíticas, como quitinases, celulases e proteases como forma de obter vantagem competitiva contra fungos fitopatogênicos (Kumar, Dubey, & Maheshwari, 2012; Mubarik *et al.*, 2010); Diversas PGPR também produzem antibióticos (Haas & Défago, 2005), sideróforos (Radzki *et al.*, 2013) e compostos orgânicos voláteis (VOCs) supressores de bactérias fitopatogênicas (Tahir *et al.*, 2017). Devido à ação destes mecanismos, uma variedade de rizobactérias tem sido utilizada comercialmente como inoculantes para plantas cultivadas.

1.3. PGPR como biofertilizantes

Biofertilizantes são inoculantes microbianos geralmente definidos como uma preparação contendo microrganismos eficientes para fixação de nitrogênio e solubilização de fosfato (Mengual *et al.*, 2014). Ao contrário dos fertilizantes químicos, os biofertilizantes contém microrganismos viáveis que não são a fonte de nutrientes, mas auxiliam ativamente na nutrição vegetal na rizosfera (Umesha, Singh, & Singh, 2018). Na última década, os biofertilizantes vêm ganhando força devido à demanda por práticas sustentáveis, apresentando vantagens na manutenção da saúde do solo e na redução da poluição ambiental decorrente do uso de produtos químicos na agricultura (Umesha, Singh, & Singh, 2018).

Diversos microrganismos são comumente usados como biofertilizantes, incluindo PGPR e fungos. Por exemplo, espécies rizóbias (*Rhizobium*, *Mesorhizobium* e *Bradyrhizobium*) têm sido utilizadas mundialmente na fixação de nitrogênio em leguminosas (Laranjo, Alexandre, & Oliveira, 2014). Bactérias diazotróficas não-simbióticas como *Azotobacter*, *Azospirillum* e *Bacillus* também são

utilizadas para inocular sementes e extensas porções de terras aráveis ao redor do mundo com o objetivo de aumentar a disponibilidade de nitrogênio e aumentar a produtividade das plantas (Babalola, 2010).

Além disso, bactérias solubilizadoras de fosfato, como algumas espécies dos gêneros *Bacillus* e *Paraburkholderia*, têm sido aplicadas ao solo para melhorar especificamente a captação de fósforo para as plantas (Alori, Glick, & Babalola, 2017). Uma variedade de formulações comerciais de biofertilizantes encontra-se atualmente disponível e diversas estratégias já foram adotadas para preservar a viabilidade dos microrganismos nestas formulações (Bashan *et al.*, 2014). Entretanto, nas últimas décadas, a ocorrência de alguns destes microrganismos (ou cepas relacionadas) em ambientes hospitalares tem aumentado, despertando preocupação acerca da segurança das PGPRs em aplicações biotecnológicas. Como medida de segurança, o uso das PGPR tem se restringido aos grupos de baixo risco de patogenicidade à humanos, conforme critérios da Organização Mundial da Saúde (OMS) (Keswani *et al.*, 2019). Desta forma, algumas cepas promissoras amplamente testadas em laboratório permanecem indisponíveis comercialmente (Singh, Sarma, & Keswani, 2017).

1.4. Patógenos oportunistas emergentes

Bactérias oportunistas são aquelas que acometem indivíduos com predisposição a infecções, particularmente os gravemente debilitados, imunocomprometidos ou portadores de fibrose cística (Blanchard & Waters, 2022; Steinkamp *et al.*, 2005). Embora estas bactérias apresentem tipicamente baixa virulência, o impacto das infecções oportunistas na população aumentou dramaticamente nas últimas décadas (Cornejo-Juárez *et al.*, 2015; Nazik *et al.*, 2018).

As bactérias oportunistas têm sido responsáveis pela maioria das infecções associadas a casos de óbito em pacientes suscetíveis na Europa e na América do Norte (Teplitski, Leitão, & Sela, 2011). Em unidades de terapia intensiva (UTI) na Europa, 45% dos pacientes apresentaram infecções nosocomiais oportunistas (Welte, 2013). Na Alemanha, cerca de um milhão de casos de infecções nosocomiais são registrados por ano, das quais cerca de 40 mil são fatais, enquanto que no Brasil cerca de 10% dos pacientes hospitalizados em UTI são acometidos por infecções nosocomiais oportunistas (Nangino *et al.*, 2012). Devido à grande diversidade, pouco se sabe sobre a ecologia e fisiologia destes patógenos emergentes, embora alguns sejam encontrados em ambientes aquáticos, por viverem como comensais humanos ou por se originar de ecossistemas terrestres (e.g. rizosfera) (Mendes, Garbeva, & Raaijmakers, 2013).

Rizobactérias do gênero *Stenotrophomonas* têm sido associadas a várias manifestações clínicas como endocardite e infecções do trato respiratório em seres humanos (Sánchez, 2015). Embora não altamente virulento, *Stenotrophomonas maltophilia* tem emergido como um importante patógeno nosocomial associado a taxas de mortalidade entre 14 e 69% em pacientes com bacteremia e 30,7% em pacientes queimados acometidos pela infecção (Brooke, 2012; Tsai *et al.*, 2006).

Cepas do gênero *Alcaligenes* também têm sido cada vez mais associadas a casos de infecções nosocomiais (Aisenberg, Rolston, & Safdar, 2004). Uma pesquisa realizada em 2013 na unidade de saúde da Nigéria, *Alcaligenes faecalis* foi encontrado em amostras de pacientes acometidos por sepse (6%), bacteremia e meningite (1%). Além disso, casos de coinfecção e multirresistência a antibióticos também tem sido reportado para *A. faecalis* (Agarwal *et al.*, 2017; Hasan, Nizhu, & Rabbani, 2019; Huang, 2020).

Assim como *Alcaligenes* e *Stenotrophomonas*, espécies do gênero *Cupriavidus, Pandoraea* e *Burkholderia* também possuem isolados reportados em interações bivalentes com hospedeiros humanos e vegetais (Tabela 1-1). Em comum, estes gêneros possuem espécies de interesse biotecnológico com isolados tanto ambientais quanto clínicos, abrindo perspectivas importantes na investigação para a discriminação genética destes isolados e que contribuam no entendimento da ecologia e evolução das espécies.

Tabela 1-1. Ocorrência de isolados ambientais e clínicos de gêneros descritos como colonizadores de ambientes rizosféricos**.**

Espécie	Ambiental			Espécie	Clínico (humano)	
	Origem	Organismo	Referência		Origem	Referência
Alcaligenes faecallis Bk1	Rizosfera	Arroz	Kakar et al., 2017	*Alcaligenes faecallis AF1*	Urina	Momtaz et al., 2018
Burkholderia sp. AU4i	Rizosfera	Ervilha	Usha, Devi, et al., 2015	Burkholderia cenocepacia K56-2	Escarro	García-Romero et al., 2020
Cupriavidus agavae ASC-9842	Rizosfera	*Agave L.*	Arroyo-Herrera, Ivan, et al., 2020	*Cupriavidus pauculus* MF1	Pia de hospital	Butler, James, et al., 2022
Pandoraea sp. ISTKB	Rizosfera	-	Kumar, Madan, et al., 2016	*Pandoraea fibrosis* 6399	Escarro	Pitt, Miranda E., et al., 2020
Stenotrophomonas maltophilia NAA11	Rizosfera	Milho	Babalola, Olubukola Oluranti, et al., 2022	*Stenotrophomonas maltophilia* D457	Sangue	Pak et al., 2015

1.5. Genômica comparativa de procariotos

Os avanços das tecnologias de sequenciamento genômico a partir de 2005 resultaram no aumento de escala e redução de custos, tornando tais tecnologias mais acessíveis (Buermans & Den Dunnen, 2014). Desta forma, as tecnologias de sequenciamento de nova geração proporcionaram um aumento significativo no número genomas sequenciados e disponibilizados em bancos de dados públicos (Tonder et al., 2014). No entanto, embora um genoma traga informações relevantes, é sua comparação com outros genomas que tipicamente possibilita compreender os aspectos evolutivos e adaptativos dos organismos (Hardison, 2003).

As análises por genômica comparativa em procariotos tipicamente visam determinar um conjunto de genes homólogos ou xenólogos (i.e. transferidos, horizontalmente), visando entender a variação gênica e dinâmica evolutiva das

espécies (Vernikos *et al.*, 2015). Sawana e colaboradores (2014), por exemplo, realizaram análises genômicas comparativas em larga escala com base na detecção de marcadores moleculares em genes constitutivos (*housekeeping genes*) e propuseram a divisão do gênero *Burkholderia* em dois grupos distintos, compreendendo espécies clínicas e ambientais, resultando na identificação de um novo gênero composto de espécies ambientais (*Paraburkholderia*). Estudos como este são atualmente essenciais na avaliação de novos isolados para aplicações biotecnológicas (Bernabeu *et al.*, 2018).

As abordagens comparativas também oferecem detalhes adicionais que podem auxiliar na descoberta de genes de interesse biotecnológico, biomédico ou ambiental. Através destas abordagens, Noweel e colaboradores (2016) identificaram um grupo de genes específicos associados à colonização vegetal pelo fitopatógeno *Pseudomonas syringae*, demonstrando a adaptação exclusiva deste microrganismo em relação às espécies mais próximas. Além disto, Matteoli e colaboradores (2018), estudando genomas de *S. mascescens* clínicas e ambientais, relataram a alta prevalência de genes de síntese de prodigiosina (*pig*), um metabólito com potencial antifúngico, entre cepas ambientais. Estes achados indicam a importância destes genes na competição em microbiotas complexas (e.g. rizosfera), além de dar suporte ao seu uso como marcador molecular, o que é particularmente interessante se considerarmos que a prodigiosina é um pigmento de coloração marcante.

Estudos comparativos baseados em análises de pangenoma têm se tornado uma abordagem eficaz na identificação de genes correlatos com diferentes estilos de vida em diferentes níveis taxonômicos (Vernikos *et al.*, 2015). As análises de pangenoma consistem na determinação de um núcleo de genes presentes em todas as estirpes (*core* genoma), um núcleo de genes distribuídos entre algumas espécies (genoma acessório) e de um terceiro conjunto de genes que exclusivos de uma das estirpes (genoma único). O *core* genoma é composto por genes constitutivos, provavelmente herdados de um ancestral comum e pouco propensos a transferência horizontal de genes (HGT- *Horizontal gene transfer*). Ao genoma acessório são atribuídos os genes relacionados à adaptação e colonização de determinados ambientes, que são usualmente responsáveis pelo estilo de vida dos organismos. No genoma único estão presentes os genes específicos de cada estirpe, frequentemente adquiridos por HGT (Rouli *et al.*, 2015).

Além do pangenoma, diferentes abordagens têm sido utilizadas em análises comparativas de procariotos. Dentre estas, a classificação filogenômica se destaca como uma etapa essencial para a determinação de relações taxonômicas (Jain *et al.*, 2018). Nesta abordagem são utilizados algoritmos para inferir a média de identidade nucleotídica ou ANI (*Average Nucleotide Identity*) entre genomas para determinar se estes pertencem ou não a uma mesma espécie (Han, Qiang, & Zhang, 2016); ou métodos estatísticos que estimam a semelhança e diversidade dos genomas baseados em coeficiente de similaridade de Jaccard (Katz, Griswold, & Carleton, 2017; Ondov *et al.*, 2016). Este segundo método apresenta vantagens em comparações genômicas em maior escala devido ao seu tempo de execução consideravelmente menor.

Em 2019, Faoro e colaboradores através de análises de ANI e pangenoma evidenciaram diferentes grupos gênicos no genoma acessório de cepas clínicas e ambientais de *Herbaspirillum seropedicae*, espécie utilizada como inoculante microbiano. Os autores também relataram a ausência de genes cruciais relacionados à fixação de nitrogênio (genes *nif*) e ao sistema de secreção do tipo 3 nos isolados clínicos, bem como a presença de genes exclusivos relacionados com a colonização em hospedeiros humanos. Os resultados deste trabalho forneceram informações importantes a respeito da capacidade adaptativa da espécie na colonização de diferentes hospedeiros através da perda e aquisição de genes específicos (Faoro *et al.*, 2019).

A presente tese buscou investigar e comparar genomas pertencentes aos gêneros *Stenotrophomonas*, *Alcaligenes* e gêneros da família *Burkholderiaceae* como *Burkholderia*, *Paraburkholderia*, *Cupriavidus*, *Trinickia*, *Caballeronia*, *Ralstonia*, *Pandoraea*, que possuem ocorrências de isolados clínicos e de interesse biotecnológico, focando particularmente em conjuntos de genes que contribuam para a ocupação de tais nichos, bem como realizar uma análise crítica de grandes tendências compartilhadas ou divergentes entre os gêneros. Os capítulos seguintes tratam de compilações dos resultados obtidos, incluindo de um artigo já publicado.

1.6. REFERÊNCIAS BIBLIOGRÁFICAS

Agarwal, A., et al. (2017). First reported case of Alcaligenes faecalis isolated from bronchoalveolar lavage in a patient with dengue hemorrhagic fever. *The Journal of Association of Chest Physicians, 5*(1), 51.

Ahemad, M., Khan, M. S. (2012). *Assessment of Plant Growth Promoting Activities of Rhizobacterium Pseudomonas putida under Insecticide-Stress* (Vol. 62).

Aisenberg, G.,Rolston, K. V., Safdar, A. (2004). Bacteremia caused by Achromobacter and Alcaligenes species in 46 patients with cancer (1989–2003). *Cancer, 101*(9), 2134-2140.

Alori, E. T.,Glick, B. R., Babalola, O. O. (2017). Microbial phosphorus solubilization and its potential for use in sustainable agriculture. *Frontiers in microbiology, 8*, 971.

Babalola, O. O. (2010). Beneficial bacteria of agricultural importance. *Biotechnol Lett, 32*(11), 1559-1570.

Bais, H. P., et al. (2006). The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual Review of Plant Biology, 57*(1), 233-266.

Bashan, Y., et al. (2014). Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). *Plant and Soil, 378*(1-2), 1-33.

Beneduzi, A.,Ambrosini, A., Passaglia, L. M. P. (2012). Plant growth-promoting rhizobacteria (PGPR): their potential as antagonists and biocontrol agents. *Genetics and Molecular Biology, 35*, 1044-1051. Retrieved from

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572012000600020&nrm=iso

Berg, G., et al. (2013). The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Molecular Microbial Ecology of the Rhizosphere, 1st ed.; de Bruijn Frans, J., Ed*, 1209-1216.

Bernabeu, P. R., et al. (2018). Assessment of bacterial inoculant formulated with Paraburkholderia tropica to enhance wheat productivity. *World J Microbiol Biotechnol, 34*(6), 81. doi:10.1007/s11274-018-2461-4

Bishnoi, U. (2015). Chapter Four - PGPR Interaction: An Ecofriendly Approach Promoting the Sustainable Agriculture System. In H. Bais & J. Sherrier (Eds.), *Advances in Botanical Research* (Vol. 75, pp. 81-113): Academic Press.

Blanchard, A. C., Waters, V. J. (2022). Opportunistic pathogens in cystic fibrosis: epidemiology and pathogenesis of lung infection. *Journal of the Pediatric Infectious Diseases Society, 11*(Supplement_2), S3-S12.

Brooke, J. S. (2012). Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen. *Clin Microbiol Rev, 25*(1), 2-41. doi:10.1128/cmr.00019-11

Buermans, H., Den Dunnen, J. (2014). Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1842*(10), 1932-1941.

Cornejo-Juárez, P., et al. (2015). The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *International Journal of Infectious Diseases, 31*, 31-34.

Duca, D., et al. (2014). Indole-3-acetic acid in plant–microbe interactions. *Antonie Van Leeuwenhoek, 106*(1), 85-125.

Faoro, H., et al. (2019). Genome comparison between clinical and environmental strains of Herbaspirillum seropedicae reveals a potential new emerging bacterium adapted to human hosts. *BMC Genomics, 20*(1), 630.

Glick, B. R. (2012). Plant Growth-Promoting Bacteria: Mechanisms and Applications. *Scientifica, 2012*, 15.

Glick, B. R. (2014). Bacteria with ACC deaminase can promote plant growth and help to feed the world. *Microbiological Research, 169*(1), 30-39.

Gopalakrishnan, S., et al. (2015). Plant growth promoting rhizobia: challenges and opportunities. *3 Biotech, 5*(4), 355-377.

Gouda, S., et al. (2017). Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture. *Microbiological Research*.

Gupta, G., et al. (2015). Plant Growth Promoting Rhizobacteria (PGPR): Current and Future Prospects for Development of Sustainable Agriculture. *Journal of Microbial & Biochemical Technology, 7*, 96-102.

Haas, D., Défago, G. (2005). Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nature Reviews Microbiology, 3*, 307.

Habib, S. H.,Kausar, H., Saud, H. M. (2016). Plant Growth-Promoting Rhizobacteria Enhance Salinity Stress Tolerance in Okra through ROS-Scavenging Enzymes. *Biomed Res Int, 2016*.

Han, N.,Qiang, Y., Zhang, W. (2016). ANItools web: a web tool for fast genome comparison within multiple bacterial strains. *Database (Oxford), 2016*. doi:10.1093/database/baw084

Hardison, R. C. (2003). Comparative genomics. *PLoS Biol, 1*(2), E58. doi:10.1371/journal.pbio.0000058

Hasan, M. J.,Nizhu, L. N., Rabbani, R. (2019). Bloodstream infection with pandrug-resistant Alcaligenes faecalis treated with double-dose of tigecycline. *IDCases, 18*, e00600.

Huang, C. (2020). Extensively drug-resistant Alcaligenes faecalis infection. *BMC infectious diseases, 20*(1), 1-11.

Jain, C., et al. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications, 9*(1), 5114. doi:10.1038/s41467-018-07641-9

Kakar, K. U., et al. (2018). Rhizosphere-associated Alcaligenes and Bacillus strains that induce resistance against blast and sheath blight diseases, enhance plant growth and improve mineral content in rice. *Journal of applied microbiology, 124*(3), 779-796. doi:doi:10.1111/jam.13678

Katz, L.,Griswold, T., Carleton, H. (2017). *Generating WGS Trees with Mashtree.* Paper presented at the Poster presented at: American Society for Microbiology Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines.

Keswani, C., et al. (2019). Re-addressing the biosafety issues of plant growth promoting rhizobacteria. *Science of The Total Environment, 690*, 841-852.

Kumar, P.,Dubey, R. C., Maheshwari, D. K. (2012). Bacillus strains isolated from rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. *Microbiological Research, 167*(8), 493-499.

Laranjo, M.,Alexandre, A., Oliveira, S. (2014). Legume growth-promoting rhizobia: An overview on the Mesorhizobium genus. *Microbiological Research, 169*(1), 2-17. doi:https://doi.org/10.1016/j.micres.2013.09.012

Matteoli, F. P., et al. (2018). Genome sequencing and assessment of plant growth-promoting properties of a Serratia marcescens strain isolated from vermicompost. *BMC Genomics, 19*(1), 750. doi:10.1186/s12864-018-5130-y

McNear-Jr, D. H. (2013). The Rhizosphere - Roots, Soil and Everything In Between. *Nature Education Knowledge, 4*(3).

Mendes, R.,Garbeva, P., Raaijmakers, J. M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS microbiology reviews, 37*(5), 634-663. doi:10.1111/1574-6976.12028

Mengual, C., et al. (2014). Microbial inoculants and organic amendment improves plant establishment and soil rehabilitation under semiarid conditions. *Journal of environmental management, 134*, 1-7.

Mishra, J.,Singh, R., Arora, N. K. (2017). Alleviation of Heavy Metal Stress in Plants and Remediation of Soil by Rhizosphere Microorganisms. *Frontiers in microbiology, 8*, 1706. doi:10.3389/fmicb.2017.01706

Mubarik, N., et al. (2010). *Chitinolytic Bacteria Isolated from Chili Rhizosphere: Chitinase Characterization and Its Application as Biocontrol for Whitefly (Bemisia tabaci Genn.)* (Vol. 5).

Nangino, G. O., et al. (2012). Impacto financeiro das infecções nosocomiais em unidades de terapia intensiva em hospital filantrópico de Minas Gerais. *Revista Brasileira de Terapia Intensiva, 24*, 357-361. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-507X2012000400011&nrm=iso

Nazik, S., et al. (2018). Nosocomial Burkholderia cepacia infection in a tertiary hospital; Five-year surveillance: A retrospective cross-sectional study. *Urology, 6*, 10-13.

Nowell, R. W., et al. (2016). Comparative genomics reveals genes significantly associated with woody hosts in the plant pathogen Pseudomonas syringae. *Molecular Plant Pathology, 17*(9), 1409-1424. doi:doi:10.1111/mpp.12423

Ondov, B. D., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology, 17*(1), 132. doi:10.1186/s13059-016-0997-x

Powlson, D. S., et al. (2011). Soil management in relation to sustainable agriculture and ecosystem services. *Food Policy, 36*, S72-S87. doi:https://doi.org/10.1016/j.foodpol.2010.11.025

Radzki, W., et al. (2013). Bacterial siderophores efficiently provide iron to iron-starved tomato plants in hydroponics culture. *Antonie Van Leeuwenhoek, 104*(3), 321-330.

Rouli, L., et al. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections, 7*, 72-85.

Sánchez, M. B. (2015). Antibiotic resistance in the opportunistic pathogen Stenotrophomonas maltophilia. *Frontiers in microbiology, 6*, 658.

Sawana, A.,Adeolu, M., Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov. harboring environmental species. *Frontiers in genetics, 5*, 429.

Singh, H. B.,Sarma, B. K., Keswani, C. (2017). *Advances in PGPR research*: CABI.

Steinkamp, G., et al. (2005). Prospective evaluation of emerging bacteria in cystic fibrosis. *Journal of Cystic Fibrosis, 4*(1), 41-48. doi:https://doi.org/10.1016/j.jcf.2004.10.002

Suhag, M. (2016). Potential of Biofertilizers to Replace Chemical Fertilizers. *International Advanced Research Journal in Science, Engineering and Technology, 3*(5).

Tahir, H. A. S., et al. (2017). Plant Growth Promotion by Volatile Organic Compounds Produced by Bacillus subtilis SYST2. *Frontiers in microbiology, 8*, 171.

Teplitski, M.,Leitão, J. H., Sela, S. (2011). Microbial Ecology and Global Health. *Int J Microbiol, 2011*. doi:10.1155/2011/564701

Tsai, W. P., et al. (2006). Stenotrophomonas maltophilia bacteremia in burn patients. *Burns, 32*(2), 155-158. doi:10.1016/j.burns.2005.08.016

Umesha, S.,Singh, K. P., Singh, P. R. (2018). Chapter 6 - Microbial Biotechnology and Sustainable Agriculture. In R. L. Singh & S. Mondal (Eds.), *Biotechnology for Sustainable Agriculture* (pp. 185-205): Woodhead Publishing.

United Nations. (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. *ESA/P/WP/248*.

Vernikos, G., et al. (2015). Ten years of pan-genome analyses. *Current Opinion in Microbiology, 23*, 148-154. doi:https://doi.org/10.1016/j.mib.2014.11.016

Welte, T. (2013). Nosocomial Infections–a Present and Future Challenge. *Deutsches Ärzteblatt International, 110*(38), 625.

2. CAPÍTULO 1: Genome sequencing of the vermicompost strain *Stenotrophomonas maltophilia* UENF-4GII and population structure analysis of the *S. maltophilia* Sm3 genogroup

Francisnei Pedrosa-Silva[a], Filipe P. Matteoli[a], Hemanoel Passarelli-Araujo[a,b], Fabio L. Olivares[c, d], Thiago M. Venancio[a, *]

[a] Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Brazil; [b] Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; [c] Núcleo de Desenvolvimento de Insumos Biológicos para a Agricultura (NUDIBA), UENF, Brazil; [d] Laboratório de Biologia Celular e Tecidual, Centro de Biociências e Biotecnologia, UENF, Brazil.

* Corresponding author:

Thiago M. Venancio; Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF); Av. Alberto Lamego 2000, P5 / sala 217; Campos dos Goytacazes, Rio de Janeiro, Brazil. E-mail: thiago.venancio@gmail.com.

Artigo publicado no periódico *Microbiological Research*
(DOI: doi.org/10.1016/j.micres.2021.126923).

2.1. ABSTRACT

The *Stenotrophomonas maltophilia* complex (Smc) is a cosmopolitan bacterial group that has been proposed an emergent multidrug-resistant pathogen. Taxonomic studies support the genomic heterogeneity of Smc, which comprises genogroups exhibiting a range of phenotypically distinct strains from different sources. Here, we report the genome sequencing and in-depth analysis of *S. maltophilia* UENF-4GII, isolated from vermicompost. This genome harbors a unique region encoding a penicillin-binding protein (*pbp*X) that was carried by a transposon, as well as horizontally-transferred genomic islands involved in anti-phage defense via DNA modification, and pili glycosylation. We also analyzed all available Smc genomes to investigate genes associated with resistance and virulence, niche occupation, and population structure. *S. maltophilia* UENF-4GII belongs to genogroup 3 (Sm3), which comprises three phylogenetic clusters (PC). Pan-GWAS analysis uncovered 471 environment-associated and 791 PC-associated genes, including antimicrobial resistance (e.g. *bla*L1 and *bla*R1) and virulence determinants (e.g. *tre*S and *kat*G) that provide insights on the resistance and virulence potential of Sm3 strains. Together, the results presented here provide the grounds for more detailed clinical and ecological investigations of *S. maltophilia*.

Keywords: *Stenotrophomonas maltophilia*; genomic island; resistome; virulence factor; pan-GWAS.

2.2. INTRODUCTION

Vermicomposting is a non-thermophilic biodegradation technique used to manage organic waste (Lim, Lee, & Wu, 2016). The process involves synergistic interactions between earthworms and microorganisms to biodegrade different organic waste into a humus-like material known as vermicompost, a nutrient-rich organic amendment used to enhance soil microbial diversity and plant development (Domínguez *et al.*, 2019; Pathma & Sakthivel, 2012). Vermicomposts harbor soil bacteria from various genera, such as *Bacillus*, *Pseudomonas*, *Serratia,* and *Stenotrophomonas,* which may engage in beneficial interactions with plants (Matteoli *et al.*, 2018; Pathma & Sakthivel, 2013).

Stenotrophomonas maltophilia is a Gram-negative bacillus found in a wide range of natural habitats, including water sources, soils, rhizospheres, animal microbiotas, including humans (Brooke, 2012; Ryan *et al.*, 2009). *S. maltophilia* has also been used as part of bioremediation and biocontrol strategies (Mukherjee & Roy, 2016; Ryan *et al.*, 2009). However, *S. maltophilia* has been reported as a global multidrug resistant opportunistic pathogen associated with significant mortality rates of up to 37.5%, mainly because of bacteremia and respiratory tract infections in severely debilitated, immunosuppressed or chronic lung disease patients (Falagas *et al.*, 2009; Zemanick *et al.*, 2017). *S. maltophilia* is intrinsically resistant to multiple classes of antibiotics, such as aminoglycosides, carbapenems, and macrolides, posing a therapeutic challenge and delay the administration of proper antibiotics (Brooke, 2012).

S. maltophilia present high intraspecific variability (Gherardi *et al.*, 2015) and, along with the closely related species *S. pavanii*, comprise the *S. maltophilia* complex (Smc) (Hauben *et al.*, 1999; Rhee *et al.*, 2013; Svensson-Stadler, Mihaylova, & Moore, 2012). Phylogenic studies based on multilocus sequencing typing (MLST) and whole-genome sequencing revealed the organization of Smc members in several genogroups (Ochoa-Sánchez & Vinuesa, 2017; Patil *et al.*, 2018; Steinmann *et al.*, 2018). A study using whole-genome multilocus sequence typing (wgMLST) and average nucleotide identity (ANI) analyses showed the genetic organization of Smc in 23 monophyletic genogroups with different virulence and resistance characteristics (Gröschel *et al.*, 2020). Among those, genogroup 3 (Sm3) exhibits a myriad of phenotypically distinct strains from different sources. These strains remain poorly explored, hampering investigations on the genetic determinants underlying the physiology and niche occupation of *S. maltophilia* isolates.

Here we present the genome sequencing of *S. maltophilia* UENF-4GII, the first *S. maltophilia* isolated from vermicompost. *S. maltophilia* UENF-4GII belongs to Sm3 and harbors a set of interesting horizontally-transferred regions, including genomic islands (GIs) involved in phage resistance via DNA modification, and pili glycosylation. ANI analysis and SNP-based phylogenetic reconstructions allowed us to reclassify six publicly available *Stenotrophomonas* spp. genomes as *S. maltophilia* from Sm3. We also used the *S. maltophilia* UENF-4GII genome with those of other 66 *S. maltophilia* isolates to thoroughly characterize of the population structure, virulence, and

resistance profiles of Sm3. Finally, a pangenome-wide association study (pan-GWAS) of Sm3 allowed us to identify genes associated with niche occupation and phylogenetic clusters.

2.3. METHODS

2.3.1. Bacterial isolation and identification

The bacterium was isolated from mature manure vermicompost produced from cattle manure at Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil. In summary, serial dilutions (10^{-1} to 10^{-7}) were performed on a solution prepared by adding 10 g of vermicompost in 90 mL of saline (8.5 $g \cdot L^{-1}$ NaCl), followed by shaking for 60 min. Then, 100 µL of the final dilution (10^{-7}) were taken and spread on plates containing solid Nutrient Broth (NB) with 8 $g \cdot L^{-1}$ of NB and 15 $g \cdot L^{-1}$ of agar in 1 L of distilled water. After incubation at 30 °C for 7 days, different colony types could be identified and, for purification, individual colonies were transferred to Petri plates with Dygs solid media acquired from Vetec (São Paulo, Brazil). After isolation and purification on Dygs solid medium, a yellowish, circular, convex elevation, punctiform and smooth surface bacterial colony was selected. Light microscopy revealed a Gram-negative strain and the presence of rod-shaped motile was confirmed under phase contrast microscopy. This isolate, named UENF-4GII, was stored in a 16 mL glass flask containing 5 mL of Nutrient Broth solid medium covered with mineral oil and later grown in liquid Dygs medium under rotatory shaker at 150 rpm and 30 °C for 36 h.

2.3.2. Genome sequencing and annotation

Genomic DNA was extracted using QIAamp® DNA Mini Kit (Qiagen) and quantified with *an Agilent Bioanalyzer 2100 instrument (Agilent, California, USA).* Paired-end libraries were previously prepared with the TruSeq Nano DNA LT Library Prep (Illumina) and sequenced on an Illumina HiSeq 2500 sequencing system at the Life Sciences Core Facility (LaCTAD; UNICAMP, Campinas, Brazil). Sequencing reads (2 x 100 bp) were assembled *de novo* with SPAdes v.10.3.1 (Bankevich *et al.*, 2012) and scaffolded with Gfinisher v.1.4 (Guizelini *et al.*, 2016), using an alternative assembly generated with Velvet 1.2.10 (Zerbino & Birney, 2008) and the complete genome of *S. maltophilia* JV3 (GCF_000223885.1) as reference. The assembly statistics were assessed with QUAST v.3.0 (Gurevich *et al.*, 2013). Genome completeness was assessed with BUSCO v.4.0 (Simão *et al.*, 2015), using the *Gammaproteobacteria* dataset as reference. PlasmidSpades (Antipov *et al.*, 2016) was used to predict plasmids. The assembled genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova *et al.*, 2016). The UENF-4GII genome was deposited into DDBJ/EMBL/GenBank under the accession number JABUNQ000000000. Genes involved in antimicrobial metabolite biosynthesis were predicted using antiSMASH (Medema *et al.*, 2011). Insertion sequences (ISs) and GIs were predicted with ISEscan v.1.5.4 (Xie & Tang, 2017) and Islandviewer4 (Bertelli *et al.*, 2017), respectively. Bacteriophage signatures were analyzed with PHASTER (Arndt *et al.*, 2016).

2.3.3. Genome similarity assessment

We downloaded 641 genomes of *Stenotrophomonas* available in the RefSeq database in January, 2021 (O'Leary *et al.*, 2016; Pruitt, Tatusova, & Maglott, 2007). Genome completeness was assessed with BUSCO 4.0 (Simão *et al.*, 2015), using quality score ≥ 90% and the *Gammaproteobacteria* dataset as reference. We excluded assemblies with more than 500 contigs. A preliminary genome distance estimation analysis of isolate UENF-4GII and RefSeq genomes was performed using Mash (Ondov *et al.*, 2016) and a distance tree was generated using MASHtree v.0.50 (Katz, Griswold, & Carleton, 2017). All-against-all ANI based on MUMmer alignment (ANIm) was performed with pyani v.0.27 (Pritchard *et al.*, 2016). To assess the concordance between ANI and Mash estimates, we performed a linear correlation analysis using Pearson's correlation coefficient. For this analysis, the Mash distances were converted into Mash scores (1-Mash distance) to allow a direct comparison with ANI values.

2.3.4. Pangenome analysis

The Sm3 genogroup pangenome was performed with Roary v.3.6, using 95% identity threshold to determine gene clusters (A. J. Page *et al.*, 2015). Core genes were aligned with MAFFT v.7.394 (Katoh & Standley, 2013). SNPs were extracted from the core-genome alignment using SNP-sites v.2.3.3 (Andrew J Page *et al.*, 2016) and maximum likelihood phylogenetic reconstructions were performed with IQ-tree (Nguyen *et al.*, 2014), with ascertainment bias correction under the model GTR+ASC. The bootstrap support was evaluated using the ultrafast bootstrap method with 1000 replicates (Minh, Nguyen, & von Haeseler, 2013). The resulting phylogenetic tree was visualized with iTOL v4 (Letunic & Bork, 2019).

As a complementary approach, we performed a pangenome-wide association study (pan-GWAS) on the on the Sm3 dataset using Scoary. The pan-GWAS was computed using the Roary output to find genes associated with isolation source (clinical or non-clinical) and to establish which genes were typical of each phylogenetic cluster (PC), while correcting for population structure using the core-genome phylogenetic tree (command –n tree). False-discovery rate was estimated by Benjamini–Hochberg adjusted p-value provided in Scoary. We only reported the results with specificity > 70% and Benjamini–Hochberg corrected p-value < 0.05. The binary heatmaps of trait-associated genes were rendered using R package tidyverse (Wickham *et al.*, 2019).

2.3.5. Virulome and resistome analysis

Virulence and antimicrobial resistance genes were predicted using Usearch v.11.0.667 screened against the Virulence Factors of Pathogenic Bacteria Database (VFDB) and the Comprehensive Antibiotic Resistance Database (CARD). Minimum identity and coverage thresholds of 60% and 50% were used in these searches, respectively. The predicted genes were assigned into two groups: core, containing genes present in more than 95 % of the Sm3 genogroup; and accessory, harboring genes present in fewer than 95 % of the Sm3 genogroup. The presence/absence

profiles of virulence and resistance-associated genes were rendered using R package tidyverse (Wickham *et al.*, 2019).

2.4. RESULTS AND DISCUSSION

2.4.1. Genome analysis of *S. maltophilia* UENF-4GII

During the characterization of culturable bacteria from mature cattle vermicompost, we identified a bacterium that was preliminarily characterized as *Stenotrophomonas* sp. using 16S rRNA sequencing. We submitted the genome of this isolate to whole-genome sequencing (see methods for details). The 30,093,445 paired-end reads were assembled into 3 scaffolds, with an N_{50} of 1.3 Mbp, encompassing a total of 4.4 Mbp with 66.55% GC content. The genome coverage was 1,342x with the breadth of coverage achieved in 98.44% to the final assembly. No plasmids were detected. BUSCO assessment recovered 437 (96.68%) single copy genes of *Gammaprotebacteria* dataset, supporting the completeness and high quality of the assembled genome. The genome harbors 3,932 protein-coding genes, 70 tRNA genes, and 2 rRNA operons.

The *S. maltophilia* UENF-4GII genome carries 18 complete IS elements: nine of the IS481 family, seven of the IS3 family and a single copy of the IS5 and IS21 families (Supplementary figure S1). In addition, we also found an incomplete prophage region that is closely related with a temperate bacteriophage from *Burkholderia pseudomallei* (Burkho_phi1026b) (Yuzenkova *et al.*, 2003). AntiSMASH prediction revealed that UENF-4GII possesses four gene clusters involved in the production of two bacteriocin-like compounds, an aryl polyene related to Xanthomonadin (Goel *et al.*, 2002), and a non-ribosomal peptide synthetase (NRPS) gene cluster involved in the biosynthesis of siderophore (Supplementary figure S1). Siderophores are the most important iron uptake systems in *S. maltophilia* (Kalidasan *et al.*, 2018). In NRPS gene cluster, we identified six genes (HRE58_11480 to 11500, and HRE58_11505) encoding an specific siderophore of *S. maltophilia* (Nas & Cianciotto, 2017). These enzymes are similar to those involved in enterobactin biosynthesis in enteric bacteria (e.g. *E. coli*) (Kalidasan *et al.*, 2018).

2.4.2. Genome similarity of UENF-4GII

In order to investigate the genomic relatedness of UENF-4GII strain, we downloaded all available *Stenotrophomonas* genomes in Refseq database (n=641, January 2021), out of which 627 were retained after BUSCO quality filtering (see method for details). Genome identity was computed using Mash and ANI. The mash-distance phylogeny highlights the diversity of the *Stenotrophomonas* genus, corroborating the *S. maltophilia* complex organization in different genogroups (Figure 2-1A), as previously described by Gröschel et al. (Gröschel *et al.*, 2020).

We identified 17 genogroups according to the distance estimation of well-characterized representative genomes, including the affiliation of UENF-4GII to Sm3, along with other 67 clinical and non-clinical strains (Supplementary table S1). In addition, this analysis allowed us to reclassify six *Stenotrophomonas* spp. as *S.*

maltophilia Sm3 isolates, including the environmental strains *Stenotrophomonas* sp. DDT-1 (GCF_001580555.1) and *Stenotrophomonas* sp. Pemsol (GCF_003586545.1), reported as specific biodegraders of organochlorine pesticide (DDT) and polycyclic aromatic hydrocarbons (PAH), respectively. Mash results are supported by ANI analysis, which showed genomic identity above 95% within the same genogroup (Figure 2-1B). Moreover, the strong positive linear correlation (r = 0.98, p-value < 2.2e-16) between ANI and Mash results (Figure 2-1C) confirm a similar level of resolution for species delineation within the *Stenotrophomonas* genus.

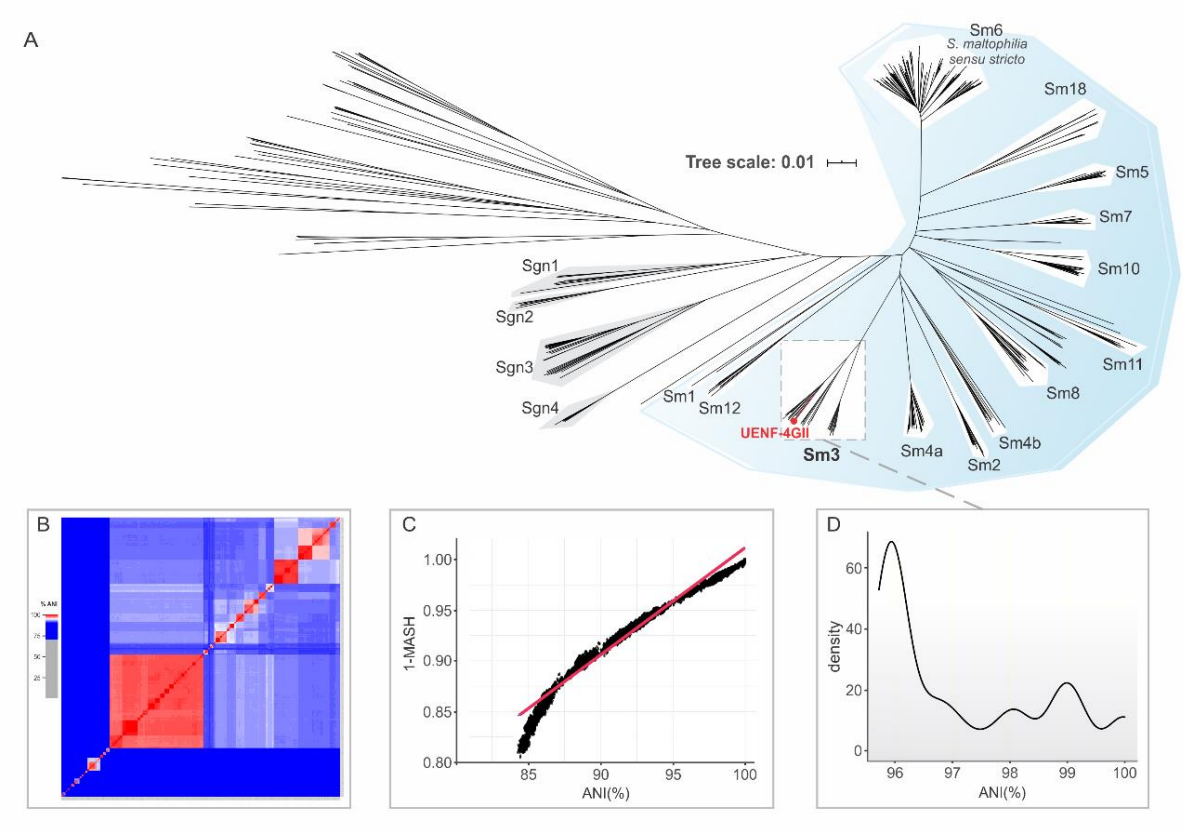


Figure 2-1. Genomic diversity of Stenotrophomonas genus. (A) Mash-distance-based phylogeny of *Stenotrophomonas*, built using 627 publicly genomes and that of UENF-4GII (red color). The *S. maltophilia sensu lato* clade is shaded in light blue. (B) Pairwise average nucleotide identity (ANI) calculated with 627 *Stenotrophomonas* genomes. Colors depict the degree of genome identity. (C) Correlation between ANI and Mash methods. (D) Density plot of pairwise ANI within the Sm3 genogroup.

2.4.3. Pangenome analysis

The pangenome of the 67 Sm3 strains comprises 13,380 genes, with 2,754 core genes (i.e. present in at least 95% of the strains), and 10,626 accessory genes. The accessory genome is composed by 1,958 high-frequency genes (present from 15% to 95% of the strains), and 8,668 low-frequency genes (present in up to 15% of the strains). The heap law estimate supports an open pangenome (alpha= 0.65) (Figure 2-2A), which typically reflects a high genetic diversity through the acquisition of exogenous DNA (Tettelin & Medini, 2020). The genomic fluidity (φ) Sm3 was

estimated in 0.18 (± 0.05), indicating that an average of 18% of unique gene families between in a given pair of genomes.

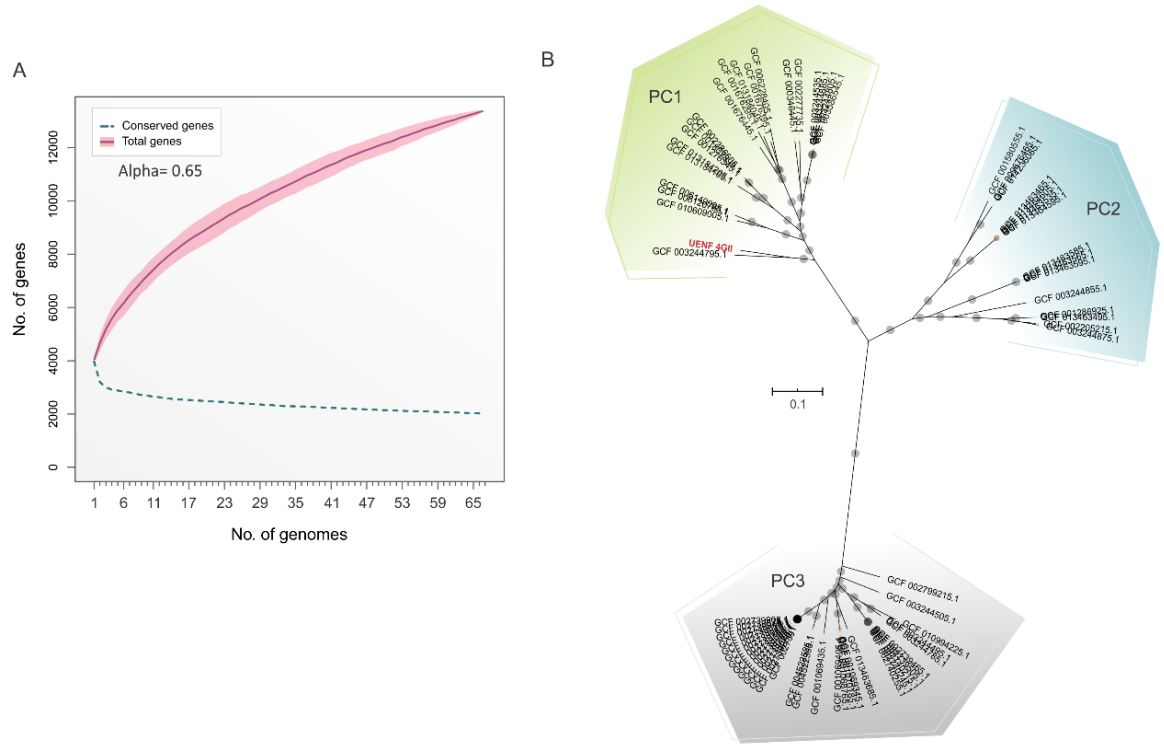


Figure 2-2. Pangenome and phylogeny of Sm3 genogroup. (A) Number of gene families in the Sm3 pangenome. The cumulative curve (in dark-red) and alpha value of the Heap law less than one (0.65) supports an open pangenome. (B) cgMLST of Sm3 genogroup showing three phylogenetic groups. SNPs extracted from the core genome were used to build a maximum likelihood phylogenetic tree using IQ-tree (see methods for details). Bootstrap values below and above 70% are represented by orange and gray points, respectively.

We also performed a maximum-likelihood phylogenetic reconstruction using 184,365 SNPs extracted from core genes (i.e. core-genome multilocus sequencing typing, cgMLST). This analysis revealed that Sm3 comprises three distinct and highly supported phylogenetic clusters (PC) (Figure 2-2B). PC1 (n= 21) and PC2 (n= 15) encompass clinical and non-clinical strains, while PC3 (n= 31) has only clinical strains. Interestingly, 10 of the 21 PC1 strains are non-clinical, including *S. maltophilia* UENF-4GII.

We identified 75 unique genes in *S. maltophilia* UENF-4GII (Supplementary table S2), including *pbp*X (HRE58_12720), which encodes a low molecular weight penicillin-binding protein (LMW-PBP). This gene is closely located to other unique genes: a gene with unknown function (HRE58_12725) and two transposase genes (HRE58_12730, HRE58_12735), flanked by a perfect 26-bp inverted repeat (IR) sequence, supporting their acquisition by horizontal gene transfer (Supplementary figure 2). LMW-PBPs are important enzymes involved in cell-wall recycling and considered the major molecular targets for β-lactam antibiotics (Ealand, Machowski, & Kana, 2018). Alterations in the structure of LMW-PBPs are associated with reduced

susceptibility to penicillin and other β-lactams and can induce the expression of β-lactamases (Ealand, Machowski, & Kana, 2018; Huang *et al.*, 2017; Sarkar, Chowdhury, & Ghosh, 2010). The presence of an LMW-PBP composite transposon likely confers a competitive advantage for *S. maltophilia* UENF-4GII to thrive in microbial communities containing penicillin.

2.4.4. UENF-4GII genomic islands

To further understand genome plasticity, we identified and analyzed 17 putative GIs (GI1-17) in *S. maltophilia* UENF-4GII, which probably represent recent horizontal gene transfers (Supplementary table S3). In addition, we compared these GIs with other Sm3 genomes (Figure 2-3). Among these regions, GI3 harbors an operon (HRE58_02625-02660) containing genes encoding GalE, HldD, and glycosyltransferases involved in lipopolysaccharide biosynthesis. This operon is present in 68% of the PC3 genomes. GI16 contains a class C beta-lactamase and GI10 have blue-light sensor genes that are absent in PC3 genomes.

Interestingly, 60% of the *S. maltophilia* UENF-4GII unique genes are located within 7 GIs related with DNA metabolism (GI6 and GI17), virulence (GI7, GI14 and GI15) and unknown functions (GI2 and GI9). GI6 has 9 unique genes, including a restriction-modification (R-M) system with two type I restriction endonucleases (HRE58_05925 and HRE58_05935) and a DNA methyltransferase (HRE58_05920). R-M systems defend the bacterial genomes against bacteriophages and other types of exogenous genetic elements (Dimitriu, Szczelkun, & Westra, 2020) and, hence, their acquisition might be adaptive by boosting resistance against phages (Tock & Dryden, 2005; Zheng *et al.*, 2010). The GI17 has 41,433 bp and comprises 14 unique genes, including a DNA polymerase-like protein (HRE58_18840) and an operon containing an 8-oxoguanine DNA glycosylase (HRE58_18870), a nitroreductase (HRE58_18875), a 7-cyano-7-deazaguanine synthase (HRE58_18880; PreQ(0)), and a nucleoside 2-deoxyribosyltransferase (HRE58_18885). These genes are likely part of a repair system for oxidative stress-mediated DNA damage (David, O'Shea, & Kundu, 2007). PreQ(0) might be involved in the insertion of 7-deazaguanine derivatives in the DNA (Hutinet, Swarjo, & de Crécy-Lagard, 2017), as part of a defense system against foreign DNA and phages (Thiaville *et al.*, 2016).

Finally, 9 of 10 genes from GI15 are unique (Supplementary table S3). This island harbor genes encoding a type IV pilus assembly ATPase (HRE58_15865; PilB), a PilE-like protein (HRE58_15910), a methyltransferase (HRE58_15890), and four glycosyltransferases (HRE58_15870-15875, HRE58_15885, HRE58_15895) potentially involved in the biosynthesis and glycosylation of major type IV pilin subunits. *pilB* acts as an important component of type IV pilus assembly, a process that involves a set of conserved *pil* genes (Leighton *et al.*, 2015), including pilC and pilD, located immediately upstream of GI15.Hence, this island is likely an important virulence determinant that mediates locomotion and bacterial adherence to biotic and abiotic surfaces. Although pilin glycosylation has been associated with immune

evasion of pathogenic bacteria (Smedley *et al.*, 2005), a recent study showed its role in the defense against phage infection (H. Harvey *et al.*, 2018). Collectively, these GIs indicate that *S. maltophilia* UENF-4GII is equipped with different horizontally-acquired genes that increased its tolerance to foreign DNA.

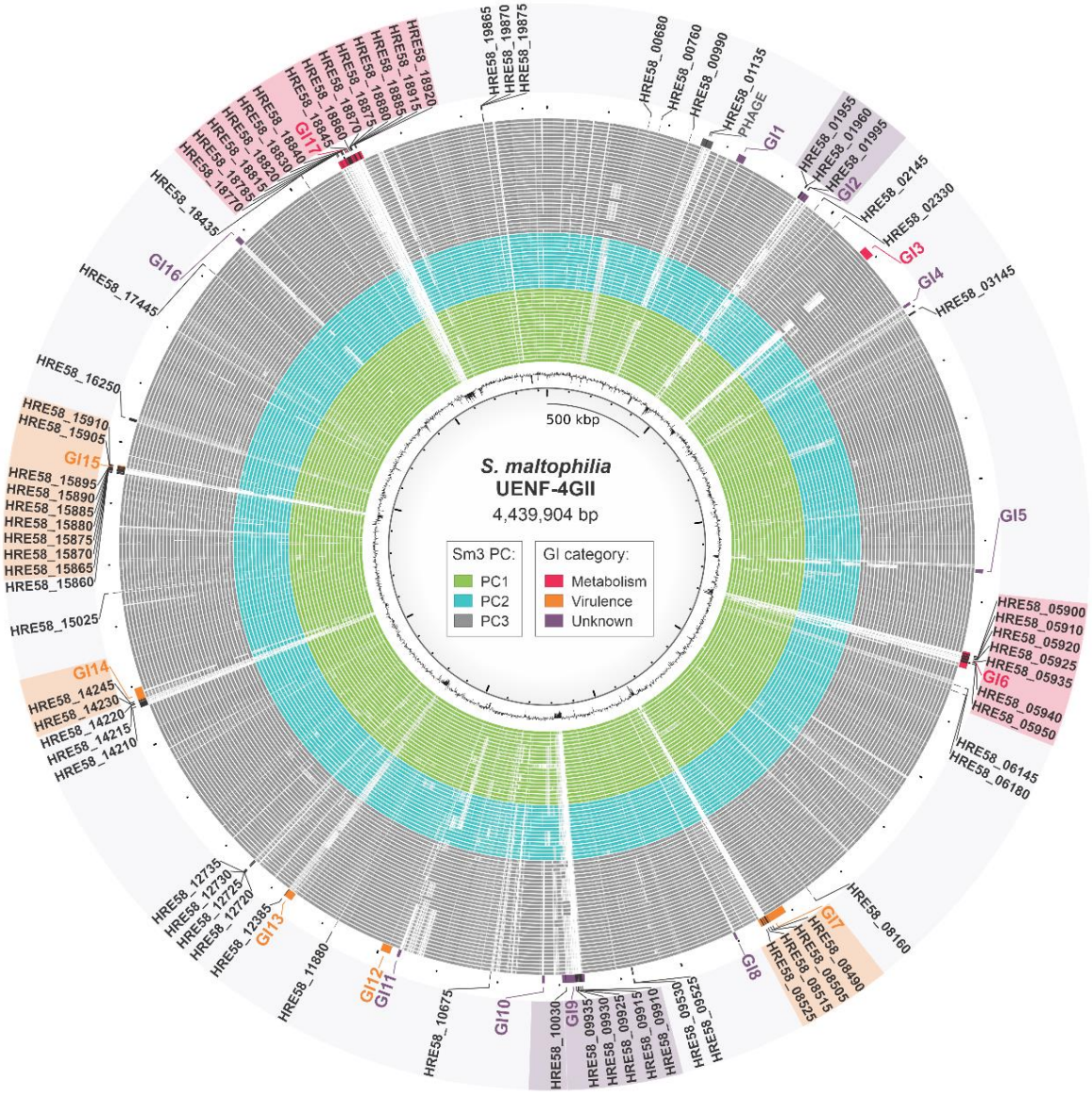


Figure 2-3. Circular genome representation of 75 unique genes and 17 genomic islands (GIs) identified in S. maltophilia UENF-4GII in comparison to other 66 Sm3 strains. The inner ring represents the UENF-4GII genome GC content. Green, blue and gray rings represent genomes from PC1, PC2 and PC3, respectively. The GIs were predicted and classified as metabolic (red), virulence (orange) or unknown (purple). Unique genes predicted within GIs were highlighted according to the GI category.

2.4.5. Pan-GWAS of Sm3 genogroup

We performed a pan-GWAS to identify genes associated with environment type (clinical and non-clinical; Supplementary table S1) and PCs in Sm3 using Scoary. We found 299 and 172 gene clusters associated with clinical and non-clinical environments, respectively (Supplementary figure S3, Supplementary table S4). Interestingly, 58% of the 471 environment-associated genes encode hypothetical proteins with unknown function.

Among the genes associated with clinical environments, we identified genes encoding a chemotaxis response regulator protein (*che*B), blue-light sensing protein (*blu*F), and trehalose synthase *tre*S, a crucial enzyme involved in trehalose biosynthesis. Several bacteria use trehalose as carbon source and stress protectant (Vanaporn & Titball, 2020). Trehalose metabolism has also been associated with the emergence of virulent human pathogens (Collins *et al.*, 2018; Schwarz & Van Dijck, 2017). Interestingly, *tre*S was also reported as a significant gene associated with clinical strains of the *S. maltophilia* complex (Mercier-Darty *et al.*, 2020). Moreover, we identified significant associations with genes involved in bacterial resistance, including multidrug efflux pumps (*mdt*A and *mdt*C), macrolide efflux (*mac*A), aminoglycoside-modifying enzyme (*aac*), β-lactamases (*bla*L1 and *bla*L2), and a β-lactam sensor gene (*bla*R1). BlaL1 and BlaL2 are well known metallo-β-lactamases in *Stenotrophomonas*. They hydrolyze almost all β-lactams, are resistant to all clinically available β-lactamase inhibitors (Mojica *et al.*, 2019; Okazaki & Avison, 2008) and have been show to increase pathogenicity in clinical settings (Han *et al.*, 2020; Mojica *et al.*, 2019). Among the genes associated with non-clinical environments, we also found genes involved in bacterial resistance, particularly in macrolide efflux (*mac*A and *mac*B), as well as three gene clusters involved in heavy metal resistance including copper (*cop*B), cobalt/zinc/cadmium (*czc*B), and mercury (*mer*R1).

As most non-clinical strains belong to PC1 and PC3 comprised only clinical strains, most of the environment-related genes are also linked with their respective phylogenetic groups. A total of 252, 97 and 442 gene clusters were associated, respectively, with PC1, PC2 and PC3 (Figure 2-4, Supplementary table S5). The majority (60%) of the PC-associated genes encoded proteins with unknown functions. PC1 was strongly associated (100% of sensitivity and specificity) with 13 gene clusters, including a type II toxin-antitoxin gene (*pas*I) and c-AMP phosphodiesterase encoding gene (*cpd*A) involved in stress resistance (Barth *et al.*, 2009; Norton & Mulvey, 2012). Further, 11 gene clusters were strongly associated to PC2, including *bes*A, involved in iron metabolism (Miethke *et al.*, 2006), and *kat*G, related with oxidative stress resistance (Chouchane *et al.*, 2003). The PC3 presented 107 associated gene clusters, with 100% sensitivity and specificity. Among these, we identified genes involved in transcriptional regulation (*cue*R, *com*R, *rst*A, and *cys*L), iron metabolism (*bes*A), antibiotic efflux (*mdt*A and *mdt*C), and β-lactam resistance (*bla*L1).

The variants found in different traits (e.g. *bes*A, *mac*A and *bla*L1) are directly associated with genomic diversity of the PCs. A closer look into the *bla*L1 variant of PC3 reveals a molecular heterogeneity with 86% mean similarity with *bla*L1 from PC1

and PC2 genomes. Although the functional implications of such diversity is not fully understood, variation in *bla*L1 have been associated with β-lactam resistance in *S. maltophilia*, which may contribute to its increased prevalence as a nosocomial pathogen (Han *et al.*, 2020; Kim *et al.*, 2020; Mojica *et al.*, 2019).

Figure 2-4. Distribution of PC-associated genes in S. maltophilia Sm3 genogroup. The cgMLST tree is annotated with two strips representing the phylogenetic cluster (PC) and the environment source. The heatmaps represent the presence (dark-gray) or absence (light-gray) of the genes identified by the pan-GWAS pipeline using phylogenetic cluster (PC) as trait. The neon red square highlights the most strongly associated genes found for each trait (i.e. 100% specificity and sensitivity).

2.4.6. Virulome and resistome analysis of the Sm3 genogroup

In order to understand the pathogenic potential within the Sm3 genogroup, we systematically investigated the distribution of virulence and antimicrobial resistance genes. The Sm3 virulome contains 54 genes (Supplementary table S6), out of which 35 constitute the core virulome. Most (21 out of 35) of the core virulence genes are involved in motility and adherence, including *pil*MNOPQ and the *pil*TU operon, which encode type IV pili subunits (Comolli *et al.*, 1999), and *tuf* (elongation factor Tu), involved in adhesion to host cells and extracellular matrix components (K. L. Harvey *et al.*, 2019). In addition, we also identified seven type II secretion system (T2SS) genes (e.g. *xps*EFG and *hxc*RS), which promote the export of enzymes during bacterial colonization. Further, the presence of T2SS genes along with other core

virulence genes (e.g. *acp*Xl, *hem*B, *hem*L, *csr*A, and *icl*), can be involved in infection and immune evasion capacity of Sm3 strains.

The Sm3 accessory virulome comprises 12 low- (genes present in up to 10% of the strains) and 7 high-frequency virulence genes (present in more than 10% of the strains), respectively (Figure 2-5). The high frequency genes are associated with adherence (*pin*N), motility (*flm*H), iron uptake (*bau*A), immune evasion (*rml*B), and stress tolerance (*clp*B, *kat*A, and *kat*G). The *clp*B gene encodes a heat shock-inducible chaperone required for bacterial tolerance to a variety of stresses, including heat, osmotic and acidic stress (Pan *et al.*, 2012). The *kat*A gene encodes a catalase that might confer tolerance against hydrogen peroxide-based disinfectants (Charoenlap *et al.*, 2019; Gröschel *et al.*, 2020) and is absent in all PC2 strains. Nevertheless, PC2 has an alternative catalase-peroxidase gene (*kat*G) that is absent in PC1 and PC3. The *kat*G gene is associated with oxidative stress control and has different roles in pathogenic bacteria (Chouchane *et al.*, 2003; Chun *et al.*, 2009). Four catalases (KatA1, KatA2, KatMn, and KatE) have been described in *S. maltophilia*, out of which KatA has been considered the main factor in conferring tolerance against hydrogen peroxide (Li *et al.*, 2020). The apparent displacement of *kat*A by *kat*G in PC2 warrants further investigation and might provide insights the response of *S. maltophilia* to oxidative stress.

The resistome analysis revealed 23 antimicrobial resistance genes (Supplementary table S7), out of which 17 comprise the core resistome, including the metallo-β-lactamase BlaL1 and the inducible Ambler class A β-lactamase BlaL2, as described above (Sánchez, 2015). Further, antibiotic efflux pumps (e.g. SmeABC, SmeDEF, SmeRS, OqxAB, GolS, EmrE, and MexK) represent the main resistance mechanism encoded by the Sm3 core resistome.

The accessory resistome encompasses six genes (Figure 2-5). No accessory genes were found in PC2, as well as in most PC1 strains, including *S. maltophilia* UENF-4GII. The accessory resistome contains two classes of aminoglycoside-modifying enzymes: aminoglycoside phosphotransferase (encoded by *aph*(6)-Id and *aph*(3″)Ib); and N-aminoglycoside acetyltransferase (encoded by *aac*(6′)-Iz and *aac*(6′)-Iak). The *aac*(6′)-Iz gene was exclusively found in 25 (81%) PC3 genomes, whereas *aac*(6′)-Iak and *aph* genes were found in four PC1 genomes. Only two strains (G4S2 and G4S2-1) have the sulfonamide resistance gene *sul*2 and the chloramphenicol resistance gene *flo*R. In summary, the Sm3 genomes do not present a large repertoire of acquired resistance genes. The low frequency of sulfonamide resistance genes suggests that this antibiotic class might be a good therapeutic strategy against *S. maltophilia* infections, as previously proposed (Gröschel *et al.*, 2020).

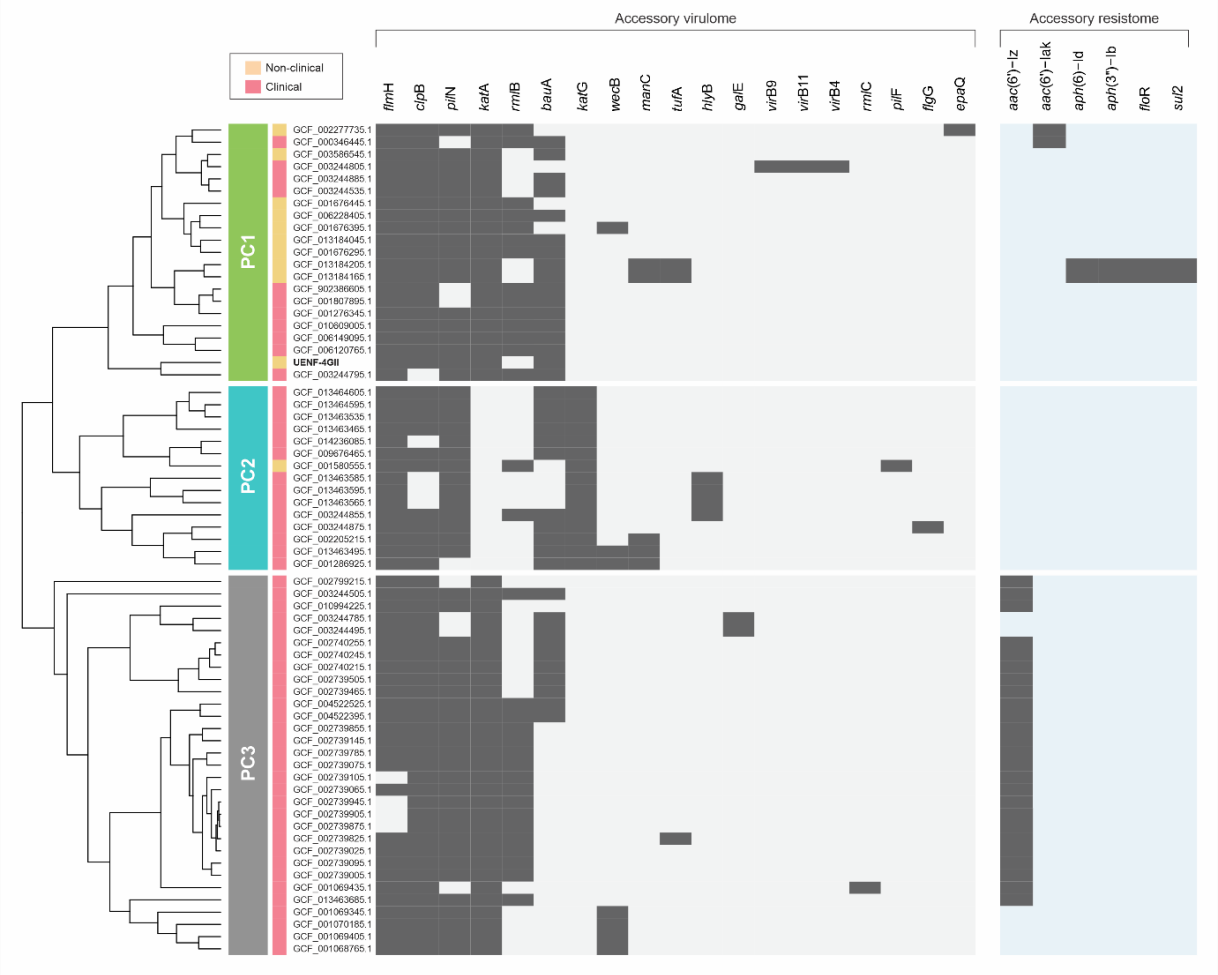


Figure 2-5. Acquired virulome and resistome of S. maltophilia Sm3 genogroup. The cgMLST tree is annotated with two strips representing the phylogenetic cluster (PC) identified and the environment source. The binary heatmaps represent the presence (dark-gray) or absence (light-gray/light-blue) of the genes identified.

2.5. CONCLUSION

In this study, we report the sequencing and in-depth analysis of the *S. maltophilia* UENF-4GII genome. This genome harbors a range of exclusive GIs associated with DNA-modification and anti-phage defense systems. Comparative analysis using pairwise genome identity metrics and cgMLST phylogeny show that *S. maltophilia* UENF-4GII belong to the Sm3 genogroup, which comprises three PCs. Using a pan-GWAS approach, we identified 131 genes as significantly associated with specific Sm3 PCs. Further, 471 genes were specifically associated with clinical and non-clinical environments. These genes could be used as biomarkers in future studies. Sm3 genomes comprise a low number of acquired virulence and resistance genes, although the presence of *kat*G and aminoglycoside resistance genes were associated with specific PCs. Collectively, our results provide important information regarding *S. maltophilia* genomic diversity that could provide the grounds for more detailed clinical and ecological investigations.

ACKNOWLEDGEMENTS

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; grants E-26/203.309/2016 and E-26/203.014/2018), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001), and Conselho Nacional de Desenvolvimento Científico e Tecnológico. The funding agencies had no role in the design of the study and collection, analysis, and interpretation of data and in writing.

AUTHOR CONTRIBUTIONS

Conceived the study: FP-S, TMV, FLO; Funding and resources: TMV, FLO; Data analysis: FP-S, HP-A; Interpretation of the results: FP-S, FPM, HP-A, TMV; Wrote the manuscript: FP-S, FPM, HP-A, TMV.

2.6. REFERENCES

Antipov, D., et al. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics, 32*(22), 3380-3387. doi:10.1093/bioinformatics/btw493

Arndt, D., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research, 44*(W1), W16-W21.

Bankevich, A., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Barth, E., et al. (2009). Interplay of cellular cAMP levels, σS activity and oxidative stress resistance in Escherichia coli. *Microbiology, 155*(Pt 5), 1680.

Bertelli, C., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research, 45*(W1), W30-W35.

Brooke, J. S. (2012). Stenotrophomonas maltophilia: an emerging global opportunistic pathogen. *Clinical microbiology reviews, 25*(1), 2-41.

Charoenlap, N., et al. (2019). Inactivation of ahpC renders Stenotrophomonas maltophilia resistant to the disinfectant hydrogen peroxide. *Antonie Van Leeuwenhoek, 112*(5), 809-814.

Chouchane, S., et al. (2003). Analysis of heme structural heterogeneity in Mycobacterium tuberculosis catalase-peroxidase (KatG). *Journal of Biological Chemistry, 278*(10), 8154-8162.

Chun, H., et al. (2009). The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *Journal of bacteriology, 191*(13), 4152-4157.

Collins, J., et al. (2018). Dietary trehalose enhances virulence of epidemic Clostridium difficile. *Nature, 553*(7688), 291-294.

Comolli, J. C., et al. (1999). Pseudomonas aeruginosa gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. *Infection and immunity, 67*(7), 3625-3630.

David, S. S.,O'Shea, V. L., Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature, 447*(7147), 941-950.

Dimitriu, T.,Szczelkun, M. D., Westra, E. R. (2020). Evolutionary Ecology and Interplay of Prokaryotic Innate and Adaptive Immune Systems. *Current Biology, 30*(19), R1189-R1202.

Domínguez, J., et al. (2019). Changes in the composition and function of bacterial communities during vermicomposting may explain beneficial properties of vermicompost. *Scientific reports, 9*(1), 1-11.

Ealand, C. S.,Machowski, E. E., Kana, B. D. (2018). β-lactam resistance: The role of low molecular weight penicillin binding proteins, β-lactamases and ld-transpeptidases in bacteria associated with respiratory tract infections. *IUBMB life, 70*(9), 855-868.

Falagas, M. E., et al. (2009). Attributable mortality of Stenotrophomonas maltophilia infections: a systematic review of the literature. *Future microbiology, 4*(9), 1103-1109.

Gherardi, G., et al. (2015). An overview of various typing methods for clinical epidemiology of the emerging pathogen Stenotrophomonas maltophilia. *Diagnostic microbiology and infectious disease, 81*(3), 219-226.

Goel, A. K., et al. (2002). Genetic locus encoding functions involved in biosynthesis and outer membrane localization of xanthomonadin in Xanthomonas oryzae pv. oryzae. *Journal of bacteriology, 184*(13), 3539-3548.

Gröschel, M. I., et al. (2020). The phylogenetic landscape and nosocomial spread of the multidrug-resistant opportunist Stenotrophomonas maltophilia. *Nature communications, 11*(1), 1-12.

Guizelini, D., et al. (2016). GFinisher: a new strategy to refine and finish bacterial genome assemblies. *Scientific reports, 6*, 34963.

Gurevich, A., et al. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Han, L., et al. (2020). Diversity of L1/L2 genes and molecular epidemiology of high-level carbapenem resistance Stenotrophomonas maltophilia isolates from animal production environment in China. *Infection, Genetics and Evolution, 86*, 104531.

Harvey, H., et al. (2018). Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. *Nature microbiology, 3*(1), 47-52.

Harvey, K. L., et al. (2019). The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. *Frontiers in microbiology, 10*, 2351.

Hauben, L., et al. (1999). Genomic diversity of the genus Stenotrophomonas. *International Journal of Systematic and Evolutionary Microbiology, 49*(4), 1749-1760.

Huang, Y.-W., et al. (2017). Impacts of penicillin binding protein 2 inactivation on β-lactamase expression and muropeptide profile in Stenotrophomonas maltophilia. *Msystems, 2*(4).

Hutinet, G.,Swarjo, M. A., de Crécy-Lagard, V. (2017). Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. *RNA biology, 14*(9), 1175-1184.

Kalidasan, V., et al. (2018). Iron and virulence in Stenotrophomonas maltophilia: all we know so far. *Frontiers in cellular and infection microbiology, 8*, 401.

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.

Katz, L.,Griswold, T., Carleton, H. (2017). *Generating WGS Trees with Mashtree.* Paper presented at the Poster presented at: American Society for Microbiology Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines.

Kim, Y., et al. (2020). Structural and biochemical analysis of the metallo-β-lactamase L1 from emerging pathogen Stenotrophomonas maltophilia revealed the subtle but distinct di-metal scaffold for catalytic activity. *Protein Science, 29*(3), 723-743.

Leighton, T. L., et al. (2015). Biogenesis of P seudomonas aeruginosa type IV pili and regulation of their function. *Environmental Microbiology, 17*(11), 4148-4163.

Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*.

Li, L.-H., et al. (2020). Protection from hydrogen peroxide stress relies mainly on AhpCF and KatA2 in Stenotrophomonas maltophilia. *Journal of biomedical science, 27*(1), 1-9.

Lim, S. L.,Lee, L. H., Wu, T. Y. (2016). Sustainability of using composting and vermicomposting technologies for organic solid waste biotransformation: recent overview, greenhouse gases emissions and economic analysis. *Journal of Cleaner Production, 111*, 262-278.

Matteoli, F. P., et al. (2018). Genome sequencing and assessment of plant growth-promoting properties of a Serratia marcescens strain isolated from vermicompost. *BMC Genomics, 19*(n. 1), 750. doi:10.1186/s12864-018-5130-y

Medema, M. H., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research, 39*(suppl_2), W339-W346.

Mercier-Darty, M., et al. (2020). Comparative Whole-Genome Phylogeny of Animal, Environmental, and Human Strains Confirms the Genogroup Organization and

Diversity of the Stenotrophomonas maltophilia Complex. *Applied and environmental microbiology, 86*(10).

Miethke, M., et al. (2006). Ferri-bacillibactin uptake and hydrolysis in Bacillus subtilis. *Molecular microbiology, 61*(6), 1413-1427.

Minh, B. Q.,Nguyen, M. A. T., von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution, 30*(5), 1188-1195.

Mojica, M. F., et al. (2019). Population structure, molecular epidemiology, and β-lactamase diversity among Stenotrophomonas maltophilia isolates in the United States. *MBio, 10*(4).

Mukherjee, P., Roy, P. (2016). Genomic potential of Stenotrophomonas maltophilia in bioremediation with an assessment of its multifaceted role in our environment. *Frontiers in microbiology, 7*, 967.

Nas, M. Y., Cianciotto, N. P. (2017). Stenotrophomonas maltophilia produces an EntC-dependent catecholate siderophore that is distinct from enterobactin. *Microbiology, 163*(11), 1590.

Nguyen, L.-T., et al. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution, 32*(1), 268-274.

Norton, J. P., Mulvey, M. A. (2012). Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic Escherichia coli. *PLoS Pathog, 8*(10), e1002954.

O'Leary, N. A., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res, 44*(D1), D733-745. doi:10.1093/nar/gkv1189

Ochoa-Sánchez, L. E., Vinuesa, P. (2017). Evolutionary genetic analysis uncovers multiple species with distinct habitat preferences and antibiotic resistance phenotypes in the Stenotrophomonas maltophilia complex. *Frontiers in microbiology, 8*, 1548.

Okazaki, A., Avison, M. B. (2008). Induction of L1 and L2 β-lactamase production in Stenotrophomonas maltophilia is dependent on an AmpR-type regulator. *Antimicrobial agents and chemotherapy, 52*(4), 1525.

Ondov, B. D., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology, 17*(n. 1), 132. doi:10.1186/s13059-016-0997-x

Page, A. J., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693. doi:10.1093/bioinformatics/btv421

Page, A. J., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics, 2*(4).

Pan, H., et al. (2012). The clpB gene is involved in the stress response of Myxococcus xanthus during vegetative growth and development. *Microbiology, 158*(Pt 9), 2336.

Pathma, J., Sakthivel, N. (2012). Microbial diversity of vermicompost bacteria that exhibit useful agricultural traits and waste management potential. *SpringerPlus, 1*(1), 26.

Pathma, J., Sakthivel, N. (2013). Molecular and functional characterization of bacteria isolated from straw and goat manure based vermicompost. *Applied soil ecology, 70*, 33-47.

Patil, P. P., et al. (2018). Taxonogenomics reveal multiple novel genomospecies associated with clinical isolates of Stenotrophomonas maltophilia. *Microbial genomics, 4*(8).

Pritchard, L., et al. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods, 8*(1), 12-24.

Pruitt, K. D.,Tatusova, T., Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res, 35*(Database issue), D61-65. doi:10.1093/nar/gkl842

Rhee, J.-Y., et al. (2013). Distinct groups and antimicrobial resistance of clinical Stenotrophomonas maltophilia complex isolates from Korea. *Journal of medical microbiology, 62*(5), 748-753.

Ryan, R. P., et al. (2009). The versatility and adaptation of bacteria from the genus Stenotrophomonas. *Nature Reviews Microbiology, 7*(7), 514-525.

Sánchez, M. B. (2015). Antibiotic resistance in the opportunistic pathogen Stenotrophomonas maltophilia. *Frontiers in Microbiology, 6*, 658.

Sarkar, S. K.,Chowdhury, C., Ghosh, A. S. (2010). Deletion of penicillin-binding protein 5 (PBP5) sensitises Escherichia coli cells to β-lactam agents. *International journal of antimicrobial agents, 35*(3), 244-249.

Schwarz, S., Van Dijck, P. (2017). Trehalose metabolism: a sweet spot for Burkholderia pseudomallei virulence. *Virulence, 8*(1), 5-7.

Simão, F. A., et al. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Smedley, J. G., et al. (2005). Influence of pilin glycosylation on Pseudomonas aeruginosa 1244 pilus function. *Infection and immunity, 73*(12), 7922-7931.

Steinmann, J., et al. (2018). Analysis of phylogenetic variation of Stenotrophomonas maltophilia reveals human-specific branches. *Frontiers in microbiology, 9*, 806.

Svensson-Stadler, L. A.,Mihaylova, S. A., Moore, E. R. (2012). Stenotrophomonas interspecies differentiation and identification by gyrB sequence analysis. *FEMS microbiology letters, 327*(1), 15-24.

Tatusova, T., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research, 44*(14), 6614-6624.

Tettelin, H., Medini, D. (2020). The pangenome: Diversity, dynamics and evolution of genomes.

Thiaville, J. J., et al. (2016). Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proceedings of the National Academy of Sciences, 113*(11), E1452-E1459.

Tock, M. R., Dryden, D. T. (2005). The biology of restriction and anti-restriction. *Current Opinion in Microbiology, 8*(4), 466-472.

Vanaporn, M., Titball, R. W. (2020). Trehalose and bacterial virulence. *Virulence, 11*(1), 1192-1202.

Wickham, H., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.

Xie, Z., Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics, 33*(21), 3340-3347.

Yuzenkova, J., et al. (2003). Genome of Xanthomonas oryzae bacteriophage Xp10: an odd T-odd phage. *Journal of molecular biology, 330*(4), 735-748.

Zemanick, E. T., et al. (2017). Airway microbiota across age and disease spectrum in cystic fibrosis. *European Respiratory Journal, 50*(5).

Zerbino, R., Birney, E. (2008). *Velvet: algorithms for de novo short read assembly using de Bruijn graphs* (Vol. 18).

Zheng, Y., et al. (2010). A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Research, 38*(16), 5527-5534.

3. CAPÍTULO 2: Comparative genomics reveals novel species and insights into biotechnological potential, virulence and resistance of *Alcaligenes*

Francisnei Pedrosa-Silva[a, *], Thiago M. Venancio[a, *]

[a] Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Brazil.

* Corresponding authors:

Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF); Av. Alberto Lamego 2000, P5 / sala 217; Campos dos Goytacazes, Rio de Janeiro, Brazil. TMV: thiago.venancio@gmail.com; FP-S: francisneipedrosa@gmail.com.

3.1. ABSTRACT

Alcaligenes is a cosmopolitan bacterial genus that exhibits diverse beneficial to plants. However, the genomic versatility of *Alcaligenes* has also been associated with the ability to cause opportunistic infections in humans, raising concerns about the safety of these microorganisms in biotechnological applications. Here, we report an in-depth comparative analysis of *Alcaligenes* species using all publicly available genomes to investigate genes associated with species, biotechnological potential, virulence, and resistance to multiple antibiotics. Phylogenomic analysis revealed that *Alcaligenes* consists of at least seven species, including three novel species. Pan-GWAS analysis uncovered 389 species-associated genes, including cold shock proteins (e.g., *csp*A) and aquaporins (e.g., *aqp*Z) found exclusively in the water isolated species, *A. aquatilis*. Functional annotation of plant growth-promoting traits revealed enrichment of genes for auxin biosynthesis, siderophores, and organic acids. Genes involved in xenobiotic degradation and heavy metal tolerance were also identified. Virulome and resistome profiles provide insight into selective pressures exerted in the clinical settings. Taken together, the results presented here provide the grounds for more detailed clinical and ecological studies of the genus *Alcaligenes*.

3.2. INTRODUCTION

The genus *Alcaligenes* belongs to the family *Alcaligenaceae* and consists of motile, Gram-negative, and rod- or coccus-shaped bacteria. The bacteria of this genus are widely found in various environments, such as soil, water, plant, and hospital settings. Some *Alcaligenes* species exhibit plant-beneficial properties and have demonstrated capacity to promote plant growth through production of siderophores (Ray *et al.*, 2016; Sayyed & Chincholkar, 2010), solubilization of phosphate, and to antagonize phytopathogenic microorganisms (Ray *et al.*, 2016). In addition, members of *Alcaligenes* genus have demonstrated capacity to remove trace metals such as Cadmium (Kumar, Singh Cameotra, & Gupta, 2012; Ndeddy Aka & Babalola, 2016) and degrade toxic pollutants such as phenol (Rehfuss & Urban, 2005), polyaromatic hydrocarbons (PHAs)(Singha, Kotoky, & Pandey, 2017), and pesticides (Sagarkar *et*

al., 2014), (Yoon *et al.*, 2009; Zhang *et al.*, 2018), demonstrating potential agricultural and industrial benefits.

In contrast to its closely related sister genus *Bordetella*, *Alcaligenes* species are typically non-pathogenic. However, some species have been associated with opportunistic infections (Adabi, Hashemi, & Bakhtiyari, 2022; Tan *et al.*, 2002). The type species of the genus, *Alcaligenes faecalis*, has been associated with nosocomial infections and was detected in clinical material such as blood, respiratory secretions, and urine (Bizet & Bizet, 1997; Dubois *et al.*, 2006). In some cases, *A. faecalis* infection is difficult to treat due to high resistance to multiple antibiotics (Dubois *et al.*, 2006; Huang, 2020; Puah & Chua, 2019).

In recent decades, several isolates of Alcaligenes spp. have been extensively investigated. A comparative genome analysis of *A. aquatilis* QD168 and 25 other strains of *Alcaligenes* spp. identified gene features for abiotic stress and aromatic compound degradation (Durán *et al.*, 2019). Recently, comparison of *Alcaligenes* sp. Mc250 and 13 Alcaligenes spp. genomes revealed genes of biotechnological interest, including for denitrification, benzene degradation, and metabolism of metals such as zinc, cadmium, and arsenic (Felestrino *et al.*, 2020). Although these characteristics have been focused in isolates, the relationships among species within the genus are still not fully explored, especially those associated with biofertilization, bioremediation and resistance to multiple antibiotics.

Here, we report a comprehensive comparative genomic analysis of *Alcaligenes*, which allowed us to uncover important features of the phylogenetic relationships, biotechnological potential, virulence, and resistance profiles of the different species of the genus.

3.3. METHODS

3.3.1. Dataset and genome curation

We downloaded 1,621 *Alcaligenaceae* family genomes from the NCBI Genbank database on May 2022. The genome distance estimation analysis of *Alcaligenaceae* genomes was performed using Mash v.2.2.1 (Ondov *et al.*, 2016) and a family-wide distance tree was generated using MASHtree v.0.50 (Katz, Griswold, & Carleton, 2017). *A. faecalis* DSM 30030 type strain was used as reference to find *Alcaligenes* genomes with maximum Mash distances of 0.15 (~85% Average Nucleotide Identity, ANI (Passarelli-Araujo, Franco, & Venancio, 2022)). Genome quality was evaluated with CheckM v.1.0.13 (Parks *et al.*, 2015), using a minimum of 90% completeness and a maximum of 10% contamination. Genomes with more than 500 contigs were removed, and contigs smaller than 500 bp were removed from the remaining genomes. In order to remove near-identical redundant genomes, we used *in house* scripts to cluster genomes with pairwise Mash distances smaller than 0.005 (~99.95% ANI identity) and keep that with the greatest N50 as the cluster representative. Genome-wide nucleotide identity values were estimated using all-against-all ANI based on MUMmer alignment (ANIm) with pyANI v.0.27 (Pritchard *et al.*, 2016).

3.3.2. Genomic features and phylogeny

All *Alcaligenes* genomes were re-annotated with Prokka v.1.12 (Seemann, 2014) to avoid bias in the identification of protein families. Plasmids and genomic islands were predicted with Plasforest (Pradier *et al.*, 2021) and Islandviewer4 (Bertelli *et al.*, 2017), respectively. Orthologous genes were identified with Orthofinder (Emms & Kelly, 2019). All predicted single-copy orthologous genes present in all isolates were used for maximum-likelihood phylogenetic analysis with IQ-tree (Nguyen *et al.*, 2014), using the best-fit model, JTT+F+I+G4, inferred with ModelFinder. Bootstrap support values were estimated with the ultrafast bootstrap method with 1000 replicates (Minh, Nguyen, & von Haeseler, 2013). The resulting phylogenetic tree was visualized and rendered with iTOL v4 (Letunic & Bork, 2019).

3.3.3. Pangenome analysis

The *Alcaligenes* pangenome was computed with Roary v.3.6, using a 95% identity threshold to determine gene clusters (Page *et al.*, 2015). The pangenome-wide association study (pan-GWAS) analysis was performed with Scoary (Brynildsrud *et al.*, 2016), using the Roary output to establish which genes were typical of *Alcaligenes* groups containing at least five genomes, while correcting for population structure using the phylogenetic tree. False-discovery rate was estimated by Benjamini–Hochberg adjusted p-value provided in Scoary. We only reported the results with specificity > 90% and Benjamini–Hochberg corrected p-value ≤ 0.05. The heatmaps of trait-associated genes were rendered using R package tidyverse (Wickham *et al.*, 2019).

3.3.4. Plant growth-promoting traits analysis

Genes associated with plant growth-promoting traits (PGPT) and of biotechnological interest were predicted using Usearch v.11.0.667 (Edgar, 2010) to search the Plant-associated bacterium database, PLaBAse (Patz *et al.*, 2021), with minimum 50% and 80% identity and coverage thresholds, respectively. PGPT density was calculated by dividing the PGPT count by the total number of coding sequences (CDS) of each genome. Comparison of PGPT number was plotted as a z-scaled heatmap using the R package tidyverse.

3.3.5. Resistome and virulome analysis

Antimicrobial resistance and virulence genes were predicted using Usearch v.11.0.667 to search the *Alcaligenes* genomes against the Comprehensive Antibiotic Resistance Database (CARD) and the Virulence Factors of Pathogenic Bacteria Database (VFDB) databases, respectively. Minimum identity and coverage thresholds of 50% and 80% were used in these searches, respectively. The presence/absence profiles of virulence and resistance-associated genes were rendered using R package tidyverse.

3.4. RESULTS AND DISCUSSION

3.4.1. Data selection and genus classification

In order to perform a systematic comparative analysis of *Alcaligenes* with minimal genome misclassification, we recovered 1,621 *Alcaligenaceae* genomes available in GenBank (May 2022). We computed the pairwise distances between each genome using Mash (Ondov *et al.*, 2016) and generated a distance tree with MASHtree (Katz, Griswold, & Carleton, 2017). We used the type strain *A. faecalis* DSM 30030 (GCF_002443155.1) as an anchor-isolate to evaluate the Mash-distances values from other genomes in the *Alcaligenaceae* family tree. This analysis allowed the identification of a monophyletic group of 68 genomes corresponding to the *Alcaligenes* genus (Figure 3-1A), which showed Mash-distances values up to 0.12 (Figure 3-1B), and ANI densities above 87% (Figure 3-1C). Moreover, the sorted distribution of MASH values of *A. faecalis* DSM 30030 showed an abrupt break around 0.13 (Figure 3-1B), suggesting that 0.15 is an effective threshold for delineating the *Alcaligenes* genus. *Alcaligenes* sp. RS4 (GCA_004028405.1) was removed as a misclassified *Alcaligenes* genome. The set genomes were further filtered to remove highly fragmented and near identical redundant genomes (see methods for more details). The final dataset comprising 50 high-quality *Alcaligenes* genomes was submitted to a series of downstream analyses, as described below.

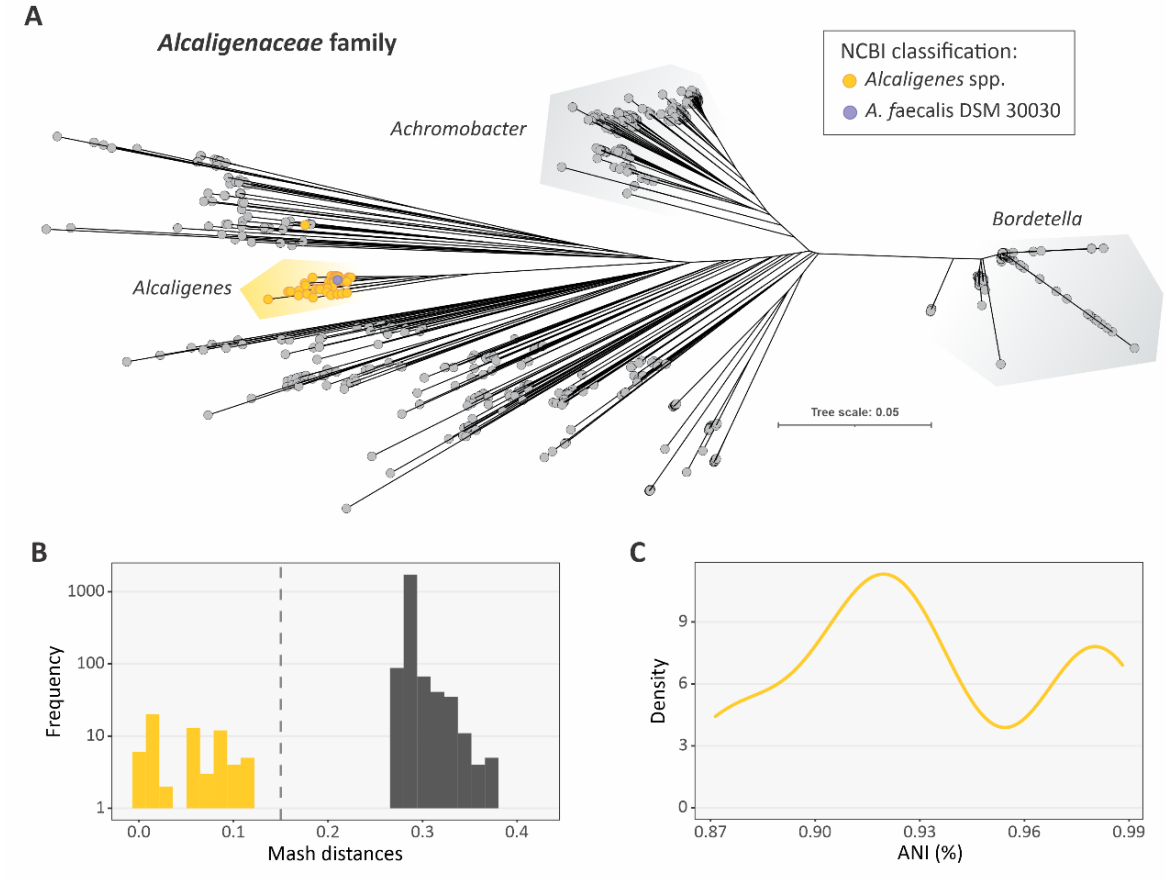


Figure 3-1. Genomic diversity of *Alcaligenaceae* family. (A) Mash-distance-based phylogeny of *Alcaligenaceae*, built using 1,621 publicly available genomes. The two genera

with the greatest number of representative genomes are shaded in gray. The *Alcaligenes* clade is shaded in yellow. (B) Mash- distance between 67 *Alcaligenes* genomes and the type strain *A. faecalis* DSM 30030. The maximum Mash-distance threshold (0.15) used to select genomes is represented by dotted line. (C) Density plot of pairwise ANI of *A. faecalis* DSM 30030 with *Alcaligenes* group.

3.4.2. Phylogenetic analysis of *Alcaligenes*

To uncover the phylogenetic relationships among the *Alcaligenes* genomes, we performed a maximum-likelihood phylogenetic reconstruction using the proteins encoded by 1,172 single-copy orthologous genes found with Orthofinder. Our analysis revealed that *Alcaligenes* comprises at least seven highly supported phylogenetic groups (A1 to A7) (Figure 3-2A). This result is supported by ANIm analysis, which showed genomic identity above 95% between the genomes of each group and allowed us hypothesize that each group comprises a *Alcaligenes* species (Figure 3-2B).

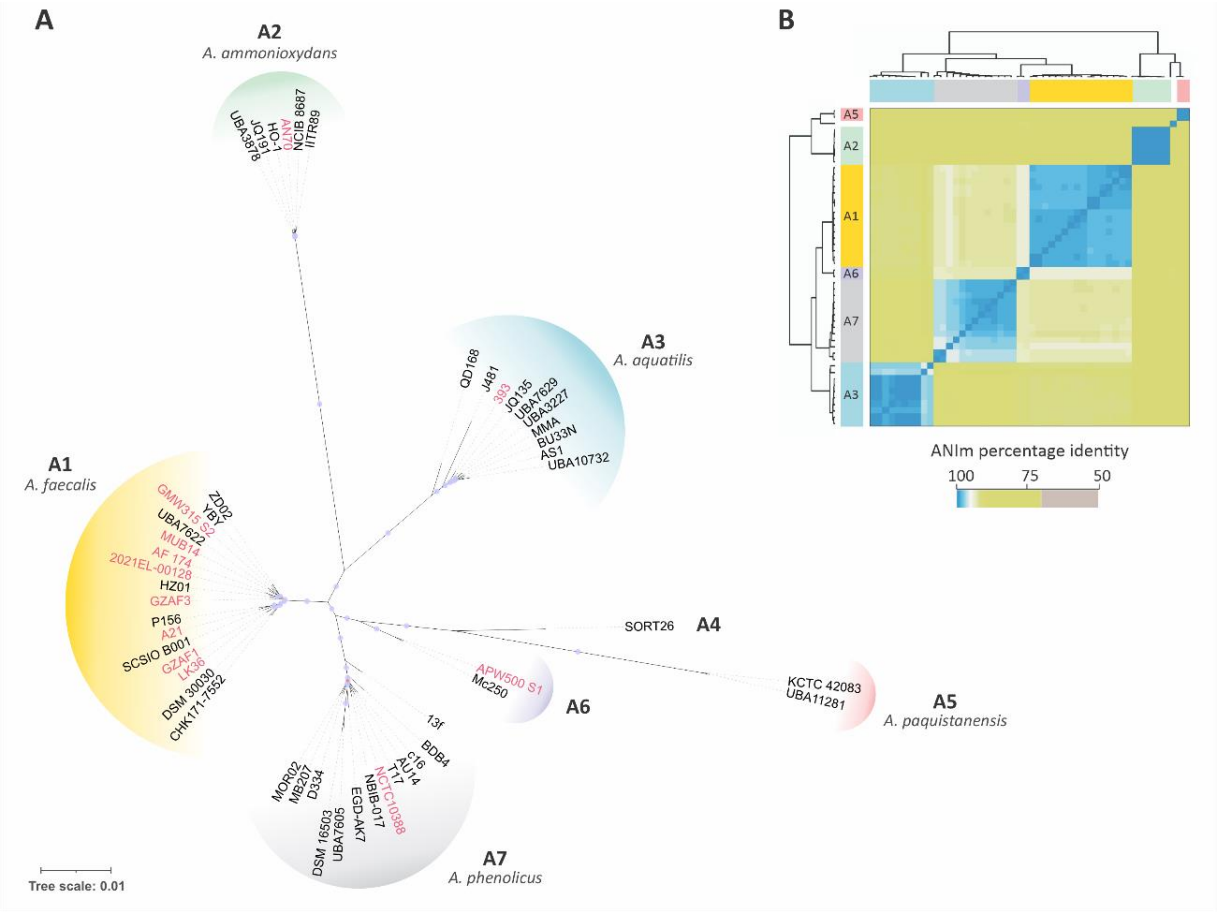


Figure 3-2. Phylogeny analysis and genomic diversity of *Alcaligenes* genus. (A) Phylogenetic tree of *Alcaligenes* genomes showing seven groups. 1,172 single-copy orthologous genes were used to build a maximum likelihood phylogenetic tree using IQ-tree (see methods for details). (B) Pairwise average nucleotide identity (ANI) calculated with 68 *Alcaligenes* genomes.

The *Alcaligenes* group A1 corresponds to *A. faecalis* and contains 16 strains, including the type strain *A. faecalis* DSM 30030 (Table 3-1). The phylogenetic group

A2 corresponds to the recently characterized species, *A. ammonioxydans* (Wu *et al.*, 2021) and comprises six strains, including four misclassified strains (*A. faecalis* AN70, *A. faecalis* subsp. *phenolicus* IITR89, *A. faecalis* subsp. *faecalis* NCIB 8687, and *A. faecalis* UBA7838). *Alcaligenes* group A3 corresponds with *A. aquatilis* species and contains 10 strains, including six misclassified strains (*A. faecalis* J481, *A. faecalis* JQ135, *A. faecalis* UBA10732, *A. faecalis* UBA3227 and *A. faecalis* UBA7629, and *Alcaligenes* sp. MMA). Groups A4 and A6 represent potentially novel species represented by one (*Alcaligenes* sp. SORT26) and two (*A. faecalis* MB250 and *A. faecalis* APW500_S1) isolates, respectively. The phylogenetic group A5 corresponds to *A. pakistanensis* and includes two strains, the type strain *A. pakistanensis* KCTC 42083 and the misclassified strain *A. faecalis* UBA 11281. Finally, the *Alcaligenes* A7 comprising 13 misclassified strains. Based on the presence of the *A. faecalis* subsp. *phenolicus* DSM 16503, also supported by ANI analysis, we proposed to designate this group as *Alcaligenes phenolicus* and reclassify all genomes accordingly. Further, clinical isolates were identified in all groups, except for A3 and A4. Furthermore, *A. faecalis* has the largest frequency of clinical isolates, corresponding to 50% (8) of *Alcaligenes* group A1.

Table 3-1. *Alcaligenes* isolates used in this study and their respective phylogenetic group (PG).

Accession	Strain	NCBI classification	Reclassified as	Source	Location	Environment	PG
GCA_018066525.1	2021EL-00128	*A. faecalis*	-	*Homo sapiens*	USA	Clinical	A1
GCA_015905185.1	A21	*A. faecalis*	-	*Homo sapiens*	Nigeria	Clinical	A1
GCA_003939865.1	AF_174	*A. faecalis*	-	Hospital washroom sink	Pakistan	Clinical	A1
GCA_020741625.1	CHK171-7552	*A. faecalis*	-	Poultry Feces	Bulgaria	Non-clinical	A1
GCA_002443155.1	DSM 30030	*A. faecalis*	-	Activated sludge	unknown	Non-clinical	A1
GCA_019836945.1	GMW315_S2	*A. faecalis*	-	Hospital effluent	South Africa	Clinical	A1
GCA_002120075.1	GZAF1	*A. faecalis*	-	*Homo sapiens*	Israel	Clinical	A1
GCA_002119995.1	GZAF3	*A. faecalis*	-	*Homo sapiens*	Israel	Clinical	A1
GCA_018682645.1	HZ01	*A. faecalis* subsp. *faecalis*	*A. faecalis*	*M. chubuense*	Brazil	Non-clinical	A1
GCA_008373885.1	LK36	*A. faecalis*	-	*Homo sapiens*	USA	Clinical	A1
GCA_010092625.1	MUB14	*A. faecalis*	-	*Homo sapiens*	Poland	Clinical	A1
GCA_001641975.2	P156	*A. faecalis*	-	Soil	China	Non-clinical	A1
GCA_016446305.1	SCSIO B001	*A. faecalis*	-	Fungi	-	Non-clinical	A1
GCA_002484125.1	UBA7622	*A. faecalis*	-	Wood	USA	Non-clinical	A1
GCA_003122065.1	YBY	*A. faecalis*	-	Activated sludge	China	Non-clinical	A1
GCA_000967305.2	ZD02	*A. faecalis*	-	Nematode	China	Non-clinical	A1
GCA_004319585.1	AN70	*A. faecalis*	*A. ammonioxydans*	*Homo sapiens*	China	Clinical	A2
GCA_019343455.1	HO-1	*A. ammonioxydans*	-	Wastewater	China	Non-clinical	A2
GCA_001516865.1	IITR89	*A. faecalis* subsp. *phenolicus*	*A. ammonioxydans*	River water	India	Non-clinical	A2
GCA_022436505.1	JQ191	*A. ammonioxydans*	-	Soil	China	Non-clinical	A2
GCA_000275465.1	NCIB 8687	*A. faecalis* subsp. *faecalis*	*A. ammonioxydans*	-	-	Non-clinical	A2
GCA_002392125.1	UBA3878	*A. faecalis*	*A. ammonioxydans*	Wood	USA	Non-clinical	A2
GCA_003938225.2	393	*A. aquatilis*	-	Hospital washroom sink	Pakistan	Clinical	A3
GCA_023373785.1	AS1	*A. aquatilis*	-	Activated sludge	China	Non-clinical	A3
GCA_003076515.1	BU33N	*A. aquatilis*	-	Sediment	Tunisia	Non-clinical	A3

GCA_003716855.1	J481	*A. faecalis*	*A. aquatilis*	Salt marsh sediment	USA	Non-clinical	A3
GCA_002242175.1	JQ135	*A. faecalis*	*A. aquatilis*	Wastewater	China	Non-clinical	A3
GCA_020907215.1	MMA	*Alcaligenes* sp.	*A. aquatilis*	River water	India	Non-clinical	A3
GCA 003671915.1	QD168	*A. aquatilis*	-	Marine sediment	Chile	Non-clinical	A3
GCA_003511485.1	UBA10732	*A. faecalis*	*A. aquatilis*	-	-	Non-clinical	A3
GCA_002362965.1	UBA3227	*A. faecalis*	*A. aquatilis*	Metal/plastic	USA	Non-clinical	A3
GCA_002484005.1	UBA7629	*A. faecalis*	*A. aquatilis*	Metal	USA	Non-clinical	A3
GCA_017377875.1	SORT26	*Alcaligenes* sp.	-	Residential yard	USA	Non-clinical	A4
GCA_014652815.1	KCTC 42083	*A. pakistanensis*	-	Industrial wastewater	-	Non-clinical	A5
GCA_003521065.1	UBA11281	*A. faecalis*	*A. pakistanensis*	-	-	Non-clinical	A5
GCA_019693795.1	APW500_S1	*A. faecalis*	*Alcaligenes* sp.	Hospital effluent	South Africa	Clinical	A6
GCA_009497775.1	Mc250	*A. faecalis*	*Alcaligenes* sp.	Plant	Brazil	Non-clinical	A6
GCA_020496585.1	13f	*Alcaligenes* sp.	*A. phenolicus*	Soil	Thailand	Non-clinical	A7
GCA_005311025.1	AU14	*A. faecalis*	*A. phenolicus*	Plant	Israel	Non-clinical	A7
GCA_002205415.1	BDB4	*A. faecalis*	*A. phenolicus*	Soil	India	Non-clinical	A7
GCA_016807785.1	c16	*A. faecalis*	*A. phenolicus*	Wastewater	China	Non-clinical	A7
GCA_023101245.1	D334	*A. faecalis*	*A. phenolicus*	Mangrove	Viet Nam	Non-clinical	A7
GCA_000429385.1	DSM 16503	A. faecalis subsp. phenolicus	*A. phenolicus*	Wastewater bioprocessor	USA	Non-clinical	A7
GCA_000465875.3	EGD-AK7	*Alcaligenes* sp.	*A. phenolicus*	Soil	India	Non-clinical	A7
GCA_002082085.1	MB207	A. faecalis subsp. phenolicus	*A. phenolicus*	Tannery effluent	Pakistan	Non-clinical	A7
GCA_000770015.1	MOR02	*A. faecalis*	*A. phenolicus*	Nematode	Mexico	Non-clinical	A7
GCA_001530325.1	NBIB-017	*A. faecalis*	*A. phenolicus*	Soil	China	Non-clinical	A7
GCA_900445215.1	NCTC10388	A. faecalis subsp. faecalis	*A. phenolicus*	*Homo sapiens*	United Kingdom	Clinical	A7
GCA_022343965.1	T17	*A. faecalis*	*A. phenolicus*	Sediment	China	Non-clinical	A7
GCA_002476455.1	UBA7605	*A. faecalis*	*A. phenolicus*	Wood	USA	Non-clinical	A7

3.4.3. Pangenome and Pan-GWAS analyses of *Alcaligenes*

Aiming to better understand genomic traits and dynamics at the genus level, we computed the *Alcaligenes* pangenome. The pangenome is defined as the total number of non-redundant genes present in a given set of genomes (Brockhurst *et al.*, 2019). In our analysis, a total of 13,410 gene clusters were identified. The core genome comprises 2,658 genes present in at least 90% of the genomes. These genes correspond to 19.82% of the pangenome and are typically associated with intrinsic physiological traits.

The accessory genome is composed by 1,726 high-frequency genes (present from 15% to 90% of the genomes) and 9,026 low-frequency genes (present in up to 15% of the genomes), corresponding to 12.87% and 69.30% of the pangenome (Figure 3-3A). The abundance of low-frequency genes further reflects the genome plasticity of *Alcaligenes* and probably plays roles in niche adaptation between species. The Heaps´ law estimate supports an open pangenome (alpha= 0.42), indicating a high level of genetic diversity and allowing us to predict that many more additional gene clusters will be detected as new genomes are sequenced (Figure 3-3B). The open pangenome of *Alcaligenes* is also in line with previous reports (Basharat *et al.*, 2018; Felestrino *et al.*, 2020). Although the *Alcaligenes* pangenome analyses reported here help us understand evolution and genomic dynamics of the genus, a clearer

picture will be available only when a significant number of genomes become sequenced for each species (i.e. phylogenetic group), as divergent genome fluidity and pangenome openness estimates can be found in species from the same genus (Henaut-Jacobs, Passarelli-Araujo, & Venancio, 2022).

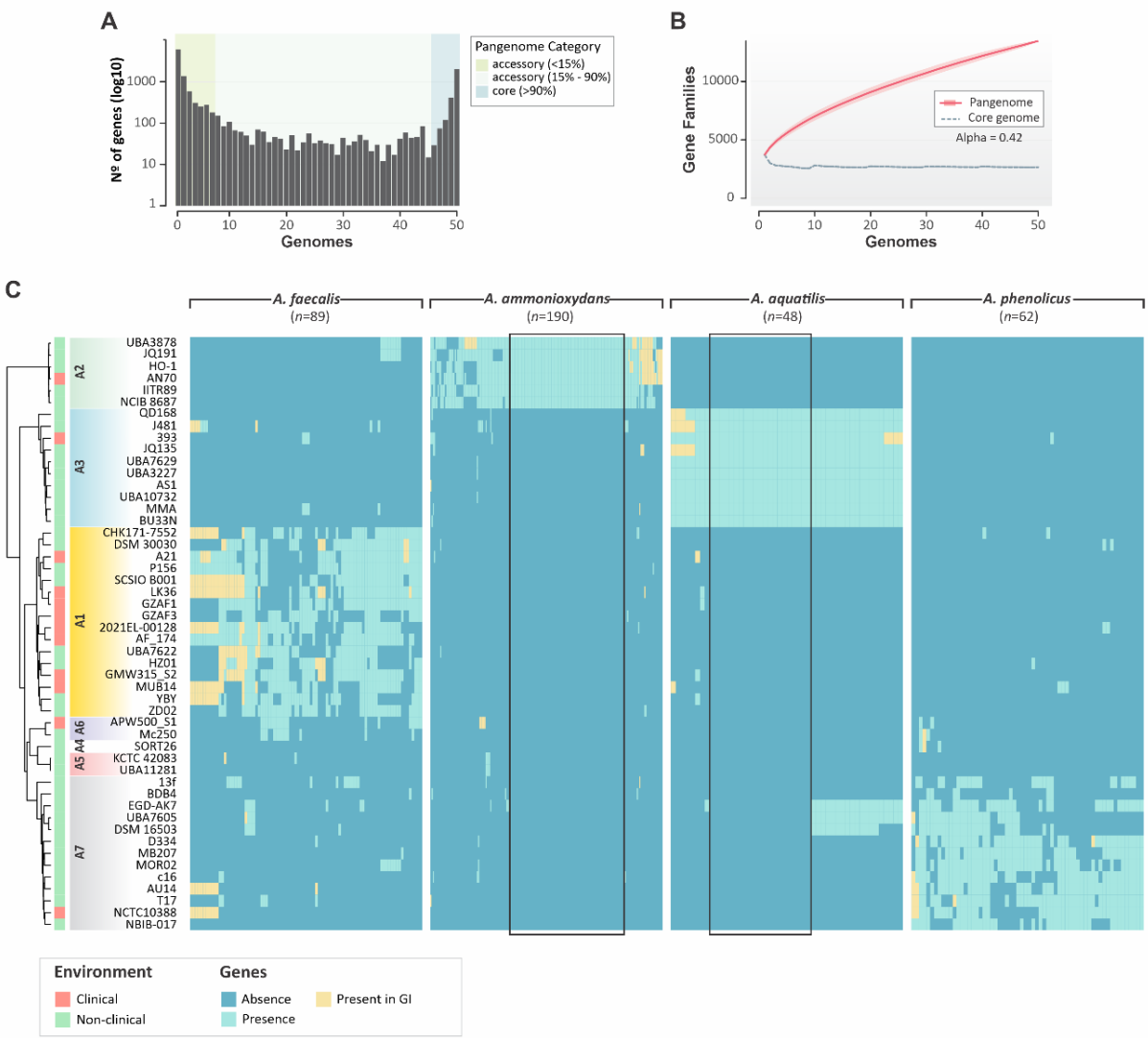


Figure 3-3. Pangenome and pan-GWAS of *Alcaligenes*. (A) Gene frequency of the *Alcaligenes* pangenome. (B) Number of gene families in the *Alcaligenes* pangenome. The cumulative curve (in red) and alpha value of the Heaps' law (0.42) supports an open pangenome. (C) Distribution of species-associated genes in *Alcaligenes*. The heatmaps represent the presence or absence of the genes identified by the pan-GWAS pipeline using *Alcaligenes* groups (species) as traits. The black square highlights the most strongly associated genes found for each trait (i.e., 100% specificity and sensitivity).

We conducted a pan-GWAS analysis to find accessory genes significantly associated to *Alcaligenes* groups. We found a total of 389 genes, consisting in 89, 190, 48, and 62 gene associated with A1, A2, A3, and A7 groups, respectively (Figure 3-3C). Interestingly, most of these genes (63.89%) encode unknown proteins. Further, *A. faecalis* (A1) and *A. phenolicus* (A7) were strongly associated (100% of sensitivity and specificity) with only one gene each, which encodes hypothetical proteins.

However, this analysis revealed important genes for *A. ammonioxidans* (A2) and *A. aquatilis* (A3), mainly related to different stress resistance. *A. ammonioxydans* presented 104 strongly associated genes, including genes involved in amino acid transport (*sfp*, *aax*C and *asp*T), repair of UV-induced DNA damage (*phr*B), nitrosative stress resistance (*hmp*, *nor*R), and salinity stress resistance through trehalose biosynthesis (*ots*AB).

Further, 25 genes were strongly associated to *A. aquatilis*, including a low conductance mechanosensitive channel (*yna*I), an aquaporin Z (*aqp*Z), and a cold shock protein (*csp*A). The *yna*I gene is related with cell protection against hypoosmotic stress (Edwards *et al.*, 2012), *aqp*Z encodes a channel that mediates rapid water influx or efflux in response to abrupt osmolarity changes (Akai *et al.*, 2012), and *csp*A gene encodes a highly conserved DNA-binding protein that is released upon abrupt temperature downshift (Moon *et al.*, 2009). The differential presence of genes involved in resistance to osmotic stress and low temperatures in *A. aquatilis* appears to be related to the aquatic lifestyles of seven isolates from this species (70%) and thus has the potential to be used as molecular markers. Nevertheless, the analysis of a larger number of genomes is warranted to validate this hypothesis.

3.4.4. Functional annotation with Plant-associated bacterium database

Biochemical and molecular studies have shown that *Alcaligenes* species have the potential to promote biofertilization, bioremediation, and heavy metal tolerance (Behera *et al.*, 2017; Dixit *et al.*, 2020; Fatima *et al.*, 2020; Haouas *et al.*, 2021; Sodhi, Kumar, & Singh, 2020; Yoon *et al.*, 2009; Zhang *et al.*, 2018). We investigated the potential of *Alcaligenes* species for these properties by integrating data from plant growth-promoting traits (PGPTs) available in the PlaBAse database. We found between 37.5% to 45% of PGPT density in *Alcaligenes* genomes (Figure 3-4A), which is lower than the average PGPT density (50% to 57%) found in *rhizobiales* species (Ashrafi *et al.*, 2022). No differences in PGPT count were found in clinical and non-clinical genomes, suggesting that the presence of genes associated with plant growth promotion cannot be used to discriminate between beneficial and potentially pathogenic isolates.

The PlaBAse database is classified on six hierarchical functional subclass levels. The level 2 comprises eight PGPT classes, including biofertilization and bioremediation. Comparative analysis of *Alcaligenes* based on PGPT level 2 revealed enrichment biofertilization and bioremediation traits in *A. faecalis* (A1), *A. phenolicus* (A7), and *Alcaligenes* A4 group, while *A. ammonioxydans* (A2) is depleted of such traits (Figure 3-4B). The main genes related to these PGPT are described in the following sections.

3.4.4.1. Biofertilization potential of *Alcaligenes*

Indole-3-acetic acid (IAA) is an auxin-class phytohormone and one of the most important plant growth regulators, known to be produced by several rhizobacteria (Mishra *et al.*, 2016). All the genes required for the synthesis of tryptophan, the IAA

precursor, were identified in all genomes of *Alcaligenes* (Suppl. figure S1). Tryptophan biosynthesis begins with anthranilate synthase (TrpEG) which catalyzes the conversion of chorismate to anthranilate. Anthranilate is converted to indole-3-glycerol phosphate in three steps by anthranilate phosphoribosyltransferase (TrpD), phosphoribosylanthranilate isomerase (TrpF), and indole-3-glycerol phosphate synthase (TrpC). Finally, tryptophan synthase (TrpAB) acts in the last step converting this compound into tryptophan.

The bacterial IAA biosynthesis can occur by at least four tryptophan-dependent pathways, classified according to their intermediates: indole-3-pyruvic acid (IPyA), tryptamine (TRY), indole-3-acetamide (IAM) and indole-3-acetonitrile (IAN) (Spaepen & Vanderleyden, 2011). We found that almost all *Alcaligenes* genomes (except CHK171-7552 and BDB4 strains) have the indole-3-acetamide hydrolase encoding gene (*iaa*H) that catalyzes the last step of IAM to IAA conversion (Suppl. figure S1, Suppl. figure S2). Further, the nitrilase (*nit*A) was identified in all *Alcaligenes* genomes, suggesting that the genus is capable of producing IAA via the IAN pathway.

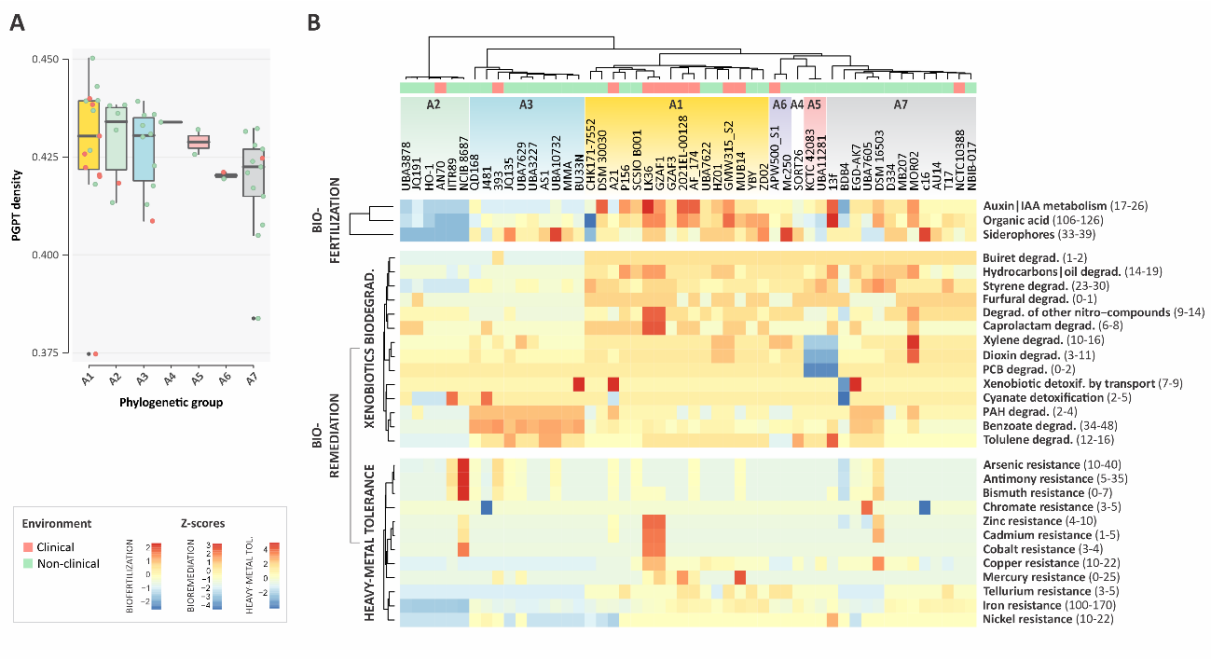


Figure 3-4. Functional plant growth promoting traits (PGPT) of *Alcaligenes* based on PlaBase annotations. (A) PGPT density found in *Alcaligenes*. (B) Heatmap highlighting PGPT abundant differences in functional classes and major genetic traits of *Alcaligenes* genomes. Reddish color indicates enriched and bluish color indicates decreased number of genes based on a trait-specific z-scale. The *Alcaligenes* phylogenetic tree is placed in the top to allow a clearer interpretation of the results.

Phosphate is often a limiting factor for plant growth because it is largely insoluble in the soil, making it unavailable to plants. In this context, some plant growth promoting rhizobacteria (PGPR) solubilize unavailable inorganic phosphate through the secretion of organic acids (Vyas & Gulati, 2009). We identified a set of 106 to 126 genes in *Alcaligenes* involved in the biosynthesis of 19 different organic acids (Suppl. figure S1), including: citric acid (*glt*A, *acn*AB, *prp*C), oxalic acid (*mdh*, *ace*B *dct*A), and

malic acid (*fum*AB, *fum*C), reported as the most effectives in phosphate solubilization by *Alcaligenes faecalis* (Behera *et al.*, 2017). The ribulose 1,5-bisphosphate carboxylase/oxidase (Rubisco) encoding gene (*rbc*L), involved in carbon dioxide fixation and associated with glycolic acid metabolism, was identified only in *A. faecalis* (A1) and *A. phenolicus* (A7) genomes.

Iron is an essential micronutrient for plant growth, required in a number of vital processes such as respiration and photosynthesis. PGPR can enhance iron uptake by plants through the secretion of siderophores that chelate iron from the soil with high affinity (Lurthy *et al.*, 2020). We found 39 genes involved in the biosynthesis and transport of siderophores in all *Alcaligenes* genomes, including the *ent*ABCEF operon, which encodes an enterobactin-like siderophore produced by *Escherichia coli* (Gehring, Mori, & Walsh, 1998; Singh *et al.*, 2022).

3.4.4.2. Biodegradation of xenobiotic compounds

We found 97 genes associated with bioremediation (as per PlaBase classifications) in *Alcaligenes*, encompassing 14 PGPT pathways linked to biodegradation and metabolism of xenobiotics, including dioxin, styrene, naphthalene, polychlorinated biphenyls (PCBs), hydrocarbon, polycyclic aromatic hydrocarbons (PAH), nitro-derivatives of aromatic compounds, and benzene derivatives such as benzoate, toluene and xylene (Figure 4B). *A. pakistanensis* (A5) and *A. phenolicus* 13f presented low gene count for the degradation of xylene (10) and dioxin (3) and no genes for PCB degradation. This occurred due to the absence of 4-hydroxy-2-oxovalerate aldolase (*mph*E) and acetaldehyde dehydrogenase (*mph*F), which are involved in phenol degradation, and the absence of the important dehydrogenase (*bph*B) and HOPDA hydrolase (*bph*D) in PCB degradation pathway. Further, *A. ammonioxydans* (A2) genomes showed a restricted repertoire of genes for the degradation of benzoate (34), toluene (12 genes) and PAH (2 genes).

A. phenolicus was first described as an *A. faecalis* subspecies with the ability to degrade phenolic compounds (Rehfuss & Urban, 2005). The degradation of benzoate and intermediates can be initiated via the aerobic pathway by the action of monooxygenase or dioxygenase, leading to the formation of catechol, protocatechuate, or gentisate (Choi *et al.*, 2013). Genes encoding monooxygenase (*pox*ABCDEF), dioxygenase (*ben*ABCDE), and genes related to the degradation of catechol intermediates (*cat*ABCD) and protocatechuate (*pca*C, *pca*D, *pca*I, *pca*J, *pca*K) were found in all *Alcaligenes* genomes (Suppl. figure S3). Additionally, we identified a set of genes involved in the degradation of protocatechuate derivatives (*lig*AB, *lig*I, *lig*J) in all genomes of *A. aquatilis* and in three genomes of *A. phenolicus* (DSM16503, EGD-AK7, and UBA7605), suggesting greater versatility in the degradation of benzoate and PAHs intermediates in these species. Furthermore, *A. phenolicus* MO02 presented two genes for ethylbenzene dioxygenase (*ebd*B and *ebd*C), suggesting their capacity to conduct aerobic ethylbenzene degradation (Rabus *et al.*, 2002).

Interestingly, all PHA-related genes (*nah*, *oph*E and *pht*5) that we found in *A. phenolicus* BDB4, reported as a potential PHA-degrading bacterium (Singha, Kotoky,

& Pandey, 2017), were present in the *Alcaligenes* core genome and is therefore an intrinsic feature of the genus. Compared with other *Alcaligenes* genomes, the BDB4 strain exhibited a depleted repertoire of copies of *cyn*X (2), and *ecs*AB (7), which are associated with cyanate bioaccumulation and xenobiotic efflux, respectively.

3.4.4.3. Heavy metal tolerance

Bacteria have also become an important component in the remediation of soil and aquatic environments contaminated with heavy metals such as nickel, chromium, tellurium, arsenic, and mercury (Chen *et al.*, 2015). Further, the utilization of metal-tolerant PGPR has proven critical for improving agricultural production in soils contaminated with heavy metals (Mandal & Basu, 2021; Pandey *et al.*, 2013). Genes associated with nickel (*ddp*ABCDF), chromate (*chr*B, *chr*C, *chrR*), and tellurium (*act*P) resistance were found in all *Alcaligenes* genomes. The *A. faecalis* LK36, *A. faecalis* GKAF1, and *A. phenolicus* NCIB_8687 genomes showed greater enrichment for PGPT conferring resistance to cadmium, cobalt, and zinc due to the presence of the *czc*D, *cad*C, *czr*A genes, involved in the regulation and tolerance against these heavy metals (Figure 4B) (Ding *et al.*, 2021; Nies, 1992).

Arsenic resistance genes (*ars*) involved in arsenic detoxification were identified in all *Alcaligenes* genomes (Suppl. figure S4,). In the detoxification pathway, arsenic (As) taken up via the phosphate transferase system (*pst*ABC) is reduced to its less toxic form, arsenite As(III), through the action of As(V) reductase (*ars*C) and then transported to the extracellular medium via efflux pumps (*ars*A/B) and aquaporin Z (*aqp*Z) channels (Ben Fekih *et al.*, 2018). No *aqp*Z genes were found in *A. ammonioxydans* genomes. However, *A. ammonioxydans* NCIB_8687 exhibited the largest gene repertoire for arsenic (40) and antimony (37), resistance due to its high count of *ars*C, *ars*A, and *ars*B genes, and the presence of exclusive genes (*aox*A, *aox*B) that oxidize and mobilize these metals to less toxic forms (Muller *et al.*, 2003; Phung *et al.*, 2012).

Mercury is one of the most toxic pollutants in the environment. It is widely used as an antimicrobial agent in healthcare, agriculture, and industry (Lemire, Harrison, & Turner, 2013). Eight genomes (16%) of *A. faecalis* and *A. aquatilis* harbor mercury resistance genes (*mer*A, *mer*C, *mer*D, *mer*E, *mer*P, *mer*R, and *mer*T) (Suppl. figure S4). Interestingly, with except for *A. aquatilis* UBA7629, only clinical isolates contain these genes. Through a more detailed analysis, we found that these *mer* genes are located in putative genomic islands or plasmids containing antimicrobial resistance (AMR) genes. The *mer* operon encodes a common mercury resistance mechanism that is often carried by transposons and a wide range of plasmids that are ubiquitous in Proteobacteria (Lal & Lal, 2010; Perez-Palacios *et al.*, 2021; Schneiker *et al.*, 2001). Previous studies have reported significant co-occurrence of *mer* operon and AMR genes in mobile elements in a range of isolates, demonstrating that mercury facilitates the selection of multidrug-resistant strains (Hall *et al.*, 2020; Robas *et al.*, 2021; Yuan *et al.*, 2019). Thus, the acquisition of mercury resistance genes by *Alcaligenes* isolates is likely adaptive in clinical settings.

3.4.5. Virulence genes and their distribution across *Alcaligenes* species

We systematically investigated the repertoire of virulence factors using the VFDB database. In addition, we investigated the presence of these genes in genomic islands and plasmids. The *Alcaligenes* virulome comprises 70 genes, of which 48 (68.5%) were found in the core virulome (present in at least 90% of the genomes). No core virulome genes were detected in GI or plasmids. The core virulome comprises genes involved in motility, adherence, biofilm, iron uptake, and immune system evasion, including the *tvi*B and *tvi*C genes, related to resistance against phagocytosis through biosynthesis of VI capsular polysaccharide (Hu *et al.*, 2017). We also identified genes involved in oxidative stress response, including *sod*B and *sod*CI, which encodes an iron and copper/zinc superoxide dismutases (SOD), respectively. *sod* genes encode important metallo-oxidoreductases that convert superoxide radicals into hydrogen peroxide and molecular oxygen and have been shown to neutralize toxic levels of reactive oxygen species generated by a range of hosts, including plants and humans (Cabrejos *et al.*, 2019). For example, SOD is crucial for the endophytic colonization of rice roots by *Glucanocetobacter diazotrophicus* (Alquéres *et al.*, 2013). In humans, the *sod*CI gene has been associated with the survival of pathogenic bacteria upon the oxidative burst caused by phagocytes during infection (Cavinato *et al.*, 2020; Uzzau, Bossi, & Figueroa-Bossi, 2002). Thus, the presence of *sod*CI in *Alcaligenes* might also be involved in the virulence in different hosts.

The accessory virulome (genes present in up to 90% of genomes) comprises 22 genes related to type VI secretions system (T6SS), biofilm, exotoxin, immune modulation, and iron uptake. No exclusive virulence genes were identified. The acinetobactin gene cluster and T6SS genes were differentially distributed across *Alcaligenes* species (Figure 3-5). The siderophore acinetobactin acts as the major iron uptake mechanism in *Acinetobacter baumannii* (Sheldon & Skaar, 2020). Receptors that mediate the recognition and internalization of ferric-acinetobactin complexes are encoded by the *bau*ABCDE operon, which were found in 70% of the *Alcaligenes* genomes, including all the *A. faecalis* and *A. phenolicus* genomes (Figure 3-5). Conversely, in *A. aquatilis* and *A. ammonioxydans*, *bau*ABCDE is rare or even absent, respectively.

T6SS is a potent weapon for interbacterial competition by delivering toxins into prokaryotic cells (Coulthurst, 2019). We found that the T6SS-HSI genes (*hsi*BI and *hsi*CI), which encodes tubule-forming proteins, are absent in *Alcaligenes* group A4, A5, and *A. aquatilis* (A3, except in *A. aquatilis* QD168). Conversely, only *A. aquatilis* and *A. ammonioxydans* have the T6SS-5 genes (*tss*B-5, *tss*C-5, *tss*E-5 and *tss*H-5). Interestingly, most of the *A. aquatilis* T6SS-5 genes are present in genomic islands, indicating their acquisition through horizontal gene transfer.

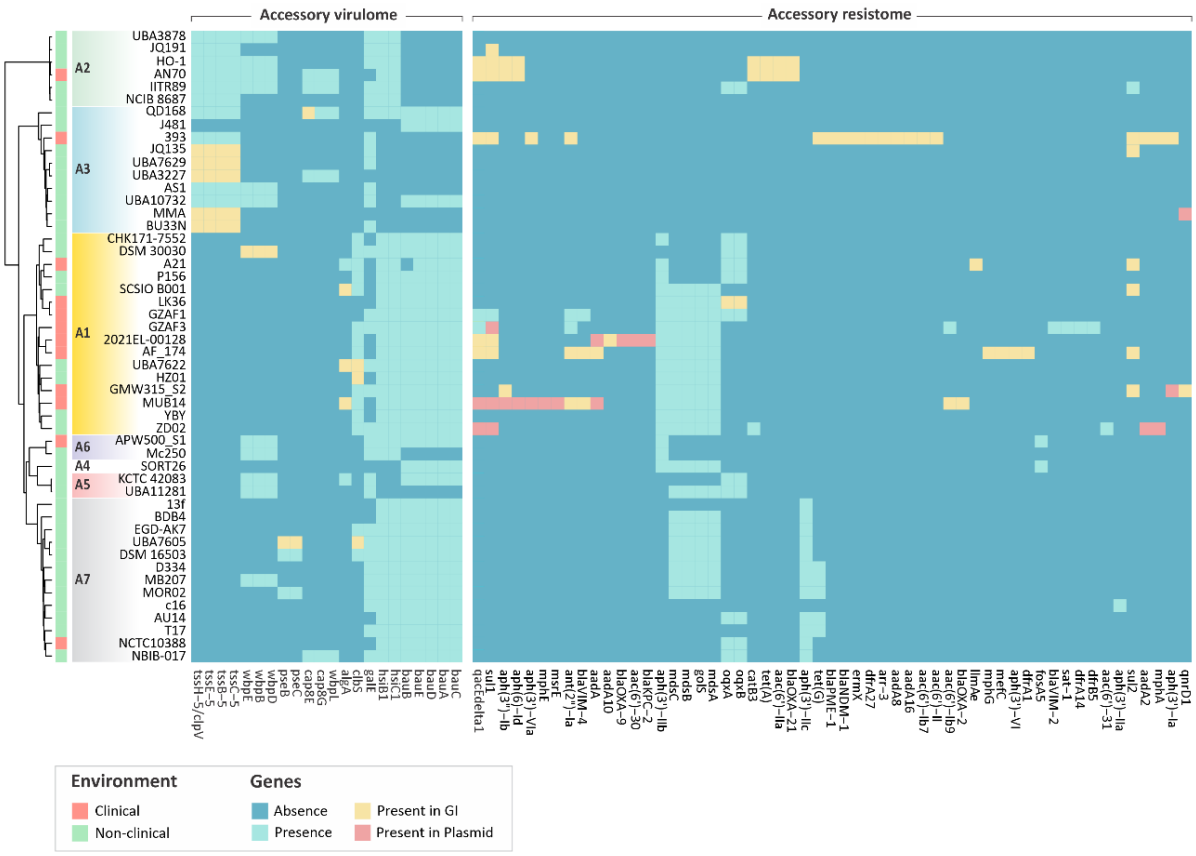


Figure 3-5. Acquired virulome and resistome of *Alcaligenes*. The respective phylogenetic groups of *Alcaligenes* are highlighted in the tree. The heatmaps represent the presence or absence of the genes identified here (see methods for details).

3.4.6. Resistance profiles of clinical and non-clinical genomes

The resistome analysis revealed 55 antimicrobial resistance genes in *Alcaligenes*. The core resistome comprises only one gene, a class A β-lactamase with an average of 54% amino acid sequence identity with BlaSCO-1, a carbenicillin-hydrolyzing β-lactamase (CARB) identified in *Escherichia coli* and *Acinetobacter* spp. (Papagiannitsis *et al.*, 2007; Poirel *et al.*, 2007). However, while BlaSCO-1 is generally found in plasmids (Papagiannitsis *et al.*, 2011; Venditti *et al.*, 2020), this core carbenicillinase is not found in mobile elements in *Alcaligenes*. The presence of CARB-type enzymes corroborates the observed intrinsic penicillin resistance of *Alcaligenes*, as shown for other Gram-negative bacteria (Chiou, Li, & Chen, 2015; Choury *et al.*, 2000).

The *Alcaligenes* accessory resistome has 54 genes, most of them (46) distributed at low frequency (in up to 20% of genomes) (Figure 3-5). The main (63.6%) acquired resistance mechanism in *Alcaligenes* involves antibiotic inactivation enzymes. Further, the genomes of *A. faecalis*, *A. phenolicus*, and the *Alcaligenes* groups A4 and A6, harbor different chromosomal aminoglycoside 3'-phosphotransferase genes (*aph*(3′)-IIa, *aph*(3′)-IIb, and *aph*(3′)-IIc), which confer

resistance against several important aminoglycosides (e.g. kanamycin, neomycin, and seldomycin), constituting an intrinsic core resistance mechanism of these groups.

Drug efflux systems play a major role in resistance of most Gram-negative pathogenic bacteria (Nishino, Latifi, & Groisman, 2006). In contrast to *Achromobacter*, a pathogenic and phylogenetically closely related genus, *Alcaligenes* has a reduced repertoire of efflux pumps (Bador *et al.*, 2013; Magallon *et al.*, 2021; Nielsen, Penstoft, & Nørskov-Lauritsen, 2019). We identified nine genes related to antibiotic efflux, including *mds*ABC and *gol*S, which encode the efflux pump MdsABC and its regulator. The *mds*ABC and *gol*S were found in *Alcaligenes* groups (except in *A. aquatilis*, *A. ammonioxydans* and *Alcaligenes* A6 groups), potentially conferring resistance against different antibiotics, such as β-lactams and phenicol (Conroy *et al.*, 2010; Pontel *et al.*, 2007). Further, the operon *oqx*AB, found in *A. pakistanensis*, *A. faecalis* and *A. phenolicus*, encodes an efflux pump that mediates resistance to multiple antibiotics such as tetracycline, nitrofuran, and fluoroquinolone (Li *et al.*, 2019). We also identified the antiseptic resistance gene *qac*Edelta1, involved in efflux of quaternary ammonium compounds (e.g., disinfecting agents and antiseptics) (Kazama *et al.*, 1999). Interestingly, *qac*Edelta1 is more prevalent in clinical genomes of *A. faecalis*, *A. ammonioxydans*, and *A. aquatilis* (Figure 5), probably due to the high selection pressure imposed by antiseptics and disinfectants in clinical settings. Finally, genes encoding efflux pumps that mediate resistance to tetracycline (i.e., *tet*(G) and *tet*(A)) were found at low frequency in *A. phenolicus* and *A. ammonioxydans*, respectively.

Our results reveal that clinical genomes (*n*=12) harbor around 4 times more AMR genes than that found in non-clinical genomes (*n*=38), with averages of 12 and 3 genes, respectively (Figure 3-6A). The clinical genome of *A. aquatilis* (strain 393) had the greatest number of acquired resistance genes (18), out of which all were exclusively found in a GI, implying their acquisition via HGT. Further, low-frequency resistance genes were mostly found near mobile genetic elements in the genomes of clinical isolates (Figure 3-6B). These genes confer resistance against a range of antibiotic classes, mainly aminoglycosides (*aad*A2, *aad*A8, *aad*A10, *aad*A16, *aac*(6')-30*, aac*(6')-Ib7, *aac*(6')—ib9*, aac*(6')-IIa*, aac*(6')-II*, ant*(2'')-Ia), sulfonamides (*sul*1 and *sul*2), and β-lactams (*bla*OXA-2, *bla*OXA-9, *bla*OXA-21, *bla*VIM-2, *bla*VIM-4, *bla*KPC-2, *bla*NDM-1*. bla*PME-1). Genes encoding resistance against other antibiotic classes, including rifamycin (*arr*-3), nucleoside (*sat*-2), and lincosamide (*llm*A) were nearly absent in *Alcaligenes*.

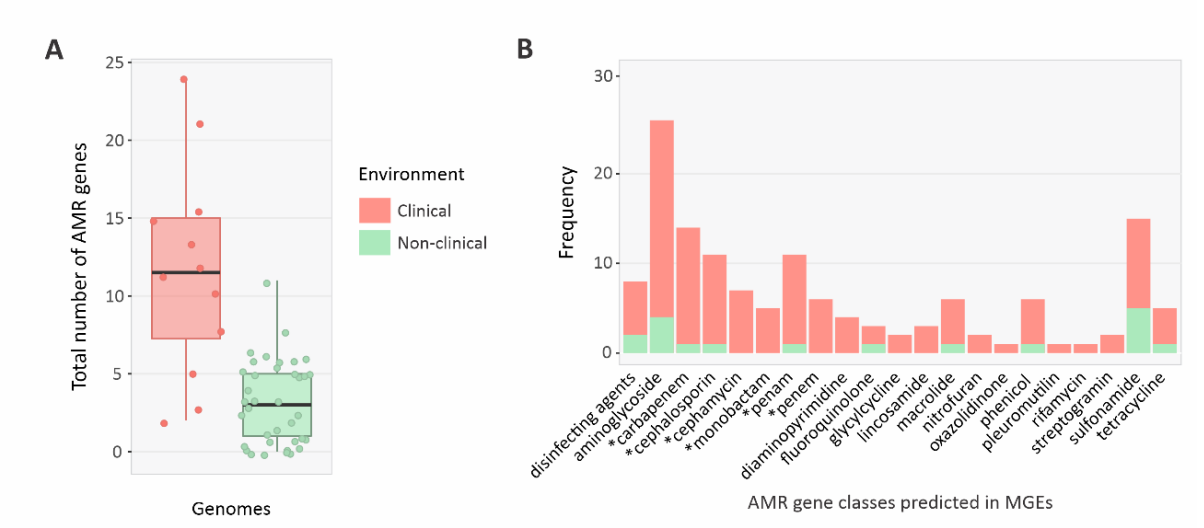


Figure 3-6. Distribution of antimicrobial resistance (AMR) genes in clinical and non-clinical isolates. (A) Total AMR genes identified in *Alcaligenes*. (B) Frequency of different classes of AMR genes in mobile genetic elements (MGE) of clinical and non-clinical genomes. β-lactams are marked with asterisks.

Acquired resistance of *Alcaligenes* has been reported sporadically, mostly associated with multidrug-resistant strains of *A. faecalis*, such as those reported by Huang (2020) in bloodstream, urinary tract, skin, and soft tissue infections. In most cases, *A. faecalis* isolates were resistant to conventional antibiotic classes such as aminoglycosides, β-lactams, and macrolides (Huang, 2020), and these characteristics were also observed in our results. Thus, the low frequency of genes encoding resistance to rifamycin, nucleosides, and lincosamides in the *Alcaligenes* resistome suggests that these antibiotics might be a good therapeutic strategy against pandrug-resistant *Alcaligenes* infections.

3.4.7. CONCLUSION

In this study, we report the comparative analysis of *Alcaligenes* genomes isolated from clinical and non-clinical environments. By assessing genome identity and reconstructing phylogeny, we properly assign *Alcaligenes* genomes to seven different species. Using a pan-GWAS approach, 471 genes were specifically associated with species that contained more than five isolates. *A. ammonioxydans* and *A. aquatilis* had exclusive sets of genes associated with different stress resistance and can be used as biomarkers in future studies. Functional annotation based on PGPT levels showed enrichment of traits for biofertilization and bioremediation in *A. faecalis*, *A. phenolicus* and *Alcaligenes* A4 group. In addition, the functional genes related to xenobiotic degradation and heavy metal tolerance are also discussed in detail. The *Alcaligenes* genomes comprise a low number of virulence and resistance genes, although most AMR genes have been associated with clinical genomes. Overall, our results provide a snapshot of the genetic diversity of the genus *Alcaligenes* and pave the way for more detailed clinical and ecological investigations.

ACKNOWLEDGEMENTS

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; grants E-26/203.309/2016 and E-26/203.014/2018), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001), and Conselho Nacional de Desenvolvimento Científico e Tecnológico. The funding agencies had no role in the design of the study and collection, analysis, and interpretation of data and in writing.

AUTHOR CONTRIBUTIONS

Conceived the study: FP-S, TMV; Funding and resources: TMV; Data analysis: FP-S; Interpretation of the results: FP-S, TMV; Wrote the manuscript: FP-S, TMV.

3.4.8. REFERENCES

Adabi, M.,Hashemi, S. H., Bakhtiyari, S. (2022). The First Study of Investigation of Clinical Isolates of Alcaligenes Xylosoxidans and Alcaligenes Faecalis by Phenotypic and Genetic Methods in Iran. *Iranian Journal of Medical Microbiology, 16*(2), 148-154.

Akai, M., et al. (2012). Aquaporin AqpZ is involved in cell volume regulation and sensitivity to osmotic stress in Synechocystis sp. strain PCC 6803. *Journal of bacteriology, 194*(24), 6828-6836.

Alquéres, S., et al. (2013). The bacterial superoxide dismutase and glutathione reductase are crucial for endophytic colonization of rice roots by Gluconacetobacter diazotrophicus PAL5. *Molecular plant-microbe interactions, 26*(8), 937-945.

Ashrafi, S., et al. (2022). Two new Rhizobiales species isolated from root nodules of common sainfoin (Onobrychis viciifolia) show different plant colonization strategies. *Microbiology Spectrum*, e01099-01022.

Bador, J., et al. (2013). Innate aminoglycoside resistance of Achromobacter xylosoxidans is due to AxyXY-OprZ, an RND-type multidrug efflux pump. *Antimicrobial agents and chemotherapy, 57*(1), 603-605.

Basharat, Z., et al. (2018). Genome sequencing and analysis of Alcaligenes faecalis subsp. phenolicus MB207. *Scientific reports, 8*(1), 1-10.

Behera, B. C., et al. (2017). Alkaline phosphatase activity of a phosphate solubilizing Alcaligenes faecalis, isolated from Mangrove soil. *Biotechnology Research and Innovation, 1*(1), 101-111.

Ben Fekih, I., et al. (2018). Distribution of arsenic resistance genes in prokaryotes. *Frontiers in microbiology, 9*, 2473.

Bertelli, C., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research, 45*(W1), W30-W35.

Bizet, J., Bizet, C. (1997). Strains of Alcaligenes faecalis from clinical material. *Journal of Infection, 35*(2), 167-169.

Brockhurst, M. A., et al. (2019). The ecology and evolution of pangenomes. *Current Biology, 29*(20), R1094-R1103.

Brynildsrud, O., et al. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology, 17*(1), 1-9.

Cabrejos, D. A. L., et al. (2019). Structural characterization of a pathogenicity-related superoxide dismutase codified by a probably essential gene in Xanthomonas citri subsp. citri. *Plos one, 14*(1), e0209988.

Cavinato, L., et al. (2020). Escaping the phagocytic oxidative burst: the role of SODB in the survival of Pseudomonas aeruginosa within macrophages. *Frontiers in microbiology, 11*, 326.

Chen, M., et al. (2015). Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: applications, microbes and future research needs. *Biotechnology advances, 33*(6), 745-755.

Chiou, J.,Li, R., Chen, S. (2015). CARB-17 family of β-lactamases mediates intrinsic resistance to penicillins in Vibrio parahaemolyticus. *Antimicrobial agents and chemotherapy, 59*(6), 3593-3595.

Choi, E. J., et al. (2013). Comparative genomic analysis and benzene, toluene, ethylbenzene, and o-, m-, and p-xylene (BTEX) degradation pathways of Pseudoxanthomonas spadix BD-a59. *Applied and environmental microbiology, 79*(2), 663-671.

Choury, D., et al. (2000). Nucleotide sequence of the bla RTG-2 (CARB-5) gene and phylogeny of a new group of carbenicillinases. *Antimicrobial agents and chemotherapy, 44*(4), 1070-1074.

Conroy, O., et al. (2010). Differing ability to transport nonmetal substrates by two RND-type metal exporters. *FEMS microbiology letters, 308*(2), 115-122.

Coulthurst, S. (2019). The Type VI secretion system: a versatile bacterial weapon. *Microbiology, 165*(5), 503-515.

Ding, J., et al. (2021). Enhanced removal of cadmium from wastewater with coupled biochar and Bacillus subtilis. *Water Science and Technology, 83*(9), 2075-2086.

Dixit, V. K., et al. (2020). Characterization of plant growth-promoting alkalotolerant Alcaligenes and Bacillus strains for mitigating the alkaline stress in Zea mays. *Antonie Van Leeuwenhoek, 113*(7), 889-905.

Dubois, V., et al. (2006). TEM-21 extended-spectrum β-lactamase in a clinical isolate of Alcaligenes faecalis from a nursing home. *Journal of Antimicrobial Chemotherapy, 57*(2), 368-369.

Durán, R. E., et al. (2019). Genomic and physiological traits of the marine bacterium Alcaligenes aquatilis QD168 isolated from Quintero Bay, Central Chile, reveal a robust adaptive response to environmental stressors. *Frontiers in microbiology, 10*, 528.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

Edwards, M. D., et al. (2012). Characterization of three novel mechanosensitive channel activities in Escherichia coli. *Channels, 6*(4), 272-281.

Emms, D. M., Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology, 20*(1), 1-14.

Fatima, T., et al. (2020). Mechanisms of halotolerant plant growth promoting Alcaligenes sp. involved in salt tolerance and enhancement of the growth of rice under salinity stress. *3 Biotech, 10*(8), 1-12.

Felestrino, É. B., et al. (2020). Complete genome sequence and analysis of Alcaligenes faecalis strain Mc250, a new potential plant bioinoculant. *Plos one, 15*(11), e0241546.

Gehring, A. M.,Mori, I., Walsh, C. T. (1998). Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry, 37*(8), 2648-2659.

Hall, J. P., et al. (2020). The impact of mercury selection and conjugative genetic elements on community structure and resistance gene transfer. *Frontiers in microbiology, 11*, 1846.

Haouas, A., et al. (2021). Alcaligenes aquatilis GTE53: Phosphate solubilising and bioremediation bacterium isolated from new biotope “phosphate sludge enriched-compost”. *Saudi Journal of Biological Sciences, 28*(1), 371-379.

Henaut-Jacobs, S.,Passarelli-Araujo, H., Venancio, T. M. (2022). Comparative genomics and phylogenomics of Campylobacter unveil potential novel species and provide insights into niche segregation. *bioRxiv*.

Hu, X., et al. (2017). Vi capsular polysaccharide: Synthesis, virulence, and application. *Critical reviews in microbiology, 43*(4), 440-452.

Huang, C. (2020). Extensively drug-resistant Alcaligenes faecalis infection. *BMC infectious diseases, 20*(1), 1-11.

Katz, L.,Griswold, T., Carleton, H. (2017). *Generating WGS Trees with Mashtree.* Paper presented at the Poster presented at: American Society for Microbiology Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines.

Kazama, H., et al. (1999). Characterization of the antiseptic-resistance gene qacE Δ 1 isolated from clinical and environmental isolates of Vibrio parahaemolyticus and Vibrio cholerae non-O1. *FEMS microbiology letters, 174*(2), 379-384.

Kumar, A.,Singh Cameotra, S., Gupta, S. (2012). Screening and characterization of potential cadmium biosorbent Alcaligenes strain from industrial effluent. *Journal of Basic Microbiology, 52*(2), 160-166.

Lal, D., Lal, R. (2010). Evolution of mercuric reductase (merA) gene: a case of horizontal gene transfer. *Microbiology, 79*(4), 500-508.

Lemire, J. A.,Harrison, J. J., Turner, R. J. (2013). Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nature Reviews Microbiology, 11*(6), 371-384.

Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*.

Li, J., et al. (2019). The nature and epidemiology of OqxAB, a multidrug efflux pump. *Antimicrobial Resistance & Infection Control, 8*(1), 1-13.

Lurthy, T., et al. (2020). Impact of bacterial siderophores on iron status and ionome in pea. *Frontiers in plant science, 11*, 730.

Magallon, A., et al. (2021). Fluoroquinolone resistance in Achromobacter spp.: substitutions in QRDRs of GyrA, GyrB, ParC and ParE and implication of the RND efflux system AxyEF-OprN. *Journal of Antimicrobial Chemotherapy, 76*(2), 297-304.

Mandal, D., Basu, A. (2021). Role of Heavy-Metal Resistant Bacteria Isolated from Rhizosphere in Bioremediation and Plant Development. In *Rhizobiology: Molecular Physiology of Plant Roots* (pp. 411-435): Springer.

Minh, B. Q.,Nguyen, M. A. T., von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution, 30*(5), 1188-1195.

Mishra, P., et al. (2016). Characterization of an indole-3-acetamide hydrolase from Alcaligenes faecalis subsp. parafaecalis and its application in efficient preparation of both enantiomers of chiral building block 2, 3-dihydro-1, 4-benzodioxin-2-carboxylic acid. *Plos one, 11*(7), e0159009.

Moon, C.-H., et al. (2009). Recombinant expression, isotope labeling, and purification of cold shock protein from Colwellia psychrerythraea for NMR study.

Muller, D., et al. (2003). Arsenite oxidase aox genes from a metal-resistant β-proteobacterium. *Journal of bacteriology, 185*(1), 135-141.

Ndeddy Aka, R. J., Babalola, O. O. (2016). Effect of bacterial inoculation of strains of Pseudomonas aeruginosa, Alcaligenes feacalis and Bacillus subtilis on germination, growth and heavy metal (Cd, Cr, and Ni) uptake of Brassica juncea. *International Journal of Phytoremediation, 18*(2), 200-209.

Nguyen, L.-T., et al. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution, 32*(1), 268-274.

Nielsen, S. M.,Penstoft, L. N., Nørskov-Lauritsen, N. (2019). Motility, biofilm formation and antimicrobial efflux of sessile and planktonic cells of Achromobacter xylosoxidans. *Pathogens, 8*(1), 14.

Nies, D. H. (1992). CzcR and CzcD, gene products affecting regulation of resistance to cobalt, zinc, and cadmium (czc system) in Alcaligenes eutrophus. *Journal of bacteriology, 174*(24), 8102-8110.

Nishino, K.,Latifi, T., Groisman, E. A. (2006). Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. *Molecular microbiology, 59*(1), 126-141.

Pandey, S., et al. (2013). Role of heavy metal resistant Ochrobactrum sp. and Bacillus spp. strains in bioremediation of a rice cultivar and their PGPR like activities. *Journal of Microbiology, 51*(1), 11-17.

Papagiannitsis, C., et al. (2007). SCO-1, a novel plasmid-mediated class A β-lactamase with carbenicillinase characteristics from Escherichia coli. *Antimicrobial agents and chemotherapy, 51*(6), 2185-2188.

Papagiannitsis, C., et al. (2011). Sequence of pR3521, an IncB plasmid from Escherichia coli encoding ACC-4, SCO-1, and TEM-1 β-lactamases. *Antimicrobial agents and chemotherapy, 55*(1), 376-381.

Parks, D. H., et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research, 25*(7), 1043-1055.

Passarelli-Araujo, H.,Franco, G. R., Venancio, T. M. (2022). Network analysis of ten thousand genomes shed light on Pseudomonas diversity and classification. *Microbiological Research, 254*, 126919.

Patz, S., et al. (2021). PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv*.

Perez-Palacios, P., et al. (2021). Co-transfer of plasmid-encoded bla carbapenemases genes and mercury resistance operon in high-risk clones of Klebsiella pneumoniae. *Applied Microbiology and Biotechnology, 105*(24), 9231-9242.

Phung, L. T., et al. (2012). Draft genome sequence of Alcaligenes faecalis subsp. faecalis NCIB 8687 (CCUG 2071). In: Am Soc Microbiol.

Poirel, L., et al. (2007). Identification of the novel narrow-spectrum β-lactamase SCO-1 in Acinetobacter spp. from Argentina. *Antimicrobial agents and chemotherapy, 51*(6), 2179-2184.

Pontel, L. B., et al. (2007). GolS controls the response to gold by the hierarchical induction of Salmonella-specific genes that include a CBA efflux-coding operon. *Molecular microbiology, 66*(3), 814-825.

Pradier, L., et al. (2021). PlasForest: a homology-based random forest classifier for plasmid detection in genomic datasets. *BMC bioinformatics, 22*(1), 1-17.

Pritchard, L., et al. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods, 8*(1), 12-24.

Puah, S. M., Chua, K. H. (2019). First report of TEM-116 and OXA-10 extended-spectrum β-lactamase in clinical isolates of Alcaligenes species from Kuala Lumpur, Malaysia. *Japanese journal of infectious diseases*, JJID. 2018.2031.

Rabus, R., et al. (2002). Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbN1. *Archives of microbiology, 178*(6), 506-516.

Ray, S., et al. (2016). Endophytic Alcaligenes isolated from horticultural and medicinal crops promotes growth in okra (Abelmoschus esculentus). *Journal of Plant Growth Regulation, 35*(2), 401-412.

Rehfuss, M., Urban, J. (2005). Alcaligenes faecalis subsp. phenolicus subsp. nov. a phenol-degrading, denitrifying bacterium isolated from a graywater bioprocessor. *Systematic and applied microbiology, 28*(5), 421-429.

Robas, M., et al. (2021). Mercury and antibiotic resistance co-selection in Bacillus sp. isolates from the Almadén mining district. *International Journal of Environmental Research and Public Health, 18*(16), 8304.

Sagarkar, S., et al. (2014). Draft genome sequence of atrazine-utilizing bacteria isolated from Indian agricultural soil. *Genome announcements, 2*(1), e01149-01113.

Sayyed, R., Chincholkar, S. (2010). Growth and siderophores production in Alcaligenes faecalis is regulated by metal ions. *Indian Journal of Microbiology, 50*(2), 179-182.

Schneiker, S., et al. (2001). The genetic organization and evolution of the broad host range mercury resistance plasmid pSB102 isolated from a microbial population residing in the rhizosphere of alfalfa. *Nucleic Acids Research, 29*(24), 5169-5181.

Sheldon, J. R., Skaar, E. P. (2020). Acinetobacter baumannii can use multiple siderophores for iron acquisition, but only acinetobactin is required for virulence. *PLoS pathogens, 16*(10), e1008995.

Singh, P., et al. (2022). In silico analysis of comparative affinity of phytosiderophore and bacillibactin for iron uptake by YSL15 and YSL18 receptors of Oryza sativa. *Journal of Biomolecular Structure and Dynamics*, 1-14.

Singha, L. P.,Kotoky, R., Pandey, P. (2017). Draft genome sequence of Alcaligenes faecalis BDB4, a polyaromatic hydrocarbon-degrading bacterium isolated from crude oil-contaminated soil. *Genome announcements, 5*(48), e01346-01317.

Sodhi, K. K.,Kumar, M., Singh, D. K. (2020). Multi-metal resistance and potential of Alcaligenes sp. MMA for the removal of heavy metals. *SN Applied Sciences, 2*(11), 1-13.

Spaepen, S., Vanderleyden, J. (2011). Auxin and plant-microbe interactions. *Cold Spring Harbor perspectives in biology, 3*(4), a001438.

Tan, K., et al. (2002). Alcaligenes infection in cystic fibrosis. *Pediatric pulmonology, 34*(2), 101-104.

Uzzau, S.,Bossi, L., Figueroa-Bossi, N. (2002). Differential accumulation of Salmonella [Cu, Zn] superoxide dismutases SodCI and SodCII in intracellular bacteria: correlation with their relative contribution to pathogenicity. *Molecular microbiology, 46*(1), 147-156.

Venditti, C., et al. (2020). Reduced susceptibility to carbapenems in a Klebsiella pneumoniae clinical isolate producing SCO-1 and CTX-M-15 β-lactamases together with OmpK35 and OmpK36 Porin deficiency. *Antimicrobial agents and chemotherapy, 64*(8), e00556-00520.

Vyas, P., Gulati, A. (2009). Organic acid production in vitro and plant growth promotion in maize under controlled environment by phosphate-solubilizing fluorescent Pseudomonas. *BMC microbiology, 9*(1), 1-15.

Wickham, H., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.

Wu, M. R., et al. (2021). Novel Alcaligenes ammonioxydans sp. nov. from wastewater treatment sludge oxidizes ammonia to N2 with a previously unknown pathway. *Environmental Microbiology, 23*(11), 6965-6980.

Yoon, I.-H., et al. (2009). Arsenite oxidation by Alcaligenes sp. strain RS-19 isolated from arsenic-contaminated mines in the Republic of Korea. *Environmental geochemistry and health, 31*(1), 109-117.

Yuan, L., et al. (2019). Mercury/silver resistance genes and their association with antibiotic resistance genes and microbial community in a municipal wastewater treatment plant. *Science of The Total Environment, 657*, 1014-1022.

Zhang, Y., et al. (2018). Complete genome sequence of Alcaligenes faecalis strain JQ135, a bacterium capable of efficiently degrading nicotinic acid. *Current microbiology, 75*(12), 1551-1554.

4. CAPÍTULO 3: Comparative and phylogenomic analysis provide insights into resistance, virulence, and plant-growth promoting potential in *Burkholderiaceae*

Francisnei Pedrosa-Silva[a, *], Thiago M. Venancio[a, *]

[a] Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Brazil.

* Corresponding authors:

Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF); Av. Alberto Lamego 2000, P5 / sala 217; Campos dos Goytacazes, Rio de Janeiro, Brazil. TMV: thiago.venancio@gmail.com; FP-S: francisneipedrosa@gmail.com.

4.1. INTRODUCTION

The *Burkholderiaceae* family belongs to the *Burkholderiales* order and consists of metabolically versatile Gram-negative bacteria. The order comprises 24 genera formally described genera, including *Burkholderia* (type genus), *Caballeronia*, *Cupriavidus*, *Polynucleobacter*, *Paraburkholderia*, *Pandoraea*, *Ralstonia*, and *Trinickia*. These genera encompass ecologically diverse organisms found in soil, water, clinical environments, and in associations with different hosts such as plants, insects, fungi, animals, and humans, from which they can develop beneficial or deleterious interactions (Coenye & Vandamme, 2003; diCenzo, Mengoni, & Perrin, 2019; Estrada-De Los Santos *et al.*, 2016).

The genus *Paraburkholderia* is mainly composed of species found in soil, rhizosphere, internal plant tissues, and legume nodules (Dall'Agnol *et al.*, 2017; Sawana, Adeolu, & Gupta, 2014). Several members of this genus, including the type species *Paraburkholderia graminis*, provide beneficial effects to host plants through the nitrogen fixation, solubilization of soil mineral nutrients, and production of phytohormones, which make them an interesting candidate for biofertilizers (Jeon *et al.*, 2021; Rahman *et al.*, 2018).

Caballeronia, Cupriavidus and *Burkholderia* have species with potential for biofertilization (Liu *et al.*, 2022; Pereira *et al.*, 2020; South, Nordstedt, & Jones, 2021). However, these genera also have potentially pathogenic species associated with hospital environments (Butler *et al.*, 2022; Carvalho *et al.*, 2007). Several species of *Pandoraea*, *Ralstonia,* and *Burkholderia* are known to cause opportunistic infections in humans (Fluit *et al.*, 2021; Mahenthiralingam, Baldwin, & Dowson, 2008; Peeters *et al.*, 2019). *Burkholderia cenocepacia* and *B. vietnamiensis* belong to the *Burkholderia cepacia complex* and affect mainly immunocompromised patients (Tavares *et al.*, 2020). Infections caused by these bacteria are difficult to treat due to significant antibiotic resistance, specially of β-lactams classes. In addition, *Burkholderia* harbors two highly pathogenic species, *B. pseudomallei* e *B. mallei*, which cause glanders and melioidosis, respectively, and are classified as potential bioterrorism agents (Rotz *et al.*, 2002). In addition, *Ralstonia* and *Burkholderia* also comprise phytopathogenic

species that cause important diseases such as bacterial panicle blight and *Ralstonia* wilt, caused by *B. glumae* and *Ralstonia solanacearum*, respectively (Ortega & Rojas, 2021; Rivard *et al.*, 2012).

The pathogenicity of several *Burkholderiaceae* has raised safety concerns about the use of beneficial microorganisms in biotechnological applications, as well as questions about the sharing of mechanisms related with pathogenicity to other members of this family. Genomic analyses of different *Burkholderiaceae* genera have been extensively applied as a strategy to identify molecular markers to discriminate pathogens from beneficial species or isolates (Bernier *et al.*, 2003; Lee *et al.*, 2016). However, systematic and family-wide studies of *Burkholderiaceae* remain scarce.

In this work, we report the phylogenomic analysis and genome structure of *Burkholderiaceae*, which was used to better understand the distribution of plant growth-promoting genes of genera. Further, we used the inferred genome structure to estimate the diversity and prevalence of resistance and virulence genes from *Burkholderiaceae*.

4.2. METHODS

4.2.1. Dataset and genome assessment

We downloaded 5,481 *Burkholderiaceae* genomes from the NCBI Genbank database in March 2022. Genome quality was assessed with CheckM v.1.0.13 (Parks *et al.*, 2015), using a minimum of 90% completeness and a maximum of 10% contamination. Genomes with more than 600 contigs were excluded and contigs shorter than 500 bp were removed. To avoid nearly identical genomes, we used *in-house* scripts to cluster genomes with Mash distances lower than 0.005 (~99.95% ANI identity) and keep the genome with the highest N50 as the cluster representative. The identification of type strains was based on descriptions found in the List of Prokaryotic Names with Standing in Nomenclature (LPSN, www.lpsn.dsmz.de).

4.2.2. Phylogenomic analysis and classification of *Burkholderiaceae* genomes

The distance estimation of *Burkholderiaceae* genomes was performed with Mash v.2.2.1 (Ondov *et al.*, 2016), using the standard *k*-mer length of 21 nt in addition to a sketch size of 5000, rather than the default of 1000, as increasing the sketch size could improve the accuracy of the estimations. Genome-wide nucleotide identity values were estimated with FastANI v.1.32 (Jain *et al.*, 2018), using a fragment length of 1,020 bp to better approximate the original ANI calculations (Goris *et al.*, 2007). FastANI was designed to estimate identities greater than 80%, which are appropriate to establish relationships between species. Conversely, Mash can handle greater divergence estimates and can be used to investigate relationships between more distantly related genomes.

To increase resolution in the comparison between genomes of different genera, we used *in-house* scripts to combine the Mash and FastANI matrices and generate a consistent family distance tree. For this analysis, ANI values were converted to dissimilarity, as represented by Mash. Values less than 20% dissimilarity in the Mash

matrix were replaced by dissimilarity values from ANI. The distance tree was built with Quicktree v.1.3 (Howe, Bateman, & Durbin, 2002) and rendered with iTOL v4 (Letunic & Bork, 2019).

4.2.3. Plant growth-promoting, resistome, and virulome analysis

All *Burkholderiaceae* genomes were re-annotated with Prokka v.1.12 (Seemann, 2014) to avoid bias in the identification of protein families. Plasmids and genomic islands (GI) were predicted with Plasforest (Pradier *et al.*, 2021) and Islandviewer4 (Bertelli *et al.*, 2017), respectively. Genes associated with plant growth promotion (PGP) were predicted using Usearch v.11.0.667 (Edgar, 2010) to search a manually curated database of genes related to nodulation (*nod*) and biofertilization, including the main genes involved in biological nitrogen fixation (*nif*), inorganic phosphate solubilization (*pqq*), auxin biosynthesis (*iaa*M, *iaa*H and *ipdC)* and reduction of ethylene levels (*acd*S).

Antimicrobial resistance and virulence genes were predicted using Usearch v.11.0.667 to search Burkholderiaceae genomes against the Comprehensive Antibiotic Resistance Database (CARD) and the Virulence Factor Databases of the bacterial pathogen database (VFDB), respectively. Minimum identity and coverage thresholds of 50% and 80% were used in all these searches, respectively. Presence/absence profiles of genes associated with PGP, virulence, and resistance were rendered using the tidyverse R package.

4.3. RESULTS AND DISCUSSION

4.3.1. Genome diversity of *Burkholderiaceae*

We obtained 5,481 *Burkholderiaceae* genomes available in Refseq in March 2022, of which 5,298 showed low contamination (less than 10%) and high completeness estimates (see methods for details). Genomes were filtered to remove redundant and highly fragmented and nearly identical genomes, comprising a final dataset of 1,921 high quality *Burkholderiaceae* genomes. This abrupt reduction in the number of genomes could be justified by the presence of sequenced genomes associated with clonal bacterial outbreaks caused mainly by *Burkholderia* species. The Mash and ANI analysis provides a raw estimate of bacterial genera and species (Parks *et al.*, 2020). To identify the phylogenomic relationships between genera in the family, we computed the pairwise distances between each genome using Mash and combined them with the mean nucleotide identity (ANI) values generated by FastANI to estimate a robust family distance tree (Figure 4-1).

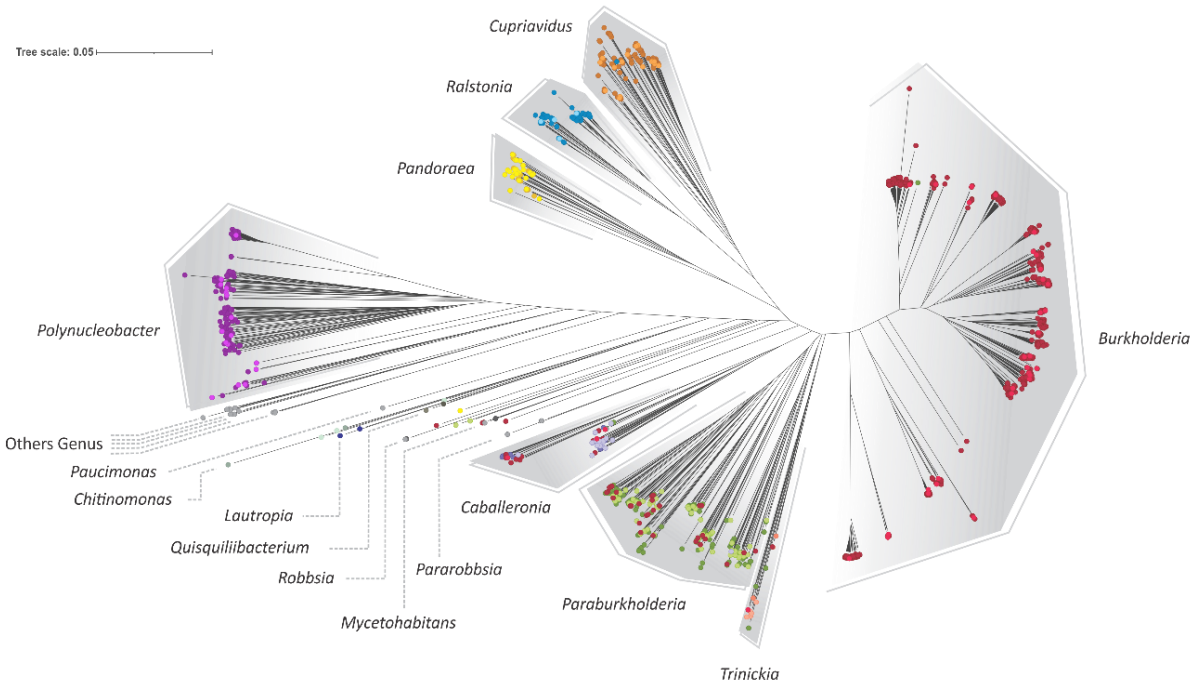


Figure 4-1. Genomic diversity of the *Burkholderiaceae* family. ANI-Mash distance-based phylogeny of *Burkholderiaceae*, built using 1921 publicly genomes.

We identified 25 phylogenetic groups corresponding to the number of genera expected within the *Burkholderiaceae* family, which is highly diverse. Based on the Mash and FastANI values, this analysis allowed us to identify and reclassify 61 genomes at the genus level (Table 1), including the genomes of *Paraburkholderia* (n=33), *Caballeronia* (n=14), *Trinickia* (n=5), *Mycetohabitans* (n=2), *Robbsia* (n=1), and *Cupriavidus* (n=1).

Table 4-1. *Burkholderiaceae* genomes reclassified in this study.

Accession	Genus	Species	Strain	Type	Source	Genus reclassification
GCF_002879885.1	*Burkholderia*	*dabaoshanensis*	GIMN1.004	Environmental	Soil	*Trinickia*
GCF_005280735.1	*Burkholderia*	sp.	DHOD12	Environmental	Soil	*Trinickia*
GCF_003367175.1	*Paraburkholderia*	sp.	DHOM06	Environmental	Soil	*Trinickia*
GCF_003628145.1	*Paraburkholderia*	sp.	7MK8-2	Environmental	Soil	*Trinickia*
GCF_013282575.1	*Paraburkholderia*	sp.	NMBU_R16	Environmental	Water	*Trinickia*
GCF_900095735.1	*Burkholderia*	sp.	JS23	Missing	Missing	*Robbsia*
GCF_000148685.1	*Burkholderia*	sp.	CCGE10	Environmental	Plant	*Paraburkholderia*
GCF_000176935.3	*Burkholderia*	sp.	CCGE10-1	Environmental	Plant	*Paraburkholderia*
GCF_000178415.2	*Burkholderia*	sp.	Ch1-1	Environmental	Soil	*Paraburkholderia*
GCF_000472445.1	*Burkholderia*	sp.	WSM2230	Environmental	Plant	*Paraburkholderia*
GCF_000472545.1	*Burkholderia*	sp.	WSM2232	Environmental	Plant	*Paraburkholderia*
GCF_000519245.1	*Burkholderia*	sp.	UYPR1.413	Environmental	Plant	*Paraburkholderia*
GCF_001931485.1	*Burkholderia*	sp.	SRS-W-2-2016	Environmental	Soil	*Paraburkholderia*
GCF_002087245.1	*Burkholderia*	sp.	A27	Environmental	Plant	*Paraburkholderia*
GCF_002196005.1	*Burkholderia*	sp.	Bk	Environmental	Soil	*Paraburkholderia*

GCF_002870125.1	*Burkholderia*	sp.	WAC0059	Environmental	Fungus	*Paraburkholderia*
GCF_005503145.1	*Burkholderia*	sp.	4M9327F10	Environmental	Water	*Paraburkholderia*
GCF_007829195.1	*Burkholderia*	sp.	604	Environmental	Missing	*Paraburkholderia*
GCF_009765935.1	*Burkholderia*	sp.	S1	Environmental	Soil	*Paraburkholderia*
GCF_011752995.1	*Burkholderia*	sp.	Ax-1724	Environmental	Fungus	*Paraburkholderia*
GCF_012275575.1	*Burkholderia*	sp.	SG-MS1	Environmental	Soil	*Paraburkholderia*
GCF_014873985.1	*Burkholderia*	sp.	OAS9	Environmental	Plant	*Paraburkholderia*
GCF_016649235.1	*Burkholderia*	sp.	R-70211	Environmental	Soil	*Paraburkholderia*
GCF_016649255.1	*Burkholderia*	sp.	R-70006	Environmental	Soil	*Paraburkholderia*
GCF_016649265.1	*Burkholderia*	sp.	R-70199	Environmental	Soil	*Paraburkholderia*
GCF_016649295.1	*Burkholderia*	sp.	R-69980	Environmental	Soil	*Paraburkholderia*
GCF_016649305.1	*Burkholderia*	sp.	R-69749	Environmental	Soil	*Paraburkholderia*
GCF_016649335.1	*Burkholderia*	sp.	R-69927	Environmental	Soil	*Paraburkholderia*
GCF_016649375.1	*Burkholderia*	sp.	R-69608	Environmental	Soil	*Paraburkholderia*
GCF_017104715.1	*Burkholderia*	sp.	Ac-20365	Environmental	Fungus	*Paraburkholderia*
GCF_900142905.1	*Burkholderia*	sp.	GAS332	Environmental	Missing	*Paraburkholderia*
GCF_900168075.1	*Burkholderia*	sp.	CF099	Environmental	Plant	*Paraburkholderia*
GCF_900230185.1	*Burkholderia*	sp.	OK233	Environmental	Plant	*Paraburkholderia*
GCF_900230235.1	*Burkholderia*	sp.	YR290	Environmental	Plant	*Paraburkholderia*
GCF_011753055.1	*Burkholderia*	sp.	Ax-1719	Environmental	Fungus	*Paraburkholderia*
GCF_000745015.1	*Burkholderia*	sp.	9120	Missing	Missing	*Paraburkholderia*
GCF_900096975.1	*Burkholderia*	*lycopersici*	TNe-862	Missing	Missing	*Paraburkholderia*
GCF_900104795.1	*Burkholderia*	sp.	WP9	Missing	Missing	*Paraburkholderia*
GCF_007474635.1	*Caballeronia*	*humi*	KEMC7302-068	Missing	Missing	*Paraburkholderia*
GCF_009765705.1	*Burkholderia*	sp.	L27(2015)	Environmental	Soil	*Burkholderiaceae* sp.
GCF_001017775.3	*Pandoraea*	*thiooxydans*	DSM 25325	Environmental	Soil	*Burkholderiaceae* sp.
GCF_009455625.1	*Paraburkholderia*	*bonniea*	BbQS859	Environmental	Amoeba	*Burkholderiaceae* sp.
GCF_009455685.1	*Paraburkholderia*	*hayleyella*	BhQS11	Environmental	Amoeba	*Burkholderiaceae* sp.
GCF_900155955.1	*Burkholderia*	sp.	B13	Missing	Missing	*Mycetohabitans*
GCF_900156195.1	*Burkholderia*	sp.	B14	Missing	Missing	*Mycetohabitans*
GCF_900104095.1	*Ralstonia*	sp.	25MFCol4.1	Missing	Missing	*Cupriavidus*
GCF_000236065.1	*Burkholderia*	sp.	YI23	Environmental	Soil	*Caballeronia*
GCF_000828875.1	*Burkholderia*	sp.	RPE67	Environmental	Insect	*Caballeronia*
GCF_001424345.1	*Burkholderia*	sp.	Leaf177	Environmental	Plant	*Caballeronia*
GCF_001557535.2	*Burkholderia*	sp.	PAMC 26561	Environmental	Lichen	*Caballeronia*
GCF_001577465.1	*Burkholderia*	sp.	PAMC 28687	Environmental	Fungus	*Caballeronia*
GCF_001988975.1	*Burkholderia*	sp.	KK1	Environmental	Water	*Caballeronia*
GCF_003269235.1	*Burkholderia*	sp.	yr520	Environmental	Plant	*Caballeronia*
GCF_009766085.1	*Burkholderia*	sp.	S171	Environmental	Soil	*Caballeronia*
GCF_009938265.1	*Burkholderia*	sp.	THE68	Environmental	Insect	*Caballeronia*
GCF_018408855.1	*Burkholderia*	sp.	NK8	Environmental	Soil	*Caballeronia*
GCF_900215135.1	*Burkholderia*	sp.	D7	Environmental	Plant	*Caballeronia*
GCF_902506145.1	*Burkholderia*	sp.	8Y	Environmental	Algae	*Caballeronia*
GCF_003097055.1	*Paraburkholderia*	sp.	OV5	Environmental	Plant	*Caballeronia*

GCF_004339005.1	*Paraburkholderia*	sp.	BL8N3	Environmental	Missing	*Caballeronia*
GCF_002097715.1	*Paraburkholderia*	*acidophila*	ATCC 31433	Environmental	Soil	*Burkholderia*

Among the most representative genera, *Burkholderia* is the genus with the greatest number of genomes, followed by *Paraburkholderia*, *Polynucleobacter*, *Cupriavidus*, *Ralstonia*, *Caballeronia,* and *Pandoraea* (Figure 4-2A). The size of the genomes of each genus corresponds to the presence of multipartite genomes (containing chromids and megaplasmids) reported in *Burkholderiaceae* (diCenzo, Mengoni, & Perrin, 2019), with *Paraburkholderia* being the genus with the highest average genome size, followed by *Caballeronia*, *Burkholderia*, *Cupriavidus* and *Trinickia* (Figure 4-2B). The genus composed of bacterioplanktons, *Polynucleobacter*, had the smallest average genome size, which is in line with its planktonic lifestyle in freshwater environments and as endosymbionts of ciliated protozoa (Boscaro *et al.*, 2013; diCenzo, Mengoni, & Perrin, 2019). Furthermore, a total of 505 clinical isolates were identified in seven genera of the family, of which 451 belong to the genus *Burkholderia*, 24 from *Pandoraea*, 14 from *Ralstonia*, five from *Cupriavidus*, four from *Caballeronia*, and one isolate present in *Lautropia* and *Paraburkholderia*.

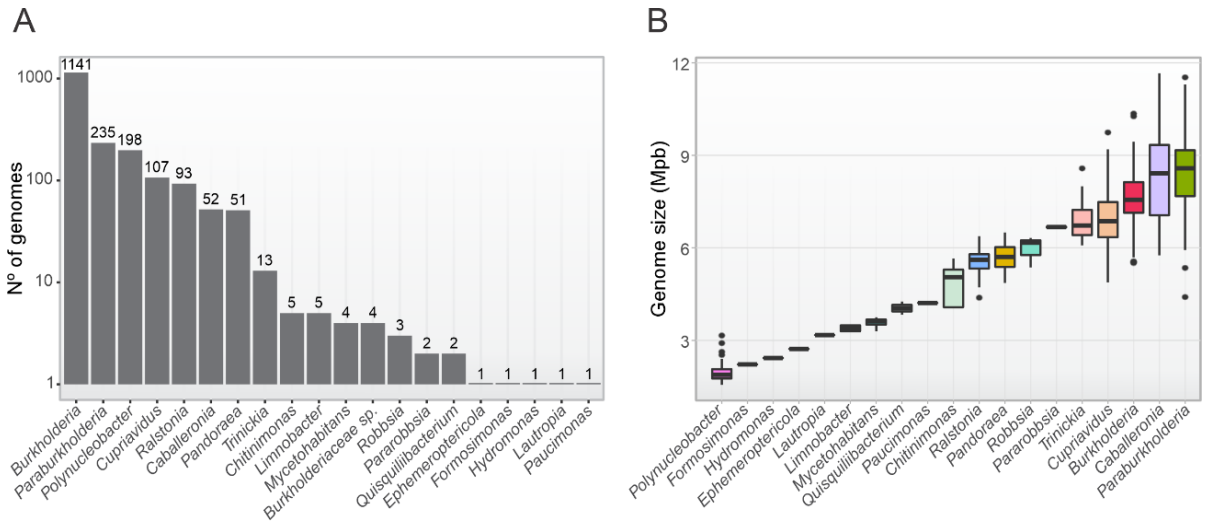


Figure 4-2. Overview of *Burkholderiaceae* genomic distribution. (A) Number of genomes used in this study classified by genus. (B) Distribution of median genome sizes of Burkholderiaceae genomes.

4.3.2. Plant growth promotion properties

Several species belonging to the *Burkholderiaceae* family is known for their ability to promote plant growth. As a way to find genera with greater PGP potential, we searched for the main PGP genes in *Burkholderiaceae* isolates. Additionally, we also report genes involved in the root nodulation. The main genes related to these properties are described in the following sections.

4.3.2.1. Ethylene biocontrol

We found the ACC deaminase gene (*acd*S) in all genomes of *Caballeronia*, *Ralstonia*, and *Trinickia*, as well as in *Burkholderia* (96.5% of the genomes),

Paraburkholderia (85%) and *Cupriavidus* (56%), including in clinical isolates and phytopathogenic species (Figure 4-3). No ACC deaminase genes were found in *Pandoraea* and *Polynucleobacter.* ACC deaminase cleaves ACC, an ethylene precursor, reducing the ethylene levels and their effects on stressed plants (Checcucci *et al.*, 2017; Murali *et al.*, 2021). Although this gene has been previously reported in *Burkholderia*, here we show that it is broadly widely distributed in the family. The functions of AcdS in opportunistic pathogens remain unknown. However, in phytopathogenic species such as *Pseudomonas thivervalensis*, AcdS is considered a factor that counteracts plant responses to infection (Checcucci *et al.*, 2017; Nascimento *et al.*, 2021).

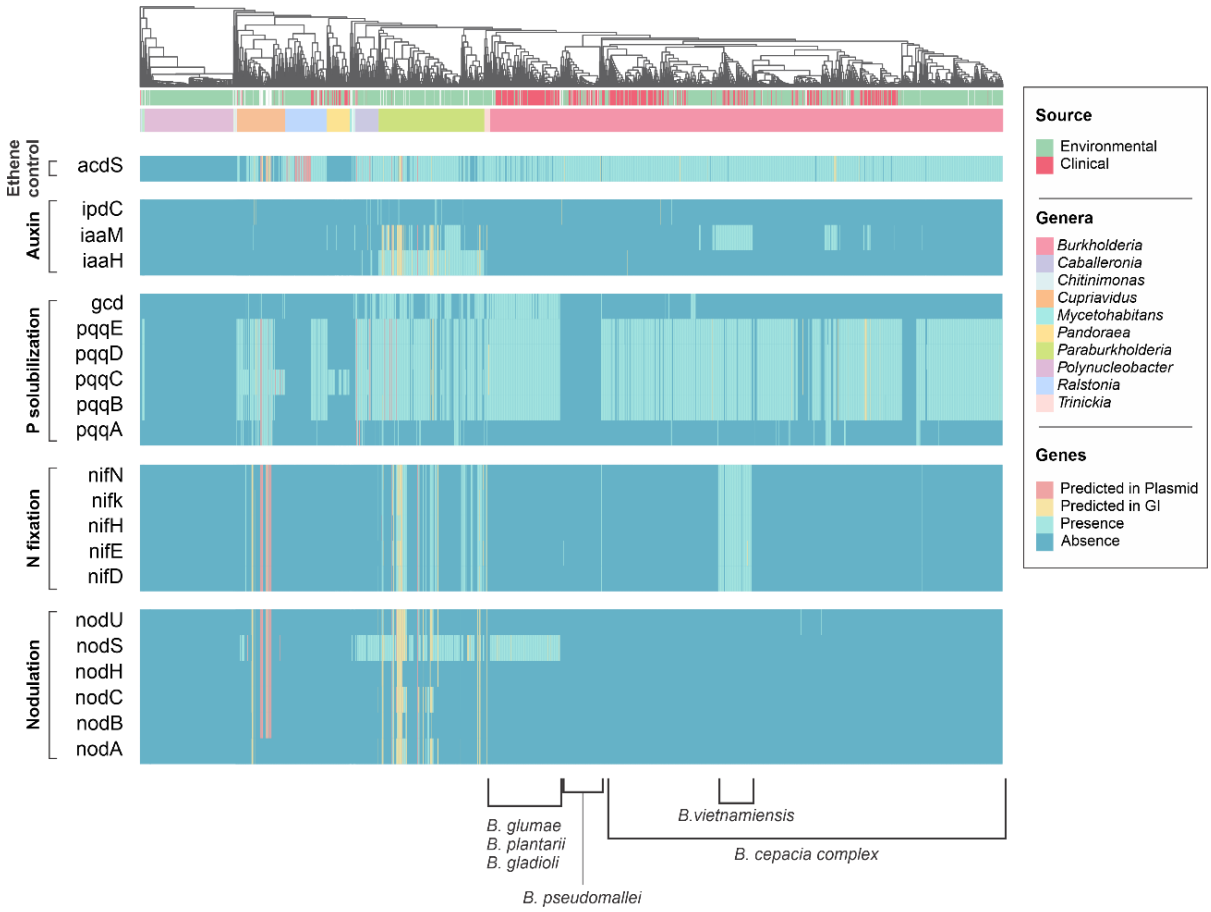


Figure 4-3. Plant growth promoting genes of *Burkholderiaceae*. The respective phylogenetic groups of *Burkholderiaceae* are highlighted in the tree. The heatmaps represent the presence or absence of the genes identified here (see methods for details).

Interestingly, only genera with secondary replicons (chromids, megaplasmids, and plasmids) possess AcdS. Considering that 95% of the identified *acd*S genes are located in regions of the genome without evidence of horizontal gene transfer (HGT), the intrinsic presence of *acd*S in *Burkholderia*, *Caballeronia*, *Cupriavidus*, *Paraburkholderia*, *Ralstonia*, and *Trinickia* can be involved in the capacity of multipartite genomes to accumulate genes involved in the adaptation of various niches (diCenzo, Mengoni, & Perrin, 2019). These characteristics have already been

observed in species with multipartite genomes such as *Sinorhizobium meliloti* (Ma, Charles, & Glick, 2004).

4.3.2.2. Auxin biosynthesis

Indole-3-acetic acid (IAA) is an auxin-class phytohormone and one of the most important plant growth regulators, known to be produced by several rhizobacteria, including symbionts (Mishra *et al.*, 2016). The bacterial IAA biosynthesis can occur by at least four tryptophan-dependent pathways, classified according to their intermediates: indole-3-pyruvic acid (IPyA), tryptamine (TRY), indole-3-acetamide (IAM) and indole-3-acetonitrile (IAN). In the IAM pathway, tryptophan is converted to IAA in two steps by the enzymes Tryptophan 2-monooxygenase (*iaa*M) and indole-3-acetamide hydrolase (*iaa*H) (Spaepen & Vanderleyden, 2011).

We identified the *iaa*M gene present in *Paraburkholderia* (44%), Trinickia (15%), *Caballeronia* (6%) and *Burkholderia* (11.3%), *Pandoraea* (4%) and *Cupriavidus* (1%). The *iaa*H gene was identified mainly in *Paraburkholderia* (99%), followed by *Trinickia* (15%) and *Caballeronia* (9.6%). The *ipd*C gene, involved in the IPyA pathway, was found in some species of *Paraburkholderia*, *Burkholderia*, and *Cupriavidus* (Figure 4-3).

A total of 103 (44%) *Paraburkholderia* genomes have *iaa*M and *iaa*H, being the only ones in the family with the complete IAM pathway. Most of these genes were found in GIs of symbiont species. These results indicate that *Paraburkholderia* could be a major contributor to IAA to plant growth compared to all members of the family.

4.3.2.3. Inorganic phosphate solubilization

Pyrroloquinoline quinone PQQ is a redox cofactor for several bacterial dehydrogenases, such as methanol, ethanol, and glucose dehydrogenases (Adachi *et al.*, 2007). When associated with glucose dehydrogenase (PQQ-GDH), PQQ acts on the periplasmic oxidation of glucose to gluconic acid (Stephen & Jisha, 2011; Vyas & Gulati, 2009). The soil acidification promoted by gluconic acid is essential for the solubilization of soil phosphate, which becomes available for plant uptake (Stella & Halimi, 2015; Wagh *et al.*, 2014).

We found the *pqq*A gene and the *pqq*BCDE operon in *Burkholderia*, *Caballeronia*, *Cupriavidus*, *Limnobacter*, *Paraburkholderia*, *Ralstonia*, and *Trinickia*, including clinical genomes (Figure 4-3). Interestingly, we identified the gene encoding PQQ-GDH (*gcd*) only in the genomes of *Paraburkholderia* (54%), *Trinickia* (39%), *Caballeronia* (37%), and some species of *Cupriavidus* (7%), and *Burkholderia* (14%) (Figure 4-3). The prevalence of this gene in *Paraburkholderia* supports the interaction of several of its species with plants.

Almost all (97%) genomes of the phytopathogenic species *B. plantarii*, *B. glumae*, and *B. gladioli* have *pqq* and *gcd* genes, including clinical isolates. Interestingly, *pqq* genes were not found in the human pathogen *B. pseudomallei*. In *Ralstonia*, *pqq* genes were absent in phytopathogenic isolates (e.g., *R.*

solanacearum). In contrast, we found the *pqq* genes in clinical *Ralstonia* isolates, although the PQQ-GDH was absent.

The presence of PQQ in *Burkholderiaceae* genomes without the enzyme PQQ-GDH may be related to the metabolic versatility of the isolates and the action of PQQ on different quinoprotein dehydrogenases, acting in processes of oxidative fermentation of substrates such as methane and ethanol, as well as in the protection against oxidative stress (Lisdat, 2020; Naveed *et al.*, 2016). The species *Ralstonia eutropha*, for example, has demonstrated the ability to degrade the xenobiotic compound tetrahydrofurfuryl alcohol through PQQ dependent alcohol dehydrogenase (Zarnt, Schräder, & Andreesen, 1997).

Collectively, these findings suggested that, *Cupriavidus*, *Trinickia*, and *Paraburkholderia* are the genera with the greatest potential for gluconic acid production in *Burkholderiaceae* and are, consequently, the most viable for studies involving the inorganic phosphate solubilization.

4.3.2.4. Biological nitrogen fixation and nodulation

Diazotrophic bacteria of the *Burkholderiaceae* family, especially *Cupriavidus* and *Paraburkholderia* species, are known as β-rhizobia due to their ability to develop symbiotic associations with plants. β-rhizobia genomes contain genes for nitrogen fixation (*nif*) and nodulation, mostly Nod factors (*nod*) (Checcucci *et al.*, 2019).

We identified the genes *nif*D, *nif*H, *nif*K, *nif*E, *nif*N in *Paraburkholderia* (34%), *Burkholderia* (6.4%), *Cupriavidus* (23.3%) and *Trinickia* (23%) (Figure 4-3). In *Burkholderia*, the *nif* genes are present mainly in the diazotrophic species *B. vietnamiensis* and *B. ubonensis*, which comprise the *B. cepacia* complex, and are also known as important opportunistic pathogens (Carvalho *et al.*, 2007).

In *Trinickia* and *Cupriavidus*, the presence of the *nif* genes is related to the presence of genes *nod*A, *nod*B, *nod*C, *nod*H *nod*S and *nod*U. However, in *Paraburkholderia*, not all genomes that have *nif* genes have *nod* genes, demonstrating a diversity of symbiotic and free-living species exclusively found in this genus (Bellés-Sancho *et al.*, 2021). Furthermore, *nif* and *nod* genes were also present in plasmids and GIs, supporting the prevalence of these essential genes for symbiosis in HGT-prone regions, as observed in other rhizobia species (Wang *et al.*, 2018).

4.3.3. Antimicrobial resistance profiles

We evaluated the composition of antibiotic resistance genes using the CARD database. To better understand the composition of antimicrobial resistance (AMR) genes in *Burkholderiaceae*, we focused on genera with more than 10 isolates, such as *Burkholderia*, *Paraburkholderia*, *Cupriavidus*, *Ralstonia*, *Caballeronia*, *Pandoraea*, *Trinickia*, and *Polynucleobacter*. We identified 30 different genes distributed in the resistome core (genes present in more than 80% of genomes) of each genus, of which 23 genes are involved in the efflux of antibiotics (Figure 4-4). No core AMR genes were found in the *Polynucleobacter*.

Five genes encoding efflux pumps (*ros*A, *mux*B, *mex*K, *kdp*E, *emr*B) and the polymyxin resistance gene (*arn*A) were present in all genera. Intrinsic resistance to polymyxin has been reported in *Burkholderia* (Loutet & Valvano, 2011). Our results show that polymyxin resistance seems to be more widespread across other *Burkholderiaceae* genera. Further, we also found other polymyxin resistance gene, *pmr*F, in the *Ralstonia* core resistome.

In addition, we identified a trimethoprim resistance gene (*dfr*A3) intrinsically present in *Ralstonia*, *Pandoraea*, and *Cupriavidus*. The *opm*38 gene, which encodes a bacterial porin involved in the reduction of permeability to β-lactams (Siritapetawee *et al.*, 2004), was found exclusively in the core resistome of *Burkholderia*, *Paraburkholderia*, *Trinickia, and Caballeronia*. We also identified the genes *bla*LRA-1, *amrA,* and *amrB* as part of the *Burkholderia* core resistome. *bla*LRA-1 encodes a β-lactamase that provides resistance to members of the penicillin structural class (Allen *et al.*, 2009), while *amr*AB encodes an efflux pump that mediates aminoglycoside resistance in clinical isolates (Jassem, Forbes, & Speert, 2014; Zulkefli *et al.*, 2021).

Figure 4-4. Distribution of core resistance genes among the most representative genera.

The accessory resistome (genes present up to 80% of genomes) of *Burkholderiaceae* comprises 266 different genes, of which 89% are distributed at a low frequency (up to 20% of genomes) in each genus, indicating that most of them are strain- or species-specific. Further, *Burkholderia* genomes tend to have more accessory AMR genes than its counterparts (Suppl. Figure S1). Further, we

investigated the frequency of each gene based on antibiotic classes that they confer resistance against across the genera (Figure 4-5). *Burkholderia* and *Cupriavidus* show greater prevalence of AMR genes involved in fluoroquinolones, while *Ralstonia* and *Caballeronia* have a greater number of genes in the resistance against β-lactams. Other genera showed a more diffuse distribution of all classes of antibiotics with a frequency lower than 40%. Further, genes encoding resistance for other antibiotic classes such as streptogramin (*vat*F), pleuromutilin (*lmr*C), and sulfonamide (*sul*) were nearly absent in *Burkholderiaceae.*

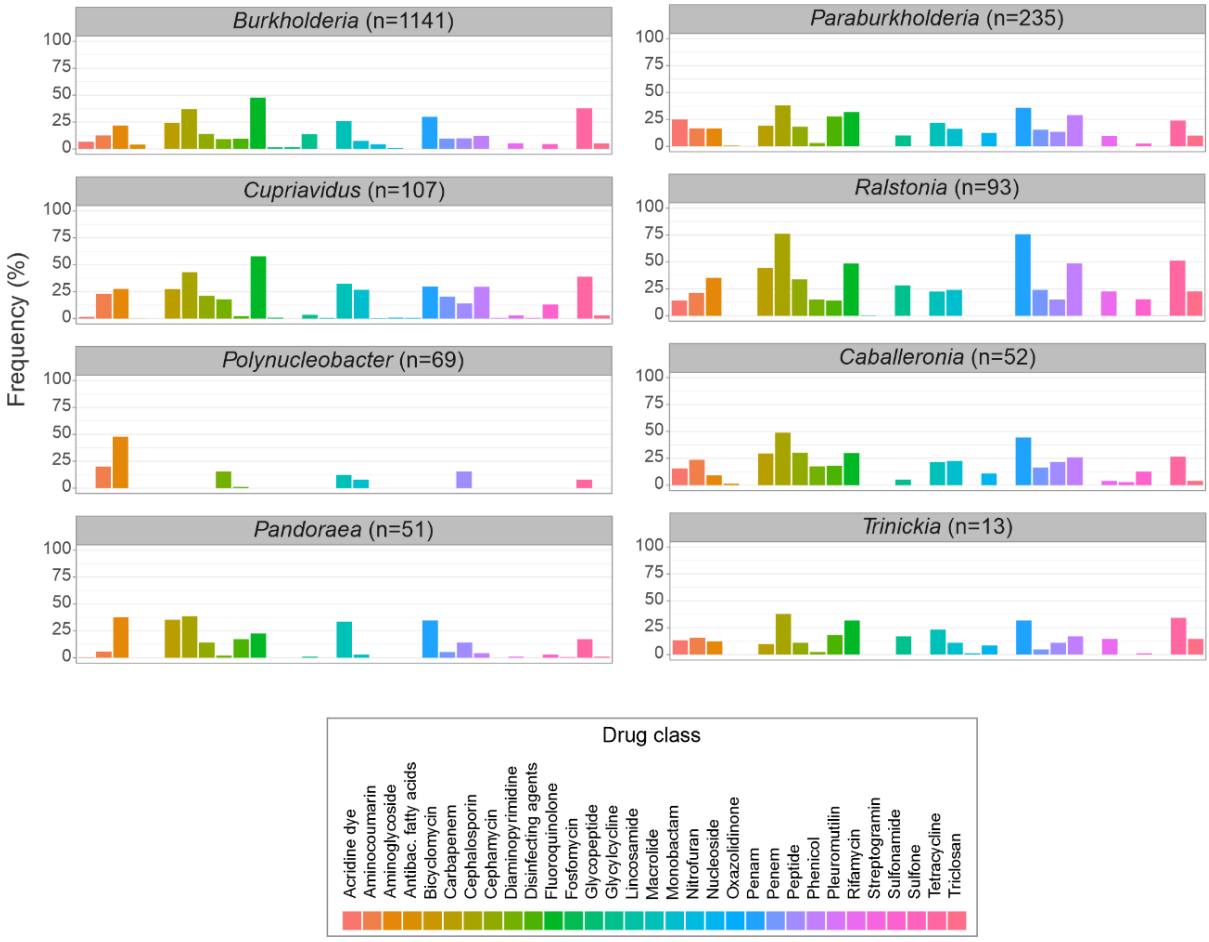


Figure 4-5. Relative acquired resistome per genus. In each box, we show the percentage of genes associated with the resistance of each class of antibiotics.

Within accessory resistome, β-lactam hydrolyzing enzymes is one of the major AMR mechanisms in *Burkholderiaceae*. β-Lactamases are structurally grouped into four classes: A, B, C, and D. Within these classes, each β-lactamase possesses a unique spectrum of activity toward penicillins, cephalosporins, monobactams, and carbapenems (Bush & Jacoby, 2010).

The distribution of Ambler class A, C and D β-lactamase genes among the genera is shown in the Figure 4-6. The genomes of *Ralstonia* and *Pandoraea* harbor different class D β-lactamases (OXA) genes (e.g. *bla*OXA-60, *bla*OXA-22, *bla*OXA-151, and *bla*OXA-158), which confer resistance against carbapenem and penicillin, constituting an intrinsic core resistance mechanism of these genera (Girlich, Naas, &

Nordmann, 2004; Schneider & Bauernfeind, 2015). In addition, *bla*OXA genes were also identified in species of the *Burkholderia cepacia* complex and in all genomes of *B. pseudomallei*.

Figure 4-6. Ambler classes β-lactamases identified in *Burkholderiaceae*. The respective genera of *Burkholderiaceae* are highlighted in the tree. The heatmaps represent the presence or absence of the genes identified here (see methods for details).

Among the class C β-lactamases, we identified the *bla*IDC genes present exclusively in *Burkholderia*, while the *bla*ACC genes showed a higher prevalence in the genomes of *Paraburkholderia* and *Caballeronia*. Genes encoding class C β-lactamases *bla*PDC were found in 49% of the *Cupriavidus* genomes. Among Amber class A, the gene for β-lactamase *bla*Y56 is present in *Burkholderia* genomes, especially in the pathogenic species, *B. glumae*, *B. plantarii*, and *B. gladioli*. The *bla*CPX-M genes were identified mainly in the pathogen *B. cenocepacia*. Finally, *bla*LRA genes, which are also part of the *Burkholderia* core resistome, were found with a low frequency in *Paraburkholderia* and *Caballeronia*.

Resistance to β-lactams in clinical isolates of the family including *Pandoraea*, *Ralstonia*, *B. pseudomallei*, *B. gladioli,* and belonging to the *Burkholderia cecapia* complex has been increasingly reported (Dlewati *et al.*, 2021; Fang *et al.*, 2019; Schneider & Bauernfeind, 2015). Additionally, *Burkholderia* isolates were identified as resistant to fluoroquinolones and tetracyclines (Somprasong *et al.*, 2021; Zhang, Li, & Poole, 2001), a finding that is also supported by our results.

Therefore, the low frequency of genes encoding resistance for streptogramin, pleuromutilin, and sulfonamide in our analysis suggests that these antibiotic classes could be tested as potential therapeutic agents against multiresistant infections caused by pathogenic species, such as *B. cepacia*, *B. vietnamiensis*, and *B. pseudomallei*. Furthermore, identification of β-lactamase Ambler among the pathogenic groups can help to determine the most effective treatment through synergistic combinations of β-lactams with different classes of antibiotics.

4.3.4. Virulence profiles

We used the VFDB database to assess the profiles of virulence genes in Burkholderiaceae. *Burkholderia* had the greatest average number of genes per genome (124) (Suppl. Figure S2). Genera with more than 10 genomes share common genes in the core virulome, mainly those related to motility (e.g. *fli*A, *fli*M, *fli*P, *flg*C, *flg*I, *flg*G, *cheB, che*Y, *and che*W), adherence (*tufA* and *htp*B), oxidative stress tolerance (*sod*B, *kat*A, *kat*G), and protein secretion (e.g. *gsp*E, *gsp*F, *gsp*G, *clp*V, *tss*B, *tss*C) (Suppl. Figure S3).

Secretion systems are potent weapons that confer an advantage for competitiveness, symbiosis, and pathogenicity (Chou *et al.*, 2022). Several pathogenic species of the family are known to have multiple secretion systems (Lennings, West, & Schwarz, 2019). We investigated the distribution of genes involved in the secretion system in *Burkholderiaceae* genomes. We found in all genera, except *Polynucleobacter* and *Pandoraea*, the operon *gsp*DEFGIK encoding the type 2 secretion system (T2SS). The type 3 (T3SS) and type 6 (T6SS) secretion systems were also widely distributed in *Burkholderiaceae*, with T3SS (*sct*) and T6SS-3 (*tss*) being the most common types among genera, including *Paraburkholderia, Pandoraea*, and *Trinickia*.

Burkholderia showed the highest prevalence of secretion systems (Figure 4-7). Among the species, *B. pseudomallei* and *B. mallei* stand out for the almost exclusive presence of T3SS-1, T3SS-2 (*hrp*), T3SS-3 (*bap*, *bid*, *bsa*, *bip*, *bop*, *bpr*), and T6SS-5 (*tss* and *tag*), which are crucial virulence factors for infection and clinical manifestation (Lennings, West, & Schwarz, 2019; Vander Broek & Stevens, 2017). In addition, phytopathogenic species of *Ralstonia* (e.g. *Ralstonia solanacearum*) showed exclusive gene groups related to T2SS (*gsp*G, *gsp*L, *gsp*N, *gsp*M, *gsp*H, and *gsp*J) and T3SS-1 (*hrp*), which are one of the main factors reported for pathogenicity.

Secretion systems are key factors for interactions with hosts. Many proteins (toxins or effector proteins) secreted by pathogens or symbionts have the ability to trigger defense responses or manipulate host cell structure and physiology, supporting bacterial colonization, nutrition, and proliferation (Torto-Alalibo *et al.*, 2009). T3SS and T6SS have already been reported to be important in plant-bacterial interactions (Tseng, Tyler, & Setubal, 2009), which is in line with their presence in *Burkholderiaceae* isolates associated with plants. However, subtypes of T3SS such as T3SS-1, T3SS-2, T3SS-3 have been directly associated with different pathogenic species from other families, including *Vibrio mimicus* (Okada *et al.*, 2010) and *Salmonella enterica* (Takemura *et al.*, 2021). Therefore, the presence of multiple secretion systems in *Burkholderiaceae* genomes and their association with pathogenesis warrant further in-depth investigations.

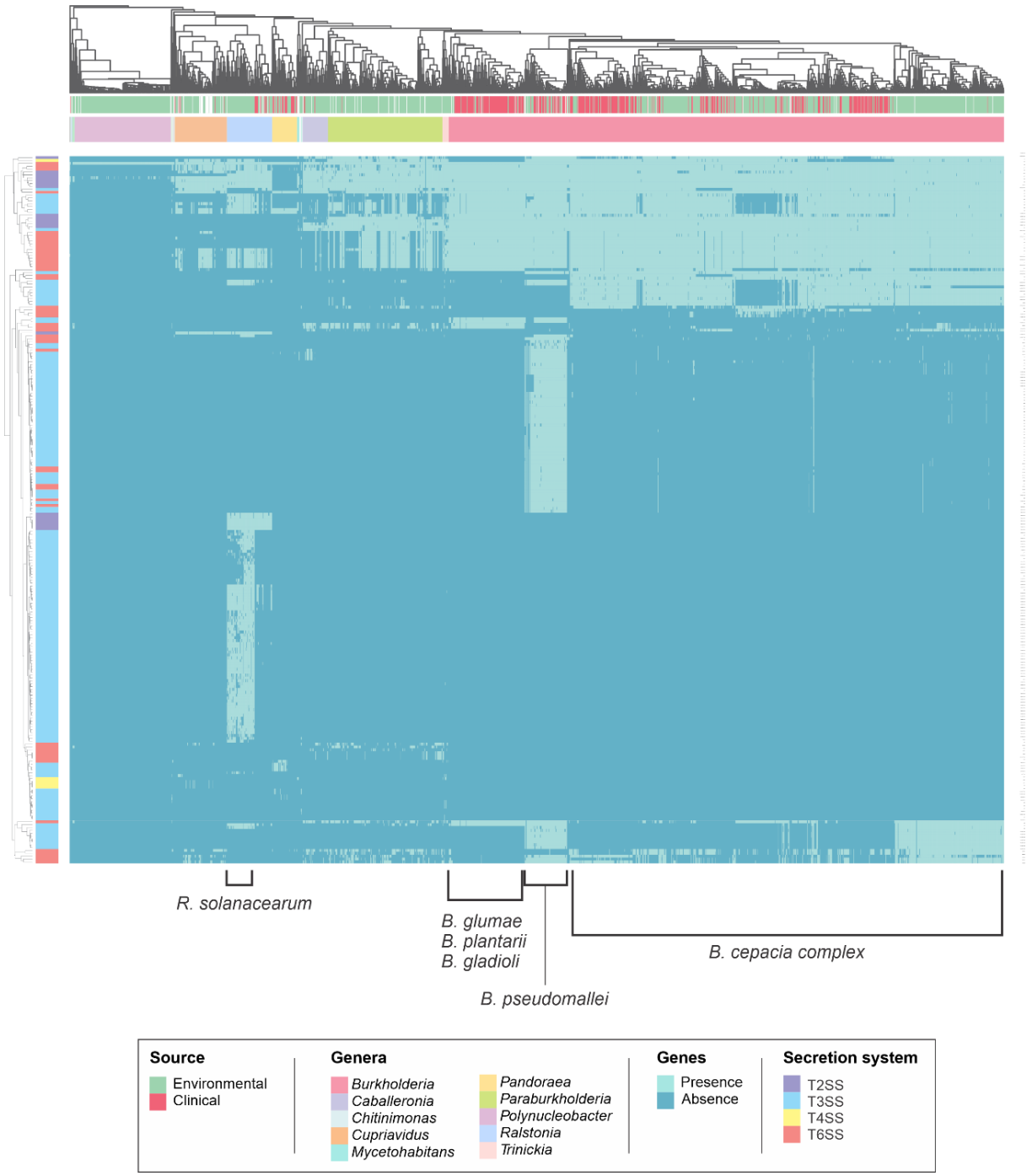


Figure 4-7. Secretion systems identified in *Burkholderiaceae*. The respective genera of *Burkholderiaceae* are highlighted in the tree. The heatmaps represent the presence or absence of genes related with secretion system.

4.4. CONCLUSION

In this work, we explored the first time the genomic diversity of the *Burkholderiaceae* family. Through a large-scale genomic analysis, we identified and reclassified 61 genus-level genomes that are currently misclassified. The phylogenomic structure of *Burkholderiaceae* includes at least 25 genera analyzed using ANI and MASH. We analyzed genes of biotechnological and clinical interest.

Paraburkholderia and *Trinickia* are the genera that show the greatest potential for bioinoculant research because they have a larger gene repertoire, including nitrogen fixation, nodulation, and especially AIA biosynthesis. In addition, these genera showed little or no relationship to clinical isolates. In terms of genes of clinical interest, *Burkholderia* exhibited unique intrinsic genes conferring resistance to β-lactams and aminoglycosides. Furthermore, we identified distinct profiles of acquired resistance between the genera of the family, which might be useful to guide the choice of antibiotics that could be used to treat infections. The genera of *Burkholderiaceae* presented a virulome with different types of secretion systems, particularly in human and plant pathogens. Together, our results provide insights into the genomic diversity of *Burkholderiaceae*, which has implications for safer selection of strains for biotechnology applications and potential treatment of human infections.

ACKNOWLEDGEMENTS

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; grants E-26/203.309/2016 and E-26/203.014/2018), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001), and Conselho Nacional de Desenvolvimento Científico e Tecnológico. The funding agencies had no role in the design of the study and collection, analysis, and interpretation of data and in writing.

AUTHOR CONTRIBUTIONS

Conceived the study: FP-S, TMV; Funding and resources: TMV; Data analysis: FP-S; Interpretation of the results: FP-S, TMV; Wrote the manuscript: FP-S, TMV.

4.5. REFERENCES

Adachi, O., et al. (2007). Biooxidation with PQQ-and FAD-Dependent Dehydrogenases. *Modern biooxidation: enzymes, reactions and applications*, 1-41.

Allen, H. K., et al. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *The ISME journal, 3*(2), 243-251.

Bellés-Sancho, P., et al. (2021). Paraburkholderia phymatum homocitrate synthase NifV plays a key role for nitrogenase activity during symbiosis with papilionoids and in free-living growth conditions. *Cells, 10*(4), 952.

Bernier, S. P., et al. (2003). Comparative analysis of plant and animal models for characterization of Burkholderia cepacia virulence. *Infection and immunity, 71*(9), 5306-5313.

Bertelli, C., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research, 45*(W1), W30-W35.

Boscaro, V., et al. (2013). Polynucleobacter necessarius, a model for genome reduction in both free-living and symbiotic bacteria. *Proceedings of the National Academy of Sciences, 110*(46), 18590-18595.

Bush, K., Jacoby, G. A. (2010). Updated functional classification of β-lactamases. *Antimicrobial agents and chemotherapy, 54*(3), 969-976.

Butler, J., et al. (2022). Hospital sink traps as a potential source of the emerging multidrug-resistant pathogen Cupriavidus pauculus: characterization and draft genome sequence of strain MF1. *Journal of medical microbiology, 71*(2), 001501.

Carvalho, A. P. D. A., et al. (2007). Burkholderia cenocepacia, B. multivorans, B. ambifaria and B. vietnamiensis isolates from cystic fibrosis patients have different profiles of exoenzyme production. *Apmis, 115*(4), 311-318.

Checcucci, A., et al. (2017). Role and regulation of ACC deaminase gene in Sinorhizobium meliloti: is it a symbiotic, rhizospheric or endophytic gene? *Frontiers in genetics, 8*, 6.

Checcucci, A., et al. (2019). Genomic diversity and evolution of rhizobia. In *Microbial diversity in the genomic era* (pp. 37-46): Elsevier.

Chou, L., et al. (2022). Modular evolution of secretion systems and virulence plasmids in a bacterial species complex. *BMC biology, 20*(1), 1-20.

Coenye, T., Vandamme, P. (2003). Diversity and significance of Burkholderia species occupying diverse ecological niches. *Environmental Microbiology, 5*(9), 719-729.

Dall'Agnol, R. F., et al. (2017). Genetic diversity of symbiotic Paraburkholderia species isolated from nodules of Mimosa pudica (L.) and Phaseolus vulgaris (L.) grown in soils of the Brazilian Atlantic Forest (Mata Atlântica). *FEMS Microbiology Ecology, 93*(4), fix027.

diCenzo, G. C.,Mengoni, A., Perrin, E. (2019). Chromids aid genome expansion and functional diversification in the family Burkholderiaceae. *Molecular biology and evolution, 36*(3), 562-574.

Dlewati, M. M., et al. (2021). Meropenem-Resistant Pandoraea Pneumonia in a Critically Ill Patient With COVID-19. *Cureus, 13*(11).

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

Estrada-De Los Santos, P., et al. (2016). To split or not to split: an opinion on dividing the genus Burkholderia. *Annals of Microbiology, 66*(3), 1303-1314.

Fang, Q., et al. (2019). Nosocomial bloodstream infection and the emerging carbapenem-resistant pathogen Ralstonia insidiosa. *BMC infectious diseases, 19*(1), 1-9.

Fluit, A. C., et al. (2021). Characterization of clinical Ralstonia strains and their taxonomic position. *Antonie Van Leeuwenhoek, 114*(10), 1721-1733.

Girlich, D.,Naas, T., Nordmann, P. (2004). OXA-60, a chromosomal, inducible, and imipenem-hydrolyzing class D β-lactamase from Ralstonia pickettii. *Antimicrobial agents and chemotherapy, 48*(11), 4217-4225.

Goris, J., et al. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology, 57*(1), 81-91.

Howe, K.,Bateman, A., Durbin, R. (2002). QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics, 18*(11), 1546-1547.

Jain, C., et al. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications, 9*(1), 1-8.

Jassem, A. N.,Forbes, C. M., Speert, D. P. (2014). Investigation of aminoglycoside resistance inducing conditions and a putative AmrAB-OprM efflux system in Burkholderia vietnamiensis. *Annals of Clinical Microbiology and Antimicrobials, 13*(1), 1-5.

Jeon, J.-S., et al. (2021). Impact of root-associated strains of three Paraburkholderia species on primary and secondary metabolism of Brassica oleracea. *Scientific reports, 11*(1), 1-14.

Lee, H.-H., et al. (2016). Understanding the direction of evolution in Burkholderia glumae through comparative genomics. *Current genetics, 62*(1), 115-123.

Lennings, J.,West, T. E., Schwarz, S. (2019). The Burkholderia type VI secretion system 5: composition, regulation and role in virulence. *Frontiers in microbiology, 9*, 3339.

Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*.

Lisdat, F. (2020). PQQ-GDH–Structure, function and application in bioelectrochemistry. *Bioelectrochemistry, 134*, 107496.

Liu, M., et al. (2022). Plant growth-promoting rhizobacteria Burkholderia vietnamiensis B418 inhibits root-knot nematode on watermelon by modifying the rhizosphere microbial community. *Scientific reports, 12*(1), 1-13.

Loutet, S. A., Valvano, M. A. (2011). Extreme antimicrobial peptide and polymyxin B resistance in the genus Burkholderia. *Frontiers in cellular and infection microbiology, 1*, 6.

Ma, W.,Charles, T. C., Glick, B. R. (2004). Expression of an exogenous 1-aminocyclopropane-1-carboxylate deaminase gene in Sinorhizobium meliloti increases its ability to nodulate alfalfa. *Applied and environmental microbiology, 70*(10), 5891-5897.

Mahenthiralingam, E.,Baldwin, A., Dowson, C. (2008). Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *Journal of applied microbiology, 104*(6), 1539-1551.

Mishra, P., et al. (2016). Characterization of an indole-3-acetamide hydrolase from Alcaligenes faecalis subsp. parafaecalis and its application in efficient preparation of both enantiomers of chiral building block 2, 3-dihydro-1, 4-benzodioxin-2-carboxylic acid. *Plos one, 11*(7), e0159009.

Murali, M., et al. (2021). Bio-prospecting of ACC deaminase producing Rhizobacteria towards sustainable agriculture: A special emphasis on abiotic stress in plants. *Applied soil ecology, 168*, 104142.

Nascimento, F. X., et al. (2021). Genomic analysis of the 1-aminocyclopropane-1-carboxylate deaminase-producing Pseudomonas thivervalensis sc5 reveals its multifaceted roles in soil and in beneficial interactions with plants. *Frontiers in microbiology, 12*.

Naveed, M., et al. (2016). The life history of pyrroloquinoline quinone (PQQ): a versatile molecule with novel impacts on living systems. *Int J Mol Biol, 1*, 1-20.

Okada, N., et al. (2010). Presence of genes for type III secretion system 2 in Vibrio mimicus strains. *BMC microbiology, 10*(1), 1-10.

Ondov, B. D., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology, 17*(1), 132. doi:10.1186/s13059-016-0997-x

Ortega, L., Rojas, C. M. (2021). Bacterial Panicle Blight and Burkholderia glumae: From pathogen biology to disease control. *Phytopathology®, 111*(5), 772-778.

Parks, D. H., et al. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature biotechnology, 38*(9), 1079-1086.

Parks, D. H., et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research, 25*(7), 1043-1055.

Peeters, C., et al. (2019). Comparative genomics of Pandoraea, a genus enriched in xenobiotic biodegradation and metabolism. *Frontiers in microbiology, 10*, 2556.

Pereira, S., et al. (2020). Plant growth-promoting rhizobacteria (PGPR) improve the growth and nutrient use efficiency in maize (Zea mays L.) under water deficit conditions. *Heliyon, 6*(10), e05106.

Pradier, L., et al. (2021). PlasForest: a homology-based random forest classifier for plasmid detection in genomic datasets. *BMC bioinformatics, 22*(1), 1-17.

Rahman, M., et al. (2018). Plant probiotic bacteria Bacillus and Paraburkholderia improve growth, yield and content of antioxidants in strawberry fruit. *Scientific reports, 8*(1), 1-11.

Rivard, C., et al. (2012). Grafting tomato to manage bacterial wilt caused by Ralstonia solanacearum in the southeastern United States. *Plant disease, 96*(7), 973-978.

Rotz, L. D., et al. (2002). Public health assessment of potential biological terrorism agents. *Emerging infectious diseases, 8*(2), 225.

Sawana, A.,Adeolu, M., Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov. harboring environmental species. *Frontiers in genetics, 5*, 429.

Schneider, I., Bauernfeind, A. (2015). Intrinsic carbapenem-hydrolyzing oxacillinases from members of the genus Pandoraea. *Antimicrobial agents and chemotherapy, 59*(11), 7136-7141.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069.

Siritapetawee, J., et al. (2004). Expression and refolding of Omp38 from Burkholderia pseudomallei and Burkholderia thailandensis, and its function as a diffusion porin. *Biochemical Journal, 384*(3), 609-617.

Somprasong, N., et al. (2021). Burkholderia ubonensis high-level tetracycline resistance is due to efflux pump synergy involving a novel TetA (64) resistance determinant. *Antimicrobial agents and chemotherapy, 65*(3), e01767-01720.

South, K. A.,Nordstedt, N. P., Jones, M. L. (2021). Identification of plant growth promoting rhizobacteria that improve the performance of greenhouse-grown petunias under low fertility conditions. *Plants, 10*(7), 1410.

Spaepen, S., Vanderleyden, J. (2011). Auxin and plant-microbe interactions. *Cold Spring Harbor perspectives in biology, 3*(4), a001438.

Stella, M., Halimi, M. (2015). Gluconic acid production by bacteria to liberate phosphorus from insoluble phosphate complexes. *J Trop Agric Food Sci, 43*(1), 41-53.

Stephen, J., Jisha, M. (2011). Gluconic acid production as the principal mechanism of mineral phosphate solubilization by Burkholderia sp.(MTCC 8369). *Journal of Tropical agriculture, 49*, 99-103.

Takemura, M., et al. (2021). A Salmonella type III effector, PipA, works in a different manner than the PipA family effectors GogA and GtgA. *Plos one, 16*(3), e0248975.

Tavares, M., et al. (2020). Burkholderia cepacia complex bacteria: a feared contamination risk in water-based pharmaceutical products. *Clin Microbiol Rev, 33*(3), e00139-00119.

Torto-Alalibo, T., et al. (2009). Common and contrasting themes in host cell-targeted effectors from bacterial, fungal, oomycete and nematode plant symbionts described using the Gene Ontology. *BMC microbiology, 9*(1), 1-8.

Tseng, T.-T.,Tyler, B. M., Setubal, J. C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC microbiology, 9*(1), 1-9.

Vander Broek, C. W., Stevens, J. M. (2017). Type III secretion in the melioidosis pathogen Burkholderia pseudomallei. *Frontiers in cellular and infection microbiology, 7*, 255.

Vyas, P., Gulati, A. (2009). Organic acid production in vitro and plant growth promotion in maize under controlled environment by phosphate-solubilizing fluorescent Pseudomonas. *BMC microbiology, 9*(1), 1-15.

Wagh, J., et al. (2014). Heterologous expression of pyrroloquinoline quinone (pqq) gene cluster confers mineral phosphate solubilization ability to Herbaspirillum seropedicae Z67. *Applied Microbiology and Biotechnology, 98*(11), 5117-5129.

Wang, X., et al. (2018). Comparative symbiotic plasmid analysis indicates that symbiosis gene ancestor type affects plasmid genetic evolution. *Letters in applied microbiology, 67*(1), 22-31.

Zarnt, G.,Schräder, T., Andreesen, J. R. (1997). Degradation of tetrahydrofurfuryl alcohol by Ralstonia eutropha is initiated by an inducible pyrroloquinoline quinone-dependent alcohol dehydrogenase. *Applied and environmental microbiology, 63*(12), 4891-4898.

Zhang, L.,Li, X.-Z., Poole, K. (2001). Fluoroquinolone susceptibilities of efflux-mediated multidrug-resistant Pseudomonas aeruginosa, Stenotrophomonas maltophilia and Burkholderia cepacia. *Journal of Antimicrobial Chemotherapy, 48*(4), 549-552.

Zulkefli, N. J., et al. (2021). Genomic comparison and phenotypic profiling of small colony variants of Burkholderia pseudomallei. *Plos one, 16*(12), e0261382.

5. CONCLUSÕES GERAIS

Nesse trabalho, realizamos análises comparativas para avaliar a distinção de genomas PGPR e associadas com infecções oportunistas em três conjuntos de dados que se referem à diferentes níveis taxonômicos, sendo a espécie *S. maltophilia*, o gênero *Alcaligenes* e a família *Burkholderiaceae*. Para cada um dos grupos genômicos analisados, utilizamos abordagens diferentes de acordo com o poder de resolução que cada ferramenta oferece para investigação de diferentes grupos taxonômicos.

Em *S. maltophilia* traçamos um paralelo consistente entre o genoma *S. maltophilia* UENF-4GII e outros genomas disponíveis publicamente e aprofundamos a filogenia de isolados da espécie, resultando em um bom grau de diferenciação entre isolados clínicos e não-clínicos através da identificação da prevalência de genes envolvidos na resistência antimicrobiana e grupos gênicos distintos (e.g. *treS*) nos genomas clínicos.

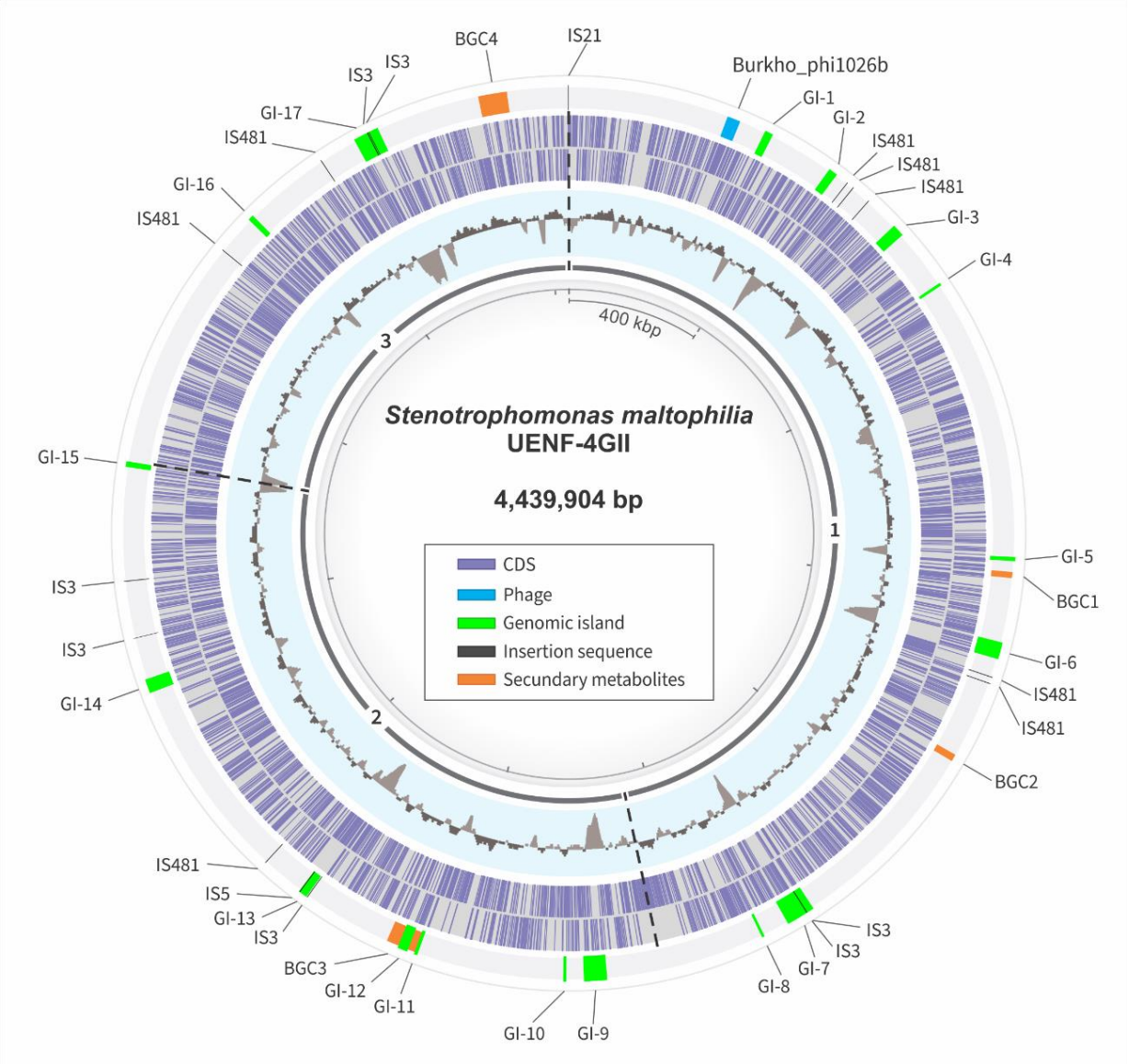
Em *Alcaligenes,* as principais diferenças entre genomas PGPR e clínicos consiste na composição e quantidade dos genes de resistência antimicrobiana adquiridos por HGT. Além disso, identificamos espécies com maior potencial biotecnológico e identificamos espécies contendo genes exclusivos relacionados com o estilo de vida.

Em *Burkholderiaceae*, identificamos os gêneros que possuem espécies com maior potencial biotecnológico. *Paraburkholderia* foi o único gênero que apresentou capacidade para produção de compostos indólicos. Além disso, identificamos o gênero *Burkholderia* com o maior repertório gênico sistemas de secreção e resistência a diferentes classes de antibióticos, especialmente para beta-lactâmicos e aminoglicosídeos, ao qual descobrimos que possui resistência intrínseca.

Coletivamente, nossas análises suportam resultados positivos para prospecção de PGPR e possíveis patógenos em diferentes níveis taxonômicos, além de uma compreensão sobre os aspectos dinâmicos e evolutivos dos organismos abordados. Essas investigações abrem perspectivas para a prospecção e utilização de rizobactérias de forma mais segura, mesmo que estas apresentem relações taxonômicas com isolados patogênicos. No entanto, experimentos complementares além da genômica devem ser sempre considerados antes da liberação de novos produtos.

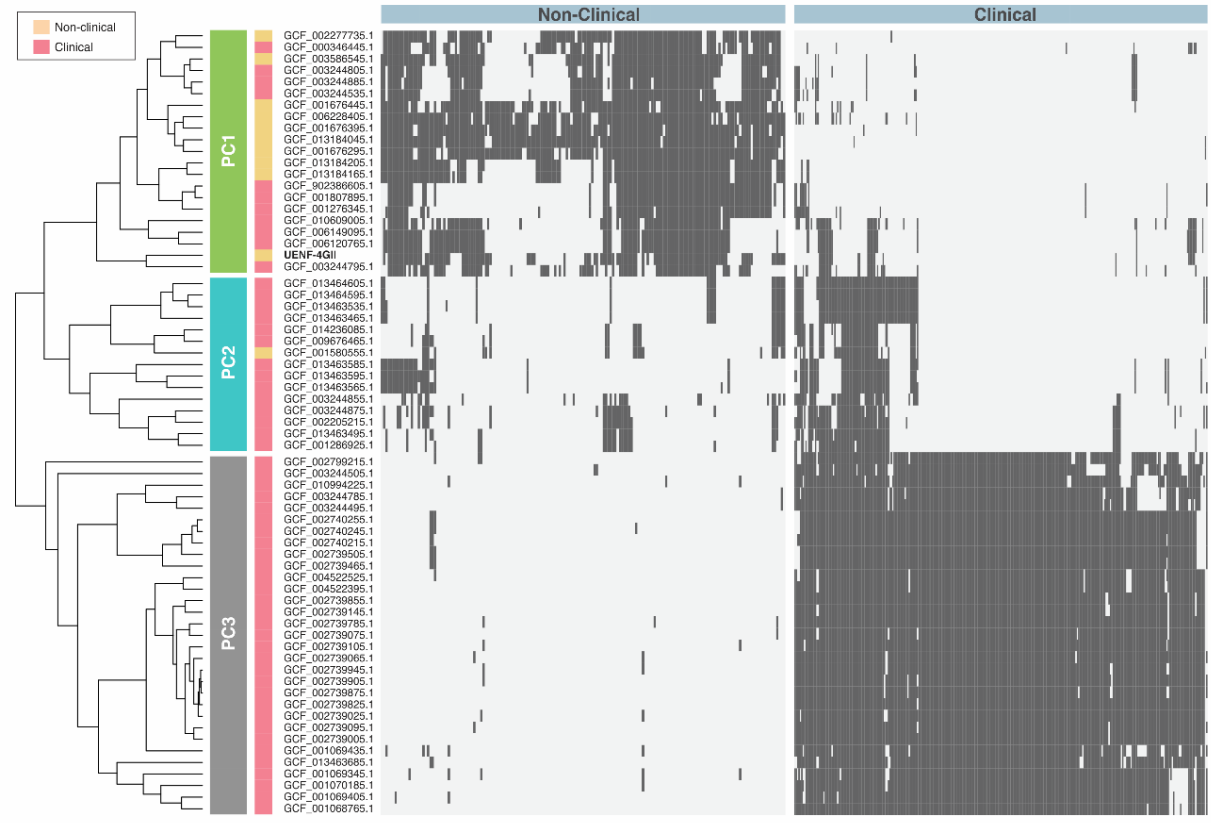
6. APÊNDICES

6.1. APÊNDICE A – Figuras suplementares do capítulo 1.

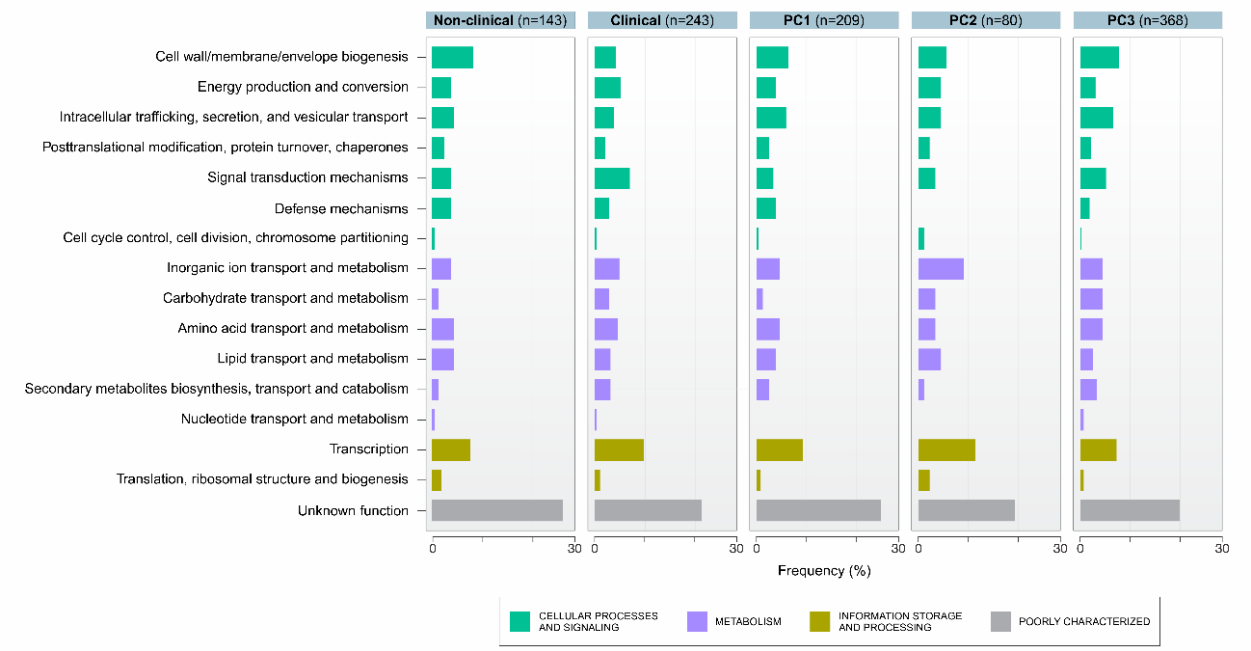


Supplementary figure S1. Circular representation of *S. maltophilia* UENF-4GII genome. From inner to outer: the first ring represents the scaffold number. The second ring represents GC content; light-grey indicates GC content higher than average; dark-grey indicates less than average. The third circle ring represents predicted CDS on the plus and minus strands, respectively. The fourth ring represents genome features such as phage, genomic islands, insertion sequence and secondary metabolites.

Supplementary figure S2. Genetic context of *S. maltophilia* UENF-4GII composite transposon. The IR represents two perfect 26-bp inverted repeat sequence. The transposon contains a *pbp*X gene winch encodes a penicillin-binding protein associated with penicillin resistance; two different transposases-encoding genes; and two unknown function genes.

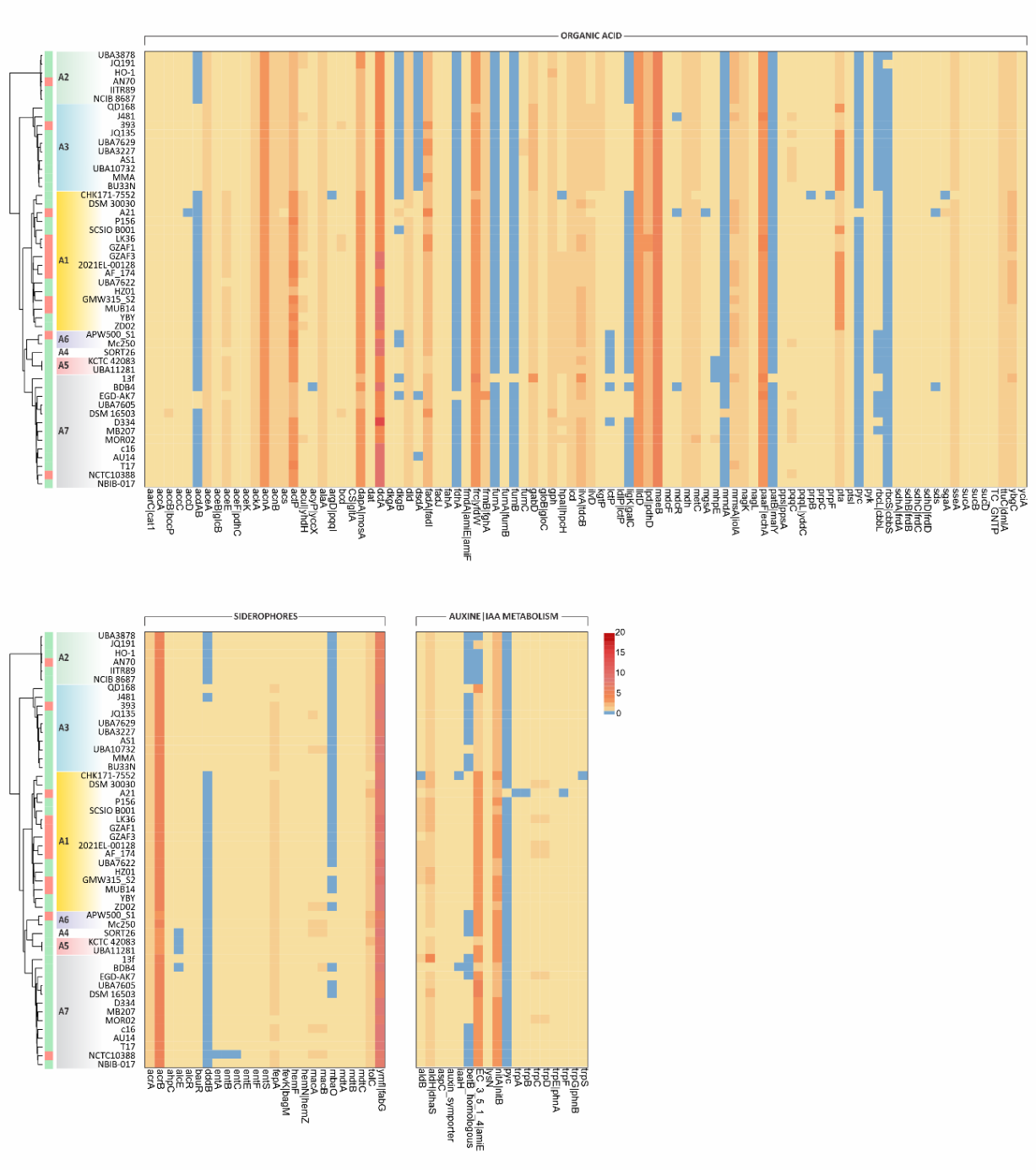


Supplementary figure S3. Distribution of environmental-associated genes in Sm3 genogroup. The cgMLST tree is annotated with two colored strips representing the phylogenetic cluster (PC) and the environment source. The heatmaps represent the presence (dark-gray) or absence (light-gray) of the genes identified by the pan-GWAS pipeline using the environment source as trait.

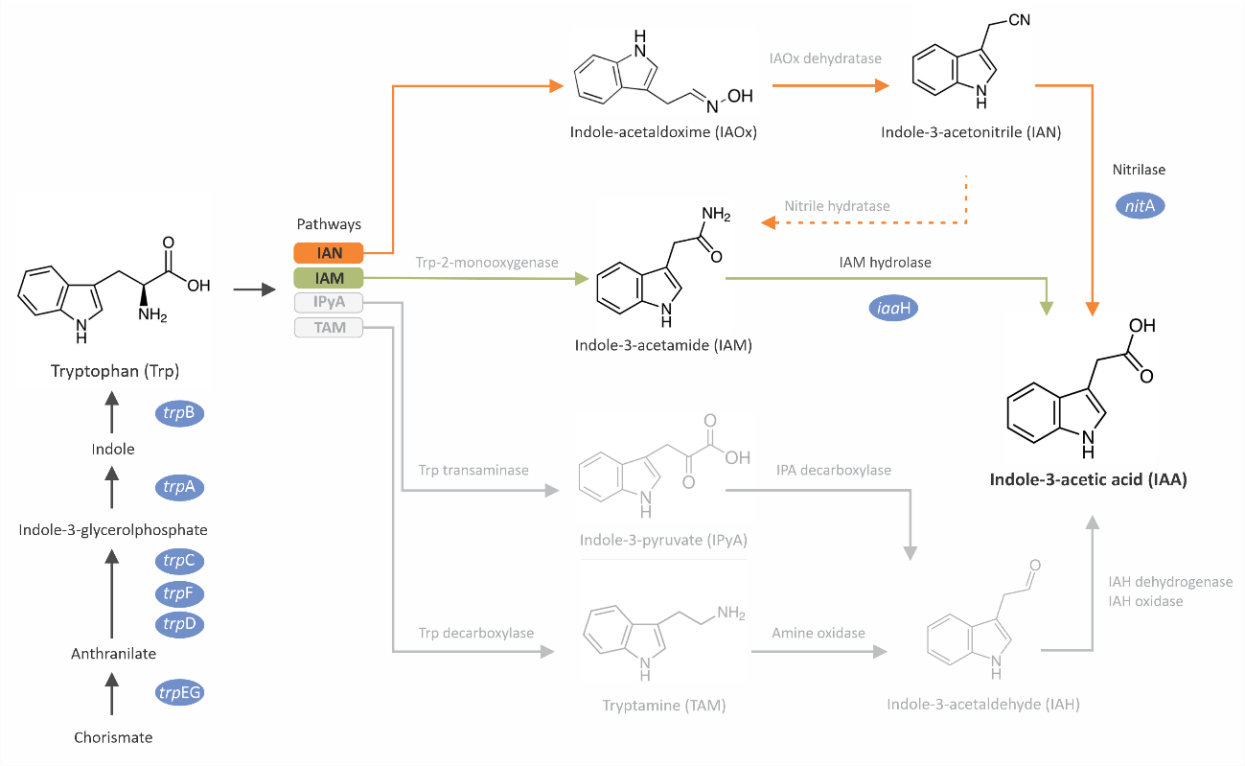


Supplementary figure S4. Functional categories identified by EggNOG annotation of environment and PC-associated genes.

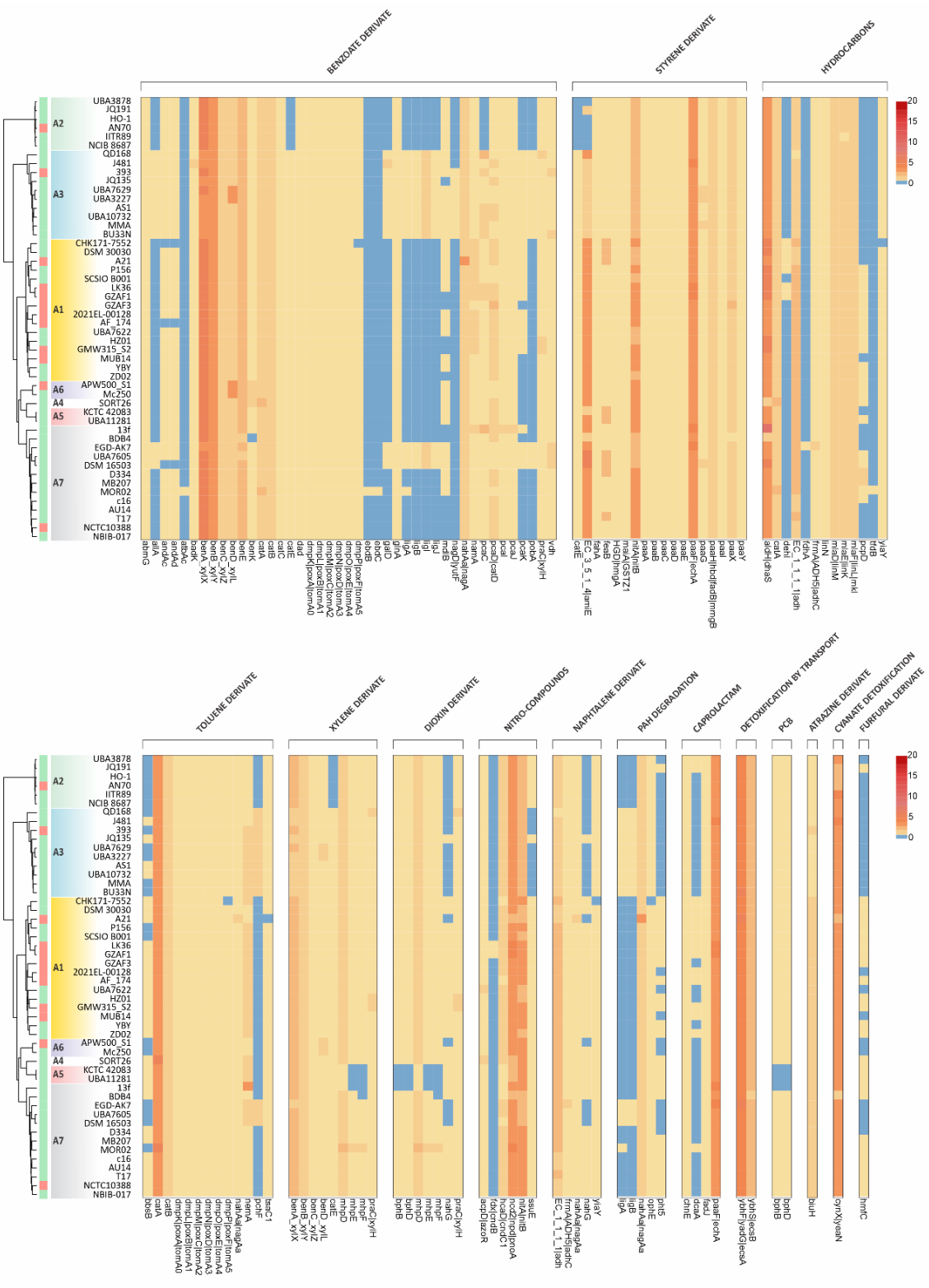
6.2. APÊNDICE B – Figuras suplementares do capítulo 2.



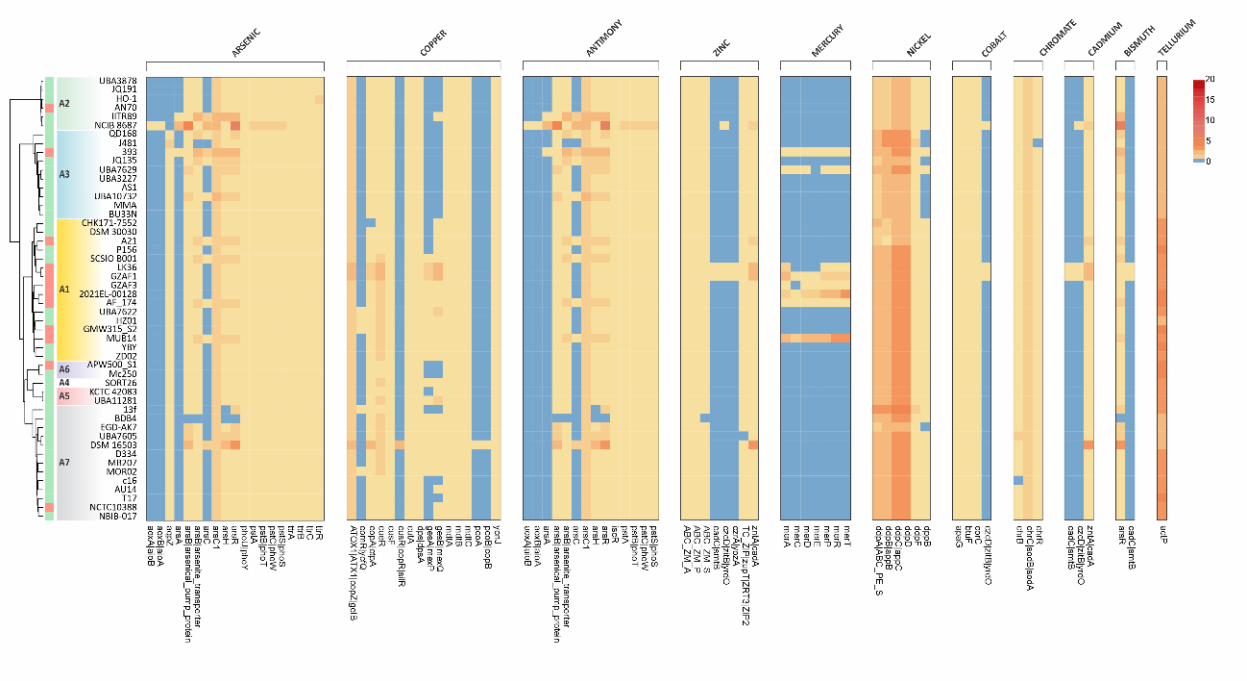
Supplementary figure S1. Distribution of Biofertilization genes in *Alcaligenes*. The respective phylogenetic groups of *Alcaligenes* are highlighted in the tree. The heatmaps show the number of genes found for each genome based on PlaBAse annotations.



Supplementary figure S2. Deduced pathways for IAA biosynthesis in *Alcaligenes*. The tryptophan (Trp) biosynthetic pathway is shown on the left. Orange and green colors indicate the principal pathways for IAA biosynthesis, and light-gray colors the pathways that are not found; see Supplementary table S1 for details). The names of the pathways indicate the name of their first products. Genes identified are highlighted in blue circles.

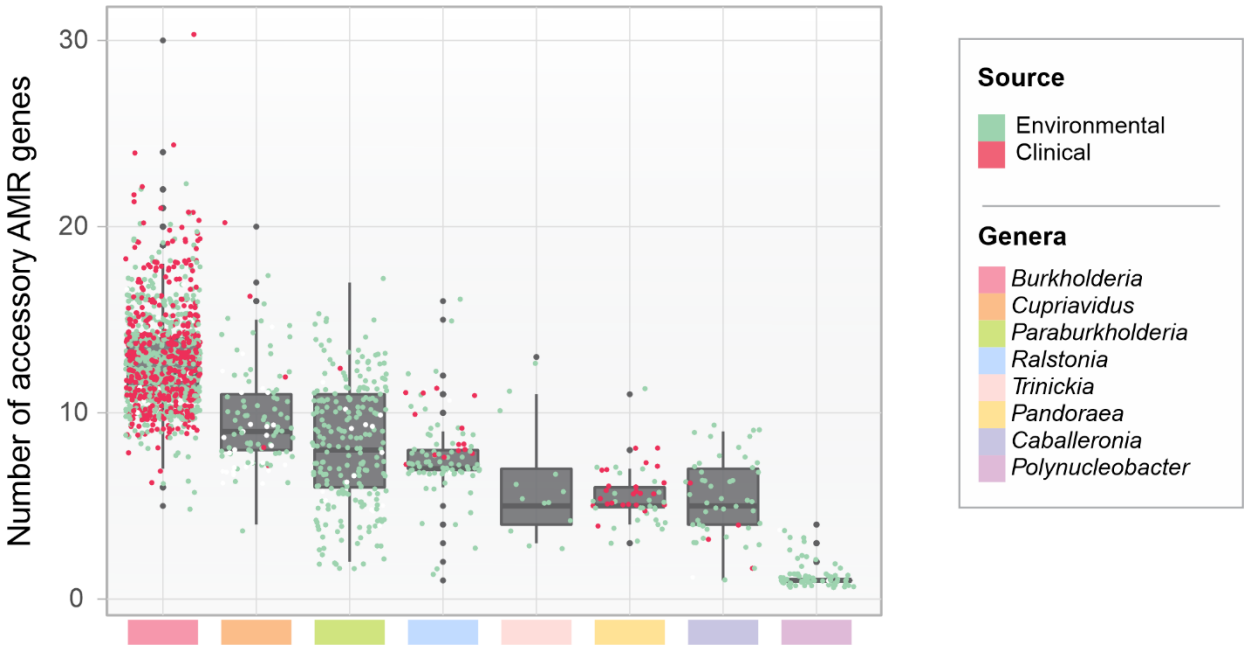


Supplementary figure S3. Distribution of xenobiotic degradation genes in *Alcaligenes*. The respective phylogenetic groups of *Alcaligenes* are highlighted in the tree. The heatmaps show the number of genes found for each genome based on PlaBAse annotations.

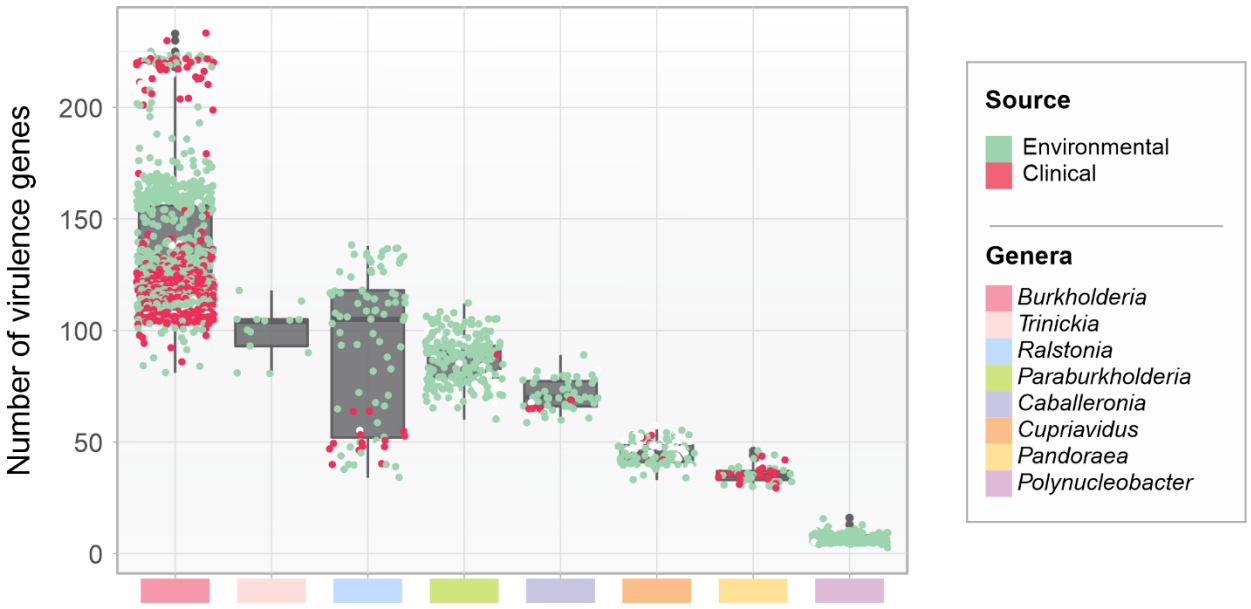


Supplementary figure S4. Distribution of heavy-metal resistance genes in *Alcaligenes*. The respective phylogenetic groups of *Alcaligenes* are highlighted in the tree. The heatmaps show the number of genes found for each genome based on PlaBAse annotations.

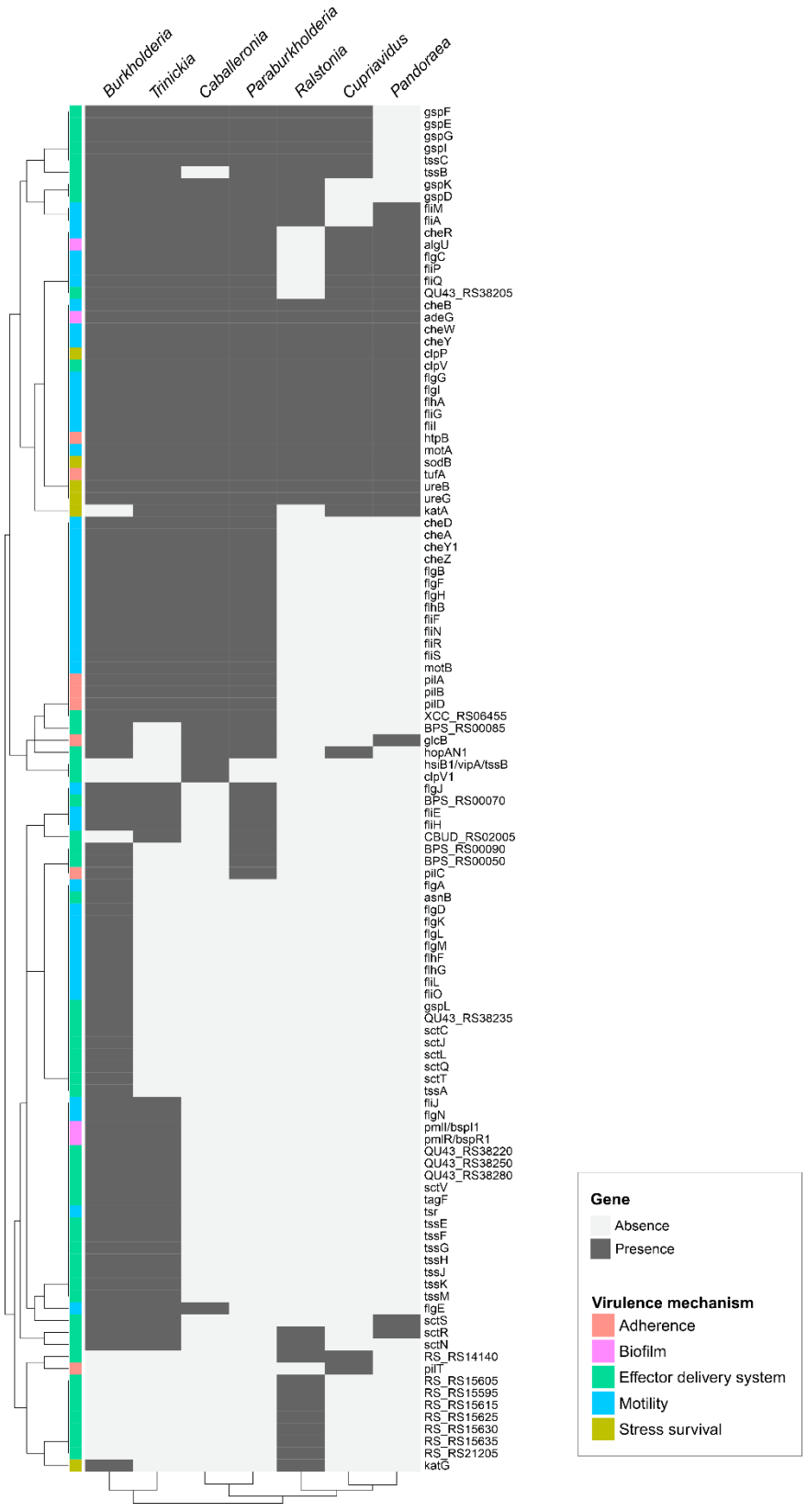
6.3. APÊNDICE C – Figuras suplementares do capítulo 3.



Supplementary figure S1. Number of accessory resistance genes by genus.



Supplementary figure S2. Total number of virulence genes by genus.



Supplementary figure S3. Distribution of core virulence genes among the most representative genera.