

https://dspace.org/

Pan-genome Analysis of Pantoea agglomerans Strains to Investigate their Ecological and Functional Diversity

Savrimuthu, Charét Arianna

2025-05

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/48284

Savrimuthu, Charét Arianna. (2025). Pan-genome Analysis of Pantoea agglomerans Strains to Investigate their Ecological and Functional Diversity. [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/48284

Downloaded from a DSpace repository. https://dspace.org/



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Pan-genome Analysis of *Pantoea agglomerans* Strains to Investigate their Ecological and Functional Diversity

By

Charét Arianna Savrimuthu

(1839966)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

In

Molecular and Cell Biology

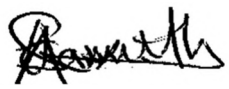
In the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Professor Pieter de Maayer

May 2025

Declaration

I, Charét Savrimuthu, declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science (Dissertation) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Signed on the fourteenth (14th) day of May 2025 at 10:00.

Preface

Pantoea agglomerans is a Gram-negative, facultatively anaerobic, rod-shaped bacterial species in the genus *Pantoea* and the family *Erwiniaceae*. Strains of this species have been isolated from a broad range of ecological niches globally and are often found in association with both plant and animal hosts. Individual strains demonstrate potential for the use in a range of biotechnological applications, including bioremediation, biological control and the production of biopharmaceuticals and biofertilizers. However, the commercial use of this species has been hampered by its implication in opportunistic human infections and plant pathogenicity. In this study, a comprehensive comparative genomic analysis was undertaken to elucidate the ecological and functional diversification of *P. agglomerans*, with the aim to dissociate between pathogenic strains and strains of biotechnological value.

Chapter one encompasses a comprehensive review of literature on the biology of *P. agglomerans*. This covers the taxonomic history of *P. agglomerans*, isolationist ecological diversity as well as what is known about its pathogenesis to both plant and animal hosts and its biotechnological potential.

Chapter two outlines the comparative genomic methodologies employed in this study. This involved the analysis of *P. agglomerans* from the pan-genome perspective of the species, through to the genomes of individual strains and finally to the minutiae of its pathogenesis, environmental resistance and secondary metabolite production.

The results of this study, and their interpretations, are documented in chapter three. This revealed that *P. agglomerans* delineates into two subspecies (*P. agglomerans* subsp. 1 and subsp. 2) and has an open pan-genome. This open pan-genome can be attributed to the expansive and variable mobilome of *P. agglomerans*, comprised of plasmid elements, bacteriophages, integrative and conjugative elements (ICEs) and transposable elements that confer various phenotypes underpinning the ecological success of this species. *P. agglomerans* harbours multiple secretion systems, which facilitates interactions with varying hosts and environmental niches. The secretome of *P. agglomerans* displayed resistance to a wide range of antimicrobials, antibiotics and environmental pollutants. *P. agglomerans* produces an array of secondary metabolites furthering its biotechnological potential.

Chapter four provides a succinct summary of the key findings in this study and future perspectives. The comprehensive comparative genomic study of this fascinating species adds towards the body of knowledge of its ecological and lifestyle diversification, contributing towards the end goals of a deeper understanding of its pathogenesis towards a broad range of hosts and driving its biotechnological application.

In memory of my mother

Rochelle Antonette Joseph Savrimuthu

30 November 1973 – 25 December 2020

Acknowledgements

Firstly, I thank God for His protection, mercy, favour and healing over my life. Due to a corneal ulcer, I lost the vision in my left eye in September 2023, and it is only through Him that my sight was restored. I would also like to express my sincere gratitude to my supervisor, Prof. Pieter de Maayer, for his constant support and guidance. Thank you for your patience and understanding and for the knowledge and experience I have gained over the years.

My dear Wits friends (Rea, Dylan, Kiara, Precious, Tereka and Aaron), thank you for all the laughs and quiz nights. Rea, Dylan and Kiara, thank you for everything. Your support, concern, friendship and love are answered prayers. To my Rivers Church (Kyalami) friends, thank you for the prayers, weekly check-ins, laughs and memories. Wednesday night rehearsals (even when I am not serving) have become my favourite mid-week break!

Thank you to my family that have believed in me, supported me and encouraged me throughout this journey. To my aunt, Bernadette (Gaynell), I am so grateful for you, our bond and everything that you are. Thank you for answering every phone call, holding me accountable, reminding me to stay focused and to prioritise my work. You are my aunt, my friend and sometimes the other half of my brain! I love doing life with you. Thank you to my cousin and his wife, Christopher and Justine. You guys have been instrumental in my faith journey and have supported me ever since I moved to Johannesburg. Thank you for taking me to church with you (which led to me joining Rivers), for encouraging me to partner and start serving and to become more prayerful, faithful and trusting in His plan. Thank you for supporting me through every health struggle and always being one phone call away. I am blessed to have you in my life and as role models.

Lastly, thank you to my dad and brother (Lael). I would not have been able to complete this degree if it were not for you. Dad, thank you for believing in me, supporting me, for all your sacrifices and unconditional love. You have never pressurised me, despite my own concerns, and allowed me to study for the last 7.5 years. Thank you for allowing me to adopt Tunzi. Thank you for answering my 10 phone calls a day, even if it was just for me to tell you that I am tired or if it suddenly started to rain. You understand more than anyone how I struggle with my eyes, and you encouraged (or insisted) that I take breaks and was a voice of reason when I would push myself past my limits. Lael, thank you for stepping up when I needed you to. Thank you for every meal, thing to drink, running around and looking after Tunzi. Thank you for supporting me throughout this degree, especially in the last two months, and listening to everything that I needed to discuss even though you had no idea what I was doing.

Thank you.

Table of Contents

Preface 3
Dedication 4
Acknowledgements 5
Table of Contents 6
List of Figures 8
List of Tables 9
List of Abbreviations and Symbols 10
Chapter One: *Pantoea agglomerans*, a versatile bacterium of all sorts and sources 13
1.1 *Pantoea agglomerans* 14
1.1.1 Taxonomic history of *Pantoea agglomerans* 14
1.1.2 *Pantoea agglomerans* in the context of the genus *Pantoea* 15
1.1.3 *P. agglomerans* in the environment 18
1.1.4 *P. agglomerans* in association with plants 18
1.1.5 Associations of *P. agglomerans* strains with animal hosts 20
1.1.6 Biotechnological potential of *P. agglomerans* and known determinants 22
1.1.7 Hurdles to the biotechnological applicability of *P. agglomerans* 24
1.1.8 Lessons from the genome of *Pantoea agglomerans* 24
1.2 Problem Statement 26
1.3 Aims and Objectives 27
1.3.1 Aims 27
1.3.2 Objectives 27
Chapter Two: Methodology 28
2.1 *P. agglomerans* genomes 29
2.2 Structural annotation and genome completeness analyses 29
2.3 Orthologous group identification 30
2.4 Phylogenomic analyses 30
2.5 Pan-genome analyses 31
2.6 Identification of molecular drivers of diversification 32
2.7 Classification of protein representatives into functional categories 33
2.8 Insight into the pathogenic potential of *P. agglomerans* 33
2.9 Prediction of *P. agglomerans* resistome 34
2.10 Antimicrobial production in *P. agglomerans* 34
Chapter Three: Results and Discussion 35
3.1 *Pantoea agglomerans* metadata 36
3.2 Phylogenomic analyses support the delineation of two potential *P. agglomerans* subspecies 38
3.3 *P. agglomerans* displays an open pan-genome 44

3.4 Pan-genome analysis of *P. agglomerans* subsp. 1 and subsp. 2 47
3.5 Characterisation of the *P. agglomerans* mobilome – drivers of diversification 50
3.5.1 Plasmids in *P. agglomerans* 50
3.5.2 Integrated bacteriophage elements 54
3.5.3 Integrative and Conjugative Elements 56
3.5.4 Transposable elements 59
3.6 Global insights into the functional diversity of *P. agglomerans* 60
3.7 Genomic insights into the pathogenicity (pathogenome) of *P. agglomerans* 65
3.7.1 Human pathogenicity potential of *P. agglomerans* 65
3.7.2 Phytopathogenicity of *P. agglomerans* 66
3.7.3 Secretion systems in *P. agglomerans* 68
3.7.4 Genomic analysis of factors underlying environmental resistance 73
3.8 Secondary metabolite production in *P. agglomerans* 78
Chapter Four: Summary 82
Reference List 85
Appendices 109
Appendix A 110
Appendix B 112
Appendix C 114
Appendix D 117
Appendix E 151
Appendix F 153
Appendix G 179

List of Figures

Figure 1.1: Taxonomy of *Pantoea agglomerans* 17

Figure 1.2: Depiction of the direct and indirect PGP mechanisms of *P. agglomerans* 19

Figure 3.1: Core genome phylogeny of *P. agglomerans* with outgroup taxa. 40

Figure 3.2: Core genome phylogeny of 100 *P. agglomerans* strains. 41

Figure 3.3: Neighbour-joining tree constructed on the basis of pairwise ANI values for *P. agglomerans.* 42

Figure 3.4: Neighbour-joining tree constructed on the basis of pairwise dDDH values for *P. agglomerans.* 43

Figure 3.5: The distribution of orthogroups in the pan-genome fractions of *P. agglomerans*. 45

Figure 3.6: Pan- and core genome graph of the 100 *P. agglomerans* strains. 46

Figure 3.7: Representation of the pan-genome fractions across the two *P. agglomerans* subspecies. 47

Figure 3.8: Pan- and core genome graph of *P. agglomerans* subsp. 1 and subsp. 2. 49

Figure 3.9: The distribution of plasmids, according to their size categories, in *P. agglomerans*. 51

Figure 3.10: Distribution of prophage elements identified in the *P. agglomerans* genomes. 56

Figure 3.11: Representative proportions of proteins according to the COG supra-functional categories in *P. agglomerans* pan-genome fractions. 61

Figure 3.12: Conserved and distinct functions of the proteome of *P. agglomerans.* 63

Figure 3.13: Distinct functions encoded by both *P. agglomerans* subsp. 1 and subsp. 2. 64

Figure 3.14: Relative proportions of proteins in different subcellular localisations in the *P. agglomerans* cell. 68

Figure 3.15: Relative numbers of the distinct types of secretion systems in *P. agglomerans* subsp. 1 and subsp. 2. 73

Figure 3.16: Distribution of antimicrobial and antibiotic resistance genes present in the 100 *P. agglomerans* genomes, across both predicted subspecies. 74

List of Tables

Table 3.1: Summary of isolation sources of the 100 compared *P. agglomerans* strains. 36

Table 3.2: Summary of the genome metrics of the 100 *P. agglomerans* strains. 37

Table 3.3: *P. agglomerans* pan-genome extrapolations in accordance with Heap's law. 46

Table 3.4: Pan-genome extrapolations in accordance with Heap's Law for *P. agglomerans* subsp. 1 and subsp. 2. 48

Table 3.5: Overall contribution of plasmid DNA to *P. agglomerans* genomes. 51

Table 3.6: Plasmid categories isolated from the 100 *P. agglomerans* genomes. 52

Table 3.7: Overall contribution of prophages to *P. agglomerans* genomes. 55

Table 3.8: Individual prophages in *P. agglomerans* genomes. 56

Table 3.9: Overview of ICEs characteristics in *P. agglomerans*. 57

Table 3.10: Predicted transposable elements in the *P. agglomerans* genomes. 59

Table 3.11: Probability of *P. agglomerans* strains being a human pathogen. 65

Table 3.12: Presence of known pathogenicity determinants in *P. agglomerans*. 66

Table 3.13: The relative proportions of subcellular localisations in the proteome of *P. agglomerans*. 69

Table 3.14: Antimicrobial resistance genes in *P. agglomerans*. 74

Table 3.15: Heavy metal resistance genes in *P. agglomerans*. 76

Table 3.16: Presence of secondary metabolites in *P. agglomerans* genomes. 78

List of Abbreviations and Symbols

%: Percent

*: Multiplied by

/: Out of

~: Approximately

<: Less than

=: Equals to

>: Greater Than

±: Approximately

≥: Greater than or equal to

°C: Degrees Celsius

ABC: ATP Binding Cassette

Acetyl-CoA: Acetyl Coenzyme A

AGA: D-alanylgriseoluteic Acid

AHLs: Acyl-Homoserine Lactones

AMR: Antimicrobial Resistance

ANI: Average Nucleotide Identity

antiSMASH: Antibiotics and Secondary Metabolite Analysis Shell

APH: Aminoglycoside Phosphotransferases

ATP: Adenosine Triphosphate

Bacmet: Antibacterial Biocide and Metal Resistance Genes

Blast: Basic Local Alignment Search Tool

BlastN: Nucleotide Search

Blastp: Protein Blast

BPGA: Bacteria Pan-Genome Analysis

BSL: BioSafety Level

BUSCO: Benchmarking Using Single-Copy Orthologues

COGs: Clusters of Orthologous Groups

dDDH: digital DNA-DNA Hybridization

DNA: Deoxyribonucleic Acid

E. amylovora: *Erwinia amylovora*

E. coli: Escherichia coli

e.g.: Example

EggNOG - Evolutionary Gene Genealogy Non-supervised Orthologous Groups

EPS: Exopolysaccharide

et al: Et alia (and others)

FastANI: Fast alignment-free computation of whole-genome average nucleotide identity

G + C: Guanine and Cytosine

GGDC: Genome-to-Genome Distance Calculator

GTDB: Genome Taxonomy Database

GTDB-tk: Genome Taxonomy Database Toolkit

HGT: Horizontal Gene Transfer

HSPs: High-scoring Segment Pieces

IAA: Indole Acetic Acid

ICEs: Integrative Conjugative Elements

IgA: Immunoglobulin A

INP: Ice Nucleating Particles

kb: Kilobases

LPP: Large *Pantoea* Plasmids

LPS: Lipopolysaccharide Layer

MacSyFinder: Macromolecular System Finder

Mb: Megabases

MeDuSa -Multi-Draft Based Scaffolder

MFP: Membrane Fusion Protein

ML: Maximum Likelihood

MPP: Medium *Pantoea* Plasmids

Multi-CSAR: Multiple reference-based Contig Scaffolder using Algebraic Rearrangements

NCBI: National Centre for Biotechnology Information

NI: Non-Ribosomal Peptide Synthesis Independent

No.: Number

Nr: non-redundant

NRPS: Non-Ribosomal Peptide Synthetase

OMP: Outer Membrane Protein

P. agglomerans: Pantoea agglomerans

P. ananatis: *Pantoea ananatis*

P. dispersa: Pantoea dispersa

P. eucalypti: Pantoea eucalypti

P. septica: Pantoea septica

P. stewartii: Pantoea stewartii

PanGP: Pan-genome Profile

PGP: Plant Growth Promotion

pH: Potential of Hydrogen

PHASTEST: Phage Search Tool with Enhanced Sequence Translation

PHI-Base: Pathogen-Host Interaction Database

PHYLIP: Phylogeny Inference Package

PNPs: *Pantoea* Natural Products

PQQ: Pyrroquinoline Quinone

PSORTb: Bacterial Protein Sorting

pv.: Pathovar (pathogenic variant)

QS: Quorum Sensing

ResFinder: Resource for identification of antimicrobial resistance genes

RiPPs: Ribosomally Synthesized and Post-translationally modified Peptides

RNA: Ribonucleic Acid

RND: Resistance-Nodulation-Division

RRE: Rev Response Element

S. enterica: Salmonella enterica

SCO: Single Copy Orthologue

SPP: Small *Pantoea* Plasmids

Subsp.: Subspecies

T1SSs: Type 1 Secretion Systems

T3Es: Type III Effectors

T3SS: Type III Secretion Systems

T4P: Type IV pili

T4SS: Type 4 Secretion System

T5SSs: Type V Secretion Systems

T6SSs: Type VI Secretion Systems

TnComp_finder: Transposon Finder

TpsB: Translocation Protein

UFboot: ultra-fast bootstrapping

v.: version

β: Beta

γ: Gamma

Chapter One:

Pantoea agglomerans, a versatile bacterium of all sorts and sources

1.1 *Pantoea agglomerans*

P. agglomerans is a Gram-negative, rod-shaped bacterium belonging to the genus *Pantoea*, family *Erwiniaceae* and order Enterobacterales (Adeolu *et al.*, 2016; Büyükcam *et al.*, 2018). Structurally, these strains are non-capsulated, have peritrichous flagella, are non-spore forming and are 0.5-0.1 by 1.0-3.0 μm in size (Büyükcam *et al.*, 2018; Dutkiewicz *et al.*, 2015; Manulis and Barash, 2003). The species name, *agglomerans* originates from the Latin word *agglomerare*, translating to "forming a ball" or "to mass together" (Dutkiewicz *et al.*, 2015; Palmer and Coutinho, 2022) . This nomenclature is linked to the ability of this species to form long aggregates that are surrounded by symplasmata, which contributes to its plant colonisation (Dutkiewicz *et al.*, 2015; Palmer and Coutinho, 2022). *P. agglomerans* strains produce yellow-pigmented colonies, are positive for acetoin and β-galactosidase production and do not produce indole (Palmer and Coutinho, 2022). This species is categorised as facultative anaerobic, catalase positive and oxidase negative (Mardaneh and Dallal, 2013; Penner *et al.*, 2022; Sulja *et al.*, 2022). *P. agglomerans* uses D-glucuronate and D-tartrate as its main carbon sources, but can also metabolise *cis*-aconitate, D-arabitol, citrate, glycerol, maltose, L-rhamnose and trehalose, with some strains able to use D-cellobiose, lactose and raffinose (Palmer and Coutinho, 2022).

Strains of this species have been isolated from many environments, including soil and water, and plant and animal hosts (Abbas et al., 2017). Primarily, *P. agglomerans* is known for its association with plants, either as an epiphyte or endophyte, but also a plant pathogen (Walterson and Stavrinides, 2015). Presented here are some of the pertinent literature on *P. agglomerans* and its ecological versatility.

1.1.1 Taxonomic history of *Pantoea agglomerans*

P. agglomerans was initially isolated from a plant tuber in 1888 as *Bacillus agglomerans* based on the bacterial shape, structure and metabolic activity observed from culturing (Beijerinck, 1888). Almost a century later, biochemical and physiological tests revealed that *Bacillus agglomerans* should be transferred to the genus *Enterobacter* (Ewing and Fife, 1972). This analysis tested carbohydrate fermentation, which is important in genus-level differentiation, acetoin production, which is indicative of *Enterobacter* species, nitrate reduction, growth characteristics, indole production, starch hydrolysis and a negative Gram reaction (Ewing and Fife, 1972). On this basis, *Bacillus agglomerans* was transferred to the genus *Enterobacter* and re-classified as *Enterobacter agglomerans* in 1972 (Ewing and Fife, 1972).

A subsequent study on strains of the species *Erwinia herbicola*, *Erwinia milletiae* and *Enterobacter agglomerans* showed that strains within these species belonged to one DNA hybridisation group, with

a relative binding ratio range of 72-100%, and shared an almost identical protein pattern and phenotype (Beji *et al.*, 1988). This high DNA relatedness and protein pattern similarity suggested that the separate species should rather all be subsumed into *Enterobacter agglomerans* (Beji *et al.*, 1988). Due to the synonymous nature of the species explored and its genetic diversity, there was a need for a change in nomenclature, with the formation of a new genus, *Pantoea*, where strains belonging to the previous genus/species combinations were combined in the species *P. agglomerans* (Gavini *et al.*, 1989).

1.1.2 *Pantoea agglomerans* in the context of the genus *Pantoea*

The genus *Pantoea* belongs to the recently established family *Erwiniaceae* and the order Enterobacterales (Figure 1.1) (Adeolu *et al.*, 2016). Until 2016, the order Enterobacterales comprised of only one family, namely *Enterobacteriaceae*; however, Adeolu and colleagues constructed genome-based phylogenies of members in the order and proposed that the family *Enterobacteriaceae* be split into seven families (Adeolu *et al.*, 2016; Janda and Abbott, 2021). This has since grown to eight validly published families with the family *Erwiniaceae* currently comprising nine validly published genera, including the widely studied genera *Erwinia* and *Pantoea* (Figure 1.1a, 1.1b) (Palmer *et al.*, 2018; Parte *et al.*, 2020; Schoch *et al.*, 2020).

The genus *Pantoea* was established in 1989, after its proposal by Gavini and co-workers, and originates from the Greek word *pantoios*, meaning “of all sorts and sources” (Adriaenssens *et al.*, 2011; Gavini *et al.*, 1989). Testament to the origin of its name, *Pantoea* species are routinely isolated from a wide range of both global and environmental sources (Lv *et al.*, 2022). Furthermore, they are frequently associated with eukaryotic hosts, including insects, plants and animals, where they can have both positive and detrimental effects on the host (Walterson and Stavrinides, 2015).

Species in the genus are Gram-negative, rod-shaped bacteria that are non-spore forming and facultatively anaerobic (Asai *et al.*, 2019). When cultured, colonies are commonly formed on Nutrient Agar, Tryptone Glucose Agar, Luria-Bertani Agar, Trypticase Soy Agar and R2A Agar (Palmer and Coutinho, 2022). The colony morphology observed is a round form, convex elevation and smooth in texture with entire margins (Palmer and Coutinho, 2022). With regards to pigmentation, a variation has been seen within the genus, ranging from yellow-pigmented colonies (as seen with *P. agglomerans*) to beige and some even being non-pigmented (Palmer and Coutinho, 2022). Moreover, these organisms are oxidase negative where they utilise other compounds such as D- galactose, maltose, D-xylose, and D-ribose as a source of carbon in its energy metabolism (Mani and Nair, 2021). From a genome perspective, *Pantoea* species have an average genome size of 4.9 megabases, with a G+C content ranging from 51.8 to 59.6% (Palmer and Coutinho, 2022).

The genus currently encompasses twenty validly published species, with *P. agglomerans* as the type species (Delétoile *et al.*, 2009; Parte *et al.*, 2020). A recent study encompassing core genome and phylogenomic analyses, however, proposed the establishment of 29 new species and 25 new within-species lineages, suggesting a much broader phylogenetic diversity for the genus (Crosby *et al.*, 2023).

Aside from *P. agglomerans,* well known species of *Pantoea* include *P. ananatis*, *P. stewartii*, *P. eucalypti*, *P. septica* and *P. dispersa* (Figure 1.1c). *P. ananatis* is best known as a pathogen on a broad range of plant hosts (Weller-Stuart et al., 2017). Pathogenic strains have been isolated from rice, onions and maize, with a range of disease symptoms such as foliar necrosis and centre rot in onions, necrotic spots and streaks in maize and palea browning of rice (Coutinho and Venter, 2009; de Maayer *et al.*, 2017; Stice *et al.*, 2018). Similarly, *P. stewartii* causes Stewart's wilt and leaf blight in maize plants, a devastating disease which had led to its status as a quarantine pathogen in the United States of America (de Maayer *et al.*, 2017). *P. eucalypti*, on the other hand, has been isolated from *Pinus massoniana* apical buds, where it promoted the growth of pine seedlings and improved their survival rate (Song *et al.*, 2020). From a clinical perspective, *P. septica* has been isolated from humans in wounds, stool samples, bodily fluids as well as specific sites such as the trachea or urethra (Soutar and Stavrinides, 2019). Human pathogenic strains of *P. dispersa* have been associated with bloodstream diseases (Yu *et al.*, 2025). Other *P. dispersa* strains, however, have been isolated from plants, such sweet potato plants, where it functions as a potential biocontrol agent of *Ceratocystis fimbriata*, the causative agent of sweet potato black rot (Jiang *et al.*, 2019). This highlights the dual role of strains of the same species in both positive and negative interactions with various hosts.

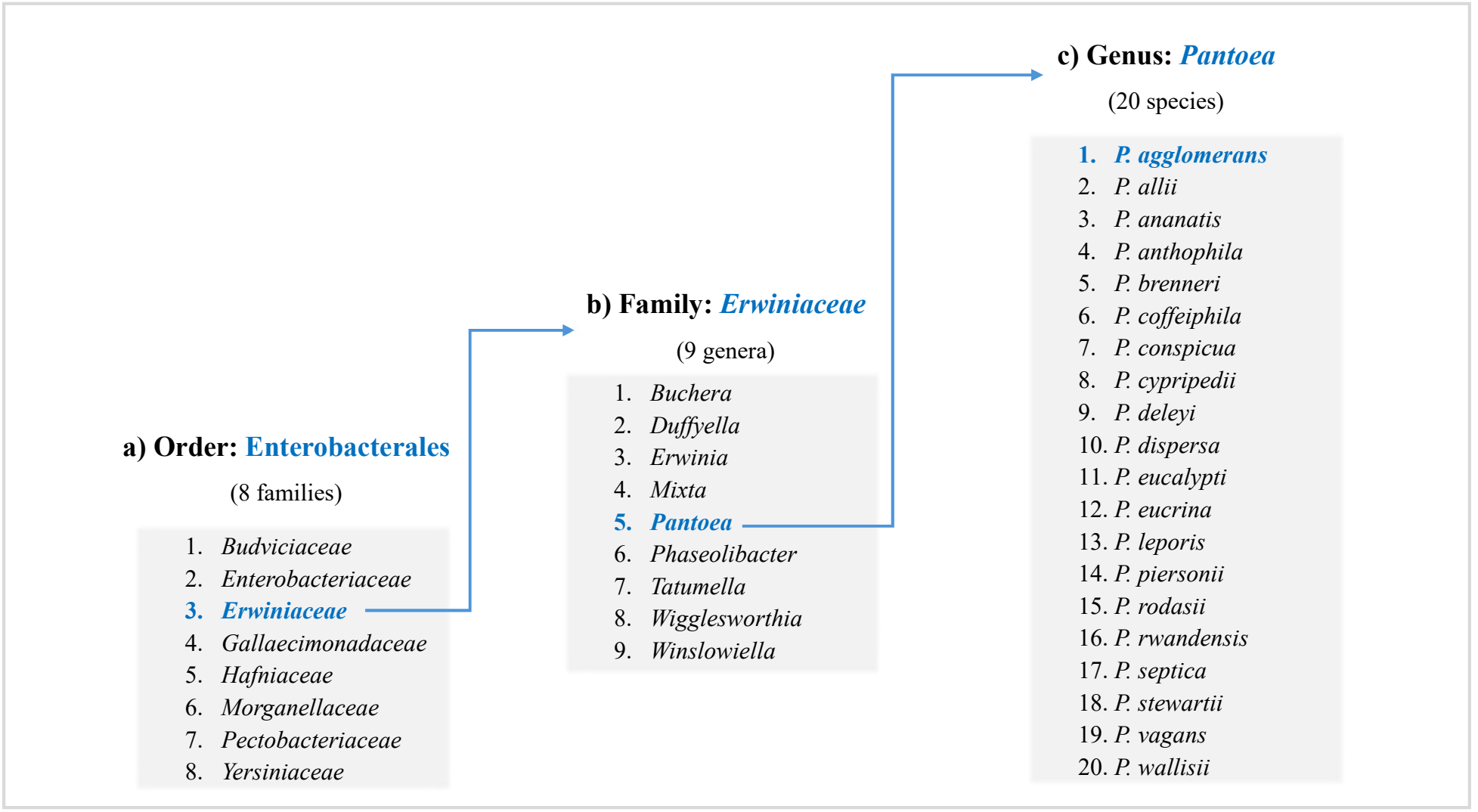


Figure 1.1: Taxonomy of *Pantoea agglomerans* (Adapted from previous studies and constructed based on The List of Prokaryotic names with Standing in Nomenclature (LPSN) (Adeolu *et al.*, 2016; Parte *et al.*, 2020)). This is the classification of all validly published (with correct names) (a) families in the order Enterobacterales, (b) genera in the family *Erwiniaceae* and (c) species in the genus *Pantoea*, which is the lineage of the species of interest in this study, *P. agglomerans*.

1.1.3 *P. agglomerans* in the environment

P. agglomerans strains have been isolated from a variety of environments on every continent, including the cold continent of Antarctica. A study on bacteria isolated from the Johann Gregor Mendel Station in Antarctica identified *P. agglomerans* as the most common Gram-negative isolate, indicating the tolerance of this bacterium at extreme temperatures (Ševčíková *et al.*, 2011). However, *P. agglomerans* typically does not grow at temperatures above 42°C (Abbas *et al*., 2017; Palmer and Coutinho, 2022). Furthermore, this bacterium is characterised by its growth over a wide pH range from acidic (pH 4.0) to slightly alkaline (pH 8.0) conditions (Silini-Che *et al.*, 2012). *P. agglomerans* can also survive under hypersaline conditions, as evidenced by the commercially available biological control agent *P. agglomerans* E325 (Pusey and Wend, 2011).

Given the dynamism of *P. agglomerans* to survive under a broad range of environmental conditions, this species has been isolated from a wide range of environmental sources, including aquatic environments such as rivers (Irawati *et al*., 2022), lakes (Dahiya *et al.*, 2024), salt marshes (Pajuelo *et al.*, 2021) and marine waters (Paulina *et al.*, 2014) as well as a broad range of soils including tropical (Valbuena-Rodríguez *et al.*, 2024), agricultural and desert soils (Shariati *et al.*, 2017), as well as foods such as Baechu kimchi (Guevarra *et al.*, 2021). *P. agglomerans* can exist in extremely controlled and sterile environments, with isolates found on the surfaces of the International Space Station (Bharadwaj *et al.*, 2020) and can even survive in clouds (Huang *et al.*, 2021). *P. agglomerans* is one of the most effective and highly active biological ice nucleating particles (INPs) that are involved in the formation of ice crystals in clouds (Huang *et al.*, 2021). The biological INPs that are on the surface of water, vegetation and soil are lifted via wind and suspended in the air, where they are involved in cloud and ice crystal formation. This cloud seeding subsequently results in precipitation which facilitates the dissemination of *P. agglomerans* to other niches and environmental hosts (Huang *et al.*, 2021).

While *P. agglomerans* is frequently isolated from a wide array of environmental sources, strains are also commonly found in association with plant, animal and other eukaryotic hosts, with which they form intimate relationships in both a positive and negative sense (Walterson and Stavrinides, 2015).

1.1.4 *P. agglomerans* in association with plants

P. agglomerans strains are most frequently isolated from plants and the plant environment, where they can occur as either an endophyte or epiphyte (Bruisson *et al.*, 2019). Epiphytes are found on the surface of the plant where they have specific adaptations allowing them to exist in the phyllosphere, on fruit and vegetable structures and disperse through the water cycle (Bruisson *et al.*, 2019; Koskella, 2020). Epiphytic populations of *P. agglomerans* are found on fruits such as tomatoes and peppers, vegetables such as cabbages, as well as on grains, specifically wheat and rice seeds (Kizheva *et al.*,

2022; Walterson and Stavrinides, 2015). On grain, it constitutes up to 75-99% of the total epiphytic microbiota (Dutkiewicz *et al.*, 2016).

Endophytes refer to those microorganisms that can penetrate the surface of the plant and live between its cells (Bruisson *et al.*, 2019). Endophytic *P. agglomerans* are commonly found within seeds of *Eucalyptus*, rice and wheat and can colonise the roots upon sprouting (Soluch *et al.*, 2021). An endophytic diazotrophic strain of *P. agglomerans* YS19 was previously isolated from rice cultivars in China (Feng *et al*., 2006). The olive plant is home to endophytic *P. agglomerans* that resides in the olive knots caused by the bacterial pathogen *Pseudomonas savastanoi* pv. *savastanoi*, where these two bacterial taxa exist in a synergistic relationship (Hosni *et al.*, 2011).

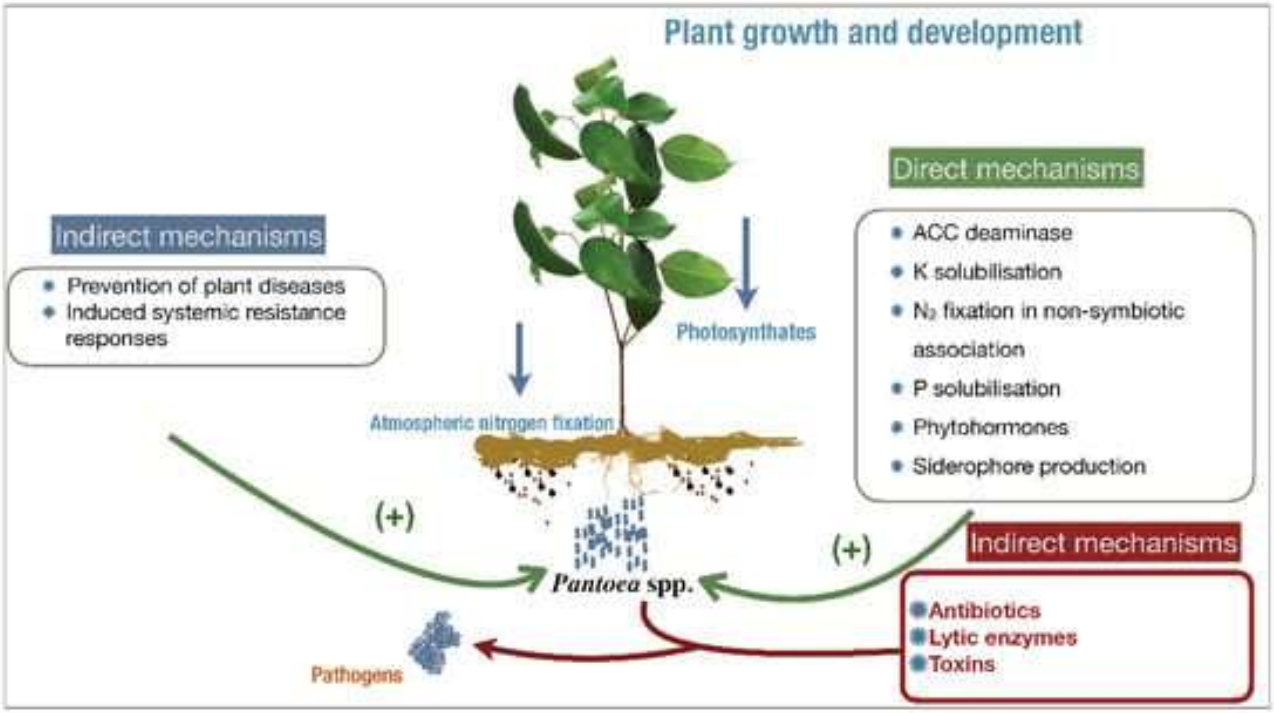


Figure 1.2: Depiction of the direct and indirect PGP mechanisms of *P. agglomerans* (Lorenzi *et al.*, 2022). Plant growth and development is stimulated directly through hormone production, metabolite synthesis and solubilisation of insoluble compounds. Indirectly, strains can produce antibiotics or even increase plant disease resistance.

The relationship between *P. agglomerans* and its plant host is generally benign and has been associated with plant growth promotion (PGP), both directly and indirectly (Figure 1.2) (Lorenzi *et al.*, 2022; Luziatelli *et al.*, 2020). PGP has been linked to the synthesis of metabolites by *P. agglomerans*, specifically the plant hormone indole-3-acetic acid (IAA) and volatile organic compounds such as acetoin and 2,3-butanediol (Kumar *et al.*, 2022; Luziatelli *et al.*, 2020; Shariati *et al.*, 2017). These metabolites promote plant growth by encouraging root formation and increasing resistance to plant diseases (Shariati *et al.*, 2017).

The biostimulant, *P. agglomerans* C1, is involved in the rooting of *Solanum lycopersicum* (tomato) and woody crops as well as directly inducing the earlier root emergence and modifying root

morphology in *Pyrus communis* (pear) (Luziatelli *et al.*, 2020). *P. agglomerans* CPHN2, isolated from *Cicer arietinum* (chickpea), was further shown to promote plant growth through phosphate solubilisation and nitrogen metabolism (Kumar *et al.*, 2022). Similarly, *P. agglomerans* UADEC20 was isolated from agricultural fields in Mexico where it increased nitrogen and phosphorus supply by solubilising ammonia and inorganic phosphates, and further enhanced plant growth by producing phytohormones (Elizondo-Reyna *et al.*, 2025; Singh *et al*., 2020).

While in most cases *P. agglomerans* strains have a benign or positive relationship, they are well known to cause diseases on a broad range of host plants, including crops of agronomic importance (Dutkiewicz *et al.*, 2016). Apricot and plum trees, specifically the Yili wild apricot forest in China, have experienced severe reduction in fruit being produced as a result of pathogenic *P. agglomerans* strains causing spot-hole on the leaves, cankers on the stems and scars on the fruit (Shu *et al.*, 2022; Xu *et al.*, 2022). Furthermore, *P. agglomerans* has been described as the causative agent of cabbage soft rot in China, necrotic disease in jujube, mango and plum, as well as blight and vascular wilt in corn (Guo *et al.*, 2020; Zamorano *et al.*, 2022). Walnut production has been impacted by pathogenic *P. agglomerans* that cause blight (necrosis of fruit) and induce premature fruit drop (An *et al.*, 2023). Other plant pathogenic strains have been identified as causing dieback and necrotic lesions in pistachio trees as well as impacting seed germination, seedling growth and decreasing the nutritional integrity of oat seeds (Wang *et al.*, 2022; Zamorano *et al.*, 2022). In the United States of America, *P. agglomerans* has been isolated as the causal agent of soft rot and centre rot to the bulbs of *Allium cepa* (onions) (Hu, 2019; Shin *et al.*, 2025).

While the phytopathogenicity determinants in *P. agglomerans* remain largely unknown, table beet (*Beta vulgaris*) and *Gypsophila paniculata*-specific strains have been shown to use Type III secretion systems, encoded by the *hrp/hrc* gene cluster, that secrete effector proteins which drive host-specific symptom development, mainly in the formation of galls around the crown region of the stem (Manulis and Barash, 2003). Type III secretion systems (T3SS) are bacterial structures that provide Gram-negative pathogens with a unique virulence mechanism to inject effector proteins (T3Es) directly into the cytoplasm of a host cell (Coburn *et al*., 2007). Gall formation in beets and *Gypsophila* have also been linked to the production of phytohormones, such as IAA and cytokinins, by *P. agglomerans* (Barash and Manulis-Sasson, 2007), suggesting these metabolites serve as double-edged swords, aiding plant growth in one host and phytopathogenesis in another.

1.1.5 Associations of *P. agglomerans* strains with animal hosts

P. agglomerans has also frequently been isolated from animal hosts, both vertebrates and invertebrates. The relationship between *P. agglomerans* and insect or arthropod hosts is generally symbiotic, and most commonly, mutualistic (Dutkiewicz *et al.*, 2016). Wood-eating termites contain

nitrogen fixing strains of *P. agglomerans* in their gut which provide a source of available nitrogen for the host (Potrikus and Breznak, 1977). In the gut of locusts, *P. agglomerans* is involved in the production of guaiacol-containing pheromones, which are excreted in faecal pellets and promote swarming (Dillon *et al*., 2000). *P. agglomerans* strains have been identified in pollen loads, honey sacs and plant blossoms visited by honeybees as well as the bees themselves, indicating their ability to utilise the honeybee as a means of dispersion (Loncaric *et al.*, 2009). Arthropods, on the other hand, are not as common an isolation source with a few reports in ticks, crabs and millipedes (Dutkiewicz, Mackiewicz, Kinga Lemieszek, *et al.*, 2016).

In humans, *P. agglomerans* is commonly identified in faeces, suggesting that it forms a commensalistic relationship with this host (Cicchetti *et al.*, 2006). However, *P. agglomerans* has also been reported as an opportunistic pathogen, mainly in immunocompromised patients, causing wound, blood and urinary tract infections (Penner *et al.*, 2022; Tiwari and Beriha, 2015). There have been cases of bacteraemia in patients with chronic obstructive pulmonary disease as well as inflammation of the eyes, heart muscle and bone, such as septic arthritis that resulted from an exogenous infection (Dutkiewicz *et al.*, 2016; Shrestha *et al.*, 2021). It is important to note that fatal cases of hospital-introduced septicaemia have been reported in several countries and amongst both adults and children (Dutkiewicz *et al.*, 2016). Although more rare, *P. agglomerans* infections have also been derived from contaminated intravenous fluids or blood products within hospitals (Cruz *et al*., 2007). Typically, *P. agglomerans* infections can be treated expediently with effective antibiotics (Tobarran *et al.*, 2023).

P. agglomerans infections are most frequently acquired through wounds or skin lacerations that result from piercing with plant parts, thorns or splinters which allow bacteria to enter, replicate and spread within the human host (Okwundu and Mercer, 2019; Olmos-Alpiste *et al*., 2021). There have been reports of *P. agglomerans* causing 'cotton fever' in intravenous drug users, where the cotton filter may harbour exotoxins and pyrogens from the bacteria (Tobarran *et al.*, 2023). This results in symptoms similar to those of septicaemia, occurring shortly after injection with a cotton filter. Cotton mill workers are also exposed to *P. agglomerans* through inhalation of cotton dust, resulting in respiratory disorders (Dutkiewicz *et al.*, 2015). *P. agglomerans* can also be introduced through ingestion of fruit, vegetables, and grains. Lesions in the gastrointestinal tract, along with a pH imbalance, will allow for the bacteria to be absorbed and its growth stimulated, respectively (Kaur *et al.*, 2020). Further, when farmers have physical contact with the bacteria it can cause allergic dermatitis, as observed in farmers working with infected cows (Dutkiewicz *et al.*, 2016).

Strains of *P. agglomerans* have also been isolated from a wide range of animals and form a broad scope of interactions with these hosts. Pathogenic strains have been isolated from cows, where they play a role in allergic pulmonary disease as well as on the teat skin of healthy cows where no

symptoms are observed (Dutkiewicz *et al.*, 2016; Gaeta *et al.*, 2018). Furthermore, *P. agglomerans* has been linked to haemorrhagic disease in the kidneys of dolphins (Dutkiewicz *et al.*, 2016). A positive interaction between the bacterium and host has been reported in both brown and rainbow trout, as *P. agglomerans* is an antagonist of the oomycete pathogen *Saprolegnia parasitica* (Carbajal-González *et al.*, 2013). This latter antagonism has added to the body of research into the use of *P. agglomerans* as a biological control agent against a range of pathogenic microorganisms and for other biotechnological purposes.

1.1.6 Biotechnological potential of *P. agglomerans* and known determinants

P. agglomerans displays a range of biotechnological applications, such as biological remediation of contaminated environments, biological control of a range of plant pathogens and pests, the production of biopharmaceuticals and as a biofertilizer. Soil-borne isolates from Korea, Venezuela, and Iran have been shown to alter the solubility of inorganic phosphates, making them more easily available to crops (Dutkiewicz *et al.*, 2016). *P. agglomerans* further produces the enzyme phytase that degrades phytate, a phosphorus storage compound in soil, which can only be degraded by some ruminant animals, and as such increases the phosphorous sources available to plants and some animals (Dutkiewicz *et al.*, 2016). Similar biofertilizing potential has been observed in strains isolated from the soybean rhizosphere (Son *et al.*, 2006).

Other *P. agglomerans* isolates, found in the soil of *Avena sativa* (oat) plants, are able to detoxify the environment by degrading hydrocarbons and reducing the heavy metals present (Pishchik *et al.*, 2009). A trial on heavy metal contaminated soil in Nigeria revealed that *P. agglomerans* degraded 99.6%, 96% and 60% of lead, iron and copper in the soil, respectively (Audu *et al*., 2020). *P. agglomerans* strains have been observed to form a thin biofilm layer that prevented heavy metal nanoparticles, such as zinc oxide particles, from migrating downwards and as such, strains in the soil could prevent harmful contaminants from penetrating into deeper soil layers (Kurlanda-Witek *et al*., 2015). *P. agglomerans* LMAE-2 was found in seabed sediment with high concentrations of copper. This strain harboured *cop* genes that contribute to copper resistance and homeostasis, indicating its potential for bioremediation (Corsini *et al.*, 2016). This potential is also seen in other *Pantoea* strains that can degrade herbicides, such as mesotrione, without producing toxic by-products (Walterson and Stavrinides, 2015).

The majority of research focusing on the biotechnological application of *P. agglomerans* has focused on its capabilities as a biological control agent. *P. agglomerans* strains have been used on a variety of fruits to control postharvest fungi such as *Rhizopus tolonifera* and *Penicillium expansum* (Adriaenssens *et al.*, 2011). For example, grapevine trunk disease, which affects the productivity and

longevity of vineyards, is caused by the pathogenic fungus *Neofusicoccum parvum* that can be inhibited by the antifungal volatile compound, phenylethyl alcohol produced by *P. agglomerans* S1 (Haidar *et al.*, 2021). *P. agglomerans* 10IL31 and *P. agglomerans* PSV1-7 counteract *Phytophthora erythroseptica* (pink rot) in potato plants and the soil-borne pathogen, *Aphanomyces euteiches*, which causes root rot of pea plants in Western Canada, respectively (Adiyaman *et al.*, 2011; Godebo et al., 2020).

P. agglomerans strains also produce a range of antimicrobials that can be applied in the control of bacterial plant diseases (Walterson and Stavrinides, 2015). For example, it can counteract fire blight, a disease in pear and apple trees which is caused by the closely related bacterial pathogen *Erwinia amylovora* (Holtappels *et al.*, 2018). Herbicolin O, a histidine-reversible antibiotic, is produced by some strains of *P. agglomerans* and is an effective biocontrol agent that counteracts *E. amylovora* (Ishimaru *et al.*, 2017). A further two antibiotics, pantocin A and pantocin B, have been demonstrated to inhibit enzymes involved in histidine and arginine biosynthesis pathways in *E. amylovora*, respectively (Wright *et al.*, 2001, 2006). On cantaloupe melons, *P. agglomerans* ASB05 can grow, persist and counteract *Salmonella enterica* by producing a phenazine type antibiotic (Lee *et al.*, 2023). Other phenazine antibiotics produced by *P. agglomerans* R190 isolated from apple orchards, have been shown to be effective against *S. enterica* and *Escherichia coli* (Lim *et al.*, 2014). Another *P. agglomerans*-derived antibiotic, andrimid, has shown antimicrobial potential against clinical pathogens in the genera *Enterococcus* and *Staphylococcus* (Williams *et al.*, 2020).

Pantoea agglomerans produces two antibiotic gene clusters called *Pantoea* Natural Products (PNPs) that are effective against Gram-negative and Gram-positive bacteria (Duchateau *et al.*, 2024). PNP-2, produced by *P. agglomerans* Tx10, was demonstrated to be effective against many drug-resistant pathogens such as, *E. coli* and *Enterobacter*, *Staphylococcus* and *Streptococcus* species (Robinson *et al.*, 2020). PNP-3, produced by *P. agglomerans* SN01080 and *P. agglomerans* 3581, inhibited pathogens such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii*(Robinson *et al.*, 2020; Williams and Stavrinides, 2020).

With regards to the pharmaceutical field, *P. agglomerans* ZMR7 has been identified as a possible biopolymer candidate (Al-Qaysi *et al.*, 2021). This strain produces levan-type exopolysaccharides that have antiparasitic and antioxidant activities against *Leishmania tropica*. Most pertinently, this study showed that these exopolysaccharides have antitumor activities, decreasing the viability of rhabdomyosarcoma and breast cancer cells (Al-Qaysi *et al.*, 2021). Symbiotic engineered *P. agglomerans* strains have also been used to explore the treatment of malaria, by delivering anti-malaria effector molecules to the mosquito through the *E. coli* hemolysin A secretion system (Wang

et al., 2012). The study revealed that these strains inhibited the development of malaria parasites as well as decreasing the proportion of mosquitoes carrying the parasites (Wang *et al.*, 2012).

1.1.7 Hurdles to the biotechnological applicability of *P. agglomerans*

While *P. agglomerans* is a key target organism for a wide range of biotechnological applications, the infrequent isolation of this bacterium as a clinical pathogen and its role in plant disease hampers its use. It is currently listed as a BioSafety Level (BSL) 2 organism in Europe and the United States of America, meaning it is safe to use but has a moderate hazard potential and the ability to cause mild disease in humans, impacting the extent of its commercial use (Sulja *et al.*, 2022; Tiwari and Beriha, 2015) . This is in contrast to all other *Pantoea* species, which are classified in the BSL1 category (Völksch *et al.*, 2009). Furthermore, *P. agglomerans* has also been deemed resistant to antibiotics such as aztreonam, cefepime, cefotaxime, cefoxitin, ceftazidime, ciprofloxacin, gentamicin, imipenem, piperacillin/ tazobactam, rifampicin, tetracycline, tobramycin, trimethoprim and vancomycin (Palmer and Coutinho, 2022). As such, the commercial use of *P. agglomerans* strains with wide-spread antibiotic resistance poses as a risk for disease management in the case of infection.

Key to the classification of *P. agglomerans* as an organism of concern is that the biochemical panels which are commonly employed in medical diagnosis of the causative agent of disease are unable to differentiate *P. agglomerans* from other species in the genus or even organisms in closely related genera such as *Phytobacter, Enterobacter,* and *Serratia* (Rezzonico et al., 2012). As such, pathogenic organisms initially presumed to be *P. agglomerans* have often been shown to be representatives of other clinical species using molecular and genotyping approaches (Rezzonico *et al.*, 2009; Soutar and Stavrinides, 2019).

1.1.8 Lessons from the genome of *Pantoea agglomerans*

The advent of genome sequencing has provided unparalleled insights into the biology of *P. agglomerans*, its metabolic and ecological versatility, its pathogenicity and its biotechnological applications. There are currently 295 *P. agglomerans* genomes available in the National Centre for Biotechnology Information genome database, with 43 of these being complete (Sayers *et al.*, 2021). *P. agglomerans* genomes consist of a circular chromosome, with an average genome size of 4.95 megabases that encodes an average of 4,820.80 proteins per strain and an average G+C content of 55.22%, (Dahiya *et al.*, 2024).

Analysing individual *P. agglomerans* genomes has shed light on the genomic contributors to their specific functional and phenotypic adaptations. The soil inoculant, *P. agglomerans* C1, has a genome size of 4.9 megabases and a G+C content of 55.2% (Luziatelli *et al.*, 2020). Genome analysis revealed that this bacterium acquired genes through horizontal gene transfer (HGT) that relate to resistance of

heavy metals, such as arsenic and copper (Luziatelli *et al.*, 2020). *P. agglomerans* C1 can also produce IAA, solubilise phosphatase, inhibit the plant pathogen *E. amylovora* and can act as a biostimulant to tomato cuttings as a result of its siderophore production (Luziatelli *et al.*, 2020). As a result, P. *agglomerans* C1 displays bioremediation potential because of its ability to detoxify soil, secrete plant hormones, increase uptake of required minerals and inhibit plant pathogens.

Similarly, *P. agglomerans* UADEC20 aids in phosphate solubilisation whilst also promoting exopolysaccharide (EPS) production (Elizondo-Reyna *et al.*, 2025). The genome of this strain, comprising 4.2 megabases of chromosomal DNA and two closed plasmids, encodes 57 and 58 proteins involved in phosphate solubilisation and EPS production, respectively (Elizondo-Reyna *et al.*, 2025). This, together with its low virulence and antibiotic resistance levels, makes *P. agglomerans* UADEC20 a potential PGP target for commercial applications.

Genome sequencing has also provided insights into the pathogenicity determinants and resistome of *P. agglomerans*. Genome analysis revealed that *P. agglomerans* pv. *gypsophilae* and *P. agglomerans* pv. *betae* have acquired an ancestral pathogenic plasmid (pPATH) (Weinthal *et al.*, 2007). This plasmid carries the *hrp/hrc* gene cluster encoding a Type III secretion system and the biosynthetic pathways of auxin and cytokinins, modifying these strains to become pathovars (Geraffi *et al.*, 2023; Weinthal *et al.*, 2007). The genome of *P. agglomerans* CHTF15, which is 4.82 megabases in size and harbours three plasmid elements that harbours 600 pathogenicity or virulence genes (the majority relating to iron uptake) (Xiao *et al.*, 2023). These plasmids also harbours 285 genes associated with drug resistance, including tetracycline and macrolide antibiotics (Xiao *et al.*, 2023). Antibiotic resistance is also seen in *P. agglomerans* KM1 (genome size of 4.04 megabases, two prophages and three plasmids), in the form of penicillin G, fosfomycin, vancomycin, rifampicin and bacitracin resistance (Guevarra *et al.*, 2021). *P. agglomerans* KM1 also harbours resistance to oxidative stress, temperature shocks and osmotic lysis (Guevarra *et al.*, 2021). Furthermore, the genome analysis of *P. agglomerans* KM1 revealed virulence factors such as type VI secretion systems (T6SSs), hemolysin and iron sequestration and uptake genes (Guevarra *et al.*, 2021).

There are now rapid, high-throughput and inexpensive DNA sequencing technologies available, together with thousands of bacterial genomes, including many strains of individual species (Giani *et al.*, 2020). This has improved the possibility to effectively compare the genomes of a large number of isolates of this species. On this basis, one can elucidate whether there are genotypic differences that could differentiate between clinical, plant pathogenic and biotechnologically relevant strains of the same species. For example, comparative genomics of commensalistic and pathogenic *E. coli* strains revealed possible treatment paths of the gastrointestinal pathogens (Conway and Cohen, 2015). *E. coli* strains display metabolic flexibility in the gut meaning that the metabolites used

changes based on other *E. coli* strains present (Conway and Cohen, 2015). As such, commensal *E. coli* strains can be engineered to be a defence mechanism against pathogenic strains, such as O157:H7, that cause intestinal infections (Conway and Cohen, 2015).

One of the most comprehensive comparisons so far is the pan-genome analysis of nineteen *P. agglomerans* strains to elucidate on their adaptability to different ecological niches (Dahiya *et al.*, 2024). A pan-genome refers to the entire set of genes within a group consisting of both a core and an accessory (shared accessory and a singleton) genome (Park *et al.*, 2019). The core genome is the group of genes that are conserved across all analysed strains, whereas the shared accessory and singleton genomes consists of genes that are present in some but not all strains and those in only one of the strains, respectively (Matthews *et al.*, 2024). This analysis revealed the competitive environmental advantage of *P. agglomerans* by highlighting genes associated with specific pathways or functions. These include carbohydrate metabolism (97 genes), chemotaxis/ motility (72 genes), iron uptake and siderophore production (36 genes), phosphate and sulphur utilisation (56 genes), stress tolerance (58 genes) and signal transduction systems (57 genes) (Dahiya *et al.*, 2024). Furthermore, protein-protein interactome genes were highlighted for those involved in secretion system elements (77 genes), stress response (31 genes) and plant growth promotion (49 genes) (Dahiya *et al.*, 2024). This indicated how a diverse set of genes, relating to different pathways and functions, were present in the pan-genome. However, while this study provided a comprehensive insight into the genomic flexibility of *P. agglomerans*, it focused on a restricted number (nineteen strains) of isolates and thus, may not have captured the full genotypic and phenotypic diversity of this species.

1.2 Problem Statement

There is an extensive body of literature on *P. agglomerans* and its isolation from a broad range of sources and in association with a variety of different hosts. However, much of this literature is of a descriptive nature, while there has been limited focus on what makes this bacterium so ecologically successful. While the availability of whole genome sequences of *P. agglomerans* have paved the way for understanding factors such as its metabolism, pathogenesis, mechanisms underlying its interaction with specific hosts and has facilitated the identification of novel biomolecules for biotechnological application, much of this research has been isolate-centric, looking at the genotypes and phenotypes of the single isolate. There is thus a knowledge gap on what makes different *P. agglomerans* strains perform different functions and display distinct phenotypes. This paves the way for the application of comparative genomic approaches to elucidate on the diversity and ecological success of this species. This body of knowledge could then be used to inform on what makes one strain a human pathogen, another a pythopathogen, and a third a benign soil-borne organism. Further it will facilitate the safe

usage of *P. agglomerans* strains for a wide range of biotechnological applications, such as biological control of phytopathogenic bacteria and fungi, bioremediation of contaminated environments and the production of medically relevant bioproducts.

1.3 Aims and Objectives

P. agglomerans strains are known to thrive in various environments, under a broad range of conditions and have various interactions with a wide scope of eukaryotic hosts. As such, *P. agglomerans* has evolved various mechanisms for their environmental survival. Furthermore, individual *P. agglomerans* strains possess many traits that could be of biotechnological value, including the production of antibiotics and environmentally stable biomolecules and enzymes. However, the association of *P. agglomerans* with plant disease and as opportunistic human pathogen hampers exploitation of its biotechnological potential. By means of comparative genomic and pan-genome analysis, the proposed project targets to address the following two aims:

1.3.1 Aims

1. To analyse the pan-genome dynamics of *P. agglomerans*.
2. To dissociate pathogenic (plant pathogens and clinical pathogens) from beneficial (biological control and antibiotic producing) strains on the basis of their genome content.
3. To identify factors driving the ecological success of *P. agglomerans*.
4. To identify potential novel biomolecules for biotechnological application.

These aims will be addressed according to the following objectives:

1.3.2 Objectives

a. Improve the assembly and annotate the genomes of >100 strains of *P. agglomerans*.
b. Perform global pan-genome analysis and construct graphs to analyse the evolution with new strains sequenced.
c. Identify key drivers behind ecological diversification (e.g., prophage and plasmid elements).
d. Identify specific proteins and genes that are unique to either individual strains or strains of similar functionality (e.g., plant pathogens, clinical isolates, biocontrol strains).
e. Explore and identify any known and possible novel pathogenic determinants and biotechnological targets.

Chapter Two:
Methodology

2.1 *P. agglomerans* genomes

The genome assemblies of 123 *P. agglomerans* strains were obtained from the NCBI database (Sayers *et al.*, 2021). This included twenty-one complete and 102 draft genomes. Furthermore, the genome assemblies of the type strains of *P. eucalypti* (LMG 24197^T) and *P. vagans* (LMG 24199^T) and were included as outgroup taxa. The taxonomy of all strains utilised in the study was confirmed using GTDB-tk v.4.0 (Chaumeil *et al.*, 2020) against the Genome Taxonomy Database (GTDB) (Parks *et al.*, 2022). The NCBI Bioproject and Biosample databases and their linked literature were used to build a metadata set for each of the taxa (Barrett *et al.*, 2012). This included information on each of the strains such as their isolation source, country of origin and host interactions.

Average Nucleotide Identity (ANI) values between all pairs of taxa were determined using FastANI v.1.33 (Jain *et al.*, 2018). The ANI values (%) were converted into a distance matrix (1-(ANI%/100)) and a neighbour joining tree was generated using the *neighbour.exe* function in the PHYLIP v.3.698 package (Felsenstein, 1993). Based on this tree, the relatedness of the *P. agglomerans* strains was determined. Those taxa with complete genomes, which clustered closely with taxa for which only a draft genome is available, were selected as a reference for improving the draft genome assemblies (Sahlin *et al*., 2016). This was done by combining contigs into scaffolds in the query genomes on the basis of the reference genomes using MeDuSa v.1.6 (Bosi *et al.*, 2015) and Multi-CSAR (Liu *et al*., 2022). Gap regions within scaffolds, indicated by stretches of Ns, were subjected to BlastN (Altschul *et al.*, 1990) analysis against the NCBI non-redundant (nr) nucleotide database and any overlaps were corrected.

2.2 Structural annotation and genome completeness analyses

The final dataset of complete and improved draft genome assemblies were structurally annotated using Prodigal v.2.6.3 (Hyatt *et al.*, 2010). Structural annotation refers to the identification of specific functional elements such as ribosomal RNAs, transfer RNAs and protein-coding genes (and their corresponding amino acid sequences), within each of the genome sequences (Bright *et al.*, 2009).

The resultant protein datasets were utilised to assess genome completeness with Benchmarking Using Single-Copy Orthologues (BUSCO) v.5.2.2 (Manni *et al.*, 2021). The genomes were searched using the enterobacterales_ odb10.2019-04-24 lineage dataset comprising 440 BUSCO markers. These markers refer to genes (or the proteins they encode) present in at least 90% of the chosen dataset (of which 90% are single copies) and a BUSCO score is assigned based on the proportion of marker genes that are found in the assembled dataset (Manni *et al.*, 2021). To ensure consistency in the subsequent analyses, only those genomes with a BUSCO genome completeness of >97% (427/440

markers) were retained. Finally, the genomes of 100 *P. agglomerans* strains with the best assembly metrics (the least number of contigs and a BUSCO score >97%) were selected for further analyses.

2.3 Orthologous group identification

The protein datasets (proteomes) of the 100 *P. agglomerans* and two outgroup taxa were subjected to OrthoFinder v.2.5.5 (Emms and Kelly, 2019) for the identification of protein orthologues (proteins conserved among genomes that are derived from a common ancestor) and paralogues (proteins derived through gene duplication in the same genome) (Altenhoff *et al.*, 2013; Koonin, 2005). OrthoFinder was run using default parameters, with pairwise BlastP (Altschul, 1997) for orthology searches. The resultant orthogroup (referring to set of genes, that have descended from a single gene in the most recent common ancestor, across the considered species) matrix was used for further phylogenomic and pan-genome analyses (Emms and Kelly, 2019).

2.4 Phylogenomic analyses

Single copy orthologous (SCOs) proteins (proteins shared by all compared taxa with no evidence of paralogy) were identified from the OrthoFinder output and extracted from the proteomes for each genome for the construction of a core genome phylogeny. The SCOs were individually aligned in m-Coffee, an implementation of T-Coffee v.13.46.0.919e8cb6 (Wallace *et al.*, 2006). M-Coffee simultaneously computes multiple sequence alignments using a set of pair-wise and multiple sequence alignment algorithms and combines the separate outputs into one alignment that is, on average, more accurate than its individual components (Moretti *et al.*, 2007). The resultant alignment of the individual SCOs were then concatenated using an in-house Perl script. To improve the alignment, poorly aligned regions were removed using Gblocks v.0.91 (Dereeper *et al.*, 2008).

The curated amino acid alignments were used to construct a maximum likelihood phylogeny using IQ-Tree2 v.2.3.6 (Minh *et al.*, 2020; Nguyen *et al.*, 2015) , with the optimal evolutionary model predicted using Modelfinder (Kalyaanamoorthy *et al.*, 2017) and branch support using ultra-fast bootstrapping (UFboot: n = 1000 replicates) (Minh *et al*., 2013).

To support the core genome phylogeny, two phylogenomic metrics were calculated on the genome dataset, namely digital DNA-DNA hybridization (dDDH) and Average Nucleotide Identity (ANI) values. The dDDH values (%) were determined using the Genome-to-Genome Distance Calculator v.3.0 (Meier-Kolthoff *et al.*, 2022), where each *P. agglomerans* genome was compared against all others in the dataset. ANI values were determined using FastANI (Jain *et al.*, 2018), as discussed in section 2.1. These values assess the sequence identity of orthologous genome fragments between two comparator strains (Arahal, 2014; Jain *et al.*, 2018). The dDDH values represents the DNA similarity

between comparator strains, which is calculated according to GGDC formula 2 (Meier-Kolthoff and Göker, 2019). This calculation represents the identities found in high-scoring segment pairs (HSPs) compared to the overall HSP length (Kangale et al., 2021). Both phylogenomic metrics, ANI and dDDH, have defined species boundaries of between 95%-96% and >70%, respectively, with dDDH also having an inter-subspecies threshold of 80% (Goris *et al.*, 2007; Konstantinidis and Tiedje, 2005; Meier-Kolthoff *et al.*, 2014). The resultant dDDH and ANI matrices were then converted into distance matrices and neighbour joining trees were generated using the *neighbour.exe* function in the PHYLIP v.3.698 package (Felsenstein, 1993).

2.5 Pan-genome analyses

The OrthoFinder output matrix was converted to a presence/absence matrix, where the presence of a protein orthologue (for each orthologous group) in each genome was coded as a 1 and the absence as 0. The resultant matrix was analysed using the Bacteria Pan-Genome Analysis (BPGA) pipeline v.1.3 (Chaudhari *et al*., 2016) to generates a Pan-genome file. In short, the presence (1) and absence (0) of orthologues for each orthogroup is determined for each combination of genomes, commencing with two genomes and progressively including up to 100 genomes in the analysis.

A pan-genome refers to the entire set of genes within a group, which in this case is a species, consisting of both a core and an accessory (shared accessory and a singleton) genome (Park *et al.*, 2019). The core genome is the group of genes that are conserved across all strains within the species (100 taxa), whereas the shared accessory and singleton genomes consists of genes that are present in some but not all strains (2-99 taxa) and those in only one of the strains (1 taxon), respectively (Matthews *et al.*, 2024). The pan- and core genomes were subsequently plotted in Excel 365 on the basis of the Pan-genome file. Phylogenomic analyses (section 2.4) indicated that the *P. agglomerans* strains form two phylogenetically distinct clusters. As such, the pan- and core genomes representative of the taxa in the two clades were also plotted using the same approach.

The BPGA Pan-genome files were further analysed in PanGP v.1.0.1 (Zhao *et al.*, 2014) to determine the parametric values for extrapolation of both the pan- and core genome. By this means it could be predicted what would happen to the pan-genome and core genome if 500 or 1,000 genomes were to be included in the analysis. PanGP determines these parametric values by using the curve fitting functions in accordance with Heap's Law (Tettelin *et al.*, 2008). For the pan-genome and core genome the formulas $y = Ax^B + C$ and $y = Ae^{Bx} + C$ were used, respectively, where y is the size of the pan- and core genome, respectively, x represents the number of genomes included in the analysis and A, B and C are curve fitting parameters. Further, the number of new genes that would be added to the pan-genome when an increased number of genomes is added to the analysis was determined using

the curve fitting function $y = Ax^B$, where y represents the number of new genes added to the pan-genome, x represents the number of genomes included in the analysis and A and B are curve fitting parameters.

The core, accessory and singleton proteins were identified from the OrthoFinder output, and a Venn diagram was constructed based on these pan-genome elements. Representative proteins for each orthogroup in the different genome fractions (core, shared accessory and singleton) were extracted and functionally annotated using EggNOG-mapper v.2.1.12 (Cantalapiedra *et al.*, 2021) against the EggNOG Database v.5.0 (Huerta-Cepas *et al.*, 2019). This was done for all 100 *P. agglomerans* strains as well as separately for the two clades. EggNOG-mapper assigns functions to the selected proteins based on the precomputed clusters and phylogenies in the EggNOG database, which integrates data from 5,090 organisms, 2,502 viruses and 4.4 million orthologous groups (Huerta-Cepas *et al.*, 2017, 2019).

2.6 Identification of molecular drivers of diversification

Genome sequences of *P. agglomerans* strains were run in PlasmidFinder v.2.1 (Carattoli and Hasman, 2020), which consists of 133 unique plasmid replicon sequences in *Enterobacteriaceae*. The complete plasmid elements and contigs were extracted from the genomes of *P. agglomerans* on the basis of the presence of plasmid replication proteins as per the EggNOG functional annotations (section 2.5). Shorter sequences in improved-draft genomes were further compared to the NCBI nr nucleotide database (Sayers *et al.*, 2021) to identify hits to known plasmids. The resultant replicon nucleotide sequences were used to calculate plasmid sizes, the G+C contents and relative genomic proportions. Proteins linked to the predicted plasmid replicons/plasmid contigs were extracted and OrthoFinder was used to determine protein conservation among the different plasmid replicons.

Integrated bacteriophage elements, or prophages, were predicted by comparing the genome nucleotide sequences to the PHASTEST web server (Wishart *et al.*, 2023). The relative genomic proportions of the prophage elements were determined along with the identity of the prophage elements on the genomes which was determined by sequence homology to known bacteriophages in the PHASTEST database (Wishart *et al.*, 2023).

Integrative Conjugative Elements (ICEs) were observed to be key drivers of diversification in the closely related species *Pantoea ananatis* (de Maayer *et al.*, 2015). These are mobile genetic elements that integrate into the host genome and mediates the DNA transfer to other cells through functional conjugation systems (Johnson and Grossman, 2015). ICEs were predicted by uploading the genome nucleotide sequences into ICEfinder v.2.0, which compares predicted ICEs against the ICEberg v.3.0 database (Wang *et al.*, 2023). The sizes of the ICE elements and the number of proteins they encode,

as well as predicted functions encoded on the ICEs (such as virulence factors, acquired antibiotic resistance or antiphage systems) were recorded. Transposable elements were predicted using Transposon Finder (TnComp_finder) v.1.0.0 (Alvarenga and Varani, 2019), which performs a local BlastP analysis with the proteome against a custom transposon database.

2.7 Classification of protein representatives into functional categories

Protein representatives of each orthogroup in the core, shared accessory and singleton genome fractions were classified according to the functional categories they belong to. This classification was done on the basis of the Clusters of Orthologous Groups (COGs) output from EggNOG_mapper (Galperin *et al.*, 2021)(section 2.5). The relative proportions of proteins involved in each functional category in the core, shared accessory and singleton genome fractions were calculated. This was done for all 100 *P. agglomerans* strains and also separately for the two clades.

2.8 Insight into the pathogenic potential of *P. agglomerans*

P. agglomerans has been linked to opportunistic infections in humans and animal hosts (Penner *et al.*, 2022; Tiwari and Beriha, 2015). To analyse the human-pathogenic potential of the comparator strains, the *P.* agglomerans genome sequences were uploaded on the PathogenFinder v.1.1 server (Cosentino *et al.*, 2013). PathogenFinder compares genomic fragments against a protein family database comprising groups of proteins which are associated with 372 human pathogens and 513 non-pathogenic bacteria. By identifying pathogen- or non-pathogen-specific proteins in the target genome, the probability of a particular organism being a human pathogen was predicted.

To further evaluate the pathogenic potential of the *P. agglomerans* strains to human, animal and plant hosts, the presence of known pathogenicity determinants in the *P. agglomerans* strains was predicted by BlastP comparison of the proteomes against the Pathogen-Host Interaction Database (PHI-Base) v.5.0 (Urban *et al.*, 2021). PHI-Base contains information on pathogenicity-related genes that have proven to affect the phenotypic outcome of pathogen-host interactions (Urban *et al.*, 2019; Winnenburg *et al.*, 2005). This database includes >10,000 genes and 7,082 proteins from 295 pathogens affecting 246 hosts (involving 22,415 interactions) (Urban *et al.*, 2024). Only those proteins sharing >70% amino acid identity and >70% query/subject coverage with the reference proteins in PHI-Base v. 5.0, were retained.

Secreted proteins play a key role in bacterial interactions with their host or environment (Green and Mecsas, 2016). The subcellular localisations of the entire proteome were determined using PSORTb v.3.0.3 (Yu *et al.*, 2010). This data was then used to map the secretome associated with each *P. agglomerans* strain. To further elucidate the secretome, secretion systems were predicted using

Macromolecular System Finder (MacSyFinder) v.2.1.4 (Néron *et al.*, 2023). This tool applies a Hidden Markov Model prediction model to identify macromolecular systems in a bacterial proteome. In this case, secretion systems were predicted in the *P. agglomerans* proteomes using the TXSS model (Yoon, 2009).

2.9 Prediction of *P. agglomerans* resistome

Pivotal to their ecological success is the ability of microorganisms to cope with both natural and man-made compounds they may encounter in the environment, including heavy metals and antimicrobial agents (Ejaz *et al.*, 2022). As such, the resistome of *P. agglomerans* was analysed using several different tools. Resistance to antibiotics and antimicrobial agents were determined using both Resfinder 4.6.0 (Bortolaia *et al.*, 2020) and the NCBI Antimicrobial Resistance gene Finder (AMRFinder) v.4.0.3 (Feldgarden *et al.*, 2019). Resfinder identifies acquired genes or mutations that facilitate antimicrobial resistance phenotypes, whereas AMRFinder identifies the AMR genes present by comparing the genome sequences to the Bacterial Antimicrobial Resistance Reference Gene database (Bortolaia *et al.*, 2020; Feldgarden *et al.*, 2019) . Resistance to heavy metals and environmental antibacterial biocides among the *P. agglomerans* strains were assessed by performing local BlastP analysis against the Antibacterial Biocide and Metal Resistance Genes dataset (Bacmet) v.2.0 (Pal *et al.*, 2013), using the database with experimentally validated reference proteins. Only those protein hits sharing >70% amino acid identity and >70% query/subject coverage with the reference proteins in Bacmet v.2.0, were retained.

2.10 Antimicrobial production in *P. agglomerans*

Certain *P. agglomerans* strains are known to produce distinct antimicrobial compounds, and as such have been of interest as biological control agents against a range of plant pathogens and pests (Duchateau *et al.*, 2024). The 100 *P. agglomerans* strains, were thus screened for the production of secondary metabolites using the antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) v.7.0 tool (Blin *et al.*, 2023) against the antiSMASH database v.4.0 (Blin *et al.*, 2024). This database consists of 231,534 biosynthetic gene cluster regions across 35,726 bacterial, 236 fungal and 592 archaeal genomes (Blin *et al.*, 2024).

Chapter Three:
Results and Discussion

3.1 *Pantoea agglomerans* metadata

The genomes of 100 *P. agglomerans* strains were analysed in this study. These strains were derived from various environmental sources globally. The majority of *P. agglomerans* strains originated from North America (52%), followed by Europe (28%), Asia (14%) and Africa (3%), while only a single strain was isolated from South America and Oceania, respectively (Appendix A). With reports of strains also isolated from Antarctica (Ševčíková *et al.*, 2011), this stands testament to the global distribution of this species.

Most of the *P. agglomerans* strains incorporated in the study are host-associated, with only six and three strains from soil and water, respectively, indicating the preponderance of research on host-microbe interactions. The majority of *P. agglomerans* strains (82/100) were isolated from plant material, with 41 described as epiphytic, twenty-one strains described as endophytic and eight identified as plant pathogens (Table 3.1). In association with non-plant hosts, two strains were isolated from insects, while four strains were found in association with fungi (Table 3.1). While *P. agglomerans* is recognised as a clinical pathogen, the genomes of only three human-associated isolates are available (Table 3.1). These include the type strains *P. agglomerans* FDAARGOS 1447^{T} (= ATCC 27155^{T} = DSM 3493 = LMG 1286) isolated from a knee laceration in Zimbabwe in 1888, *P. agglomerans* Tx10 isolated from the sputum of a cystic fibrosis patient in the USA (Robinson *et al.*, 2020) and *P. agglomerans* NCTC10500 isolated from the cervix of a patient in the United Kingdom (Appendix A).

The selected strains show a broad distribution, both geographically and ecologically, thus providing an effective dataset for the analysis of the factors underlying the ecological diversification of *P. agglomerans*. This distribution is also mirrored in other species within the same genus, such as *P. ananatis*, which has been isolated across all continents and has been found in association with plants, humans, insects and the environment (Weller-Stuart *et al*., 2017).

Table 3.1: Summary of isolation sources of the 100 compared *P. agglomerans* strains.

Isolation Source	Lifestyle	Frequency
Plants	Epiphytes	41
	Endophytes	21
	Pathogens	8
	Other	12
Environment	Soil	6
	Water	3
Fungi		4
Insects		2
Humans	Clinical isolates	3
	Total	100

The genomes of the compared *P. agglomerans* strains are on average 4.89 megabases (Mb) in size (Table 3.2). This is slightly larger than the genomes of both closely related outgroup taxa *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T (4.80 and 4.79 Mb, respectively) included in the study (Appendix A). Comparison against other taxa whose genomes are available in the NCBI genome database, showed that the *P. agglomerans* genomes are substantially larger than the average for members of the family *Erwiniaceae* (4.26 Mb) and the type species of the sister genus *Erwinia*, *E. amylovora* (3.81 Mb). The genomes of the comparator *P. agglomerans* strains, however, are similar in size to the genus *Pantoea* (4.80 Mb) as a whole and the species *P. ananatis* (4.94 Mb) and *P. dispersa* (4.91 Mb), for which a large number of genomes are available. The *P. agglomerans* strains do show extensive variation in genome size, with the largest genome, belonging to *P. agglomerans* DAPP-PG734, being ~ 870 kilobases (kb) larger than that of *P. agglomerans* NC, both of which are found in association with plant hosts (Table 3.2; Appendix A). Similarly, extensive variation in the number of proteins encoded on the *P. agglomerans* genomes was observed, with 896 less proteins encoded on the genome of the plant isolate *P. agglomerans* T1 than the strain with the largest genome, *P. agglomerans* DAPP-PG734 (Table 3.2).

The *P. agglomerans* genomes have an average G+C content of 55.54%, similar to that of the outgroup taxa *P. eucalypti* LMG 24197^T (54.57%) and *P. vagans* LMG 24199^T (55.62%) (Appendix A). However, extensive variation (±3.45%) was observed across the *P. agglomerans* genomes (Table 3.2; Appendix A). Given that bacterial genomes have little intergenic DNA and are densely packed with coding genes, the varying G+C content would have an influence on the expression and functioning of the resultant proteins (Bohlin *et al*., 2017; Hildebrand *et al*., 2010; Sela *et al*., 2016). A study on *Escherichia coli* showed a strong relationship between G+C content and fitness, demonstrating that the higher proportions of guanines and cytosines in the genomes of *E. coli* strains are associated with increased evolutionary pressures (Bohlin *et al*., 2017). The extensive variation in genome size, G+C content and protein complement of the *P. agglomerans* strains suggest extensive genomic flexibility exists within this species, which could drive their expansive diversification to the distinct ecological niches and lifestyles adopted by members of this species.

Table 3.2: Summary of the genome metrics of the 100 *P. agglomerans* strains.

	Genome		**Chromosome**		**Genome G+C content**		**Genome Proteins**	
	Size (Mb)	Strain	Size (Mb)	Strain	GC%	Strain	Total	Strain
Average	4.89		4.05		55.54		4,446.69	
Minimum	4.53	NC	3.92	PMG_056	54.18	Pa39-27	4,091	T1
Maximum	5.40	DAPP-PG734	4.41	DAPP-PG734	57.63	CFBP8756	4,960	DAPP-PG734

3.2 Phylogenomic analyses support the delineation of two potential *P. agglomerans* subspecies

To elucidate the taxonomic placement of the compared *P. agglomerans* strains, several genome-based taxonomic approaches were employed. The *P. agglomerans* proteomes were subjected to Orthofinder (Emms and Kelly, 2019) and a total of 3,206 single copy orthologous (SCOs) proteins were identified which are conserved among all comparator strains, as well as the two outgroup taxa, *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T (Figure 3.1). A core genome maximum likelihood (ML) phylogeny encompassing these SCOs was constructed, which shows the cohesive clustering of the *P. agglomerans* strains away from the outgroup taxa (Figure 3.1). However, bifurication within the *P. agglomerans* strains suggests that they form two distinct clades. The species and clade clustering is supported by a bootstrap values of 100%, indicating the accuracy and reliability of the delineation, with values >95% being statistically significant (Russo and Selvatti, 2018; Soltis and Soltis, 2003).

An unrooted core genome ML phylogeny, comprising only the 100 *P. agglomerans* strains (based on 3,331 SCOs), further demonstrates the clear delineation into two distinct clades, with Clade 1 comprising 78 taxa, and Clade 2 comprising twenty-two taxa, respectively (Figure 3.2; Appendix B). This agrees with a recent study of *P. agglomerans* strains found in association with onion bulbs, which were observed to form at least two distinct phylogroups (Shin *et al.*, 2025). The isolation sources (from plants, human, insect, fungal hosts or the environment) of the *P. agglomerans* strains were mapped on the phylogeny. This showed that taxa associated with human, plant and fungal hosts are distributed in both clades, along with the environmental isolates. Insect-associated isolates are restricted to Clade 1, but this only encompasses two strains. As such, association with different hosts does not appear to be a driver of the observed phylogenetic clustering.

The core genome phylogenies were further supported by Average Nucleotide Identity (ANI) and digital DNA-DNA Hybridisation (dDDH) values. Pair-wise ANI values between the *P. agglomerans* strains and the outgroup taxa, ranged between 90.46% - 90.75% (average 90.63%) and 91.09% - 91.34% (average 91.20%) with *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T, respectively. Pair-wise dDDH values between the *P. agglomerans* strains and the outgroup taxa, ranged between 41.30% – 42% (average 41.70%) and 42.90% - 43.70% (average 43.20%) with *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T, respectively. This clearly delineates *P. agglomerans* as a distinct species from the outgroup taxa, where ANI values of 95-96% and dDDH values of 70% are recognised as effective species boundaries (Goris *et al.*, 2007; Konstantinidis and Tiedje, 2005; Meier-Kolthoff *et al.*, 2014). Intra-specific ANI values of 97.03% - 99.99% (average 98.09%) and dDDH values of 77.30% – 100% (average 85.60%) demonstrate that all 100 *P. agglomerans* strains belong to the same species.

The two *P. agglomerans* clades observed in the core genome phylogeny, were also evaluated on the basis of their ANI and dDDH values. Neighbour-joining distance-based phylogenies constructed on the basis of these metrics also group the *P. agglomerans* strains into two distinct clades (Figure 3.3, 3.4). While intra-clade ANI values of 98.14% – 99.98% (average 98.53%) and 98.10% - 99.99% (average 98.50%) are observed for Clade 1 and Clade 2, respectively, inter-clade values are much lower, ranging from 97.03% – 97.55% (average 97.32%). Similarly, intra-clade dDDH values of 85.60% – 99.90% (average 89.30%) and 86.50% – 100% (average 88.90%) were observed for Clade 1 and Clade 2, respectively, while the average inter-clade value of 79.30% (range: 77.30% – 80.70%) was substantially lower.

A previous study on *Escherichia coli* strains suggested that a dDDH boundary of 80% can be used to delineate bacterial subspecies (Meier-Kolthoff *et al.*, 2014). While there were a few strains sharing an inter-clade dDDH value >80% (maximum of 80.70%), the average dDDH value of 79.30% suggests that Clade 1 and Clade 2 could be considered as distinct subspecies of *P. agglomerans*. Thus, for the purpose of this study, the taxa were further treated as such, where Clade 1 (78 strains) represents *P. agglomerans* subspecies 1 and Clade 2 (twenty-two strains) represents *P. agglomerans* subspecies 2 (Appendix B).

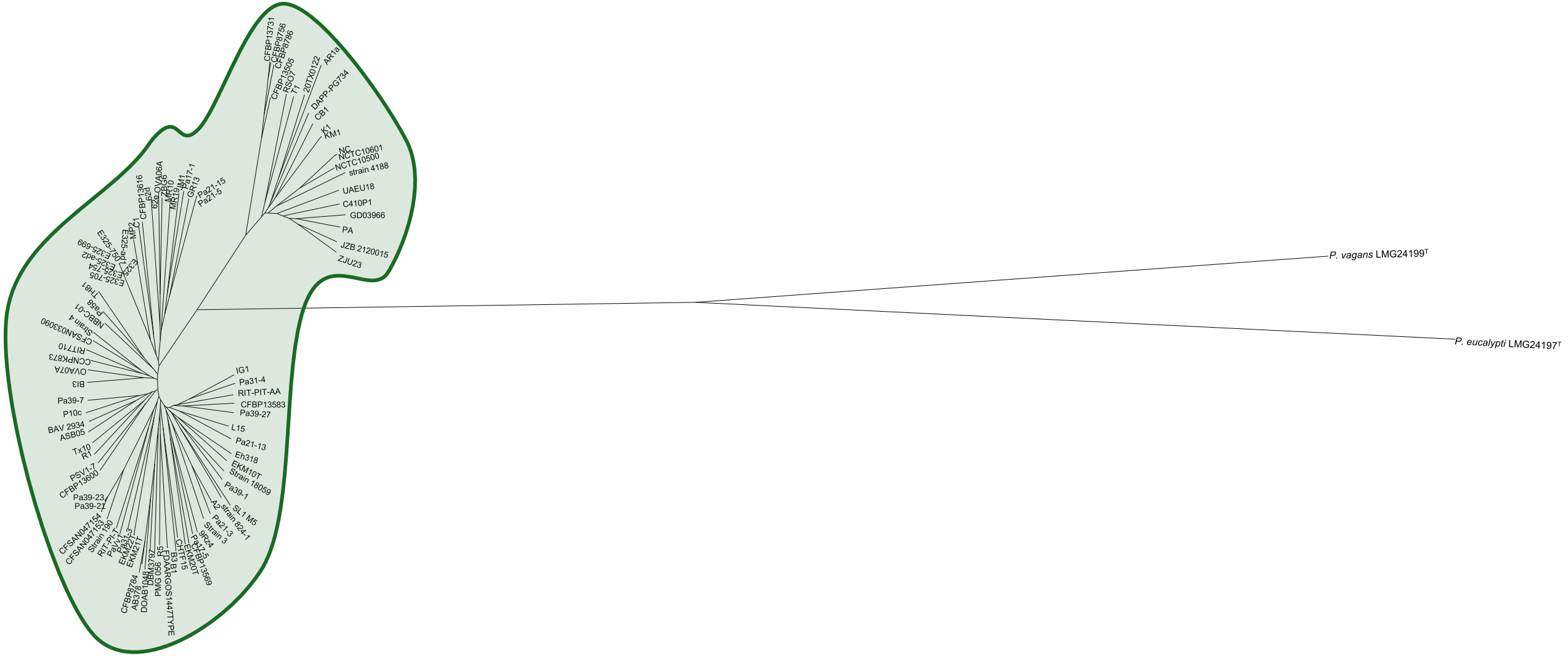


Figure 3.1: Core genome phylogeny of *P. agglomerans* with outgroup taxa. The ML phylogeny was constructed on the basis of 3,206 SCOs, comprising a curated alignment of 1,037,441 amino acid positions. The tree was constructed using the optimal evolutionary model HIVb+F+I+G4 with bootstrap support (n=1,000 replicates).

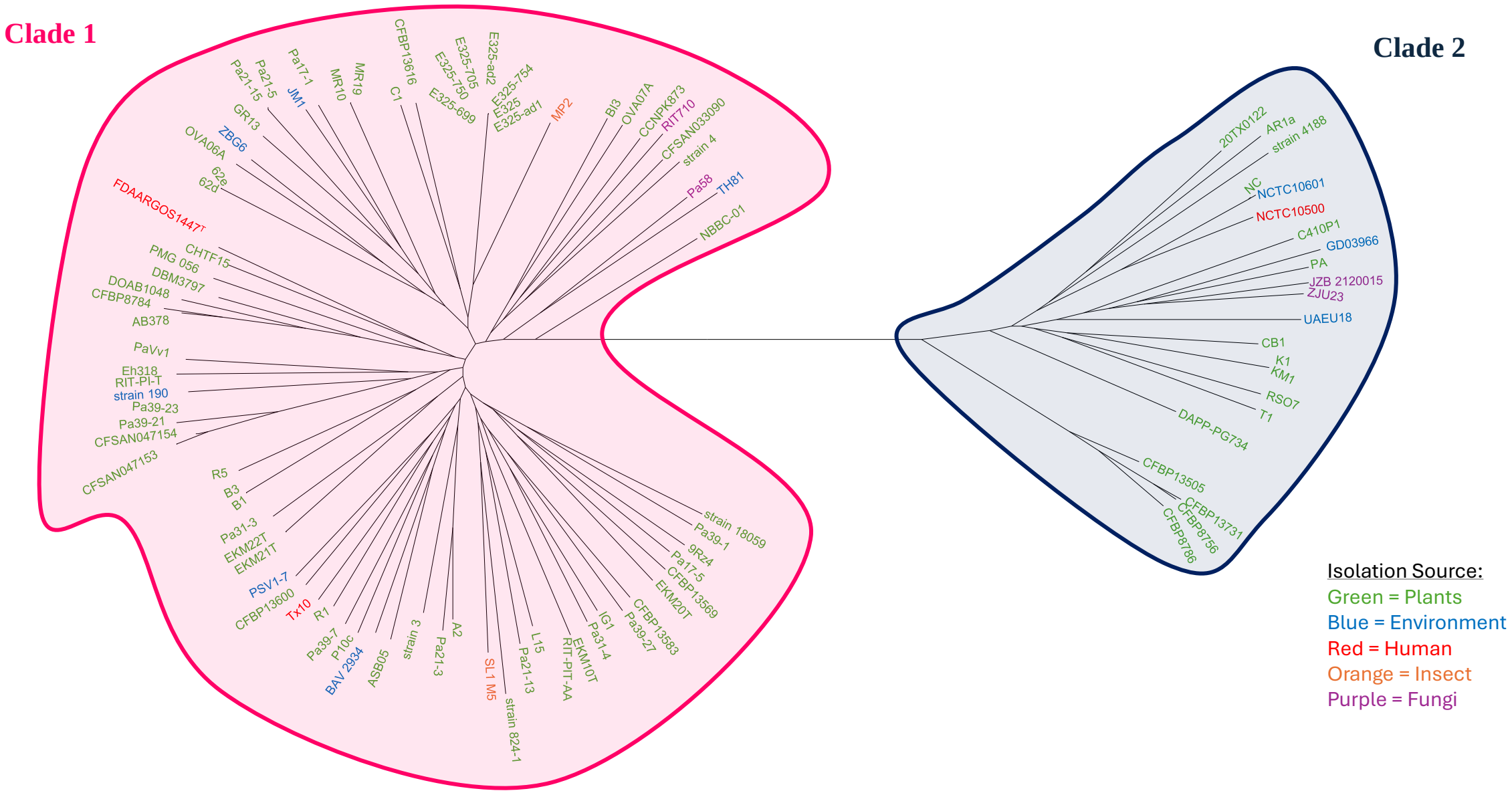


Figure 3.2: Core genome phylogeny of 100 *P. agglomerans* strains. The ML phylogeny was constructed on the basis of 3,331 concatenated SCOs with the curated alignment comprising 1,082,573 amino acid positions. The tree was constructed using the optimal evolutionary model HIVb+F+I+G4 with bootstrap support (n = 1 000 replicates). *P. agglomerans* clade 1 (pink) consists of 78 strains, with *P. agglomerans* clade 2 (blue) encompassing the remaining twenty-two strains.

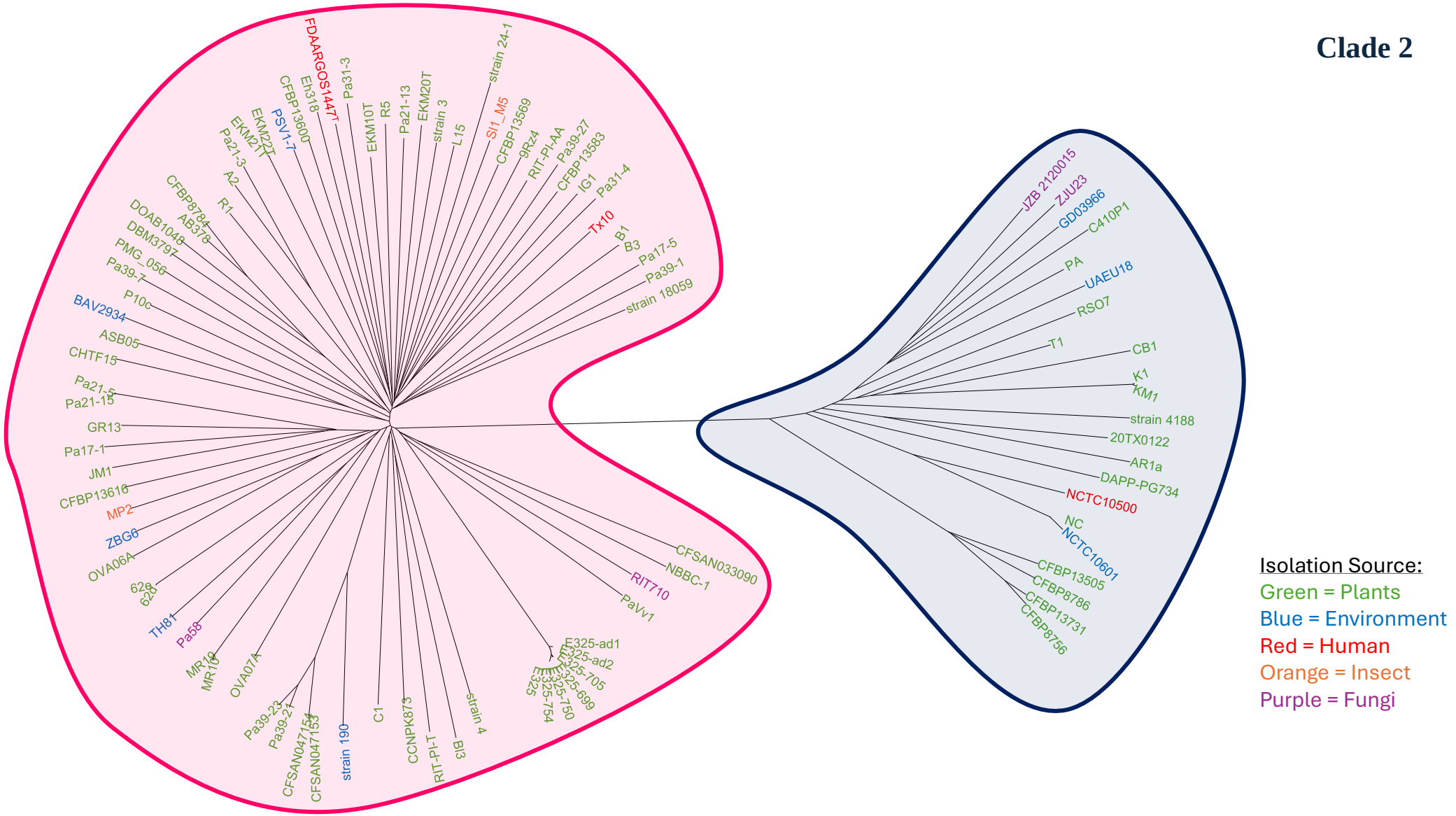


Figure 3.3: Neighbour-joining tree constructed on the basis of pairwise ANI values for *P. agglomerans*. *P. agglomerans* clade 1 (pink) has 78 strains and clade (blue) has twenty-two strains.

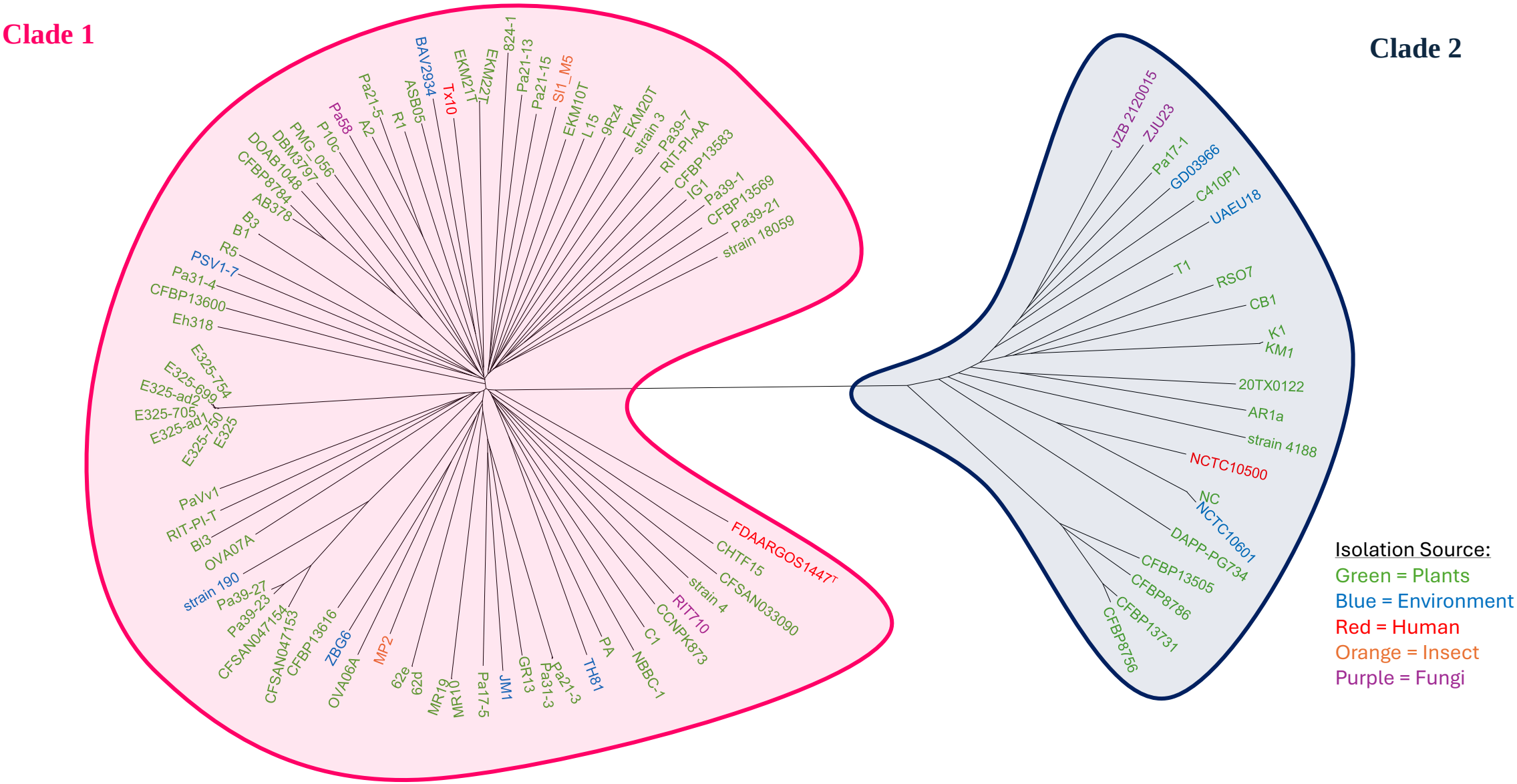


Figure 3.4: Neighbour-joining tree constructed on the basis of pairwise dDDH values for *P. agglomerans.* *P. agglomerans* clade 1 (pink) has 78 strains and clade (blue) has twenty-two strains.

3.3 *P. agglomerans* displays an open pan-genome

The pan-genome, comprising the core (shared by all taxa), shared accessory (shared by some but not all taxa) and singleton (unique to a single comparator taxon) proteins, was analysed on the basis of the OrthoFinder (Emms and Kelly, 2019) output. The 100 compared *P. agglomerans* strains have an expansive pan-genome, encompassing 11,666 protein orthogroups. However, only a small portion (29.43%) of this pan-genome is core to all 100 strains (Figure 3.5). A similar analysis of the pan-genome of 273 *Cronobacter sakazakii* strains demonstrated a pan-genome comprising 17,158 protein orthogroups, of which only 19.5% were core to all compared strains (Lee and Andam, 2019). By contrast, the accessory fraction of *P. agglomerans* comprises of 70.57% of the orthogroups.

The pan-genome of different bacterial species can be considered as either open (where increasing the number of genomes included leads to the continual increase in the total number of genes in the pan-genome) or closed (the total number of genes in the pan-genome plateaus as more genomes are compared) (Matthews *et al.*, 2024). The degree to which a pan-genome is considered open can be assessed with the core/pan-genome ratio (%), with higher percentages (≥ 89%) reflecting closed pan-genomes (Rouli *et al.*, 2015).

The pan-genome ratio can also be suggestive of the lifestyle of a bacterial taxon and whether it is a sympatric or allopatric species (Rouli *et al.*, 2015). Allopatric species are specialised bacteria that live in a narrow range of ecological niches with reduced gene flow, typically have small genomes and display a closed pan-genome (Liu *et al.*, 2022; Rouli *et al.*, 2015). A pan-genome analysis of twenty-seven distinct species identified a core/pan-genome ratio of ≥ 89% (representing a closed pan-genome) in nine of the species, all of which were allopatric (Rouli *et al.*, 2015). On the other hand, sympatric bacteria live in a community, have open and broad pan-genomes, with high rates of horizontal gene transfer (Liu *et al.*, 2022; Rouli *et al.*, 2015).

For *P. agglomerans*, a ratio of 29.43% (3,433/11,666*100) is observed which is an indication of the ability of this species to acquire extrachromosomal DNA through its open pan-genome. When considering the accessory genome of the 100 comparator *P. agglomerans* strains, 52.37% (5,111 orthogroups) of the orthogroups are shared by two or more taxa, while 47.63% (3,122 orthogroups) are unique to single strains (singletons) (Figure 3.5). This is further suggestive of various routes of gene acquisition and the exposure to foreign genetic material.

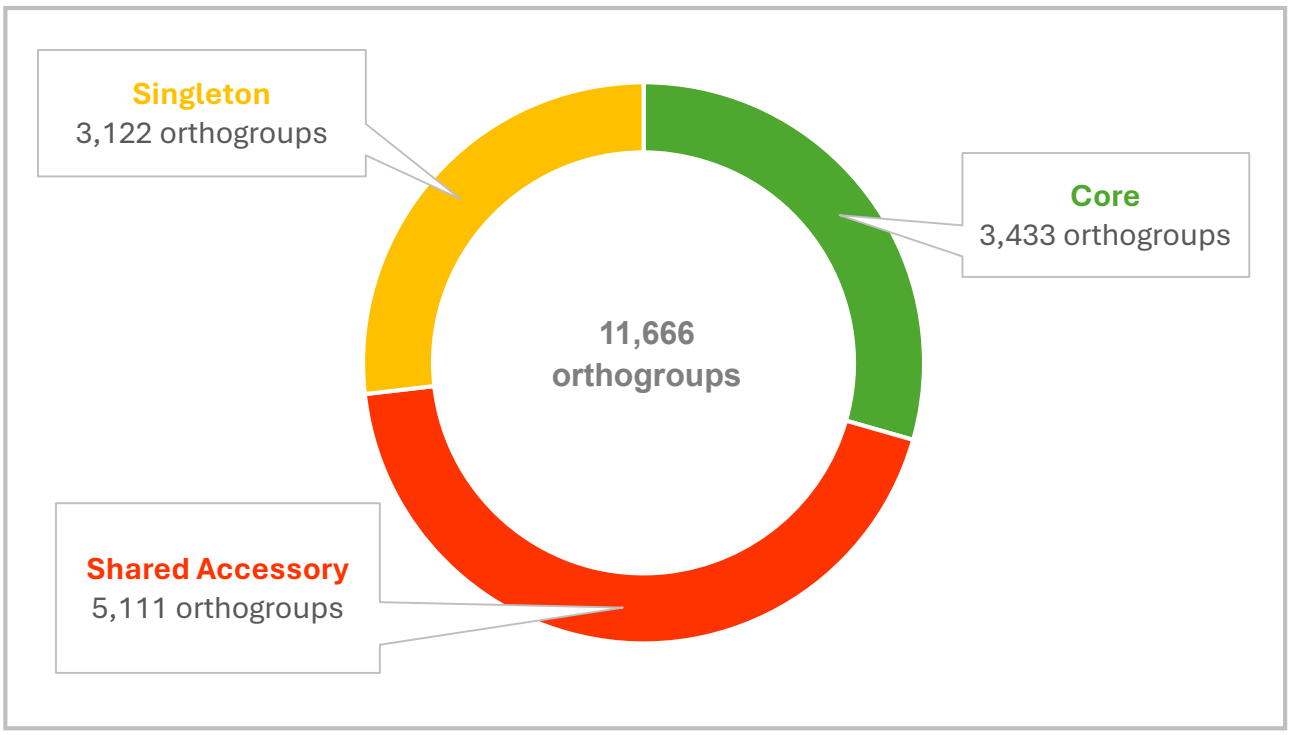


Figure 3.5: The distribution of orthogroups in the pan-genome fractions of *P. agglomerans*. A total of 11,666 orthogroups are identified in the core (green), shared accessory (orange) and singleton (yellow) genome fractions of *P. agglomerans*.

The pan- and core genomes of the 100 *P. agglomerans* strains were plotted as curves, descriptive of the pan- (core, accessory and singleton genes) and core genomes when comparing single and all combinations of strains up to 100 genomes (Figure 3.6; Appendix C). The pan-genome curve continues to increase even, when 100 genomes were compared, further supporting an open pan-genome for this species (Figure 3.6). By contrast, the core stabilised, reaching ~3,514 orthogroups when 100 genomes were compared. Similar analyses of 81 and seventeen strains of the closely related *P. ananatis* (Agarwal *et al.*, 2021a) and *P. stewartii* subsp. *indologenes* (Agarwal *et al.*, 2021b), respectively, showed that these taxa likewise display open pan-genomes and a stabilising core genome. Dips could be observed in both the core and pan-genome curves. This is likely due to the distinct *P. agglomerans* subspecies, which may have distinct degrees of genomic flexibility (Goyal, 2018).

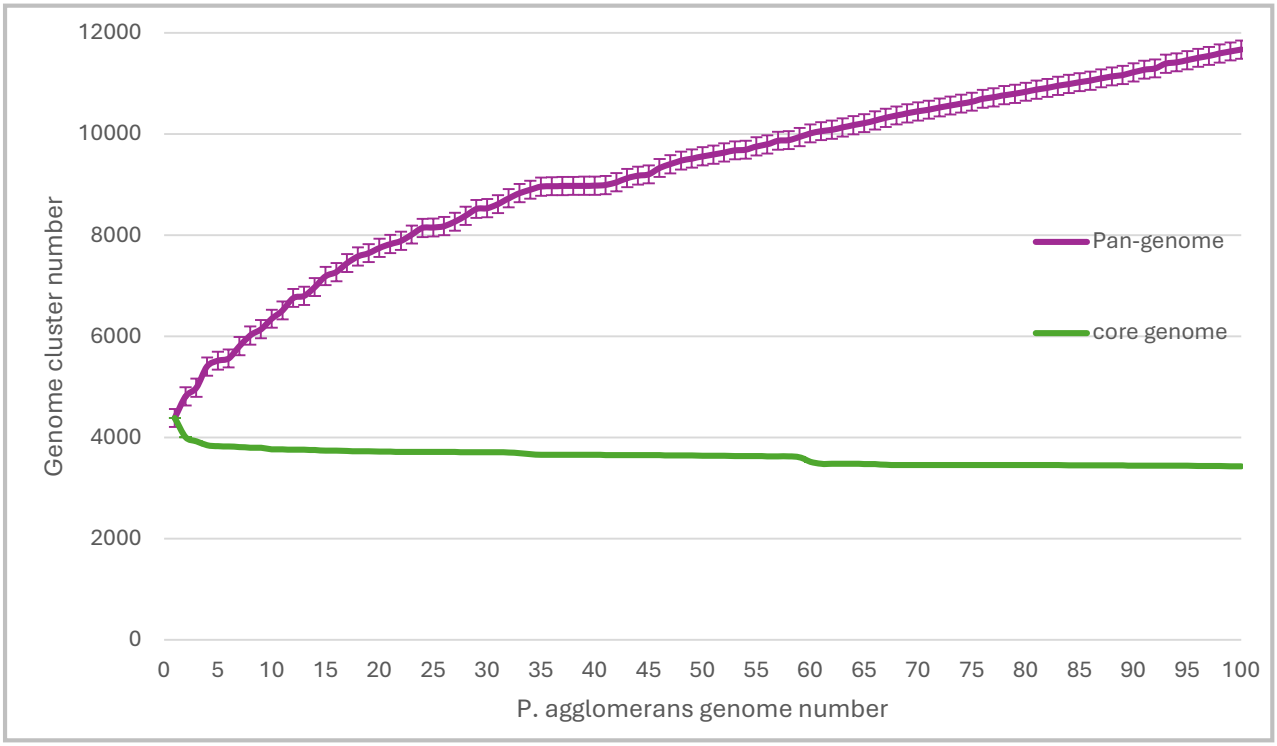


Figure 3.6: Pan- and core genome graph of the 100 *P. agglomerans* strains. The pan-genome (purple) and core genome (green) data, produced in BPGA, was plotted using Excel 365.

The *P. agglomerans* pan-genome was further extrapolated, using the curve fitting functions to model Heap's Law in PanGP (Tettelin *et al.*, 2005). By this means, the dynamics of the pan-genome if 500 or 1,000 genomes were added to the analysis could be modelled. This extrapolation showed that the pan-genome size continues to increase, with the pan-genome of 1,000 *P. agglomerans* being nearly double (1.9 x) that of 100 *P. agglomerans* strains (Table 3.3). While the 100th genome is predicted to add ~30 new genes to the pan-genome, eight new genes would be added with the 1000th genome. By contrast, the core genome stabilises between genome 500 and 1,000.

The pan-genome extrapolation should, however, be viewed with caution, given that it would assume that the 100 genomes represent effective sampling across the entire species. Given the extensive set of singletons among the 100 compared strains, the extrapolation may not be reflective of the broader species (Medini *et al.*, 2020). This can be observed for the true pan-genome as predicted for the 100 *P. agglomerans* strains (11,666 orthogroups), which is ~ 500 proteins larger than when extrapolation is applied (11,165.51).

Table 3.3: *P. agglomerans* pan-genome extrapolations in accordance with Heap's law.

Function		Number of genomes analysed		
		100	**500**	**1,000**
Pan-genome	$y= 1{,}951.08x^{0.33} + 2{,}247.35$	11,165.51	17,415.59	21,314.03
Core genome	$y= 663.08e^{-0.01x} + 3{,}271.03$	3,514.96	3,275.50	3,271.06
New genes	$y= 389.14x^{-0.56}$	29.52	11.99	8.13

3.4 Pan-genome analysis of *P. agglomerans* subsp. 1 and subsp. 2

Comparison of the pan-genomes of the two *P. agglomerans* subspecies showed a substantially larger pan-genome for *P. agglomerans* subsp. 1 (10,065 orthogroups), compared to *P. agglomerans* subsp. 2 (7,635 orthogroups), which likely reflects the smaller sample size (twenty-two strains) of the latter subspecies (Figure 3.7). Notably, the singleton proportion of the accessory genome of *P. agglomerans* subsp. 2 (57.84% of the accessory genome) was substantially larger than that of *P. agglomerans* subsp. 1 (45.50%), suggesting greater genetic variability in *P. agglomerans* subsp. 2 (Figure 3.7). When assessing the core genomes, *P. agglomerans* subsp. 2 incorporated nine proteins conserved in all twenty-two strains that were unique to this subspecies, while only a single unique protein is shared by all 78 *P. agglomerans* subsp. 1 strains (Figure 3.7).

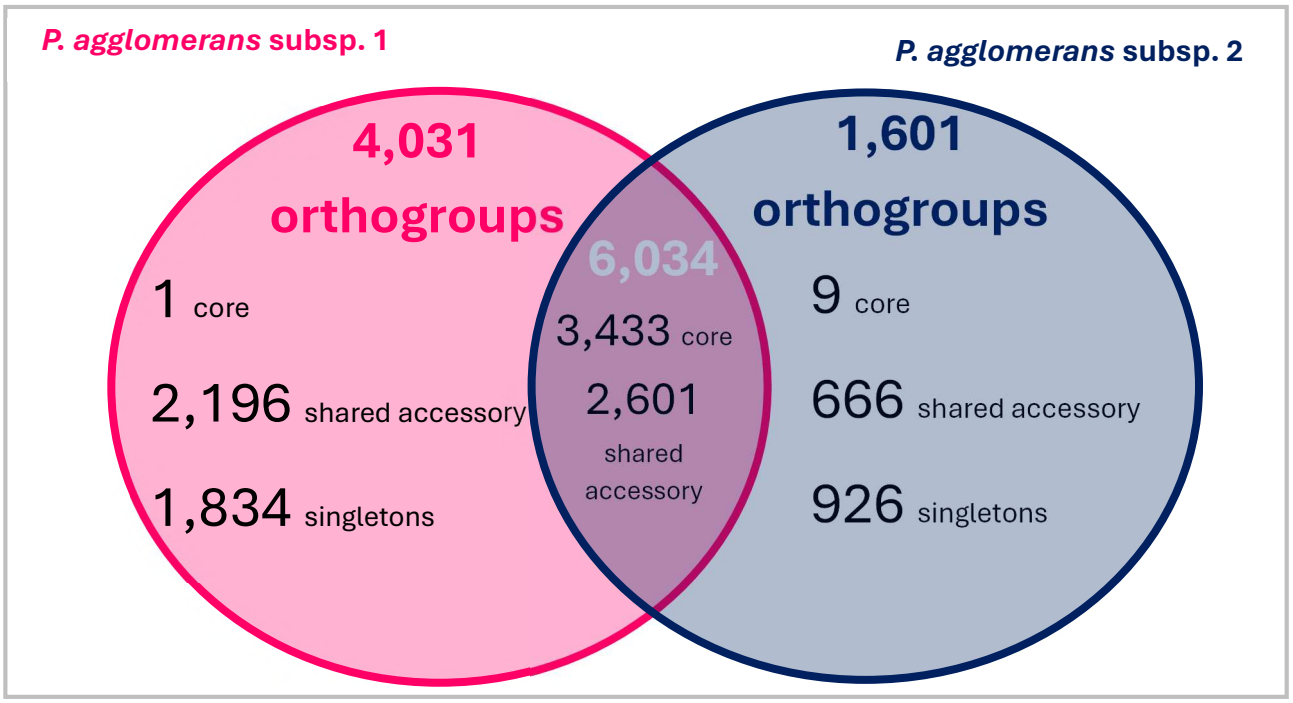


Figure 3.7: Representation of the pan-genome fractions across the two *P. agglomerans* subspecies. Of the 11,666 pan-genome orthogroups in *P. agglomerans*, 6,034 orthogroups occur in both *P. agglomerans* subsp. 1 and 2. Subsp. 1 (pink) has 4,031 orthogroups, of which 2,196 are shared between two and 77 taxa, 1,834 is in only one of the 78 strains and one orthogroup is shared across all 78 strains. Subsp. 2 (blue) has 1,601 orthogroups in it is pan-genome, consisting of nine orthogroups core to all twenty-two strains, 666 orthogroups present in two to twenty-one of the taxa and 926 orthogroups in the singleton genome fraction.

Functional analysis of these proteins revealed that the single core protein of *P. agglomerans* subsp. 1 was of unknown function, whereas all nine *P. agglomerans* subspecies 2-unique proteins were assigned functional roles. This included a signal transduction histidine kinase (located in bacterial membranes and enabling phosphorylation), a DNA-binding response regulator (OmpR), a tripartite tricarboxylate transport system (TctABC), two copies of the type I fimbrial protein FimA and two cytochrome bd-type quinol oxidases. OmpR is a cytoplasmic transcriptional regulator, which has been shown to influence the expression levels of porin proteins (OmpF and OmpC) in response to

osmolarity changes in *E. coli* (Brzóstkowska *et al*., 2012). FimA is a key component of type 1 fimbriae, adhesive structures extending from the bacterial surface that promote bacterial colonisation and adhesion to host cells and surfaces (Puorger *et al.*, 2011). The TctABC system, comprising the transmembrane permease TctA, helper protein TctB and extracytoplasmic receptor TctC, plays a crucial role in the uptake of tricarboxylates, such as citrate, from the environment (Winnen *et al*., 2003). Cytochrome bd-oxidases are transmembrane, quinol-dependent, terminal oxidases that are exclusive to prokaryotes (Boot *et al.*, 2017). They have a high oxygen affinity and play a key role in the oxidative phosphorylation pathway and can counteract bacterial stress caused by hypoxia, antibacterial drugs and reactive oxygen and nitrogen species (Boot *et al.*, 2017).

Plotting of the pan- and core genomes of *P. agglomerans* subsp. 1 and 2 (up to twenty-two comparator genomes) showed that, while both subspecies display open pan-genomes, *P. agglomerans* subsp. 2 attained the larger pan-genome, likely due to the large proportion of singleton proteins (Figure 3.8; Appendix C). The core genome of this subspecies was smaller across twenty-two compared genomes than that of *P. agglomerans* subsp. 1.

Extrapolation of the pan-genomes up to 500 and 1,000 genomes, however, showed that *P. agglomerans* subsp. 1 is predicted to attain the bigger pan-genome when 1,000 genomes are compared, 2,840 proteins larger than that of the entire species (Table 3.4). By contrast, the *P. agglomerans* subsp. 2 pan-genome is predicted to be 33% smaller than that of *P. agglomerans* subsp. 1 when comparing 1,000 genomes. This suggests greater genetic diversity and genome flexibility for *P. agglomerans* subsp. 1 across the entire population. The greater genomic flexibility is evident from the number of new genes added to the pan-genome by the 1000th strain, where *P. agglomerans* subsp. 1 displays ~6x the number of new genes compared to *P. agglomerans* subsp. 2 (Table 3.4). Again, however, extrapolation of open pan-genomes across smaller sample sets should be viewed with caution.

Table 3.4: Pan-genome extrapolations in accordance with Heap's Law for *P. agglomerans* subsp. 1 and subsp. 2.

	Function		Number of genomes analysed		
			100	**500**	**1,000**
subsp. 1	Pan-genome	$y= 858.84x^{0.46} + 3,552.21$	10,695.73	18,529.68	24,154.35
	Core genome	$y= 409.78e^{-0.06x} + 3,685.42$	3,686.44	3,685.42	3,685.42
	New genes	$y= 245.52x^{-0.44}$	32.37	15.94	11.75
subsp. 2	Pan-genome	$y= 3,582.73x^{0.21}+ 744.12$	10,167.66	13,957.01	16,027.31
	Core genome	$y= 977.04e^{-0.34x} + 3,610.73$	3,610.73	3,610.73	3,610.73
	New genes	$y= 1,004.65x^{-0.92}$	14.52	3.30	1.75

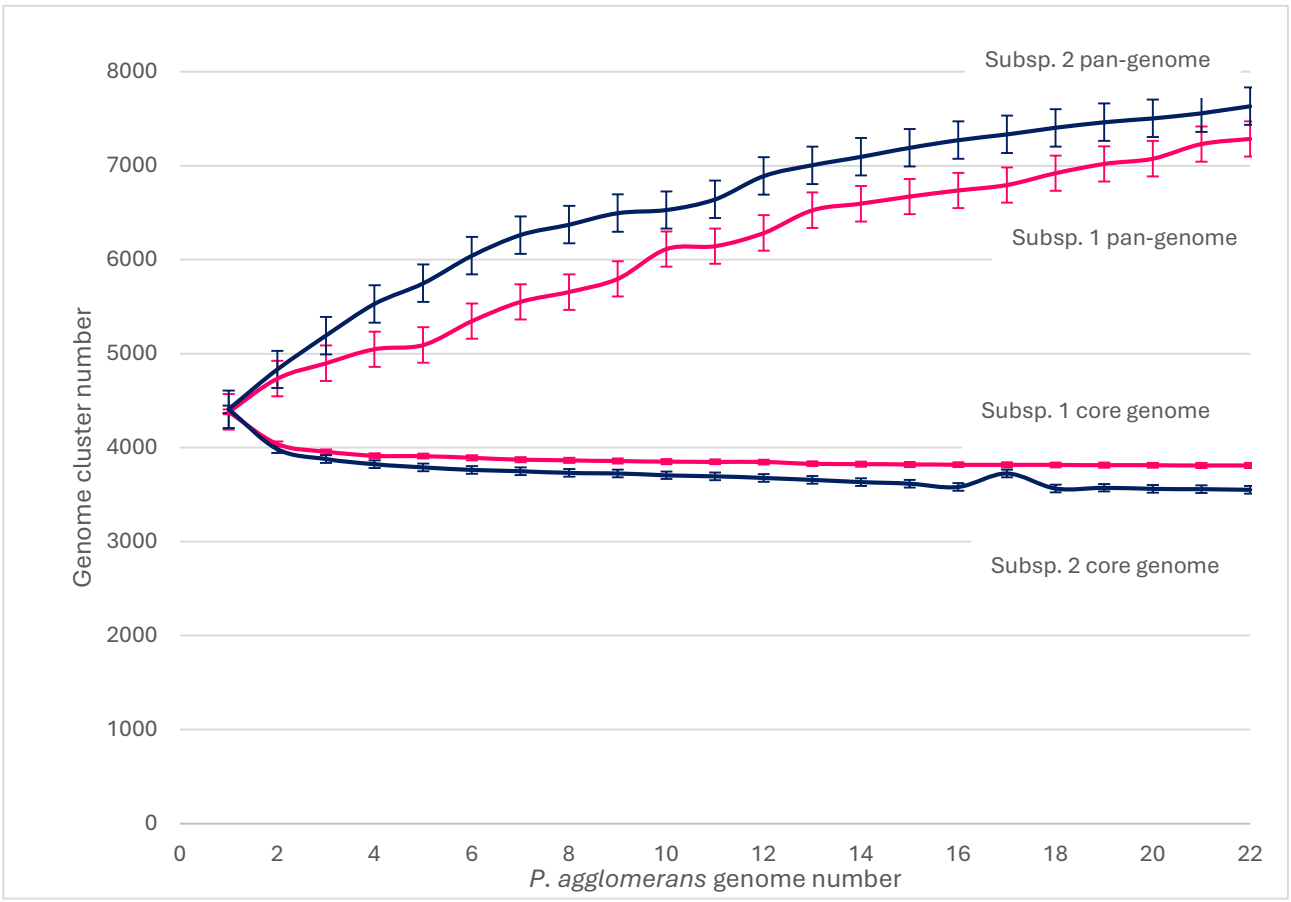


Figure 3.8: Pan- and core genome graph of *P. agglomerans* subsp. 1 and subsp. 2. The pan-genome and core genome data for *P. agglomerans* subsp. 1 (pink) and subsp. 2 (blue), produced in BPGA, was plotted using Excel 365.

3.5 Characterisation of the *P. agglomerans* mobilome – drivers of diversification

The open pan-genome of *P. agglomerans* and the diversity in terms of genome sizes and G+C content among the 100 comparator *P. agglomerans* strains suggest extensive genomic flexibility and evolution (Table 3.2). Key to this genomic diversification is the mobilome, which includes transferable genetic elements such as plasmids, transposable elements, bacteriophages and integrative and conjugative elements (ICEs) *(*Carr *et al*., 2021; Vale *et al*., 2022). Here, several tools were employed to elucidate the *P. agglomerans* mobilome and its contribution towards ecological diversification.

3.5.1 Plasmids in *P. agglomerans*

Plasmids are highly heterogeneous elements, that range from less than a kilobase to megabases in DNA size, and are incorporated into bacterial genomes through the horizontal gene transfer (HGT) mechanisms conjugation, transduction and transformation (Carr *et al.*, 2021). Multiple studies on individual *P. agglomerans* strains have demonstrated that they have integrated plasmid elements in their genome, which contribute to their functional and ecological diversity (Crosby *et al.*, 2023; Guevarra *et al.*, 2021; Wang *et al.*, 2023; Xiao *et al.*, 2023). Here, on the basis of the functional annotations of the proteomes of the 100 *P. agglomerans* genomes and the OrthoFinder (Emms and Kelly, 2019) output, plasmid elements were identified.

A total of 306 distinct plasmid elements were identified among the *P. agglomerans* strains. On average, plasmid DNA contributes 17.15% of the total genome and encompass an average of 841.40 kb of DNA (Table 3.5). The relative proportion of plasmid DNA ranges between 10.84% (490.77 kb) in *P. agglomerans* NC and 23.16% (1,213.46 kb) in the *P. agglomerans* B3 genome. The largest sum of plasmid DNA in a single genome, 1,213.49 kb was observed in *P. agglomerans* B1, contributing 23.12% of its genome (Appendix D). In the outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, plasmids contribute 15.90% (763 kb) and 15.45% (740.16 kb) of the total genomic DNA, respectively (Appendix D). The number of plasmid elements found in each strain ranged from a single plasmid in *P. agglomerans* NC and T1 to seven (*P. agglomerans* B1, B3 and Pa39-23), with an average plasmid size of 272.50 kb across the 306 plasmid elements (Appendix D). However, there is extensive variability in the DNA size of individual plasmid elements, with a single plasmid DNA contributing 0.02% (1.19 kb) and 13.62% (678.17 kb) to the genomic content of *P. agglomerans* Pa39-23 and A2, respectively.

Table 3.5: Overall contribution of plasmid DNA to *P. agglomerans* genomes.

	Total plasmid DNA (kb)	Proportion of Genome (%)	Strain
Average	841.40	17.20	
Minimum	490.77	10.84	NC
Maximum	1,213.46	23.16	B3

The *P. agglomerans* plasmids were placed into three distinct categories on the basis of size. Large *Pantoea* Plasmids (LPPs) were categorised as those plasmid elements which are larger than 100 kb in size, while Medium *Pantoea* Plasmids (MPPs) ranged between 20 kb and 100 kb in size and Small *Pantoea* Plasmids (SPPs) were < 20 kb in size. The majority of the 306 plasmids belong to the LPP category (89.22%; 273/306), with 6.54% (20/306) and 4.25% (13/306) belonging to the MPP and SPP categories, respectively (Figure 3.9; Appendix D). These were further subclassified according to the conservation of orthogroups among the plasmid elements, resulting in fourteen subcategories of LPPs (LPP-1 to -14), thirteen subcategories of MPPs (MPP-1 to -13) and twelve subcategories of SPPs (SPP-1 to -12) (Table 3.6).

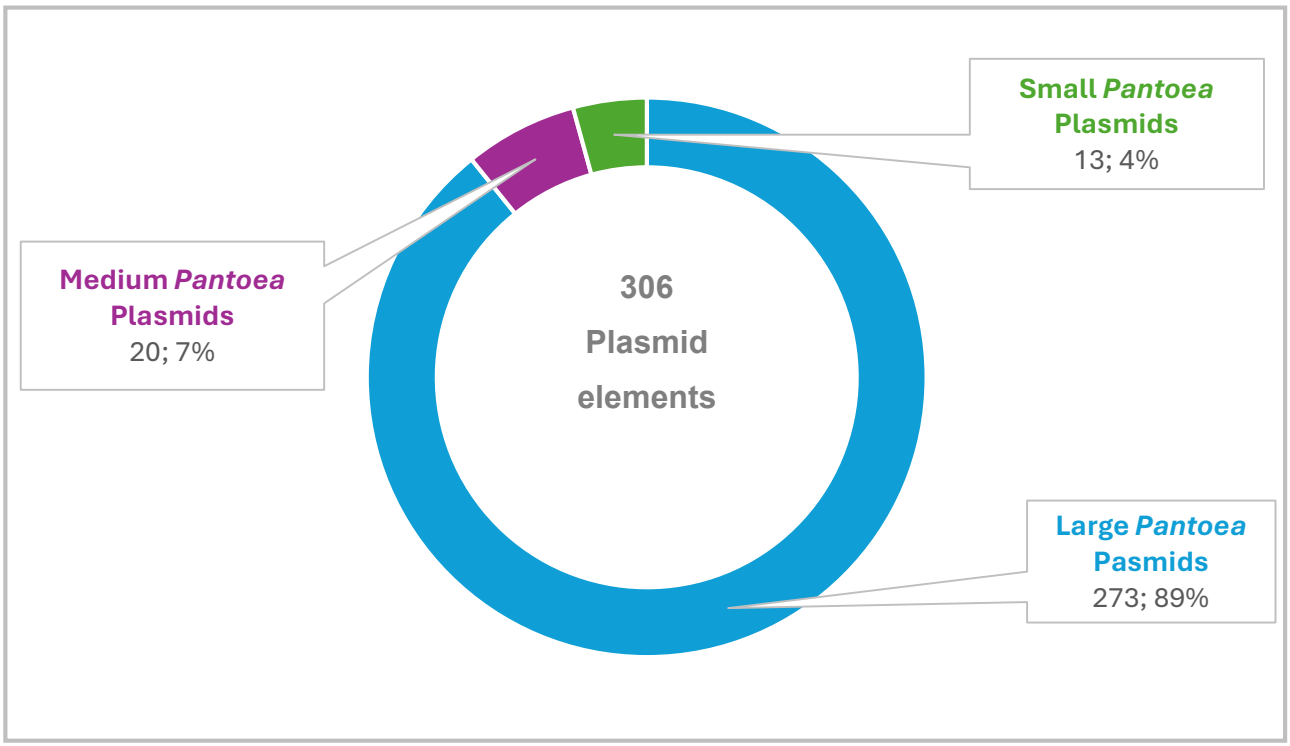


Figure 3.9: The distribution of plasmids, according to their size categories, in *P. agglomerans*. A total of 306 plasmid elements were identified in the 100 *P. agglomerans* genomes. This was further categorised into size with 273/306 being LPPs (blue), 20/306 MPPs (purple) and 13/306 SPPs (green).

Only a single plasmid, LPP-1, is conserved among all 100 compared *P. agglomerans* strains (Table 3.6). This plasmid element was also found in the genomes of the two outgroup taxa. The analysis of twenty *Pantoea* strains, spanning over seven species, indicated that LPP-1 is conserved within the genus (de Maayer *et al.*, 2012). The latter study showed that the LPP-1 plasmids ranged from 281 to 794 kb in size, carry between 238 to 750 coding sequences each and has undergone extensive

diversification, with only 2.20% of the encoded proteins being conserved across all analysed plasmids (de Maayer *et al.*, 2012). In the compared *P. agglomerans* strains, the LPP-1 plasmids were an average size of 556 kb, ranging from 460.68 kb in *P. agglomerans* NCTC10500 to 678.17 kb in *P. agglomerans* A2. A total of 1,620 orthogroups were identified across all 100 LPP-1s of *P. agglomerans*. This comprises 259 core (15.99%), 862 shared accessory (53.21%) and 499 singletons (30.80%) (Appendix D). As such, LLP-1 likely plays a key role in the diversification of *P. agglomerans*.

Table 3.6: Plasmid categories isolated from the 100 *P. agglomerans* genomes.

Plasmid	No. of strains	Plasmid size (kb)			G+C content (%)		
		Average	Minimum	Maximum	Average	Minimum	Maximum
LPP-1	100	556	460.68	678.17	53.30	52.45	53.71
LPP-2	97	181.91	129.01	220.09	52.63	51.62	53.52
LPP-3	28	144.69	107.59	229.91	52.89	50.05	56.52
LPP-4	20	97.44	59.03	205.25	51.80	50.19	53.80
LPP-5	10	146.43	111.02	205.53	50.22	48.83	51.98
LPP-6	4	138.48	130.01	141.56	49.75	49.66	49.81
LPP-7	3	101.43	86.85	109.33	53.66	53.32	54.27
LPP-8	2	182.74	182.74	182.74	48.38	48.38	48.38
LPP-9	2	142.08	101.43	182.74	51.02	48.38	53.66
LPP-10	2	124.87	110.25	139.50	50.30	49.61	50.99
LPP-11	2	110.15	110.15	110.15	47.83	47.83	47.83
LPP-12	1	152.52	-	-	48.87	-	-
LPP-13	1	142.67	-	-	51.81	-	-
LPP-14	1	121.14	-	-	50.34	-	-
MPP-1	3	553.45	48.78	59.92	51.41	50.40	52.11
MPP-2	2	64.67	64.60	64.73	50.48	50.48	50.48
MPP-3	2	40.56	40.56	40.56	49.20	49.20	49.20
MPP-4	2	38.09	38.09	38.09	44.92	44.91	44.92
MPP-5	2	35.19	35.18	35.21	46.48	46.47	46.49
MPP-6	2	31.87	28.15	35.60	45.03	43.79	46.27
MPP-7	1	49.69	-	-	46.29	-	-
MPP-8	1	45.54	-	-	49.97	-	-
MPP-9	1	51.20	-	-	51.60	-	-
MPP-10	1	66.48	-	-	50.41	-	-
MPP-11	1	64.61	-	-	52	-	-
MPP-12	1	61.48	-	-	48.93	-	-
MPP-13	1	35.69	-	-	47.30	-	-
SPP-1	2	4.90	4.40	5.39	47.20	45.48	48.89
SPP-2	1	1.19	-	-	49.61	-	-
SPP-3	1	8.67	-	-	54.97	-	-
SPP-4	1	7.78	-	-	46.72	-	-
SPP-5	1	6.88	-	-	50.52	-	-
SPP-6	1	5.71	-	-	52.63	-	-
SPP-7	1	5.39	-	-	44.73	-	-
SPP-8	1	14.55	-	-	34.15	-	-
SPP-9	1	3.40	-	-	47.54	-	-
SPP-10	1	2.97	-	-	38.31	-	-
SPP-11	1	2.86	-	-	48.58	-	-
SPP-12	1	2.55	-	-	46.55	-	-

The second most prevalent plasmid, LPP-2, is present in 97/100 *P. agglomerans* strains (Table 3.6, Appendix D). The three isolates without an LPP-2 plasmid present are *P. agglomerans* NC, NCTC10500 and T1 (Appendix D). This plasmid has an average size of 181.91 kb and ranging from 129.01 kb (2.69% of total genome) in *P. agglomerans* Pa21-15 to 220.09 kb (4.62% of the total genome) in *P. agglomerans* RIT710. Other LPP-type plasmids were less distributed, with LPP-3, LPP-4 and LPP-5 present in the genomes of 30, twenty and ten strains, respectively (Table 3.6). The remaining LPP subcategories were less prevalent, with three plasmids, namely LPP-12, LPP-13 and LPP-14 restricted to a single strain, namely *P. agglomerans* TH81, 9Rz4 and CFSAN033090, respectively (Appendix D). The MPP and SPP plasmids also show a more restricted distribution. MPP-1 is found in three strains (9Rz4, PaVv1 and ZBG6), while MPP-2 (B1 and B3), MPP-3 (B1 and B3), MPP-4 (CFBP13600 and NBBC-01), MPP-5 (B1 and B3) and MPP-6 (BI3 and Pa39-23) are observed in two *P. agglomerans* strains each. The remaining seven MPP plasmids are restricted to a single strain. Of the thirteen SPP plasmids, only one, SPP-1, is found in two *P. agglomerans* strains, namely Pa39-23 and ZBG6.

Notably, some plasmids are restricted in distribution to either *P. agglomerans* subspecies 1 or 2. LPP-8 and LPP-11 were found in the genomes of the same two *P. agglomerans* subsp. 1 strains, namely B1 and B3. Similarly, LPP-10 was also found in two *P. agglomerans* subsp. 1 genomes, namely A2 and SL1_M5. The LPP-6 and LPP-7 plasmids are restricted to *P. agglomerans* subsp. 2 and were found in four (CFBP8756, CFBP8786, CFBP13505 and CFBP13731) and three (CFBP8756, CFBP13731 and UAEU18) strains, respectively. The twenty MPP plasmids were predominantly found in *P. agglomerans* subsp. 1, with only MPP-9 being in the genome of a single *P. agglomerans* subsp. 2 strain (20TX0122) (Appendix D). The thirteen SPP plasmids were each found in the genome of a single *P. agglomerans* strain, with only SPP-1 occurring in two *P. agglomerans* subsp. 1 strains (Table 3.6). A total of five SPP plasmids were restricted to *P. agglomerans* subsp. 2, with SPP-12 and SPP-6 found on the genome of *P. agglomerans* 20TX0122. The remaining six SPP plasmids (SPP-2, 3, 4, 7, 9 and 10) were found on the genomes of individual *P. agglomerans* subsp. 1 strains.

A recent study on *P. agglomerans* strains associated with onions identified two phylogroups, with those causing severe onion symptoms in phylogroup II. This was linked to the HiVir gene cluster (Shin *et al.*, 2025), comprising seventeen genes (*hvrBCDEFGHKJKLMNOPQR*) which are involved in the biosynthesis of pantaphos (2-hydroxy[phosphono-methyl]-maleate) (Asselin *et al*., 2018; Polidore *et al.*, 2021; Shin *et al.*, 2025). This potent phosphonate phytotoxin is responsible for the hallmark onion centre rot lesions cause by *P. ananatis* (Asselin *et al*., 2018; Polidore *et al.*, 2021), as well as *P. agglomerans* (Shin *et al.*, 2025). In *P. agglomerans*, this gene cluster was observed to be encoded on two plasmids/plasmid groups, pAggl and pOnion. The latter, pOnion, also harbours two

gene clusters, the eleven gene *alt* cluster (*altBCADERGHI-gorB-altJ*) and four gene *cop* cluster (*copABCD*) which provide resistance to the onion thiosulfinate, allicin and copper, respectively (Shin *et al.*, 2025; Stice *et al.*, 2020) . Blastn searches with the HiVir, *alt* and *cop* gene clusters against the 100 *P. agglomerans* genomes, showed that the HiVir locus is indeed located on two plasmid groups in only three *P. agglomerans* subsp. 2 strains which are isolated from plant hosts, namely the LPP-3 plasmid of *P. agglomerans* CB1 and AR1a, and the LPP-2 plasmid of *P. agglomerans* 20TX0122.

The LPP-9 plasmid was only observed in *P. agglomerans* 4188 (*Betae vulgaris*-specific) and 824-1 (*Gypsophila paniculata*-specific*)* and represents the pPATH plasmid, which plays a key role in phytopathogenesis in these host-specific strains (Manulis and Barash, 2003). These strains have been shown to use a Type III secretion system, encoded on the pPATH plasmid to secrete effector proteins which drive host-specific symptom development, mainly the formation of galls around the crown region of the stem (Manulis and Barash, 2003). Besides phytopathogenesis, genes on the pPATH plasmid has also been linked to phytohormone (IAA and cytokinins) production in the analysed *P. agglomerans* strains (Barash and Manulis-Sasson, 2007; Geraffi *et al.*, 2023; Weinthal *et al.*, 2007) .

3.5.2 Integrated bacteriophage elements

Bacteriophages, viruses that replicate within bacteria, are acquired through transduction and can either be lytic (destroy the host cell) or lysogenic (integrate its DNA into the host genome) (Carr *et al.*, 2021). Integrated bacteriophages, forming prophage elements, drive genome-wide evolution and increase the diversity and divergence of strains within a species (Braga *et al.*, 2018; Moura De Sousa *et al.*, 2021), with strong selective pressure on the host phenotype (Somerville *et al.*, 2022). As such, bacteriophages have been shown to confer a range of phenotypes, including resistance to antibiotics (Pfeifer *et al*., 2022). Bacteriophages have a limited host range, and as of 2023 there are nineteen *Pantoea*-infecting bacteriophages with complete genomes available, twelve of which are in association with *P. agglomerans*. These bacteriophages range from 36.79 to 149.91 kb in size and have a G+C content of 31 to 55.33% (Zrelovs *et al.*, 2023). The *P. agglomerans* LS5-2 genome revealed three new *Pantoea* prophage elements (Nifs112, Nufs112 and Naffs113), which are the first to be isolated from insect-associated *Pantoea* species (Zrelovs *et al.*, 2023).

A total of 226 prophages were identified in the genomes of 94/100 *P. agglomerans* strains using the PHASTEST server (Wishart *et al.*, 2023), with an average of 2.40 prophages per genome, coding for an average of 95.61 proteins (Appendix D). These prophages are predominantly integrated in the chromosome, but also in some plasmid elements (Appendix D). Overall, there were six strains (*P. agglomerans* MR19, Pa39-1, PMG056, R1, T1 and Tx10) for which no prophages were predicted. The most phage elements were observed in *P. agglomerans* Pa31-3, which harbours six distinct

prophage elements in its genome, coding for 208 proteins. The prophage elements code for as few as twenty proteins, as in *P. agglomerans* MR10 (harbouring one 12.50 kb prophage element - *Klebsiella* phage ST13_OXA48), to as many as 295 proteins in *P. agglomerans* SL1_M5.

Prophages contribute between 0.25% (*P. agglomerans* MR10; 12.50 kb) and 4.77% (*P. agglomerans* SL1_M5; 236 kb) of the total genomic DNA, with an average contribution of 1.62% (79.33 kb) across the 94 *P. agglomerans* strains (Table 3.7). This is similar to the outgroup taxa, *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T, where prophage elements contribute 1.68% and 1.54% of total genomic DNA, respectively (Appendix D).

Table 3.7: Overall contribution of prophages to *P. agglomerans* genomes.

	Size (kb)	Proportion of Genome (%)	Strain
Average	79.33	1.62	
Minimum	12.50	0.25	MR10
Maximum	236.00	4.77	SL1_M5

Phage elements were further identified on the basis of gene conservation against the PHASTEST server (Wishart *et al.*, 2023). A total of 45 distinct prophage elements were identified across the *P. agglomerans* genomes. Of these prophages, three occurred the most frequently, contributing 42.92% of the total prophage elements (Figure 3.10; Appendix D). The most common phages were *Erwinia* phage ENT90 (NC_019932; 46/226), *Klebsiella* phage ST13_OX48 (NC_049453; 27/226) and *Salmonella* phage 118970_sal3 (NC_031940; 24/226). The *Erwinia* phage ENT90 was found in 41 strains, with five strains having two copies of the prophage element, whereas *Klebsiella* phage ST13_OX48 and *Salmonella* phage 118970_sal3 occur in each strain (27 and 24, respectively) only once.

Extensive variability in the size and gene content of the prophage elements was observed (Table 3.8). The largest prophage element (*Enterobacter* phage BP_4795; 62.60 kb) was observed in *P. agglomerans* OVA07A (1.3% of total genomic DNA content; 67 proteins). This prophage element is ~14x the size of the smallest prophage element, *Erwinia* phage ENT90, integrated in the *P. agglomerans* CCNPK873 chromosome (4.50 kb in size and coding for eight proteins) (Table 3.8). The genomes of *P. agglomerans* B1 and B3 housed the prophage element that encoded the most proteins (*Cronobacter* phage ENT47670; 84 proteins), as well phage element *Erwinia* phage vB_EhrS_59 in *P. agglomerans* CFBP 13569 (84 proteins) (Appendix D).

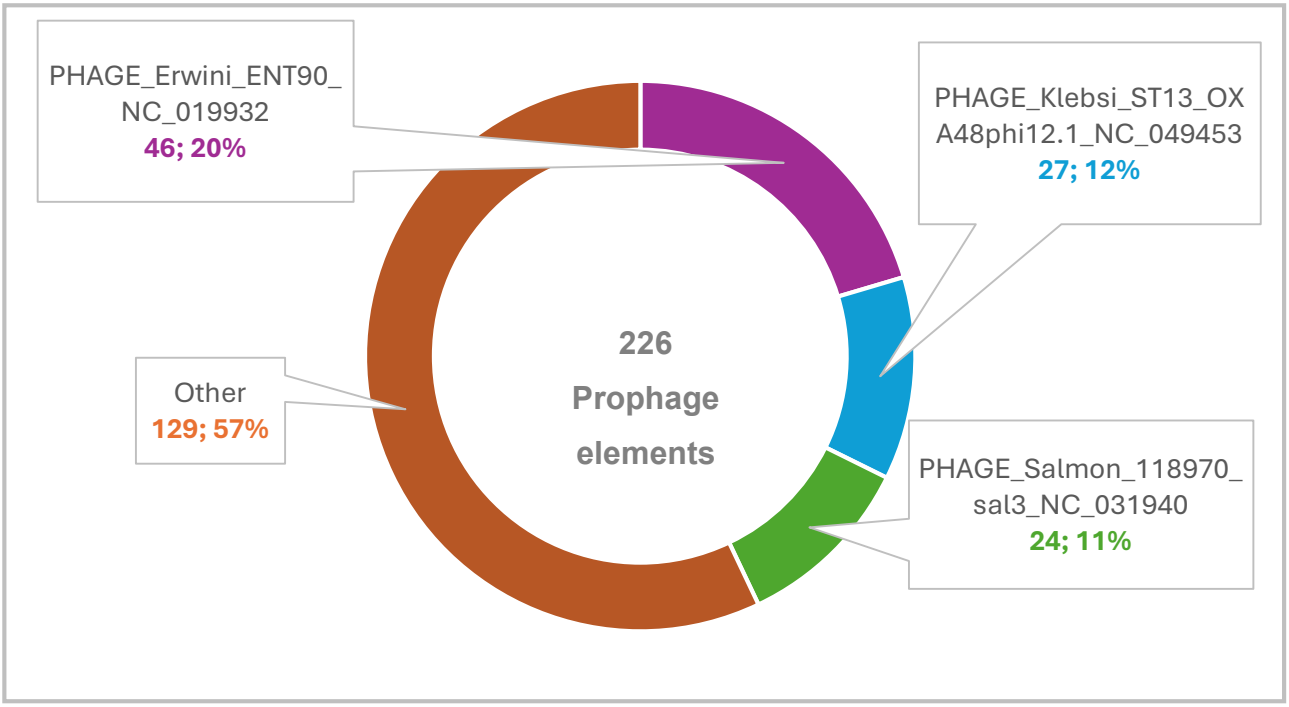


Figure 3.10: Distribution of prophage elements identified in the *P. agglomerans* genomes. Analysis of the individual prophage elements revealed that the three most frequently occurring phages (42.92% of all prophages) were *Erwinia* phage ENT90 (46/226) (purple), followed by *Klebsiella* phage ST13_OXA48 (27/226) (blue) and *Salmonella* phage 118970_sal3 (24/226) (green).

Table 3.8: Individual prophages in *P. agglomerans* genomes.

	Size (kb)	Prophage element	Proportion of Genome (%)	Strain
Average	33.00		0.67	
Minimum	4.50	*Erwinia* phage ENT90	0.09	CCNPK873
Maximum	62.60	*Enterobacter* phage BP_4795	1.30	OVA07A

3.5.3 Integrative and Conjugative Elements

Integrative and conjugative elements (ICEs) are mobile genetic elements that integrate into the host genome and mediate the transfer of DNA to other cells through functional conjugation systems (Johnson and Grossman, 2015). They remain passive in the host chromosome, undergo vertical transmission and are activated when they are excised from one chromosome, transferred and integrated into another host chromosome (McKeithen-Mead and Grossman, 2023). ICEs contain three core modules that are essential for their functioning. This includes integrase genes, which ensure the effective integration of ICEs into the host chromosome, Type IV secretion systems (T4SSs) that are necessary for transport between the donor and recipient cell as well as regulatory proteins for the effective vertical transmission within the host chromosome (de Maayer *et al.*, 2015). ICEs frequently carry cargo genes that influence the various phenotypes in the bacterial hosts including pathogenesis, antibiotic and heavy metal resistance and secondary metabolite production (de Maayer *et al.*, 2015).

The transfer of these beneficial advantages to the host microbial cell impacts the ecological success of both Gram-negative and Gram-positive bacteria (Delavat *et al.*, 2017). ICEs were identified as prominent features in *P. ananatis*, where they carry genes involved in stress response and antibiosis (de Maayer *et al.*, 2015).

A total of 55 ICE regions were identified in the 100 comparator *P. agglomerans* using ICEfinder (Wang *et al.*, 2023), where they were observed in the genomes of 41/100 *P. agglomerans* strains and were absent from the genomes of both outgroup taxa. On average, the *P. agglomerans* ICEs are 83 kb in size (contribute 1.7% of the total genomic DNA content), but range in size from 40.37 kb (*P. agglomerans* DAPP-PG734; 0.8% of genomic DNA content) to 235.65 kb (*P. agglomerans* DAPP-PG734; 4.37% of genomic DNA content) (Table 3.9; Appendix D). These elements code for an average of 48.40 proteins, with the large ICE of *P. agglomerans* DAPP-PG734 coding for 240 proteins. Of the 41 *P. agglomerans* strains that have ICEs, eleven strains have two ICE regions in each, *P. agglomerans* DAPP-PG734 has 3 ICEs and the remaining twenty-nine strains each have one ICE. ICEs are more prevalent in *P. agglomerans* subsp. 2 strains (12/22 strains; ~ 55%) than *P. agglomerans* subsp. 1 strains (29/78 strains; ~ 37%).

Table 3.9: Overview of ICEs characteristics in *P. agglomerans*.

	Size (kb)	Strain	Proportion of Genome (%)	Number of proteins encoded	Strain
Average	83.06		1.70	48.40	
Minimum	40.37	20TX0122	0.80	42	E325-699, E325-754
Maximum	235.65	DAPP-PG734	4.37	240	DAPP-PG734

ICEfinder (Wang *et al.*, 2023) indicated the presence of genes on the ICEs conferring a wide range of phenotypes among the *P. agglomerans* strains. Out of the ICE positive strains, the ICEs of fourteen *P. agglomerans* were predicted to confer resistance to a range of heavy metals, including copper, arsenic and silver. Copper is frequently used as pesticide, for example on onion fields (Tho *et al.*, 2019), while silver is used in therapeutic strategies due to its antimicrobial properties (Li and Xu, 2024). By contrast, arsenic resistant phyto-associated bacteria can prevent the morphological, physiological and biochemical effects of this heavy metal on its plant host (Abbas *et al.*, 2018).

The ICE of *P. agglomerans* BAV2934 is predicted to encode proteins involved in zinc and nickel uptake, which is an important virulence factor in bacterial pathogens (Maier and Benoit, 2019). The ferric aerobactin receptor and siderophore import system is a virulence factor that enables bacteria to absorb iron, that is essential for bacterial growth, in iron-deficient environments (Li *et al.*, 2021) and was predicted on the ICE elements of three *P. agglomerans* strains (Pa39-7, Pa39-21 and Pa39-23). ICE elements of *P. agglomerans* K1 and KM1 code for Caf1M and Caf1A proteins. Caf1M and Caf1A

are predicted in *Yersinia pestis* that harbours an F1 capsule antigen, where Caf1M is a chaperone that plays a role in the posttranslational folding and secretion of F1 and Caf1A is an outer membrane protein involved in its anchoring (Zhou *et al*., 2006). As such, these proteins are important for the virulence of the human pathogen as the capsule infers high resistance to phagocyte uptake (Zhou *et al*., 2006).

An ICE of *P. agglomerans* DAPP-PG734 is predicted to carry genes for the biosynthesis of rhamnolipid biosurfactants. Rhamnolipid biosurfactants disrupt biofilm formation and adhesions, induce apoptosis and are useful in industry as they are easily degradable, less toxic, hydrophilic and tolerant to varying levels of pH, temperature and salinity (Chong and Li, 2017). The large ICE element of *P. agglomerans* DAPP-PG734 carries the genes *hrpN* and *wtsE*, which code for T3SS effector proteins. HrpN, a harpin protein, induces a hypersensitive response in plant, leading to rapid cell death (Choi *et al.*, 2013), while WtsE is essential for Stewart's wilt caused by *P. stewartii* subsp. *stewartii* in some plants, causing cell death and water-soaked, necrotic lesions on maize plants (Jin *et al.*, 2016). The VirB protein, predicted to be encoded on the ICEs of 38 *P. agglomerans* strains (excludes *P. agglomerans* CFBP8756, CFBP13731 and UAEU18), is a transcriptional activator of virulence found on the plasmid pINV in *Shigella flexneri* (Antar and Gruber, 2023).

The ICE elements of nine *P. agglomerans* strains are predicted to harbour antiphage systems such as the CBASS, Gao, Lamassu and Mokosh systems (Duncan-Lowey and Kranzusch, 2022; Millman *et al.*, 2022). These systems represent antiviral defence mechanisms that inhibit phage replication and can even kill the infected host cell (abortive infection proteins) (Georjon and Bernheim, 2023; Patel and Maxwell, 2023). This poses a constant evolutionary struggle between bacteria and bacteriophages to coexist, adding to the fluidity of their interactions (Koskella, 2020). Restriction modification defence systems (predicted on the ICE elements of five *P. agglomerans* strains) also limit foreign DNA acquisition which can further contribute to antibiotic resistance (Vasu and Nagaraja, 2013).

The *P. agglomerans* ICE elements are further predicted to confer resistance against a wide range of antibiotics including chloramphenicol, virginiamycin and bleomycin (*P. agglomerans* DAPP-PG734) as well as aminoglycosides (*P. agglomerans* Pa21-15 and Pa21-5) and macrolides (macrolide efflux system MacAB; five *P. agglomerans* strains) (Dinos, 2017; Paul *et al.*, 2023). Furthermore, ICEs on *P. agglomerans* DAPP-PG734 and strain 4 are predicted to code proteins (DdaD and DdaF) involved in dapdiamide biosynthesis, which is an antibiotic effective against the fire blight pathogen *Erwinia amylovora* (Dawlaty *et al.*, 2010). The presence of beta-glucosidase (predicted on ICEs of *P. agglomerans* MR10 and Mr19) aids in nutritional acquisition, ecological association and is a source of carbon (Magwaza *et al*., 2024). This is similar to the functions of proteins involved in sorbitol

metabolism and transport, which are predicted on the ICEs of six *P. agglomerans* (Pérez-Ramos *et al.*, 2017). Levansucrase (encoded on ICE of *P. agglomerans* Pa31-3) converts sucrose to fructooligosaccharides and levan (Martínez-Fleites *et al.*, 2005).

The wide range of phenotypes predicted to be encoded on the ICE elements in *P. agglomerans* strains, including resistance to antimicrobial compounds, virulence factors, anti-phage elements and metabolic capacities suggest that ICEs play a major role in the diversification of *P. agglomerans*, providing ICE-carrying taxa with competitive advantages.

3.5.4 Transposable elements

Transposable elements (transposons) are DNA sequences that can change position within a given genome, via transposases (Bourque *et al.*, 2018). Transposons have levels of preference for insertion in the host genome, balancing between propagation and mitigating deleterious effects on cellular functioning (Bourque *et al.*, 2018). Transposable elements were predicted using TnComp_finder (Alvarenga and Varani, 2019). On average, each *P. agglomerans* genome is predicted to encode 26.64 transposases. The genome of *P. agglomerans* 4188 encodes the most transposases, with a total of 84 distinct transposases (Table 3.10). Previous studies of other *Pantoea* species have identified 101 transposases in *P. ananatis* NN08200 and over 400 transposases in the genome of *P. stewartii* subsp. *stewartii* DC283 (Duong *et al*., 2018; Zeng *et al.*, 2020), suggesting they play a major role in the diversification of members of the genus. By contrast, the genomes of the outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, harbour only twenty-one and eight transposases, respectively (Appendix D).

Table 3.10: Predicted transposable elements in the *P. agglomerans* genomes.

	Frequency	Strain
Average	26.64	
Minimum	9	IG1
Maximum	84	4188

3.6 Global insights into the functional diversity of *P. agglomerans*

The extensive variability in genome metrics and mobilome of the compared *P. agglomerans* strains indicate extensive genome flexibility for this species, which is predicted to translate into functional diversity. Representative proteins of each orthogroup in the *P. agglomerans* core, accessory and singleton proteins were classified according to their COG categories (Galperin *et al.*, 2021) by comparison against the EggNOG database (Huerta-Cepas *et al.*, 2019) using EggNOG-mapper (Cantalapiedra *et al.*, 2021). Proteins of unknown function (COG category S) and general function prediction only (COG category R) were excluded from the analysis.

The relative proportions of proteins involved in the other categories were calculated and summarised into the COG supra-functional categories 1) Information storage & processing, 2) Metabolism and 3) Cellular processes (Figure 3.11; Appendix E) (de Maayer *et al.*, 2014). This analysis showed that Metabolism was best represented in the core genome (54.14%), while the highest proportions of proteins involved in the COG supra-functional categories Cellular processes and Information storage & processing are in the accessory (33.74%) and singleton (43.28%) pan-genome fractions, respectively (Figure 3.11). This is similar to what was observed in a pan-genome analysis of *P. ananatis* which demonstrated that the core genome comprises a greater proportion of proteins involved in metabolic and cellular processes than the accessory genome (de Maayer *et al.*, 2014).

While representatives of all 18 COG categories were observed in the core, accessory and singleton pan-genome elements, differences in their relative abundance were observed for a number of categories (Figure 3.12). Higher proportions of proteins involved in metabolic functions such as energy production (COG category C), amino acid (COG category E) and carbohydrate (COG category G) transport and metabolism occur in the core genome than the accessory genomic elements, likely due to shared central metabolic pathways. However, a number of proteins involved in carbohydrate transport and metabolism (shared accessory: 6.92%; singleton: 4.42%) and amino transport and metabolism (shared accessory: 5.99%; singleton: 6.92%) in the accessory and singleton pan-genome fractions do suggest metabolic versatility among the compared *P. agglomerans* strains.

A greater proportion of proteins involved in DNA replication, recombination and repair (L) proteins occur in the accessory genome (14.40%) and singleton (23.81%) elements than the core genome (4.58%). This can be linked to the observed differences in the mobilome, where proteins involved in plasmid and ICE mobility, phage integration and transposable elements belong to this category. Similarly, many of these mobilome element proteins are involved in intracellular trafficking, secretion and vesicular transport (COG category U) and hence a larger proportion of proteins involved in these functions are observed in the accessory (5.27%) and singleton (4.14%) pan-genome fractions than

the core genome (2.10%) Proteins involved in defence mechanisms (COG category V) occurred most frequently in the singleton genome fraction (7.82%) compared to the core (1.35%) and shared accessory (4.40%) fractions. This can be due to the sympatric nature of *P. agglomerans* in that individual strains will acquire tolerance and resistance mechanisms that allow them to tolerate antimicrobial and other stresses in their given environment. A similar result was observed in species in the family *Lactobacillaceae*, where the singleton genome fraction had a much higher proportion of proteins involved in various defence mechanisms (Rajput *et al.*, 2023).

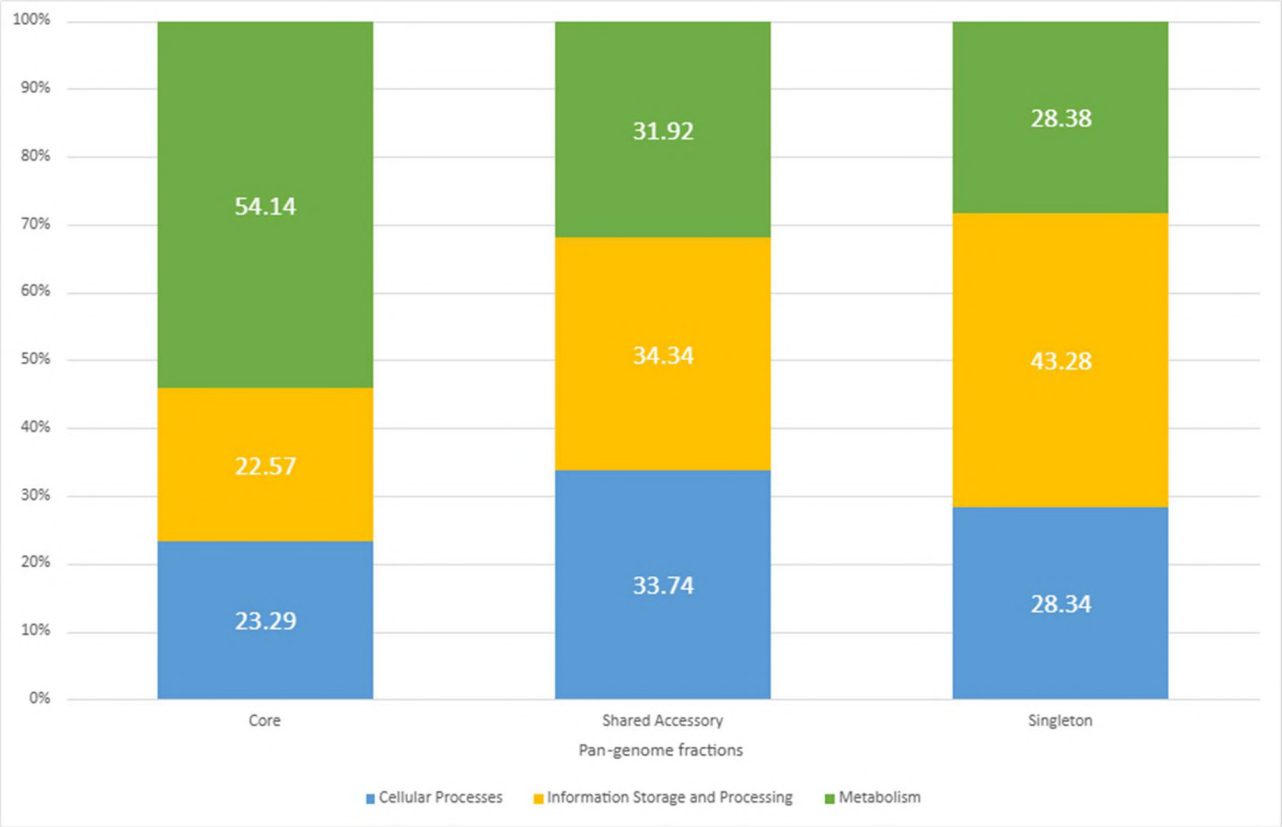


Figure 3.11: Representative proportions of proteins according to the COG supra-functional categories in *P. agglomerans* pan-genome fractions. The COG supra-functional categories, Metabolism (green), Cellular processes (blue) and Information storage & processing (yellow), was summarised from the EggNOG annotations for *P. agglomerans* across the core (left), shared accessory (middle) and singleton (right) genome fractions.

To identify gross functional differences between the two predicted *P. agglomerans* subspecies, the proteins unique to either *P. agglomerans* subsp. 1 or *P. agglomerans* subsp. 2 were categorised according to their COG functions (Figure 3.13). Similar to what was observed for the pan-genome of the entire species, the subspecies-specific fraction was dominated by proteins involved in DNA replication, recombination and repair (COG category L), contributing 20.54% and 15.62% of the categorised proteins of *P. agglomerans* subsp. 1, and subsp. 2, respectively. *P. agglomerans* subsp. 2 had higher proportions of proteins involved in intracellular trafficking, secretion and vesicular transport (COG category U; 7.37%) and posttranslational modification, protein turnover and chaperone proteins (COG category O; 4.94%) than *P. agglomerans* subsp. 1 (3.93% and 1.88%,

respectively). *P. agglomerans* subsp. 1, on the other hand, had a higher proportion of proteins involved in signal transduction mechanisms (COG category T; 4.32%) than *P. agglomerans* subsp. 2 (2.74%).

The pan-genome based functional analysis highlighted substantial functional diversification, both at the species and sub-species level. To gain further insights into pertinent phenotypes linked to *P. agglomerans*, further comparative genomic analysis of the pathogenicity determinants (with specific focus on the secretome), resistome and secondary metabolite biosynthesis pathways of the comparator *P. agglomerans* strains was performed.

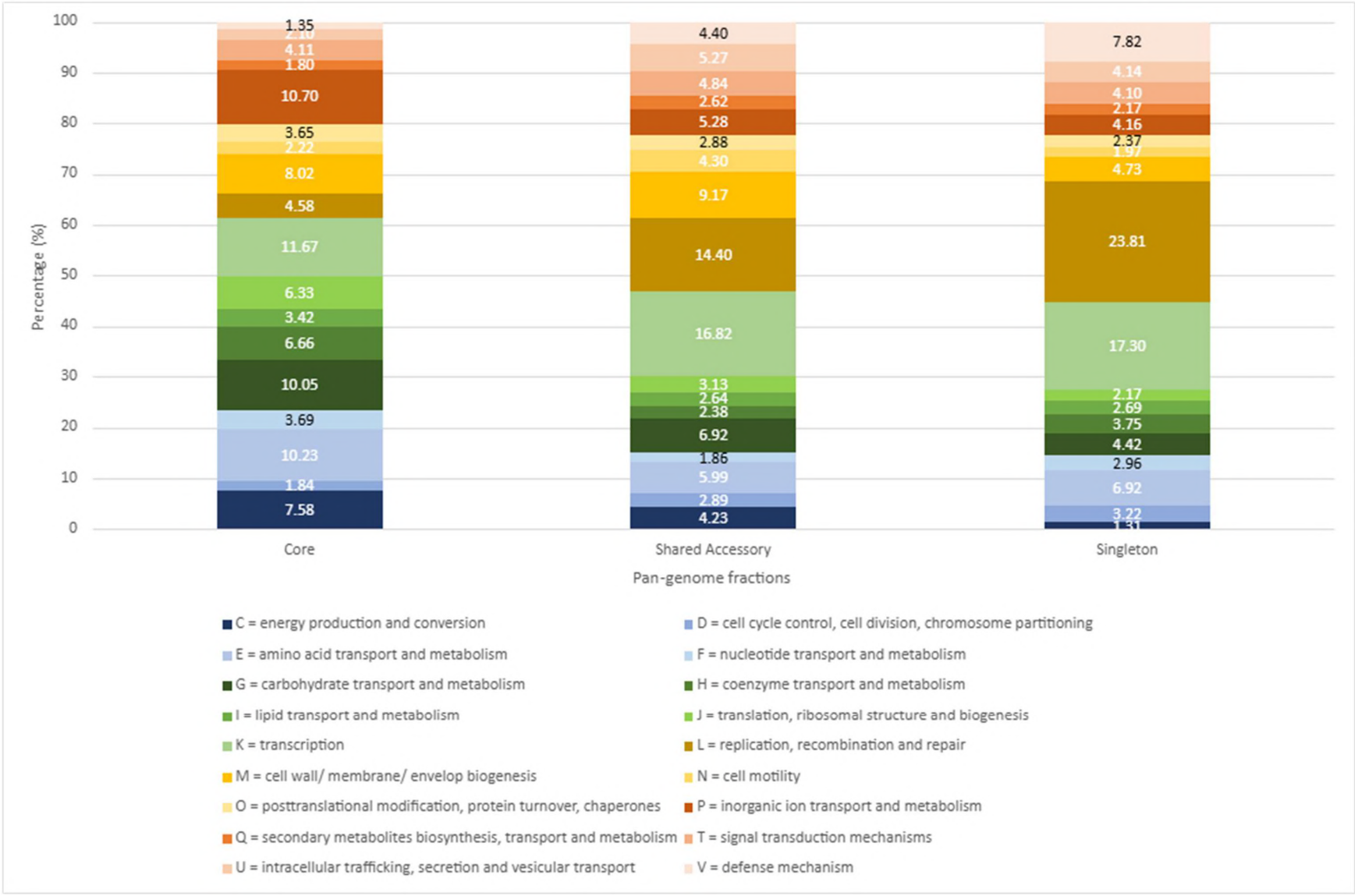


Figure 3.12: Conserved and distinct functions of the proteome of *P. agglomerans.* This was constructed on the basis of the EggNOG annotations for the core (left), shared accessory (middle) and singleton (right) genome fractions of the 100 *P. agglomerans* genomes.

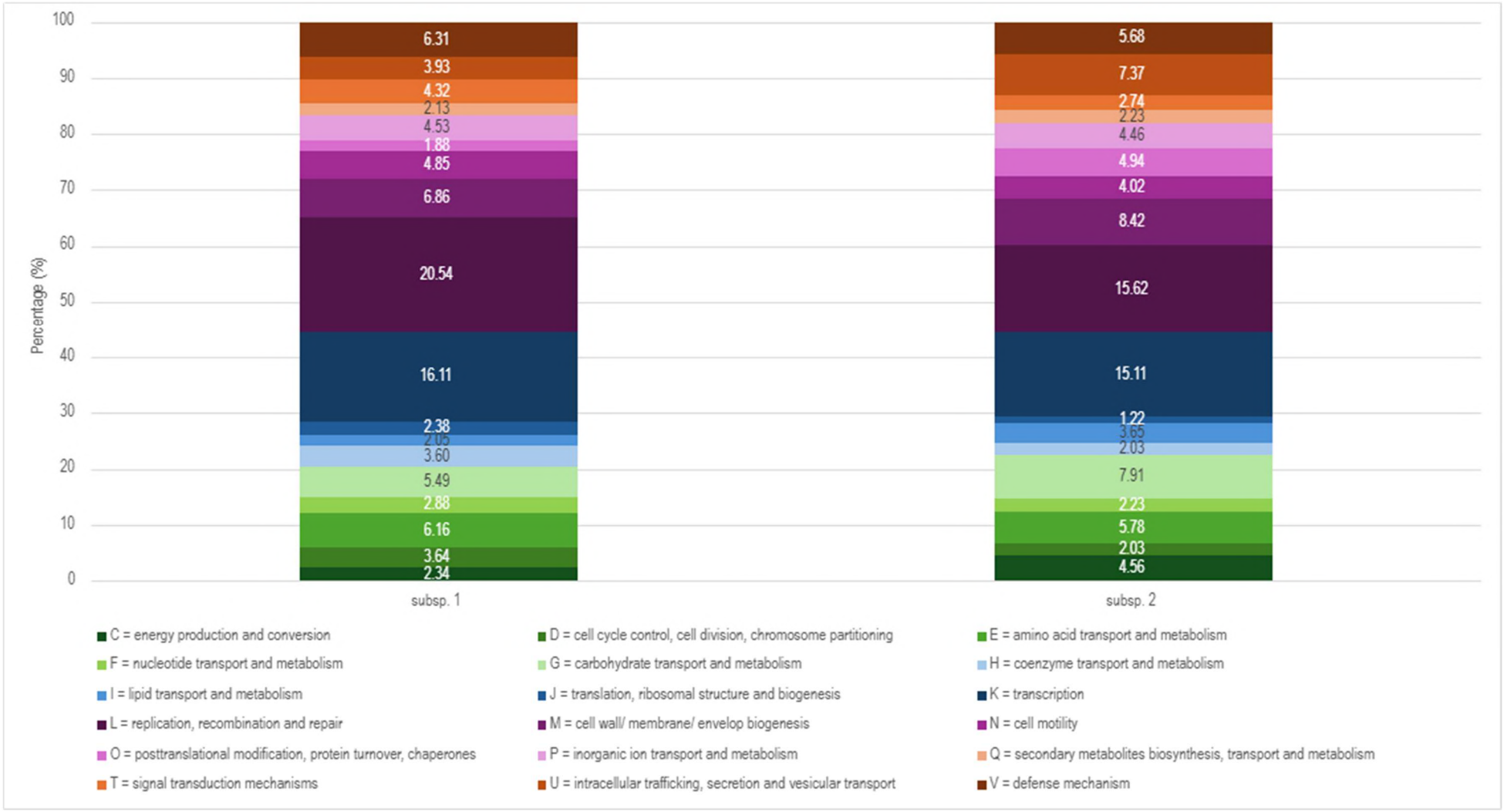


Figure 3.13: Distinct functions encoded by both *P. agglomerans* subsp. 1 and subsp. 2. This was constructed on the basis of the EggNOG annotations for the accessory genomes of *P. agglomerans* subsp. 1 (left) and subsp. 2 (right).

3.7 Genomic insights into the pathogenicity (pathogenome) of *P. agglomerans*

3.7.1 Human pathogenicity potential of *P. agglomerans*

P. agglomerans is recognised as a bacterial species with extensive biotechnological potential. However, the use of *P. agglomerans* for bioindustrial purposes is hampered by its recognition as an opportunistic human pathogen (Penner *et al.*, 2022; Tiwari and Beriha, 2015). The probability of *P. agglomerans* as human pathogen was predicted using PathogenFinder v. 1.1 (Cosentino *et al.*, 2013). This demonstrated that all 100 comparator strains, including the recognised clinical isolates *P. agglomerans* FDAARGOS 1447^{T}, NCTC10500 and Tx10, were predicted as non-pathogenic to human hosts. An average pathogen probability score of 0.30 was attained for all 100 strains, with the highest probability score for *P. agglomerans* CFBP8786 (0.36; Table 3.11). The human isolates, *P. agglomerans* FDAARGOS 1447^{T}, NCTC10500 and Tx10 displayed pathogen probability scores of 0.33, 0.32 and 0.28, respectively (Appendix F).

These results are consistent with those obtained for non-human pathogens, such as *Lactiplantibacillus plantarum* ZBK1-5 and 299V, which had pathogen probability scores of 0.20 and 0.19, respectively (Choovet *et al.*, 2024). By contrast, known human pathogens such as *E. coli* O11:H12 and *Citrobacter werkmanii* Ak-8 displayed pathogen probability scores of 0.92 and 0.87, respectively (Parvez *et al.*, 2020; Yibar *et al*., 2024). These pathogens also matched with 720 and 91 pathogenic protein families in the PathogenFinder database, respectively, with up to 100% protein coverage. By contrast, average pathogen protein coverage values of 0.91% (0.56%-1.04%) were observed for the compared *P. agglomerans* strains (Table 3.11).

A study on extended spectrum β-lactamase producing *E. coli*, isolated from pigs, suggested a pathogen probability score threshold of 0.9 for a significant probability of human pathogenicity (Founou *et al.*, 2022). The low pathogenicity probability scores and pathogenic protein family coverage of *P. agglomerans* suggests that members of this species have a low clinical pathogenicity potential, testament to its description as an opportunistic human pathogen (Penner *et al.*, 2022; Tiwari and Beriha, 2015).

Table 3.11: Probability of *P. agglomerans* strains being a human pathogen.

	Human Pathogenicity Potential		Input Protein Coverage	
	Proportion	**Strain**	**%**	**Strain**
Average	0.30		0.91	
Minimum	0.03	PA	0.56	IG1
Maximum	0.36	CFBP8786	1.04	JM1

While *P. agglomerans* may demonstrate limited pathogenic potential against human hosts, the species is renowned for its role as a plant pathogen, causing a variety of disease symptoms on a broad range of plant hosts (Dutkiewicz, Mackiewicz, Kinga Lemieszek, *et al.*, 2016). Among the 100 compared *P. agglomerans* strains, eight were isolated from diseased plants. To further elucidate on its phytopathogenicity, the proteomes for the 100 *P. agglomerans* comparators were compared against the Pathogen-Host Interaction database (PHIBase) (Urban *et al.*, 2021).

3.7.2 Phytopathogenicity of *P. agglomerans*

A total of 28,990 proteins involved in pathogen-host interaction were identified across the 100 *P. agglomerans* strains (Table 3.12). *P. agglomerans* NCTC10500, a clinical isolate, had the lowest number of PHI hits (280). The other clinical isolates, *P. agglomerans* FDAARGOS 1447^{T} and Tx10, produced PHI hits of 288 and 283, respectively, which is below the calculated average (289.90) for *P. agglomerans* strains. On the contrary, the highest number of PHI hits was 301 in the plant epiphytes, *P. agglomerans* BI3 and Pa39-23.

The genomes of the eight phytopathogenic strains harboured genes encoding a total of 2,328 PHI proteins (average = 291) (Appendix F). Phytopathogens causing *Allium cepa* bulb/centre rot (*P. agglomerans* 20TX0122, AR1a, CB1) attained 285, 289 and 287 PHI hits, respectively. The walnut blight and leaf rot pathogenic strains *P. agglomerans* CFSAN033090 and *P. agglomerans* CHTF15 generated 290 and 291 PHI hits, respectively, while the highest number of hits were observed for *P. agglomerans* 4188 (PHI = 299) and 824-1 (PHI = 298), which form galls on beets and *Gypsophia*, respectively.

Table 3.12: Presence of known pathogenicity determinants in *P. agglomerans*.

	Pathogen-Host Interactions	
	Frequency	**Strain**
Average	289.90	
Minimum	280	NCTC10500
Maximum	301	BI3, Pa39-23

Among the 341 distinct pathogenicity determinants (PHI proteins), 249 (73.02%) were shared by all *P. agglomerans* strains and twenty (5.87%) were restricted to a single strain. Some of these unique PHIs (eight; 2.35%) had multiple copies in a strainwith functions relating to type VI secretion systems, cold shock response, chemotaxis and protein biosynthesis. Within the identified PHI proteins, *P. agglomerans* subsp. 1 and subsp. 2 were predicted to have twenty-one and fourteen distinct pathogenicity determinants, respectively.

Of the twenty-one distinct PHI proteins unique to *P. agglomerans* subsp. 1, none were conserved across all 78 strains, with the most frequent being the acyltransferase PagP predicted in 25/78 strains.

This enzyme, encoded on the phage vB_PagP-SK1, plays a role in the transfer of palmitate to the lipid A component of the outer membrane lipopolysaccharide layer (LPS), enhancing *P. agglomerans* resistance to antimicrobial peptides and contributing towards virulence (McDougall *et al.*, 2020). The arabinose transferase ArnT (unique to *P. agglomerans* E325-699), transfers a sugar group from a lipid carrier to the LPS lipid A component, inferring resistance to polymxin antibiotics (Petrou *et al.*, 2016). Among the *P. agglomerans* subsp. 1-unique PHI proteins, six were predicted in the genome of *P. agglomerans* BI3 only. This included five proteins encoded by the gene cluster *iroBCDEN* (Puorger *et al.*, 2011; Zhu *et al.*, 2005). This gene cluster codes for salmochelin (IroB protein), catecholate siderophore esterases (IroD and IroE proteins), Ton-B-dependent outer membrane siderophore receptor protein (IroN) and the ABC transporter (IroC protein). Together, these proteins are involved in the glucosylation of enterobactin and the uptake of iron-rich salmochelins (Zhu *et al.*, 2005). Thirteen *P. agglomerans* subsp. 1 strains housed sorbitol-6-phosphate dehydrogenases (SrlD) and the type II sorbitol transport complexes (SrlA and SrlE). This enables bacteria to utilise sorbitol which would normally be used for carbohydrate transport in host plants (Salomone-Stagni *et al.*, 2018). *P. agglomerans* B1 and B3 each included a catalase (KatN) which was previously proven to increase hydrogen peroxide resistance in *Salmonella* (Robbe-Saule *et al.*, 2001).

Similar to *P. agglomerans* subsp. 1, there are no PHI proteins conserved across all fourteen *P. agglomerans* subsp. 2 strains. The genome of *P. agglomerans* 4188 harbours three of the fourteen *P. agglomerans* subsp. 2-unique PHI proteins. This included RovA, a transcriptional regulator of virulence genes (Ellison and Miller, 2006), as well orthologues of the HopQ1 and HopR1 effector proteins. These T3Es enhance bacterial virulence in *Pseudomonas syringae* pv. *tomato* by interacting with host proteins, affecting the immune response of the host plant and facilitating bacterial survival and proliferation (Kvitko *et al.*, 2009; Li *et al.*, 2013).

Pathogenic strains can also adapt to the host environment and resist host immunity due to macrophage subversion factor MgtC (Belon *et al.*, 2015). This protein (predicted in the *P. agglomerans* subsp. 2 strains CFBP 13731, CFBP 8756 and CFBP 8786) improves pathogen survival and replication within macrophages by enabling bacterial adaptation to the low magnesium concentration of macrophages (Belon *et al.*, 2015). The Toll/IL1 receptor proteins, specifically TcpY1 in *P. agglomerans* DAPP-PG734 and RSO7, has previously been shown to be associated with immune systems of plants, animals and bacteria (Li *et al.*, 2023).

3.7.3 Secretion systems in *P. agglomerans*

The comparison against the PHI database highlighted the role of several secreted enzymes in *P. agglomerans* pathogenesis. The subcellular localisation of the *P. agglomerans* proteome was determined using PsortB (Yu *et al.*, 2010). The majority of the total 448,314 *P. agglomerans* proteins are localised in the cytoplasm (40.92%), while 25.78%, 3.74% and 1.75% are localised in the cytoplasmic membrane, periplasmic space and outer cell wall, respectively (Figure 3.14). Furthermore, 1.08% of the proteins are predicted to be secreted extracellularly, while 26.73% of the proteins have an unspecified localisation (Figure 3.14). The number of extracellular proteins ranged between 33 (1.30% of proteome) and 61 proteins (0.80% of the total proteome) in *P. agglomerans* 824-1 and T1, respectively.

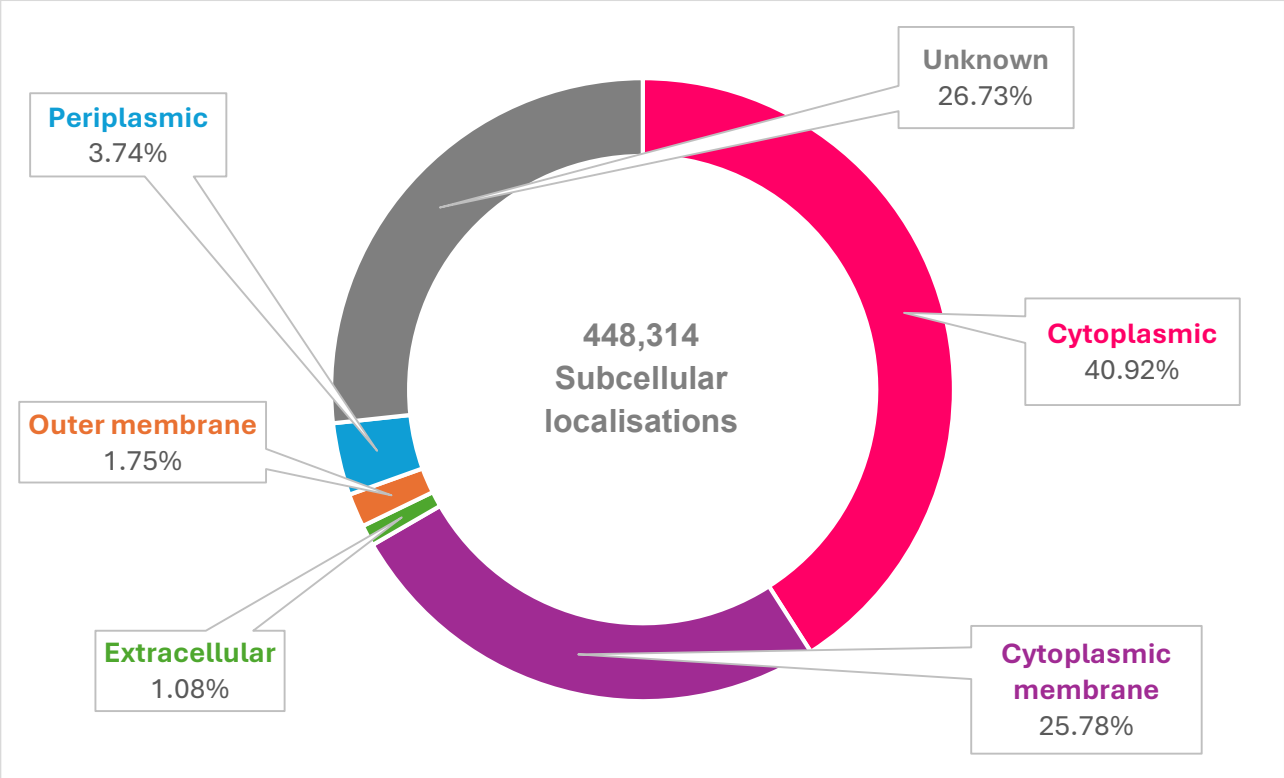


Figure 3.14: Relative proportions of proteins in different subcellular localisations in the *P. agglomerans* cell. The subcellular localisations of protein were distributed across the cytoplasm (pink), cytoplasmic membrane (purple), outer membrane (orange), periplasmic space (blue) and extracellular space (green) with 26.73% being unknown (grey).

Albeit in relatively low proportions, the presence of extracellular proteins in the *P. agglomerans* genomes implies that they possess secretion systems, which serve as mechanisms for their extrusion (Table 3.13). These secretion systems enable *P. agglomerans* to interact with the external environment in the broad ecological niches they occupy, extracting nutrients and minerals, extruding wastes and coping with both biotic and abiotic stresses as well as interact with the hosts with which they are associated (Ebner and Götz, 2019; Ellory and Hewitt, 2011; Garcia-Garcera and Rocha, 2020; Molina-Santiago *et al.*, 202; Rollauer *et al.*, 2015).

Table 3.13: The relative proportions of subcellular localisations in the proteome of *P. agglomerans*.

	Cytoplasmic		Cytoplasmic membrane		Outer Membrane		Periplasmic		Extracellular	
	%	Strain	%	Strain	%	Strain	%	Strain	%	Strain
Average	40.94		25.79		1.75		3.75		1.08	
Minimum	38.89	C410P1	24.76	B3	1.62	CFSAN033090	3.34	824-1	0.80	T1
Maximum	42.38	T1	26.76	PA	1.96	KM1	4.02	T1	1.30	824-1

A total of 1,141 secretion systems were identified across the 100 *P. agglomerans* genomes using Macsyfinder (Néron *et al.*, 2023). The secretion systems belonged to seven distinct types, namely the Type I, Type III, Type IV, Type V, Type VI (TXSS), Type IV pilus and flagellar secretion systems (Figure 3.15). Of these different types, five were observed in 60/100 of the strains, while 38 strains harbour six secretion system types. Only two *P. agglomerans*, namely *P. agglomerans* BI3 and Pa39-23, harbour all seven secretion system types. The genomes of the outgroup taxa, *P. eucalypti* LMG 24197^T and *P. vagans* LMG 24199^T, harbour six and eleven secretion systems across five of the seven secretion system types, respectively.

Type I secretion systems (T1SSs) are the most frequently occurring in *P. agglomerans*, at an average of 2.96 T1SS encoded per genome. A total of 296 T1SSs were identified in the *P. agglomerans* genomes, with 238 and 58 in *P. agglomerans* subsp. 1 (average 3.05/genome) and subsp. 2 (2.64/genome), respectively (Figure 3.15). Only a single T1SS was identified in the genomes of five *P. agglomerans* strains, while twenty strains harboured four distinct T1SS. The T1SS is comprised of three large proteins, an ATP binding cassette (ABC) transporter, an inner membrane fusion protein (MFP) and an outer membrane protein (OMP) (Anlauf *et al.*, 2024). The substrate protein, in unfolded state, binds to the ABC transporter, passes through the periplasmic space via the MFP and subsequently the exit channel provided by the OMP, before folding to its native state (Spitz *et al.*, 2019). The substrates include a broad range of different proteins and include adhesins, lipases, proteases, pore-forming toxins and iron-scavenger proteins, which play various roles in nutrient acquisition and virulence (Anlauf *et al.*, 2024; Spitz *et al.*, 2019) .

Type III secretion systems (T3SSs) were the least common secretion system carried on the *P. agglomerans* genomes of both subspecies, with 39 T3SS distributed across 32 and seven *P. agglomerans* subsp. 1 and subsp. 2 strains, respectively (Figure 3.15). T3SSs play various roles in virulence as they inject effector proteins directly into the cytoplasm of a range of eukaryotic hosts (Deng *et al.*, 2017; Wagner *et al.*, 2018). This is achieved through a basal body (with a cytosolic ATPase complex and cytoplasmic ring), a needle complex and a translocation pore (Deng *et al.*, 2017). T3SS influence bacterial communications and interactions with both plant and animal hosts (Puhar and Sansonetti, 2014). Type III effector proteins (T3Es), which include effectors that disrupt normal cellular functions, leading to symptom development, enable colonisation by altering the host

cell's cytoskeleton and signalling pathways, as well as T3Es that allow them to respond to environmental signals and adapt to different host environments, thereby expanding their host range (Deng *et al.*, 2017).

T3SSs are key virulence in phytopathogenic *Pantoea* species, exemplified by *P. stewartii* subsp. *stewartii,* where T3Es are responsible for the watersoaked lesions and tissue necrosis that typified Stewart's wilt of maize (Correa *et al.*, 2012). This has been linked to the T3E WtsE, which causes vascular occlusion and the development of water-soaked lesions on maize plants (Jin *et al.*, 2016). An orthologue was found to be encoded on the ICE element in *P. agglomerans* DAPP-PG734, along with the harpin HrpN, which induces a hypersensitive response in plants (Choi *et al.*, 2013). The T3SS of *P. agglomerans* pv. *gypsophilae* and *P. agglomerans* pv. *betae* contributes to gall formation (in gypsophila and beet and gypsophila, respectively) via the T3Es HsvB, HSvG , PseB and PthG (Nissan *et al.*, 2018). HsvB and HsvG are host-specific transcription factors of pathovar *betae* and pathovar *gypsophilae*, respectively, and bind to the promoter of a host gene, inducing gall formation (Nissan *et al.*, 2019). PseB also plays a role in gall development in both pathovars, while PthG is involved in modifying the host plant response to auxin and cytokinin hormones (Nissan *et al.*, 2019).

The flagella secretion system is similar to the structure of T3SSs and is predicted to be the progenitor of the latter system (Zhu *et al.*, 2023). This multi-gene motility structure comprises motor, anchored in a basal body and located in the bacterial cell wall. The latter connects a hook to the whip-like filament which serves as an injectisome for the growing filament (Diepold and Armitage, 2015; Nakamura and Minamino, 2024). Flagellar secretion systems are present in all *P. agglomerans* strains, testament to their capacity for swimming motility (Figure 3.6) (Palmer and Coutinho, 2022). Two strains, namely *P. agglomerans* B1 and B3, harbour two flagellar secretion systems on their genomes. Comparative genomics of the order Enterobacterales revealed that 20% of the studied taxa harboured dual flagellar systems (de Maayer *et al*., 2020). Aside from the primary flagellar system (*flag-1*), some taxa also incorporate one of four other *flag* loci, which show distinct genetic organisation and are predicted to serve different functions (de Maayer *et al*., 2020). *P. agglomerans* B1 and B3 harbour both the *flag-1* and a *flag3* locus, with the function of the latter yet to be determined (de Maayer *et al*., 2020).

Type IV pili (T4P) systems are long filaments that are divided into two different subgroups, T4aP and T4bP. These pili differ in the proteins involved in their assembly and function (Roux *et al*., 2012). T4aP pili are involved in twitching motility, DNA uptake and surface attachment (Izadi-Pruneyre *et al*., 2024). T4bP pili are typically involved in other functions, including structuring of biofilms, adhesion to eukaryotic host cells and conjugative DNA transfer (Roux *et al*., 2012). Both machineries

comprise an inner membrane platform, a pilus filament, an outer membrane secretin and an ATPase, providing energy for the rapid and repeated pili extension and retraction (Silva *et al.*, 2020). All 100 *P. agglomerans* strains have T4aP, which correlatives with its prevalence in almost all bacterial phyla (Pelicic, 2023). T4bP pilus loci were only predicted in five *P. agglomerans* subsp. 2 strains (*P. agglomerans* C410P1, CFBP13731, CFBP8756, DAPP-PG734 and UAEU18), which is characterised as only being present in the phylum Proteobacteria (Pelicic, 2023). The outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, harbour one T4aP system each and no T4bP locus.

The Type IV secretion system (T4SS) facilitates the transport of DNA, proteins and DNA-protein complexes into a host (Yin *et al.*, 2019). These complexes contribute to pathogenesis and adaptations to the cellular niche in the host cell (Yin *et al.*, 2019). More specifically, the genomes of three *P. agglomerans* subsp. 1 strains (*P. agglomerans* BI3, CFSAN033090 and Pa39-23) house pT4SSt loci, a subtype T4SS involved in the secretion of proteins and is particularly pertinent in conjugative transfer of genetic material between bacterial cells, but also the delivery of effector proteins into host cells (Green and Mecsas, 2016). The function of the T4SS systems in *P. agglomerans* remains unclear.

Type V secretion systems (T5SSs) deliver receptor-binding proteins and toxins into the host cell (Green and Mecsas, 2016). These T5SSs have been subclassified into three main types (T5SSa, b and c) on the basis of the secretion system mechanism (Green and Mecsas, 2016). *P. agglomerans* harbours 444 T5SSs, across the three subtypes, namely the T5aSS, T5bSS and T5cSS. T5aSSs are classical autotransporters that secrete themselves through a translocator domain, passenger domain, linker domain and, in some cases, a protease domain (Green and Mecsas, 2016). Well-known examples of the T5aSS are the IgA protease of *Neisseria meningitidis* and the EstA lipase of *P. aeruginosa*, which are known pathogenicity determinants in these bacteria (Meuskens *et al.*, 2019). The 100 *P. agglomerans* strains harbour a total 225 T5aSSs, with between one and four T5aSS/genome. The five *P. agglomerans* strains with four T5aSS each include the *P. agglomerans* subsp. 1 strains B1, B3 and R5, as well as the *P. agglomerans* subsp. 2 strains K1 and KM1.

The transport of large proteins is carried out through the two-partner secretion systems encompassed in the T5bSS subtype. This consists of a translocation protein (TpsB) and a secreted passenger protein (TpsA) (Meuskens *et al.*, 2019). The separate translocator and passenger proteins enables the latter to be released into the host cell, after transport, without the use of proteolytic cleavage enzymes (Meuskens *et al.*, 2019). After translocation, some TpsBs stay in a non-covalent bond with the outer membrane, as with HMW1 and HMW2 in *Haemophilus influenzae*, whilst others are secreted, such as ShlA in *Serratia marcescens* (Meuskens *et al.*, 2019) A total of 218 T5bSS are incorporated in the *P. agglomerans* genomes (between one and four T5bSS/genome), with 164 of these in *P. agglomerans*

subsp. 1 and 54 T5bSSs in *P. agglomerans* subsp. 2. The majority of strains incorporate two T5bSSs on their genomes, while three strains (*P. agglomerans* PA, 4188 and ZJU23), in *P. agglomerans* subsp. 2, incorporate four distinct copies of this system.

T5cSSs are trimeric autotransporter adhesins that have the same topology as T5aSSs, with the exception of forming trimeric structures (Da Costa *et al.*, 2024; Meuskens *et al.*, 2019). This subtype is common in Gram-negative bacteria with the most widely studied being the YadA adhesin of *Yersinia enterocolitca* and *Yersinia pseudotuberculosis* (Meuskens *et al.*, 2019). Other adhesins have also been studied in *Bartonella henselae* (BadA), *Salmonella typhimurium* (SadA and SadB) as well as *H. influenzae* (Hia and Hsf) (Luo *et al.*, 2023). Of the 100 comparator *P. agglomerans* strains, *P. agglomerans* Pa39-23 was the only strain to harbour a T5cSS.

Type VI secretion systems (T6SSs) are contact dependent and play a role in bacterial communications and environment interactions (Habich *et al.*, 2025; Unni *et al.*, 2022). The delivered effectors can kill or inhibit fungal and bacterial cell growth, manipulate eukaryotes and influence nutrient uptake (Unni *et al.*, 2022). T6SSs provide a competitive survival advantage as they deliver toxic proteins to adjacent cells and share structural similarities with the puncturing devices identified in bacteriophages (Crisan and Goldberg, 2022; Singh and Kumari, 2023). Comparative genomics of *Pantoea* and *Erwinia* species revealed that strains harbour between two and four T6SSs per genome, with T6SS-1 and T6SS-2 being conserved across all studied genomes of both genera (de Maayer *et al.*, 2011). *P. agglomerans* pv. *betae* has antibacterial T6SS that release lysozyme-like effectors to ward off competing bacteria (Carobbi *et al.*, 2022). Generally, T6SSs encompass 13 conserved proteins (TssA-M) which form a membrane complex, a baseplate with a needle spike and sheath (Unni *et al.*, 2022). T6SSs are classified into three main types with T6SSi harboured mainly in *Proteobacteria*, T6SSii in *Francisella* and T6SSiii in *Bacteroidetes* (Morgado and Vicente, 2022). A total of 152 T6SSi loci were identified on the genomes of the 100 *P. agglomerans* strains, with 123 and 29 of these in *P. agglomerans* subsp. 1 and 2, respectively. Two T6SSi loci occurred in the genomes of 52 of these strains.

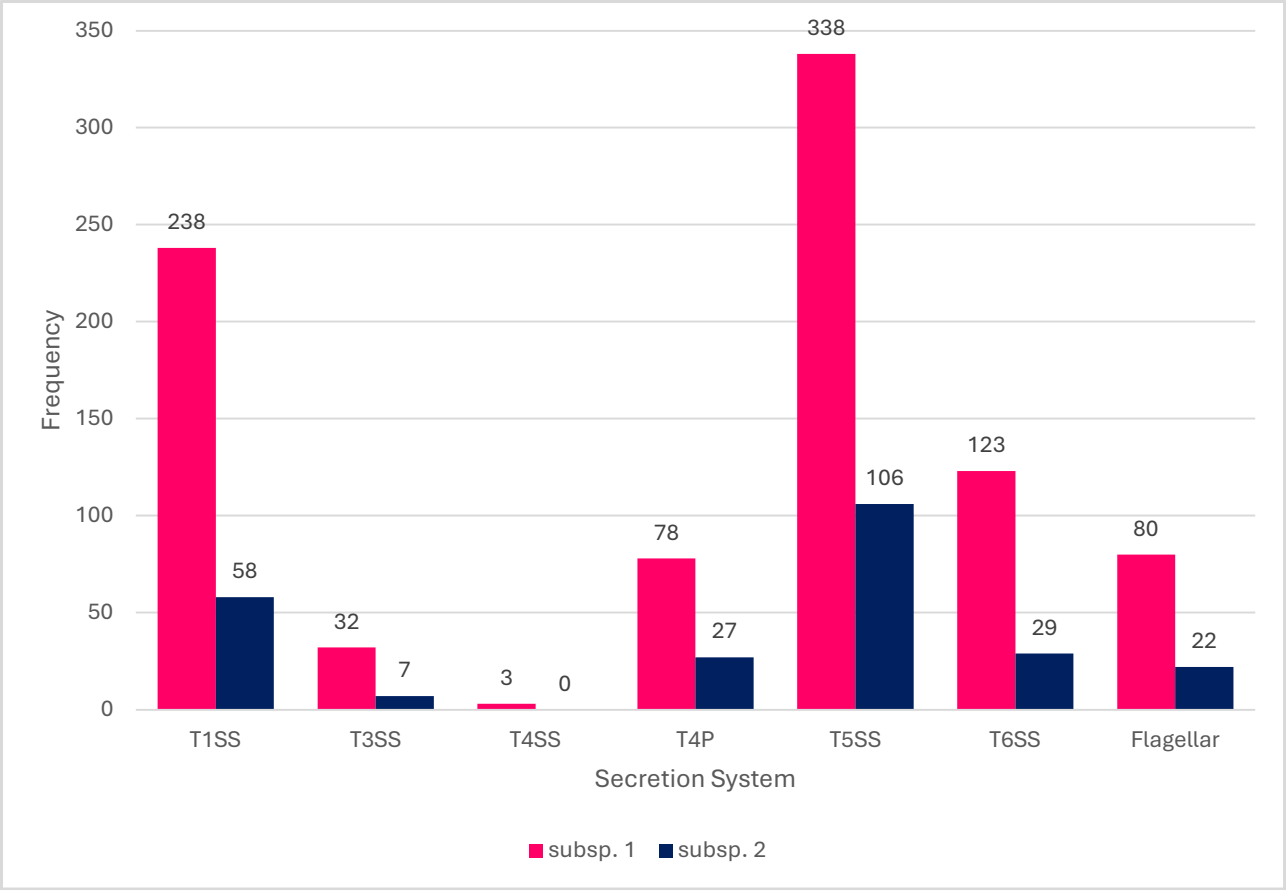


Figure 3.15: Relative numbers of the distinct types of secretion systems in *P. agglomerans* subsp. 1 and subsp. 2. Seven distinct types of secretion systems were identified in P. agglomerans subsp. 1 (pink) and subsp. 2 (blue).

3.7.4 Genomic analysis of factors underlying environmental resistance

P. agglomerans has been isolated from a broad range of ecological niches, and as such need effective mechanisms of resistance against environmental toxins, such as antimicrobials and heavy metals, as well as antibiotics, which are commonly employed in the agricultural and anthropogenic setting (Acioly *et al.*, 2017). Heavy metal (mercury, lead, copper, arsenic, chromium and cadmium) resistant *P. agglomerans* have been isolated from agricultural fields in India (Lee and Cho, 2024). Similarly, *P. agglomerans* C1 displayed resistance to arsenic, copper and cadmium, and *P. agglomerans* GN2, tolerance to the metal compounds, iron sulfate and copper sulfate which is a commonly used pesticide (Hamid *et al.*, 2023; Luziatelli, Ficca, *et al.*, 2020). Genome analysis of *P. agglomerans* KM1 revealed antimicrobial resistance genes effective against penicillin G, rifampicin, vancomycin, fosfomycin and bacitracin (Guevarra *et al.*, 2021).

A comparison of the 100 *P. agglomerans* comparator genomes against the ResFinder database (Bortolaia *et al.*, 2020) did not identify resistance phenotypes against any of the 92 evaluated antibiotics. This is likely because the database was designed on the basis of well-known pathogens in the family *Enterobacteriaceae*, while antibiotic resistance genes in members of the family *Erwiniaceae*, which includes *Pantoea* species, are sufficiently different at the sequence level to be

overlooked. Subsequently, the *P. agglomerans* proteomes were evaluated using AMRFinder-Plus (Feldgarden *et al.*, 2019). On average, each *P. agglomerans* genome harbours 4.15 antimicrobial resistance (AMR) genes (Table 3.14). The lowest number of AMR genes (two genes) was observed on the genome of *P. agglomerans* CFBP8784. By contrast, 31 *P. agglomerans* strains encoded a maximum of five AMR genes. A total of four and three AMR genes were identified in the genomes of the outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, respectively. The 415 AMR genes identified in *P. agglomerans* belonged to seven distinct categories (Figure 3.16).

Table 3.14: Antimicrobial resistance genes in *P. agglomerans*.

	No. of AMR genes	Strain
Average	4.15	
Minimum	2	CFBP 8784
Maximum	5	31 strains

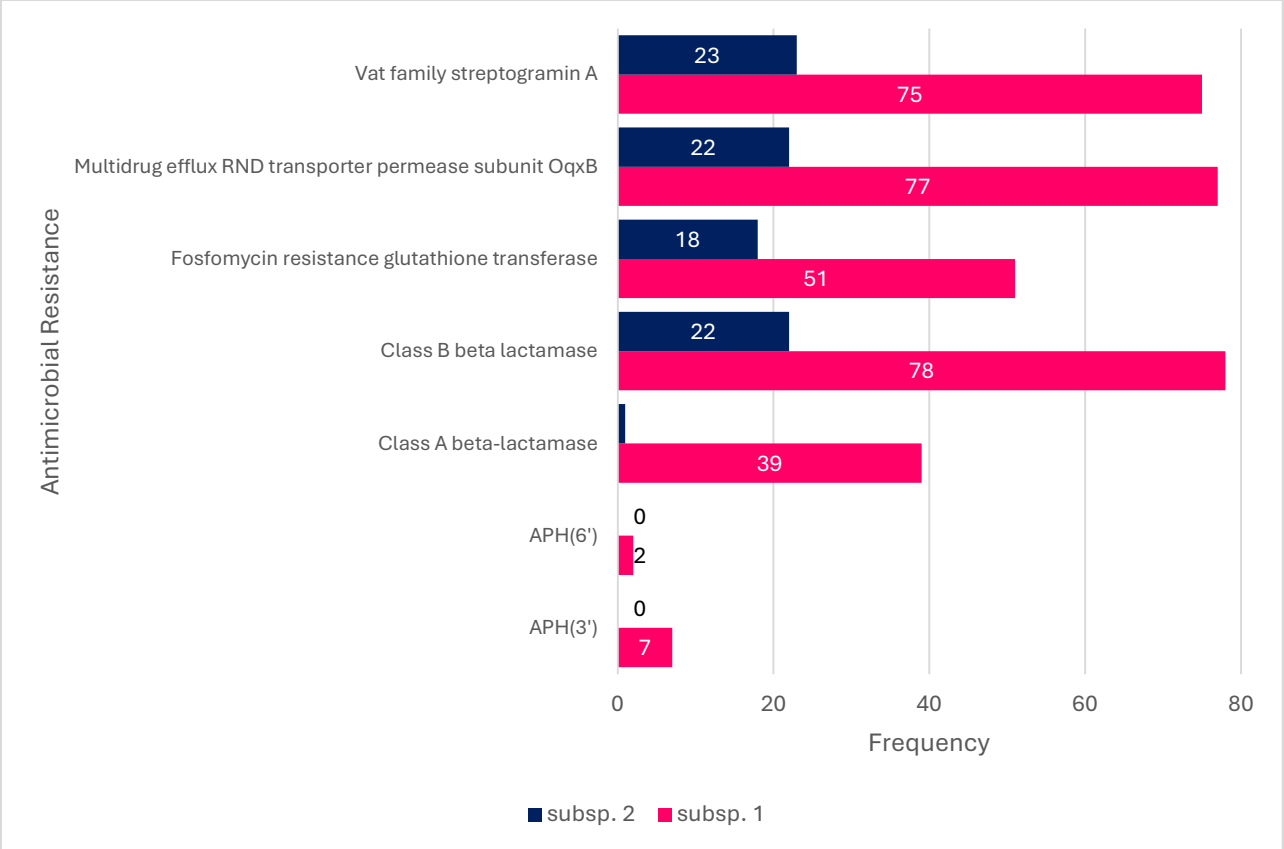


Figure 3.16: Distribution of antimicrobial and antibiotic resistance genes present in the 100 *P. agglomerans* genomes, across both predicted subspecies.

Aminoglycoside phosphotransferases (APH) catalyse the transfer of the terminal phosphoryl group in ATP to the hydroxyl group of the aminoglycoside antibiotics, resulting in the loss of their biological activity (Kaplan *et al.*, 2016; Kim *et al.*, 2006). The APHs are subdivided, into seven subclasses, on the basis of the location of the modified carbon on the aminoglycoside antibiotics (Kaplan *et al.*, 2016). The genomes of seven *P. agglomerans*, belonging to *P. agglomerans* subsp. 1, harbour *aph(3')*

genes which codes for the enzyme transferring a phosphoryl group to carbon 3 on the aminoglycoside antibiotics. Two of these strains, *P. agglomerans* Pa21-5 and Pa21-15, also harbour *aph(6')* genes coding for the enzyme that catalyses a phosphoryl group transfer to carbon 6 on the aminoglycoside antibiotics.

β-lactams are widespread antibiotics used to treat Gram-negative bacterial infections (Pandey *et al*., 2023). β-lactamase enzymes hydrolyses the amide bond of the β-lactam ring, deeming the antibiotic ineffective (Pandey *et al*., 2023). Class A and C β-lactamases both have serine active sites on their target β-lactams antibiotics, however, class A β-lactamases are mainly plasmid-based, whereas class C are primarily localised on the chromosome (Bush and Jacoby, 2010). Class C β-lactamases are encoded on the genomes of all 100 *P. agglomerans* comparators. On the other hand, class A β-lactamases are present in 39 *P. agglomerans* subsp. 1 strains and only a single subsp. 2 strain (*P. agglomerans* AR1a).

Fosfomycin is a broad spectrum antibiotic that is inactivated by the addition of glutathione, catalysed by FosA (Rigsby *et al.*, 2005). FosA is a Mn(II)-dependent metalloenzyme, which requires potassium to function optimally (Rigsby *et al.*, 2005). This enzyme is present in 51 and 18 *P. agglomerans* subsp. 1 and subsp. 2 strains, respectively. Virginiamycin O-acetyltransferase (Vat) enzymes inactivate virginiamycin-like antibiotics by transferring an acetyl group from acetyl-CoA to the secondary alcohol of streptogramin (on either component A or B) (Stogios *et al.*, 2014). Copies of the Vat family protein Streptogramin A were identified in the genomes of 75 *P. agglomerans* subsp. 1 strains (except *P. agglomerans* CFBP8784, Pa39-1 and R5) and all 22 *P. agglomerans* subsp. 2 strains (two copies in *P. agglomerans* DAPP-PG734). The multidrug efflux RND transporter permease subunit gene *oqxB* was identified in 99/100 *P. agglomerans* strains and was lacking from the genome of *P. agglomerans* subsp. 1 strain CFSAN033090. Efflux pumps reduce the intracellular antibiotic concentration and promotes site mutation accumulation (Li *et al.*, 2019).

In addition to antimicrobial resistance, proteins involved in heavy metal resistance were identified by comparing the *P. agglomerans* proteomes against the BacMet database (Pal *et al.*, 2013). On average 58.38 heavy metal resistance protein hits were identified in the compared *P. agglomerans* genomes, ranging from 54 to 71) (Table 3.15). A total of twenty-two strains harbour 54 heavy metal resistance genes, of which only one strain belonged to *P. agglomerans* subsp. 2 (*P. agglomerans* T1). The P. *agglomerans* subsp. 1 strains, *P. agglomerans* MR10 and MR19, each carry 71 heavy metal resistance genes. The outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, harbour 57 and 56 heavy metal resistance genes, respectively.

Table 3.15: Heavy metal resistance genes in *P. agglomerans*.

	No. of heavy metal resistance genes	No. of strains present in
Average	58.58	
Minimum	54	22
Maximum	71	2

A total of 82 distinct heavy metal resistance genes were identified in the comparator datasets, of which 50 are found on the genomes of all 100 *P. agglomerans* strains. Four heavy metal resistance genes were found only in *P. agglomerans* subsp. 2. This included two genes (*ncrC* and *nirB*) involved in nickel resistance, found in two copies on the genome of *P. agglomerans* NCTC10601, a gene (*sitD*) involved in manganese, iron and hydrogen peroxide resistance restricted to *P. agglomerans* PA, and a gene (*modA*) involved in tungsten and molybdenum resistance, which is restricted to *P. agglomerans* GD03966, JZB_2120015 and ZJU23, respectively (Pal *et al.*, 2013). *P. agglomerans* subsp. 1 uniquely harbours seven heavy metal resistance proteins across ten strains. A gene (*baeR*) inferring zinc and tungsten resistance was identified in *P. agglomerans* AB378 and CFBP8784 (Pal *et al.*, 2013). The gene *baeR* was also associated with resistance to the disinfectant sodium deoxycholate (Appendix F) (Pal *et al.*, 2013). Similarly, a gene (*cepA*) restricted to *P. agglomerans* Pa31-4, expressed resistance to the disinfectant chlorhexidine (Pal *et al.*, 2013). Four heavy metal resistance genes (*terB*, *terC*, *terD* and *terE*), inferring tellurium resistance, were only identified on the genomes of *P. agglomerans* B1, B3, MP2 and PSV1-7 (Pal *et al.*, 2013). The gene *arsB* was associated with inferring arsenic and antimony resistance in *P. agglomerans* MR10 and MR19 (Pal *et al.*, 2013).

Most pertinent is the resistance to copper, which is formulated in bactericides to manage onion centre rot caused by *P. agglomerans* (Koirala *et al.*, 2023). Copper-tolerance has been linked to the gene cluster *copABCDRS*, which was observed to be on pOnion plasmids in onion *P. agglomerans* strains (Shin *et al.*, 2025). Blastn analysis against the *P. agglomerans* comparators in this study, identified these gene clusters in the genomes of twenty-three strains (fourteen and nine in *P. agglomerans* subsp. 1 and 2, respectively), showing a broader tolerance to copper for the species. In the comparator strains the *cop* locus is found on a wide range of plasmid types, including LPP-1, LPP-3, LPP-5 and MPP-5, dependent on the strain.

Aside from antimicrobial and heavy metal resistance, the success of *P. agglomerans* on plant hosts is also dependent on their ability to withstand plant antimicrobials. While this has been the focus of limited research, previous research has identified a genetic locus in *P. ananatis*, the *alt* locus, which provides tolerance against thiosulfinates, volatile antimicrobial compounds produced by members of the genus *Allium* (e.g. onion and garlic), upon cellular damage (Stice *et al.*, 2020). The *alt* locus (*altB-J*) derives its name from *al*licin *t*olerance, the key thiosulfinate produced by onions and garlic (Stice

et al., 2020). Previous research identified the *alt* locus in onion centre rot strains of *P. agglomerans*, linking it the pOnion plasmid also known to carry the HiVir locus (Shin *et al.*, 2025). While the HiVir locus was found to have a limited distribution (on the LPP-3 plasmid of *P. agglomerans* CB1 and AR1a, and the LPP-2 plasmid of *P. agglomerans* 20TX0122), the *alt* locus showed a more broad distribution, being found on the LPP-3 plasmids of seven *P. agglomerans* strains, namely 20TX0122, 824-1, AR1a, CB1, CFBP8756, CFBP13731, K1 and KM1. The limited range of taxa showing allicin resistance and the HiVir locus suggests specialisation of this restricted strain set to *Allium*-type hosts.

3.8 Secondary metabolite production in *P. agglomerans*

Secondary metabolites are specialised, auxiliary compounds that are beneficial for interactions between microorganisms and their environment (Sharrar *et al.*, 2020). Secondary metabolites can influence nutrient acquisition, bacterial communication and pathogenesis and as such have become of interest in a broad range of biotechnological applications (Sharrar *et al.*, 2020). *P. agglomerans* strains are known to produce a wide range of secondary metabolites with biotechnological potential. This includes antibiotics such as herbicolin A and B (lipopeptide antibiotics), herbicolin I (dapdiamide antibiotics), andrimid (pseudopeptide antibiotics), phenazines, microcin (peptide antibiotics), agglomerins A-D (acidic antibiotics) and pantocin A and B (Dutkiewicz *et al*., 2016; Wright *et al.*, 2006)

The 100 *P. agglomerans* strains were screened for the production of secondary metabolites using the antiSMASH server (Blin *et al.*, 2023). The 100 comparator *P. agglomerans* are predicted to produce an average of 7.36 secondary metabolites/genome. This ranges from six (*P. agglomerans* GD03966, IG1 and Pa17-5) to ten secondary metabolites (*P. agglomerans* CFSAN047153, CFSAN047154 and MR10) (Table 3.16; Appendix G). The outgroup taxa, *P. eucalypti* LMG 24197^{T} and *P. vagans* LMG 24199^{T}, are predicted to produce six and eight secondary metabolites, respectively. A total of fourteen distinct secondary metabolites were identified from the genomes of *P. agglomerans*.

Table 3.16: Presence of secondary metabolites in *P. agglomerans* genomes.

	No. of secondary metabolites	Strain
Average	7.36	
Minimum	6	GD03966, IG1, Pa17-5
Maximum	10	CFSAN047153, CFSAN047154, MR10

Loci for the biosynthesis of a non-ribosomal peptide synthetase (NRPS)-independent (NI) siderophore, non-ribosomal peptide synthase-dependent (NRPS)siderophore, thiopeptide, acyl-homoserine lactones (AHLs), terpene and redox co-factors were universally present on the genomes of all compared *P. agglomerans* strains. Siderophores are ferric iron chelators, which enable bacteria to obtain iron, essential for microbial growth, even in iron depauperate environments (Carroll and Moore, 2018). The NI-siderophore, localised on plasmid LPP-1 in the *P. agglomerans* strains, was identified as desferrioxamine E (antiSMASH similarity score 0.92 – *P. agglomerans* BGC0001572), which has been shown to have antibacterial activity in *P. agglomerans,* allowing to outcompete other microorganisms for restricted iron sources (Choi *et al.*, 2022). The NRPS-siderophore, identified as enterobactin (antiSMASH similarity score: 1.12 – *E. coli* K-12 substr. MG1655 BGC0002476), is chromosomally localised in all *P. agglomerans* strains and is a high-affinity siderophore (Soutar and

Stavrinides, 2018). A third siderophore aerobactin, is found only in *P. ananatis, P. septica* and *P. stewartii* (Soutar and Stavrinides, 2018) and was absent from all comparator *P. agglomerans* strains.

The redox co-factor locus comprises the genes *pqqABCDEF*, which are involved in the synthesis of pyrroquinoline quinone (PQQ), a cofactor in a wide array of enzymatic reactions. PQQ has been shown to play a role in plant growth promotion, through the solubilisation of phosphate in soils (Carreño-López, Alatorre-Cruz and Marín-Cevada, 2019). *N*-acyl-homoserine lactones (AHLs) are small volatile signal molecules that facilitate quorum sensing (QS), a form of bacterial communication that allows bacterial communities to respond to population density and control expression of a wide range of genes underlying physiological functions such as epiphytic growth in plants and virulence (Chalupowicz *et al.*, 2008; von Bodman and Farrand, 1995). This concept has been particularly well-studied in the genus *Pantoea*, where AHLs regulate exopolysaccharide production in *P. stewartii* subsp. *stewartii*, resulting in vascular occlusion in maize plants (Von Bodman and Farrand, 1995) and in *P. agglomerans* pv. *gypsophilae*, where AHL production regulates expression of the T3SS and phytohormone production (and thus gall formation) (Chalupowicz *et al.*, 2008). Production of two distinct AHLs (the major *N*-butanoyl-L-homoserine lactone and minor *N*-hexanoyl-L-homoserine) in the latter pathogen was linked to the AHL-synthase PagI and its cognate transcriptional regulator PagR (Chalupowicz *et al.*, 2008). Two distinct AHL biosynthetic loci were identified in the comparator *P. agglomerans* genomes. The first of these is conserved among all 100 strains. The second AHL biosynthetic locus, including the genes coding for the PagIR combination, is found in 98/100 strains, being absent from *P. agglomerans* GD03966 and IG1. The latter strain was isolated as pear epiphytes, suggesting the universally conserved AHL locus may serve similar functions as the PagIR-encoding locus. The terpene biosynthetic locus, conserved among all 100 comparator taxa, is predicted to encode a carotenoid (antiSMASH similarity score: 0.62 – *P. ananatis* BGC0000638), the yellow pigment that provides antioxidant properties and photoprotection in *P. agglomerans* (Dussault *et al.*, 2008). Notable, carotenoid production has been observed to increase upon *P. agglomerans* exposure to γ-radiation, suggesting this pigment provides radiotolerance, allowing this species to survive in harsh environments (Dussault *et al.*, 2008).

Other secondary metabolite biosynthetic loci are less prevalent among the 100 comparator *P. agglomerans* strains. These secondary metabolites correlate to virulence and antibiotic biosynthesis phenotypes of individual strains. Only a single comparator strain, *P. agglomerans* EKM10T incorporates a furan biosynthetic locus on its genome. BlastP analysis of the biosynthetic protein showed it to belong to the AvrD family (pfam05655). The AvrD protein of *P. syringae* serves as an elicitor of the hypersensitive response in plants carrying the complimentary resistance gene, thereby determining host range (Keen *et al.*, 1991). Phosphonate biosynthetic loci only occur on the plasmids

of *P. agglomerans* 20TX0122, AR1a and CB1. These correlate to the HiVir biosynthetic genes for the phosphonate phytotoxin pantaphos (Shin *et al.*, 2025).

The other secondary metabolite biosynthetic loci are involved in the production of a broad range of antimicrobial substances. Three distinct, plasmid-localised, loci for the production of ribosomally synthesized and post-translationally modified peptides (RiPPs) are present on the genomes of a restricted set of *P. agglomerans* strains. The genomes of five *P. agglomerans* subsp. 1 strains, namely *P. agglomerans* BI3, BAV2934, CFSAN047153, CFSA047154 and RIT-PI-T harbour biosynthetic locus for an uncharacterised lassopeptide, a class of RiPPs with a wide range of biological activities, including antimicrobial activity against a range of bacterial pathogens (Cheng and Hua, 2020; Wright *et al.*, 2024; Zhu *et al.*, 2016). Lassopeptides can further be applied in a wide array of biotechnological applications including as therapeutics against cancer, gastrointestinal diseases and HIV (Cheng and Hua, 2020). The uncharacterised lassopeptide of these strains thus warrants further investigation. Furthermore, the genomes of these same five strains encode a colicin-like RiPP on plasmid LPP-4. Colicins, a type of bacteriocin, kill or inhibit bacterial strains closely related to the producing taxon, thereby achieving a competitive advantage (Cascales *et al.*, 2007). Additionally, a RiPP-like locus, unique to *P. agglomerans* SL1_M5 (among the compared strains), was identified by antiSMASH, but its function remains unknown.

Phenazines are a group of antibiotics with broad spectrum activity that can supress phytopathogens and kill nematodes (Mavrodi *et al.*, 2010). The D-alanylgriseoluteic acid (AGA) phenazine antibiotic is encoded on the LPP-2 plasmid of eight *P. agglomerans* subsp. 1 strains (*P. agglomerans* ASB05, BAV2934, CFSAN047153, CFSAN047154, Pa39-7, Pa39-21, Pa39-23 and strain 190). The locus comprises sixteen genes (*ephA-O*) and the encoded antibiotic has particular relevance to the biological control of the fire blight pathogen *E. amylovora* (Giddens *et al*., 2002; Kearns and Mahanty, 1998).

Previously, *Pantoea* species, producing histidine-reversible antibiotics, were assessed as a fire blight biological control agent (Smits *et al.*, 2019). The most commonly isolated histidine-reversible antibiotics include herbicolin O, MccEh252 and pantocin A (Smits *et al.*, 2019). Pantocin A loci are found in six *P. agglomerans*, namely *P. agglomerans* 20TX0122, Eh318, Pa31-3, Pa39-1, P10c and Tx10. Pantocin A is a ribosomally encoded, post-translationally modified peptide that is effective against the closely related enterobacterial phytopathogens *E. amylovora*, *Erwinia rhapontici*, *Pectobacterium carotovorum* and *Dickeya dadantii* (Smits *et al.*, 2019). A second RRE-containing locus is restricted to *P. agglomerans* CFBP13600 and NBBC-01 (of *P. agglomerans* subsp. 1) as well as *P. stewartia* subsp. *indologenes* LMG 2671^{T}. This locus shows some similarity with the pantocin A locus (antiSMASH similarity score: 0.33 – *P. agglomerans* BGC0002317) but remains

uncharacterised. Although not directly identified by antiSMASH, the genomes of four strains (*P. agglomerans* C410P1*,* DAPP-PG734, SL1_M5 and strain 4) contain the nine-gene biosynthetic locus, *dapABCDEFGHI*, located on a plasmid-borne ICE element, that is involved in the biosynthesis of the antibiotic dapdiamide E (or herbicolin I), which was shown to be effective against the fire blight pathogen *E. amylovora* (Kamber *et al.*, 2012; Sulja *et al.*, 2022).

Chapter Four:
Summary

Pantoea agglomerans is a geographically, ecologically and functionally diverse species, which is reflected in the genomes of the 100 comparator strains in this study. Phylogenomic analyses revealed that *P. agglomerans* delineates into two distinct subspecies, although this delineation did not align to their source of isolation or lifestyles as commensal or pathogen. Both subspecies, and the species as a whole, display a wide-open pan-genome, even when extrapolated to a more expansive dataset, supporting their ability to acquire genes within their environments and expand their phenotypes.

A significant proportion of the genomes of *P. agglomerans* encompass plasmid elements, bacteriophages, integrative and conjugative elements (ICEs) and transposable elements. These elements infer variation and adaptations at a species, subspecies and strain-specific level, providing them with selective competitive advantages within the environments they occupy, and potentially allowing them to occupy novel ecological niches. Functional analysis of *P. agglomerans* revealed that the core genome houses a higher proportion of proteins relating to basic cellular and metabolic functioning whereas the accessory genome comprised of proteins, primarily linked to the mobilome, that allow adaptation to specific niches and environmental conditions. This includes strain-specific metabolic pathways and defense mechanisms against various environmental stresses, contributing towards the functional diversity of the species. In particular, the genomes of *P. agglomerans* harbour mechanisms of resistance against a broad array of antimicrobials agents and environmental pollutants such as heavy metals.

Despite *P. agglomerans* being recognised as an opportunistic clinical pathogen, all strains demonstrate a low predicted human pathogenicity potential, which is below the threshold for clinical significance. However, the identification of a broad range of pathogenicity determinants, particularly those associated with phytopathogens, provided insight into the specialisation of phytopathogenic strains towards their plant hosts. This was particularly evidenced by the variable presence of secretion systems that allow strains to interact with their environment and their hosts.

A wide variety of secondary metabolites are produced by distinct strains of *P. agglomerans*, many of which could be of biotechnological value. In particular, an array of antibiotic agents, including novel biosynthetic loci, were identified which can be exploited in biological control strategies against different pathogens and pests.

This study explored the molecular determinants underlying the ecological success and functional diversification of *P. agglomerans*. The open pan-genome and mobile elements of *P. agglomerans* are supportive of the continual acquisition of genes that relay niche-specific adaptations. The genomes of *P. agglomerans* harbour strain-specific metabolic pathways, defense mechanisms, resistance to environment pollutants and pathogenicity determinants that are key to the ecological success of the

species and enable the dissociation of pathogenic from biotechnologically beneficial strains. Furthermore, specific strains of *P. agglomerans*, producing biomolecules and secondary metabolites with biotechnological importance, were identified and can be exploited for commercial applications.

Reference List

Abbas, G. *et al.* (2018) 'Arsenic Uptake, Toxicity, Detoxification, and Speciation in Plants: Physiological, Biochemical, and Molecular Aspects', *International Journal of Environmental Research and Public Health*, 15(1), p. 59. Available at: https://doi.org/10.3390/ijerph15010059.

Abbas, Z.R., Authman, S.H. and Al-Ezee, A.M.M. (2017) 'Temperature effects on growth of the biocontrol agent *Pantoea agglomerans* (An oval isolate from Iraqi soils)', *Journal of Advanced Laboratory Research in Biology*, 8(4), pp. 85–88.

Acioly, L. *et al.* (2017) 'An Investigation on the Heavy Metal Tolerance and Antibiotic Profile of the *Pantoea agglomerans* UCP1320', *International Journal of Current Microbiology and Applied Sciences*, 6(11), pp. 4145–4151. Available at: https://doi.org/10.20546/ijcmas.2017.611.486.

Adeolu, M. *et al.* (2016) 'Genome-based phylogeny and taxonomy of the "Enterobacteriales": proposal for Enterobacterales ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov.', *International Journal of Systematic and Evolutionary Microbiology*, 66(12), pp. 5575–5599. Available at: https://doi.org/10.1099/ijsem.0.001485.

Adiyaman, T. *et al.* (2011) 'Selection of Biocontrol Agents of Pink Rot Based on Efficacy and Growth Kinetics Index Rankings', *Plant Disease*, 95(1), pp. 24–30. Available at: https://doi.org/10.1094/PDIS-04-10-0284.

Adriaenssens, E.M. *et al.* (2011) 'Bacteriophages LIMElight and LIMEzero of *Pantoea agglomerans*, Belonging to the "phiKMV-Like Viruses"', *Applied and Environmental Microbiology*, 77(10), pp. 3443–3450. Available at: https://doi.org/10.1128/AEM.00128-11.

Agarwal, G. *et al.* (2021a) 'Pan-Genome-Wide Analysis of *Pantoea ananatis* Identified Genes Linked to Pathogenicity in Onion', *Frontiers in Microbiology*, 12, p. 684756. Available at: https://doi.org/10.3389/fmicb.2021.684756.

Agarwal, G., Gitaitis, R.D. and Dutta, B. (2021b) 'Pan-Genome of Novel *Pantoea stewartii* subsp. *indologenes* Reveals Genes Involved in Onion Pathogenicity and Evidence of Lateral Gene Transfer', *Microorganisms*, 9(8), p. 1761. Available at: https://doi.org/10.3390/microorganisms9081761.

Al-Qaysi, S.A.S. *et al.* (2021) 'Bioactive Levan-Type Exopolysaccharide Produced by *Pantoea agglomerans* ZMR7: Characterization and Optimization for Enhanced Production', *Journal of Microbiology and Biotechnology*, 31(5), pp. 696–704. Available at: https://doi.org/10.4014/jmb.2101.01025.

Altenhoff, A.M. *et al.* (2013) 'Inferring Hierarchical Orthologous Groups from Orthologous Gene Pairs', *PLoS ONE*. Edited by O. Lespinet, 8(1), p. e53786. Available at: https://doi.org/10.1371/journal.pone.0053786.

Altschul, S. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. Available at: https://doi.org/10.1093/nar/25.17.3389.

Altschul, S.F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

An, X.-H. *et al.* (2023) 'Physiological and transcriptomic analyses of response of walnuts (Juglans regia) to *Pantoea agglomerans* infection', *Frontiers in Plant Science*, 14, p. 1294643. Available at: https://doi.org/10.3389/fpls.2023.1294643.

Anlauf, M.T. *et al.* (2024) 'Type 1 secretion necessitates a tight interplay between all domains of the ABC transporter', *Scientific Reports*, 14(1), p. 8994. Available at: https://doi.org/10.1038/s41598-024-59759-0.

Antar, H. and Gruber, S. (2023) 'VirB, a transcriptional activator of virulence in *Shigella flexneri*, uses CTP as a cofactor', *Communications Biology*, 6(1), p. 1204. Available at: https://doi.org/10.1038/s42003-023-05590-8.

Arahal, D.R. (2014) 'Whole-Genome Analyses', in *Methods in Microbiology*. Elsevier, pp. 103–122. Available at: https://doi.org/10.1016/bs.mim.2014.07.002.

Asai, N. *et al.* (2019) '*Pantoea dispersa* bacteremia in an immunocompetent patient: a case report and review of the literature', *Journal of Medical Case Reports*, 13(1), p. 33. Available at: https://doi.org/10.1186/s13256-019-1969-z.

Asselin, J.A.E., Bonasera, J.M. and Beer, S.V. (2018) 'Center Rot of Onion (*Allium cepa*) Caused by *Pantoea ananatis* Requires *pepM* , a Predicted Phosphonate-Related Gene', *Molecular Plant-Microbe Interactions®*, 31(12), pp. 1291–1300. Available at: https://doi.org/10.1094/MPMI-04-18-0077-R.

Audu, K.E., Adeniji, S.E. and Obidah, J.S. (2020) 'Bioremediation of toxic metals in mining site of Zamfara metropolis using resident bacteria (*Pantoea agglomerans*): A optimization approach', *Heliyon*, 6(8), p. e04704. Available at: https://doi.org/10.1016/j.heliyon.2020.e04704.

Barash, I. and Manulis-Sasson, S. (2007) 'Virulence mechanisms and host specificity of gall-forming *Pantoea agglomerans*', *Trends in Microbiology*, 15(12), pp. 538–545. Available at: https://doi.org/10.1016/j.tim.2007.10.009.

Barrett, T. *et al.* (2012) 'BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata', *Nucleic Acids Research*, 40(D1), pp. D57–D63. Available at: https://doi.org/10.1093/nar/gkr1163.

Beji, A. *et al.* (1988) 'Subjective Synonymy of *Erwinia herbicola, Erwinia milletiae,* and *Enterobacter agglomerans* and Redefinition of the Taxon by Genotypic and Phenotypic Data', *International Journal of Systematic Bacteriology*, 38(1), pp. 77–88. Available at: https://doi.org/10.1099/00207713-38-1-77.

Belon, C. *et al.* (2015) 'A Macrophage Subversion Factor Is Shared by Intracellular and Extracellular Pathogens', *PLOS Pathogens*. Edited by A.R. Hauser, 11(6), p. e1004969. Available at: https://doi.org/10.1371/journal.ppat.1004969.

Bharadwaj, A.R. *et al.* (2020) 'Draft Genome Sequences of *Enterobacteriales* Strains Isolated from the International Space Station', *Microbiology Resource Announcements*. Edited by D. Rasko, 9(37), pp. e00817-20. Available at: https://doi.org/10.1128/MRA.00817-20.

Blin, K. *et al.* (2023) 'antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation', *Nucleic Acids Research*, 51(W1), pp. W46–W50. Available at: https://doi.org/10.1093/nar/gkad344.

Blin, K. *et al.* (2024) 'The antiSMASH database version 4: additional genomes and BGCs, new sequence-based searches and more', *Nucleic Acids Research*, 52(D1), pp. D586–D589. Available at: https://doi.org/10.1093/nar/gkad984.

Boot, M. *et al.* (2017) 'A fluorescence-based reporter for monitoring expression of mycobacterial cytochrome bd in response to antibacterials and during infection', *Scientific Reports*, 7(1), p. 10665. Available at: https://doi.org/10.1038/s41598-017-10944-4.

Bortolaia, V. *et al.* (2020) 'ResFinder 4.0 for predictions of phenotypes from genotypes', *Journal of Antimicrobial Chemotherapy*, 75(12), pp. 3491–3500. Available at: https://doi.org/10.1093/jac/dkaa345.

Bosi, E. *et al.* (2015) 'MEDUSA: a multi-draft based scaffolder', *Bioinformatics*, 31(15), pp. 2443–2451. Available at: https://doi.org/10.1093/bioinformatics/btv171.

Bourque, G. *et al.* (2018) 'Ten things you should know about transposable elements', *Genome Biology*, 19(1), p. 199. Available at: https://doi.org/10.1186/s13059-018-1577-z.

Braga, L.P.P. *et al.* (2018) 'Bacterial Diversification in the Light of the Interactions with Phages: The Genetic Symbionts and Their Role in Ecological Speciation', *Frontiers in Ecology and Evolution*, 6, p. 6. Available at: https://doi.org/10.3389/fevo.2018.00006.

Bright, L.A. *et al.* (2009) 'Structural and functional-annotation of an equine whole genome oligoarray', *BMC Bioinformatics*, 10(S11), p. S8. Available at: https://doi.org/10.1186/1471-2105-10-S11-S8.

Bruisson, S. *et al.* (2019) 'Endophytes and Epiphytes From the Grapevine Leaf Microbiome as Potential Biocontrol Agents Against Phytopathogens', *Frontiers in Microbiology*, 10, p. 2726. Available at: https://doi.org/10.3389/fmicb.2019.02726.

Brzóstkowska, M., Raczkowska, A. and Brzostek, K. (2012) 'OmpR, a response regulator of the two-component signal transduction pathway, influences inv gene expression in *Yersinia enterocolitica* O9', *Frontiers in Cellular and Infection Microbiology*, 2. Available at: https://doi.org/10.3389/fcimb.2012.00153.

Bush, K. and Jacoby, G.A. (2010) 'Updated Functional Classification of β-Lactamases', *Antimicrobial Agents and Chemotherapy*, 54(3), pp. 969–976. Available at: https://doi.org/10.1128/AAC.01009-09.

Büyükcam, A. *et al.* (2018) 'Clinical and microbiological characteristics of *Pantoea agglomerans* infection in children', *Journal of Infection and Public Health*, 11(3), pp. 304–309. Available at: https://doi.org/10.1016/j.jiph.2017.07.020.

Cantalapiedra, C.P. *et al.* (2021) 'eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale', *Molecular Biology and Evolution*. Edited by K. Tamura, 38(12), pp. 5825–5829. Available at: https://doi.org/10.1093/molbev/msab293.

Carattoli, A. and Hasman, H. (2020) 'PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS)', in F. De La Cruz (ed.) *Horizontal Gene Transfer*. New York, NY: Springer US (Methods in Molecular Biology), pp. 285–294. Available at: https://doi.org/10.1007/978-1-4939-9877-7_20.

Carbajal-González, M. *et al.* (2013) 'Adhesion to brown trout skin mucus, antagonism against cyst adhesion and pathogenicity to rainbow trout of some inhibitory bacteria against *Saprolegnia parasitica*', *Diseases of Aquatic Organisms*, 104(1), pp. 35–44. Available at: https://doi.org/10.3354/dao02582.

Carobbi, A. *et al.* (2022) 'An antibacterial T6SS in *Pantoea agglomerans* pv. *betae* delivers a lysozyme-like effector to antagonize competitors', *Environmental Microbiology*, 24(10), pp. 4787–4802. Available at: https://doi.org/10.1111/1462-2920.16100.

Carr, V.R. *et al.* (2021) 'Probing the Mobilome: Discoveries in the Dynamic Microbiome', *Trends in Microbiology*, 29(2), pp. 158–170. Available at: https://doi.org/10.1016/j.tim.2020.05.003.

Carreño-López, R., Alatorre-Cruz, J.M. and Marín-Cevada, V. (2019) 'Pyrroloquinoline quinone (PQQ): Role in Plant-Microbe Interactions', in H.B. Singh et al. (eds) *Secondary Metabolites of Plant Growth Promoting Rhizomicroorganisms*. Singapore: Springer Singapore, pp. 169–184. Available at: https://doi.org/10.1007/978-981-13-5862-3_9.

Carroll, C.S. and Moore, M.M. (2018) 'Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases', *Critical Reviews in Biochemistry and Molecular Biology*, 53(4), pp. 356–381. Available at: https://doi.org/10.1080/10409238.2018.1476449.

Cascales, E. *et al.* (2007) 'Colicin Biology', *Microbiology and Molecular Biology Reviews*, 71(1), pp. 158–229. Available at: https://doi.org/10.1128/MMBR.00036-06.

Chalupowicz, L. *et al.* (2008) 'Quorum-Sensing System Affects Gall Development Incited by *Pantoea agglomerans* pv. *gypsophilae*', *Molecular Plant-Microbe Interactions®*, 21(8), pp. 1094–1105. Available at: https://doi.org/10.1094/MPMI-21-8-1094.

Chaudhari, N.M., Gupta, V.K. and Dutta, C. (2016) 'BPGA- an ultra-fast pan-genome analysis pipeline', *Scientific Reports*, 6(1), p. 24373. Available at: https://doi.org/10.1038/srep24373.

Chaumeil, P.-A. *et al.* (2020) 'GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database', *Bioinformatics*. Edited by J. Hancock, 36(6), pp. 1925–1927. Available at: https://doi.org/10.1093/bioinformatics/btz848.

Cheng, C. and Hua, Z.-C. (2020) 'Lasso Peptides: Heterologous Production and Potential Medical Application', *Frontiers in Bioengineering and Biotechnology*, 8, p. 571165. Available at: https://doi.org/10.3389/fbioe.2020.571165.

Choi, M.-S. *et al.* (2013) 'Harpins, Multifunctional Proteins Secreted by Gram-Negative Plant-Pathogenic Bacteria', *Molecular Plant-Microbe Interactions®*, 26(10), pp. 1115–1122. Available at: https://doi.org/10.1094/MPMI-02-13-0050-CR.

Choi, O. *et al.* (2022) 'Negatively Regulated Aerobactin and Desferrioxamine E by Fur in *Pantoea ananatis* Are Required for Full Siderophore Production and Antibacterial Activity, but Not for Virulence', *Applied and Environmental Microbiology*. Edited by G. Alexandre, 88(6), pp. e02405-21. Available at: https://doi.org/10.1128/aem.02405-21.

Chong, H. and Li, Q. (2017) 'Microbial production of rhamnolipids: opportunities, challenges and strategies', *Microbial Cell Factories*, 16(1), p. 137. Available at: https://doi.org/10.1186/s12934-017-0753-2.

Choovet, N. *et al.* (2024) 'Antimicrobial activity, genome analysis, and gene cluster encoding the plantaricin production of *Lactiplantibacillus plantarum* ZBK1-5', *ScienceAsia*, 50(2), p. 1. Available at: https://doi.org/10.2306/scienceasia1513-1874.2024.046.

Cicchetti, R. *et al.* (2006) '*Pantoea agglomerans* Sepsis After Rotavirus Gastroenteritis':, *The Pediatric Infectious Disease Journal*, 25(3), pp. 280–281. Available at: https://doi.org/10.1097/01.inf.0000202211.64017.c6.

Coburn, B., Sekirov, I. and Finlay, B.B. (2007) 'Type III Secretion Systems and Disease', *Clinical Microbiology Reviews*, 20(4), pp. 535–549. Available at: https://doi.org/10.1128/CMR.00013-07.

Conway, T. and Cohen, P.S. (2015) 'Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut', *Microbiology Spectrum*. Edited by T. Conway and P. Cohen, 3(3), p. 3.3.24. Available at: https://doi.org/10.1128/microbiolspec.MBP-0006-2014.

Correa, V.R. *et al.* (2012) 'The Bacterium *Pantoea stewartii* Uses Two Different Type III Secretion Systems To Colonize Its Plant Host and Insect Vector', *Applied and Environmental Microbiology*, 78(17), pp. 6327–6336. Available at: https://doi.org/10.1128/AEM.00892-12.

Corsini, G. *et al.* (2016) 'Draft Genome Sequence of a Copper-Resistant Marine Bacterium, *Pantoea agglomerans* Strain LMAE-2, a Bacterial Strain with Potential Use in Bioremediation', *Genome Announcements*, 4(3), pp. e00525-16. Available at: https://doi.org/10.1128/genomeA.00525-16.

Cosentino, S. *et al.* (2013) 'PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data', *PLoS ONE*. Edited by J.D. Barbour, 8(10), p. e77302. Available at: https://doi.org/10.1371/journal.pone.0077302.

Coutinho, T.A. and Venter, S.N. (2009) '*Pantoea ananatis* : an unconventional plant pathogen', *Molecular Plant Pathology*, 10(3), pp. 325–335. Available at: https://doi.org/10.1111/j.1364-3703.2009.00542.x.

Crisan, C.V. and Goldberg, J.B. (2022) 'Antibacterial contact-dependent proteins secreted by Gram-negative cystic fibrosis respiratory pathogens', *Trends in Microbiology*, 30(10), pp. 986–996. Available at: https://doi.org/10.1016/j.tim.2022.03.009.

Crosby, K.C. *et al.* (2023) 'Genomic delineation and description of species and within-species lineages in the genus *Pantoea*', *Frontiers in Microbiology*, 14, p. 1254999. Available at: https://doi.org/10.3389/fmicb.2023.1254999.

Cruz, A.T., Cazacu, A.C. and Allen, C.H. (2007) '*Pantoea agglomerans* , a Plant Pathogen Causing Human Disease', *Journal of Clinical Microbiology*, 45(6), pp. 1989–1992. Available at: https://doi.org/10.1128/JCM.00632-07.

Da Costa, R.M. *et al.* (2024) 'Type 5 secretion system antigens as vaccines against Gram-negative bacterial infections', *npj Vaccines*, 9(1), p. 159. Available at: https://doi.org/10.1038/s41541-024-00953-6.

Dahiya, P. *et al.* (2024) 'Comparative Genomic and Functional Analyses for Insights into *Pantoea agglomerans* Strains Adaptability in Diverse Ecological Niches', *Current Microbiology*, 81(8), p. 254. Available at: https://doi.org/10.1007/s00284-024-03763-0.

Dawlaty, J. *et al.* (2010) 'Dapdiamides, Tripeptide Antibiotics Formed by Unconventional Amide Ligases', *Journal of Natural Products*, 73(3), pp. 441–446. Available at: https://doi.org/10.1021/np900685z.

de Maayer, P. *et al.* (2011) 'Comparative genomics of the type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins', *BMC Genomics*, 12(1), p. 576. Available at: https://doi.org/10.1186/1471-2164-12-576.

de Maayer, P. *et al.* (2012) 'The large universal *Pantoea* plasmid LPP-1 plays a major role in biological and ecological diversification', *BMC Genomics*, 13(1), p. 625. Available at: https://doi.org/10.1186/1471-2164-13-625.

de Maayer, P. *et al.* (2014) 'Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts', *BMC Genomics*, 15(1), p. 404. Available at: https://doi.org/10.1186/1471-2164-15-404.

de Maayer, P. *et al.* (2015) 'Integrative conjugative elements of the ICEPan family play a potential role in *Pantoea ananatis* ecological diversification and antibiosis', *Frontiers in Microbiology*, 6. Available at: https://doi.org/10.3389/fmicb.2015.00576.

de Maayer, P. *et al.* (2017) 'Phylogenomic, Pan-genomic, Pathogenomic and Evolutionary Genomic Insights into the Agronomically Relevant Enterobacteria *Pantoea ananatis* and *Pantoea stewartii*', *Frontiers in Microbiology*, 8, p. 1755. Available at: https://doi.org/10.3389/fmicb.2017.01755.

de Maayer, P., Pillay, T. and Coutinho, T.A. (2020) 'Flagella by numbers: comparative genomic analysis of the supernumerary flagellar systems among the Enterobacterales', *BMC Genomics*, 21(1), p. 670. Available at: https://doi.org/10.1186/s12864-020-07085-w.

Delavat, F. *et al.* (2017) 'The hidden life of integrative and conjugative elements', *FEMS Microbiology Reviews*, 41(4), pp. 512–537. Available at: https://doi.org/10.1093/femsre/fux008.

Delétoile, A. *et al.* (2009) 'Phylogeny and Identification of *Pantoea* Species and Typing of *Pantoea agglomerans* Strains by Multilocus Gene Sequencing', *Journal of Clinical Microbiology*, 47(2), pp. 300–310. Available at: https://doi.org/10.1128/JCM.01916-08.

Deng, W. *et al.* (2017) 'Assembly, structure, function and regulation of type III secretion systems', *Nature Reviews Microbiology*, 15(6), pp. 323–337. Available at: https://doi.org/10.1038/nrmicro.2017.20.

Dereeper, A. *et al.* (2008) 'Phylogeny.fr: robust phylogenetic analysis for the non-specialist', *Nucleic Acids Research*, 36(Web Server), pp. W465–W469. Available at: https://doi.org/10.1093/nar/gkn180.

Diepold, A. and Armitage, J.P. (2015) 'Type III secretion systems: the bacterial flagellum and the injectisome', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1679), p. 20150020. Available at: https://doi.org/10.1098/rstb.2015.0020.

Dillon, R.J., Vennard, C.T. and Charnley, A.K. (2000) 'Exploitation of gut bacteria in the locust', *Nature*, 403(6772), pp. 851–851. Available at: https://doi.org/10.1038/35002669.

Dinos, G.P. (2017) 'The macrolide antibiotic renaissance', *British Journal of Pharmacology*, 174(18), pp. 2967–2983. Available at: https://doi.org/10.1111/bph.13936.

Duchateau, S. *et al.* (2024) 'The plant-associated *Pantoea* spp. as biocontrol agents: Mechanisms and diversity of bacteria-produced metabolites as a prospective tool for plant protection', *Biological Control*, 188, p. 105441. Available at: https://doi.org/10.1016/j.biocontrol.2024.105441.

Duncan-Lowey, B. and Kranzusch, P.J. (2022) 'CBASS phage defense and evolution of antiviral nucleotide signaling', *Current Opinion in Immunology*, 74, pp. 156–163. Available at: https://doi.org/10.1016/j.coi.2022.01.002.

Duong, D.A., Stevens, A.M. and Jensen, R.V. (2018) 'Expanded Analysis of the *Pantoea stewartii* subsp. *stewartii* DC283 Complete Genome Reveals Plasmid-borne Virulence Factors'. Bioinformatics. Available at: https://doi.org/10.1101/407825.

Dussault, D. *et al.* (2008) 'Carotenoids' influence on radiotolerance of *Pantoea agglomerans* , a plant pathogen', *Letters in Applied Microbiology*, 47(3), pp. 208–213. Available at: https://doi.org/10.1111/j.1472-765X.2008.02410.x.

Dutkiewicz, J. *et al.* (2015) '*Pantoea agglomerans*: a marvelous bacterium of evil and good. Part I. Deleterious effects: Dust-borne endotoxins and allergens – focus on cotton dust', *Annals of Agricultural and Environmental Medicine*, 22(4), pp. 576–588. Available at: https://doi.org/10.5604/12321966.1185757.

Dutkiewicz, J., Mackiewicz, B., Lemieszek, M.K., Golec, M., Skórska, C., *et al.* (2016) '*Pantoea agglomerans*: a mysterious bacterium of evil and good. Part II. Deleterious effects: Dust-borne endotoxins and allergens – focus on grain dust, other agricultural dusts and wood dust', *Annals of Agricultural and Environmental Medicine*, 23(1), pp. 6–29. Available at: https://doi.org/10.5604/12321966.1196848.

Dutkiewicz, J., Mackiewicz, B., Kinga Lemieszek, M., *et al.* (2016) '*Pantoea agglomerans*: a mysterious bacterium of evil and good. Part III. Deleterious effects: infections of humans, animals and plants', *Annals of Agricultural and Environmental Medicine*, 23(2), pp. 197–205. Available at: https://doi.org/10.5604/12321966.1203878.

Dutkiewicz, J., Mackiewicz, B., Lemieszek, M.K., Golec, M. and Milanowski, J. (2016) '*Pantoea agglomerans*: a mysterious bacterium of evil and good. Part IV. Beneficial effects', *Annals of Agricultural and Environmental Medicine*, 23(2), pp. 206–222. Available at: https://doi.org/10.5604/12321966.1203879.

Ebner, P. and Götz, F. (2019) 'Bacterial Excretion of Cytoplasmic Proteins (ECP): Occurrence, Mechanism, and Function', *Trends in Microbiology*, 27(2), pp. 176–187. Available at: https://doi.org/10.1016/j.tim.2018.10.006.

Ejaz, H. *et al.* (2022) 'Multiple Antimicrobial Resistance and Heavy Metal Tolerance of Biofilm-Producing Bacteria Isolated from Dairy and Non-Dairy Food Products', *Foods*, 11(18), p. 2728. Available at: https://doi.org/10.3390/foods11182728.

Elizondo-Reyna, E. *et al.* (2025) 'Insights from a Genome-Wide Study of *Pantoea agglomerans* UADEC20: A Promising Strain for Phosphate Solubilization and Exopolysaccharides Production', *Current Issues in Molecular Biology*, 47(1), p. 56. Available at: https://doi.org/10.3390/cimb47010056.

Ellison, D.W. and Miller, V.L. (2006) 'Regulation of virulence by members of the MarR/SlyA family', *Current Opinion in Microbiology*, 9(2), pp. 153–159. Available at: https://doi.org/10.1016/j.mib.2006.02.003.

Ellory, C. and Hewitt, A. (2011) 'Essentials of cell physiology', *Surgery (Oxford)*, 29(3), pp. 103–106. Available at: https://doi.org/10.1016/j.mpsur.2010.11.017.

Emms, D.M. and Kelly, S. (2019) 'OrthoFinder: phylogenetic orthology inference for comparative genomics', *Genome Biology*, 20(1), p. 238. Available at: https://doi.org/10.1186/s13059-019-1832-y.

Ewing, W.H. and Fife, M.A. (1972) '*Enterobacter agglomerans* (Beijerinck) comb. nov. (the Herbicola-Lathyri Bacteria)', *International Journal of Systematic Bacteriology*, 22(1), pp. 4–11. Available at: https://doi.org/10.1099/00207713-22-1-4.

Feldgarden, M. *et al.* (2019) 'Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates', *Antimicrobial Agents and Chemotherapy*, 63(11), pp. e00483-19. Available at: https://doi.org/10.1128/AAC.00483-19.

Feng, Y., Shen, D. and Song, W. (2006) 'Rice endophyte *Pantoea agglomerans* YS19 promotes host plant growth and affects allocations of host photosynthates', *Journal of Applied Microbiology*, 100(5), pp. 938–945. Available at: https://doi.org/10.1111/j.1365-2672.2006.02843.x.

Founou, L.L. *et al.* (2022) 'Genome Analysis of ESBL-Producing *Escherichia coli* Isolated from Pigs', *Pathogens*, 11(7), p. 776. Available at: https://doi.org/10.3390/pathogens11070776.

Gaeta, N.C. *et al.* (2018) 'Bacterial pathogens of the lower respiratory tract of calves from Brazilian rural settlement herds and their association with clinical signs of bovine respiratory disease', *Pesquisa Veterinária Brasileira*, 38(3), pp. 374–381. Available at: https://doi.org/10.1590/1678-5150-pvb-5323.

Galperin, M.Y. *et al.* (2021) 'COG database update: focus on microbial diversity, model organisms, and widespread pathogens', *Nucleic Acids Research*, 49(D1), pp. D274–D281. Available at: https://doi.org/10.1093/nar/gkaa1018.

Gavini, F. *et al.* (1989) 'Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as Pantoea agglomerans comb. nov. and Description of *Pantoea dispersa* sp. nov.', *International Journal of Systematic Bacteriology*, 39(3), pp. 337–345. Available at: https://doi.org/10.1099/00207713-39-3-337.

Georjon, H. and Bernheim, A. (2023) 'The highly diverse antiphage defence systems of bacteria', *Nature Reviews Microbiology*, 21(10), pp. 686–700. Available at: https://doi.org/10.1038/s41579-023-00934-x.

Geraffi, N. *et al.* (2023) 'Comparative sequence analysis of pPATH pathogenicity plasmids in *Pantoea agglomerans* gall-forming bacteria', *Frontiers in Plant Science*, 14, p. 1198160. Available at: https://doi.org/10.3389/fpls.2023.1198160.

Giani, A.M. *et al.* (2020) 'Long walk to genomics: History and current approaches to genome sequencing and assembly', *Computational and Structural Biotechnology Journal*, 18, pp. 9–19. Available at: https://doi.org/10.1016/j.csbj.2019.11.002.

Giddens, S.R., Feng, Y. and Mahanty, H.K. (2002) 'Characterization of a novel phenazine antibiotic gene cluster in *Erwinia herbicola* Eh1087', *Molecular Microbiology*, 45(3), pp. 769–783. Available at: https://doi.org/10.1046/j.1365-2958.2002.03048.x.

Godebo, A.T., Germida, J.J. and Walley, F.L. (2020) 'Isolation, identification, and assessment of soil bacteria as biocontrol agents of pea root rot caused by *Aphanomyces euteiches*', *Canadian Journal of Soil Science*. Edited by M.A. Naeth, 100(3), pp. 206–216. Available at: https://doi.org/10.1139/cjss-2019-0133.

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), pp. 81–91. Available at: https://doi.org/10.1099/ijs.0.64483-0.

Goyal, A. (2018) 'Metabolic adaptations underlying genome flexibility in prokaryotes', *PLOS Genetics*. Edited by D. Hughes, 14(10), p. e1007763. Available at: https://doi.org/10.1371/journal.pgen.1007763.

Green, E.R. and Mecsas, J. (2016) 'Bacterial Secretion Systems: An Overview', *Microbiology Spectrum*. Edited by I.T. Kudva, 4(1), p. 4.1.13. Available at: https://doi.org/10.1128/microbiolspec.VMBF-0012-2015.

Guevarra, R.B. *et al.* (2021) 'Comprehensive genomic analysis reveals virulence factors and antibiotic resistance genes in *Pantoea agglomerans* KM1, a potential opportunistic pathogen', *PLOS ONE*. Edited by F. Gao, 16(1), p. e0239792. Available at: https://doi.org/10.1371/journal.pone.0239792.

Guo, M. *et al.* (2020) 'First Report of Bacterial Soft Rot Caused by *Pantoea agglomerans* on Chinese Cabbage in China', *Plant Disease*, 104(1), p. 277. Available at: https://doi.org/10.1094/PDIS-04-19-0746-PDN.

Habich, A. *et al.* (2025) 'Distribution of the four type VI secretion systems in *Pseudomonas aeruginosa* and classification of their core and accessory effectors', *Nature Communications*, 16(1), p. 888. Available at: https://doi.org/10.1038/s41467-024-54649-5.

Haidar, R. *et al.* (2021) 'Application methods and modes of action of *Pantoea agglomerans* and *Paenibacillus* sp. to control the grapevine trunk disease-pathogen, *Neofusicoccum parvum*', *OENO One*, 55(3). Available at: https://doi.org/10.20870/oeno-one.2021.55.2.4530.

Hamid, B. *et al.* (2023) 'Heavy-metal tolerant bacterial strains isolated from industrial sites and scrap yards in Kashmir, India', *Journal of Basic Microbiology*, 63(12), pp. 1361–1372. Available at: https://doi.org/10.1002/jobm.202300349.

Holtappels, M. *et al.* (2018) 'Fire blight host-pathogen interaction: proteome profiles of *Erwinia amylovora* infecting apple rootstocks', *Scientific Reports*, 8(1), p. 11689. Available at: https://doi.org/10.1038/s41598-018-30064-x.

Hosni, T. *et al.* (2011) 'Sharing of quorum-sensing signals and role of interspecies communities in a bacterial plant disease', *The ISME Journal*, 5(12), pp. 1857–1870. Available at: https://doi.org/10.1038/ismej.2011.65.

Hu, J. (2019) 'First Report of Onion Bulb Rot Caused by *Pantoea agglomerans* in Arizona', *Plant Disease*, 103(6), pp. 1408–1408. Available at: https://doi.org/10.1094/PDIS-10-18-1749-PDN.

Huang, S. *et al.* (2021) 'Overview of biological ice nucleating particles in the atmosphere', *Environment International*, 146, p. 106197. Available at: https://doi.org/10.1016/j.envint.2020.106197.

Huerta-Cepas, J. *et al.* (2017) 'Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper', *Molecular Biology and Evolution*, 34(8), pp. 2115–2122. Available at: https://doi.org/10.1093/molbev/msx148.

Huerta-Cepas, J. *et al.* (2019) 'eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses', *Nucleic Acids Research*, 47(D1), pp. D309–D314. Available at: https://doi.org/10.1093/nar/gky1085.

Hyatt, D. *et al.* (2010) 'Prodigal: prokaryotic gene recognition and translation initiation site identification', *BMC Bioinformatics*, 11(1), p. 119. Available at: https://doi.org/10.1186/1471-2105-11-119.

Irawati, W., Tahya, C.Y. and Greisnaningsi, G. (2022) '*Pantoea agglomerans, Klebsiella pneumoniae*, and *Shigella flexneri* isolated from the Cisadane River as multiresistant bacteria to copper and dyes', *Indonesian Journal of Biotechnology*, 27(4), p. 179. Available at: https://doi.org/10.22146/ijbiotech.66103.

Ishimaru, C.A. *et al.* (2017) 'THE HISTIDINE-REVERSIBLE ANTIBIOTIC HERBICOLIN O PRODUCED BY *PANTOEA VAGANS* C9-1 IS PANTOCIN A', *Journal of Plant Pathology*, 99(Vol 99 (2017): Special issue – First International Symposium on Fire Blight of Rosaceous Plants). Available at: https://doi.org/10.4454/jpp.v99i0.3910.

Izadi-Pruneyre, N., Karami, Y. and Nilges, M. (2024) 'Structure and Dynamics of Type 4a Pili and Type 2 Secretion System Endopili', in J.R. Harris and J. Marles-Wright (eds) *Macromolecular Protein Complexes V*. Cham: Springer International Publishing (Subcellular Biochemistry), pp. 549–563. Available at: https://doi.org/10.1007/978-3-031-58843-3_21.

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

Janda, J.M. and Abbott, S.L. (2021) 'The Changing Face of the Family *Enterobacteriaceae* (Order: " *Enterobacterales* "): New Members, Taxonomic Issues, Geographic Expansion, and New Diseases and Disease Syndromes', *Clinical Microbiology Reviews*, 34(2), pp. e00174-20. Available at: https://doi.org/10.1128/CMR.00174-20.

Jiang, L. *et al.* (2019) 'Potential of *Pantoea dispersa* as an effective biocontrol agent for black rot in sweet potato', *Scientific Reports*, 9(1), p. 16354. Available at: https://doi.org/10.1038/s41598-019-52804-3.

Jin, L. *et al.* (2016) 'Direct and Indirect Targeting of PP2A by Conserved Bacterial Type-III Effector Proteins', *PLOS Pathogens*. Edited by P. Birch, 12(5), p. e1005609. Available at: https://doi.org/10.1371/journal.ppat.1005609.

Johnson, C.M. and Grossman, A.D. (2015) 'Integrative and Conjugative Elements (ICEs): What They Do and How They Work', *Annual Review of Genetics*, 49(1), pp. 577–601. Available at: https://doi.org/10.1146/annurev-genet-112414-055018.

Kalyaanamoorthy, S. *et al.* (2017) 'ModelFinder: fast model selection for accurate phylogenetic estimates', *Nature Methods*, 14(6), pp. 587–589. Available at: https://doi.org/10.1038/nmeth.4285.

Kamber, T. *et al.* (2012) 'Characterization of the Biosynthetic Operon for the Antibacterial Peptide Herbicolin in *Pantoea vagans* Biocontrol Strain C9-1 and Incidence in *Pantoea* Species', *Applied*

and Environmental Microbiology, 78(12), pp. 4412–4419. Available at: https://doi.org/10.1128/AEM.07351-11.

Kangale, L.J., Raoult, D. and Pierre-Edouard, F. (2021) '*Pedobacter ghigonii* sp. nov., Isolated from the Microbiota of the Planarian *Schmidtea mediterranea*', *Microbiology Research*, 12(2), pp. 268–287. Available at: https://doi.org/10.3390/microbiolres12020019.

Kaplan, E. *et al.* (2016) 'Aminoglycoside binding and catalysis specificity of aminoglycoside 2″-phosphotransferase IVa: A thermodynamic, structural and kinetic study', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(4), pp. 802–813. Available at: https://doi.org/10.1016/j.bbagen.2016.01.016.

Kaur, I.P. *et al.* (2020) '*Pantoea agglomerans* Bacteremia: Is It Dangerous?', *Case Reports in Infectious Diseases*, 2020, pp. 1–4. Available at: https://doi.org/10.1155/2020/7890305.

Kearns, L.P. and Mahanty, H.K. (1998) 'Antibiotic Production by *Erwinia herbicola* Eh1087: Its Role in Inhibition of *Erwinia amylovora* and Partial Characterization of Antibiotic Biosynthesis Genes', *Applied and Environmental Microbiology*, 64(5), pp. 1837–1844. Available at: https://doi.org/10.1128/AEM.64.5.1837-1844.1998.

Keen, N. *et al.* (1991) 'Avirulence Gene D from *Pseudomonas Syringae* pv. Tomato and Its Interaction with Resistance Gene Rpg4 in Soybean', in H. Hennecke and D.P.S. Verma (eds) *Advances in Molecular Genetics of Plant-Microbe Interactions Vol. 1*. Dordrecht: Springer Netherlands (Current Plant Science and Biotechnology in Agriculture), pp. 37–44. Available at: https://doi.org/10.1007/978-94-015-7934-6_6.

Kim, C. *et al.* (2006) 'Hydrolysis of ATP by Aminoglycoside 3′-Phosphotransferases', *Journal of Biological Chemistry*, 281(11), pp. 6964–6969. Available at: https://doi.org/10.1074/jbc.M513257200.

Kizheva, Y. *et al.* (2022) 'Cross-Over Pathogenic Bacteria Detected in Infected Tomatoes (*Solanum lycopersicum* L.) and Peppers (*Capsicum annuum* L.) in Bulgaria', *Pathogens*, 11(12), p. 1507. Available at: https://doi.org/10.3390/pathogens11121507.

Koirala, S. *et al.* (2023) 'Evaluating Options to Increase the Efficacy of Biocontrol Agents for the Management of *Pantoea* spp. Under Field Conditions', *Plant Disease*, 107(9), pp. 2701–2708. Available at: https://doi.org/10.1094/PDIS-11-22-2710-RE.

Konstantinidis, K.T. and Tiedje, J.M. (2005) 'Genomic insights that advance the species definition for prokaryotes', *Proceedings of the National Academy of Sciences*, 102(7), pp. 2567–2572. Available at: https://doi.org/10.1073/pnas.0409727102.

Koonin, E.V. (2005) 'Orthologs, Paralogs, and Evolutionary Genomics', *Annual Review of Genetics*, 39(1), pp. 309–338. Available at: https://doi.org/10.1146/annurev.genet.39.073003.114725.

Koskella, B. (2020) 'The phyllosphere', *Current Biology*, 30(19), pp. R1143–R1146. Available at: https://doi.org/10.1016/j.cub.2020.07.037.

Kumar, P. *et al.* (2022) 'Whole genome analysis for plant growth promotion profiling of *Pantoea agglomerans* CPHN2, a non-rhizobial nodule endophyte', *Frontiers in Microbiology*, 13, p. 998821. Available at: https://doi.org/10.3389/fmicb.2022.998821.

Kurlanda-Witek, H., Ngwenya, B.T. and Butler, I.B. (2015) 'The influence of biofilms on the mobility of bare and capped zinc oxide nanoparticles in saturated sand and glass beads', *Journal of Contaminant Hydrology*, 179, pp. 160–170. Available at: https://doi.org/10.1016/j.jconhyd.2015.06.009.

Kvitko, B.H. *et al.* (2009) 'Deletions in the Repertoire of *Pseudomonas syringae* pv. tomato DC3000 Type III Secretion Effector Genes Reveal Functional Overlap among Effectors', *PLoS Pathogens*. Edited by J.L. Dangl, 5(4), p. e1000388. Available at: https://doi.org/10.1371/journal.ppat.1000388.

Lee, I.P.A. and Andam, C.P. (2019) 'Pan-genome diversification and recombination in Cronobacter sakazakii, an opportunistic pathogen in neonates, and insights to its xerotolerant lifestyle', *BMC Microbiology*, 19(1), p. 306. Available at: https://doi.org/10.1186/s12866-019-1664-7.

Lee, S.I. *et al.* (2023) 'Use of *Pantoea agglomerans* ASB05 as a biocontrol agent to inhibit the growth of *Salmonella enterica* on intact cantaloupe melons', *Journal of Applied Microbiology*, 134(10), p. lxad235. Available at: https://doi.org/10.1093/jambio/lxad235.

Lee, S.Y. and Cho, K.-S. (2024) 'Isolation of Heavy Metal-Tolerant and Anti-Phytopathogenic Plant Growth-Promoting Bacteria from Soils', *Journal of Microbiology and Biotechnology*, 34(11), pp. 2252–2265. Available at: https://doi.org/10.4014/jmb.2407.07013.

Li, C. *et al.* (2021) 'Aerobactin-Mediated Iron Acquisition Enhances Biofilm Formation, Oxidative Stress Resistance, and Virulence of *Yersinia pseudotuberculosis*', *Frontiers in Microbiology*, 12, p. 699913. Available at: https://doi.org/10.3389/fmicb.2021.699913.

Li, H. and Xu, H. (2024) 'Mechanisms of bacterial resistance to environmental silver and antimicrobial strategies for silver: A review', *Environmental Research*, 248, p. 118313. Available at: https://doi.org/10.1016/j.envres.2024.118313.

Li, J. *et al.* (2019) 'The nature and epidemiology of OqxAB, a multidrug efflux pump', *Antimicrobial Resistance & Infection Control*, 8(1), p. 44. Available at: https://doi.org/10.1186/s13756-019-0489-3.

Li, S. *et al.* (2023) 'Toll/interleukin-1 receptor domains in bacterial and plant immunity', *Current Opinion in Microbiology*, 74, p. 102316. Available at: https://doi.org/10.1016/j.mib.2023.102316.

Li, W. *et al.* (2013) 'The *Pseudomonas syringae* Effector HopQ1 Promotes Bacterial Virulence and Interacts with Tomato 14-3-3 Proteins in a Phosphorylation-Dependent Manner', *Plant Physiology*, 161(4), pp. 2062–2074. Available at: https://doi.org/10.1104/pp.112.211748.

Lim, J.-A. *et al.* (2014) 'Draft genome sequence of *Pantoea agglomerans* R190, a producer of antibiotics against phytopathogens and foodborne pathogens', *Journal of Biotechnology*, 188, pp. 7–8. Available at: https://doi.org/10.1016/j.jbiotec.2014.07.440.

Liu, S.-C., Ju, Y.-R. and Lu, C.L. (2022) 'Multi-CSAR: a web server for scaffolding contigs using multiple reference genomes', *Nucleic Acids Research*, 50(W1), pp. W500–W509. Available at: https://doi.org/10.1093/nar/gkac301.

Liu, Y. *et al.* (2022) 'Patterns of genomic divergence in sympatric and allopatric speciation of three Mihoutao (*Actinidia*) species', *Horticulture Research*, 9, p. uhac054. Available at: https://doi.org/10.1093/hr/uhac054.

Loncaric, I. *et al.* (2009) 'Typing of *Pantoea agglomerans* isolated from colonies of honey bees (*Apis mellifera*) and culturability of selected strains from honey', *Apidologie*, 40(1), pp. 40–54. Available at: https://doi.org/10.1051/apido/2008062.

Lorenzi, A.S. *et al.* (2022) 'Opposite Sides of *Pantoea agglomerans* and Its Associated Commercial Outlook', *Microorganisms*, 10(10), p. 2072. Available at: https://doi.org/10.3390/microorganisms10102072.

Luo, Y. *et al.* (2023) 'The Love and Hate Relationship between T5SS and Other Secretion Systems in Bacteria', *International Journal of Molecular Sciences*, 25(1), p. 281. Available at: https://doi.org/10.3390/ijms25010281.

Luziatelli, F., Ficca, A.G., *et al.* (2020) 'A Genetic and Metabolomic Perspective on the Production of Indole-3-Acetic Acid by *Pantoea agglomerans* and Use of Their Metabolites as Biostimulants in Plant Nurseries', *Frontiers in Microbiology*, 11, p. 1475. Available at: https://doi.org/10.3389/fmicb.2020.01475.

Luziatelli, F., Gatti, L., *et al.* (2020) 'Metabolites Secreted by a Plant-Growth-Promoting Pantoea agglomerans Strain Improved Rooting of *Pyrus communis* L. cv Dar Gazi Cuttings', *Frontiers in Microbiology*, 11, p. 539359. Available at: https://doi.org/10.3389/fmicb.2020.539359.

Lv, L. *et al.* (2022) 'Beneficial Effect and Potential Risk of *Pantoea* on Rice Production', *Plants*, 11(19), p. 2608. Available at: https://doi.org/10.3390/plants11192608.

Magwaza, B., Amobonye, A. and Pillai, S. (2024) 'Microbial β-glucosidases: Recent advances and applications', *Biochimie*, 225, pp. 49–67. Available at: https://doi.org/10.1016/j.biochi.2024.05.009.

Maier, R. and Benoit, S. (2019) 'Role of Nickel in Microbial Pathogenesis', *Inorganics*, 7(7), p. 80. Available at: https://doi.org/10.3390/inorganics7070080.

Mani, S. and Nair, J. (2021) '*Pantoea* Infections in the Neonatal Intensive Care Unit', *Cureus* [Preprint]. Available at: https://doi.org/10.7759/cureus.13103.

Manni, M., Berkeley, Matthew R., *et al.* (2021) 'BUSCO: Assessing Genomic Data Quality and Beyond', *Current Protocols*, 1(12), p. e323. Available at: https://doi.org/10.1002/cpz1.323.

Manni, M., Berkeley, Matthew R, *et al.* (2021) 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes', *Molecular Biology and Evolution*. Edited by J. Kelley, 38(10), pp. 4647–4654. Available at: https://doi.org/10.1093/molbev/msab199.

Manulis, S. and Barash, I. (2003) '*Pantoea agglomerans* pvs. *gypsophilae* and *betae* , recently evolved pathogens?', *Molecular Plant Pathology*, 4(5), pp. 307–314. Available at: https://doi.org/10.1046/j.1364-3703.2003.00178.x.

Mardaneh, J. and Dallal, M.M.S. (2013) 'Isolation, identification and antimicrobial susceptibility of *Pantoea* (*Enterobacter*) *agglomerans* isolated from consumed powdered infant formula milk (PIF) in NICU ward: First report from Iran', *Iranian Journal of Microbiology*, 5(3), pp. 263–267.

Martínez-Fleites, C. *et al.* (2005) 'Crystal structure of levansucrase from the Gram-negative bacterium *Gluconacetobacter diazotrophicus*', *Biochemical Journal*, 390(1), pp. 19–27. Available at: https://doi.org/10.1042/BJ20050324.

Matthews, C.A. *et al.* (2024) 'A gentle introduction to pangenomics', *Briefings in Bioinformatics*, 25(6), p. bbae588. Available at: https://doi.org/10.1093/bib/bbae588.

Mavrodi, D.V. *et al.* (2010) 'Diversity and Evolution of the Phenazine Biosynthesis Pathway', *Applied and Environmental Microbiology*, 76(3), pp. 866–879. Available at: https://doi.org/10.1128/AEM.02009-09.

McDougall, D.L. *et al.* (2020) 'Isolation and Characterization of vB_PagP-SK1, a T7-Like Phage Infecting *Pantoea agglomerans*', *PHAGE*, 1(1), pp. 45–56. Available at: https://doi.org/10.1089/phage.2019.0012.

McKeithen-Mead, S.A. and Grossman, A.D. (2023) 'Timing of integration into the chromosome is critical for the fitness of an integrative and conjugative element and its bacterial host', *PLOS Genetics*. Edited by I. Matic, 19(2), p. e1010524. Available at: https://doi.org/10.1371/journal.pgen.1010524.

Medini, D. *et al.* (2020) 'The Pangenome: A Data-Driven Discovery in Biology', in H. Tettelin and D. Medini (eds) *The Pangenome*. Cham: Springer International Publishing, pp. 3–20. Available at: https://doi.org/10.1007/978-3-030-38281-0_1.

Meier-Kolthoff, J.P. *et al.* (2014) 'Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy', *Standards in Genomic Sciences*, 9(1), p. 2. Available at: https://doi.org/10.1186/1944-3277-9-2.

Meier-Kolthoff, J.P. *et al.* (2022) 'TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes', *Nucleic Acids Research*, 50(D1), pp. D801–D807. Available at: https://doi.org/10.1093/nar/gkab902.

Meier-Kolthoff, J.P. and Göker, M. (2019) 'TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy', *Nature Communications*, 10(1), p. 2182. Available at: https://doi.org/10.1038/s41467-019-10210-3.

Meuskens, I. *et al.* (2019) 'Type V Secretion Systems: An Overview of Passenger Domain Functions', *Frontiers in Microbiology*, 10, p. 1163. Available at: https://doi.org/10.3389/fmicb.2019.01163.

Millman, A. *et al.* (2022) 'An expanding arsenal of immune systems that protect bacteria from phages'. Microbiology. Available at: https://doi.org/10.1101/2022.05.11.491447.

Minh, B.Q. *et al.* (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*. Edited by E. Teeling, 37(5), pp. 1530–1534. Available at: https://doi.org/10.1093/molbev/msaa015.

Minh, B.Q., Nguyen, M.A.T. and Von Haeseler, A. (2013) 'Ultrafast Approximation for Phylogenetic Bootstrap', *Molecular Biology and Evolution*, 30(5), pp. 1188–1195. Available at: https://doi.org/10.1093/molbev/mst024.

Molina-Santiago, C., De Vicente, A. and Romero, D. (2021) 'Bacterial extracellular matrix as a natural source of biotechnologically multivalent materials', *Computational and Structural Biotechnology Journal*, 19, pp. 2796–2805. Available at: https://doi.org/10.1016/j.csbj.2021.05.008.

Moretti, S. *et al.* (2007) 'The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods', *Nucleic Acids Research*, 35(Web Server), pp. W645–W648. Available at: https://doi.org/10.1093/nar/gkm333.

Morgado, S. and Vicente, A.C. (2022) 'Diversity and distribution of Type VI Secretion System gene clusters in bacterial plasmids', *Scientific Reports*, 12(1), p. 8249. Available at: https://doi.org/10.1038/s41598-022-12382-3.

Moura De Sousa, J.A. *et al.* (2021) 'Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa', *Molecular Biology and Evolution*. Edited by I. Arkhipova, 38(6), pp. 2497–2512. Available at: https://doi.org/10.1093/molbev/msab044.

Nakamura, S. and Minamino, T. (2024) 'Structure and Dynamics of the Bacterial Flagellar Motor Complex', *Biomolecules*, 14(12), p. 1488. Available at: https://doi.org/10.3390/biom14121488.

Néron, B. *et al.* (2023) 'MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes', *Peer Community Journal*, 3, p. e28. Available at: https://doi.org/10.24072/pcjournal.250.

Nguyen, L.-T. *et al.* (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32(1), pp. 268–274. Available at: https://doi.org/10.1093/molbev/msu300.

Nissan, G. *et al.* (2018) 'Revealing the inventory of type III effectors in *Pantoea agglomerans* gall-forming pathovars using draft genome sequences and a machine-learning approach', *Molecular Plant Pathology*, 19(2), pp. 381–392. Available at: https://doi.org/10.1111/mpp.12528.

Nissan, G. *et al.* (2019) 'Two *Pantoea agglomerans* type III effectors can transform nonpathogenic and phytopathogenic bacteria into host-specific gall-forming pathogens', *Molecular Plant Pathology*, 20(11), pp. 1582–1587. Available at: https://doi.org/10.1111/mpp.12860.

Okwundu, N. and Mercer, J. (2019) '*Pantoea agglomerans* cutaneous infection', *Journal of Dermatology and Dermatologic Surgery*, 23(1), p. 41. Available at: https://doi.org/10.4103/jdds.jdds_43_18.

Olmos-Alpiste, F., Ezquerra, G.M. and Pujol, R.M. (2021) 'Wound infection by *Pantoea agglomerans* after penetrating plant injury', *Indian Journal of Dermatology, Venereology and Leprology*, 88, pp. 633–635. Available at: https://doi.org/10.25259/IJDVL_1069_19.

Pajuelo, E. *et al.* (2021) 'Coastal Ecosystems as Sources of Biofertilizers in Agriculture: From Genomics to Application in an Urban Orchard', *Frontiers in Marine Science*, 8, p. 685076. Available at: https://doi.org/10.3389/fmars.2021.685076.

Pal, C. *et al.* (2013) 'BacMet: antibacterial biocide and metal resistance genes database', *Nucleic Acids Research*, 42(D1), pp. D737–D743. Available at: https://doi.org/10.1093/nar/gkt1252.

Palmer, M. *et al.* (2018) '*Mixta* gen. nov., a new genus in the *Erwiniaceae*', *International Journal of Systematic and Evolutionary Microbiology*, 68(4), pp. 1396–1407. Available at: https://doi.org/10.1099/ijsem.0.002540.

Palmer, M. and Coutinho, T.A. (2022) '*Pantoea*', in W.B. Whitman (ed.) *Bergey's Manual of Systematics of Archaea and Bacteria*. 1st edn. Wiley, pp. 1–25. Available at: https://doi.org/10.1002/9781118960608.gbm01157.pub2.

Pandey, D., Singhal, N. and Kumar, M. (2023) 'β-LacFamPred: An online tool for prediction and classification of β-lactamase class, subclass, and family', *Frontiers in Microbiology*, 13, p. 1039687. Available at: https://doi.org/10.3389/fmicb.2022.1039687.

Park, S.-C. *et al.* (2019) ‘Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size’, *Frontiers in Microbiology*, 10, p. 834. Available at: https://doi.org/10.3389/fmicb.2019.00834.

Parks, D.H. *et al.* (2022) ‘GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy’, *Nucleic Acids Research*, 50(D1), pp. D785–D794. Available at: https://doi.org/10.1093/nar/gkab776.

Parte, A.C. *et al.* (2020) ‘List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ’, *International Journal of Systematic and Evolutionary Microbiology*, 70(11), pp. 5607–5612. Available at: https://doi.org/10.1099/ijsem.0.004332.

Parvez, S. *et al.* (2020) ‘An insight into the genome of extensively drug-resistant and uropathogenic Citrobacter werkmanii’, *Journal of Global Antimicrobial Resistance*, 22, pp. 785–791. Available at: https://doi.org/10.1016/j.jgar.2020.06.014.

Patel, P.H. and Maxwell, K.L. (2023) ‘Prophages provide a rich source of antiphage defense systems’, *Current Opinion in Microbiology*, 73, p. 102321. Available at: https://doi.org/10.1016/j.mib.2023.102321.

Paul, D. *et al.* (2023) ‘Antibiotic Potentiation as a Promising Strategy to Combat Macrolide Resistance in Bacterial Pathogens’, *Antibiotics*, 12(12), p. 1715. Available at: https://doi.org/10.3390/antibiotics12121715.

Paulina, P. *et al.* (2014) ‘*Pantoea agglomerans* an Agent to Remove Residual Copper from Aquaculture Activity’, *Advanced Materials Research*, 945–949, pp. 3479–3482. Available at: https://doi.org/10.4028/www.scientific.net/AMR.945-949.3479.

Pelicic, V. (2023) ‘Mechanism of assembly of type 4 filaments: everything you always wanted to know (but were afraid to ask): This article is part of the Bacterial Cell Envelopes collection.’, *Microbiology*, 169(3). Available at: https://doi.org/10.1099/mic.0.001311.

Penner, M. *et al.* (2022) ‘Successful Treatment of *Pantoea agglomerans* Bacteremia Using Oral Antibiotics’, *Case Reports in Infectious Diseases*, 2022, p. 6136265. Available at: https://doi.org/10.1155/2022/6136265.

Pérez-Ramos, A. *et al.* (2017) ‘Characterization of the Sorbitol Utilization Cluster of the Probiotic Pediococcus parvulus 2.6: Genetic, Functional and Complementation Studies in Heterologous Hosts’, *Frontiers in Microbiology*, 8, p. 2393. Available at: https://doi.org/10.3389/fmicb.2017.02393.

Petrou, V.I. *et al.* (2016) ‘Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation’, *Science*, 351(6273), pp. 608–612. Available at: https://doi.org/10.1126/science.aad1172.

Pfeifer, E., Bonnin, R.A. and Rocha, E.P.C. (2022) ‘Phage-Plasmids Spread Antibiotic Resistance Genes through Infection and Lysogenic Conversion’, *mBio*. Edited by G.H. Goldman, 13(5), pp. e01851-22. Available at: https://doi.org/10.1128/mbio.01851-22.

Pishchik, V.N. *et al.* (2009) ‘Interactions between plants and associated bacteria in soils contaminated with heavy metals’, *Microbiology*, 78(6), pp. 785–793. Available at: https://doi.org/10.1134/S0026261709060162.

Polidore, A.L.A. *et al.* (2021) 'A Phosphonate Natural Product Made by *Pantoea ananatis* is Necessary and Sufficient for the Hallmark Lesions of Onion Center Rot', *mBio*. Edited by F.M. Ausubel, 12(1), pp. e03402-20. Available at: https://doi.org/10.1128/mBio.03402-20.

Potrikus, C.J. and Breznak, J.A. (1977) 'Nitrogen-fixing Enterobacter agglomerans isolated from guts of wood-eating termites', *Applied and Environmental Microbiology*, 33(2), pp. 392–399. Available at: https://doi.org/10.1128/aem.33.2.392-399.1977.

Puhar, A. and Sansonetti, P.J. (2014) 'Type III secretion system', *Current Biology*, 24(17), pp. R784–R791. Available at: https://doi.org/10.1016/j.cub.2014.07.016.

Puorger, C. *et al.* (2011) 'Structure, Folding and Stability of FimA, the Main Structural Subunit of Type 1 Pili from Uropathogenic *Escherichia coli* Strains', *Journal of Molecular Biology*, 412(3), pp. 520–535. Available at: https://doi.org/10.1016/j.jmb.2011.07.044.

Pusey, P.L. and Wend, C. (2011) 'SURVIVAL OF *PANTOEA AGGLOMERANS* E325 AS FIRE BLIGHT BIOCONTROL AGENT WHEN OSMOADAPTED IN HIGH-SALINE MEDIUM', *Acta Horticulturae*, (896), pp. 467–470. Available at: https://doi.org/10.17660/ActaHortic.2011.896.68.

Rajput, A. *et al.* (2023) 'Pangenome analysis reveals the genetic basis for taxonomic classification of the *Lactobacillaceae* family', *Food Microbiology*, 115, p. 104334. Available at: https://doi.org/10.1016/j.fm.2023.104334.

Rezzonico, F. *et al.* (2009) 'Genotypic comparison of *Pantoea agglomerans* plant and clinical strains', *BMC Microbiology*, 9(1), p. 204. Available at: https://doi.org/10.1186/1471-2180-9-204.

Rezzonico, F., Smits, T.H.M. and Duffy, B. (2012) 'Misidentification slanders Pantoea agglomerans as a serial killer', *Journal of Hospital Infection*, 81(2), pp. 137–139. Available at: https://doi.org/10.1016/j.jhin.2012.02.013.

Rigsby, R.E. *et al.* (2005) 'Fosfomycin Resistance Proteins: A Nexus of Glutathione Transferases and Epoxide Hydrolases in a Metalloenzyme Superfamily', in *Methods in Enzymology*. Elsevier, pp. 367–379. Available at: https://doi.org/10.1016/S0076-6879(05)01023-2.

Robbe-Saule, V. *et al.* (2001) 'Identification of a non-haem catalase in *Salmonella* and its regulation by RpoS (σ^S)', *Molecular Microbiology*, 39(6), pp. 1533–1545. Available at: https://doi.org/10.1046/j.1365-2958.2001.02340.x.

Robinson, L.J. *et al.* (2020) 'A broad-spectrum antibacterial natural product from the cystic fibrosis isolate, *Pantoea agglomerans* Tx10', *Microbiological Research*, 237, p. 126479. Available at: https://doi.org/10.1016/j.micres.2020.126479.

Rollauer, S.E. *et al.* (2015) 'Outer membrane protein biogenesis in Gram-negative bacteria', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1679), p. 20150023. Available at: https://doi.org/10.1098/rstb.2015.0023.

Rouli, L. *et al.* (2015) 'The bacterial pangenome as a new tool for analysing pathogenic bacteria', *New Microbes and New Infections*, 7, pp. 72–85. Available at: https://doi.org/10.1016/j.nmni.2015.06.005.

Roux, N., Spagnolo, J. and De Bentzmann, S. (2012) 'Neglected but amazingly diverse type IVb pili', *Research in Microbiology*, 163(9–10), pp. 659–673. Available at: https://doi.org/10.1016/j.resmic.2012.10.015.

Russo, C.A.D.M. and Selvatti, A.P. (2018) ‘Bootstrap and Rogue Identification Tests for Phylogenetic Analyses’, *Molecular Biology and Evolution*. Edited by K. Tamura, 35(9), pp. 2327–2333. Available at: https://doi.org/10.1093/molbev/msy118.

Sahlin, K., Chikhi, R. and Arvestad, L. (2016) ‘Assembly scaffolding with PE-contaminated mate-pair libraries’, *Bioinformatics*, 32(13), pp. 1925–1932. Available at: https://doi.org/10.1093/bioinformatics/btw064.

Salomone-Stagni, M. *et al.* (2018) ‘Structural and functional analysis of Erwinia amylovora SrlD. The first crystal structure of a sorbitol-6-phosphate 2-dehydrogenase’, *Journal of Structural Biology*, 203(2), pp. 109–119. Available at: https://doi.org/10.1016/j.jsb.2018.03.010.

Sayers, E.W. *et al.* (2021) ‘Database resources of the National Center for Biotechnology Information’, *Nucleic Acids Research*, 49(D1), pp. D10–D17. Available at: https://doi.org/10.1093/nar/gkaa892.

Schoch, C.L. *et al.* (2020) ‘NCBI Taxonomy: a comprehensive update on curation, resources and tools’, *Database*, 2020, p. baaa062. Available at: https://doi.org/10.1093/database/baaa062.

Ševčíková, A. *et al.* (2011) ‘Microbial diversity of internal environment of Johann Gregor Mendel Station, Antarctica’, *Czech Polar Reports*, 1(1), pp. 34–41. Available at: https://doi.org/10.5817/CPR2011-1-3.

Shariati, V. *et al.* (2017) ‘Comprehensive genomic analysis of a plant growth-promoting rhizobacterium *Pantoea agglomerans* strain P5’, *Scientific Reports*, 7(1), p. 15610. Available at: https://doi.org/10.1038/s41598-017-15820-9.

Sharrar, A.M. *et al.* (2020) ‘Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type’, *mBio*. Edited by J.E. Davies, 11(3), pp. e00416-20. Available at: https://doi.org/10.1128/mBio.00416-20.

Shin, G.Y. *et al.* (2025) ‘Plasmids encode and can mobilize onion pathogenicity in *Pantoea agglomerans*’, *The ISME Journal*, p. wraf019. Available at: https://doi.org/10.1093/ismejo/wraf019.

Shrestha, B. *et al.* (2021) ‘*Pantoea agglomerans*: An Elusive Contributor to Chronic Obstructive Pulmonary Disease Exacerbation’, *Cureus* [Preprint]. Available at: https://doi.org/10.7759/cureus.18562.

Shu, R. *et al.* (2022) ‘Detection and Control of *Pantoea agglomerans* Causing Plum Bacterial Shot-Hole Disease by Loop-Mediated Isothermal Amplification Technique’, *Frontiers in Microbiology*, 13, p. 896567. Available at: https://doi.org/10.3389/fmicb.2022.896567.

Silini-Che, H. *et al.* (2012) ‘Isolation and Characterization of Plant Growth Promoting Traits of a Rhizobacteria: *Pantoea agglomerans* lma2’, *Pakistan Journal of Biological Sciences*, 15(6), pp. 267–276. Available at: https://doi.org/10.3923/pjbs.2012.267.276.

Silva, Y.R.D.O. *et al.* (2020) ‘Bacterial secretins: Mechanisms of assembly and membrane targeting’, *Protein Science*, 29(4), pp. 893–904. Available at: https://doi.org/10.1002/pro.3835.

Singh, D., Yadav, D.K. and Fatima, F. (2020) ‘Characterization and genetic diversity of *Pantoea agglomerans* isolates having dual potentiality to suppress growth of *Ralstonia solanacearum* and plant growth promoting ability’, *Indian Phytopathology*, 73(4), pp. 643–653. Available at: https://doi.org/10.1007/s42360-020-00268-1.

Singh, R.P. and Kumari, K. (2023) ‘Bacterial type VI secretion system (T6SS): an evolved molecular weapon with diverse functionality’, *Biotechnology Letters*, 45(3), pp. 309–331. Available at: https://doi.org/10.1007/s10529-023-03354-2.

Smits, T.H.M. *et al.* (2019) ‘Pantocin A, a peptide-derived antibiotic involved in biological control by plant-associated *Pantoea* species’, *Archives of Microbiology*, 201(6), pp. 713–722. Available at: https://doi.org/10.1007/s00203-019-01647-7.

Soltis, P.S. and Soltis, D.E. (2003) ‘Applying the Bootstrap in Phylogeny Reconstruction’, *Statistical Science*, 18(2), pp. 256–267.

Soluch, R. *et al.* (2021) ‘Colonization dynamics of *Pantoea agglomerans* in the wheat root habitat’, *Environmental Microbiology*, 23(4), pp. 2260–2273. Available at: https://doi.org/10.1111/1462-2920.15430.

Somerville, V. *et al.* (2022) ‘Extensive diversity and rapid turnover of phage defense repertoires in cheese-associated bacterial communities’, *Microbiome*, 10(1), p. 137. Available at: https://doi.org/10.1186/s40168-022-01328-6.

Son, H.-J. *et al.* (2006) ‘Solubilization of insoluble inorganic phosphates by a novel salt- and pH-tolerant *Pantoea agglomerans* R-42 isolated from soybean rhizosphere’, *Bioresource Technology*, 97(2), pp. 204–210. Available at: https://doi.org/10.1016/j.biortech.2005.02.021.

Song, Z. *et al.* (2020) ‘Evaluation of *Pantoea eucalypti* FBS135 for pine (*Pinus massoniana*) growth promotion and its genome analysis’, *Journal of Applied Microbiology*, 129(4), pp. 958–970. Available at: https://doi.org/10.1111/jam.14673.

Soutar, C.D. and Stavrinides, J. (2018) ‘The evolution of three siderophore biosynthetic clusters in environmental and host-associating strains of *Pantoea*’, *Molecular Genetics and Genomics*, 293(6), pp. 1453–1467. Available at: https://doi.org/10.1007/s00438-018-1477-7.

Soutar, C.D. and Stavrinides, J. (2019) ‘Molecular validation of clinical *Pantoea* isolates identified by MALDI-TOF’, *PLOS ONE*. Edited by P.V. Morais, 14(11), p. e0224731. Available at: https://doi.org/10.1371/journal.pone.0224731.

Spitz, O. *et al.* (2019) ‘Type I Secretion Systems—One Mechanism for All?’, *Microbiology Spectrum*. Edited by M. Sandkvist, E. Cascales, and P.J. Christie, 7(2), p. 7.2.12. Available at: https://doi.org/10.1128/microbiolspec.PSIB-0003-2018.

Stice, S.P. *et al.* (2018) ‘*Pantoea ananatis* Genetic Diversity Analysis Reveals Limited Genomic Diversity as Well as Accessory Genes Correlated with Onion Pathogenicity’, *Frontiers in Microbiology*, 9, p. 184. Available at: https://doi.org/10.3389/fmicb.2018.00184.

Stice, S.P. *et al.* (2020) ‘Thiosulfinate Tolerance Is a Virulence Strategy of an Atypical Bacterial Pathogen of Onion’, *Current Biology*, 30(16), pp. 3130-3140.e6. Available at: https://doi.org/10.1016/j.cub.2020.05.092.

Stogios, P.J. *et al.* (2014) ‘Potential for Reduction of Streptogramin A Resistance Revealed by Structural Analysis of Acetyltransferase VatA’, *Antimicrobial Agents and Chemotherapy*, 58(12), pp. 7083–7092. Available at: https://doi.org/10.1128/AAC.03743-14.

Sulja, A. *et al.* (2022) ‘Comparative genomics to examine the endophytic potential of *Pantoea agglomerans* DAPP-PG 734’, *BMC Genomics*, 23(1), p. 742. Available at: https://doi.org/10.1186/s12864-022-08966-y.

Tettelin, H. *et al.* (2005) ‘Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae* : Implications for the microbial “pan-genome”’, *Proceedings of the National Academy of Sciences*, 102(39), pp. 13950–13955. Available at: https://doi.org/10.1073/pnas.0506758102.

Tettelin, H. *et al.* (2008) ‘Comparative genomics: the bacterial pan-genome’, *Current Opinion in Microbiology*, 11(5), pp. 472–477. Available at: https://doi.org/10.1016/j.mib.2008.09.006.

Tho, K.E. *et al.* (2019) ‘Bacteria Associated with Onion Foliage in Michigan and Their Copper Sensitivity’, *Plant Health Progress*, 20(3), pp. 170–177. Available at: https://doi.org/10.1094/PHP-03-19-0022-RS.

Tiwari, S. and Beriha, S.S. (2015) ‘*Pantoea* species causing early onset neonatal sepsis: a case report’, *Journal of Medical Case Reports*, 9(1), p. 188. Available at: https://doi.org/10.1186/s13256-015-0670-0.

Tobarran, N. *et al.* (2023) ‘Cotton fever: A case report and review of the literature’, *JEM Reports*, 2(2), p. 100030. Available at: https://doi.org/10.1016/j.jemrpt.2023.100030.

Unni, R. *et al.* (2022) ‘Presence and absence of type VI secretion systems in bacteria: This article is part of the Bacterial Cell Envelopes collection.’, *Microbiology*, 168(4). Available at: https://doi.org/10.1099/mic.0.001151.

Urban, M. *et al.* (2019) ‘PHI-base: the pathogen–host interactions database’, *Nucleic Acids Research*, p. gkz904. Available at: https://doi.org/10.1093/nar/gkz904.

Urban, M. *et al.* (2021) ‘PHI-base in 2022: a multi-species phenotype database for Pathogen–Host Interactions’, *Nucleic Acids Research*, 50(D1), pp. D837–D847. Available at: https://doi.org/10.1093/nar/gkab1037.

Urban, M. *et al.* (2024) ‘PHI-base – the multi-species pathogen–host interaction database in 2025’, *Nucleic Acids Research*, p. gkae1084. Available at: https://doi.org/10.1093/nar/gkae1084.

Valbuena-Rodríguez, J.L. *et al.* (2024) ‘Isolation and characterization of *Pantoea ananatis* and *P. agglomerans* in quinoa: *P. ananatis* as a potential fungal biocontroller and plant growth promoter’, *International Microbiology* [Preprint]. Available at: https://doi.org/10.1007/s10123-024-00608-5.

Vale, F.F., Lehours, P. and Yamaoka, Y. (2022) ‘Editorial: The Role of Mobile Genetic Elements in Bacterial Evolution and Their Adaptability’, *Frontiers in Microbiology*, 13, p. 849667. Available at: https://doi.org/10.3389/fmicb.2022.849667.

Vasu, K. and Nagaraja, V. (2013) ‘Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense’, *Microbiology and Molecular Biology Reviews*, 77(1), pp. 53–72. Available at: https://doi.org/10.1128/MMBR.00044-12.

Völksch, B. *et al.* (2009) ‘Polyphasic study of plant- and clinic-associated *Pantoea agglomerans* strains reveals indistinguishable virulence potential’, *Infection, Genetics and Evolution*, 9(6), pp. 1381–1391. Available at: https://doi.org/10.1016/j.meegid.2009.09.016.

Von Bodman, B. and Farrand, S.K. (1995) ‘Capsular polysaccharide biosynthesis and pathogenicity in *Erwinia stewartii* require induction by an N-acylhomoserine lactone autoinducer’, *Journal of Bacteriology*, 177(17), pp. 5000–5008. Available at: https://doi.org/10.1128/jb.177.17.5000-5008.1995.

Wagner, S. *et al.* (2018) ‘Bacterial type III secretion systems: a complex device for the delivery of bacterial effector proteins into eukaryotic host cells’, *FEMS Microbiology Letters*, 365(19). Available at: https://doi.org/10.1093/femsle/fny201.

Wallace, I.M. *et al.* (2006) ‘M-Coffee: combining multiple sequence alignment methods with T-Coffee’, *Nucleic Acids Research*, 34(6), pp. 1692–1699. Available at: https://doi.org/10.1093/nar/gkl091.

Walterson, A.M. and Stavrinides, J. (2015) ‘*Pantoea:* insights into a highly versatile and diverse genus within the *Enterobacteriaceae*’, *FEMS Microbiology Reviews*. Edited by J.R. Van Der Meer, 39(6), pp. 968–984. Available at: https://doi.org/10.1093/femsre/fuv027.

Wang, J. *et al.* (2022) ‘Pantoea agglomerans, a seed-borne plant pathogenic bacterium, decreased seed germination, seedling growth and seed quality of oat’, *European Journal of Plant Pathology*, 162(3), pp. 667–679. Available at: https://doi.org/10.1007/s10658-021-02430-5.

Wang, M. *et al.* (2023) ‘ICEberg 3.0: functional categorization and analysis of the integrative and conjugative elements in bacteria’, *Nucleic Acids Research*, 52(D1), pp. D732–D737. Available at: https://doi.org/10.1093/nar/gkad935.

Wang, S. *et al.* (2012) ‘Fighting malaria with engineered symbiotic bacteria from vector mosquitoes’, *Proceedings of the National Academy of Sciences*, 109(31), pp. 12734–12739. Available at: https://doi.org/10.1073/pnas.1204158109.

Weinthal, D.M. *et al.* (2007) ‘Distribution and Replication of the Pathogenicity Plasmid pPATH in Diverse Populations of the Gall-Forming Bacterium *Pantoea agglomerans*’, *Applied and Environmental Microbiology*, 73(23), pp. 7552–7561. Available at: https://doi.org/10.1128/AEM.01511-07.

Weller-Stuart, T., De Maayer, P. and Coutinho, T. (2017) ‘*Pantoea ananatis* : genomic insights into a versatile pathogen’, *Molecular Plant Pathology*, 18(9), pp. 1191–1198. Available at: https://doi.org/10.1111/mpp.12517.

Williams, A.N. *et al.* (2020) ‘The Integration of Genome Mining, Comparative Genomics, and Functional Genetics for Biosynthetic Gene Cluster Identification’, *Frontiers in Genetics*, 11, p. 600116. Available at: https://doi.org/10.3389/fgene.2020.600116.

Williams, A.N. and Stavrinides, J. (2020) ‘*Pantoea* Natural Product 3 is encoded by an eight-gene biosynthetic gene cluster and exhibits antimicrobial activity against multi-drug resistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa*’, *Microbiological Research*, 234, p. 126412. Available at: https://doi.org/10.1016/j.micres.2020.126412.

Winnen, B., Hvorup, R.N. and Saier, M.H. (2003) ‘The tripartite tricarboxylate transporter (TTT) family’, *Research in Microbiology*, 154(7), pp. 457–465. Available at: https://doi.org/10.1016/S0923-2508(03)00126-8.

Winnenburg, R. *et al.* (2005) ‘PHI-base: a new database for pathogen host interactions’, *Nucleic Acids Research*, 34(90001), pp. D459–D464. Available at: https://doi.org/10.1093/nar/gkj047.

Wishart, D.S. *et al.* (2023) ‘PHASTEST: faster than PHASTER, better than PHAST’, *Nucleic Acids Research*, 51(W1), pp. W443–W450. Available at: https://doi.org/10.1093/nar/gkad382.

Wright, G. *et al.* (2024) ‘A Broad Spectrum Lasso Peptide Antibiotic Targeting the Bacterial Ribosome’. In Review. Available at: https://doi.org/10.21203/rs.3.rs-5058118/v1.

Wright, S.A.I. *et al.* (2001) '*Pantoea agglomerans* Strain EH318 Produces Two Antibiotics That Inhibit *Erwinia amylovora* In Vitro', *Applied and Environmental Microbiology*, 67(1), pp. 284–292. Available at: https://doi.org/10.1128/AEM.67.1.284-292.2001.

Wright, S.A.I. *et al.* (2006) 'THE BIOSYNTHETIC GENES OF PANTOCIN A AND PANTOCIN B OF *PANTOEA AGGLOMERANS* EH318', *Acta Horticulturae*, (704), pp. 313–320. Available at: https://doi.org/10.17660/ActaHortic.2006.704.45.

Xiao, J. *et al.* (2023) 'Complete Genome Sequence of *Pantoea agglomerans* CHTF15, a Walnut Pathogen', *Molecular Plant-Microbe Interactions®*, 36(2), pp. 134–137. Available at: https://doi.org/10.1094/MPMI-10-22-0216-A.

Xu, F. *et al.* (2022) 'Isolation, Identification, and Whole-Genome Sequence Analysis of *Pantoea Agglomerans* Causing Leaf Spot-Hole Disease in Wild Apricot in Xinjiang, China'. In Review. Available at: https://doi.org/10.21203/rs.3.rs-1178673/v1.

Yibar, A., Ajmi, N. and Duman, M. (2024) 'First report and genomic characterization of *Escherichia coli* O111:H12 serotype from raw mussels in Türkiye', *BMC Genomics*, 25(1), p. 1027. Available at: https://doi.org/10.1186/s12864-024-10945-4.

Yin, Z. *et al.* (2019) 'Comparative genomic analysis of the *Hafnia* genus reveals an explicit evolutionary relationship between the species *alvei* and *paralvei* and provides insights into pathogenicity', *BMC Genomics*, 20(1), p. 768. Available at: https://doi.org/10.1186/s12864-019-6123-1.

Yoon, B.-J. (2009) 'Hidden Markov Models and their Applications in Biological Sequence Analysis', *Current Genomics*, 10(6), pp. 402–415. Available at: https://doi.org/10.2174/138920209789177575.

Yu, N.Y. *et al.* (2010) 'PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes', *Bioinformatics*, 26(13), pp. 1608–1615. Available at: https://doi.org/10.1093/bioinformatics/btq249.

Yu, W. *et al.* (2025) 'The First Infant Bloodstream Infection Caused by *Pantoea dispersa* in China: A Case Report and Literature Review', *Infection and Drug Resistance*, Volume 18, pp. 661–667. Available at: https://doi.org/10.2147/IDR.S496299.

Zamorano, A. *et al.* (2022) '*Pantoea agglomerans*-Induced Dieback in Pistachio in Chile', *Horticulturae*, 8(11), p. 1052. Available at: https://doi.org/10.3390/horticulturae8111052.

Zeng, Q. *et al.* (2020) 'Complete Genome Sequence of *Pantoea ananatis* Strain NN08200, an Endophytic Bacterium Isolated from Sugarcane', *Current Microbiology*, 77(8), pp. 1864–1870. Available at: https://doi.org/10.1007/s00284-020-01972-x.

Zhao, Y. *et al.* (2014) 'PanGP: A tool for quickly analyzing bacterial pan-genome profile', *Bioinformatics*, 30(9), pp. 1297–1299. Available at: https://doi.org/10.1093/bioinformatics/btu017.

Zhou, D., Han, Y. and Yang, R. (2006) 'Molecular and physiological insights into plague transmission, virulence and etiology', *Microbes and Infection*, 8(1), pp. 273–284. Available at: https://doi.org/10.1016/j.micinf.2005.06.006.

Zhu, M. *et al.* (2005) 'Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization', *Microbiology*, 151(7), pp. 2363–2372. Available at: https://doi.org/10.1099/mic.0.27888-0.

Zhu, S. *et al.* (2016) 'The B1 Protein Guides the Biosynthesis of a Lasso Peptide', *Scientific Reports*, 6(1), p. 35604. Available at: https://doi.org/10.1038/srep35604.

Zhu, S. *et al.* (2023) 'The common origin and degenerative evolution of flagella in *Actinobacteria*', *mBio*. Edited by I.B. Jouline, 14(6), pp. e02526-23. Available at: https://doi.org/10.1128/mbio.02526-23.

Zrelovs, N. *et al.* (2023) 'Three Phages One Host: Isolation and Characterization of *Pantoea agglomerans* Phages from a Grasshopper Specimen', *International Journal of Molecular Sciences*, 24(3), p. 1820. Available at: https://doi.org/10.3390/ijms24031820.

Appendices

Appendix A

Table 1: Metadata set

#Organism/Name	0AStrain	TaxID	BioProject Accession	BioSample Accession	Assembly Accession	BUSCO completeness	Country of origin		Isolation Source	Lifestyle	Group	SubGroup	WGS	Genome size (bp)	Ave GC%	Chromosome size (bp)	Total no. of proteins
Pantoea agglomerans	20TX0122	549	PRJNA813887	SAMN26527452	GCA_022585235.1	0.996	USA	Plant	Allium cepa (Onion)	Plant pathogen - bulb rot	Proteobacteria	Gammaproteobacteria	JAKZMS01	5015400	55.2	4109768	4581
Pantoea agglomerans	62d	549	PRJNA547115	SAMN12024746	GCA_014138195.1	0.995	USA	Plant	N/A	N/A	Proteobacteria	Gammaproteobacteria	JACGXH01	4753502	55.58	3951557	4312
Pantoea agglomerans	62e	549	PRJNA547116	SAMN12024817	GCA_014138175.1	0.995	USA	Plant	N/A	N/A	Proteobacteria	Gammaproteobacteria	JACGXI01	4751230	55.57	3952064	4314
Pantoea agglomerans	9Rz4		PRJNA861005	SAMN29880859	GCA_030177195.1	0.995	Germany	Plant	Oilseed rape (Rapseed)	Rhizosphere	Proteobacteria	Gammaproteobacteria	JANHBZ01	4977562	55.43	4061525	4500
Pantoea agglomerans	A2	549	PRJNA578090	SAMN13563990	GCA_010667965.1	0.995	Germany	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSSP01	4980702	55.54	4026248	4555
Pantoea agglomerans	AB378		PRJNA903092	SAMN31782636	GCA_026625085.1	0.995	Switzerland	Plant	Red topaz apple blossom	Biocontrol agent (antimicrobial activity)	Proteobacteria	Gammaproteobacteria		4781164	55.58	4042201	4303
Pantoea agglomerans	AR1a	549	PRJNA642846	SAMN15405801	GCA_017474165.1	0.996	USA	Plant	Allium cepa (Onion)	Plant pathogen - bulb rot	Proteobacteria	Gammaproteobacteria	-	5077082	55.24	4169790	4602
Pantoea agglomerans	ASB05	549	PRJNA594723	SAMN13527266	GCA_009762725.1	0.995	USA	Plant	Fresh cherries	Plant epiphyte - fruit	Proteobacteria	Gammaproteobacteria	-	4858648	55.47	4022781	4385
Pantoea agglomerans	B1	549	PRJNA578090	SAMN13563991	GCA_010667985.1	0.995	Germany	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSSQ01	5249320	55.61	4035831	4775
Pantoea agglomerans	B3	549	PRJNA578090	SAMN13563992	GCA_010667975.1	0.995	Germany	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSSR01	5239263	54.53	4025802	4777
Pantoea agglomerans	BAV 2934	549	PRJNA576872	SAMN13012102	GCA_009765475.1	0.995	USA	Environmental	Rain water	Bioncontrol agent	Proteobacteria	Gammaproteobacteria	WHOZ01	4944994	55.56	4003977	4492
Pantoea agglomerans	BI3	549	PRJNA475250	SAMN09382822	GCA_003258435.1	0.995	USA	Plant	Humulus lupulus (hops)	Plant epiphyte - roots	Proteobacteria	Gammaproteobacteria	QLAA01	5058520	55.68	4069916	4644
Pantoea agglomerans	C1	549	PRJNA523737	SAMN10986840	GCA_009759885.1	0.995	Italy	Plant	Hydroponic lettuce leaf	Plant epiphyte - PGPR	Proteobacteria	Gammaproteobacteria	SMLN01	4863707	54.98	4039687	4431
Pantoea agglomerans	C410P1	549	PRJNA335437	SAMN05444153	GCA_001709315.1	0.987	China	Plant	Lettuce	Plant epiphyte - biocontrol agent	Proteobacteria	Gammaproteobacteria	-	5115241	55.54	4182028	4659
Pantoea agglomerans	CB1		PRJNA642846	SAMN21555674	GCA_029016125.1	0.995	USA	Plant	Allium cepa (onion)	Plant pathogen - centre rot	Proteobacteria	Gammaproteobacteria		5048993	55.5	4145235	4608
Pantoea agglomerans	CCNPK 873		PRJNA798255	SAMN25064068	GCA_023614195.1	0.995	China	Plant	maize xylem sap	Biocontrol agent	Proteobacteria	Gammaproteobacteria	JAKKEQ01	4751625	55.58	3987440	4286
Pantoea agglomerans	CFBP13505	549	PRJNA459521	SAMN09062489	GCA_005233495.1	0.996	France	Plant	Raphanus sativus var. Flamboyant 5 (radish)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	QFZW01	4895124	55.36	4054063	4457
Pantoea agglomerans	CFBP13569	549	PRJNA665156	SAMN16237903	GCA_014839115.1	0.995	France	Plant	Raphanus sativus (radish)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNG01	4807031	55.67	4013073	4396
Pantoea agglomerans	CFBP13583	549	PRJNA665156	SAMN16237032	GCA_014839545.1	0.995	France	Plant	Phaseolus valguris (common bean)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNZ01	4791367	55.47	4073037	4340
Pantoea agglomerans	CFBP13600	549	PRJNA665156	SAMN16236980	GCA_014842695.1	0.995	France	Plant	Phaseolus valguris (common bean)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYVR01	4871119	55.59	4040197	4469
Pantoea agglomerans	CFBP13616	549	PRJNA665156	SAMN16237057	GCA_014839425.1	0.995	France	Plant	Phaseolus valguris (common bean)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNT01	4779714	55.53	4041412	4346
Pantoea agglomerans	CFBP13731	549	PRJNA665156	SAMN16237466	GCA_014839315.1	0.996	France	Plant	Brassica napus (rapeseed)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNO01	5033642	57.32	4095764	4601
Pantoea agglomerans	CFBP8756	549	PRJNA665156	SAMN16237653	GCA_014839205.1	0.996	France	Plant	Brassica napus (rapeseed)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNK01	5026583	57.63	4090432	4606
Pantoea agglomerans	CFBP8784	549	PRJNA665156	SAMN16237910	GCA_014839125.1	0.995	France	Plant	Raphanus sativus (radish)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNE01	4777212	55.64	4039473	4322
Pantoea agglomerans	CFBP8786	549	PRJNA665156	SAMN16237914	GCA_014839015.1	0.996	France	Plant	Raphanus sativus (radish)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JACYNA01	4975611	55.23	4132832	4563
Pantoea agglomerans	CFSAN033090	1690502	PRJNA283309	SAMN03897582	GCA_001264395.1	0.995	Chile	Plant	Walnut tree	Plant pathogen - walnut blight	Proteobacteria	Gammaproteobacteria	LGYX01	4956095	55.46	4040997	4559
Pantoea agglomerans	CFSAN047153	549	PRJNA243331	SAMN09909699	GCA_004136415.1	0.993	USA	Plant	Apple leaf	Plant epiphyte - fruit	Proteobacteria	Gammaproteobacteria	-	5158055	55.6	4047712	4720
Pantoea agglomerans	CFSAN047154	549	PRJNA243331	SAMN09909698	GCA_004117135.1	0.993	USA	Plant	Apple leaf	Plant epiphyte - fruit	Proteobacteria	Gammaproteobacteria	-	5158055	55.6	4047712	4718
Pantoea agglomerans	CHTF15		PRJNA870770	SAMN30435625	GCA_024799805.1	0.995	China	Plant	Leaves of diseased walnut trees	Plant pathogen	Proteobacteria	Gammaproteobacteria	NYF24	4820607	55.69	4012069	4379
Pantoea agglomerans	DAPP-PG734		PRJEB52945	SAMN02843220	GCA_943184895.1	0.995	Italy	Plant	Olive plant	Plant endophyte - olive knots (biocontrol activity)	Proteobacteria	Gammaproteobacteria	JNVA01	5396424	55.15	4410564	4960
Pantoea agglomerans	DBM 3797	549	PRJNA774971	SAMN22600026	GCA_020783495.1	0.995	Czech Republic	Plant	Humulus lupulus (hops)	Plant epiphyte - hops cones	Proteobacteria	Gammaproteobacteria	-	4827526	55.52	4089578	4365
Pantoea agglomerans	DOAB1048	549	PRJNA633159	SAMN14934319	GCA_013201565.1	0.995	Canada	Plant	Triticum aestivum (common wheat)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JABLUT01	4790635	55.71	4010183	4357
Pantoea agglomerans	E325	549	PRJNA658485	SAMN15872107	GCA_014353865.1	0.995	USA	Plant	Mallus domestica (apple)	Plant epiphyte - biocontrol agent	Proteobacteria	Gammaproteobacteria	JACSWT01	4783752	55.7	3963873	4345
Pantoea agglomerans	E325-699	549	PRJNA658485	SAMN15872108	GCA_014353845.1	0.995	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWS01	4812495	55.69	3983667	4345
Pantoea agglomerans	E325-705	549	PRJNA658485	SAMN15872109	GCA_014353835.1	0.997	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWR01	4695825	55.69	3978706	4235
Pantoea agglomerans	E325-750	549	PRJNA658485	SAMN15872110	GCA_014353995.1	0.997	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWQ01	4811308	55.7	3980964	4345
Pantoea agglomerans	E325-754	549	PRJNA658485	SAMN15872111	GCA_014353955.1	0.995	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWP01	4804018	55.69	3975281	4339
Pantoea agglomerans	E325-ad1	549	PRJNA658485	SAMN15872112	GCA_014353935.1	0.997	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWO01	4696086	55.68	3986179	4238
Pantoea agglomerans	E325-ad2	549	PRJNA658485	SAMN15872113	GCA_014354005.1	0.997	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWN01	4690352	55.68	3984697	4239
Pantoea agglomerans	Eh318	1408177	PRJNA222832	SAMN02768228	GCA_000687245.1	0.995	USA	Plant	Mallus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	AXOF01	5037315	55.19	4157977	4552
Pantoea agglomerans	EKM10T	2708058	PRJNA605402	SAMN14056937	GCA_011800685.1	0.995	Canada	Plant	Fresh field cucumber	Plant epiphyte - seed mucilage	Proteobacteria	Gammaproteobacteria	JAALEU01	4724926	55.52	3999316	4287
Pantoea agglomerans	EKM20T	2708059	PRJNA605422	SAMN14060167	GCA_011800485.1	0.995	Canada	Plant	Cantaloupe melon	Plant epiphyte - seed mucilage	Proteobacteria	Gammaproteobacteria	JAALFW01	4906948	55.73	4083691	4477
Pantoea agglomerans	EKM21T	2708060	PRJNA605422	SAMN14064680	GCA_011800405.1	0.995	Canada	Plant	Cantaloupe melon	Plant epiphyte - seed mucilage	Proteobacteria	Gammaproteobacteria	JAALFV01	4898830	55.42	4078058	4457
Pantoea agglomerans	EKM22T	2708061	PRJNA605422	SAMN14066448	GCA_011800475.1	0.995	Canada	Plant	Cantaloupe melon	Plant epiphyte - seed mucilage	Proteobacteria	Gammaproteobacteria	JAALFX01	4903849	55.52	4083806	4454
Pantoea agglomerans	FDAARGOS1447	549	PRJNA231221	SAMN16357589	GCA_019048385.1	0.995	Zimbabwe	Human	Knee laceration	Clinical pathogen	Proteobacteria	Gammaproteobacteria	-	4692018	55.44	3999686	4233
Pantoea agglomerans	GD03966		PRJNA868296	SAMN30525422	GCA_029838845.1	0.996	Pakistan	Environmental	surface swab used in a sink drain (hospital)		Proteobacteria	Gammaproteobacteria	JAOBZL01	4853300	55.44	4131879	4403
Pantoea agglomerans	GR13	549	PRJEB20710	SAMN02744765	GCA_900182855.1	0.995	USA	Plant	Switchgrass	Plant epiphyte	Proteobacteria	Gammaproteobacteria	FXUX01	4755241	55.32	4038190	4306
Pantoea agglomerans	IG1	1110694	PRJDB7	SAMD00036596	GCA_000241285.2	0.995	Japan	Plant	Triticum aestivum (common wheat)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	BAEF01	4836682	55.48	4043890	4412
Pantoea agglomerans	JM1	549	PRJNA386857	SAMN07136241	GCA_002222515.1	0.995	Slovakia	Environmental	Soil (contaminated with antibiotics)	N/A	Proteobacteria	Gammaproteobacteria	NHAS01	4816214	55.59	4043554	4390
Pantoea agglomerans	JZB 2120015	1592631	PRJNA513927	SAMN10716182	GCA_014156615.1	0.996	China	Fungi	Pleurotus eryngii (king trumpet mushroom) - fruiting bodies		Proteobacteria	Gammaproteobacteria	SBFD01	4910805	55.48	4115477	4464
Pantoea agglomerans	K1	549	PRJNA531280	SAMN11355438	GCA_007050955.1	0.996	South Korea	Plant	Fermented baechu kimchi	Isolated from fermented food	Proteobacteria	Gammaproteobacteria	SRSB01	5012459	55.55	4055009	4554
Pantoea agglomerans	KM1	549	PRJNA613157	SAMN14395380	GCA_012241415.1	0.996	South Korea	Plant	Fermented baechu kimchi	Isolated from fermented food	Proteobacteria	Gammaproteobacteria	JAAVXI01	5009979	55.55	4052596	4553
Pantoea agglomerans	L15	549	PRJNA386631	SAMN07109613	GCA_003860325.1	0.995	Poland	Plant	Hypericum perphoratum (St. John's wort)	Plant epiphyte - leaves	Proteobacteria	Gammaproteobacteria	-	4858869	55.58	4029228	4422
Pantoea agglomerans	MP2	549	PRJNA254768	SAMN02905153	GCA_000757415.2	0.995	South Africa	Insect	Macrotermes natalensis (fungus-growing termite)	Termite gut	Proteobacteria	Gammaproteobacteria	JPKQ01	4735277	55.63	3988816	4319
Pantoea agglomerans	MR10	549	PRJNA547119	SAMN12024877	GCA_014138245.1	0.995	USA	Plant	N/A	N/A	Proteobacteria	Gammaproteobacteria	JACGXK01	5024161	55.55	4000999	4561
Pantoea agglomerans	MR19	549	PRJNA547117	SAMN12024926	GCA_014138205.1	0.995	USA	Plant	N/A	N/A	Proteobacteria	Gammaproteobacteria	JACGXJ01	5027458	55.55	4003638	4559
Pantoea agglomerans	NBBC-01	549	PRJNA851950	SAMN29257230	GCA_030710645.1	-	China	Plant	Rottenb Potato Tubers	Plant pathogen	Proteobacteria	Gammaproteobacteria		4927193	66.6	4007306	4492
Pantoea agglomerans	NC	549	PRJNA471388	SAMN09209272	GCA_015711555.1	0.994	China	Plant	Arachis hypogaea (peanut)	N/A	Proteobacteria	Gammaproteobacteria	QIBS01	4526756	55.53	4035987	4095
Pantoea agglomerans	NCTC10500	549	PRJEB6403	SAMEA80454418	GCA_900454505.1	0.996	United Kingdom	Human	Cervix	Clinical isolate	Proteobacteria	Gammaproteobacteria	UGSM01	4655521	55.55	4090220	4253
Pantoea agglomerans	NCTC10601	549	PRJEB6403	SAMEA104338353	GCA_900454395.1	0.994	France	Environmental	Water	N/A	Proteobacteria	Gammaproteobacteria	UGSL01	4752160	55.6	4072630	4291
Pantoea agglomerans	OVA06A	549	PRJNA749625	SAMN20394466	GCA_019969965.1	0.995	USA	Plant	Vegetable	Antibiotic resistance	Proteobacteria	Gammaproteobacteria	JAHXYL01	4873762	55.58	4075337	4423
Pantoea agglomerans	OVA07A	2862677	PRJNA749625	SAMN20394467	GCA_019969975.1	0.995	USA	Plant	Vegetable	Antibiotic resistance	Proteobacteria	Gammaproteobacteria	JAHXYM01	4801905	55.67	4066654	4326
Pantoea agglomerans	P10c	549	PRJNA292535	SAMN03996017	GCA_001288285.1	0.995	New Zealand	Plant	Pear tree	Plant epiphyte - pear blossoms	Proteobacteria	Gammaproteobacteria	LIME01	4791557	55.36	4050618	4365
Pantoea agglomerans	Pa	470934	PRJNA369420	SAMN06285540	GCA_002082355.1	0.995	Algeria	Plant	Durum wheat	Plant epiphyte - rhizosphere PGPR	Proteobacteria	Gammaproteobacteria	MUJJ01	4792688	55.42	4083173	4330
Pantoea agglomerans	Pa17-1	549	PRJNA658485	SAMN15872087	GCA_014354235.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXN01	4832886	55.43	4005172	4406
Pantoea agglomerans	Pa17-5	549	PRJNA658485	SAMN15872088	GCA_014354165.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXM01	4957367	55.66	4042504	4510
Pantoea agglomerans	Pa21-13	549	PRJNA658485	SAMN15872092	GCA_014354155.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXI01	4793834	55.47	4006386	4370
Pantoea agglomerans	Pa21-15	549	PRJNA658485	SAMN15872093	GCA_014354065.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXH01	4789821	55.64	4051844	4392
Pantoea agglomerans	Pa21-3	549	PRJNA658485	SAMN15872089	GCA_014354135.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXL01	5080599	55.28	4072711	4618
Pantoea agglomerans	Pa21-5	549	PRJNA658485	SAMN15872090	GCA_014354145.1	0.995	Canada	Plant	Pyrus communis (european pear)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXK01	4847038	55.55	4046287	4392
Pantoea agglomerans	Pa31-3	549	PRJNA658485	SAMN15872095	GCA_014353705.1	0.995	Canada	Plant	Malus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXF01	5110483	55.5	4113125	4687
Pantoea agglomerans	Pa31-4	549	PRJNA658485	SAMN15872096	GCA_014354045.1	0.995	Canada	Plant	Malus domestica (apple)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXE01	4992970	55.37	4091940	4560
Pantoea agglomerans	Pa39-1	549	PRJNA658485	SAMN15872098	GCA_014353745.1	0.995	Canada	Plant	Sorbus aucuparia (rowan)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSXC01	4960295	55.71	4017879	4512
Pantoea agglomerans	Pa39-21	549	PRJNA658485	SAMN15872103	GCA_014354035.1	0.995	Canada	Plant	Crategus cuneata (hawthorn)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWX01	5040326	55.65	4013523	4604
Pantoea agglomerans	Pa39-23	549	PRJNA658485	SAMN15872104	GCA_014353945.1	0.995	Canada	Plant	Sorbus aucuparia (rowan)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWW01	5103566	55.64	4037192	4694
Pantoea agglomerans	Pa39-27	549	PRJNA658485	SAMN15872105	GCA_014353895.1	0.991	Canada	Plant	Crategus monogyna (common hawthorn)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWV01	4999069	54.18	4088307	4586
Pantoea agglomerans	Pa39-7	549	PRJNA658485	SAMN15872101	GCA_014353795.1	0.995	Canada	Plant	Sorbus aucuparia (rowan)	Plant epiphyte	Proteobacteria	Gammaproteobacteria	JACSWZ01	4969232	55.62	3974434	4527
Pantoea agglomerans	Pa58	549	PRJNA795028	SAMN24669166	GCA_022966105.1	0.995	China	Fungi	Fusarium graminearum perithecium	Biocontrol agent	Proteobacteria	Gammaproteobacteria	-	4861200	55.46	4100605	4359
Pantoea agglomerans	PaVv11	470934	PRJEB8255	SAMEA3213013	GCA_900068865.1	0.995	N/A	Plant	Grapevine	Plant endophyte	Proteobacteria	Gammaproteobacteria	CEFP01	4855725	54.89	3985932	4411
Pantoea agglomerans	PMG_056	69393	PRJNA272922	SAMN10678621	GCA_004295855.1	0.995	United Kingdom	Plant	Populus alba (silverleaf)	Plant epiphyte - leaves	Proteobacteria	Gammaproteobacteria	SECZ01	4691543	55.52	3920618	4262
Pantoea agglomerans	PSV1-7	549	PRJNA798364	SAMN25065992	GCA_021609925.1	0.995	Canada	Environmental	Soil (pea-filled)	Biocontrol agent	Proteobacteria	Gammaproteobacteria	-	4914829	55.35	4054466	4467
Pantoea agglomerans	R1	549	PRJNA578090	SAMN13563993	GCA_010667995.1	0.995	Germany	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSSS01	4741314	55.53	4004524	4264
Pantoea agglomerans	R5	549	PRJNA578090	SAMN13563994	GCA_010667945.1	0.995	Germany	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSST01	4825137	55.67	4068315	4387
Pantoea agglomerans	RIT710	549	PRJNA588411	SAMN13243132	GCA_009668105.1	0.995	USA	Fungi	Wild mushrooms	Mushroom endophyte	Proteobacteria	Gammaproteobacteria	WJYF01	4768579	55.69	4007496	4325
Pantoea agglomerans	RIT-PI-AA		PRJNA875050	SAMN30594663	GCA_025642775.1	1	USA	Plant	Toxicodendron radicans (poison ivy)	Plant endophyte	Proteobacteria	Gammaproteobacteria		4840418	55.69	4056922	4371
Pantoea agglomerans	RIT-PI-T		PRJNA875050	SAMN30594656	GCA_025642955.1	0.997	USA	Plant	Toxicodendron radicans (poison ivy)	Plant endophyte	Proteobacteria	Gammaproteobacteria	JANZKK01	4993318	55.52	4077842	4572
Pantoea agglomerans	RSO7	549	PRJEB43400	SAMEA8191770	GCA_905332865.1	0.994	Spain	Plant	Spartina maritima (common cordgrass) - halophyte	Plant epiphyte - rhizosphere	Proteobacteria	Gammaproteobacteria	CAJOSF01	4823630	55.46	4116227	4347

Pantoea agglomerans	SI1_M5	549	PRJNA49341	SAMN02441788	GCA_000220605.2	0.995	USA	Insect	Silex noctilio (invasive woodwasp)	Larval guts - cellulolysis	Proteobacteria	Gammaproteobacteria	ADWZ01	4952634	55.62	4107901	4572
Pantoea agglomerans	strain 18059	2681407	PRJEB34793	SAMEA6372292	GCA_902706165.1	0.995	England	Plant	Fagus sylvatica (beech tree)	Plant epiphyte - tree holes	Proteobacteria	Gammaproteobacteria	CACSJS01	4713717	55.57	3997515	4284
Pantoea agglomerans	strain 190	549	PRJNA248051	SAMN02786245	GCA_000731125.1	0.995	South Korea	Environmental	Apple orchard soil	Bioncontrol agent	Proteobacteria	Gammaproteobacteria	JNGC01	5000005	55.54	4072244	4564
Pantoea agglomerans	strain 3	549	PRJNA315986	SAMN04572589	GCA_001597625.1	0.993	Canada	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	LVHW01	4811315	55.69	4016073	4349
Pantoea agglomerans	strain 4	549	PRJNA255804	SAMN02929135	GCA_000743785.2	0.995	Canada	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	JPOT02	4833670	55.66	3952782	4386
Pantoea agglomerans	strain 4188		PRJNA320975	SAMN04961652	GCA_001662025.2	0.995	USA	Plant	Beta vulgaris	Plant pathogen	Proteobacteria	Gammaproteobacteria		5041798	55.41	4165783	4614
Pantoea agglomerans	strain 824-1		PRJNA320975	SAMN04961658	GCA_001661985.2	0.995	USA	Plant	Gypsophila paniculata	Plant pathogen	Proteobacteria	Gammaproteobacteria		5034205	55.28	4098036	4627
Pantoea agglomerans	T1	549	PRJNA578090	SAMN13563995	GCA_010667925.1	0.996	Turkey	Plant	Triticum aestivum (common wheat)	Plant seed endophyte	Proteobacteria	Gammaproteobacteria	WSSU01	4541858	55.6	4007545	4091
Pantoea agglomerans	TH81	549	PRJNA486198	SAMN09840060	GCA_003704305.1	0.995	USA	Environmental	Forest soil	antibiotic resistance	Proteobacteria	Gammaproteobacteria	-	4984468	55.17	4128817	4500
Pantoea agglomerans	Tx10	1332072	PRJNA202241	SAMN02471931	GCA_000475055.1	0.995	USA	Human	Sputum from a cystic fibrosis patient	Clinical isolate	Proteobacteria	Gammaproteobacteria	ASJI01	4806090	55.52	4003096	4372
Pantoea agglomerans	UAEU18	549	PRJNA602576	SAMN13899307	GCA_010523255.1	0.996	United Arab Emirates	Environmental	Desert soil	N/A	Proteobacteria	Gammaproteobacteria	-	4825350	55.55	4040629	4364
Pantoea agglomerans	ZBG6	470934	PRJNA287868	SAMN03793177	GCA_001238615.1	0.995	France	Environmental	Soil	N/A	Proteobacteria	Gammaproteobacteria	LFQL01	4768948	55.58	3986401	4305
Pantoea agglomerans	ZJU23	549	PRJNA692806	SAMN17370745	GCA_021559955.1	0.996	China	Fungi	Fusarium graminearum perithecium	Biocontrol agent	Proteobacteria	Gammaproteobacteria	-	5061267	55.52	4151813	4572
Pantoea eucalypti	LMG 24197	470933	PRJNA503936	SAMN10385900	GCA_009646115.1	1	Uraguay	Plant	Eucalyptus leaves	Plant pathogen - blight and dieback	Proteobacteria	Gammaproteobacteria	-	4798990	54.57	4035995	4394
Pantoea vagans	LMG 24199	470934	PRJNA505269	SAMN10414043	GCA_004792415.1	0.996	Uganda	Plant	Eucalyptus leaves	Plant pathogen - blight and dieback	Proteobacteria	Gammaproteobacteria	-	4790329	55.62	4050173	4304

Appendix B

Table 2: Distribution of the 100 *P. agglomerans* strains across subsp. 1 and 2 in accordance with the core genome phylogeny.

P. agglomerans subsp. 1 strains	*P. agglomerans* subsp. 2 strains
9Rz4	20TX0122
62d	AR1a
62e	C410P1
A2	CB1
AB378	CFBP8756
ASB05	CFBP8786
B1	CFBP13505
B3	CFBP13731
BAV_2934	DAPP-PG734
BI3	GD03966
C1	JZB_2120015
CCNPK873	K1
CFBP8784	KM1
CFBP13569	NC
CFBP13583	NCTC10500
CFBP13600	NCTC10601
CFBP13616	PA
CFSAN033090	RSO7
CFSAN047153	4188
CFSAN047154	T1
CHTF15	UAEU18
DBM_3797	ZJU23
DOAB1048	
E325	
E325-699	
E325-705	
E325-750	
E325-754	
E325-ad1	
E325-ad2	
Eh318	
EKM10T	
EKM20T	
EKM21T	
EKM2T	
FDAARGOS 1447^{T}	
GR13	
IG1	
JM1	
L15	
MP2	
MR10	
MR19	
NBBC-01	
OVA06A	
OVA07A	
P10c	
Pa17-1	
Pa17-5	
Pa21-3	
Pa21-5	

Pa21-13	
Pa21-15	
Pa31-3	
Pa31-4	
Pa39-1	
Pa39-7	
Pa39-21	
Pa39-23	
Pa39-27	
Pa58	
PaVv1	
PMG_056	
PSV1-7	
R1	
R5	
RIT710	
RIT-PIT-AA	
RIT-PI-T	
SL1_M5	
Strain 3	
Strain 4	
Strain 190	
824-1	
Strain 18059	
TH81	
Tx10	
ZBG6	

Appendix C

Table 3: Points plotted on the pan- and core genome graph for the 100 *P. agglomerans*.

Genome Number	Average (pan-genome)	Standard Deviation (pan-genome)	Average (core genome)	Standard Deviation (core genome)
1	4388.12	130.75	4388.12	130.75
2	4813.85	195.07	4006.71	70.68
3	4986.35	196.96	3926.96	52.55
4	5404.12	175.32	3849.40	34.83
5	5521.10	189.86	3827.85	32.43
6	5565.76	174.72	3824.62	33.87
7	5807.37	178.87	3813.86	32.11
8	6018.99	187.47	3801.79	32.20
9	6142.93	172.63	3796.65	30.69
10	6351.09	172.68	3769.69	24.93
11	6513.32	148.33	3764.32	26.20
12	6760.98	123.37	3761.17	25.00
13	6801.02	106.39	3758.27	25.62
14	6977.05	115.41	3753.83	25.19
15	7194.06	104.15	3742.01	25.69
16	7272.60	90.91	3741.23	23.00
17	7452.21	65.21	3733.90	22.94
18	7581.58	69.71	3728.84	23.72
19	7646.35	68.34	3727.26	24.10
20	7750.82	43.96	3721.82	24.66
21	7827.33	47.98	3720.19	25.34
22	7890.18	44.28	3718.96	24.78
23	8016.27	67.40	3717.68	25.38
24	8145.74	40.59	3715.79	26.01
25	8154.48	40.04	3715.10	26.51
26	8182.55	47.19	3714.56	26.99
27	8269.05	66.33	3713.51	27.47
28	8385.95	52.72	3712.79	27.93
29	8519.92	37.05	3711.54	28.28
30	8536.31	41.60	3712.50	61.07
31	8614.86	82.76	3709.89	26.48
32	8731.99	78.90	3707.05	23.33
33	8834.04	32.98	3692.65	28.04
34	8899.94	36.54	3670.33	31.81
35	8961.69	41.32	3658.74	22.74
36	8971.61	37.14	3658.58	23.13
37	8974.52	36.37	3658.42	23.27
38	8977.14	35.30	3658.06	23.63
39	8979.27	33.99	3657.34	23.82
40	8981.89	32.54	3656.96	24.07
41	8994.41	39.05	3655.78	24.22
42	9049.24	34.48	3651.92	25.07
43	9131.57	37.40	3650.70	25.05
44	9180.78	33.52	3650.51	25.20
45	9206.32	64.45	3650.29	25.55
46	9331.83	53.64	3650.17	25.73
47	9405.02	26.77	3649.33	26.00
48	9476.69	32.89	3645.90	26.60
49	9517.27	25.49	3644.52	26.68
50	9561.88	26.60	3641.81	26.57
51	9595.12	25.95	3638.90	31.33

52	9637.46	25.67	3638.16	31.68
53	9678.82	25.51	3635.75	35.27
54	9692.67	40.56	3635.51	35.55
55	9757.87	31.32	3634.61	36.38
56	9797.30	54.81	3630.94	35.76
57	9868.56	24.91	3629.88	36.48
58	9881.94	35.99	3629.18	36.74
59	9943.82	23.24	3612.08	57.27
60	10015.47	22.72	3522.02	36.21
61	10058.46	23.02	3479.50	4.56
62	10086.75	24.12	3479.50	4.56
63	10132.74	24.16	3479.47	4.61
64	10176.96	23.71	3479.43	4.65
65	10217.78	23.87	3479.04	4.85
66	10269.93	23.70	3473.53	8.59
67	10323.36	25.42	3461.45	9.74
68	10369.14	26.17	3456.44	1.52
69	10411.44	25.10	3456.44	1.52
70	10450.07	24.73	3456.43	1.53
71	10484.24	27.92	3456.34	1.59
72	10527.46	27.64	3456.34	1.59
73	10564.78	36.48	3456.30	1.60
74	10602.68	38.04	3456.23	1.64
75	10640.90	25.83	3456.18	1.65
76	10699.95	24.35	3456.15	1.65
77	10727.74	26.73	3456.13	1.66
78	10771.13	30.64	3456.05	1.69
79	10798.68	30.80	3456.03	1.71
80	10839.22	29.56	3455.92	1.77
81	10880.03	35.16	3455.67	2.11
82	10914.11	31.80	3454.33	2.59
83	10955.80	31.77	3454.26	2.58
84	10992.46	30.83	3452.66	2.14
85	11028.29	33.18	3452.26	2.17
86	11058.24	34.98	3451.88	2.21
87	11103.85	33.38	3450.44	2.54
88	11141.55	38.89	3449.97	2.69
89	11170.06	39.65	3447.91	2.55
90	11221.42	47.95	3447.21	2.57
91	11276.31	50.07	3446.44	3.08
92	11301.53	50.32	3445.28	3.20
93	11393.72	47.35	3444.50	3.56
94	11421.45	47.58	3443.94	3.77
95	11463.32	49.08	3444.00	3.95
96	11508.52	50.41	3439.97	4.25
97	11546.40	46.22	3438.60	4.15
98	11596.31	35.68	3436.34	3.64
99	11636.35484	23.76	3434.03	2.17
100	11672.00		3433.00	

Table 4: Points plotted for the pan- and core genome graph, up to 22 strains, for *P. agglomerans* subsp. 1 and subsp. 2.

	***P. agglomerans* subsp. 1**				***P. agglomerans* subsp. 2**			
Genome Number	Average (pan-genome)	Standard Deviation (pan-genome)	Average (core genome)	Standard Deviation (core genome)	Average (pan-genome)	Standard Deviation (pan-genome)	Average (core genome)	Standard Deviation (core genome)
1	4381.64	120.58	4381.64	120.58	4409.27	163.31	4409.27	163.31
2	4735.44	170.26	4038.89	45.36	4833.51	171.46	3985.03	108.99
3	4899.28	176.80	3956.60	23.41	5192.82	165.47	3879.56	83.52
4	5048.15	195.53	3914.55	17.76	5530.05	142.09	3824.10	70.78
5	5093.57	182.41	3911.19	15.56	5750.21	136.94	3790.01	69.78
6	5346.97	177.74	3892.39	15.79	6043.81	110.87	3763.51	68.88
7	5551.91	178.70	3872.84	14.64	6262.59	100.20	3749.81	67.80
8	5656.74	175.01	3865.27	13.80	6373.72	100.95	3731.66	69.01
9	5797.37	176.79	3858.23	12.53	6497.05	125.02	3725.52	68.64
10	6114.76	118.87	3851.79	12.60	6529.98	106.39	3706.95	71.05
11	6144.85	110.11	3849.76	11.91	6642.97	109.12	3695.40	68.66
12	6285.99	109.37	3847.54	11.60	6891.70	67.21	3679.01	66.64
13	6527.97	47.80	3829.37	8.67	7006.35	61.02	3657.10	65.01
14	6596.30	48.41	3825.40	8.34	7095.84	63.32	3634.97	64.45
15	6672.29	48.50	3822.33	8.54	7191.85	67.61	3618.00	47.73
16	6737.61	51.57	3819.68	8.53	7274.06	70.29	3581.73	15.09
17	6794.93	61.69	3817.17	8.67	7334.87	551.81	3726.55	94.07
18	6921.01	87.03	3816.71	8.34	7404.66	73.42	3565.38	6.50
19	7020.46	87.79	3814.69	8.36	7464.68	71.42	3574.28	20.91
20	7076.35	52.11	3812.94	8.54	7505.21	82.29	3563.17	13.03
21	7231.57	45.57	3811.78	8.70	7559.50	62.46	3558.32	9.12
22	7286.91	47.80	3811.22	8.86	7635.00		3553.00	

Appendix D

Table 5: Plasmid elements in the 100 *P. agglomerans* and two outgroup taxa.

Strain	Chromosome size (bp)	Plasmid	Plasmid size	GC%	Genome size (bp)	Total plasmid size (bp)	%plasmid
9Rz4	4061525	LPP-1	536203	53.44	4977562	916037	18.403
		LPP-2	177238	52.89			
		LPP-13	142672	51.81			
		MPP-1	59924	51.72			
20TX0122	4109768	LPP-1	555299	53.56	5015400	905632	18.057
		LPP-2	173091	51.85			
		LPP-3COMB	117777	50.56			
		MPP-9	51203	51.6			
		SPP-6	5712	52.63			
		SPP-12	2550	46.55			
62d	3951557	LPP-1	619936	53.05	4753502	801945	16.871
		LPP-2	182009	53.13			
62e	3952064	LPP-1	619936	53.05	4751230	799166	16.820
		LPP-2	179230	52.91			
Strain 18059	3997515	LPP-1	553338	53.29	4713717	716202	15.194
		LPP-2	162864	52.64			
A2	4026248	LPP-1	678173	52.45	4980702	954454	19.163
		LPP-2	166033	52.71			
		LPP-10	110248	50.99			
AB378	4042201	LPP-1	555125	53.08	4781164	738963	15.456
		LPP-2	183838	53.06			
AR1a	4169790	LPP-1	557771	53.27	5077082	907292	17.870
		LPP-2	160336	52.94			
		LPP-3A	90466	50.1			
		LPP-3B	98719	55.31			
ASB05	4022781	LPP-1	563807	52.91	4858648	835867	17.204
		LPP-2	207454	52.1			
		MPP-11	64606	52			
B1	4035831	LPP-1	577199	52.96	5249320	1213489	23.117
		LPP-2	202936	51.87			
		LPP-8	182736	48.38			
		MPP-2	64731	50.48			
		MPP-5	35179	46.47			
		LPP-11	110150	47.83			
		MPP-3	40558	49.2			
B3	3994139 31663	LPP-1	577199	52.96	5239263	1213461	23.161
		LPP-2	203008	51.88			
		LPP-8	182737	48.38			
		MPP-2	64604	50.48			
		MPP-5	35206	46.49			
		LPP-11	110150	47.83			
		MPP-3	40557	49.2			

BAV_2934	4003977	LPP-1	528933	53.35	4944994	941017	19.030
		LPP-2	203868	51.84			
		LPP-4	205248	50.19			
		SPP-10	2968	38.31			
BI3	4069916	LPP-1	505988	53.71	5058520	988604	19.543
		LPP-2	178183	53.05			
		LPP-5	205528	51.47			
		LPP-3B	70759	56.19			
		MPP-6	28146	43.79			
C1	3883860	LPP-1	645700	52.98	4863707	824020	16.942
	155827	LPP-2	178320	52.84			
C410P1	4182028	LPP-1	543504	53.24	5115241	933213	18.244
		LPP-2	173592	52.44			
		LPP-3COMB	216117	52.32			
CB1	4145235	LPP-1	569555	53.57	5048993	903758	17.900
		LPP-2	171886	52.52			
		LPP-3COMB	162317	51.04			
CCNPK873	3987440	LPP-1	587110	53.25	4751625	764185	16.083
		LPP-2	177075	52.72			
CFBP8756	4081790	LPP-1	501483	53.68	5026583	936151	18.624
	8642	LPP-2	185001	51.89			
		LPP-6	141560	49.81			
		LPP-7	108107	53.4			
CFBP8784	4039473	LPP-1	556653	53.1	4777212	737739	15.443
		LPP-2	181086	52.95			
CFBP8786	4124982	LPP-1	515999	53.54	4975611	842779	16.938
	7850	LPP-2	183142	51.89			
		LPP-6	140783	49.73			
		SPP-11	2855	48.58			
CFBP13505	2784135	LPP-1	526007	53.27	4895124	841061	17.182
	1269928	LPP-2	185041	51.91			
		LPP-6	130013	49.66			
CFBP13569	4013073	LPP-1	615683	53.02	4807031	793958	16.517
		LPP-2	178275	52.92			
CFBP13583	4073037	LPP-1	538466	53.26	4791367	718330	14.992
		LPP-2	179864	52.99			
CFBP13600	4040197	LPP-1	526898	53.43	4871119	830922	17.058
		LPP-2	180109	53.01			
		MPP-4	38088	44.92			
		LPP-4	85827	52.26			
CFBP13616	4041412	LPP-1	549949	53.01	4779714	738302	15.447
		LPP-2	188353	52.37			
CFBP13731	4088437	LPP-1	501915	53.68	5033642	937878	18.632
	7327	LPP-2	185073	51.89			
		LPP-6	141560	49.8			
		LPP-7	109330	53.32			
CFSAN033090	4040997	LPP-1	520656	53.56	4956095	915098	18.464

		LPP-2	178064	52.95			
		LPP-14	121142	50.34			
		MPP-7	49692	46.29			
		MPP-8	45544	49.97			
CFSAN047153	4047712	LPP-1	613013	53.21	5158055	1110343	21.526
		LPP-2	211326	51.91			
		LPP-5	203713	50.69			
		LPP-3B	82291	56.5			
CFSAN047154	4047712	LPP-1	613013	53.21	5158055	1110343	21.526
		LPP-2	211326	51.91			
		LPP-5	203713	50.69			
		LPP-3B	82291	56.5			
CHTF15	4012069	LPP-1	514938	53.66	4820607	808538	16.773
		LPP-2	180424	53.09			
		LPP-4	113176	52.76			
DAPP-PG734	4410564	LPP-1	530328	53.67	5396424	985860	18.269
		LPP-3COMB	174327	52.61			
		LPP-2	163706	52.63			
		LPP-4	117499	50.78			
DBM_3797	4089578	LPP-1	555522	52.97	4827526	737948	15.286
		LPP-2	182426	53.18			
DOAB1048	4010183	LPP-1	547793	53.07	4790635	780452	16.291
		LPP-2	171177	52.95			
		MPP-12	61482	48.93			
E325	3963873	LPP-1	535779	53.51	4783752	819879	17.139
		LPP-2	173079	52.69			
		LPP-5	111021	48.83			
E325-699	2131063	LPP-1	535748	53.51	4812495	828828	17.222
	1852604	LPP-2	178083	52.8			
		LPP-5	114997	48.98			
E325-705	3978706	LPP-1	535828	53.52	4695825	717119	15.271
		LPP-2	181291	53.07			
E325-750	1251337	LPP-1	535951	53.51	4811308	830344	17.258
	1505786	LPP-2	181167	53.05			
	1223841	LPP-5	113226	48.99			
E325-754	3975281	LPP-1	535806	53.52	4804018	828737	17.251
		LPP-2	179224	52.89			
		LPP-5	113707	49.04			
E325-ad1	3986179	LPP-1	536027	53.51	4696086	709907	15.117
		LPP-2	173880	52.74			
E325-ad2	3984697	LPP-1	535806	53.52	4699352	714655	15.208
		LPP-2	178849	53			
Eh318	4157977	LPP-1	557367	53.42	5037315	879338	17.456
		LPP-2	205021	52.3			
		LPP-3B	116950	50.05			
EKM10T	3999316	LPP-1	544272	53.42	4724926	725610	15.357
		LPP-2	172664	52.92			

		SPP-3	8674	54.97			
EKM20T	3812600	LPP-1	538753	53.54	4906948	823257	16.777
	271091	LPP-2	181890	53.1			
		LPP-3B	102614	53.9			
EKM21T	3284155	LPP-1	529478	53.56	4898830	820772	16.754
	793903	LPP-2	173830	52.8			
		LPP-3B	117464	52.39			
EKM22T	4083806	LPP-1	529478	53.56	4903849	820043	16.722
		LPP-2	175000	52.9			
		LPP-3B	115565	52.27			
FDAARGOS1447^{T}	3999686	LPP-1	511735	53.37	4692018	692332	14.756
		LPP-2	180597	52.61			
GD03966	4131879	LPP-1	550649	53.46	4853300	721421	14.865
		LPP-2	170772	52.56			
GR13	3163405	LPP-1	534640	53.34	4755241	717051	15.079
	874785	LPP-2	182411	53.15			
IG1	4043890	LPP-1	616430	52.65	4836682	792792	16.391
		LPP-2	176362	52.83			
JM1	4043554	LPP-1	582742	52.84	4816214	772660	16.043
		LPP-2	189918	52.75			
JZB_2120015	4115477	LPP-1	625114	53.06	4910805	795328	16.195
		LPP-2	170214	52.64			
K1	4055009	LPP-1	562782	53.6	5012459	957450	19.101
		LPP-2	170651	52.53			
		LPP-3COMB	224017	51.92			
KM1	4052596	LPP-1	562844	53.6	5009979	957383	19.110
		LPP-2	170760	52.54			
		LPP-3COMB	223779	51.89			
L15	4029228	LPP-1	583567	52.96	4858869	829641	17.075
		LPP-2	179590	52.97			
		MPP-10	66484	50.41			
MP2	2240403	LPP-1	560782	53.2	4735277	746461	15.764
	1748413	LPP-2	185679	52.49			
MR10	4000999	LPP-1	558870	53.34	5024161	1023162	20.365
		LPP-2	193574	52.17			
		LPP-5	139549	51.92			
		LPP-4	131169	51.41			
MR19	4003638	LPP-1	558870	53.34	5027458	1023820	20.365
		LPP-2	191974	52.14			
		LPP-5	141809	51.98			
		LPP-4	131167	51.41			
NBBC-01	4007306	LPP-1	531583	53.55	4927193	919887	18.670
		LPP-2	197206	52.54			
		LPP-3B	81391	55.03			
		LPP-4	71619	53.05			
		MPP-4	38088	44.91			
NC	4035987	LPP-1	490769	53.69	4526756	490769	10.842

NCTC10500	4090220	LPP-1	460684	53.46	4655521	565301	12.143
		LPP-4	90067	51.29	9.89543383		
		SPP-8	14550	34.15			
NCTC10601	4072630	LPP-1	497989	53.64	4752160	679530	14.299
		LPP-2	174662	52.06			
		SPP-5	6879	50.52			
OVA06A	4075337	LPP-1	535698	53.36	4873762	798425	16.382
		LPP-2	191384	52.72			
		LPP-4	71343	51.3			
OVA07A	1090599	LPP-1	553233	53.09	4801905	735251	15.312
	2976055	LPP-2	182018	53.11			
P10c	4050618	LPP-1	558805	53.41	4791557	740939	15.463
		LPP-2	182134	53.17			
PA	4083173	LPP-1	539122	53.57	4792688	709515	14.804
		LPP-2	170393	52.58			
Pa17-1	4005172	LPP-1	569750	52.86	4832886	827714	17.127
		LPP-2	171077	52.37			
		LPP-3B	86887	50.95			
Pa17-5	4042504	LPP-1	622481	52.91	4957367	914863	18.455
		LPP-2	174945	52.85			
		LPP-3B	117437	53			
Pa21-3	1126735	LPP-1	588534	52.82	5080599	1007888	19.838
	2945976	LPP-2	181823	53.06			
		LPP-4pt1	141557	52.03			
		LPP-4pt2	14546	47.27			
		LPP-3B	81428	56.52			
Pa21-5	4046287	LPP-1	534247	53.46	4847038	800751	16.520
		LPP-2	191691	52.26			
		LPP-4	74813	52.08			
Pa21-13	2883646	LPP-1	609077	53.3	4793834	787448	16.426
	1122740	LPP-2	174972	52.82			
		SPP-9	3399	47.54			
Pa21-15	592580	LPP-1	534403	53.47	4789821	737977	15.407
	2975385	LPP-2	129005	52.07	2.693315679		
	483879	LPP-4	74569	52.02			
Pa31-3	4113125	LPP-1	576686	53.28	5110483	997358	19.516
		LPP-2	176825	52.86			
		LPP-4	113037	52.32			
		LPP-3B	130810	50.38			
Pa31-4	4091940	LPP-1	537676	53.39	4992970	901030	18.046
		LPP-2	180339	52.91			
		LPP-3COMB	183015	52.58			
Pa39-1	655597	LPP-1	548512	53.38	4960295	942416	18.999
	473002	LPP-2	163995	52.58			
	2889280	LPP-3A	159316	50.83			

		LPP-3B	70593	54.36			
Pa39-7	3974434	LPP-1	524321	53.49	4969232	994798	20.019
		LPP-2	205738	51.93			
		LPP-3COMB	180570	53.29			
		LPP-4	84169	53.8			
Pa39-21	4013523	LPP-1	525869	53.63	5040326	1026803	20.372
		LPP-2	210155	51.81			
		LPP-3COMB	214061	52.75			
		LPP-4	76718	51.02			
Pa39-23	2499838	LPP-1	525842	53.63	5103566	1066374	20.895
	1537354	LPP-2	206369	51.62			
		LPP-3COMB	214146	52.75			
		LPP-4	78835	50.71			
		MPP-6	35597	46.27			
		SPP-2	1187	49.61			
		SPP-1	4398	48.89			
Pa39-27	3826743	LPP-1	650217	52.83	4999069	910762	18.219
	190317	LPP-2	176376	52.83			
	71247	LPP-4	84169	53.8			
Pa58	4100605	LPP-1	554101	53.13	4861200	760595	15.646
		LPP-2	206494	52.87			
PaVv1	3960194	LPP-1	626216	52.81	4855725	869793	17.913
	25738	LPP-2	180851	53.04			
		MPP-1	57334	52.11			
		SPP-7	5392	44.73			
PMG_056	3920618	LPP-1	590820	52.59	4691543	770925	16.432
		LPP-2	180105	52.91			
PSV1-7	4054466	LPP-1	621567	52.54	4914829	860363	17.505
		LPP-2	179340	53			
		LPP-4	59456	52.34			
4188	4165783	LPP-1	541337	53.41	5041798	876015	17.375
		LPP-2	178621	52.36			
		LPP-9	156057	50.82			
824-1	4098036	LPP-1	582658	53.09	5034205	936169	18.596
		LPP-2	143524	53.52			
		LPP-9	131449	50.73			
		LPP-3A	78538	52.01			
R1	4004524	LPP-1	558375	53.07	4741314	736790	15.540
		LPP-2	178415	52.91			
R5	4068315	LPP-1	533549	53.26	4825137	756822	15.685
		LPP-2	187582	52.13			
		MPP-13	35691	47.3			
RIT710	4007496	LPP-1	540995	53.54	4768579	761083	15.960
		LPP-2	220088	52.07	4.615379131		
RIT-PIT-AA	4056922	LPP-1	535530	53.46	4840418	783496	16.187
		LPP-2	177089	52.9			

		LPP-4	70877	51.68			
RIT-PI-T	4077842	LPP-1	537926	53.41	4993318	915476	18.334
		LPP-2	193687	52.93			
		LPP-5	117051	49.56			
		LPP-4	59032	50.28			
		SPP-4	7780	46.72			
RSO7	4116227	LPP-1	533176	53.64	4823630	707403	14.665
		LPP-2	174227	52.36			
SL1_M5	4107901	LPP-1	527046	53.52	4952634	844733	17.056
		LPP-2	178192	53.01			
		LPP-10	139495	49.61			
Strain 3	4016073	LPP-1	612665	53.19	4811315	795242	16.529
		LPP-2	182577	53.19			
Strain 4	3952782	LPP-1	590188	53.05	4833670	880888	18.224
		LPP-2	180841	53.17			
		LPP-3B	109859	52.29			
Strain 190	4072244	LPP-1	538297	53.42	5000005	927761	18.555
		LPP-2	212535	51.97			
		LPP-3B	176929	52.46			
T1	4007545	LPP-1	534313	53.53	4541858	534313	11.764
TH81	4128817	LPP-1	520959	53.69	4984468	855651	17.166
		LPP-2	182169	52.3			
		LPP-12	152523	48.87			
Tx10	4003096	LPP-1	621541	52.9	4806090	802994	16.708
		LPP-2	181453	53.09			
UAEU18	4040629	LPP-1	513383	53.67	4825350	784721	16.262
		LPP-2	184488	52.18			
		LPP-7	86850	54.27			
ZBG6	2829875	LPP-1	539489	53.42	4768948	782547	16.409
	905733	LPP-2	188889	52.67			
	250793	MPP-1	48776	50.4			
		SPP-1	5393	45.48			
ZJU23	4151813	LPP-1	567588	53.41	5061267	909454	17.969
		LPP-2	170989	52.58			
		LPP-3B	170877	54.06			
P. eucalypti LMG 24197^{T}	4035995	LPP-1	529303	52.25	4798990	762995	15.899
		LPP-2	138725	50.9			
		pEUC-1	94967	52.16			
P. vagans LMG 24199^{T}	4050173	LPP-1	559692	54.04	4790329	740156	15.451
		LPP-2	180464	53.06			

Table 6: Summary of plasmid types found in each *P. agglomerans* strain.

Strain	LPP	MPP	SPP	Total
9Rz4	3	1		**4**
20TX0122	3	1	2	**6**
62d	2			**2**
62e	2			**2**
18059	2			**2**
A2	3			**3**
AB378	2			**2**
AR1a	3			**3**
ASB05	2	1		**3**
B1	4	3		**7**
B3	4	3		**7**
BAV_2934	3		1	**4**
BI3	4	1		**5**
C1	2			**2**
C410P1	3			**3**
CB1	3			**3**
CCNPK873	2			**2**
CFBP8756	4			**4**
CFBP8784	2			**2**
CFBP8786	3		1	**4**
CFBP13505	3			**3**
CFBP13569	2			**2**
CFBP13583	2			**2**
CFBP13600	3	1		**4**
CFBP13616	2			**2**
CFBP13731	4			**4**
CFSAN033090	3	2		**5**
CFSAN047153	4			**4**
CFSAN047154	4			**4**
CHTF15	3			**3**
DAPP-PG734	4			**4**
DBM_3797	2			**2**
DOAB1048	2	1		**3**
E325	3			**3**
E325-699	3			**3**
E325-705	2			**2**
E325-750	3			**3**
E325-754	3			**3**
E325-ad1	2			**2**
E325-ad2	2			**2**
Eh318	3			**3**
EKM10T	2		1	**3**
EKM20T	3			**3**
EKM21T	3			**3**
EKM22T	3			**3**
FDAARGOS1447TYPE	2			**2**
GD03966	2			**2**
GR13	2			**2**
IG1	2			**2**
JM1	2			**2**
JZB_2120015	2			**2**
K1	3			**3**
KM1	3			**3**
L15	2	1		**3**

MP2	2			**2**
MR10	4			**4**
MR19	4			**4**
NBBC-01	4	1		**5**
NC	1			**1**
NCTC10500	2		1	**3**
NCTC10601	2		1	**3**
OVA06A	3			**3**
OVA07A	2			**2**
P10c	2			**2**
PA	2			**2**
Pa17-1	3			**3**
Pa17-5	3			**3**
Pa21-3	4			**4**
Pa21-5	3			**3**
Pa21-13	2		1	**3**
Pa21-15	3			**3**
Pa31-3	4			**4**
Pa31-4	3			**3**
Pa39-1	3			**3**
Pa39-7	4			**4**
Pa39-21	4			**4**
Pa39-23	4	1	2	**7**
Pa39-27	3			**3**
Pa58	2			**2**
PaVv1	2	1	1	**4**
PMG_056	2			**2**
PSV1-7	3			**3**
4188	3			**3**
824-1	4			**4**
R1	2			**2**
R5	2	1		**3**
RIT710	2			**2**
RIT-PIT-AA	3			**3**
RIT-PI-T	4		1	**5**
RSO7	2			**2**
SL1_M5	3			**3**
3	2			**2**
4	3			**3**
190	3			**3**
T1	1			**1**
TH81	3			**3**
Tx10	2			**2**
UAEU18	3			**3**
ZBG6	2	1	1	**4**
ZJU23	3			**3**
Total	**273**	**20**	**13**	**306**

Table 7: Prophage elements found on the genomes of *P. agglomerans* and both outgroup taxa.

Strain	Region	Region Length (kb)	Completeness	# Total Proteins	Most Common Phage	GC %	Average GC %	Genome size (bp)	Phage size (kb)	%
9Rz4	Chromosome	42,3	intact	65	PHAGE_Salmon_118970_sal3_NC_031940(25)	51,53	51,15	4977562	90,8	1,824
		17,8	questionable	26	PHAGE_Salmon_SEN34_NC_028699(15)	50,45				
		30,7	intact	39	PHAGE_Erwini_ENT90_NC_019932(16)	51,47				
20TX0122	Chromosome	23	questionable	27	PHAGE_Vibrio_12B8_NC_021073(4)	39,32	44,29	5015400	41,8	0,833
		18,8	incomplete	12	PHAGE_Erwini_vB_EhrS_59_NC_048198(3)	49,26				
62d	Chromosome	41,4	intact	50	PHAGE_Erwini_ENT90_NC_019932(22)	51,38	51,38	4753502	41,4	0,871
62e	Chromosome	41,4	intact	50	PHAGE_Erwini_ENT90_NC_019932(22)	51,38	51,38	4751230	41,4	0,871
strain 18059	Chromosome	35,7	incomplete	26	PHAGE_Escher_500465_1_NC_049342(8)	51,84	51,84	4713717	35,7	0,757
A2	Chromosome	37,4	intact	46	PHAGE_Erwini_ENT90_NC_019932(23)	52,31	52,31	4980702	37,4	0,751
AB378	Chromosome	18,3	incomplete	21	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	48,49	49,08	4781164	41,9	0,876
		23,6	incomplete	35	PHAGE_Erwini_vB_EhrS_59_NC_048198(4)	49,67				
AR1a	Chromosome	31,3	intact	40	PHAGE_Cronob_ESSI_2_NC_047854(24)	52,47	52,47	5077082	31,3	0,616
ASB05	Chromosome	55,7	intact	49	PHAGE_Shigel_SfII_NC_021857(13)	50,42	49,54	4858648	130,9	2,694
		36,6	questionable	36	PHAGE_Erwini_vB_EhrS_59_NC_048198(5)	46,86				
	MPP-11	38,6	incomplete	30	PHAGE_Paenib_Yerffej_NC_048714(2)	51,35				
B1	Chromosome	46	incomplete	26	PHAGE_Escher_500465_1_NC_049342(8)	52,95	53,59	5249320	151,7	2,890

		51,4	intact	48	PHAGE_Salmon_Fels_2_NC_010463(21)	54,72				
		46,9	intact	84	PHAGE_Cronob_ENT47670_NC_019927(14)	51,88				
	LPP-11	7,4	questionable	8	PHAGE_Bacill_BCD7_NC_019515(3)	54,80				
B3	Chromosome	46	incomplete	26	PHAGE_Escher_500465_1_NC_049342(8)	52,95	53,59	5239263	151,7	2,895
		51,4	intact	48	PHAGE_Salmon_Fels_2_NC_010463(21)	54,72				
		46,9	intact	84	PHAGE_Cronob_ENT47670_NC_019927(14)	51,88				
	LPP-11	7,4	questionable	8	PHAGE_Bacill_BCD7_NC_019515(3)	54,80				
BAV_2934	Chromosome	16,4	incomplete	26	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	48,29	50,84	4944994	46,7	0,944
		30,3	intact	42	PHAGE_Salmon_Fels_2_NC_010463(19)	53,39				
BI3	Chromosome	40,5	intact	63	PHAGE_Salmon_118970_sal3_NC_031940(26)	51,69	51,52	5058520	142,6	2,819
		32,7	questionable	51	PHAGE_Salmon_SEN1_NC_029003(9)	48,97				
		22,1	intact	34	PHAGE_Erwini_vB_EhrS_59_NC_048198(6)	49,35				
		47,3	intact	55	PHAGE_Salmon_SEN8_NC_047753(20)	56,05				
C1	Chromosome	8	questionable	9	PHAGE_Salmon_Fels_2_NC_010463(8)	60,51	55,89	4863707	97,6	2,007
		54,8	intact	63	PHAGE_Erwini_ENT90_NC_019932(31)	54,62				
		34,8	intact	41	PHAGE_Klebsi_ST15_OXA48phi14.1_NC_049454(14)	52,54				
C410P1	Chromosome	19,5	questionable	23	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	48,23	52,01	5115241	120,1	2,348

		23,2	questionable	27	PHAGE_Cronob_ENT47670_NC_019927(5)	54,01				
		35,7	intact	51	PHAGE_Entero_Tyrion_NC_031077(4)	54,41				
		41,7	intact	48	PHAGE_Erwini_ENT90_NC_019932(24)	51,38				
CB1	Chromosome	35,3	intact	47	PHAGE_Salmon_SEN8_NC_047753(22)	57,32	51,35	5048993	132,7	2,628
		50,7	intact	69	PHAGE_Salmon_118970_sal3_NC_031940(18)	50,32				
		46,7	intact	47	PHAGE_Erwini_vB_EhrS_59_NC_048198(11)	46,40				
CCNPK873	Chromosome	4,5	incomplete	8	PHAGE_Erwini_ENT90_NC_019932(4)	44,63	47,00	4751625	33,6	0,707
		29,1	incomplete	24	PHAGE_Erwini_vB_EhrS_59_NC_048198(7)	49,37				
CFBP8756	Chromosome	41,7	intact	56	PHAGE_Burkho_BcepMu_NC_005882(31)	55,54	53,67	5026583	91,9	1,828
		30,7	questionable	35	PHAGE_Aggreg_S1249_NC_013597(9)	53,83				
		19,5	questionable	23	PHAGE_Escher_500465_1_NC_049342(8)	51,65				
CFBP8784	Chromosome	22,8	incomplete	34	PHAGE_Erwini_vB_EhrS_59_NC_048198(4)	49,24	51,58	4777212	66,3	1,388
		43,5	intact	45	PHAGE_Erwini_ENT90_NC_019932(24)	53,92				
CFBP8786	Chromosome	21,8	incomplete	24	PHAGE_Salmon_yarpen_NC_048863(5)	53,33	51,95	4975611	106,9	2,148
		42,8	intact	66	PHAGE_Salmon_118970_sal3_NC_031940(18)	51,08				
		42,3	intact	74	PHAGE_Salmon_SEN34_NC_028699(20)	51,45				
CFBP13505	Chromosome	33,4	intact	42	PHAGE_Cronob_ESSI_2_NC_047854(23)	51,98	51,98	4895124	33,4	0,682
CFBP13569	Chromosome	46,8	intact	84	PHAGE_Erwini_vB_EhrS_59_NC_048198(12)	52,05	52,05	4807031	46,8	0,974

CFBP13583	Chromosome	40,4 13,7	intact questionable	57 23	PHAGE_Erwini_ENT90_NC_019932(27) PHAGE_Erwini_ENT90_NC_019932(6)	52,50 49,02	50,76	4791367	54,1	1,129
CFBP13600	Chromosome	55,6 40,8	intact intact	78 52	PHAGE_Salmon_SEN34_NC_028699(24) PHAGE_Erwini_ENT90_NC_019932(24)	51,26 52,96	52,11	4871119	96,4	1,979
CFBP13616	Chromosome	43,1	intact	56	PHAGE_Erwini_ENT90_NC_019932(29)	52,51	52,51	4779714	43,1	0,902
CFBP13731	Chromosome	30,7 39,6	questionable intact	35 55	PHAGE_Aggreg_S1249_NC_013597(9) PHAGE_Burkho_BcepMu_NC_005882(31)	53,83 55,31	54,57	5033642	70,3	1,397
CFSAN033090	Chromosome	17,7 48 36,9	incomplete intact intact	26 69 46	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(9) PHAGE_Salmon_118970_sal3_NC_031940(8) PHAGE_Erwini_ENT90_NC_019932(24)	48,31 48,78 52,99	50,03	4956095	102,6	2,070
CFSAN047153	Chromosome	21,6 11,2 56,6	intact questionable intact	38 19 64	PHAGE_Cronob_phiES15_NC_018454(5) PHAGE_Erwini_ENT90_NC_019932(5) PHAGE_Salmon_118970_sal3_NC_031940(23)	48,79 50,06 53,53	50,79	5158055	89,4	1,733
CFSAN047154	Chromosome	21,6 11,2 56,6	intact questionable intact	38 19 64	PHAGE_Cronob_phiES15_NC_018454(5) PHAGE_Erwini_ENT90_NC_019932(5) PHAGE_Salmon_118970_sal3_NC_031940(23)	48,79 50,06 53,53	50,79	5158055	89,4	1,733
CHTF15	Chromosome	53,4	intact	76	PHAGE_Entero_ES18_NC_006949(14)	50,66	50,66	4820607	53,4	1,108
DAPP-PG734	Chromosome	15,6	intact	26	PHAGE_Erwini_ENT90_NC_019932(6)	46,92	51,20	5396424	132,2	2,450

		39,3	intact	54	PHAGE_Erwini_ENT90_NC_019932(24)	53,40				
		35,9	incomplete	28	PHAGE_Cronob_ENT39118_NC_019934(2)	50,76				
		41,4	questionable	21	PHAGE_Stx2_c_1717_NC_011357(3)	53,73				
DBM3797	Chromosome	41,8	intact	62	PHAGE_Erwini_ENT90_NC_019932(31)	54,26	54,26	4827526	41,8	0,866
DOAB1048	Chromosome	16,1	incomplete	26	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	49,72	51,05	4790635	145,8	3,043
		25,4	incomplete	24	PHAGE_Escher_500465_1_NC_049342(8)	50,83				
		51,3	intact	69	PHAGE_Erwini_vB_EhrS_59_NC_048198(22)	50,15				
		12,2	questionable	14	PHAGE_Erwini_vB_EhrS_49_NC_048197(9)	50,38				
		40,8	intact	45	PHAGE_Klebsi_ST16_OXA48phi5.4_NC_049450(29)	54,18				
E325	Chromosome	40,8	intact	56	PHAGE_Erwini_ENT90_NC_019932(26)	52,61	52,07	4783752	77,1	1,612
		36,3	incomplete	19	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	51,53				
E325-699	Chromosome	36,3	incomplete	19	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	51,53	52,14	4812495	86,4	1,795
		50,1	intact	56	PHAGE_Erwini_ENT90_NC_019932(26)	52,74				
E325-705	Chromosome	36,3	incomplete	19	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	51,53	52,13	4695825	86,5	1,842
		50,2	intact	58	PHAGE_Erwini_ENT90_NC_019932(26)	52,73				

E325-750	Chromosome	36,3 50,1	incomplete intact	19 57	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7) PHAGE_Erwini_ENT90_NC_019932(26)	51,53 52,74	52,14	4811308	86,4	1,796
E325-754	Chromosome	40,8 36,3	intact incomplete	57 19	PHAGE_Erwini_ENT90_NC_019932(26) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	52,61 51,53	52,07	4804018	77,1	1,605
E325-ad1	Chromosome	42,4 36,3 12,9	intact incomplete questionable	41 19 15	PHAGE_Erwini_ENT90_NC_019932(21) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7) PHAGE_Erwini_ENT90_NC_019932(5)	51,89 51,53 56,06	53,16	4696086	91,6	1,951
E325-ad2	Chromosome	50,1 36,3	intact incomplete	56 19	PHAGE_Erwini_ENT90_NC_019932(26) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	52,74 51,53	52,14	4699352	86,4	1,839
Eh318	Chromosome	19,7	questionable	30	PHAGE_Salmon_SEN34_NC_028699(15)	49,87	49,87	5037315	19,7	0,391
EKM10T	Chromosome	33,9 10,4	questionable questionable	23 15	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7) PHAGE_Erwini_ENT90_NC_019932(5)	49,06 51,54	50,30	4724926	44,3	0,938
EKM20T	Chromosome	36,9 48,2 33,7	incomplete intact intact	25 71 46	PHAGE_Escher_500465_1_NC_049342(8) PHAGE_Entero_mEp235_NC_019708(10) PHAGE_Erwini_vB_ErhS_49_NC_048197(14)	53,08 49,73 49,33	50,71	4906948	118,8	2,421

EKM21T	Chromosome	49,8 33,6	intact incomplete	61 31	PHAGE_Erwini_vB_EhrS_59_NC_048198(17) PHAGE_Escher_500465_1_NC_049342(8)	52,59 50,08	51,34	4898830	83,4	1,702
EKM22T	Chromosome	29 47,2	incomplete intact	31 61	PHAGE_Escher_500465_1_NC_049342(8) PHAGE_Erwini_vB_EhrS_59_NC_048198(16)	49,00 49,26	49,13	4903849	76,2	1,554
FDAARGOS1447	Chromosome	23,5	incomplete	29	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(12)	49,33	49,33	4692018	23,5	0,501
GD03966	Chromosome	41,1 45 26,2 37,6	intact intact incomplete incomplete	54 78 28 19	PHAGE_Erwini_ENT90_NC_019932(25) PHAGE_Edward_GF_2_NC_026611(16) PHAGE_Escher_500465_1_NC_049342(8) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	51,92 52,02 49,55 51,54	51,26	4853300	149,9	3,089
GR13	Chromosome	18,5 19,6 21,9	questionable incomplete questionable	20 33 15	PHAGE_Salmon_vB_SosS_Oslo_NC_018279(9) PHAGE_Erwini_vB_EhrS_59_NC_048198(7) PHAGE_Entero_SfV_NC_003444(5)	52,56 49,61 49,53	50,57	4755241	60	1,262
IG1	Chromosome	22,2 20,7	questionable questionable	16 26	PHAGE_Entero_PsP3_NC_005340(4) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	48,58 47,73	48,16	4836682	42,9	0,887
JM1	Chromosome	41,7 41,5	intact intact	56 48	PHAGE_Salmon_118970_sal3_NC_031940(26) PHAGE_Erwini_ENT90_NC_019932(23)	50,25 51,87	51,06	4816214	83,2	1,727

JZB_2120015	Chromosome	30,5 31,6	intact intact	42 35	PHAGE_Cronob_ESSI_2_NC_047854(24) PHAGE_Erwini_ENT90_NC_019932(21)	52,52 53,23	52,88	4910805	62,1	1,265
K1	Chromosome	45,9 22,7	intact incomplete	79 26	PHAGE_Erwini_vB_EhrS_49_NC_048197(13) PHAGE_Escher_500465_1_NC_049342(8)	51,87 51,06	51,47	5012459	68,6	1,369
KM1	Chromosome	45,9 22,7	intact incomplete	79 26	PHAGE_Erwini_vB_EhrS_49_NC_048197(13) PHAGE_Escher_500465_1_NC_049342(8)	51,87 51,06	51,47	5009979	68,6	1,369
L15	Chromosome LPP-1 MPP-10	21,8 56,5 38,4 20	questionable intact intact incomplete	15 81 46 16	PHAGE_Entero_PsP3_NC_005340(4) PHAGE_Salmon_SEN34_NC_028699(23) PHAGE_Entero_mEp237_NC_019704(3) PHAGE_Salmon_vB_SosS_Oslo_NC_018279(2)	48,90 53,02 49,55 50,70	50,54	4858869	136,7	2,813
MP2	Chromosome	55	intact	77	PHAGE_Salmon_SEN34_NC_028699(21)	51,78	51,78	4735277	55	1,161
MR10	Chromosome	12,5	incomplete	20	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(6)	50,94	50,94	5024161	12,5	0,249
NBBC-01	Chromosome	46,5	intact	54	PHAGE_Escher_vB_EcoM_ECO1230_10_NC_027995(7)	50,18	50,18	4927193	46,5	0,944
NC	Chromosome	18,4 20,1	intact incomplete	30 25	PHAGE_Salmon_RE_2010_NC_019488(10) PHAGE_Shigel_SfIV_NC_022749(2)	50,64 46,82	48,73	4526756	38,5	0,850
NCTC10500	Chromosome	12,2 43,7	questionable intact	16 58	PHAGE_Escher_500465_1_NC_049342(4) PHAGE_Salmon_118970_sal3_NC_031940(27)	45,01 52,32	48,67	4655521	55,9	1,201

NCTC10601	Chromosome	28,6	intact	36	PHAGE_Salmon_RE_2010_NC_019488(10)	46,50	46,50	4752160	28,6	0,602
OVA06A	Chromosome	41,1	intact	46	PHAGE_Erwini_ENT90_NC_019932(23)	51,33	51,64	4873762	83,2	1,707
		42,1	intact	74	PHAGE_Salmon_118970_sal3_NC_031940(20)	51,94				
OVA07A	Chromosome	62,6	intact	67	PHAGE_Entero_BP_4795_NC_004813(10)	52,12	52,12	4801905	62,6	1,304
P10c	Chromosome	46,7	intact	53	PHAGE_Salmon_Fels_2_NC_010463(20)	55,28	55,28	4791557	46,7	0,975
PA	Chromosome	16,4	intact	29	PHAGE_Escher_500465_1_NC_049342(6)	47,12	48,46	4792688	71,2	1,486
		12,7	intact	18	PHAGE_Erwini_ENT90_NC_019932(4)	48,66				
		42,1	intact	52	PHAGE_Escher_vB_EcoM_ECO1230_10_NC_027995(7)	49,60				
Pa17-1	Chromosome	37,7	intact	38	PHAGE_Klebsi_ST437_OXA245phi4.2_NC_049449(13)	54,33	54,33	4832886	37,7	0,780
Pa17-5	Chromosome	38	intact	50	PHAGE_Erwini_ENT90_NC_019932(24)	53,24	51,32	4957367	97,7	1,971
		8,5	questionable	15	PHAGE_Erwini_vB_EhrS_59_NC_048198(3)	47,23				
		51,2	intact	62	PHAGE_Salmon_118970_sal3_NC_031940(18)	53,50				
Pa21-3	Chromosome	24,1	questionable	29	PHAGE_Erwini_vB_EhrS_59_NC_048198(9)	51,89	51,89	5080599	24,1	0,474
Pa21-5	Chromosome	41,6	intact	54	PHAGE_Escher_vB_EcoM_ECO1230_10_NC_027995(7)	49,37	50,12	4847038	74,8	1,543
		33,2	incomplete	27	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	50,86				

Pa21-13	Chromosome	24,1	incomplete	24	PHAGE_Phage_Gifsy_2_NC_010393(3)	49,09	49,85	4793834	39,2	0,818
	LPP-1	15,1	incomplete	26	PHAGE_Erwini_ENT90_NC_019932(2)	50,61				
Pa21-15	Chromosome	33,2	incomplete	25	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	50,86	50,12	4789821	74,8	1,562
		41,6	intact	52	PHAGE_Escher_vB_EcoM_ECO1230_10_NC_027995(8)	49,37				
Pa31-3	Chromosome	28,4	intact	29	PHAGE_Erwini_ENT90_NC_019932(12)	49,99	50,31	5110483	177,4	3,471
		17,8	questionable	26	PHAGE_Salmon_SEN34_NC_028699(15)	50,51				
		49,6	intact	62	PHAGE_Salmon_118970_sal3_NC_031940(26)	51,60				
		32,7	intact	43	PHAGE_Cronob_ESSI_2_NC_047854(24)	54,24				
		21,2	incomplete	21	PHAGE_Entero_P4_NC_001609(2)	44,52				
	LPP-1	27,7	incomplete	27	PHAGE_Bacill_SPbeta_NC_001884(3)	51,02				
Pa31-4	Chromosome	14,2	incomplete	22	PHAGE_Salmon_SW9_NC_049459(3)	48,24	51,74	4992970	143,1	2,866
		49,6	intact	61	PHAGE_Salmon_118970_sal3_NC_031940(26)	51,57				
		47,5	intact	54	PHAGE_Salmon_Fels_2_NC_010463(20)	56,03				
		31,8	incomplete	23	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	51,13				
Pa39-7	Chromosome	15	intact	22	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(9)	46,25	46,25	4969232	15	0,302

Pa39-21	Chromosome	21,6 11,2 53,9	intact questionable intact	38 19 58	PHAGE_Cronob_phiES15_NC_018454(5) PHAGE_Erwini_ENT90_NC_019932(5) PHAGE_Salmon_118970_sal3_NC_031940(23)	48,79 50,07 51,51	50,12	5040326	86,7	1,720
Pa39-23	Chromosome	11,2 41,2 21,6	questionable intact intact	19 68 38	PHAGE_Erwini_ENT90_NC_019932(5) PHAGE_Salmon_118970_sal3_NC_031940(26) PHAGE_Cronob_phiES15_NC_018454(5)	50,06 52,75 48,79	50,53	5103566	74	1,450
Pa39-27	Chromosome LPP-1	53,7 23,9 10,5	intact intact intact	66 28 14	PHAGE_Salmon_118970_sal3_NC_031940(26) PHAGE_Erwini_vB_EhrS_59_NC_048198(21) PHAGE_Entero_I2_2_NC_001332(8)	52,78 51,21 46,80	50,26	4999069	88,1	1,762
Pa58	Chromosome	36,7 20,2 44,1	intact incomplete intact	47 22 51	PHAGE_Salmon_Fels_2_NC_010463(20) PHAGE_Escher_500465_1_NC_049342(9) PHAGE_Entero_Tyrion_NC_031077(24)	54,75 48,20 52,65	51,87	4861200	101	2,078
PaVv1	Chromosome	37 11,2 20 16	intact incomplete questionable incomplete	48 32 15 21	PHAGE_Erwini_ENT90_NC_019932(25) PHAGE_Erwini_vB_EhrS_59_NC_048198(5) PHAGE_Erwini_vB_EhrS_59_NC_048198(3) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(7)	53,41 48,13 49,33 48,99	49,97	4855725	84,2	1,734
PSV1-7	LPP-1	28,9	incomplete	23	PHAGE_Salmon_118970_sal3_NC_031940(2)	48,42	48,42	4914829	28,9	0,588

4188	Chromosome	45,4	intact	69	PHAGE_Erwini_ENT90_NC_019932(29)	53,33	51,25	5041798	200,2	3,971
		30,4	incomplete	30	PHAGE_Escher_500465_1_NC_049342(8)	48,99				
		40,4	intact	55	PHAGE_Erwini_ENT90_NC_019932(25)	52,31				
		54,5	intact	64	PHAGE_Salmon_118970_sal3_NC_031940(10)	51,43				
	LPP-9	29,5	intact	36	PHAGE_Escher_RCS47_NC_042128(6)	50,19				
824-1	Chromosome	35,5	intact	48	PHAGE_Salmon_Fels_2_NC_010463(21)	55,43	53,30	5034205	85	1,688
		39,1	questionable	20	PHAGE_Erwini_vB_EhrS_59_NC_048198(4)	53,18				
		10,4	intact	16	PHAGE_Erwini_ENT90_NC_019932(5)	51,29				
R5	Chromosome	38,8	incomplete	20	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	51,86	53,18	4825137	168,2	3,486
		39,7	intact	39	PHAGE_Salmon_Fels_2_NC_010463(16)	55,18				
		35,6	intact	48	PHAGE_Erwini_ENT90_NC_019932(24)	53,64				
		54,1	intact	79	PHAGE_Salmon_SEN34_NC_028699(21)	52,04				
RIT710	Chromosome	12,8	incomplete	22	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	51,26	51,53	4768579	114,4	2,399
		35,9	intact	47	PHAGE_Salmon_Fels_2_NC_010463(18)	54,10				
		18,8	incomplete	14	PHAGE_Erwini_vB_EhrS_59_NC_048198(2)	49,11				
		46,9	intact	67	PHAGE_Salmon_118970_sal3_NC_031940(27)	51,63				

RIT-PIT-AA	Chromosome	19,3	incomplete	21	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(9)	47,85	50,00	4840418	73,4	1,516
		14,2	incomplete	24	PHAGE_Salmon_SW9_NC_049459(3)	48,24				
		39,9	intact	52	PHAGE_Salmon_Fels_2_NC_010463(25)	53,92				
RIT-PI-T	Chromosome	40,8	intact	61	PHAGE_Salmon_118970_sal3_NC_031940(25)	52,28	50,22	4993318	105,7	2,117
		17,8	questionable	26	PHAGE_Salmon_SEN34_NC_028699(15)	50,48				
		47,1	intact	52	PHAGE_Erwini_vB_EhrS_59_NC_048198(10)	47,89				
RSO7	Chromosome	17	incomplete	23	PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	47,71	49,92	4823630	39,6	0,821
		22,6	questionable	24	PHAGE_Escher_500465_1_NC_049342(8)	52,13				
SL1_M5	Chromosome	38,1	intact	53	PHAGE_Erwini_ENT90_NC_019932(24)	53,99	53,85	4952634	236	4,765
		41,4	intact	65	PHAGE_Salmon_118970_sal3_NC_031940(23)	52,08				
		56,4	intact	62	PHAGE_Entero_phi80_NC_021190(9)	52,53				
		45,8	intact	48	PHAGE_Salmon_Fels_2_NC_010463(21)	55,80				
		54,3	intact	67	PHAGE_Erwini_ENT90_NC_019932(31)	54,87				
Strain 3	Chromosome	15,4	questionable	15	PHAGE_Salmon_Fels_2_NC_010463(7)	58,23	54,39	4811315	67,3	1,399
		29,5	incomplete	25	PHAGE_Escher_500465_1_NC_049342(8)	53,03				
		22,4	intact	32	PHAGE_Klebsi_ST437_OXA245phi4.2_NC_049449(15)	51,92				
4188	Chromosome	17,7	questionable	28	PHAGE_Salmon_SW9_NC_049459(6)	49,26	49,26	4833670	17,7	0,366

Strain 190	Chromosome	41,6 21,6 11,2 18	intact intact questionable incomplete	65 38 19 25	PHAGE_Salmon_118970_sal3_NC_031940(28) PHAGE_Cronob_phiES15_NC_018454(5) PHAGE_Erwini_ENT90_NC_019932(5) PHAGE_Klebsi_ST147_VIM1phi7.1_NC_049451(7)	52,23 48,79 50,06 51,71	50,70	5000005	92,4	1,848
TH81	Chromosome LPP-12	27,7 42,5	questionable questionable	15 37	PHAGE_Erwini_vB_EhrS_59_NC_048198(3) PHAGE_Escher_RCS47_NC_042128(2)	52,00 48,31	50,16	4984468	70,2	1,408
UAEU18	Chromosome	40,2 43,8	intact incomplete	50 22	PHAGE_Salmon_118970_sal3_NC_031940(6) PHAGE_Escher_500465_1_NC_049342(8)	48,75 51,08	49,92	4825350	84	1,741
ZBG6	Chromosome	19,1 23,7	intact incomplete	32 22	PHAGE_Erwini_ENT90_NC_019932(7) PHAGE_Klebsi_ST13_OXA48phi12.1_NC_049453(8)	48,38 54,23	51,31	4768948	42,8	0,897
ZJU23	Chromosome	41,1 50,1 26,9	intact intact incomplete	46 64 30	PHAGE_Escher_vB_EcoM_ECO1230_10_NC_027995(7) PHAGE_Salmon_118970_sal3_NC_031940(19) PHAGE_Escher_500465_1_NC_049342(8)	48,68 53,31 49,47	51,12	5061267	150,6	2,976

		32,5	intact	47	PHAGE_Salmon_SEN8_NC_047753(16)	53,02				
P. eucalypti LMG24197^T	Chromosome	11,2	questionable	17	PHAGE_Salmon_SW9_NC_049459(6)	51,83	51,50	4798990	80,4	1,675
		26,1	questionable	33	PHAGE_Entero_phiP27_NC_003356(4)	46,65				
	pEUC-1	18,7	incomplete	23	PHAGE_Entero_P1_NC_005856(11)	52,56				
		24,4	intact	21	PHAGE_Salmon_SJ46_NC_031129(22)	54,97				
P. vagans LMG24199^T	Chromosome	20,5	questionable	30	PHAGE_Salmon_SEN34_NC_028699(12)	50,00	50,52	4790329	73,7	1,539
		38,7	intact	51	PHAGE_Entero_mEp460_NC_019716(7)	50,57				
		14,5	intact	19	PHAGE_Entero_P4_NC_001609(10)	50,99				

Table 8: ICE regions found in the genomes of *P. agglomerans*

Strain	Length (bp)	No. of Proteins
20TX0122	40373	44
AR1a	79716	93
B1	50770	48
BAV2934	95903	107
BI3	52427	50
	48010	49
C410P1	72654	77
CB1	84619	88
CFBP8756	107599	114
CFBP13600	56407	69
CFBP13731	108075	114
CFSAN033090	44195	67
CFSAN047153	56186	55
	80841	92
CFSAN047154	56186	55
	56855	57
CHTF15	97370	115
DAPP-PG734	235647	240
	139133	142
	95735	105
E325	45450	45
E325-699	46180	42
E325-750	46136	43
E325-754	46180	42
EKM20T	77603	74
EKM21T	87384	85
K1	210286	217
KM1	175899	187
MR10	92059	96
	113290	111
MR19	92059	97
	113284	111
NBBC-01	49884	55
	59391	68
NCTC10500	90058	95
OVA06A	43844	47
Pa17-5	74390	86
Pa21-3	100880	116
	60347	61
Pa21-5	73441	82
Pa21-15	60896	67
Pa31-3	105447	111
	107867	114
Pa39-7	169825	183
	71459	81
Pa39-21	98601	105
	64533	70
Pa39-23	98601	105
	49331	53
Pa39-27	70497	79
RIT-PIT-AA	49803	55
Strain 4	78034	70
UAEU18	80937	91
ZJU23	105879	115

Table 9: Distinct ICEs found on *P. agglomerans*

ICE	Frequency of occurrence
Phage_integrase	109
Protein UmuC	100
Protein UmuD	58
t4cp2	49
F_traF	42
Replication initiation protein	38
T_virB1	35
F_traH	31
Protein TraC, virb4	31
F_traB	30
F_traE	30
F_traG	30
F_traK	30
F_traL	30
F_traU	30
F_traW	30
F_trbC	30
Mating pair stabilization protein TraN, F_traN	30
G_tfc7	24
DNA topoisomerase 3	21
G_tfc10	21
G_tfc11	21
G_tfc12	21
G_tfc13	21
G_tfc14	21
G_tfc15	21
G_tfc18	21
G_tfc19	21
G_tfc2	21
G_tfc22	21
G_tfc23	21
G_tfc24	21
G_tfc3	21
G_tfc5	21
G_tfc8	21
G_tfc9	21
MOBH	21
virb4	21
G_tfc17	20
Protein TraV, F_traV	20
Lipoprotein YlpA	19
MOBF	19
Replicative DNA helicase	19
Transcription antitermination protein RfaH	19
Tyrosine recombinase XerD, Phage_integrase	19
DNA primase TraC	18
Phospholipase D	14
2%2C3-bisphosphoglycerate-dependent phosphoglycerate mutase	13
Copper resistance protein D	13
Enolase	13
Putative fluoride ion transporter CrcB	13
Sensor histidine kinase CusS	13
Transcriptional activator protein CopR; pcoR	13
Arsenate reductase; arsC	12

Copper resistance protein A; pcoA	12
Copper resistance protein B	12
NADPH-dependent FMN reductase ArsH; arsH	12
Putative transposon Tn552 DNA-invertase bin3	12
Copper resistance protein C	11
Coupling protein TraD, t4cp2	11
ISNCY family transposase ISSen7; ISSen7	11
Multifunctional conjugation protein TraI, MOBF	11
Sporulation initiation inhibitor protein Soj	11
Arsenical pump membrane protein; arsB	10
DNA adenine methyltransferase YhdJ	10
F_traV	10
PTS system galactitol-specific EIIC component	10
Antirestriction protein KlcA	9
hypothetical protein; RM_Type_III,Type_III_MTases	9
putative manganese-dependent inorganic pyrophosphatase	9
Hemolysin expression-modulating protein Hha	8
hypothetical protein; PD-T7-5,PD-T7-5	8
putative signaling protein	8
Single-stranded DNA-binding protein	8
Arsenic resistance transcriptional regulator ArsR2	7
DNA polymerase III subunit theta	7
HTH-type transcriptional regulator ZntR	7
putative cysteine-rich protein YhjQ	7
Toxin CcdB	7
3-oxoacyl-[acyl-carrier-protein] reductase FabG	6
Arabinose 5-phosphate isomerase GutQ	6
Arsenical pump membrane protein	6
Glucitol operon repressor	6
IS30 family transposase ISCfr4, rve; ISEch7	6
MOBP1	6
N-ethylmaleimide reductase	6
Petrobactin import ATP-binding protein FpuD	6
PTS system glucitol/sorbitol-specific EIIA component	6
PTS system glucitol/sorbitol-specific EIIB component	6
PTS system glucitol/sorbitol-specific EIIC component	6
RepFIB replication protein A	6
Sorbitol-6-phosphate 2-dehydrogenase	6
SOS response-associated protein YedK	6
Trehalose-6-phosphate synthase	6
Tyrosine recombinase XerC, Phage_integrase	6
Cation efflux system protein CusA; silA	5
Cation efflux system protein CusB; silB	5
Cation efflux system protein CusC; ibeB; silC	5
Cation efflux system protein CusF; silF	5
DNA-cytosine methyltransferase; RM_Type_II,Type_II_MTases	5
hypothetical protein; Gabija,GajA	5
Macrolide export ATP-binding/permease protein MacB	5
Macrolide export protein MacA	5
OriC-binding nucleoid-associated protein	5
putative HTH-type transcriptional regulator YybR	5
Sensor histidine kinase CusS; silS	5
Silver-binding protein SilE	5
Transcriptional regulatory protein CusR; cusR/ylcA	5
ATP-dependent DNA helicase PcrA; Gabija,GajB_2	4
Cystathionine beta-lyase PatB	4
Glutathione reductase	4

HTH-type transcriptional regulator DmlR	4
hypothetical protein; Lamassu-Fam,LmuA_effector_Cap4_nuclease_II	4
hypothetical protein; Lamassu-Fam,LmuB_SMC_Cap4_nuclease_II	4
hypothetical protein; Lamassu-Fam,LmuC_acc_Cap4_nuclease	4
hypothetical protein; RosmerTA,RmrT_2634932349	4
hypothetical protein; TnpR_TnEc1	4
IS30 family transposase ISCfr4	4
IS5 family transposase ISPs1	4
Leucine efflux protein	4
Mokosh_TypeI,MkoB2	4
Phosphinothricin N-acetyltransferase	4
Protein PsiB	4
PTS system beta-glucoside-specific EIIBCA component	4
Replication protein RepA	4
Tn3 family transposase TnEc1	4
Type-2 restriction enzyme EcoRII; RM_Type_II,Type_II_REases	4
Xylulose kinase	4
Zinc-type alcohol dehydrogenase-like protein	4
Acetyltransferase	3
Chromosome-partitioning ATPase Soj	3
Colicin V secretion protein CvaA	3
Diguanylate cyclase DgcM	3
Endoribonuclease HigB	3
Ferric aerobactin receptor	3
Flavin reductase	3
HTH-type transcriptional regulator GltC	3
HTH-type transcriptional repressor GlaR	3
HTH-type transcriptional repressor NagR	3
hypothetical protein; AbiD,AbiD	3
I_traI	3
I_traK	3
I_traL	3
I_traM	3
I_traN	3
I_traO	3
I_traP	3
I_traQ	3
I_traR	3
I_traT	3
I_traU	3
I_traW	3
I_traY	3
I_trbA	3
I_trbB	3
IS110 family transposase ISSm1; ISSm1	3
IS3 family transposase IS1222	3
ISNCY family transposase ISRor2; ISRor2	3
O-acetyltransferase OatA	3
Outer membrane usher protein FimD	3
p-hydroxyphenylacetate 3-hydroxylase%2C oxygenase component	3
Plasmid partition protein A	3
Protein TraI, MOBP1	3
putative fimbrial chaperone SfmC	3
putative MFS-type transporter EfpA	3
putative siderophore transport system permease protein YfhA	3
putative siderophore transport system permease protein YfiZ	3
putative siderophore-binding lipoprotein YfiY	3

t4cp1	3
Thiol:disulfide interchange protein DsbG	3
Toxin coregulated pilus biosynthesis protein E	3
Transcriptional repressor PifC	3
Virulence regulon transcriptional activator VirB	3
1-deoxyxylulose-5-phosphate synthase YajO	2
3-hexulose-6-phosphate isomerase	2
5'-nucleotidase SurE	2
6-phospho-beta-glucosidase BglA	2
AIDA-I autotransporter	2
Aldo-keto reductase IolS	2
Alginate lyase	2
Aminoglycoside 3'-phosphotransferase; aph(3")-Ib	2
Anaerobic C4-dicarboxylate transporter DcuA	2
Aryl-phospho-beta-D-glucosidase BglC	2
Asparagine synthetase [glutamine-hydrolyzing] 1	2
Aspartate ammonia-lyase	2
Beta-lactamase hydrolase-like protein	2
Bicarbonate transporter BicA	2
Bicyclomycin resistance protein	2
Chaperone protein caf1M	2
Chromosome partitioning protein ParA	2
Copper resistance protein C; copC	2
Cryptic beta-glucoside bgl operon antiterminator	2
Cysteine/O-acetylserine efflux protein	2
Dapdiamide A synthase	2
Dapdiamide synthesis protein DdaD	2
dCTP deaminase%2C dUMP-forming	2
Divalent-cation tolerance protein CutA	2
DNA-invertase hin; TnpR_ISMex22	2
D-tagatose-1%2C6-bisphosphate aldolase subunit KbaY	2
F1 capsule-anchoring protein	2
Fluoroacetate dehalogenase	2
Fumarate--(S)-2%2C3-diaminopropanoate ligase	2
Galactitol 1-phosphate 5-dehydrogenase	2
Glyoxal reductase	2
GTP cyclohydrolase-2	2
HTH-type transcriptional regulator GlvR	2
HTH-type transcriptional regulator PgrR	2
hypothetical protein; aph(6)-Id	2
hypothetical protein; RloC,RloC	2
hypothetical protein; Shedu,SduA	2
I_traE	2
IS110 family transposase IS1618; IS1618	2
IS110 family transposase IS5708; IS5708	2
IS3 family transposase IS1133	2
IS3 family transposase IS1133, rve; IS1133	2
IS3 family transposase IS1400	2
IS3 family transposase IS1400, rve	2
IS3 family transposase ISEal1	2
IS3 family transposase ISEal1, rve; ISPeat2	2
IS3 family transposase ISEhe3, rve	2
IS3 family transposase ISKpn34, rve	2
IS481 family transposase ISErsp1, rve; ISErsp1	2
IS630 family transposase ISSpu2; ISSpu2	2
IS66 family transposase ISCro1	2
IS66 family transposase ISCro1; ISCro1	2

IS66 family transposase ISSgsp1; ISSgsp1	2
ISNCY family transposase ISBcen27	2
ISNCY family transposase ISRor2	2
Isochorismate pyruvate lyase	2
Lactococcin-G-processing and transport ATP-binding protein LagD	2
Linear gramicidin dehydrogenase LgrE	2
Multidrug efflux pump subunit AcrB	2
N-(2-amino-2-carboxyethyl)-L-glutamate synthase	2
NADH oxidase	2
Nitroreductase NfnB	2
Outer membrane lipoprotein BfpB	2
Phosphocarrier protein HPr	2
Phosphorylated carbohydrates phosphatase	2
Protein TraR	2
putative fimbrial-like protein YcbV	2
putative membrane transporter protein	2
putative protein YcjY	2
Ribose import ATP-binding protein RbsA	2
Ribose import binding protein RbsB	2
Ribose import permease protein RbsC	2
Ribulose-phosphate 3-epimerase	2
S-fimbrial protein subunit SfaA	2
Sorbitol dehydrogenase	2
Sorbitol operon regulator	2
Thiol:disulfide interchange protein DsbD	2
Tn3 family transposase ISEc63	2
Tn3 family transposase ISEc63; ISEc63	2
Tn3 family transposase ISYps3; TnpR_TnSba14	2
Tn3 family transposase TnEc1; TnpA_TnEc1	2
Tn3 family transposase; ISRsp12	2
Toxin and drug export protein A	2
Trans-2%2C3-dihydro-3-hydroxyanthranilate isomerase	2
Transcriptional activator protein CzcR	2
Transposon gamma-delta resolvase	2
tRNA(fMet)-specific endonuclease VapC	2
tRNA-Phe(gaa)	2
Ureidoglycolate lyase	2
Zinc-transporting ATPase	2
(R)-stereoselective amidase	1
(S)-2-haloacid dehalogenase	1
2-epi-5-epi-valiolone epimerase	1
2-hydroxy-3-oxopropionate reductase	1
3-oxoacyl-[acyl-carrier-protein] synthase 2	1
Acetoacetyl-CoA reductase	1
Acetophenone carboxylase delta subunit	1
Acetophenone carboxylase gamma subunit	1
Acetylornithine aminotransferase	1
Actin cross-linking toxin VgrG1; vgrG1	1
Aldose 1-epimerase	1
Arsenic resistance transcriptional regulator ArsR1; arsR	1
Aspartate racemase	1
Carboxypeptidase G2	1
Catechol O-methyltransferase	1
Cation efflux system protein CusA	1
cGAMP-activated phospholipase; CBASS,Phospholipase	1
Chaperone protein FimC	1
Chloramphenicol 3-O phosphotransferase	1

Cyclic GMP-AMP synthase; CBASS,Cyclase_II	1
D-altritol 5-dehydrogenase	1
DNA-binding protein H-NS	1
EAMY_RS19965	1
EAMY_RS19975	1
F_traK; hrcC	1
Gao_Tmn,TmnA	1
Glucose 1-dehydrogenase	1
Harpin HrpN; hrpN	1
Hca operon transcriptional activator HcaR	1
Hexuronate transporter	1
Homoserine/homoserine lactone efflux protein	1
hrcJ	1
HTH-type transcriptional regulator LrpC	1
HTH-type transcriptional regulator NimR	1
HTH-type transcriptional regulator SrpS	1
HTH-type transcriptional regulator YofA	1
hypothetical protein; AbiH,AbiH	1
hypothetical protein; Borvo,BovA	1
hypothetical protein; CBASS,AG_E1_ThiF	1
hypothetical protein; CBASS,Jab	1
hypothetical protein; EAMY_RS20030	1
hypothetical protein; Gabija,GajB_2	1
hypothetical protein; Gao_Ppl,PplA	1
hypothetical protein; Gao_Qat,QatA	1
hypothetical protein; Gao_Qat,QatB	1
hypothetical protein; Gao_Qat,QatC	1
hypothetical protein; Gao_Tmn,TmnA	1
hypothetical protein; NLR_like_bNACHT01,NLR_like_bNACHT01	1
hypothetical protein; pbrA	1
hypothetical protein; RM_Type_III,Type_III_REases	1
hypothetical protein; wtsE	1
IS1 family transposase IS1N	1
IS110 family transposase IS884; IS884	1
IS110 family transposase ISCfr13; ISCfr13	1
IS110 family transposase ISEsa2; ISEsa2	1
IS110 family transposase ISGme8	1
IS110 family transposase ISPlu13; ISPlu13	1
IS3 family transposase ISBcen15	1
IS3 family transposase ISEhe3, rve; ISEhe3	1
IS3 family transposase ISEhe4	1
IS3 family transposase ISEhe4, rve; ISEhe4	1
IS3 family transposase ISKpn34, rve; ISKpn34	1
IS3 family transposase ISRso16, rve	1
IS3 family transposase ISSlo2, rve	1
IS3 family transposase ISSpr1	1
IS3 family transposase ISSpr1, rve	1
IS30 family transposase ISCfr4, rve	1
IS5 family transposase IS52; IS52	1
IS5 family transposase IS903; IS903	1
IS630 family transposase IS630	1
ISNCY family transposase ISSen7	1
Levansucrase	1
L-methionine sulfoximine/L-methionine sulfone acetyltransferase	1
Long-chain-fatty-acid--CoA ligase FadD15	1
L-threonine dehydratase catabolic TdcB	1
Major fimbrial subunit	1

Metal-pseudopaline receptor CntO	1
Methyl viologen resistance protein SmvA	1
Methyl-accepting chemotaxis protein II	1
mRNA interferase toxin RelE	1
Multidrug export protein AcrE	1
Multidrug resistance ABC transporter ATP-binding/permease protein BmrA	1
Multidrug resistance protein 3	1
Multifunctional conjugation protein TraI	1
Na(+)%2C Li(+)%2C K(+)/H(+) antiporter	1
Organic hydroperoxide resistance transcriptional regulator	1
Outer membrane usher protein HtrE	1
pbrA	1
Phenazine antibiotic resistance protein EhpR	1
Phosphate regulon transcriptional regulatory protein PhoB	1
Pilin	1
Plasmid segregation protein ParM	1
Proline/betaine transporter	1
Protein AfaD	1
Protein VirD4, t4cp2	1
putative deoxyribonuclease RhsA; tke2	1
putative isomerase YddE	1
putative major fimbrial subunit LpfA	1
putative metal-dependent hydrolase YjjV; Gao_Qat,QatD	1
putative protein	1
putative protein YycE	1
Pyruvate dehydrogenase complex repressor	1
Recombination-promoting nuclease RpnA; ISRor2	1
Relaxosome protein TraM	1
Rhamnolipids biosynthesis 3-oxoacyl-[acyl-carrier-protein] reductase	1
Riboflavin transport system permease protein RibX	1
Secretory immunoglobulin A-binding protein EsiB	1
Septu,PtuA_2	1
Serine--tRNA ligase	1
S-sulfocysteine synthase	1
T_virB10	1
T_virB2	1
T_virB3	1
T_virB6	1
Taurine import ATP-binding protein TauB	1
Threonine efflux protein	1
Toxin RTX-I translocation ATP-binding protein	1
Transcriptional regulator AcuR	1
Transcriptional repressor PifC; Pif,PifC	1
tRNA-Ser(cga)	1
Type IV secretion system protein PtlE, T_virB8	1
Type IV secretion system protein PtlF, T_virB9	1
Type IV secretion system protein VirB11, T_virB11	1
Type IV secretion system protein virB4, virb4	1
Virginiamycin A acetyltransferase	1

Table 10: Number of transposable elements found in the genomes of *P. agglomerans* and the two outgroup taxa

Strain	No. of TEs
strain 18059	13
20TX0122	33
4188	84
62d	14
62e	14
824-1	49
9Rz4	24
A2	30
AB378	23
AR1a	34
ASB05	27
B1	66
B3	66
BAV_2934	22
BI3	26
C1	26
C410P1	46
CB1	26
CCNPK873	14
CFBP13505	59
CFBP13569	14
CFBP13583	10
CFBP13600	26
CFBP13616	13
CFBP13731	58
CFBP8756	55
CFBP8784	23
CFBP8786	55
CFSAN033090	20
CFSAN047153	28
CFSAN047154	28
CHTF15	27
DAPP-PG734	58
DBM_3797	27
DOAB1048	23
E325.	20
E325-699	20
E325-705	18
E325-750	20
E325-754	21
E325-ad1	19
E325-ad2	20
Eh318	44
EKM10T	19
EKM20T	21
EKM21T	17
EKM22T	17
FDAARGOS 1447T	19
GD03966	17
GR13	14
IG1	9
JM1	26
JZB_2120015	22

K1	31
KM1	32
L15	18
MP2	23
MR10	43
MR19	42
NBBC-01	27
NC	24
NCTC10500	31
NCTC10601	41
OVA06A	24
OVA07A	19
P10c	19
PA	15
Pa17-1	19
Pa17-5	17
Pa21-13	21
Pa21-15	29
Pa21-3	32
Pa21-5	28
Pa31-3	37
Pa31-4	19
Pa39-1	25
Pa39-21	27
Pa39-23	28
Pa39-27	21
Pa39-7	32
Pa58	22
PaVv1	25
PMG_056	18
PSV1-7	19
R1	20
R5	22
RIT710	18
RIT-PIT-AA	16
RIT-PI-T	24
RSO7	20
SL1_M5	16
strain 190	23
strain 3	16
strain 4	14
T1	13
TH81	41
Tx10	15
UAEU18	28
ZBG6	29
ZJU23	17
P. eucalypti LMG_24197^{T}	21
P. vagans LMG_24199^{T}	8

Appendix E

Table 11: COG supra-functional category proportions in *P. agglomerans*

COG Supra-functional categories	Proportion of pan-genome fraction (%)		
Metabolism	Core	Shared Accessory	Singleton
C = energy production and conversion	7.579	4.230	1.314
P = inorganic ion transport and metabolism	10.699	5.281	4.161
Q = secondary metabolites biosynthesis, transport and metabolism	1.798	2.616	2.168
E = amino acid transport and metabolism	10.235	5.990	6.921
F = nucleotide transport and metabolism	3.693	1.858	2.957
G = carbohydrate transport and metabolism	10.054	6.919	4.424
H = coenzyme transport and metabolism	6.664	2.384	3.745
I = lipid transport and metabolism	3.422	2.641	2.694
Information Storage and Processing			
J = translation, ribosomal structure and biogenesis	6.329	3.130	2.168
K = transcription	11.665	16.817	17.302
L = replication, recombination and repair	4.576	14.397	23.806
Cellular Processes			
M = cell wall/ membrane/ envelop biogenesis	8.018	9.169	4.731
N = cell motility	2.217	4.295	1.971
O = posttranslational modification, protein turnover, chaperones	3.654	2.877	2.365
T = signal transduction mechanisms	4.105	4.837	4.095
U = intracellular trafficking, secretion and vesicular transport	2.101	5.273	4.139
V = defense mechanism	1.353	4.401	7.819
D = cell cycle control, cell division, chromosome partitioning	1.837	2.885	3.219

Table 12: Individual COG category distributions in *P. agglomerans*

	CORE		SHARED ACCESSORY		SINGLETON	
Category	Frequency	%	Frequency	%	Frequency	%
C	196	7,579	86,5	4,230	10	1,314
D	47,5	1,837	59	2,885	24,5	3,219
E	264,667	10,235	122,5	5,990	52,667	6,921
F	95,5	3,693	38	1,858	22,5	2,957
G	260	10,054	141,5	6,919	33,667	4,424
H	172,333	6,664	48,75	2,384	28,5	3,745
I	88,5	3,422	54	2,641	20,5	2,694
J	163,667	6,329	64	3,130	16,5	2,168
K	301,667	11,665	343,917	16,817	131,667	17,302
L	118,333	4,576	294,417	14,397	181,167	23,806
M	207,333	8,018	187,5	9,169	36	4,731
N	57,333	2,217	87,833	4,295	15	1,971
O	94,5	3,654	58,833	2,877	18	2,365
P	276,667	10,699	108	5,281	31,667	4,161
Q	46,5	1,798	53,5	2,616	16,5	2,168
T	106,167	4,105	98,917	4,837	31,167	4,095
U	54,333	2,101	107,833	5,273	31,5	4,139
V	35	1,353	90	4,401	59,5	7,819
Total	2586	100	2045	100	761	100

Table 13: COG category distribution in the accessory genomes of *P. agglomerans* subsp. 1 and 2.

Subsp. 1	Frequency	Percentage	Subsp. 2	Frequency	Percentage
C	28	2,341	C	22,5	4,564
D	43,5	3,637	D	10	2,028
E	73,667	6,159	E	28,5	5,781
F	34,5	2,885	F	11	2,231
G	65,667	5,491	G	39	7,911
H	43	3,595	H	10	2,028
I	24,5	2,048	I	18	3,651
J	28,5	2,383	J	6	1,217
K	192,667	16,109	K	74,5	15,112
L	245,667	20,541	L	77	15,619
M	82	6,856	M	41,5	8,418
N	58	4,849	N	19,833	4,023
O	22,5	1,881	O	24,333	4,936
P	54,167	4,529	P	22	4,462
Q	25,5	2,132	Q	11	2,231
T	51,667	4,320	T	13,5	2,738
U	47	3,930	U	36,333	7,370
V	75,5	6,313	V	28	5,680
	1196			493	

Appendix F

Table 14: Pathogen-host interactions of the 100 *P. agglomerans* strains and both outgroup taxa.

Strain	PHI hits
NCTC10500	280
3	293
4	283
190	293
4188	299
18059	286
20TX0122	285
62d	286
62e	286
824-1	298
9Rz4	283
A2	288
AB378	294
AR1a	289
ASB05	289
B1	285
B3	285
BAV2934	297
BI3	301
C1	294
C410P1	289
CB1	287
CCNPK873	288
CFBP13505	285
CFBP13569	294
CFBP13583	290
CFBP13600	293
CFBP13616	290
CFBP13731	295
CFBP8756	296
CFBP8784	292
CFBP8786	283
CFSAN033090	290
CFSAN047153	291
CFSAN047154	291
CHTF15	291
DAPP-PG734	296
DBM3797	294
DOAB1048	291
E325	292
E325-699	293
E325-705	292
E325-750	292
E325-754	292
E325-ad1	291
E325-ad2	292
Eh318	292
EKM10T	283
EKM20T	292
EKM21T	289
EKM22T	289
FDAARGOS1447	288

GD03966	285
GR13	286
IG1	292
JM1	293
JZB2120015	287
K1	292
KM1	292
L15	287
MP2	293
MR10	288
MR19	288
NBBC-01	289
NC	283
NCTC10601	288
OVA06A	295
OVA07A	289
P10c	286
PA	288
Pa17-1	286
Pa17-5	290
Pa21-13	289
Pa21-15	292
Pa21-3	292
Pa21-5	293
Pa31-3	293
Pa31-4	291
Pa39-1	293
Pa39-21	299
Pa39-23	301
Pa39-27	290
Pa39-7	288
Pa58	283
PaVv1	282
PMG056	284
PSV1-7	292
R1	292
R5	290
RIT710	283
RIT-PIT-AA	287
RIT-PI-T	292
RS07	290
SI1	284
T1	286
TH81	283
Tx10	294
UAEU18	293
ZBG6	290
ZJU23	290
P. eucalypti LMG_24197^T	285
P. vagans LMG_24199^T	290

Table 15: Distinct pathogen-host interactions in *P. agglomerans* subsp. 1 and 2.

	Total occurrence	Subsp. 1	Subsp. 2
A0A0F6B5T1#PHI:10085#STM14_RS15360#28901#Salmonella_enterica#unaffected_pathogenicity	1	0	1
A0A0F6UGK4#PHI:6941#FeoB#287#Pseudomonas_aeruginosa#increased_virulence_(hypervirulence)_loss_of_pathogenicity	1	0	1
A6TF96#PHI:7953#ArnT#573#Klebsiella_pneumoniae#unaffected_pathogenicity	1	1	0
D4I208#PHI:7079_PHI:7417#TssA-3_EAMY_3201#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	1	1	0
D4I223#PHI:7091_PHI:7432#TssM-3_EAMY_3216#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	1	1	0
E0SF07#PHI:11739#ProQ#204038#Dickeya_dadantii#unaffected_pathogenicity	1	1	0
H9L4I8#PHI:11152#Hcp2#28901#Salmonella_enterica#reduced_virulence_increased_virulence_(hypervirulence)	1	0	1
P69988#PHI:6114#RovA#632#Yersinia_pestis#reduced_virulence	1	0	1
Q0SXL9#PHI:9888#FimA#562#Escherichia_coli#unaffected_pathogenicity	1	1	0
Q6U5X7#PHI:6564#IroB#573#Klebsiella_pneumoniae#unaffected_pathogenicity	1	1	0
Q7WTI6#PHI:4861#IroE#562#Escherichia_coli#unaffected_pathogenicity_reduced_virulence	1	1	0
Q888Y7#PHI:984_PHI:2752_PHI:2879_PHI:8329#HopQ1-1_HopQ1_PtoHopQ1#317#Pseudomonas_syringae#effector_(plant_avirulence_determinant)	1	0	1
Q8G8P0#PHI:4860#IroD#562#Escherichia_coli#reduced_virulence	1	1	0
Q8Z4H1#PHI:9464#IroN#28901#Salmonella_enterica#reduced_virulence	1	1	0
Q8Z4H3#PHI:9461#IroC#28901#Salmonella_enterica#reduced_virulence	1	1	0
Q8ZFR1#PHI:8898#CobB_(Yp_1760)#632#Yersinia_pestis#reduced_virulence	1	0	1
Q9FCY7#PHI:6376#WtsE#66269#Pantoea_stewartii#effector_(plant_avirulence_determinant)	1	1	0
S6TYC4#PHI:10902#HopR1#317#Pseudomonas_syringae#effector_(plant_avirulence_determinant)	1	0	1
A0A0U1FZZ2#PHI:8757#TatA#28901#Salmonella_enterica#reduced_virulence	1	1	0
A7ZRQ3#PHI:10978#SufI#562#Escherichia_coli#reduced_virulence	1	1	0
A0A0H3NBW6#PHI:11382#KatN#28901#Salmonella_enterica#reduced_virulence	2	2	0
A7FN30#PHI:3847#TcpYI#633#Yersinia_pseudotuberculosis#reduced_virulence	2	0	2
B8F4W8#PHI:3317#HsdR#738#Glaesserella_parasuis#reduced_virulence	2	0	2
Q886L1#PHI:6365#HopAF1#317#Pseudomonas_syringae#effector_(plant_avirulence_determinant)	2	1	1
D4I0C5#PHI:3676_PHI:3682#Rmf3#552#Erwinia_amylovora#unaffected_pathogenicity	2	1	1
A0A0Q0B337#PHI:6649#IaaM#29438#Pseudomonas_savastanoi#reduced_virulence	3	1	2
P0AE88#PHI:4523_PHI:8273_PHI:9363#CpxR#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	3	0	3
Q888Y8#PHI:4005#HopD1#317#Pseudomonas_syringae#effector_(plant_avirulence_determinant)	3	1	2

Q9HVF6#PHI:4897#MgtC#287#Pseudomonas_aeruginosa#reduced_virulence	3	0	3
D4HVL8#PHI:6338#HrcU#552#Erwinia_amylovora#loss_of_pathogenicity	4	2	2
D4I233#PHI:7088_PHI:7442#TssK-2_EAMY_3226#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	4	2	2
D4IAW2#PHI:2456_PHI:3672_PHI:3678#HrpL#552#Erwinia_amylovora#loss_of_pathogenicity	4	2	2
Q46640#PHI:2468#HrpJ#552#Erwinia_amylovora#reduced_virulence	4	2	2
Q9FCZ1#PHI:123739#HrcC#66269#Pantoea_stewartii#unaffected_pathogenicity_reduced_virulence	4	2	2
Q9FD00#PHI:2473#Hrpx#66269#Pantoea_stewartii#reduced_virulence	4	2	2
Q9X3T0#PHI:3671_PHI:3677#HrpS#552#Erwinia_amylovora#loss_of_pathogenicity	4	2	2
B5XY48#PHI:4160#OmpA2#573#Klebsiella_pneumoniae#reduced_virulence	5	5	0
A0A6C8TBQ7#PHI:10668#NorV#562#Escherichia_coli#unaffected_pathogenicity_reduced_virulence	6	4	2
A0A0E1M219#PHI:9601#RstA#562#Escherichia_coli#reduced_virulence	7	4	3
P69428#PHI:10981#TatA#562#Escherichia_coli#reduced_virulence	8	8	0
D4I0H1#PHI:9328#CarA#552#Erwinia_amylovora#reduced_virulence	9	0	9
A0A0F6B5R1#PHI:10068#STM14_RS15275#28901#Salmonella_enterica#unaffected_pathogenicity	10	0	10
A0A0F6B5R3#PHI:10070#STM14_RS15285#28901#Salmonella_enterica#unaffected_pathogenicity	10	0	10
Q06951#PHI:703#RfbB#666#Vibrio_cholerae#reduced_virulence	11	8	3
Q46877#PHI:12042#NorV#562#Escherichia_coli#increased_virulence_(hypervirulence)	11	11	0
A0A0F6B5P2#PHI:10052#STM14_RS15195_(srlA)#28901#Salmonella_enterica#unaffected_pathogenicity	12	12	0
D4I220#PHI:7085_PHI:7429#TssG-2_EAMY_3213#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	12	7	5
A0A0F6B5P3#PHI:10053#STM14_RS15200_(srlE)#28901#Salmonella_enterica#unaffected_pathogenicity	13	13	0
A0A0F6B5P5#PHI:10055#STM14_RS15210_(srlD)#28901#Salmonella_enterica#unaffected_pathogenicity	13	13	0
D4I221#PHI:7083_PHI:7430#TssF-2_EAMY_3214#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	13	8	5
Q7CPD5#PHI:8726#TatC#28901#Salmonella_enterica#reduced_virulence	15	11	4
P26469#PHI:9594#RfaC#28901#Salmonella_enterica#reduced_virulence	20	19	1
A6T684#PHI:7956#PagP#573#Klebsiella_pneumoniae#unaffected_pathogenicity	25	25	0
D4I0P2#PHI:7109_PHI:7391#VgrG-3_EAMY_3003#552#Erwinia_amylovora#effector_(plant_avirulence_determinant)_unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	28	15	13
Q9K2Y6#PHI:7845#FliC#562#Escherichia_coli#unaffected_pathogenicity_reduced_virulence	36	33	3
Q9HU22#PHI:7634#RmlA#287#Pseudomonas_aeruginosa#reduced_virulence_loss_of_pathogenicity	38	31	7
Q6WEF6#PHI:10603#HrcN#554#Pectobacterium_carotovorum#unaffected_pathogenicity_reduced_virulence	39	32	7

D4I230#PHI:7108_PHI:7439#Hcp-2_EAMY_3223#552#Erwinia_amylovora#effector_(plant_avirulence_determinant)_unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	53	47	6
D4HVH6#PHI:11451#RfbX#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity	54	32	22
D4I229#PHI:7094_PHI:7438#ClpV-2_EAMY_3222#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	56	50	6
A5F6S0#PHI:4108#FadD#666#Vibrio_cholerae#reduced_virulence_unaffected_pathogenicity	60	54	6
D4I209#PHI:7082_PHI:7418#TssE-2_EAMY_3202#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	63	56	7
D4I234#PHI:7081_PHI:7443#TssC-2_EAMY_3227#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	66	58	8
D4I235#PHI:7080_PHI:7444#TssB-2_EAMY_3228#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity_increased_virulence_(hypervirulence)	71	62	9
B8F4X8#PHI:3147#RfaE#738#Glaesserella_parasuis#reduced_virulence	77	59	18
A0A0F6B1Q1#PHI:6982#Trg#28901#Salmonella_enterica#reduced_virulence	82	61	21
D4I0R5#PHI:7087_PHI:7414#TssK-1_EAMY_3026#552#Erwinia_amylovora#reduced_virulence_effector_(plant_avirulence_determinant)_unaffected_pathogenicity_increased_virulence_(hypervirulence)	82	68	14
P69423#PHI:10983#TatC#562#Escherichia_coli#reduced_virulence	85	67	18
A6TFA0#PHI:7949#ArnB#573#Klebsiella_pneumoniae#unaffected_pathogenicity	87	68	19
A3M1G3#PHI:7860#RpoB#470#Acinetobacter_baumannii#reduced_virulence	90	69	21
E0SH08#PHI:8577#CarA_(Dda3937_01390)#204038#Dickeya_dadantii#reduced_virulence	91	78	13
A0A0U1J4N7#PHI:6697#YqiC#28901#Salmonella_enterica#reduced_virulence	92	72	20
A1ABF3#PHI:3466#RstA#562#Escherichia_coli#reduced_virulence	93	74	19
W1ATL3#PHI:7534#CcmA_(VK055_1938)#573#Klebsiella_pneumoniae#unaffected_pathogenicity	93	75	18
A0A0F6AZN3#PHI:123431#OmpA#28901#Salmonella_enterica#reduced_virulence_increased_virulence_(hypervirulence)_unaffected_pathogenicity	95	73	22
P19642#PHI:4631#PafP#562#Escherichia_coli#reduced_virulence	95	78	17
A0A2K8VZT6#PHI:124154#BcsA_(=DS0823_v1_2374)#1089444#Dickeya_solani#unaffected_pathogenicity	97	76	21
D2TU76#PHI:8512_PHI:8688_PHI:123999#CpxR#67825#Citrobacter_rodentium#loss_of_pathogenicity_reduced_virulence_unaffected_pathogenicity	97	78	19
D2TJQ0#PHI:4473#CroI#67825#Citrobacter_rodentium#increased_virulence_(hypervirulence)	98	77	21
A6TF98#PHI:7951#ArnA#573#Klebsiella_pneumoniae#unaffected_pathogenicity	98	79	19

A6TF99#PHI:7950#ArnC#573#Klebsiella_pneumoniae#unaffected_pathogenicity	98	79	19
Q8ZKE2#PHI:10047#FumB#28901#Salmonella_enterica#reduced_virulence	99	78	21
A0A0F6KRN4#PHI:11987#SlyA#615#Serratia_marcescens#increased_virulence_(hypervirulence)	99	78	21
A0A0F7J4C9#PHI:11220#TomB#28901#Salmonella_enterica#reduced_virulence	99	78	21
A0A8E6Y3T8#PHI:123638#DcuR#28901#Salmonella_enterica#increased_virulence_(hypervirulence)	99	78	21
C6DEJ2#PHI:7469#RplY#554#Pectobacterium_carotovorum#reduced_virulence	99	77	22
P0A2F2#PHI:5572#Cob#28901#Salmonella_enterica#unaffected_pathogenicity	99	78	21
P0AC98#PHI:9230#SatP#562#Escherichia_coli#unaffected_pathogenicity	99	78	21
P30745#PHI:10987#MoaA#562#Escherichia_coli#unaffected_pathogenicity	99	78	21
Q6D803#PHI:8933#RsmS_(ECA1172)#29471#Pectobacterium_atrosepticum#increased_virulence_(hypervirulence)	99	78	21
Q7VTC1#PHI:11120#RplN#520#Bordetella_pertussis#unaffected_pathogenicity_increased_virulence_(hypervirulence)	99	77	22
U6ZHS1#PHI:4179#ExpI#1089444#Dickeya_solani#reduced_virulence	99	77	22
A0A086IQC3#PHI:8968#ProV#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
C6DD61#PHI:10578#PccS1_002702#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
E0SGC8#PHI:8570#MetB_(Dda3937_03888)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
P0ACF8#PHI:6369#Hns#562#Escherichia_coli#increased_virulence_(hypervirulence)	100	78	22
Q8Z8K9#PHI:9460#EntA#28901#Salmonella_enterica#reduced_virulence	100	78	22
W9B570#PHI:7538#FepA_(VK055_1934)#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
W9B580#PHI:7486#EntE_(VK055_1924)#573#Klebsiella_pneumoniae#unaffected_pathogenicity_reduced_virulence	100	78	22
W9BNM8#PHI:6565_PHI:6652_PHI:7479#EntB_EntB_(VK055_1923)#573#Klebsiella_pneumoniae#unaffected_pathogenicity_reduced_virulence	100	78	22
A0A076ZMY5#PHI:7520#VK055_1952#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
A0A076ZPI5#PHI:7533#VK055_1939#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
A0A076ZPI8#PHI:7528#VK055_1944#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
A0A0D6I7L0#PHI:11417#PpiA#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0E9B8Z9#PHI:6361#MotA#553#Pantoea_ananatis#reduced_virulence	100	78	22
A0A0F6B0W4#PHI:6893_PHI:10099#SufB_STM14_RS07705_(sufB)#28901#Salmonella_enterica#unaffected_pathogenicity_reduced_virulence	100	78	22
A0A0F6B0W7#PHI:10102#STM14_RS07720_(sufS)#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F6B0X1#PHI:10106#STM14_RS07740_(lpp)#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F6B1X0#PHI:7658#PspB#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F6B1X1#PHI:7657#PspA#28901#Salmonella_enterica#reduced_virulence_unaffected_pathogenicity	100	78	22
A0A0F6B259#PHI:123379#NarK#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F6B5S8#PHI:10082#STM14_RS15345#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22

A0A0F6B5S9#PHI:10083#STM14_RS15350#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22
A0A0F6B5T0#PHI:10084#STM14_RS15355#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22
A0A0F6B941#PHI:9436#AtpB#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F7DI09#PHI:11418#PpiB#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0F7DJX4#PHI:11419#PpiC#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A0H2V4D2#PHI:11344#C2516#562#Escherichia_coli#reduced_virulence	100	78	22
A0A0H3B2K4#PHI:11187#MAM_(YPK_1779)#633#Yersinia_pseudotuberculosis#increased_virulence_(hypervirulence)	100	78	22
A0A0H3Q0Y8#PHI:9843#AckA#666#Vibrio_cholerae#unaffected_pathogenicity	100	78	22
A0A0M2J080#PHI:6894#SufC#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22
A0A0U1YU79#PHI:6363#PilT#553#Pantoea_ananatis#reduced_virulence	100	78	22
A0A0U1YU94#PHI:6362#PilA#553#Pantoea_ananatis#reduced_virulence	100	78	22
A0A0U1YUA0#PHI:6360#FlgK#553#Pantoea_ananatis#reduced_virulence	100	78	22
A0A0W8AWM7#PHI:11411#YhbU#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
A0A164J6A6#PHI:6447#RarA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
A0A378CFB4#PHI:11408#YaaA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
A0A380PIP1#PHI:11030#BipA#632#Yersinia_pestis#reduced_virulence	100	78	22
A0A518VKX7#PHI:10587#KduD#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
A0A5W0R9X0#PHI:123166#PhoU#28901#Salmonella_enterica#reduced_virulence_loss_of_pathogenicity	100	78	22
A0A631R7G5#PHI:124019#GalU#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A8E5JM30#PHI:12047#AcrD#28901#Salmonella_enterica#reduced_virulence	100	78	22
A0A8E6NAW3#PHI:123168#Pta#28901#Salmonella_enterica#reduced_virulence_loss_of_pathogenicity	100	78	22
A1AEE9#PHI:4218#SmpB#562#Escherichia_coli#reduced_virulence	100	78	22
A1AEN2#PHI:3731#LuxS#562#Escherichia_coli#reduced_virulence	100	78	22
A1JQ74#PHI:7671#YbeY#630#Yersinia_enterocolitica#reduced_virulence	100	78	22
A3M645#PHI:10380#RecA#470#Acinetobacter_baumannii#reduced_virulence	100	78	22
A6TAW6#PHI:11407#GntR#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
B5F7U8#PHI:8862#LeuB#28901#Salmonella_enterica#reduced_virulence	100	78	22
B5QZL1#PHI:5109#Prt#28901#Salmonella_enterica#reduced_virulence	100	78	22
B5XQD2#PHI:124344#IhfA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
B5XR94#PHI:8967#MdtI#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
B5Y006#PHI:7521#BetI_(VK055_1951)#573#Klebsiella_pneumoniae#unaffected_pathogenicity	100	78	22
B5Y1W2#PHI:8962#LeuA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
C3T032#PHI:11759#Ppk1#562#Escherichia_coli#reduced_virulence	100	78	22
C5BAT0#PHI:2961#GcvP#67780#Edwardsiella_ictaluri#reduced_virulence	100	78	22
C5BEV2#PHI:2962#GlyA#67780#Edwardsiella_ictaluri#reduced_virulence	100	78	22
C5BF98#PHI:2959#Mdh#67780#Edwardsiella_ictaluri#reduced_virulence	100	78	22

C6DC70#PHI:10595#ClpB#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
C6DEY6#PHI:10573#RsmA#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
C6DI28#PHI:10580#PccS1_003516#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
C6DIJ6#PHI:10577#PccS1_001523#554#Pectobacterium_carotovorum#unaffected_pathogenicity	100	78	22
C9XTL5#PHI:5504#RpfR#413502#Cronobacter_turicensis#unaffected_pathogenicity	100	78	22
C9XTL6#PHI:5503#RpfF#413502#Cronobacter_turicensis#reduced_virulence	100	78	22
D0Z9U0#PHI:3074#HupA#636#Edwardsiella_tarda#reduced_virulence	100	78	22
D2TI49#PHI:8690#DegP_(ROD_01651)#67825#Citrobacter_rodentium#loss_of_pathogenicity_unaffected_pathogenicity	100	78	22
D2TI55#PHI:9334#Map#67825#Citrobacter_rodentium#effector_(plant_avirulence_determinant)	100	78	22
D2TM47#PHI:124002#YebE#67825#Citrobacter_rodentium#unaffected_pathogenicity	100	78	22
D2TNF0#PHI:10537#ArgR#67825#Citrobacter_rodentium#reduced_virulence	100	78	22
D2TPE2#PHI:10905#TolB#67825#Citrobacter_rodentium#reduced_virulence	100	78	22
D2TR63#PHI:10536#ArtP#67825#Citrobacter_rodentium#reduced_virulence	100	78	22
D2TRN3#PHI:124003#YgiB#67825#Citrobacter_rodentium#unaffected_pathogenicity	100	78	22
D2TU77#PHI:8513_PHI:8689_PHI:124000#CpxA#67825#Citrobacter_rodentium#loss_of_pathogenicity_reduced_virulence_unaffected_pathogenicity	100	78	22
D4GNZ9#PHI:9566#Hfq_(PANA_RS17940)#553#Pantoea_ananatis#loss_of_pathogenicity	100	78	22
D4HUC7#PHI:9309#PfkA#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HUG0#PHI:9325#MetF#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4HUG1#PHI:9322#Ppc#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4HUG4#PHI:9316#ArgG#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HUI1#PHI:9318#IlvA#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HUM6#PHI:9326#MetR#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4HUM7#PHI:9315#MetE#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HUY4#PHI:3674_PHI:3680#YhbH#552#Erwinia_amylovora#loss_of_pathogenicity	100	78	22
D4HUY5#PHI:2906_PHI:3673_PHI:3679#RpoN#552#Erwinia_amylovora#loss_of_pathogenicity	100	78	22
D4HW76#PHI:9324#HisA#552#Erwinia_amylovora#unaffected_pathogenicity_reduced_virulence	100	78	22
D4HWM2#PHI:9323#CysI3#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4HX24#PHI:3336#RpoS#552#Erwinia_amylovora#reduced_virulence_unaffected_pathogenicity	100	78	22
D4HX67#PHI:7161#CsrA#552#Erwinia_amylovora#reduced_virulence_loss_of_pathogenicity	100	78	22
D4HX89#PHI:3675_PHI:3681#YfiA#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4HXG4#PHI:9317#PurF#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HXP5#PHI:9311#PtsH#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HXP6#PHI:9312#PtsI#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HY24#PHI:7160#Lon#552#Erwinia_amylovora#increased_virulence_(hypervirulence)_loss_of_pathogenicity	100	78	22
D4HYB2#PHI:9306#GuaB#552#Erwinia_amylovora#reduced_virulence	100	78	22

D4HYF4#PHI:9320#PurL#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4HYW4#PHI:9327#AsnB#552#Erwinia_amylovora#unaffected_pathogenicity	100	78	22
D4I0P6#PHI:7101_PHI:7395#TssF-1_EAMY_3007#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
D4I0P7#PHI:7100_PHI:7396#TssE-1_EAMY_3008#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
D4I0Q0#PHI:7399#EAMY_3011#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
D4I0R0#PHI:7098_PHI:7409#TssB-1_EAMY_3021#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
D4I0R3#PHI:7090_PHI:7412#TssM-1_EAMY_3024#552#Erwinia_amylovora#reduced_virulence_effector_(plant_avirulence_determinant)_unaffected_pathogenicity_increased_virulence_(hypervirulence)	100	78	22
D4I0R4#PHI:7103_PHI:7413#TssL-1_EAMY_3025#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
D4I0W8#PHI:9307#PyrC#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4I2T0#PHI:9310#Pgi#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4I2T4#PHI:9313#MetA#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4I307#PHI:3126_PHI:9305#ArgD#552#Erwinia_amylovora#loss_of_pathogenicity_reduced_virulence_unaffected_pathogenicity	100	78	22
D4I3E6#PHI:9321#TrpB#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4I4J9#PHI:10919#EaAdeP_(EAMY_3685)#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4I4L2#PHI:9308#GlmS#552#Erwinia_amylovora#reduced_virulence	100	78	22
D4ICB6#PHI:2459#NlpI#552#Erwinia_amylovora#reduced_virulence	100	78	22
E0SDK2#PHI:8574#CysJ_(Dda3937_00209)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E0SEI0#PHI:8581#LysA_(Dda3937_02577)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E0SHD8#PHI:8575#DegQ_(Dda3937_00697)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E0SIM9#PHI:8572#FlhC_(Dda3937_02784)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E0SMW8#PHI:8568#ClpS#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E0SMW9#PHI:8569#ClpA_(Dda3937_02363)#204038#Dickeya_dadantii#reduced_virulence	100	78	22
E1AP33#PHI:2479#CorA#554#Pectobacterium_carotovorum#reduced_virulence	100	78	22
F4N4J3#PHI:124123#TldD#630#Yersinia_enterocolitica#unaffected_pathogenicity	100	78	22
H3RB93#PHI:12256#Lrp#66269#Pantoea_stewartii#reduced_virulence	100	78	22

H3RDJ5#PHI:12255#RcsA#66269#Pantoea_stewartii#reduced_virulence	100	78	22
H3RDZ6#PHI:12254#Nac#66269#Pantoea_stewartii#unaffected_pathogenicity	100	78	22
H3RGK7#PHI:12253#IscR#66269#Pantoea_stewartii#unaffected_pathogenicity	100	78	22
H3RHG9#PHI:12252#NsrR#66269#Pantoea_stewartii#unaffected_pathogenicity	100	78	22
H3RI38#PHI:3092#SoxR#66269#Pantoea_stewartii#reduced_virulence	100	78	22
H3RJ93#PHI:3091#OxyR#66269#Pantoea_stewartii#unaffected_pathogenicity	100	78	22
H9L460#PHI:11380#KatE#28901#Salmonella_enterica#reduced_virulence	100	78	22
K8DPN9#PHI:4584#FkpA#535744#Cronobacter_universalis#reduced_virulence	100	78	22
O85806#PHI:12271_PHI:12272_PHI:12275_PHI:12276#FlhD#615#Serratia_marcescens#unaffected_pathogenicity	100	78	22
P04042#PHI:6461#CheB#28901#Salmonella_enterica#reduced_virulence	100	78	22
P05416#PHI:10051#GlgA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P07014#PHI:7965#SdhB#562#Escherichia_coli#reduced_virulence	100	78	22
P07363#PHI:6535#CheA#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P07800#PHI:10050#CheZ_Chez#28901#Salmonella_enterica#reduced_virulence	100	78	22
P07801#PHI:11916#CheR#28901#Salmonella_enterica#increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
P08179#PHI:8652#PurN#562#Escherichia_coli#unaffected_pathogenicity_reduced_virulence	100	78	22
P08331#PHI:7020#CpdB#562#Escherichia_coli#reduced_virulence	100	78	22
P08870#PHI:7662#PyrE#28901#Salmonella_enterica#reduced_virulence_unaffected_pathogenicity	100	78	22
P08982#PHI:2686#EnvZ#28901#Salmonella_enterica#reduced_virulence	100	78	22
P09152#PHI:10512#NarG#562#Escherichia_coli#reduced_virulence	100	78	22
P09158#PHI:9153#SpeE#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P0A1D7#PHI:3040#ClpP#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A1G5#PHI:6508#DksA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A1P6#PHI:7742#GlnA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A251#PHI:11383#AhpC#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A2F0#PHI:2680#RpoE#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A2P8#PHI:2672#FruR#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0A6T3#PHI:6269#GalK#562#Escherichia_coli#reduced_virulence	100	78	22
P0A6Z3#PHI:6472#HtpG#562#Escherichia_coli#reduced_virulence	100	78	22
P0A7F6#PHI:9152#SpeD#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P0A917#PHI:6250_PHI:124233#OmpX#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P0A964#PHI:6536#CheW#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P0A9E5#PHI:4876_PHI:7248#FNR_Fnr#562#Escherichia_coli#reduced_virulence	100	78	22
P0A9G2#PHI:3267#NhaR#562#Escherichia_coli#reduced_virulence	100	78	22
P0A9Q1#PHI:6532_PHI:124163#ArcA#562#Escherichia_coli#reduced_virulence	100	78	22
P0A9X1#PHI:11007#ZnuC#562#Escherichia_coli#reduced_virulence	100	78	22
P0AA19#PHI:2685#OmpR#28901#Salmonella_enterica#reduced_virulence	100	78	22

P0AA28#PHI:2644#Thioredoxin_1#28901#Salmonella_enterica#reduced_virulence	100	78	22
P0AB06#PHI:11001#YcbK#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P0AB24#PHI:10991#EfeO#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P0AC41#PHI:7964#SdhA#562#Escherichia_coli#reduced_virulence	100	78	22
P0AC44#PHI:7963#SdhD#562#Escherichia_coli#reduced_virulence	100	78	22
P0ACR9#PHI:5407_PHI:7646#MprA#562#Escherichia_coli#loss_of_pathogenicity_reduced_virulence	100	78	22
P0AEC3#PHI:124164#ArcB#562#Escherichia_coli#reduced_virulence	100	78	22
P0AEV1#PHI:124165#RssB#562#Escherichia_coli#reduced_virulence	100	78	22
P0AF28#PHI:124167#NarL#562#Escherichia_coli#reduced_virulence	100	78	22
P0AF54#PHI:9228#YjcH#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P0AFV4#PHI:11093#MepS#562#Escherichia_coli#reduced_virulence	100	78	22
P0AG24#PHI:123920#SpoT#562#Escherichia_coli#reduced_virulence	100	78	22
P11349#PHI:10513#NarH#562#Escherichia_coli#reduced_virulence	100	78	22
P11350#PHI:10515#NarI#562#Escherichia_coli#reduced_virulence	100	78	22
P11557#PHI:6729#DamX#562#Escherichia_coli#reduced_virulence	100	78	22
P13035#PHI:123652#GlpD#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P16437#PHI:124366#FlgB#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22
P17750#PHI:11381#KatG#28901#Salmonella_enterica#reduced_virulence	100	78	22
P19786#PHI:9459#AroA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P21169#PHI:9151#SpeC#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P21345#PHI:124242#GltP#562#Escherichia_coli#reduced_virulence	100	78	22
P23865#PHI:9896#Prc#562#Escherichia_coli#reduced_virulence	100	78	22
P25535#PHI:7028#UbiI#562#Escherichia_coli#reduced_virulence	100	78	22
P27550#PHI:9227#Acs#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P31545#PHI:10992#EfeB#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P31550#PHI:123649#YabL#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P32705#PHI:9229#ActP#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P33221#PHI:8653#PurT#562#Escherichia_coli#unaffected_pathogenicity_reduced_virulence	100	78	22
P36556#PHI:4497#PmrA#28901#Salmonella_enterica#increased_virulence_(hypervirulence)	100	78	22
P37902#PHI:124243#GltI#562#Escherichia_coli#reduced_virulence	100	78	22
P39832#PHI:11006#ZnuB#562#Escherichia_coli#reduced_virulence	100	78	22
P40120#PHI:10993#MdoD#562#Escherichia_coli#unaffected_pathogenicity	100	78	22
P40729#PHI:8731#FlhA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P43019#PHI:11376#SodA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P50334#PHI:124230#DctA#28901#Salmonella_enterica#unaffected_pathogenicity	100	78	22
P60651#PHI:9150#SpeB#562#Escherichia_coli#reduced_virulence_unaffected_pathogenicity	100	78	22
P63798#PHI:123208#ClsA#28901#Salmonella_enterica#unaffected_pathogenicity_reduced_virulence	100	78	22

P64036#PHI:6420#EF-P#28901#Salmonella_enterica#increased_virulence_(hypervirulence)	100	78	22
P64281#PHI:7130#GreA#28901#Salmonella_enterica#loss_of_pathogenicity	100	78	22
P64394#PHI:2673#HimD#28901#Salmonella_enterica#reduced_virulence	100	78	22
P65882#PHI:2625_PHI:9416#PurA#28901#Salmonella_enterica#reduced_virulence_loss_of_pathogenicity	100	78	22
P66796#PHI:6882#QseB#28901#Salmonella_enterica#increased_virulence_(hypervirulence)	100	78	22
P67347#PHI:124212#MsrP#28901#Salmonella_enterica#reduced_virulence	100	78	22
P69054#PHI:7962_PHI:123646#SdhC#562#Escherichia_coli#reduced_virulence	100	78	22
P96058#PHI:560#SirA#28901#Salmonella_enterica#reduced_virulence	100	78	22
P96320#PHI:2499_PHI:4479#RcsB#552#Erwinia_amylovora#loss_of_pathogenicity	100	78	22
Q0WEH7#PHI:4563#TyrR#632#Yersinia_pestis#reduced_virulence	100	78	22
Q0WF94#PHI:4185#Ypo2062#632#Yersinia_pestis#reduced_virulence	100	78	22
Q46631#PHI:2471#Ams#552#Erwinia_amylovora#loss_of_pathogenicity	100	78	22
Q4UZD3#PHI:3942#XC_0509#339#Xanthomonas_campestris#reduced_virulence	100	78	22
Q54001#PHI:6553#MdsABC#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q56056#PHI:3724#RNase_III#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q5KQZ9#PHI:2652#MagA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
Q665I9#PHI:3133#PtsN#633#Yersinia_pseudotuberculosis#reduced_virulence	100	78	22
Q66A26#PHI:3134#PykF#633#Yersinia_pseudotuberculosis#reduced_virulence	100	78	22
Q66EI1#PHI:3135#PdhR#633#Yersinia_pseudotuberculosis#reduced_virulence	100	78	22
Q6CZA0#PHI:2695#MetJ#29471#Pectobacterium_atrosepticum#reduced_virulence	100	78	22
Q6D4B5#PHI:11149#Eda#29471#Pectobacterium_atrosepticum#reduced_virulence	100	78	22
Q7CJ94#PHI:5089#RovM#632#Yersinia_pestis#reduced_virulence	100	78	22
Q7CQY3#PHI:11509#Fur#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q7CR17#PHI:11219#Hha#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q7CR43#PHI:123534#PhoB#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q7WTQ9#PHI:2469#AcrB#552#Erwinia_amylovora#reduced_virulence	100	78	22
Q8PBY5#PHI:9833#ClpX#339#Xanthomonas_campestris#reduced_virulence	100	78	22
Q8Z447#PHI:9161#RelA#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8Z7H2#PHI:9465#PhoP#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8ZLJ2#PHI:7131#GreB#28901#Salmonella_enterica#loss_of_pathogenicity	100	78	22
Q8ZLT3#PHI:3878#Pnp#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8ZMB9#PHI:8729#AmiC#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8ZMF9#PHI:6187#Pimt#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8ZMX2#PHI:5571_PHI:6366#Pat#28901#Salmonella_enterica#reduced_virulence_increased_virulence_(hypervirulence)	100	78	22
Q8ZNK8#PHI:11609#YeiE#28901#Salmonella_enterica#reduced_virulence	100	78	22
Q8ZPL7#PHI:10048#FumC#28901#Salmonella_enterica#reduced_virulence	100	78	22

Q8ZRE1#PHI:9434#PhoR#28901#Salmonella_enterica#increased_virulence_(hypervirulence)_reduced_virulence	100	78	22
Q9F496#PHI:5564_PHI:8654#OpgG#204038#Dickeya_dadantii#reduced_virulence	100	78	22
Q9I0J4#PHI:8937#NuoI_(PA2644)#287#Pseudomonas_aeruginosa#unaffected_pathogenicity_reduced_virulence	100	78	22
Q9I690#PHI:3166#PA0423#287#Pseudomonas_aeruginosa#reduced_virulence	100	78	22
U5U790#PHI:6448#OqxA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
U6ZHA8#PHI:4175#KdgR#1089444#Dickeya_solani#unaffected_pathogenicity	100	78	22
V6A547#PHI:4074#WecA#615#Serratia_marcescens#reduced_virulence	100	78	22
W8DUB8#PHI:3407#TssA#553#Pantoea_ananatis#unaffected_pathogenicity	100	78	22
W8DVN3#PHI:3408#TssD#553#Pantoea_ananatis#unaffected_pathogenicity	100	78	22
W8ZK20#PHI:9198#YegN#562#Escherichia_coli#reduced_virulence	100	78	22
W9B383#PHI:11405#DedA#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
W9B4Q6#PHI:6449#OqxB#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
W9BB63#PHI:6616#CRP#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
W9BBU8#PHI:8964#LeuC#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
W9BQ45#PHI:8965#LeuD#573#Klebsiella_pneumoniae#reduced_virulence	100	78	22
D4I0P4#PHI:7106_PHI:7393#ClpV-1_EAMY_3005#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	101	79	22
D4I0Q9#PHI:7099_PHI:7408#TssC-1_EAMY_3020#552#Erwinia_amylovora#unaffected_pathogenicity_increased_virulence_(hypervirulence)_reduced_virulence	101	78	23
P0A9Y9#PHI:7262#CspC#28901#Salmonella_enterica#reduced_virulence_loss_of_pathogenicity_unaffected_pathogenicity	101	78	23
P0A2D5#PHI:8732_PHI:10049#CheY#28901#Salmonella_enterica#reduced_virulence	102	80	22
Q9I1W9#PHI:10946#PA2146#287#Pseudomonas_aeruginosa#increased_virulence_(hypervirulence)	197	156	41
W8TRT9#PHI:10282#TufA#1280#Staphylococcus_aureus#increased_virulence_(hypervirulence)_unaffected_pathogenicity	200	157	43
A0A6C7HPB2#PHI:10387#SEN1538#28901#Salmonella_enterica#reduced_virulence	216	157	59
A0A0F6AYC2#PHI:7263#CspE#28901#Salmonella_enterica#reduced_virulence_loss_of_pathogenicity_unaffected_pathogenicity	399	312	87

Table 16: Human pathogenicity potential of the 100 *P. agglomerans* strains and both outgroup taxa.

Strain	Probability of being a human pathogen	Input proteome coverage (%)
9Rz4	0,31	0,93
20TX0122	0,273	1
62d	0,296	0,92
62e	0,296	0,92
18059	0,334	1,02
A2	0,299	0,87
AB378	0,308	0,97
AR1a	0,27	0,91
ASB05	0,275	0,91
B1	0,316	0,9
B3	0,316	0,9
BAV2934	0,317	0,97
BI3	0,333	0,94
C1	0,293	0,92
C410P1	0,318	0,92
CB1	0,295	1,01
CCNPK873	0,303	0,9
CFBP8756	0,323	0,95
CFBP8784	0,309	0,97
CFBP8786	0,355	0,94
CFBP13505	0,306	0,93
CFBP13569	0,319	0,99
CFBP13583	0,303	0,89
CFBP13600	0,317	0,89
CFBP13616	0,315	0,94
CFBP13731	0,323	0,95
CFSAN033090	0,296	0,89
CFSAN047153	0,335	0,84
CFSAN047154	0,335	0,84
CHTF15	0,294	0,93
DAPP-PG734	0,285	0,96
DBM3797	0,338	0,98
DOAB1048	0,295	0,91
E325	0,304	0,89
E325-699	0,305	0,87
E325-705	0,304	0,91
E325-750	0,323	0,89
E325-754	0,305	0,87
E325-ad1	0,305	0,89
E325-ad2	0,305	0,89
Eh318	0,296	0,87
EKM10T	0,293	0,95
EKM20T	0,307	0,93
EKM21T	0,299	0,89
EKM22T	0,299	0,89
FDAARGOS1447	0,331	1,01
GD03966	0,285	0,86
GR13	0,319	0,9
IG1	0,243	0,56
JM1	0,338	1,04
JZB2120015	0,282	0,87
K1	0,285	0,83
KM1	0,285	0,83
L15	0,315	0,9

MP2	0,339	0,92
MR10	0,301	0,87
MR19	0,301	0,87
NBBC-01	0,29	0,9
NC	0,317	0,92
NCTC10500	0,279	0,91
NCTC10601	0,283	0,88
OVA06A	0,3	0,88
OVA07A	0,295	0,94
P10c	0,31	0,93
PA	0,0282	0,9
Pa17-1	0,311	0,86
Pa17-5	0,314	0,9
Pa21-3	0,299	0,85
Pa21-5	0,319	0,88
Pa21-13	0,276	0,93
Pa21-15	0,315	0,9
Pa31-3	0,316	0,87
Pa31-4	0,314	0,96
Pa39-1	0,303	0,95
Pa39-7	0,306	0,92
Pa39-21	0,319	0,84
Pa39-23	0,335	0,85
Pa39-27	0,327	0,91
Pa58	0,32	0,88
PaVv1	0,278	0,9
PMG056	0,296	0,93
PSV1-7	0,299	0,89
4188	0,337	1,03
824-1	0,321	0,88
R1	0,295	0,93
R5	0,302	0,88
RIT710	0,295	0,92
RIT-PIT-AA	0,295	0,91
RIT-PI-T	0,33	0,91
RS07	0,281	1,01
SI1	0,32	0,84
3	0,325	0,89
4	0,298	0,91
190	0,351	1,02
T1	0,295	0,98
TH81	0,321	0,88
Tx10	0,318	0,91
UAEU18	0,313	0,93
ZBG6	0,276	0,92
ZJU23	0,307	0,89
P. eucalypti LMG_24197^T	0,311	1,13
P. vagans LMG_24199^T	0,34	1,2

Table 17: Subcellular localisations of the proteome of the 100 *P. agglomerans* and both outgroup taxa.

Strain	Cytoplasmic	Cytoplasmic Membrane	Extracellular	Outer Membrane	Periplasmic	Unknown
9Rz4	1851	1166	53	75	167	1229
20TX0122	1907	1178	44	84	177	1218
62d	1797	1135	42	75	164	1125
62e	1796	1135	42	75	164	1128
18059	1786	1139	46	78	167	1095
A2	1900	1162	47	83	168	1225
AB378	1795	1139	45	77	166	1119
AR1a	1912	1189	43	82	180	1233
ASB05	1836	1137	46	77	166	1158
B1	1934	1198	58	84	179	1369
B3	1935	1196	59	83	179	1379
BAV2934	1860	1163	49	82	166	1218
BI3	1892	1175	51	83	177	1314
C1	1826	1137	50	77	170	1199
C410P1	1869	1226	51	94	172	1394
CB1	1915	1166	41	82	167	1258
CCNPK873	1804	1130	45	75	160	1109
CFBP8756	1884	1176	49	78	163	1309
CFBP8784	1803	1133	52	78	164	1125
CFBP8786	1873	1152	51	82	165	1284
CFBP13505	1848	1149	47	79	163	1221
CFBP13569	1814	1148	53	76	167	1175
CFBP13583	1803	1140	56	74	161	1132
CFBP13600	1838	1154	51	75	170	1206
CFBP13616	1810	1150	47	74	165	1131
CFBP13731	1894	1171	47	77	162	1299
CFSAN033090	1862	1163	49	75	163	1304
CFSAN047153	1909	1182	49	83	184	1355
CFSAN047154	1811	1149	53	76	162	1194
CHTF15	1836	1157	41	77	169	1134
DAPP-PG734	1997	1243	56	87	177	1442
DBM3797	1810	1137	54	79	169	1149
DOAB1048	1766	1136	53	77	164	1199
E325	1830	1130	44	74	172	1134
E325-699	1830	1133	46	74	169	1127
E325-705	1783	1116	46	73	164	1086
E325-750	1832	1130	50	74	170	1135
E325-754	1831	1128	45	74	169	1125
E325-ad1	1784	1117	50	75	164	1106
E325-ad2	1784	1119	45	73	163	1092
Eh318	1897	1162	55	79	166	1235
EKM10T	1792	1132	51	73	166	1102
EKM20T	1822	1169	53	79	177	1211
EKM21T	1828	1167	46	79	172	1204
EKM22T	1826	1167	46	79	173	1198
FDAARGOS1447	1788	1110	46	77	162	1076
GD03966	1801	1162	45	81	162	1183
GR13	1786	1143	49	73	162	1148
IG1	1827	1142	46	77	169	1181
JM1	1808	1150	55	73	165	1173
JZB_2120015	1834	1166	44	82	168	1198
K1	1861	1182	46	90	178	1231
KM1	1861	1182	44	90	176	1233

L15	1826	1145	53	74	167	1195
MP2	1791	1124	47	74	161	1154
MR10	1862	1184	58	78	172	1243
MR19	1863	1184	57	78	171	1242
NBBC-01	1857	1172	37	75	167	1231
NC	1731	1086	41	80	165	1032
NCTC10500	1774	1107	40	78	162	1112
NCTC10601	1787	1143	46	83	168	1109
OVA06A	1824	1149	48	75	161	1203
OVA07A	1798	1142	51	78	167	1118
P10c	1816	1154	49	73	167	1137
PA	1780	1165	43	81	160	1124
Pa17-1	1835	1136	49	80	174	1177
Pa17-5	1852	1162	51	82	171	1243
Pa21-3	1877	1188	50	80	177	1294
Pa21-5	1808	1145	52	76	167	1190
Pa21-13	1807	1150	51	74	163	1165
Pa21-15	1806	1146	50	75	167	1192
Pa31-3	1883	1213	46	81	175	1334
Pa31-4	1855	1196	55	78	171	1258
Pa39-1	1859	1194	53	83	174	1197
Pa39-7	1846	1179	53	81	170	1238
Pa39-21	1874	1190	51	81	174	1283
Pa39-23	1894	1197	50	82	174	1339
Pa39-27	1870	1179	52	81	165	1280
Pa58	1811	1149	53	76	162	1194
PaVv1	1811	1162	46	82	167	1176
PMG056	1784	1120	43	76	161	1105
PSV1-7	1835	1157	51	78	175	1215
4188	1851	1171	54	86	171	1324
824-1	1907	1173	61	78	157	1330
R1	1796	1136	45	72	164	1079
R5	1809	1132	52	80	166	1183
RIT710	1803	1136	52	77	161	1123
RIT-PIT-AA	1810	1158	45	75	166	1145
RIT-PI-T	1864	1158	47	76	169	1292
RSO7	1818	1153	40	81	169	1111
SL1_M5	1887	1163	57	77	162	1282
3	1803	1140	47	76	162	1147
4	1806	1151	41	76	170	1176
190	1855	1175	51	77	172	1289
T1	1741	1087	33	78	165	1004
TH81	1848	1173	48	79	165	1249
Tx10	1813	1145	45	77	165	1155
UAEU18	1799	1150	45	85	169	1150
ZBG6	1795	1135	45	75	162	1122
ZJU23	1861	1204	47	88	169	1246
P. eucalypti LMG_24197^T	1833	1152	37	67	161	1185
P. vagans LMG_24199^T	1770	1146	48	80	164	1125

Table 18: Secretion systems in the 100 *P. agglomerans* and outgroup taxa.

Strain	Flagellar	T1SS	T3SS	T4aP	T5aSS	T5bSS	T6SSi	T4bP	T5cSS	pT4SSt
Strain 18059	1	2	1	1	1	2	1	0	0	0
20TX0122	1	2	0	1	1	2	1	0	0	0
62d	1	4	0	1	2	2	1	0	0	0
62e	1	4	0	1	2	2	1	0	0	0
9Rz4	1	3	0	1	2	1	1	0	0	0
A2	1	2	0	1	2	3	1	0	0	0
AB378	1	3	1	1	3	2	2	0	0	0
AR1a	1	1	0	1	2	3	1	0	0	0
ASB05	1	3	0	1	3	2	1	0	0	0
B1	2	3	0	1	4	1	1	0	0	0
B3	2	3	0	1	4	1	1	0	0	0
BAV2934	1	3	1	1	3	2	1	0	0	0
BI3	1	3	1	1	2	3	2	0	0	1
C1	1	3	0	1	2	1	2	0	0	0
C410P1	1	3	1	1	3	3	1	1	0	0
CB1	1	3	0	1	2	3	1	0	0	0
CCNPK873	1	3	0	1	2	3	1	0	0	0
CFBP13505	1	2	0	1	3	1	1	0	0	0
CFBP13569	1	3	1	1	2	2	2	0	0	0
CFBP13583	1	3	1	1	2	2	2	0	0	0
CFBP13600	1	3	0	1	3	1	2	0	0	0
CFBP13616	1	3	0	1	2	2	2	0	0	0
CFBP13731	1	2	0	1	3	1	2	1	0	0
CFBP8756	1	2	0	1	3	1	2	1	0	0
CFBP8784	1	3	1	1	3	2	2	0	0	0
CFBP8786	1	2	0	1	3	1	1	0	0	0
CFSAN033090	1	3	0	1	2	1	1	0	0	1
CFSAN047153	1	3	1	1	2	2	2	0	0	0
CFSAN047154	1	3	1	1	2	2	2	0	0	0
CHTF15	1	3	0	1	2	2	2	0	0	0
DAPP-PG734	1	3	1	1	2	2	2	1	0	0
DBM3797	1	4	1	1	3	2	2	0	0	0
DOAB1048	1	2	1	1	3	2	1	0	0	0
E325	1	4	0	1	2	2	2	0	0	0
E325-699	1	4	0	1	2	2	2	0	0	0

E325-705	1	4	0	1	2	2	2	0	0	0
E325-750	1	4	0	1	2	2	2	0	0	0
E325-754	1	4	0	1	2	2	2	0	0	0
E325-ad1	1	4	0	1	2	2	2	0	0	0
E325-ad2	1	4	0	1	2	2	2	0	0	0
Eh318	1	4	0	1	2	2	2	0	0	0
EKM10T	1	2	1	1	2	1	1	0	0	0
EKM20T	1	3	1	1	2	3	2	0	0	0
EKM21T	1	3	1	1	2	2	2	0	0	0
EKM22T	1	3	1	1	2	2	2	0	0	0
FDAARGOS1447	1	3	0	1	3	3	1	0	0	0
GD03966	1	3	1	1	2	3	1	0	0	0
GR13	1	3	0	1	2	2	1	0	0	0
IG1	1	1	1	1	2	2	2	0	0	0
JM1	1	3	0	1	3	2	2	0	0	0
JZB_2120015	1	3	1	1	3	3	1	0	0	0
K1	1	3	0	1	4	3	2	0	0	0
KM1	1	3	0	1	4	3	2	0	0	0
L15	1	3	1	1	2	1	1	0	0	0
MP2	1	4	0	1	2	1	2	0	0	0
MR10	1	3	0	1	2	3	2	0	0	0
MR19	1	3	0	1	2	3	2	0	0	0
NBBC-01	1	1	0	1	1	2	1	0	0	0
NC	1	2	0	1	2	2	1	0	0	0
NCTC10500	1	2	0	1	1	1	1	0	0	0
NCTC10601	1	3	0	1	2	2	1	0	0	0
OVA06A	1	3	0	1	2	3	2	0	0	0
OVA07A	1	4	1	1	2	3	2	0	0	0
P10c	1	3	0	1	2	2	1	0	0	0
PA	1	4	1	1	2	4	1	0	0	0
Pa17-1	1	2	0	1	3	3	1	0	0	0
Pa17-5	1	2	0	1	2	3	2	0	0	0
Pa21-13	1	3	1	1	2	1	1	0	0	0
Pa21-15	1	3	0	1	2	3	2	0	0	0
Pa21-3	1	3	0	1	3	3	2	0	0	0
Pa21-5	1	3	0	1	2	3	2	0	0	0
Pa31-3	1	4	1	1	2	2	2	0	0	0

Pa31-4	1	3	1	1	2	2	1	0	0	0
Pa39-1	1	3	1	1	1	3	1	0	0	0
Pa39-21	1	4	1	1	2	2	2	0	0	0
Pa39-23	1	3	1	1	2	2	2	0	1	1
Pa39-27	1	1	1	1	1	3	2	0	0	0
Pa39-7	1	3	0	1	2	3	1	0	0	0
Pa58	1	3	0	1	2	2	1	0	0	0
PaVv1	1	3	0	1	2	3	1	0	0	0
PMG056	1	3	0	1	2	2	1	0	0	0
PSV1-7	1	4	1	1	3	2	1	0	0	0
4188	1	4	1	1	2	4	1	0	0	0
824-1	1	1	1	1	1	2	1	0	0	0
R1	1	3	0	1	3	2	2	0	0	0
R5	1	3	0	1	4	3	2	0	0	0
RIT710	1	4	0	1	2	3	1	0	0	0
PIT-PI-T	1	3	1	1	2	1	2	0	0	0
RIT-PIT-AA	1	3	1	1	2	2	2	0	0	0
RSO7	1	3	0	1	3	3	1	0	0	0
SL1_M5	1	3	1	1	2	1	1	0	0	0
Strain 190	1	3	1	1	3	2	2	0	0	0
Strain 3	1	3	1	1	2	2	2	0	0	0
Strain 4	1	3	0	1	2	2	1	0	0	0
T1	1	2	0	1	1	3	2	0	0	0
TH81	1	3	0	1	3	3	1	0	0	0
TX10	1	4	0	1	3	1	2	0	0	0
UAEU18	1	3	0	1	2	2	2	1	0	0
ZBG6	1	3	0	1	2	2	1	0	0	0
ZJU23	1	3	1	1	2	4	1	0	0	0
P. eucalypti LMG24197^T	1	1	0	1	2	0	1	0	0	0
P. vagans LMG24199^T	1	4	0	1	1	2	2	0	0	0

Table 19: Number of AMR genes in the genomes of the 100 *P. agglomerans* and two outgroup taxa.

Strain	No. of AMR genes
9Rz4	5
20TX0122	4
62d	4
62e	4
18059	4
A2	4
AB378	4
AR1a	5
ASB05	5
B1	5
B3	5
BAV2934	5
BI3	4
C1	5
C410P1	3
CB1	4
CCNPK873	5
CFBP8756	4
CFBP8784	2
CFBP8786	4
CFBP13505	3
CFBP13569	4
CFBP13583	4
CFBP13600	3
CFBP13616	4
CFBP13731	4
CFSAN033090	4
CFSAN047153	4
CFSAN047154	4
CHTF15	5
DAPP-PG734	5
DBM3797	4
DOAB1048	4
E325	4
E325-699	5
E325-705	4
E325-750	5
E325-754	5
E325-ad1	5
E325-ad2	5
Eh318	3
EKM10T	4
EKM20T	5
EKM21T	4
EKM22T	4
FDAARGOS1447	3
GD03966	4
GR13	4
IG1	5
JM1	4
JZB2120015	3
K1	4
KM1	4
L15	5

MP2	3
MR10	4
MR19	4
NBBC-01	4
NC	4
NCTC10500	4
NCTC10601	4
OVA06A	5
OVA07A	4
P10c	5
PA	3
Pa17-1	4
Pa17-5	4
Pa21-3	4
Pa21-5	5
Pa21-13	3
Pa21-15	5
Pa31-3	5
Pa31-4	4
Pa39-1	3
Pa39-7	4
Pa39-21	4
Pa39-23	4
Pa39-27	4
Pa58	4
PaVv1	5
PMG056	5
PSV1-7	4
4188	4
824-1	4
R1	4
R5	3
RIT710	5
RIT-PIT-AA	4
RIT-PI-T	5
RS07	4
SI1	3
3	5
4	4
190	4
T1	4
TH81	4
Tx10	4
UAEU18	4
ZBG6	5
ZJU23	4
LMG24197	4
LMG24199	3

Table 20: Number of heavy metal resistance genes in the 100 *P. agglomerans* and both outgroup taxa.

Strain	No. of genes
9Rz4	65
20TX0122	60
62d	58
62e	58
18059	55
A2	61
AB378	55
AR1a	65
ASB05	54
B1	61
B3	61
BAV2934	54
BI3	54
C1	65
C410P1	58
CB1	61
CCNPK873	58
CFBP8756	57
CFBP8784	55
CFBP8786	57
CFBP13505	57
CFBP13569	58
CFBP13583	54
CFBP13600	65
CFBP13616	58
CFBP13731	57
CFSAN033090	59
CFSAN047153	58
CFSAN047154	58
CHTF15	67
DAPP-PG734	59
DBM3797	55
DOAB1048	57
E325	58
E325-699	58
E325-705	58
E325-750	58
E325-754	58
E325-ad1	58
E325-ad2	58
Eh318	57
EKM10T	54
EKM20T	56
EKM21T	65
EKM22T	65
FDAARGOS1447	54
GD03966	59
GR13	58
IG1	54
JM1	58
JZB2120015	64
K1	60
KM1	60
L15	54

MP2	61
MR10	71
MR19	71
NBBC-01	58
NC	65
NCTC10500	65
NCTC10601	69
OVA06A	54
OVA07A	58
P10c	58
PA	59
Pa17-1	58
Pa17-5	54
Pa21-3	55
Pa21-5	60
Pa21-13	57
Pa21-15	60
Pa31-3	65
Pa31-4	54
Pa39-1	54
Pa39-7	57
Pa39-21	54
Pa39-23	54
Pa39-27	60
Pa58	58
PaVv1	66
PMG056	54
PSV1-7	62
4188	58
824-1	58
R1	54
R5	59
RIT710	58
RIT-PIT-AA	54
RIT-PI-T	54
RS07	58
SI1	54
3	58
4	54
190	54
T1	54
TH81	58
Tx10	58
UAEU18	60
ZBG6	65
ZJU23	61
LMG24197	57
LMG24199	56

Table 21: Distinct heavy metal resistance genes found in *P. agglomerans* subsp. 1 and 2.

Distinct gene	Total no. of strains	Subsp. 1	Subsp. 2	Resistance
BAC0005|acrA|sp|P0AE06|ACRA_ECOLI	100	78	22	Acriflavine, Phenol, Triclosan, p-xylene, Cyclohexane, Pentane
BAC0006|acrB|sp|P31224|ACRB_ECOLI	100	78	22	Acriflavine, Phenol, Triclosan, p-xylene, Cyclohexane, Pentane
BAC0034|arsH|tr|E8PS81|E8PS81_YERPE	80	57	23	Arsenic (As)
BAC0039|baeR|sp|P69228|BAER_ECOLI	2	2	0	Zinc (Zn), Tungsten (W), Sodium Deoxycholate
BAC0041|bcr|sp|P28246|BCR_ECOLI	100	78	22	Acriflavine
BAC0061|cepA|sp|Q8RR17|FIEF_KLEPN	1	1	0	Chlorhexidine
BAC0088|corC|sp|P0A2L3|CORC_SALTY	100	78	22	Cobalt (Co), Magnesium (Mg)
BAC0105|cueR/ybbI|sp|P0A9G4|CUER_ECOLI	100	78	22	Copper (Cu)
BAC0111|cusR/ylcA|sp|P0ACZ8|CUSR_ECOLI	13	10	3	Copper (Cu), Silver (Ag)
BAC0115|cutE/lnt|sp|P23930|LNT_ECOLI	100	78	22	Copper (Cu)
BAC0137|dsbB|sp|P0A6M2|DSBB_ECOLI	100	78	22	Cadmium (Cd), Mercury (Hg)
BAC0146|emmdR|tr|D5CJ69|D5CJ69_ENTCC	99	77	22	Benzylkonium Chloride (BAC), Ethidium Bromide, Acriflavine, Rhodamine 6G
BAC0147|emrA|sp|P27303|EMRA_ECOLI	100	78	22	Phenylmercury Acetate, 2-Chlorophenylhydrazine, Carbonylcyanide m-chlorophenyl hydrazone (CCCP), Tetrachlorosalicylanilide (TCS)
BAC0148|emrB|sp|P0AEJ0|EMRB_ECOLI	100	78	22	Phenylmercury Acetate, 2-Chlorophenylhydrazine, Carbonylcyanide m-chlorophenyl hydrazone (CCCP)
BAC0156|fabI|sp|P0AEK4|FABI_ECOLI	100	78	22	Triclosan
BAC0166|fetB/ybbM|sp|P77307|YBBM_ECOLI	13	10	3	Iron (Fe), Hydrogen Peroxide (H2O2) [class: Peroxides]
BAC0167|fieF/yiip|sp|P69380|FIEF_ECOLI	99	77	22	Iron (Fe), Zinc (Zn), Cobalt (Co), Cadmium (Cd), Nickel (Ni)
BAC0181|glpF|sp|P0AER0|GLPF_ECOLI	100	78	22	Antimony (Sb), Arsenic (As), Glycerol
BAC0194|ibpA|sp|P0C054|IBPA_ECOLI	100	78	22	Hydrogen Peroxide (H2O2)
BAC0196|iclR|sp|P16528|ICLR_ECOLI	100	78	22	Sodium acetate
BAC0211|mdtB/yegN|sp|P76398|MDTB_ECOLI	100	78	22	Sodium Deoxycholate, Hydrochloric acid (HCl)
BAC0212|mdtC/yegO|sp|P76399|MDTC_ECOLI	100	78	22	Sodium Deoxycholate
BAC0218|mdtK/ydhE|sp|P37340|MDTK_ECOLI	100	78	22	Tetraphenylphosphonium (TPP), Sodium Deoxycholate, Ethidium Bromide, Benzylkonium Chloride (BAC), Acriflavine
BAC0251|mntH/yfeP|sp|P0A769|MNTH_ECOLI	100	78	22	Manganese (Mn), Iron (Fe), Cadmium (Cd), Cobalt (Co), Zinc (Zn)
BAC0253|mntR|sp|P0A9F1|MNTR_ECOLI	100	78	22	Manganese (Mn), Magnesium (Mg)
BAC0265|ncrC|tr|D5CKG5|D5CKG5_ENTCC	2	0	2	Nickel (Ni)
BAC0277|nirB|tr|Q6RUG2|Q6RUG2_KLEOX	2	0	2	Nickel (Ni)
BAC0294|oqxA|tr|Q69HW3|Q69HW3_ECOLX	100	78	22	Benzylkonium Chloride, Clorine Dioxide, Triclosan, Cetrimide, Chlorhexidine, Sodium Dodecyl Sulfate, Acriflavine

BAC0295|oqxB|tr|Q69HW2|Q69HW2_ECOLX	100	78	22	Benzylkonium Chloride, Clorine Dioxide, Triclosan, Cetrimide, Chlorhexidine, Sodium Dodecyl Sulfate, Acriflavine
BAC0298|pbrA|tr|Q58AJ6|Q58AJ6_RALME	7	1	6	Lead (Pb)
BAC0303|pcoA|sp|Q47452|PCOA_ECOLX	23	14	9	Copper (Cu)
BAC0308|pcoR|sp|Q47456|PCOR_ECOLX	23	14	9	Copper (Cu)
BAC0312|pitA|sp|P0AFJ7|PITA_ECOLI	100	78	22	Zinc (Zn), Tellurium (Te)
BAC0315|pstA|sp|P07654|PSTA_ECOLI	100	78	22	Arsenic (As)
BAC0316|pstB|sp|P0AAH0|PSTB_ECOLI	100	78	22	Arsenic (As)
BAC0317|pstC|sp|P0AGH8|PSTC_ECOLI	100	78	22	Arsenic (As)
BAC0318|pstS|sp|P0AG82|PSTS_ECOLI	100	78	22	Arsenic (As)
BAC0334|robA|sp|P0ACI0|ROB_ECOLI	100	78	22	Silver (Ag), Mercury (Hg), Cadmium (Cd), Cyclohexane, Pentane, n-hexane, Diphenyl Ether
BAC0335|rpoS|sp|P35540|RPOS_SHIFL	100	78	22	Hydrochloric acid (HCl), Sodium hydroxide (NaOH)
BAC0341|silA|sp|Q9ZHC9|SILA_SALTM	13	10	3	Silver (Ag)
BAC0342|silB|sp|Q9ZHD0|SILB_SALTM	13	10	3	Silver (Ag)
BAC0343|silC|sp|Q9ZHD2|SILC_SALTM	13	10	3	Silver (Ag)
BAC0345|silF|tr|Q9ZHD1|Q9ZHD1_SALTM	13	10	3	Silver (Ag)
BAC0348|silS|sp|Q9ZHD4|SILS_SALTM	13	10	3	Silver (Ag)
BAC0349|sitA|tr|Q9XCS2|Q9XCS2_SALTI	100	78	22	Manganese (Mn), Iron (Fe), Hydrogen Peroxide (H2O2)
BAC0350|sitB|tr|Q9XCS1|Q9XCS1_SALTM	100	78	22	Manganese (Mn), Iron (Fe), Hydrogen Peroxide (H2O2)
BAC0351|sitC|tr|Q9XCS0|Q9XCS0_SALTM	100	78	22	Manganese (Mn), Iron (Fe), Hydrogen Peroxide (H2O2)
BAC0352|sitD|tr|Q9XCR9|Q9XCR9_SALTM	1	0	1	Manganese (Mn), Iron (Fe), Hydrogen Peroxide (H2O2)
BAC0353|smdA|tr|A7VN01|A7VN01_SERMA	100	78	22	4,6-diamidino-2-phenylindole (DAPI), Hoechst 33342
BAC0354|smdB|tr|A7VN02|A7VN02_SERMA	100	78	22	4,6-diamidino-2-phenylindole (DAPI), Hoechst 33342
BAC0355|ruvB|sp|Q51426|RUVB_PSEAE	100	78	22	Chromium (Cr), Tellurium (Te), Selenium (Se)
BAC0368|sodA|sp|P00448|SODM_ECOLI	100	78	22	Selenium (Se), Hydrogen Peroxide (H2O2)
BAC0387|terB|sp|P18779|TERB_ALCSP	4	4	0	Tellurium (Te)
BAC0388|terC|sp|P18780|TERC_ALCSP	4	4	0	Tellurium (Te)
BAC0389|terD|sp|P18781|TERD_ALCSP	4	4	0	Tellurium (Te)
BAC0390|terE|sp|P18782|TERE_ALCSP	4	4	0	Tellurium (Te)
BAC0393|tolC|sp|P02930|TOLC_ECOLI	100	78	22	Phenol, Triclosan, Sodium Deoxycholate, Sodium Cholate, Sodium Taurodeoxycholate, Sodium Dodecyl Sulfate, Proflavin, Acriflavine, Ethidium Bromide, Tetraphenylphosphonium, Rhodamine 6G, Tetraphenylarsonium, Cetrimide, Dequalinium, Benzylkonium Chloride, Plumbagin
BAC0459|zitB/ybgR|sp|P75757|ZITB_ECOLI	100	78	22	Zinc (Zn)
BAC0462|zntR/yhdM|sp|P0ACS5|ZNTR_ECOLI	99	77	22	Zinc (Zn)
BAC0464|znuB/yebI|sp|P39832|ZNUB_ECOLI	100	78	22	Zinc (Zn)

BAC0465|znuC/yebM|sp|P0A9X1|ZNUC_ECOLI	100	78	22	Zinc (Zn)
BAC0477|kpnF|tr|C4X7Z4|C4X7Z4_KLEPN	100	78	22	Sodium Dodecyl Sulfate, Sodium Deoxycholate, Benzylkonium Chloride, Chlorhexidine, Triclosan, Methyl Viologen, Hydrogen Peroxide
BAC0492|kexD|tr|A6TA71|A6TA71_KLEP7	2	1	1	Rhodamine 6G, Tetraphenylphosphonium, Ethidium Bromide, Sodium Cholate, Sodium Deoxycholate
BAC0530|phoB|tr|C4X6T6|C4X6T6_KLEPN	100	78	22	Benzylkonium Chloride, Chlorhexidine
BAC0532|cpxA|tr|C4WZK5|C4WZK5_KLEPN	100	78	22	Hydrogen Peroxide, Benzylkonium Chloride, Chlorhexidine
BAC0533|cpxR|tr|C4WZK6|C4WZK6_KLEPN	100	78	22	Hydrogen Peroxide, Benzylkonium Chloride, Chlorhexidine
BAC0536|oxyRkp|tr|C4WZN6|C4WZN6_KLEPN	100	78	22	Hydrogen Peroxide, Sodium Deoxycholate, Acriflavinem Rhodamine 6G, Ethidium Bromide, Sodium Dodecyl Sulfate
BAC0540|nfsA|sp|P17117|NFSA_ECOLI	100	78	22	Chromium (Cr)
BAC0559|emrR|sp|P0ACR9|MPRA_ECOLI	100	78	22	Carbonyl cyanide 3-chlorophenylhydrazone (CCCP)
BAC0563|acrD|tr|Q8ZN77|Q8ZN77_SALTY	100	78	22	Copper (Cu), Zinc (Zn)
BAC0564|actP/yjcG|sp|P32705|ACTP_ECOLI	100	78	22	Sodium Glycocholate, Sodium acetate
BAC0576|arsB|sp|P52146|ARSB2_ECOLX	2	2	0	Arsenic (As), Antimony (Sb)
BAC0577|arsB|sp|P74985|ARSB_YEREN	45	35	10	Arsenic (As)
BAC0578|arsB|tr|O50594|O50594_ACIMU	38	24	14	Arsenic (As), Antimony (Sb)
BAC0582|arsC|sp|P08692|ARSC1_ECOLX	37	22	15	Arsenic (As), Antimony (Sb)
BAC0584|arsC|sp|O50595|ARSC_ACIMA	43	34	9	Arsenic (As), Antimony (Sb)
BAC0588|arsR|tr|P74986|P74986_YEREN	45	35	10	Arsenic (As)
BAC0596|baeR|tr|D0ZNE3|D0ZNE3_SALT1	98	76	22	Copper (Cu), Zinc (Zn), Tungsten (W), Sodium Deoxycholate
BAC0609|modA|sp|P37329|MODA_ECOLI	3	0	3	Tungsten (W), Molybdenum (Mo)
BAC0610|modB|sp|P0AF01|MODB_ECOLI	100	78	22	Tungsten (W), Molybdenum (Mo)
BAC0641|corA|sp|P0A2R8|CORA_SALTY	100	78	22	Magnesium (Mg), Cobalt (Co), Nickel (Ni), Manganese (Mn)
BAC0644|corD|sp|Q56017|APAG_SALTY	100	78	22	Cobalt (Co), Magnesium (Mg)

Appendix G

Table 22: Secondary metabolites produced in the 100 *P. agglomerans* strains.

Strain	Furan	Homoserine lactone	Lassopeptide	NI-siderophore	NRP-metallophore	NRPS	NRPS-like	Phenazine	Phosphonate	Redox cofactor	RiPP-like	RRE-containing	Pantocin A	Terpene	Thiopeptide	Total
20TX0122		2		1	1				1	1			1	1	1	9
62d		2		1	1					1				1	1	7
62e		2		1	1					1				1	1	7
9Rz4		2		1	1	1				1				1	1	8
A2		2		1	1					1				1	1	7
AB378		2		1	1					1				1	1	7
AR1a		2		1	1	1			1	1				1	1	9
ASB05		2		1	1			1		1				1	1	8
B1		2		1	1					1				1	1	7
B3		2		1	1					1				1	1	7
BAV2934		2		1	1			1		1				1	1	8
BI3		2	1	1	1					1				1	1	8
C1		2		1	1					1				1	1	7
C410P1		2		1	1		2			1				1	1	9
CB1		2		1	1				1	1				1	1	8
CCNPK873		2		1	1					1				1	1	7
CFBP13505		2		1	1					1				1	1	7
CFBP13569		2		1	1					1				1	1	7
CFBP13583		2		1	1					1				1	1	7
CFBP13600		2		1	1					1		1		1	1	8
CFBP13616		2		1	1					1				1	1	7
CFBP13731		2		1	1					1				1	1	7
CFBP8756		2		1	1					1				1	1	7
CFBP8784		2		1	1					1				1	1	7
CFBP8786		2		1	1					1				1	1	7
CFSAN033090		2		1	1					1				1	1	7
CFSAN047153		2	1	1	1			1		1				1	1	9
CFSAN047154		2	1	1	1			1		1				1	1	9
CHTF15		2		1	1					1				1	1	7
DAPP-PG734		2		1	1		1			1				1	1	8
DBM3797		2		1	1					1				1	1	7
DOAB1048		2		1	1					1				1	1	7
E325		2		1	1					1				1	1	7
E325-699		2		1	1					1				1	1	7

E325-705		2		1	1					1				1	1	7
E325-750		2		1	1					1				1	1	7
E325-754		2		1	1					1				1	1	7
E325-ad1		2		1	1					1				1	1	7
E325-ad2		2		1	1					1				1	1	7
Eh318		2		1	1					1			1	1	1	8
EKM10T	1	2		1	1					1				1	1	8
EKM20T		2		1	1					1				1	1	7
EKM21T		2		1	1					1				1	1	7
EKM22T		2		1	1					1				1	1	7
FDAARGOS1447		2		1	1					1				1	1	7
GD03966		1		1	1					1				1	1	6
GR13		2		1	1					1				1	1	7
IG1		1		1	1					1				1	1	6
JM1		2		1	1					1				1	1	7
JZB2120015		2		1	1					1				1	1	7
K1		2		1	1					1				1	1	7
KM1		2		1	1					1				1	1	7
L15		2		1	1					1				1	1	7
MP2		2		1	1					1				1	1	7
MR10		2		1	1	2	1			1				1	1	10
MR19		2		1	1	2				1				1	1	9
NBBC-01		2		1	1					1		1		1	1	8
NC		2		1	1					1				1	1	7
NCTC10500		2		1	1					1				1	1	7
NCTC10601		2		1	1					1				1	1	7
OVA06A		2		1	1					1				1	1	7
OVA07A		2		1	1					1				1	1	7
P10c		2		1	1					1			1	1	1	8
PA		2		1	1					1				1	1	7
Pa17-1		2		1	1					1				1	1	7
Pa17-5		2		1	1					1				1	1	7
Pa21-13		2		1	1					1				1	1	7
Pa21-15		2		1	1					1				1	1	7
Pa21-3		2		1	1					1				1	1	7
Pa21-5		2		1	1					1				1	1	7
Pa31-3		2		1	1					1			1	1	1	8
Pa31-4		2		1	1					1				1	1	7
Pa39-1		2		1	1					1			1	1	1	8
Pa39-21		2		1	1			1		1				1	1	8
Pa39-23		2		1	1			1		1				1	1	8

Pa39-27		2		1	1					1				1	1	7
Pa39-7		2		1	1			1		1				1	1	8
Pa58		2		1	1					1				1	1	7
PaVv1		2		1	1					1				1	1	7
PMG056		2		1	1					1				1	1	7
PSV1-7		2		1	1					1				1	1	7
R1		2		1	1					1				1	1	7
R5		2		1	1					1				1	1	7
RIT710		2		1	1					1				1	1	7
RIT-PI-T		2		1	1					1				1	1	7
RIT-PIT-AA		2		1	1					1				1	1	7
RS07		2		1	1					1				1	1	7
SI1		2		1	1		1			1	1			1	1	9
strain 18059		2		1	1					1				1	1	7
strain 190		2		1	1			1		1				1	1	8
strain 3		2		1	1					1				1	1	7
strain 4		2		1	1		1			1				1	1	8
4188		2		1	1					1				1	1	7
824-1		2		1	1					1				1	1	7
T1		2		1	1					1				1	1	7
TH81		2		1	1					1				1	1	7
Tx10		2		1	1					1			1	1	1	8
UAEU18		2		1	1					1				1	1	7
ZBG6		2		1	1					1				1	1	7
ZJU23		2		1	1	2				1				1	1	9