PHYLOGENOMIC AND MOLECULAR SIGNATURE-BASED APPROACHES FOR RESOLVING THE EVOLUTIONARY RELATIONSHIPS AMONG *PSEUDOMONAS* SPECIES

By

BASHUDEV RUDRA, M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2025) McMaster University

(Biochemistry) Hamilton, Ontario

TITLE: Phylogenomic and Molecular Signature-Based Approaches for Resolving the Evolutionary Relationships Among *Pseudomonas* Species

AUTHOR: Bashudev Rudra, M.Sc. (University of Chittagong)

SUPERVISOR: Professor Dr. Radhey S. Gupta

NUMBER OF PAGES: xxiv, 202

DEDICATION

"To my beloved mother, Parbati Rudra, and my late father, Ranjit Kumar Rudra, who inspired me to study Molecular Biology."

ABSTRACT

The genus *Pseudomonas* includes genetically diverse groups of species that do not share a common evolutionary history. My research focused on analyzing the genome sequences of different *Pseudomonas* species to robustly elucidate their evolutionary relationships using multiple independent approaches, which include: (i) Construction of phylogenetic trees based on several large data sets of conserved proteins, and 16S rRNA gene sequences (ii) Determination of pairwise genomic similarities based on AAI and POCP matrices, (iii) Identification of molecular markers such as Conserved Signature Indels (CSIs) and Conserved Signature Proteins (CSPs), specific for different *Pseudomonas* species clades supported by other methods. Our Phylogenomic analyses revealed three major lineages/groups within *Pseudomonas*: Aeruginosa, Fluorescens, and Pertucinogena. While the Aeruginosa and Fluorescens lineages include multiple distinct clades, no molecular or biochemical traits were previously known to differentiate them. Our analyses identified >160 CSIs specific to these clades/groups, providing molecular means for their reliable demarcation. Based on phylogenomic evidence, AAI and POCP values, and clade-specific CSIs, we proposed restricting the genus *Pseudomonas* only to the Aeruginosa clade of species. Prior to this, based on our analyses, we reclassified the Pertucinogena lineage of species as a novel genus, *Halopseudomonas*, and reclassified several misclassified species into their related genera. Further analyses led to the reclassification of the Aeruginosa lineage of species into 12 novel and emended genera. Ongoing studies on the Fluorescens lineage, comprising 13 clades, have identified CSIs for several of them. Additionally, our studies led to the discovery of a novel species, *Pseudomonas paraeruginosa*. The resulting CSI-based phylogenetic framework offers a stable,

predictive system for classifying new or uncharacterized *Pseudomonas* species. Using the predictive ability of CSIs, we predicted assigning ~300 uncharacterized strains into 14 *Pseudomonadaceae* genera. Besides systematic studies, these conserved markers hold promise for diagnostic applications and deeper insights into microbial evolution and function.

ACKNOWLEDGMENTS

I express my deepest gratitude to my graduate supervisor, Professor Dr. Radhey S. Gupta, for his invaluable mentorship, patient guidance, and unflagging support throughout my Ph.D. journey. His vast knowledge, insightful advice, and innovative approach have profoundly shaped my academic and research development, inspiring me to think critically and push the boundaries of scientific inquiry. I am incredibly grateful for the intellectual freedom he provided, allowing me to explore diverse computational and laboratory techniques, participate in workshops and conferences, establish meaningful scholarly connections, and take the lead in several research projects. This freedom enriched my learning experience and helped me cultivate the confidence and skills necessary to become an independent researcher. His meticulous attention to detail, constructive feedback, and encouragement at every stage have shaped my PhD thesis over the years and broadened my perspective on scientific challenges. I am especially indebted to him for his kindness and support during my difficult times, particularly during the COVID-19 quarantine, the birth of my son, and the Cholecystectomy of my wife. His understanding and flexibility during these personal challenges meant a great deal to me and made my Ph.D. journey more manageable. It has been an absolute privilege to work under his supervision, and I will always cherish the knowledge, skills, and values I have gained from his mentorship.

I would also like to express my heartfelt gratitude to my Supervisory Committee members, Professor Dr. Herb E. Schellhorn and Professor Dr. Russell Bishop, for their time, guidance, and invaluable scientific insights. Their thoughtful discussions and advice during my numerous committee meetings have been instrumental in shaping different aspects of my thesis.

I sincerely appreciate the contributions of both past and present members of Dr. Gupta's Laboratory. I am especially grateful to Dr. Bijendra Khadka for his intellectual support and engaging discussions during his stay in the lab. My heartfelt thanks also go to my colleague Sarah Bello, whose insightful conversations made my time in the lab both productive and enjoyable. Additionally, I am thankful to the undergraduate research students who have been part of Dr. Gupta's lab.

I am sincerely grateful to the faculty and staff of the Department of Biochemistry and Biomedical Sciences at McMaster University for their support throughout my academic journey. I would also like to extend my appreciation to Lisa Kush, Nadia Bhatt, Toluwalase Dayo-Olaide, and Ana Pereira for their invaluable administrative and technical assistance.

Outside the lab, I would like to express my very special thanks to my wife, Bornali Rudra, for taking care of our family, looking after our beloved son Rishik Rudra, and supporting me in every possible way. I have been fortunate to form wonderful friendships in Hamilton with people from Bangladesh, Canada, India, Nepal, and Sri Lanka. I sincerely appreciate their support, the great times we shared, and the unforgettable memories that made my stay memorable. Finally, I want to express my heartfelt gratitude to my mother, Parbati Rudra, for her unwavering love and sacrifices, which gave me the freedom to pursue my dreams. I am also deeply thankful to my parents-in-law, Surit Kumar Rudra and Panu Rani Paul, for their unconditional love and support throughout my studies.

PREFACE

This thesis follows a sandwich format. Chapter 1 provides an introduction, including background information and research objectives. Chapters 2, 3, 4, and 5 consist of unaltered manuscripts published between 2021 and 2025. Chapter 6 is an altered manuscript that will be submitted for publication in the coming months. Chapter 7 summarizes the findings, discusses their significance, and outlines future research directions. References for Chapters 1, 6, and 7 are listed at the end of the thesis. Each chapter includes a preface describing the details of the published and ongoing work and my contributions to co-authored work. All chapters have been reproduced with the consent of the co-authors. Additionally, an irrevocable, non-exclusive license has been granted to McMaster University and the National Library of Canada. Copies of permissions and licenses have been submitted to the School of Graduate Studies.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF ABBREVIATIONS

aa …….. Amino Acid
AAI …… Average Amino Acid Identity
AF ……. Alignment Fraction
ANI …….. Average Nucleotide Identity
BI …… Bayesian Inference
BLAST …… Basic Local Alignment Search Tool
BLASTp …… Protein vs Protein BLAST search
CSI …… Conserved Signature Indel
CSP …… Conserved Signature Protein
Del ……. Deletion
DNA ……Deoxyribonucleic Acid
DDH ……DNA-DNA Hybridization
dDDH ……..digital DNA-DNA Hybridization
DnaK ……..Chaperone DnaK (Hsp60)
E value ……Expect value
GBDP ……Genome BLAST Distance Phylogeny
GC or G+C ……Guanine and Cysteine
GGDC …… Genome to Genome Distance Calculator
GTDB ……Genome Taxonomy Database
GroEL ……Chaperonin GroEL (Hsp70)
HGT ……Horizontal Gene Transfer
HMMs ……Hidden Markov Model
HSP ……Heat Shock Protein
HSPs……High-Scoring Segment Pairs
Hsp60 …… Heat shock protein of the 60 kDa size
Hsp70 ……Heat shock protein of the 70 kDa size
Indel …… Insertion/Deletion
Ins …… Insertion
ICNP ……International Code of Nomenclature of Prokaryotes
LPSN ……List of Prokaryotic Names with Standing in Nomenclature
MAFTT ……Multiple Alignment using Fast Fourier Transform
MEGA ……Molecular Evolutionary Genetics Analysis
ML ……Maximum-Likelihood
MLSA ……Multi-Locus Sequence Analysis
MLST ……Multi-Locus Sequence Typing
MP ……Maximum-Parsimony
MUMi ……Maximal Unique Matches index
MSA ……Multiple Sequence Alignment
MUSCLE …… Multiple Sequence Comparison by Log-Expectation

NCBI …… National Center for Biotechnology Information
NCTC …… The National Collection of Type Cultures
NGS …… Next Generation Sequencing
NJ …… Neighbour-Joining
PCR …… Polymerase Chain Reaction
POCP …… Percentage of Conserved Proteins
QS …… Quorum sensing
RAxML …… Randomized Accelerated Maximum Likelihood
RED ……Relative Evolutionary Divergence
RNA ……Ribonucleic Acid
RpoA …… DNA-dependent RNA polymerase α-subunit
RpoB …… DNA-dependent RNA polymerase β-subunit
SH …… Shimodaira and Hasegawa
SMRT …… Single-Molecule Real-Time
SOLiD …… Sequencing by Oligonucleotide Ligation and Detection
SSU rRNA …… Small subunit ribosomal ribonucleic acid
T3SS ……Type III Secretion System
rRNA …… Ribosomal RNA
tRNA …… .. Transfer RNA
3G …… Third Generation
4G …… Fourth Generation

GLOSSARY

16S ribosomal RNA or 16S rRNA: The small subunit of the 30S ribosomal complex plays a vital role in protein synthesis. It is highly conserved across species and exhibits strong resistance to lateral gene transfer.

Archaea or Archaebacteria: Prokaryotes belong to one of the three domains of life and are distinct from bacteria based on genetic analysis. They lack peptidoglycan in their cell walls and possess unique membrane lipids.

Ancestor: An organism, population, or species from which another organism, population, or species has evolved or descended.

Average Amino Acid Identity: Assesses the degree of similarity in amino acid sequences of proteins across different organisms, providing insight into their evolutionary relationships and aiding in taxonomic classification.

Average Nucleotide Identity: The average percentage of identical nucleotides in gene alignments between two organisms, used in microbiology to assess genome similarity, define species boundaries, and confirm taxonomic identities, especially in prokaryotes.

Bacteria or Eubacteria: One of the three domains of life, consisting of prokaryotic organisms, distinguishable from Archaea by genetic variations and the presence of peptidoglycan in their cell walls.

Bergey's Manual: A commonly used resource in microbiology, particularly for the classification and identification of bacteria and archaea.

Bootstrapping: A statistical method used to evaluate the reliability of a result (typically a phylogenetic tree) by repeatedly sampling data with replacement from the original set.

Clade: A monophyletic group that includes an ancestor and all its descendants.

Comparative Genomics: A branch of biological research that examines genomic features, including gene sequences, proteins, gene organization, and regulatory elements, across different organisms to explore their evolutionary relationships and differences.

Concatenation of Genes: Integrating genetic data into a sequence and analyzing it as a single gene.

Conserved Signature Indel (CSI): An insertion or deletion of a specific size in a particular region of gene or protein sequences unique to a group of interest and absent in other bacterial groups. This molecular change (insertion/deletion) is flanked by conserved residues on both sides, ensuring its reliability.

Convergent Evolution: The process by which distantly related bacterial species independently develop similar traits as a result of adapting to similar environmental pressures.

Duplication: A process by which a molecular sequence is duplicated during evolution.

Core Genome: A term used to describe the essential sets of genes or proteins common to all members of a defined group. These genes are involved in fundamental cellular functions and are typically conserved across individuals within a group.

Degenerate Oligonucleotide Primers: Primers designed to amplify the same region in related organisms. Their sequence covers a range that includes different nucleotide sequences present in the amplification region across different organisms.

Genomic Distance: A measure of divergence between two genomes.

Hidden Markov Model: A statistical tool used to predict sequences of events based on hidden factors. It helps make predictions when the influencing factors are not directly observable.

Homologs or Homologous genes/proteins: Similar Genes or Proteins in different organisms that are evolutionarily related by descent from a common ancestor.

Horizontal Gene Transfer: A process where an organism transfers genetic material to another organism, bypassing the usual inheritance from parent to offspring. This can occur between different species or organisms of the same species.

International Code of Nomenclature of Bacteria or Bacteriological Code: The system or set of rules that governs the scientific naming of Bacteria and Archaea.

Likelihood Ratio Test or SH-Like Test: A test that compares the likelihood of a null model (no specific relationship between organisms) with an alternative model (organisms X and Y are more closely related than X and Z) to assess how well the alternative model fits the data.

Lineage: A line of descent or ancestry, tracing the continuous path of organisms from a common ancestor to its descendants over time.

Long branch attraction: A phenomenon in phylogenetic analysis, particularly in maximum-parsimony methods, where rapidly evolving lineages are mistakenly considered closely related, regardless of their actual evolutionary connections.

Maximum likelihood tree: A phylogenetic tree constructed using the maximum likelihood method, which identifies the tree topology with the highest probability of being generated from the given alignment.

Monophyletic clade: A group of organisms that includes a common ancestor and all of its descendants, with no other organisms outside the group included.

Multilocus Sequence Analysis: Analyzing multiple genes (loci) to understand the phylogeny and evolutionary relationships among organisms.

Multilocus Sequence Typing: Analyzing multiple genes (loci) to identify and classify microorganisms.

Neighbour-joining tree: A method for constructing a phylogenetic tree based on a neighbor-joining approach that clusters organisms using their genetic distance.

Multiple Sequence Alignments (MSA): Used to align three or more sequences (DNA, RNA, or protein) to identify similarities and differences among them, helping to understand the evolutionary relationships.

Orthologs or Orthologous genes/proteins: Sequences from different species that originated from a common ancestral sequence and have evolved separately due to speciation events over time.

Outgroup: A species/group of species identified as the earliest to diverge in a phylogenetic analysis. It is included to establish the root position of the tree.

Paralogs or Paralogous genes/proteins: Genes or proteins originating from a common ancestor within the same organism through gene duplication and may evolve to perform different functions.

Paraphyletic: A group consisting of the group's last common ancestor and some, but not all, of the descendants of that ancestor.

Phylogenetic tree: A branching diagram that illustrates the evolutionary relationships among organisms based on their biological and molecular similarities and differences.

Polyphasic Taxonomy: An Approach incorporating different data types, including phenotypic, genotypic, molecular, and biochemical characteristics, in taxonomy.

Polyphyletic: Refers to a group of organisms that are grouped together in a phylogenetic tree based on similar traits but do not share a common ancestor exclusive to that group.

Protein Family: A group of proteins with a shared evolutionary origin, characterized by similar functions and sequence or structural similarities.

SILVA: A curated database of 16s rRNA gene sequences named "SILVA" after the Latin word silva, meaning forest.

Supermatrix: A concatenated set of all genes/proteins found in the core genome.

Supertree: A consensus phylogenomic tree created by combining the phylogenetic trees of all genes or proteins in the core genome.

Single gene/protein phylogenetic tree: Constructing a phylogenetic tree by comparing homologous sequences of a single gene or protein.

Synapomorphy: Characteristic or trait shared by two or more species and inherited from a common ancestor. This trait helps to identify and define a group of organisms.

Systematics: A branch of biology that focuses on the diversity of organisms. It is typically divided into two areas: phylogeny, the study of evolutionary relationships, and taxonomy, the classification and naming of organisms.

Taxonomic Framework: The system and structure are used to classify and name a group of organisms.

Taxonomic Ranks: The levels in the classification system used to organize and categorize organisms. These ranks include domain, kingdom, phylum, class, order, family, genus, and species, each representing a different level of relatedness and specificity.

Tree topology: The branching pattern in a phylogenetic tree, representing the evolutionary relationships between species or genes.

Valid Publication: A prokaryotic name is deemed validly published if it is included in the Approved Lists of Bacterial Names or published in a Validation List within the International Journal of Systematic and Evolutionary Microbiology or the International Journal of Systematic Bacteriology.

CHAPTER 1

INTRODUCTION AND BACKGROUND STUDY

1.1 A Brief History of Prokaryotic Systematics and Taxonomic Research

"... the sure and definite determination (of each species of bacteria) requires so much time, so much acumen of eye and judgment, so much perseverance and patience that there is hardly anything so difficult"

(Müller and Fabricius, 1786).

The history of life on Earth began approximately 3.8 billion years ago (Schopf, 1978; Schopf and Packer, 1987; Kasting, 1993; Farías-Rico and Mourra-Díaz, 2022). Understanding the origins and complexities of simple microorganisms, such as prokaryotes, has remained a profound enigma in modern science (Winslow et al., 1920; Stanier and Van Niel, 1941; Sagan, 1967; Zotin et al., 1975; Woese and Fox, 1977; Schwartz and Dayhoff, 1978; Gupta, 1998). To understand the evolutionary history of living organisms, scientists have developed a hierarchical system (*viz.* Phylum, Class, Order, Family, Genus, and Species) known as Systematics or Taxonomy, which classifies organisms based on shared characteristics (Linnaeus, 1735; Darwin, 1859). The term "Taxonomy" originates from Ancient Greek, where "táxis" means order or arrangement, and "nomos" means law.

Establishing a reliable taxonomic framework for prokaryotes has been an arduous challenge for taxonomists. In the 18th century, Carl Linnaeus, who introduced the framework of biological taxonomy (the organization of organisms into ranked categories) and nomenclature (the guidelines for assigning names to diverse groups of organisms), set the basis of modern taxonomy (Linnaeus, 1735). Through his seminal work *Systema Naturae*, Carl Linnaeus grouped all microscopic forms as "*Infusoria*" into a single species, which he named *Chaos infusoria* (Linnaeus, 1789). However, there

was minimal progress in microbial classification following Linnaeus's system due to the absence of advanced observational tools and a theoretical framework suited to explaining life at the microscopic level (Ehrenberg, 1838; Haeckel, 1866; Gram, 1884). In 1872, Ferdinand Cohn pioneered the prokaryotic classification by proposing six bacterial genera and grouping them within the plant kingdom, mainly based on their morphological features, growth conditions, and pathogenic potentials (Cohn, 1872; Cohn, 1875). The first edition of Bergey's Manual of Determinative Bacteriology (now known as 'Bergey's Manual of Systematic Bacteriology'), published in 1923, introduced a systematic method for classifying bacteria based on their phenotypic traits, such as morphology, pathogenic characters, and culturing conditions. Bacteria were organized in a hierarchy to show their levels of relatedness, starting from broader groups like Class and Order to narrower ones like Families, Genera, and Species. This edition classified bacteria as ''typically unicellular plants'' or "*Schizomycetes*" (Orla-Jensen, 1909; Bergey DH, 1923).

In the late 19th and early 20th centuries, microbiologists began discovering an increasing variety of prokaryotes that differed widely in shape, physiology, survival strategies, and life cycles. This growing diversity led researchers to incorporate biochemical, physiological, and morphological characteristics into the organism descriptions and classification efforts (Migula, 1900; Orla-Jensen, 1909; Pringsheim, 1923; Buchanan, 1925; Kluyver and Van Niel, 1936; Stanier and Van Niel, 1941). These efforts eventually led to the creation of a universal Code of Bacteriological Nomenclature, which was officially adopted at the 4th International Congress for Microbiology in 1947 (Huddleson, 1947).

However, the biochemical and phenotypic properties used in bacterial classification were of limited use and found to exhibit a high degree of convergence with unrelated organisms (Stanier and van Niel, 1962; Whittaker, 1969; Stanier et al., 1976). Moreover, due to the high diversity in prokaryotic species, simple morphologies, sizes, and sharing of characters through convergent evolution, it was difficult to establish a reliable and clearly demarcating prokaryotic classification system based solely on morphological characteristics (van Niel, 1946; Buchanan and Gibbons, 1974; Sneath et al., 1986). These difficulties in bacterial classification based on phenotypic criteria were discussed widely and acknowledged as "The Dark Age" (Kluyver and Van Niel, 1936; Stanier and Van Niel, 1941; van Niel, 1946; Stanier and van Niel, 1962; Woese, 1992; Gupta, 1998; Sapp, 2006; Oren, 2010). Stanier and Van Niel also highlighted the challenges in defining and classifying bacteria during the 1940s to 1960s. They stated, "…. *Any good biologist finds it intellectually distressing to devote his life to the study of a group that cannot be readily and satisfactorily defined in biological terms, and the abiding intellectual scandal of bacteriology has been the absence of a clear concept of a bacterium*…." (Stanier and van Niel, 1962). These discussions highlighted the need for developing molecular sequence-based, more reliable methods for differentiating and classifying prokaryotes.

1.2 The Genomics Era: Advancing Comparative Genomics and Evolutionary Systematics Studies

The inadequacy of the bacterial classification system that relied on phenotypic and biochemical traits during the mid-20th century gave rise to the emergence of

alternative approaches for prokaryotic classification. Concurrently, the middle of the 20th century witnessed the unveiling of the role of Deoxyribonucleic Acid (DNA) in information transfer and its structural composition (Hershey and Chase, 1952; Watson and Crick, 1953; Crick, 1958; Zuckerkandl and Pauling, 1965; Crick, 1970). This revelation introduced a novel molecular target encompassing all the information governing an organism's phenotypic, physiological, and biochemical characteristics. Thus, one of the earliest nucleic acid-based methods used in the field of prokaryotic taxonomy was DNA-DNA hybridization (DDH) (Hall and Spiegelman, 1961; Schildkraut et al., 1961; McCarthy and Bolton, 1963). In the DDH method, denatured DNA is immobilized in a solid phase (gel of agar), where it is unable to re-nature but can hybridize with free complementary single-strand of DNA. The strength of hybridization between two strands of the DNA duplex is proportional to the similarity of DNA sequences between two organisms. This can be calculated by the dissociation temperature (melting temperature) of the hybridized DNA molecule (McCarthy and Bolton, 1963). The standardized definition of a species and strains that are closely related to it is that they share > 70% DDH, correlated with a hybridized DNA melting point of $<5^0$C (Wayne et al., 1987; Stackebrandt and Goebel, 1994; Rosselló-Mora, 2006; Tindall et al., 2010). However, determining DDH values is an extremely laborious process for which only a few laboratories are properly equipped (Grimont et al., 1980; Huss et al., 1983; Rosselló-Mora, 2006; Goris et al., 2007). Moreover, it is not an accessible method for classifying non-culturable prokaryotes (Rosselló-Mora, 2006; Yarza et al., 2014). In addition, this method provides a rough estimate of genetic relationships, distinguishing only closely related species or subspecies with over 90% genome similarity (Goris et al., 2007; Schleifer, 2009). Lastly, because the DDH method

relies on an experimental approach that uses a specific threshold without producing sequence data, it cannot support the creation of incremental databases that can identify taxa of different ranks (Stackebrandt and Goebel, 1994; Schleifer, 2009; Oren and Garrity, 2014).

In the late 1960s, a major revolution and advancement in the field of taxonomy and systematics came with the advent of determining molecular sequences by the development of experimental, computational, and mathematical methods, offering a new approach for understanding the evolutionary history of genes and organisms (Sanger, 1959; Zuckerkandl and Pauling, 1965; Eck and Dayhoff, 1966; Fitch and Margoliash, 1967). Molecular data soon proved to be a more reliable and objective means of classifying organisms than traditional morphological and biochemical approaches. An important breakthrough came with the work of Zuckerkandl and Pauling (1965), who introduced the idea that molecular sequences serve as historical records of an organism's evolutionary past, allowing for phylogenetic reconstruction. Their insights greatly strengthened the concept of inferring evolutionary relationships through molecular comparisons. Subsequently, Carl Woese and his coworkers introduced the use of 16S ribosomal RNA (rRNA) gene sequence, a component of the prokaryotic small ribosomal subunit or 30S subunit (SSU rRNA) and a universally conserved component of the protein synthesis machinery, for understanding evolutionary relationships (Fox et al., 1977; Woese and Fox, 1977; Woese, 1987). The 16S rRNA sequence contains three structural domains, which are designated as 'U' (Universally conserved), 'S' (semi-conserved), and 'V' (Variable or non-conserved). The 'U' segment is highly conserved and interspersed with variable regions. PCR primers targeting the conserved regions amplify the variable sequences. The 'S'

segment is less conserved than ‘V’ and more restricted in occurrence. ‘V’ segments vary markedly in length, primary sequence, and secondary structure, even within a given lineage, and may be recognizably similar only in very closely related species (Gray et al., 1984). Thus, these conserved and variable regions facilitate the classification of both closely related and highly divergent groups of organisms (Fox et al., 1977; Woese and Fox, 1977; Gray et al., 1984; Woese, 1987; Tindall et al., 2010). Besides these, 16S rRNA offers some notable advantages, making it well-suited for taxonomic and systematic study. Beyond its ubiquity, the 16S rRNA gene is easy to isolate. It is a part of the large ribosomal complex, functionally equivalent and evolutionarily homologous in bacteria, archaea, mitochondria, plastids, and the nucleus, which is unlikely to undergo lateral gene transfer (Olsen et al., 1994; Patel, 2001; Janda and Abbott, 2007). Studies on prokaryotic classification based on 16S rRNA sequences revolutionized bacterial taxonomy, and for the first time, prokaryotes were classified based on their phylogenetic relatedness (Woese and Fox, 1977; Woese, 1987; Woese et al., 1990). The studies by Woese and coworkers resulted in the proposal for the three-domain classification, in which the three domains, Bacteria, Archaea, and Eukaryota, are considered coequal and fundamental divisions of life on earth (Woese et al., 1990). The three-domain model remains the dominant model for biological classification. However, Woese expressed concern that the microbial phylogenetic framework relied too heavily on a single molecule, which he felt was inadequate for accurately representing the complexity of microbial relationships (Woese, 1991).

Over the last 25 to 30 years, 16S rRNA has become the foundation of modern prokaryotic classification (Olsen and Woese, 1993; Garrity et al., 2005; Yarza et al., 2008; Kämpfer, 2012). Bergey’s Manual of Systematics of Archaea and Bacteria

(Whitman et al., 2015), the updated version of Bergey's Manual of Determinative Bacteriology, also adopted the 16S rRNA sequence-based phylogenetic framework for classifying prokaryotic microorganisms. The All-Species Living Tree, which has become the *de facto* tree of life for systematic studies, is also based on the analyses of 16S rRNA gene sequences (Yarza et al., 2008; Yilmaz et al., 2014; Ludwig et al., 2021). A 16S rRNA gene sequence similarity value of >97% is thought to correlate with the 70% DDH threshold for species demarcation (Stackebrandt and Goebel, 1994; Stackebrandt, 2006; Tindall et al., 2010; Meier-Kolthoff et al., 2013). SILVA 16S rRNA database has also suggested thresholds of 94.5%, 86.5%, 82%, 78.5%, and 75% 16S rRNA gene sequence similarity for the demarcation of prokaryotic taxa at the level of Genus, Family, Order, Class, and Phylum, respectively, providing novel guidance for 16S rRNA-based classification (Quast et al., 2013; Yarza et al., 2014).

Despite being considered the gold standard for prokaryotic phylogeny and systematics, the 16S rRNA gene does have some limitations, which include: (a) 16S rRNA is not distinctive at the species level. It has limited capacity to differentiate among closely related species due to high sequence conservation (Fox et al., 1992; Tang et al., 1998; Mignard and Flandrois, 2006; Janda and Abbott, 2007; Reller et al., 2007). (b) It lacks specific biochemical, molecular, or physiological properties distinctive of prokaryotic taxa and other groups (Gupta, 1998; Gao and Gupta, 2012a; Zhi et al., 2012). (c) It also fails to define the branching order and inter-relationships among higher prokaryotic taxa, which is core to understanding the origin of life and its diversification (Gupta, 1998; Gupta and Griffiths, 2002; Yarza et al., 2008; Garrity, 2010; Gupta, 2016) (d) Prokaryotic organisms often have multiple copies of the 16S rRNA gene, differing by 1–2% or more in sequence, which can complicate the accurate inference of

evolutionary relationships (Klappenbach et al., 2001; Boucher et al., 2004; Sun et al., 2013). (e) Moreover, the structure of the 16S rRNA gene is fixed and cannot change freely, causing sudden shifts rather than gradual changes. This can lead to misleading conclusions about the relationships among prokaryotes (Gupta, 1998; Ludwig et al., 1998; Griffiths and Gupta, 2001; Ludwig and Klenk, 2005). Therefore, there was growing interest in identifying and using other genes or proteins to address evolutionary questions that the 16S rRNA gene sequence analysis could not fully resolve (Gupta, 1998; Gupta and Griffiths, 2002).

Since 16S rRNA sequencing cannot reliably distinguish closely related bacterial strains/species, another sequence-based method, Multilocus Sequence Typing (MLST), was introduced to identify clonal relationships among bacteria (Maiden et al., 1998). MLST analyzes DNA sequences from internal regions of multiple housekeeping genes (*viz.*, RNA polymerase β-subunit (*rpoB*), the β-subunit of DNA gyrase (*gyrB*), recombinase A (*recA*), sigma 70 factor (*rpoD*), tRNA modification GTPase ThdF or TrmE (*thdF*), β-subunit of ATP synthase (*atpD*), translation initiation factor IF-2 (*infB*), and chaperonin GroEL (*groEL*) to classify and characterize microbial isolates (Kampfer and Glaeser, 2011; Maiden et al., 2013; Glaeser and Kampfer, 2015; Gomila et al., 2015). It provides greater resolution in characterizing and distinguishing closely related species than 16S rRNA gene-based analysis (Rokas et al., 2003; Jolley et al., 2004; Ciccarelli et al., 2006; Cody et al., 2014).

1.3 Impact of Whole Genome Sequencing in Prokaryotic Classification

Genome sequences provide a new platform in prokaryotic classification and systematics using multiple independent approaches (Gupta, 1998; Danchin, 2003; Coenye et al., 2005; Konstantinidis and Tiedje, 2005a; Klenk and Goker, 2010; Parks et al., 2018). Before 2000, sequencing remained expensive, time-consuming, and limited to a few sequencing centers. The first whole genome sequencing of *Haemophilus influenzae* was done in 1995 (Fleischmann et al., 1995). After 2000, the introduction of high-throughput next-generation sequencing (NGS) technologies, such as 454 parallel pyrosequencing, sequencing by Oligonucleotide Ligation and Detection (SOLiD), Ion Semiconductor sequencing, Illumina dye sequencing, Third-generation (3G) methods, such as Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing approach, (Schadt et al., 2010; van Dijk et al., 2018) and fourth-generation (4G) methods, such as Oxford Nanopore Technologies (Ke et al., 2016; Jain et al., 2018) substantially lowered the cost of sequencing. This development led to a vast increase in the number of whole genome sequences of different organisms. As of December 2024, the NCBI genome database consists of >2.40 M bacterial genomes, out of which >2.03M are annotated (Sayers et al., 2019).

Over the last 15 years, several whole-genome-based methods have been utilized in prokaryotic systematics studies. These methods include Average Nucleotide Identity (ANI), which measures the sequence identity of shared genes and has an established 95-96% sequence identity threshold for species-level demarcation (Konstantinidis and Tiedje, 2005a). The ANI value of 95%-96% is found to be equivalent to 70% DDH or 98.65% 16S rRNA sequence similarity (Konstantinidis and Tiedje, 2005a; Goris et al.,

2007; Richter and Rossello-Mora, 2009; Kim et al., 2014; Varghese et al., 2015). Besides ANI, other approaches such as Maximum Unique Exact Match index (MUEMi) (Deloger et al., 2009) and tetranucleotide regression (Richter and Rossello-Mora, 2009) can also help evaluate a strain's species status. The MUEMi measures genomic distance based on core genome conservation and shared DNA. Tetranucleotide regression looks at differences between observed and expected frequencies of all 256 tetranucleotide combinations (A, T, G, C) (Konstantinidis and Tiedje, 2005b; Thompson et al., 2013; Varghese et al., 2015). Genomes BLAST Distance Phylogeny (GBDP) (Henz et al., 2005) uses BLAST to identify High-Scoring Segment Pairs (HSPs) between genomes, which are then used to calculate distances for constructing phylogenetic trees. This method generates distance matrices to assess genome relatedness and is used mainly for creating evolutionary trees based on whole-genome comparisons. On the other hand, the Genome-to-Genome Distance Calculator (GGDC) (Deloger et al., 2009) calculates genomic distances by comparing whole genomes using BLAST-derived HSPs and expresses genetic relatedness as pairwise percentages. This method is mainly used in microbial taxonomy to assess genomic similarity. Several other methods include Average Amino acid Identity (AAI) (Konstantinidis and Tiedje, 2005b) which measures the amino acid in shared proteins and provides greater resolution for more distant comparisons than ANI (Konstantinidis and Tiedje, 2005a; Thompson et al., 2013), Percentage of Conserved Proteins (POCP) which measure the proportion of proteins/genes shared by two genomes (Qin et al., 2014). Two species are considered members of the same genus if they share >50% POCP values (Qin et al., 2014). However, the AAI and POCP values are limited to defining the genus level and cannot be used for species classification or demarcating higher taxonomic ranks. Additionally,

their values often overlap between ingroup and outgroup species, making it difficult to establish clear genus boundaries (Gupta, 2019; Barco et al., 2020; Rudra and Gupta, 2024).

1.4 Construction of Phylogenetic Trees for the Study of Systematics

Phylogenetic trees depict evolutionary relationships by comparing biological and molecular characteristics, forming the basis of natural classification and guiding systematic research for the past 25 years (Woese et al., 1990; Stackebrandt and Goebel, 1994; Doolittle, 1999; Yilmaz et al., 2014; Parks et al., 2022). Phylogenetic trees can be constructed based on either nucleic acids or protein sequences. Trees based on non-coding sequences, such as rRNA, tRNA, and introns, use only nucleic acid sequences, while those based on coding sequences can use either nucleic acid or protein sequences (Dayhoff et al., 1974; Hasegawa and Hashimoto, 1993; Olsen and Woese, 1993; Hashimoto et al., 1994). However, nucleotide-based analyses are considered less reliable than protein-based phylogenetic analyses, as nucleotide-based analyses can be affected by several factors, including codon biases such as differences in G+C content among lineages and the influence of genetic code degeneracy (Steel et al., 1993; Karlin et al., 1995; Gupta, 1998).

The first step in constructing a sequence-based phylogenetic tree is aligning genomic or proteomic sequences, which organize homologous sequences and serve as the foundation for further analysis. As this is an important step in phylogenetic reconstruction, several Multiple Sequence Alignment (MSA) algorithms have been developed to enhance phylogenetic accuracy while balancing speed and precision.

Notable examples include the Clustal series (e.g., ClustalX and Clustal Omega) (Chenna et al., 2003; Sievers et al., 2011), MUSCLE (Edgar, 2004), MAFFT (Katoh et al., 2005), and T-Coffee (Notredame et al., 2000). Subsequently, the aligned data are processed with different clustering approaches (distance matrix or character-based) such as Neighbor-Joining (NJ) method (Saitou and Nei, 1987), Maximum-Parsimony (MP) (Fitch, 1971), Maximum-Likelihood (ML) (Felsenstein, 1981), and Bayesian Inference (BI) (Rannala and Yang, 1996). In the NJ (Saitou and Nei, 1987), a distance matrix-based method, similar organisms are grouped together based on genetic or genomic distances. In contrast, character-based methods, i.e., MP, ML, and BI, focus on optimizing tree scores. ML aims to identify the tree topology that best explains the observed traits of tip species by maximizing their probability under a given evolutionary model (Felsenstein, 1981). On the contrary, MP selects the tree requiring the fewest evolutionary changes to account for the observed character states of tip species (Fitch and Margoliash, 1967; Fitch, 1971). Bayesian Inference (BI) of phylogeny combines prior knowledge with data likelihood to calculate the posterior probability of trees, representing the probability that a tree is correct based on the data, prior assumptions, and the chosen likelihood model (Yang and Rannala, 2012). The ML method offers notable advantages over others i.e., distance-based or parsimony methods for inferring sequence evolution. Unlike these approaches, ML incorporates a broader range of information from the sequences, including positional variability, transition-to-transversion ratios, and the probability of character states at each position, among other factors (Felsenstein, 1981; 2004). However, one limitation of the ML method is its high computational demand. The accuracy of the relationships shown in the tree is evaluated using statistical tests such as bootstrap and jackknife resampling (Quenouille, 1949;

Efron, 1992) or likelihood ratio tests (Shimodaira and Hasegawa, 1999; Anisimova and Gascuel, 2006).

In the early period, for microbial phylogeny, scientists used ferredoxins and cytochrome protein sequence-based phylogenetic trees (Hall et al., 1973; Saeki et al., 1989). Later, 16S rRNA was used as a phylogenetic marker molecule to construct phylogenetic trees (Woese and Fox, 1977; Woese, 1987). Subsequently, multiple prokaryotic housekeeping genes were used to construct more robust phylogenetic trees, called multi-locus sequence analysis or MLSA (Gupta, 2000; Maiden, 2006; Gao et al., 2009a; Kampfer and Glaeser, 2011; Naushad and Gupta, 2013; Glaeser and Kampfer, 2015). Over the past decade, the availability of genome sequence data has allowed the application of bacterial core genes in phylogenomic analysis extensively (Daubin et al., 2002; Rokas et al., 2003; Ciccarelli et al., 2006; Wu et al., 2009; Gao and Gupta, 2012a; Gao and Gupta, 2012b; Ankenbrand and Keller, 2016; Gupta et al., 2018; Na et al., 2018). These core genes are conserved and present in many bacterial genomes. Studies have demonstrated that phylogenomic analysis using core genes yields more accurate and robust results than traditional methods, which typically rely on a single gene (e.g., 16S rRNA) or a small group of genes (e.g., MLST/MLSA) (Rokas et al., 2003; Jeffroy et al., 2006; Wu et al., 2009). Two main approaches are used to construct genome-based phylogenetic trees. The first method creates individual trees for each gene or protein in the core genome, which are combined into a consensus tree or supertree (Bininda-Emonds, 2004; Puigbo et al., 2009; Lang et al., 2013). This approach is computationally efficient and provides the supertree and individual gene trees for further analysis. The second method aligns all shared genes into a supermatrix, which is used to build a more robust phylogenetic tree (Brown et al., 2001; Snel et al., 2005; Segata et al., 2013; Hug

et al., 2016). This method improves the resolution of relationships and allows statistical techniques, like bootstrap resampling, to validate the tree's structure. The supermatrix method offers notable advantages over the supertree approach, such as enhanced resolution of organismal relationships and the ability to apply statistical methods, including bootstrap resampling and likelihood ratio analysis, for evaluating tree topology (Gadagkar et al., 2005; Lang et al., 2013).

In the present age, the Genome Taxonomy Database (GTDB) (http://gtdb.ecogenomic.org/) is a widely used valuable resource for taxonomic inferences (Parks et al., 2022). It is based on phylogenetic analyses of 120 universally conserved single-copy bacterial marker genes and 122 archaeal marker genes, providing a robust framework for classifying microbial genomes. This online database is an initiative to establish a standardized microbial taxonomy based on genome phylogeny. The genomes utilized for constructing this phylogeny are sourced from RefSeq and GenBank. GTDB utilizes Relative Evolutionary Divergence (RED) values to define taxonomic ranks above the species level. The concept behind RED is that taxa of the same rank should have originated at approximately the same point in evolutionary history (Parks et al., 2022). It also uses the List of Prokaryotic Names with Standing in Nomenclature (LPSN) (lpsn.dsmz.de/) (Parte, 2018) as the primary reference resource to ensure naming priorities and nomenclature consistency. To date (January 2025), GTDB classified 584,382 bacterial genome sequences into 175 Phyla, 538 Classes, 4,870 Families, 23,112 Genera, and 107,235 Species, and 12,477 archaea sequences into 19 Phyla, 64 Classes, 166 Orders, 564 Families, 1,847 Genera, and 5,869 Species.

While widely used in prokaryotic systematics, phylogenetic analysis based on either the 16S rRNA gene, conserved multiple genes, or core genome sequences has several limitations. 1) Phylogenetic branching is influenced by several factors, including Horizontal Gene Transfer (HGT) between/among divergent bacterial species/lineages, which is a very common phenomenon in shaping bacterial genomes (Gupta, 1998; Nelson et al., 1999; Philippe and Douady, 2003; Griffiths and Gupta, 2006a; Bapteste et al., 2009). Additionally, the long-branch attraction effect can result in distantly related species being incorrectly inferred as closely related (Tateno et al., 1994; Gupta, 1998; Philippe et al., 2005; Susko and Roger, 2021). 2) Branching of the species in phylogenetic trees is a continuum; it is often difficult to reliably delimit the boundaries of different clades. Except for the branching of species in the phylogenetic trees, the phylogenetic trees provide no information regarding any genetic, biochemical, or molecular properties that are specific to different taxonomic clades of species (Gupta and Griffiths, 2002; Ludwig, 2005; Gupta, 2016). 3) Phylogenetic trees based on 16S rRNA and other gene or protein sequences often lack the resolution needed to clearly determine the evolutionary relationships and branching order among higher-level prokaryotic groups (Gupta, 2000; Ciccarelli et al., 2006; Wu et al., 2009; Yarza et al., 2010; Segata et al., 2013; Adeolu et al., 2016; Gupta, 2016). Hence, to develop a more informative and reliable prokaryotic classification system, it is important to identify biochemical and molecular properties specific to different groups of organisms. The distinguishing properties or characters inherited from a common parent/ancestor and specific to a group of organisms are the most helpful information for classification purposes and for establishing evolutionary relationships among different organisms

(Woese, 1991; Baldauf and Palmer, 1993; Gupta, 1998; Rokas and Holland, 2000; Gupta and Griffiths, 2002; Bhandari et al., 2012; Gao and Gupta, 2012a; Gupta, 2014).

1.5 Impact of Molecular Signatures in Evolutionary and Taxonomic Studies

The abundance of genomic data offers a valuable resource for identifying molecular markers or synapomorphies shared by evolutionarily related groups of organisms. Two important categories of these molecular markers for understanding microbial phylogeny and systematics, whose discovery has been pioneered by our lab, are Conserved Signature Insertions or Deletions (Indels) (CSIs) in molecular sequences and Conserved Signature Proteins (CSPs) (Gupta, 1998; Griffiths and Gupta, 2006b; Gupta, 2006; Naushad et al., 2014). These molecular markers provide novel and powerful means for the definitive demarcation of different groups of species and aid in understanding their branching order as well as interrelationships (Gupta, 1998; Griffiths and Gupta, 2001; Gao et al., 2006; Bhandari et al., 2013; Gupta, 2016; Hu et al., 2018). The CSIs are amino acid insertions or deletions of fixed lengths, present at a specific position within a conserved sequence region in an evolutionarily related group of organisms. These molecular characteristics result from highly specific genetic changes confined to a monophyletic group of organisms. Because of the rare and highly specific nature of these genetic changes, they are less likely to occur independently in different organisms (Gupta, 1998; Rokas and Holland, 2000; Bhandari et al., 2012; Gao and Gupta, 2012a; Gupta, 2016). The most parsimonious explanation for the occurrence of CSIs that are specific for monophyletic groups of organisms is that the genetic changes leading to the CSIs first occurred in a common ancestor of the group, followed by their

vertical inheritance to all descendants (Adeolu and Gupta, 2013; Naushad and Gupta, 2013; Gupta, 2014; Hu et al., 2019). CSIs are important markers in evolutionary and classification studies for the following reasons. First, they are discrete characters specific to monophyletic groups of organisms, easily detected due to being flanked by conserved regions (Griffiths and Gupta, 2002; Gao and Gupta, 2005). Second, these markers are not influenced by factors, including evolutionary rate differences, compositional biases, or long-branch attraction, that can affect phylogenetic tree accuracy (Gupta, 1998; Rokas and Holland, 2000; Gupta, 2016), and hence they exhibit high degree of predictive ability to be found in other related organisms (Gupta, 2014; 2016; Gupta and Kanter-Eivin, 2023). Finally, although CSIs in both nucleic acids and proteins are informative, most research work on them has focused on protein sequences, where even a single amino acid indel arises from a rare 3 bp insertion or deletion, making these changes in conserved regions most suitable for evolutionary/taxonomy studies (Gupta, 1998; Bhandari et al., 2012; Gao and Gupta, 2012a; Gao and Gupta, 2012b; Gupta, 2016; Hu et al., 2018). Additionally, examining the presence or absence of CSIs in outgroup species makes it possible to determine whether a given CSI is an insertion or deletion in a given group or organisms. However, it is important to acknowledge that the shared occurrence of a CSI within distant organisms in some cases can also result from non-specific processes like HGT or convergent evolution, where similar genetic changes have occurred independently in unrelated lineages (Griffiths and Gupta, 2006a; Gao and Gupta, 2012b; Naushad and Gupta, 2013; Khadka et al., 2020).

CSIs are unique to specific groups of organisms and are flanked by conserved sequences, highlighting their functional importance and suggesting that they are

maintained by strong selective pressure (Gao et al., 2009b; Gupta et al., 2015a; Khadka et al., 2020). The functional importance of CSIs found in bacterial GroEL and DnaK proteins was experimentally demonstrated by Singh and Gupta (Singh and Gupta, 2009). This study showed that CSIs present within the GroEL and DnaK proteins of different bacteria are crucial for their growth, with their removal or alterations in their sequences resulting in failure of cell growth (Singh and Gupta, 2009). Another notable characteristic of the CSIs, which is of much importance for classification purposes, is that they exhibit a high degree of predictive ability to be found in other (uncharacterized or unidentified) members of a given group/taxon (Barbour et al., 2017; Gupta et al., 2020; Gupta and Kanter-Eivin, 2023; Rudra and Gupta, 2024). The work from our lab led by Dr. Gupta, over the past few decades has used CSIs to address several critical issues in microbial phylogeny and systematics (Griffiths and Gupta, 2006b; Naushad et al., 2015; Adeolu et al., 2016; Bello et al., 2022a; Malhotra et al., 2024). Recently, this work has also led to the development of a web-based tool/server (AppIndels.com) (Gupta and Kanter-Eivin, 2023) that uses the information for the presence/absence of known taxon-specific CSIs in a genome sequence to predict the taxonomic affiliation of any submitted genome. The utility of this server for taxonomic purposes was demonstrated by its ability to correctly predict the taxonomic affiliation of 651 uncharacterized *Bacillus* spp. genomes into 29 different genera/families for which CSI information was present in the AppIndels.com database (Gupta and Kanter-Eivin, 2023). My thesis, Chapters 2, 3, 4, 5, and 6, presents the contribution of CSIs in resolving important taxonomic and evolutionary questions within the family *Pseudomonadaceae*.

Another category of molecular markers that are useful for prokaryotic systematics and taxonomic studies is Conserved Signature Proteins (CSPs). CSPs are a group of proteins restricted to a phylogenetically well-defined group of organisms (i.e. monophyletic group) and are introduced during speciation or strain divergence (Gao et al., 2006; Griffiths et al., 2006; Dutilh et al., 2008a; Dutilh et al., 2008b; Bhandari et al., 2012; Naushad et al., 2014; Gupta et al., 2015b). While the mechanisms behind the origin and evolution of these clade/lineage-specific proteins remain unclear (Dutilh et al., 2008b; Kuo and Ochman, 2008), their conserved presence across all or most species/strains within a monophyletic clade and absence in other lineages, suggests that these genes originated in a common ancestor. This ancestor likely passed the genes down vertically to its descendants (Woese et al., 1984; Gao et al., 2006; Gupta, 2006; Gupta and Griffiths, 2006; Dutilh et al., 2008b; Fang et al., 2008; Narra et al., 2008; Gao and Gupta, 2012a). Similar to the CSIs, CSPs are present for species at different phylogenetic/taxonomic depths (Gupta, 2006; Gupta and Mathews, 2010). Several studies indicate that lineage-specific proteins are typically smaller and have a higher prevalence of transmembrane domains than other proteins (Hemm et al., 2008; Knopp et al., 2019). The presence of transmembrane helices and signal peptides indicates they may function as membrane-associated or extracellular proteins, often involved in transport or interactions with cells and the environment (Hassan and Gupta, 2018). In specialized environments, such as those of halophiles, these proteins likely evolved to address specific environmental challenges (Hemm et al., 2008; Hassan and Gupta, 2018; Knopp et al., 2019; Méheust et al., 2022). Due to the lineages/clade/species/strain specificities of CSPs, extensive work has been conducted in our lab on identifying CSPs specific for several taxa and using them to clarify evolutionary relationships and

taxonomy of multiple groups of prokaryotic organisms (Griffiths et al., 2006; Gupta and Griffiths, 2006; Gao et al., 2009b; Gupta and Mathews, 2010; Bhandari et al., 2012; Naushad et al., 2014; Gupta et al., 2015b). In my thesis (Chapter 3), I have also used the CSPs and CSIs to distinguish the strains of *P. paraeruginosa* from those of *P. aeruginosa*.

2. My Research Focus

My graduate research focuses on analyzing the genome sequences of *Pseudomonas* species to robustly elucidate their evolutionary relationships using multiple independent approaches. The approaches that I will use include: (i) Construction of phylogenetic trees based on several large data sets of conserved proteins, and also based on 16S rRNA gene sequences, (ii) Determination of overall genomic similarity between these species based on AAI and POCP matrices, (iii) Comprehensive analyses of protein sequences from *Pseudomonas* species to identify molecular markers such as CSIs and CSPs specific for different *Pseudomonas* species clades supported by other methods. A robust phylogenetic framework for *Pseudomonas* species developed using these approaches will then be used to reclassify *Pseudomonas* into different novel taxa (genera) so that species from all proposed taxa are evolutionarily related, sharing multiple unique molecular characteristics.

Although the focus of my thesis work is on *Pseudomonas* species, during my Ph.D. work, I also used similar approaches to clarify the evolutionary relationships amongst other groups of prokaryotic organisms. A listing of these studies (published) is provided in Appendix A.

3. Introduction to the Genus *Pseudomonas*

3.1. Background Information

Genus *Pseudomonas* (Migula, 1894), which is one of the earliest and best-studied bacterial genera, consists of a large assemblage of motile, rod-shaped, aerobic, non-spore-forming, Gram-negative bacteria (Palleroni, 2005; 2015). The term *Pseudomonas* was first introduced by the German botanist Walter Migula, deriving from the Greek words “pseudo” (false) and “monas” (unit). Although the exact reasoning behind the name was not clarified, it is suggested that Migula chose it due to the bacteria's resemblance in size and motility to the non-flagellate cells of Monas (D'Agata, 2015; Parte et al., 2020). *P. aeruginosa* is the type species of the genus *Pseudomonas* (Schroeter, 1872; Migula, 1894), which serves as the type genus of the family *Pseudomonadaceae*, encompassing several other genera such as *Atopomonas, Azomonas, Azorhizophilus, Azotobacter, Entomomonas, Mesophilobacter, Oblitimonas, Permianibacter, Rhizobacter, Rugamonas, Stutzerimonas,* and *Thiopseudomonas* (Parte et al., 2020). The phylum name *Pseudomonadota* has recently been derived from the genus *Pseudomonas* (Parte et al., 2020; Oren and Garrity, 2022).

With over 300 validly published species, *Pseudomonas* exhibits remarkable genetic and metabolic diversity (Parte et al., 2020). This diversity enables its species to adapt to diverse environments, including soil, air, water, plants, and animal tissues (Schroth et al., 2006; Morris et al., 2008; Lister et al., 2009; Peix et al., 2009; Kidd et al., 2011; Silby et al., 2011; Scales et al., 2014; Palleroni, 2015). Some species from this genus function as opportunistic pathogens in humans, animals, and plants, while others play crucial roles in economic and ecological processes (Palleroni, 2005; Lund-Palau et

al., 2016; Winsor et al., 2016; Xin et al., 2018; Rossi et al., 2021). *P. aeruginosa* is one of the best-studied opportunistic human pathogens capable of causing a wide array of life-threatening acute and chronic diseases, including cystic fibrosis, nosocomial infections, eye and ear infections, and multiple sepsis syndromes (Stover et al., 2000; Planquette et al., 2013; Lund-Palau et al., 2016; Freschi et al., 2018; Spagnolo et al., 2021; Qin et al., 2022). Besides infecting humans, it is also known to cause diseases in livestock and companion animals, such as urinary tract infections in dogs (Harada et al., 2012; Haenni et al., 2015), mastitis in dairy cows (Osborne et al., 1981; Silby et al., 2011; Banerjee et al., 2022), and endometritis in horses (Hughes et al., 1966; Kidd et al., 2011). It exhibits intrinsic resistance to multiple antibiotics and has a strong ability to develop new resistance mechanisms (Lister et al., 2009; Planquette et al., 2013; Moradali et al., 2017; Pang et al., 2019; Qin et al., 2022). Although less virulent than *P. aeruginosa*, another species, *P. fluorescens,* can cause opportunistic acute infections in humans and has been identified in clinical samples from the mouth, stomach, lungs, and bloodstream (Scales et al., 2014; Liu et al., 2021a; Wu and Jing, 2024). Another important species, *P. syringae*, ranks first among the top 10 plant pathogenic bacteria and is responsible for diseases like blast and black pit in diverse plants, including citrus (Lelliott et al., 1966; Klingner et al., 1976; Preston, 2000; Xin et al., 2018). The species comprises around 50 pathovars, many targeting different hosts (Gardan et al., 1992; Schroth et al., 2006). *P. syringae* also plays a role in the water cycle by acting as an ice nucleus in clouds and has been identified in rain, snow, lakes, and plants (Hirano and Upper, 2000; Morris et al., 2008). Other phytopathogenic *Pseudomonas* species include *P. amygdali, P. avellanae, P. cannabina, P. caricapapayae, P. ficuserectae, P. meliae, P.*

savastanoi, P. tremae, and *P. viridiflava* (Lelliott et al., 1966; Gardan et al., 2002; Lopez et al., 2012; Beiki et al., 2016).

On the other hand, different *Pseudomonas* species, including *P. fluorescens, P. mendocina, P. putida, P. stutzeri, P. syringae,* etc., play key roles in biotechnology, contributing to plant growth promotion, bioremediation agents, detectors of food spoilage agents in milk, and the production of valuable secondary metabolites (Scales et al., 2014; Madhaiyan et al., 2017; Vasconcellos et al., 2017; Hassen et al., 2018; Lee et al., 2022; Mora et al., 2022; Mehmood et al., 2023). *P. fluorescens* benefits plant growth and is commonly used in agriculture to enhance sustainability and control plant diseases (Panpatte et al., 2016; Garrido-Sanz et al., 2017; David et al., 2018; Raio, 2024). *P. putida* is widely used as an industrial biocatalyst (Loeschcke and Thies, 2015; Nikel and de Lorenzo, 2018; Weimer et al., 2020). *P. chlororaphis* has demonstrated its potential as a biocontrol agent for managing peanut stem rot disease (Johnsson et al., 1998; Garrido-Sanz et al., 2017; Liu et al., 2022), *P. pertucinogena* and related species contain genes for the production of enzymes, including esterases, dehalogenases, and transaminases, as well as secondary metabolites (Bollinger et al., 2020; Kruse et al., 2024). Their widespread presence and importance in ecology have resulted in a consistent annual increase in identified *Pseudomonas* species. As many environments remain unexplored, this trend is expected to continue, with new species and strains being discovered rapidly. As a result, this genus has become one of the fastest-growing bacterial groups (Girard et al., 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022; Mulet et al., 2024).

However, the inclusion of genetically and phenotypically unrelated species makes the genus *Pseudomonas* highly complex and polyphyletic, as not all species share a common evolutionary history (Hesse et al., 2018; Peix et al., 2018; Girard et al., 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022). Although all these species bear the genus name *Pseudomonas*, they exhibit substantial variation, which contradicts the fundamental principles of microbial taxonomy. In prokaryotic taxonomy, the genus and species are the primary classification units, and a genus name typically implies that species within it are closely related, sharing common genetic, phenotypic, and functional traits, such as pathogenic potential. These shared characteristics distinguish them from species in other genera (Rosselló-Mora and Amann, 2001; Gupta et al., 2018; Gupta et al., 2020; Gupta, 2021; Lalucat et al., 2022). In contrast, the current *Pseudomonas* classification groups genetically unrelated species, many of which are distantly related to the type species *P. aeruginosa* (Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020). This system lacks a clear phylogenetic framework for understanding the relationships among species. It also fails to organize the species into taxonomic units based on their evolutionary history and relationships. The absence of a reliable and informative phylogenetic framework for *Pseudomonas* species hinders understanding the roles of these species in disease causation in animals and plants, their ecological importance, and the production of metabolites and enzymes, which are helpful for diverse purposes. As a result, there is a growing consensus that the *Pseudomonas* classification system should be revised to more accurately reflect the evolutionary relationships among these species based on their shared history and other commonly shared traits (Hesse et al., 2018; Saati-Santamaria et al., 2021; Lalucat et al., 2022).

3.2 Taxonomic History of the Genus *Pseudomonas* and Research Gaps:

Understanding the evolutionary relationships among *Pseudomonas* species and developing a reliable taxonomic framework for this genus has been a persistent challenge, primarily due to the absence of consistent phenotypic or genotypic traits shared by all its species (Anzai et al., 1997; Mulet et al., 2010; Garrido-Sanz et al., 2016; Peix et al., 2018; Lalucat et al., 2022). Initially described by Migula in 1894 (Migula, 1894), the genus includes some species distinguishable by specific features, such as producing a green, fluorescent pigment and being polarly flagellated, strictly oxidative, Gram-negative rods (Cowan and Liston, 1974). However, many *Pseudomonas* species lack these traits, and similar characteristics are observed in other genera, making classification difficult. By the 1960s, *Pseudomonas* taxonomy had become highly disorganized, with nearly 800 species names in use (Stanier and van Niel, 1962; Palleroni, 2010). To address this issue, Stanier and colleagues introduced a systematic approach in 1966 (Stanier et al., 1966), utilizing biochemical and substrate utilization tests initially developed by Den Dooren de Jong (1926). This extensive analysis examined 165 phenotypic traits across 401 recognized *Pseudomonas* strains. Building on this, Stanier, Doudoroff and Palleroni (Doudoroff, 1974) advanced *Pseudomonas* taxonomy by incorporating polyphasic methods, including analyses of G+C content and DNA-DNA hybridization. In addition, Palleroni et al. (Palleroni, 1984) divided Pseudomonads into five rRNA subgroups based on RNA-DNA hybridization. They identified Group I, represented by *P. aeruginosa*, as the core of the genus, while the remaining groups were reclassified into separate genera within the same or related families.

An important breakthrough in *Pseudomonas* taxonomy occurred in the 1980s when Woese and colleagues introduced the use of 16S ribosomal RNA gene sequencing for bacterial classification. This work placed *Pseudomonas* within the *Gammaproteobacteria* (Woese et al., 1984). The most pointed changes impacting the classification of the genus *Pseudomonas* emerged in the 2000s, starting with a study by Anzai et al., (2000). This research analyzed the 16S rRNA gene sequences of 128 *Pseudomonas* species, revealing that many species fell outside the *Pseudomonas sensu stricto* cluster, which corresponds to the rRNA group I defined by Palleroni (1984). However, the limited resolution of 16S rRNA at the species level necessitated using alternative markers, such as concatenated sequences of multiple housekeeping genes, for more precise classification. Over the last two decades, phylogenomic techniques, including MLSA based on multiple housekeeping genes, i.e., 16S rRNA, *gyrB, rpoD*, and *rpoB*, and core genome comparisons, have become essential tools in studying *Pseudomonas* taxonomy (Yamamoto et al., 2000; Hilario et al., 2004; Mulet et al., 2010; Hesse et al., 2018; Lalucat et al., 2022).

Extensive research using MLSA and core genome analysis has been undertaken to understand the evolutionary relationships among *Pseudomonas* species (Mulet et al., 2010; Jun et al., 2016; Hesse et al., 2018; Peix et al., 2018; Saati-Santamaria et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). The findings from these studies have consistently identified several unresolved issues within the genus. 1) *Pseudomonas* species do not form a monophyletic lineage and group into three unrelated lineages: the Aeruginosa, Fluorescens, and Pertucinogena lineages (each based on representative species names), suggesting they do not share a common ancestor. 2) The type species of *Pseudomonas*, *P. aeruginosa*, along with a few others, form a clade known as the

'Aeruginosa clade,' which is distinct from the rest of the *Pseudomonas* species clades. 3) Within these two lineages (Aeruginosa and Fluorescens), several distinct genus-level groups or clades were observed which includes Alcaligenes, Anguilliseptica, Flexibilis, Fluorescens, Kuykendallii, Linyingensis, Lutea, Massiliensis, Oleovorans, Oryzihabitans, Putida, Resinovorans, Rhizosphaerae, Straminea, Stutzeri, and Syringae clades consisting different *Pseudomonas* species. Each clade is named after the representative species name. 4) A deep-branching lineage, the Pertucinogena lineage (named for *P. pertucinogena*), is observed, which branches outside of all *Pseudomonas* species/clades. 5) Several studies also showed that species from other genera within the *Pseudomonadaceae* family, such as *Azotobacter, Azomonas*, and *Chryseomonas*, branch in between different *Pseudomonas* species clades, making the genus polyphyletic. Furthermore, some *Pseudomonas* species, i.e., *P. acidophila, P. cissicola*, and *P. geniculata*, are more closely related to several distant genera, including *Paraburkholderia, Xanthomonas*, and *Stenotrophomonas*, respectively. 6) No taxonomic, biochemical, or molecular markers that clearly distinguish all genus-level clades within the genus *Pseudomonas* have been identified. For taxonomic markers to be effective, they must be stable and shared by all members of the clades/genera. In addition, the markers should possess strong predictive power, enabling the determination of the taxonomic affiliation of uncharacterized strains based on their presence or absence.

Thus, the challenges mentioned above highlight the need to revise the *Pseudomonas* classification by developing a reliable, informative marker-based taxonomic framework that reflects the evolutionary relationships among *Pseudomonas* species and resolves the existing taxonomic issues within the genus *Pseudomonas*. In my thesis, I have used

multiple genome sequence-based approaches, *viz.*, construction of phylogenomic trees based on large datasets of core proteins, identification of molecular markers such as CSIs and CSPs, which are specific for different *Pseudomonas* species clades/genera and other whole-genome-based approaches (*viz.,* construction of AAI and POCP matrices) to address the challenges related to understanding evolutionary relationships and re-classification of *Pseudomonas* species in different thesis chapters.

In addition to genus-level reclassification, my research also focuses on species or strain-level demarcation. I used molecular marker-based approaches to propose a new species, *P. paraeruginosa*, distinct from *P. aeruginosa* strains.

P. aeruginosa, which is a part of the "ESKAPE" group of pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* species), poses a serious threat to human health and is prioritized for novel antimicrobial development (Pendleton et al., 2013; Miller and Arias, 2024). *P. aeruginosa* drives its pathogenicity through numerous virulence factors, either located on the cell surface or secreted into the surroundings (Lyczak et al., 2000; Wolfgang et al., 2003; Hauser, 2009; Rutherford and Bassler, 2012; Elsen et al., 2014; Garcia-Reyes et al., 2020). One of the key virulence factors is Type IV pili (T4P) (Hood et al., 2010; Burrows, 2012; Basso et al., 2017), which are responsible for twitching motility. This type of motility plays a crucial role in biofilm formation and surface exploration (Burrows, 2012). Other key virulence factors in *P. aeruginosa* include the Type I (T1SS) (Filloux, 2011; Qin et al., 2022), Type II (T2SS) (Jyot et al., 2011), and Type III (T3SS) secretion systems (Yahr et al., 1996; Engel and Balachandran, 2009; Nadal Jimenez et al., 2012; Elsen et al., 2014; Toska et al., 2014).

T1SS secretes alkaline protease, which inhibits fibrin formation and aids bacterial spread (Filloux, 2011; Qin et al., 2022). T2SS releases exotoxin A, phospholipase C, protease IV, and elastase, which contribute to cytotoxicity, inflammation, and colonization (Jyot et al., 2011; Wiener-Kronish and Pittet, 2011). The Type 3 secretion system (T3SS) injects exotoxins directly into host cells. These include ExoU (a phospholipase causing apoptosis and necrosis), ExoY (an adenylate cyclase disrupting endothelial cell function), and ExoT and ExoS (bifunctional proteins that impair DNA synthesis and alter cell morphology). While all strains carry T3SS genes, only some secrete these effectors, and T3SS expression is linked to worse clinical outcomes (El-Solh et al., 2012; Ledizet et al., 2012; Elsen et al., 2014; Toska et al., 2014; Qin et al., 2022). Several other virulence factors include quorum sensing (QS), which controls cell communication and biofilm formation, and endotoxin (lipopolysaccharide), located on the outer membrane, providing resistance to host defenses (Ramsey et al., 2005; Rutherford and Bassler, 2012).

Studies using phylogenomic, computational, and experimental approaches have shown that *P. aeruginosa* strains are pathogenically distinct, grouping into two clades in the phylogenetic analysis. These clades are called “Classical clades” and “Outlier clades”. Based on the presence/absence of the T3SS, the more pathogenic “Classical” clade, containing T3SS, is represented by strain PAO1, while the less pathogenic “Outlier” clade, which lacks T3SS, is represented by strain PA7 (Roy et al., 2010; Sood et al., 2019; Sood et al., 2020). We also performed extensive phylogenomic and comparative genomic analyses to distinguish the strains of these two clades. Based on our findings, we reclassified the strains from the “Outlier clade” as a new species, *Pseudomonas paraeruginosa*. This work is detailed in Chapter 4 of this thesis.

3.3 Thesis Chapters Overview

As previously mentioned, *Pseudomonas* species are categorized into three main groups or lineages: Pertucinogena, Aeruginosa, and Fluorescens. My thesis is structured into seven chapters, each addressing key aspects of my research to clarify the evolutionary relationships among *Pseudomonas* species from these lineages using phylogenomic and molecular signature-based approaches. Chapter 1 introduces the genus *Pseudomonas*, highlights current research challenges, and lays the foundation for the studies presented in subsequent chapters. Chapters 2-5 include my published studies, presented in their original manuscript form. Chapter 2 focuses on the Pertucinogena group of species, which branches outside of all *Pseudomonas* species clades, underscoring that they are separate from them. Our analyses of these *Pseudomonas* species have led to the reclassification of most of the species from the Pertucinogena group into two new genera, *Halopseudomonas* and *Atopomonas*. Furthermore, we have also merged the genus *Oblitimonas* with *Thiopseudomonas* and reclassified several misclassified *Pseudomonas* species into their respective genera. Chapter 3 focuses on the Aeruginosa lineage or group of species. Aeruginosa lineage/group consists of over 13 distinct *Pseudomonas* species clades in addition to *Azomonas* and *Azotobacter* species. This chapter highlights the delineation of these clades through comprehensive phylogenomic analyses, the identification of CSIs, and the application of additional comparative genomic methods such as AAI and POCP. Based on the results from these analyses, the genus *Pseudomonas* was redefined to include only species within the Aeruginosa clade, containing the type species *P. aeruginosa*. Twelve novel and emended genera were described to represent the remaining clades within the Aeruginosa lineage. Chapter 4 investigates many *P.*

aeruginosa strains and identifies two distinct clades containing *P. aeruginosa* strains: Classical and Outlier. Through phenotypic and genotypic analyses, we established that the Outlier clade represents a new species, which we named *P. paraeruginosa*. Chapter 5 highlights the predictive potential of CSIs as a reliable tool for determining the taxonomic affiliations of ~ 300 unclassified *Pseudomonas* strains. This chapter underscores the utility of CSIs in refining microbial taxonomy. Chapter 6 presents the results of phylogenomic studies to clarify the evolutionary relationships within species from the Fluorescens lineage. It also reports the identification of CSIs specific for some of the observed clades. This study, after further work, will form the basis for reclassifying species from the Fluorescens lineage into different novel genera. Finally, Chapter 7 concludes by summarizing the overall significance of my research and discussing potential future directions for advancing our understanding of prokaryotic taxonomy, classification, and evolution.

CHAPTER 2

Phylogenomic and Comparative Genomic Analyses of Species of the Family *Pseudomonadaceae*: Proposals for the Genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., Merger of the Genus *Oblitimonas* with the Genus *Thiopseudomonas*, and Transfer of some Misclassified Species of the Genus *Pseudomonas* into Other Genera.

This chapter describes phylogenomic and comparative genomics approaches to clarify the evolutionary relationships of the Pertucinogena lineage of species, which groups outside of all other *Pseudomonas* species clades. Comparative analysis of protein sequences across *Pseudomonas* species led to the identification of CSIs unique to the Pertucinogena group of species, along with other misclassified *Pseudomonas* species clades. These CSIs, in conjunction with phylogenetic analyses, provide insights into the evolutionary relationships among different *Pseudomonas* species, leading to the establishment of two novel genera, *Halopseudomonas* and *Atopomonas*. Furthermore, this study presents molecular evidence for the misclassification of several *Pseudomonas* species, supporting their reassignment to more closely related genera. My contributions to this chapter include the construction of phylogenetic trees, the identification of CSIs, drafting and revising the manuscript, and producing all main and supplemental figures and tables.

Due to space constraints, supplementary figures and tables are not included in this chapter but are available alongside the entire manuscript at:

Rudra B., Gupta R. S. (2021). Int. J. Syst. Evol. Microbiol.71(9):005011.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC NOTE
Rudra and Gupta, *Int. J. Syst. Evol. Microbiol.* 2021;71:005011
DOI 10.1099/ijsem.0.005011

MICROBIOLOGY SOCIETY

Phylogenomic and comparative genomic analyses of species of the family *Pseudomonadaceae*: Proposals for the genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., merger of the genus *Oblitimonas* with the genus *Thiopseudomonas*, and transfer of some misclassified species of the genus *Pseudomonas* into other genera

Bashudev Rudra and Radhey S. Gupta*

Abstract

The evolutionary relationships among species of the family *Pseudomonadaceae* were examined based on 255 available genomes representing >85% of the species from this family. In a phylogenetic tree based on concatenated sequences of 118 core proteins, most species of the genus *Pseudomonas* grouped within one large cluster which also included members of the genera *Azotobacter* and *Azomonas*. Within this large cluster 18–30 clades/subclades of species of the genus *Pseudomonas* consisting of between 1 and 36 species, were observed. However, a number of species of the genus *Pseudomonas* branched outside of this main cluster and were interspersed among other genera of the family *Pseudomonadaceae*. This included a strongly supported clade (Pertucinogena clade) consisting of 19 mainly halotolerant species. The distinctness of this clade from all other members of the family *Pseudomonadaceae* is strongly supported by 24 conserved signature indels (CSIs) in diverse proteins that are exclusively found in all members of this clade. Nine uncharacterized members of the genus *Pseudomonas* also shared these CSIs and they branched within the Pertucinogena clade, indicating their affiliation to this clade. On the basis of the strong evidence supporting the distinctness of the Pertucinogena clade, we are proposing transfer of species from this clade into a novel genus *Halopseudomonas* gen. nov. *Pseudomonas caeni* also branches outside of the main cluster and groups reliably with *Oblitimonas alkaliphila* and *Thiopseudomonas denitrificans*. Six identified CSIs are uniquely shared by these three species and we are proposing their integration into the emended genus *Thiopseudomonas*, which has priority over the name *Oblitimonas*. We are also proposing transfer of the deep-branching *Pseudomonas hussainii*, for which 22 exclusive CSIs have been identified, into the genus *Atopomonas* gen. nov. Lastly, we present strong evidence that the species *Pseudomonas cissicola* and *Pseudomonas geniculata* are misclassified into the genus *Pseudomonas* and that they are specifically related to the genera *Xanthomonas* and *Stenotrophomonas*, respectively. In addition, we are also reclassifying '*Pseudomonas acidophila*' as *Paraburkholderia acidicola* sp. nov. (Type strain: G-6302=ATCC 31363=BCRC 13035).

INTRODUCTION

The genus *Pseudomonas* first described by Migula in 1894 [1], consists of a group of rod-shaped, aerobic, non-spore forming, Gram-negative bacteria generally possessing some polar flagella that assist in their movement [2, 3]. *Pseudomonas* is the type genus of the family *Pseudomonadaceae*, which harbours several other genera, including *Azomonas*, *Azorhizophilus*, *Azotobacter*, *Entomomonas*, *Mesophilobacter*, *Oblitimonas*, *Permianibacter*, *Rhizobacter*, *Rugamonas* and *Thiopseudomonas* [4–6]. The genus *Pseudomonas*, which

Author affiliations: [1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada.
***Correspondence:** Radhey S. Gupta, gupta@mcmaster.ca
Keywords: *Pseudomonadaceae*; *Pseudomonas*; Phylogenomic and comparative genomic analyses; Conserved signature indels (CSIs); Pertucinogena clade; Caeni-Thiopseudomanas clade; *Pseudomonas hussainii*; *Pseudomonas cissicola*; *Pseudomonas geniculata*; '*Pseudomonas acidophila*'; classification.
Abbreviations: CSIs, conserved signature indels; GTDB, Genome Taxonomy Database; MLSA, multilocus sequence analysis.
The accession numbers for the 16S rRNA gene sequence and the draft genome sequence for the type strain G-6302 of *Paraburkholderia acidicola* are MW628182 and ASM236231v1 (GCF_002362315.1), respectively.
A supplementary table and 53 supplementary figures are available with the online version of this article.

contains more than 240 species with validly published names, is the largest genus within the family *Pseudomonadaceae* and its members span enormous genetic and metabolic diversity and inhabit a wide variety of environments including soil, water, animal and plant tissues [2, 3, 7, 8]. However, several species of the genus *Pseudomonas* are opportunistic pathogens of humans, animals and plants and these are often the most studied [2, 8]. The best studied species within this genus include *Pseudomonas aeruginosa*, an opportunistic human pathogen [9–11], *Pseudomonas syringae*, a plant pathogen [12], *Pseudomonas putida*, a soil bacterium and *Pseudomonas fluorescens,* a plant growth-promoting bacterium [2]. The ubiquitous nature of these organisms is leading to a steady increase in the number of species of the genus *Pseudomonas* every year [8, 13, 14], and as many environments remain to be explored, this trend is expected to continue [8, 15, 16].

Over the years, extensive work has been carried out on clarifying the classification and evolutionary relationships among species of the genus *Pseudomonas* on the basis of different characteristics [2, 3, 17, 18]. While the earlier studies in this regard were based on physiological, biochemical and chemotaxonomic features [3, 7, 19], subsequent studies have utilized phylogenetic analysis based on 16S rRNA gene sequences [17] and multilocus sequence analysis (MLSA) based on concatenated sequences of several conserved genes and/or proteins, such as 16S rRNA, *gyrB, rpoD* and *rpoB* [3, 8, 16, 20–22]. Although these studies differ in terms of the number of species that were analysed, and also have used different evolutionary models for phylogenetic tree reconstructions, in most of these studies the examined species of the genus *Pseudomonas* group into two or three main clusters or lineages. The two main clusters observed in these studies are referred to as the *P. fluorescens* lineage and *P. aeruginosa* lineage [3, 8, 16, 20]. Within these two main lineages, a number of phylogenetic groups or clades named after the following species *P. fluorescens*, *P. lutea*, *P. syringae*, *P. rhizosphaerae*, *P. putida*, *P. anguilliseptica*, *P. straminea*, *P. luteola*, *P. oryzihabitans*, *P. stutzeri*, *P. oleovorans* and *P. aeruginosa* have been observed in different studies [3, 8, 16, 20–23]. However, the numbers of these groups as well as their branching positions and species composition often vary in different studies [3, 8, 16, 20–22]. Additionally, besides these two main lineages, another deep-branching lineage, referred to as the Pertucinogena clade (based on its containing *P. pertucinogena*) is also observed in most studies [8, 16]. In an important advancement towards understanding the evolutionary relationships among species of the genus *Pseudomonas*, Hesse *et al.* [24] have sequenced the genomes of 118 type strains of species of the genus *Pseudomonas*. Using these genomes and other available genome sequences, these authors reconstructed a phylogenomic tree for 163 species of the genus *Pseudomonas* based on concatenated sequences of 100 single-copy genes [24]. The highly-resolved tree reconstructed in this study also confirmed the existence of at least 13 distinct clades of species of the genus *Pseudomonas* (generally similar to those identified by MLSA) and their work also indicated that the known species of the genus *Pseudomonas* do not adequately represent the overall genetic diversity of this genus [24].

Most of the studies on examining the evolutionary relationships among species of the family *Pseudomonadaceae* are limited to members of the genus *Pseudomonas* and they do not include species from other genera of the family *Pseudomonadaceae*. However, it is known from earlier work that the species from a number of genera of the family *Pseudomonadaceae*, such as *Azotobacter* and *Azomonas,* branch in between the species of the genus *Pseudomonas* making this genus polyphyletic [23, 25–29]. Furthermore, earlier work indicates that a number of other species of the genus *Pseudomonas* such as '*P. acidophila*', *P. cissicola* and *P. geniculata,* are more closely related to other distantly related genera (*Paraburkholderia, Xanthomonas* and *Stenotrophomonas*) [17, 30–32] than to the genus *Pseudomonas*, but they are still placed within the genus *Pseudomonas* [4, 6] due to lack of reliable evidence to reclassify them. In view of the aforementioned taxonomic anomalies, it is important to carry out further detailed phylogenomic and comparative genomic studies on species of the family *Pseudomonadaceae* to reliably discern their evolutionary relationships.

Due to rapid advancements in genome sequencing technology, genome sequences are now available for more than 240 species of the genus *Pseudomonas* in the NCBI genome sequence database [14] (https://www.ncbi.nlm.nih.gov/genome/). In addition, genome sequences are also available for a number of species from other genera of the family *Pseudomonadaceae* providing a valuable resource for undertaking detailed studies to clarify the evolutionary relationships among species of the genus *Pseudomonas* and the family *Pseudomonadaceae*. Using these genome sequences, we have reconstructed a highly-resolved phylogenetic tree based on concatenated sequences of 118 conserved proteins. This tree provides a phylogenetic framework for understanding the evolutionary relationships among the members of the family *Pseudomonadaceae* and the genus *Pseudomonas*. In addition, we also describe the results of our comparative genomic analysis of protein sequences which have identified multiple molecular signatures, consisting of conserved signature indels (CSIs), which are distinctive characteristics of a number of clades of species of the family *Pseudomonadaceae*, enabling their reliable demarcation in molecular terms. One of these strongly supported clades, referred to in earlier work as the Pertucinogena clade [8, 16, 20, 24], branches deeply in phylogenetic trees, and it is comprised of halophilic and/or halotolerant species. We are proposing the transfer of species from this distinct clade into a novel genus *Halopseudomonas* gen. nov. We also present reliable evidence that the species *P. caeni, Oblitimonas alkaliphila* and *Thiopseudomonas denitrificans* form a monophyletic grouping and they should be integrated into the genus *Thiopseudomonas*. Evidence presented here also supports the placement of deep branching *P. hussainii* into a novel genus. Lastly, the results from our phylogenetic studies and molecular signatures robustly establish that the species *P. cissicola, P. geniculata* and '*P. acidophila*' are specifically related to the species from the genera *Xanthomonas*, *Stenotrophomonas* and *Paraburkholderia*, respectively, and they should be reclassified into these genera.

METHODS

Reconstruction of phylogenetic trees and genomic analysis of the genus *Pseudomonas*

Genome sequences for different named species of the family *Pseudomonadaceae* were downloaded from the NCBI genome sequence database. These included 243 sequences for named species of the genus *Pseudomonas* and 12 sequences for species from other genera of the family *Pseudomonadaceae* including the genera *Azotobacter, Azomonas, Entomomonas, Oblitimonas, Permianibacter, Rugamonas* and *Thiopseudomonas*. The genome sequences for *Moraxella bovoculi* and *Moraxella bovis* were included in the dataset for rooting the trees. Each species in our dataset is represented by a single genomic sequence, generally of the type strain when available. On the basis of these genome sequences, a rooted phylogenetic tree was reconstructed based on concatenated sequences of 118 conserved proteins. The proteins used for this tree reconstruction are a part of the phyloeco set for the class *Gammaproteobacteria* and they are single-copy genes, which, based on analyses of a limited number of genomes, were indicated as being widely distributed within the class *Gammaproteobacteria* [33]. The names and accession numbers of the proteins which were used for tree reconstruction are provided in Table S1 (available in the online version of this article), Reconstruction of the phylogenetic trees was carried out as described in our earlier work [34, 35]. Briefly, using the profile Hidden Markov Models of different proteins from the phyloeco set, the members of these protein families were identified in the input genomes using HMMer 3.1 [36]. On the basis of the results of these analyses only those protein families where the proteins share a minimum of 50% in sequence identity and sequence length and where the protein is found in at least 80% of the input genomes were retained for phylogenetic tree reconstruction. Multiple sequence alignments of these protein families were generated using the Clustal Omega [37] algorithm. TrimAl [38] was used to remove poorly aligned regions from the sequence alignments before concatenation of the sequences into a single file. The concatenated sequence alignment used for phylogenetic analysis contained a total of 41417 aligned positions. A maximum likelihood (ML) tree based on this alignment was reconstructed using the Whelan and Goldman model [39] of protein sequence evolution in FastTree 2 [40]. Optimization of the robustness of the trees was completed by conducting SH tests [41] in RAxML 8 [42] and the trees were drawn using MEGA X [43].

In addition to this comprehensive tree, another phylogenetic tree was reconstructed based on concatenated sequences (full-length) for five highly-conserved proteins (i.e. RpoA, RpoB, RpoC, GyrA and GyrB), which are commonly used for phylogenetic analyses [44]. This tree only included sequences for representative species of the genus *Pseudomonas* from different clades (except for the Pertucinogena clade for which all species were included) but also included representative sequences from some other prokaryotic genera including *Paraburkholderia, Xanthomonas* and *Stenotrophomonas* as well sequences for specific unnamed members of the genus *Pseudomonas* that were found to share conserved signature indels (CSIs) specific for the Pertucinogena clade. This tree was based on 4705 aligned positions and reconstructed as described above.

Identification of conserved signature indels (CSIs)

Identification of conserved signature indels was carried out as described in our earlier work [35, 45–47]. BLASTp searches were carried out on all protein sequences from the genomes of *P. litoralis*, *P. caeni* and *P. hussainii*, the species related to which are studied in greater detail in the present work. On the basis of the results of these BLAST searches, sequences from 12 to 15 diverse species of the genus *Pseudomonas* as well as 6–8 species from other genera of the family *Pseudomonadaceae* were retrieved for each protein and their multiple sequence alignments (MSA) were created using CLUSTALX [48]. The alignments were visually inspected for inserts or deletions of fixed lengths which were flanked on both sides by at least four to five conserved amino acids (aa) in the neighbouring 40–50 amino acids and which were specific for particular clades of species of the genus *Pseudomonas*. In the present study, our work has focused mainly on identifying conserved indels that are specific for the Pertucinogena clade of species and some other species that branch outside of the main cluster of species of the genus *Pseudomonas*. Query sequences encompassing the potential indels and flanking regions (60–100 aa long) were collected and a more detailed BLASTp search (500 or more hits) was carried out to determine the group specificities of the observed indels. Signature files for all CSIs of interest were formatted by using SIG_CREATE and SIG_STYLE programmes from the GLEANS software package (available on Gleans.net) [46, 49]. Due to space constraints, sequence information is shown for only a limited number of species in the figures. However, unless otherwise indicated the CSIs reported here are only found in all or most of the named species of the genus *Pseudomonas* from the indicated groups.

RESULTS

Phylogenetic analysis of the family *Pseudomonadaceae*

The family *Pseudomonadaceae* besides the genus *Pseudomonas* contains ten other genera with validly published names including *Azomonas*, *Azorhizophilus*, *Azotobacter*, *Entomomonas*, *Mesophilobacter*, *Oblitimonas*, *Permianibacter*, *Rhizobacter*, *Rugamonas* and *Thiopseudomonas* [4, 6, 50]. Unlike *Pseudomonas*, which is a very large and genetically diverse genus harboring more than 240 species with validly published names, the other genera within the family *Pseudomonadaceae* contain only a limited number of species (less than 20 in total) [4]. However, the interrelationships of species from these genera to the species of the genus *Pseudomonas* have not been thoroughly studied and remain unclear. Genome sequences are now available for >85% of the species with validly published names from the genus *Pseudomonas* (210 out 240) as well representatives from all other genera of the family *Pseudomonadaceae* except the

genus *Mesophilobacter*. Thus, the evolutionary relationships among species of the family *Pseudomonadaceae* can now be comprehensively examined using genome sequence data. In the present work, we have reconstructed a phylogenetic tree based on 255 species of the family *Pseudomonadaceae* whose genome sequences were available in NCBI genome sequence database (https://www.ncbi.nlm.nih.gov/genome/browse#!overview/Pseudomonadaceae) as of December 30, 2020. Using these genome sequences, we have reconstructed a phylogenetic tree based on concatenated sequences for 118 conserved proteins, which are commonly shared by members of the class *Gammaproteobacteria* [33]. The tree shown in Fig. 1, which is based on 41417 aligned amino acid positions, was rooted using sequences for species of the genus *Moraxella* (a member of another family within the order *Pseudomonadales*) and it will be referred to as the phyloeco tree. The tree shown in Fig. 1 is robust and nearly all of the observed nodes in it are supported with 100% bootstrap scores (SH values). In this tree as well as in other Figures, the names of species that are not validly published are shown within double quotation marks.

In the tree shown in Fig. 1, species of the genus *Pseudomonas* grouped into a number of distinct clusters in different parts of the trees. In addition, several species of the genus *Pseudomonas* were not part of any observed clusters. Importantly, a number of clusters of species of the genus *Pseudomonas*, or individual species that are not part of any clusters were interspersed with species from other genera of the family *Pseudomonadaceae* indicating that the genus *Pseudomonas*, as currently known, is highly divergent and polyphyletic [16, 24]. In Fig. 1, for ease of presentation, the clades corresponding to most of the species clusters of the genus *Pseudomonas*, which are not the main focus of the present work, are shown in compressed forms. However, an uncompressed tree is provided as Fig. S1 and the species compositions of different clades that are shown in compressed forms is provided in Table 1. The *Pseudomonas* clusters observed in Fig. 1 (Fig. S1, Table 1) are labelled as the Aeruginosa, Alcaligenes, Anguilliseptica, Flexibilis, Fluorescens, Kuykendallii, Linyingensis, Lutea, Massiliensis, Oleovorans, Oryzihabitans, Pertucinogena, Putida, Resinovorans, Rhizosphaerae, Straminea, Stutzeri and Syringae clusters, after specific species of the genus *Pseudomonas* that are a part of these clades. Additionally, within the Fluorescens clade, a number of subclades, similar to those reported in earlier work [8, 15, 16, 22, 24], are also observed and labelled. The grouping of species of the genus *Pseudomonas* into different clades and subclades as seen in Fig. 1 (Fig. S1 and Table 1) is very similar to that reported by Hesse *et al.* [24]. However, as our phylogenetic tree includes many additional species of the genus *Pseudomonas*, it also reveals the existence of a number of smaller species clades, namely Alcaligenes, Flexibilis, Kuykendallii, Massiliensis and Rhizosphaerae, not identified previously [24]. All of the main clades observed in our phylogenetic trees are also indicated to be distinct according to the Genome Taxonomy Database (GTDB (http://gtdb.ecogenomic.org/), which is based on phylogenetic analysis of 120 ubiquitously conserved proteins and provides an important resource for phylogenetic–taxonomic inferences [29]. The GTDB taxon names for different clades are also indicated in the tree in Fig. 1 and Table 1. Similar branching patterns and groupings of species of the genus *Pseudomonas*, with minor differences (as indicated by [16]), have also been observed in other studies, using MLSA trees based on 16S rDNA, *gyrB*, *rpoB* and *rpoD* genes [8, 15, 16, 22].

Although phylogenetic relationships among species of the genus *Pseudomonas* have been examined in a number of studies [8, 15, 16, 22, 24], most of these studies have not included members of other genera of the family *Pseudomonadaceae*. Hence, the tree shown in Fig. 1 reveals a number of novel aspects of evolutionary relationships among species of the family *Pseudomonadaceae*. In Fig. 1, while a majority of the species of the genus *Pseudomonas* are part of a large cluster (shown in blue and referred to as the *Pseudomonas* main cluster), a number of species branched, or formed distinct clades, in different parts of the phylogenetic tree and they were surrounded by species belonging to other genera of the family *Pseudomonadaceae*. Some of the anomalies observed in this regard are as follows: (i) Species from two genera of the family *Pseudomonadaceae*, *Azotobacter* and *Azomonas*, are deeply embedded within the main *Pseudomonas* cluster highlighting the polyphyletic nature of this genus and the difficulty in distinguishing it from other genera [23, 26]. (ii) The Pertucinogena clade of species branched deeply and outside of the main cluster of species of the genus *Pseudomonas*. Species from several other genera of the family *Pseudomonadaceae*, including *Azotobacter, Azomonas, Entomomonas, Oblitimonas* and *Thiopseudomonas*, were more closely related to the main cluster of species of the genus *Pseudomonas* than the Pertucinogena clade of species. The deep-branching of the Pertucinogena clade has also been consistently observed in earlier studies [8, 15, 16]. (iii) The species *P. caeni* grouped with the species *T. denitrificans* in a cluster that also included species from the genera *Oblitimonas* and *Entomomonas*. (iv) The species *P. hussainii* also branched separately and deeply from the main cluster of species of the genus *Pseudomonas*. (v) Three other species of the genus *Pseudomonas*, namely *P. cissicola*, *P. geniculata* and '*P. acidophila*' branched very deeply in the tree and they lay in between the species from other genera of the family *Pseudomonadaceae* and outgroup species.

Identification of molecular markers for some clades of species of the family *Pseudomonadaceae*

The results of our phylogenomic studies demonstrate that the genus *Pseudomonas* exhibits extensive polyphyly and its members are interspersed among other genera of the family *Pseudomonadaceae*. In the tree shown in Fig. 1, the type species of the genus *Pseudomonas* i.e. *P. aeruginosa* is a part of the Aeruginosa clade and does not seem to share common evolutionary history with numerous other clades of species of the genus *Pseudomonas* observed in

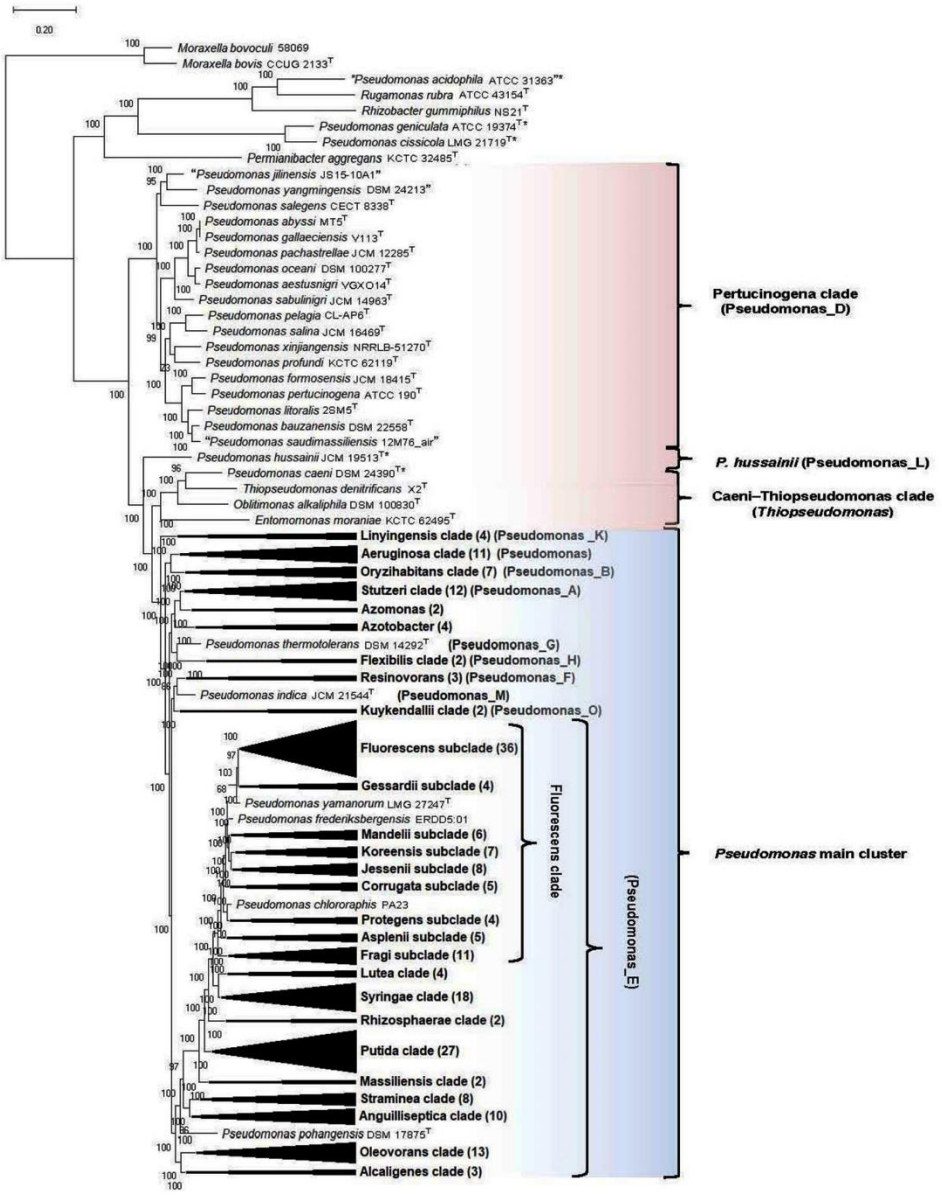

Fig. 1. A bootstrapped maximum-likelihood tree for 255 genome-sequenced species of the family *Pseudomonadaceae* based on concatenated sequences for 118 conserved proteins for members of the class *Gammaproteobacteria*. The statistical support values for different branches are indicated on the nodes. This tree was rooted by using species from the genus *Moraxella*. To aid in visualization, some of the clades of species of the genus *Pseudomonas* are shown in compressed forms. However, an uncompressed form of the tree is provided in Fig. S1. The number of species that are part of each clade and subclade are noted in parenthesis alongside the names of different clades. The Genome Taxonomy Database cluster names are also indicated for various clades (e.g. Pseudomonas_A to Pseudomonas_O). The species of the genus *Pseudomonas*, which are the main focus of this study are highlighted in pink and marked with * in the tree. Species with non-validly published names are shown within quotation (" ") marks in this and other Figures. The main cluster of species of the genus *Pseudomonas* seen in the tree is highlighted in light blue.

Table 1. Species composition of compressed clades of the genus *Pseudomonas* in Fig. 1

Clade name		Species name
Linyingensis (Pseudomonas_K)		*P. linyingensis, P. sagittaria, 'P. oryzae', P. guangdongensis*
Aeruginosa(Pseudomonas)		*P. humi, P. citronellolis, P. delhiensis, P. knackmussii, P. panipatensis, P. nitritireducens, P. nitroreducens, P. multiresinivorans, P. jinjuensis, P. aeruginosa, P. denitrificans*
Oryzihabitans (Pseudomonas_B)		*P. rhizoryzae, P. psychrotolerans, P. oryzihabitans, P. luteola, P. zeshuii, P. asuensis, P. duriflava.*
Stutzeri (Pseudomonas_A)		*P. chloritidismutans, P. kunmingensis, 'P. songnenensis', P. stutzeri, P. balearica, P. xanthomarina, P. zhaodongensis, P. nitrititolerans, P. kirkiae, 'P. saudiphocaensis', P. azotifigens, P. nosocomialis.*
Thermotolerans (Pseudomonas_G)		*P. thermotolerans*
Flexibilis (Pseudomonas_H)		*P. flexibilis, P. tuomuerensis*
Resinovorans (Pseudomonas_F)		*P. resinovorans, P. furukawaii, P. otitidis*
Indica (Pseudomonas_M)		*P. indica*
Kuykendallii (Pseudomonas_O)		*P. kuykendallii, P. matsuisoli.*
Pseudomonas_E	**Subclade**	*P. reactans, P. azotoformans, P. canadensis, P. simiae, P. allii, P. lurida, P. extremorientalis, P. palleroniana, P. rhodesiae, P. poae, P. cedrina, P. kairouanensis, P. nabeulensis, P. tolaasii, P. trivialis, P. marginalis, P. orientalis, P. libanensis, P. synxantha, P. carnis, P. lactis, P. paralactis, P. edaphica, P. salomonii, P. antarctica, P. fluorescens, P. cremoris, P. costantinii, P. sivasensis, P. kitaguniensis, P. haemolytica, P. grimontii, P. fildesensis, P. veronii, P. panacis, P. extremaustralis*
	Fluorescens	
	Gessardii	*P. gessardii, P. proteolytica, P. brenneri, P. mucidolens,*
	Yamanorum	*P. yamanorum*
	Frederiksbergensis	*P. frederiksbergensis.*
	Mandelii	*P. silesiensis, P. prosekii, P. mandelii, P. arsenicoxydans, P. migulae, P. lini.*
	Koreensis	*'P. atacamensis', P. koreensis, P. moraviensis, P. granadensis, P. baetica, P. helmanticensis, 'P. kribbensis'*
	Jessenii	*P. laurylsulfatiphila, P. jessenii, P. laurylsulfatovorans, P. vancouverensis, P. moorei, P. mohnii, P. umsongensis, P. reinekei.*
	Corrugata	*P. corrugata, P. mediterranea, P. kilonensis, P. thivervalensis, P. brassicacearum.*
	Chlororaphis	*P. chlororaphis*
	Protegens	*'P. aestus', P. piscis, P. saponiphila, P. protegens.*
	Asplenii	*P. asplenii, P. fuscovaginae, P. agarici, P. gingeri, 'P. batumici',*
	Fragi	*P. endophytica, P. helleri, P. lundensis, P. weihenstephanensis, P. versuta, P. taetrolens, P. fragi, P. bubulae, P. psychrophila, P. deceptionensis, P. saxonica.*
	Lutea	*P. graminis, P. lutea, P. bohemica, P. abietaniphila.*
	Syringae	*P. caspiana, P. ovata, P. cichorii, P. viridiflava, P. asturiensis, P. floridensis, P. coronafaciens, P. tremae, P. avellanae, P. cannabina, P. syringae, P. congelans, P. cerasi, P. amygdali, P. meliae, P. savastanoi, P. ficuserectae, P. caricapapayae,*
	Rhizosphaerae	*P. coleopterorum, P. rhizosphaerae.*
	Putida	*P. wadenswilerensis, P. donghuensis, P. vranovensis, P. alkylphenolica, P. brassicae, 'P. qingdaonensis', P. mosselii, P. soli, P. entomophila, P. sichuanensis, P. guariconensis, P. taiwanensis, 'P. hunanensis', P. juntendi, P. monteilii, P. plecoglossicida, P. shirazica, P. asiatica, P. pudica, P. inefficax, P. parafulva, P. putida, P. capeferrum, P. cremoricolorata, P. reidholzensis, P. japonica, P. laurentiana.*
	Massiliensis	*'P. typographi', 'P. massiliensis'.*
	Straminea	*P. straminea, P. fulva, P. argentinensis, P. punonensis, P. seleniipraecipitans, P. flavescens, P. daroniae, P. dryadis.*
	Anguilliseptica	*P. segetis, P. marincola, P. benzenivorans, P. taeanensis, P. anguilliseptica, P. peli, P. leptonychotis, P. guineae, P. cuatrocienegasensis, P. borbori.*
	Pohangensis	*P. pohangensis*
	Oleovorans	*P. khazarica, P. oleovorans, P. mendocina, P. hydrolytica, P. guguanensis, P. pseudoalcaligenes, 'P. indoloxydans', 'P. sediminis', P. composti, 'P. sihuiensis', P. chengduensis, P. alcaliphila, P. toyotomiensis.*
	Alcaligenes	*P. fluvialis, P. pharmafabricae, P. alcaligenes.*

the tree. However, the branching of species in phylogenetic trees is influenced by large numbers of variables and it is often not reliable [49, 51–53]. Furthermore, in the case of species of the genus *Pseudomonas*, many of the observed clades are separated by very short branches (as in the Pseudomonas_E group), which makes it difficult to reliably distinguish or demarcate species from these clades on the basis of branching in the phylogenetic tree alone. Hence, it is important to confirm the existence and genetic cohesiveness of the observed clades byindependent and robust means that are not dependent upon phylogenetic analysis. Genome sequences provide a powerful resource for identifying molecular markers which are uniquely shared by an evolutionarily related group of organisms. There is now extensive evidence showing that conserved signature insertions and deletions (CSIs) in genes or proteins provide an important class of molecular markers for evolutionary and taxonomic studies [45–47, 49, 54]. When a conserved indel of a specific length, which is present at a specific position, is specifically shared by a monophyletic group of organisms, the most parsimonious explanation is that the rare genetic change giving rise to this CSI occurred in a common ancestor of this group of organisms and then was vertically inherited by various descendants [45, 46, 49, 55]. Thus, CSIs represent molecular synapomorphic characteristics that provide reliable evidence, independent of the branching patterns in phylogenetic trees, of the common ancestry and relatedness of a given group of species. Hence, an important aspect of the present study focused on the use of CSI identification approach to reliably demarcate and clarify the evolutionary relationships of certain deep-branching lineages within the family *Pseudomonadaceae*. The specific lineages or group of species which are the focus of this study are marked with * in Fig. 1 and the results of our studies on demarcating these lineages using the CSI-identification approach are described below.

Molecular signatures specific for the Pertucinogena clade of species

As seen from Fig. 1, the Pertucinogena clade of species forms a distinct lineage outside of the main cluster of species of the genus *Pseudomonas*. The distinctness of this group of microorganisms from all other species of the genus *Pseudomonas* as well as other genera within the family *Pseudomonadaceae* is strongly supported by our identification of 24 CSIs in proteins involved in diverse functions that are specific for all of the species from this clade. Sequence information for one of these CSIs is presented in Fig. 2. In this figure we show a segment of the sequence alignment for the flagellar protein FlgN in which a two amino acid insertion (boxed) in a conserved region is specifically present in all 19 species from the Pertucinogena clade. However, this insert is not found in any other named species from either the family *Pseudomonadaceae* or other bacteria. The insert shown is present within a conserved region, indicating that it constitutes a reliable genetic/molecular characteristic. The FlgN protein is a flagellar type III export chaperone that plays an important role in controlling the motility of the bacterial flagellum [56]. Similar to the CSI shown in Fig. 2, our analyses have identified 23 other CSIs in diverse proteins that are also distinctive characteristics of the Pertucinogena clade of species. Sequence information for these CSIs is presented in Figs S2–S24 and a summary of some of their characteristics is provided in Table 2. On the basis of the exclusive presence of these CSIs in different species from the Pertucinogena clade, the genetic changes responsible for these CSIs are postulated to have occurred in a common ancestor of this group of organisms and then been vertically inherited by other members. It should be noted that in addition to the known members of the Pertucinogena clade, the CSIs specific for this clade are also present in nine uncharacterized strains of the genus *Pseudomonas*. The significance of this observation is discussed later.

Molecular markers supporting grouping of *Pseudomonas caeni* with the genera *Thiopseudomonas* and *Oblitimonas*

The species *P. caeni* exhibits deep and distinct branching in all of the reconstructed phylogenetic trees [8, 15, 16, 24]. In the tree shown in Fig. 1, this species groups within a clade with species from the genera *Thiopseudomonas*, *Oblitimonas* and also *Entomomonas*. In the GTDB taxonomy, *P. caeni* along with species from the genera *Oblitimonas* and *Thiopseudomonas* are indicated to be a part of the genus *Thiopseudomonas* [29]. Furthermore, Hesse *et al.* [24] have noted that *P. caeni* is unusual in having a much smaller genome size and lower DNA G+C content in comparison to most other species of the genus *Pseudomonas* and in these regards it is more similar to the species of the genera *Thiopseudomonas* and *Oblitimonas* [28, 57]. Our comparative genomic analysis has identified six CSIs that support a specific relationship of *P. caeni* to the species from the genera *Thiopseudomonas* and *Oblitimonas*. In Fig. 3(a), we present partial sequence alignment of an ABC transporter ATP-binding protein, where a one aa insert in a conserved region is specifically shared by *P. caeni, O. alkaliphila* and *T. denitrificans*, but not present in any other species of the genus *Pseudomonas* including those from the Pertucinogena clade. Sequence information for five other CSIs showing similar specificities is provided in Figs S25 –S29 and some characteristics are summarized in Table 3. The results presented in Figs 3a and S25–S29 also indicate that in addition to *P. caeni, O. alkaliphila* and *T. denitrificans*, the CSIs specific for this clade are also shared by another member of the genus *Pseudomonas*, strain C27(2019), indicating that this uncharacterized member of the genus *Pseudomonas* may also be a member of this clade.

Molecular markers specific for other deep-branching species of the genus *Pseudomonas*

In Fig. 1, the species *P. hussainii* [58] branches deeply and outside of the main cluster of species of the genus *Pseudomonas* as well as members of several other genera of the family *Pseudomonadaceae* including *Azotobacter,*

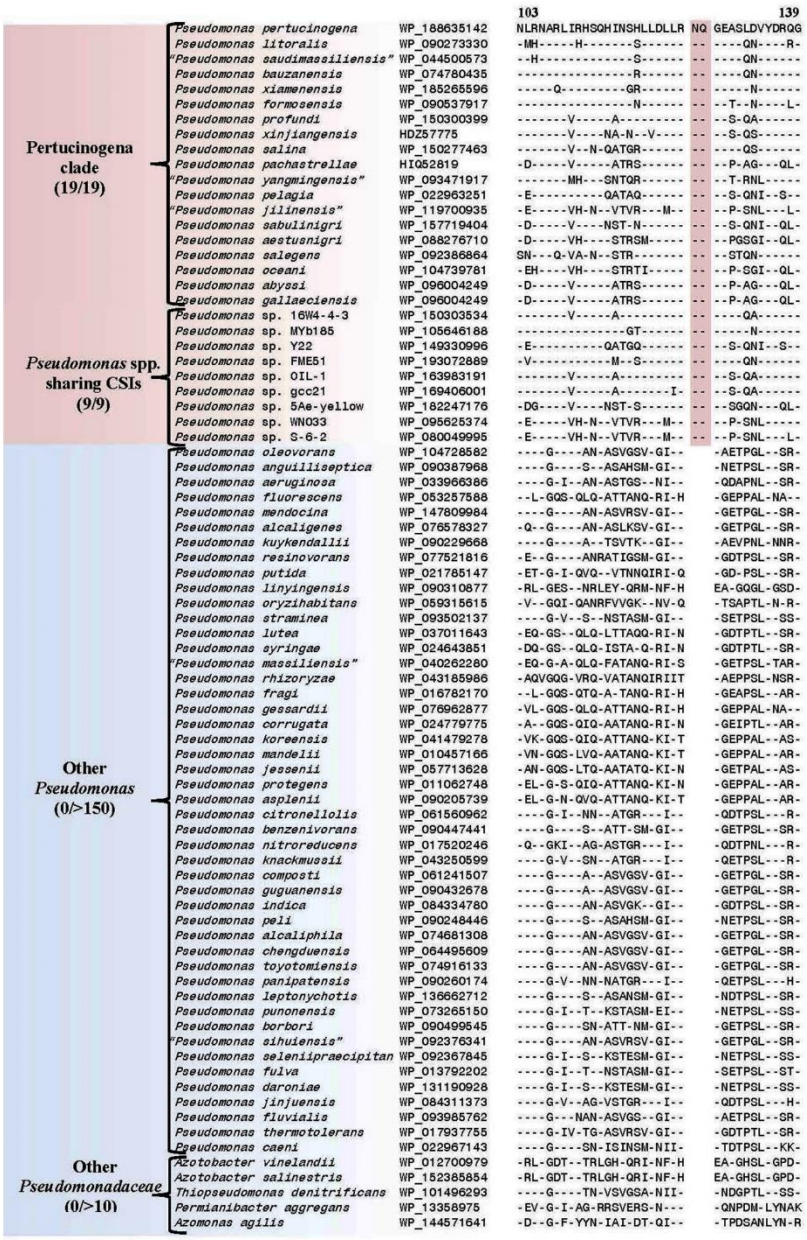

Fig. 2. Partial sequence alignment of the flagellar FlgN protein showing a two amino acid insertion within a conserved region (boxed) that is commonly shared by all members of the. Pertucinogena clade, but not found in any other named species of the family *Pseudomonadaceae* or other bacteria. The dashes (-) in the alignment indicate identity with the amino acids on the top line. Accession numbers for different sequences are indicated in the second column and the numbers at the top indicate the position of this sequence fragment within the protein sequences. In addition to the members of the Pertucinogena clade, this CSI is also present in nine other members of the genus *Pseudomonas*, which are also indicated to be a part of this clade. Sequence information for 23 other CSIs that are also specific for the Pertucinogena clade is provided in Figs S2–S24.

Table 2. Characteristics of conserved signature indels (CSIs) specific for the Pertucinogena clade*

Protein name	Accession no.	Indel size	Indel position	Figure number
Flagellar protein FlgN	WP_090273330	2 aa Ins	103–139	Fig. 2
RND family transporter	WP_188635687	1 aa Ins	532–576	Fig. S2
tRNA cyclic N6 threonylcarbamoyladenosine (37) synthase TcdA	WP_188634832	1 aa Ins	169–223	Fig. S3
Protein BatD	WP_090272452	1 aa Ins	252–306	Fig. S4
RNA polymerase-associated protein RapA	WP_188635120	1 aa Del	221–261	Fig. S5
Glyceraldehyde-3-phosphate dehydrogenase	WP_188635402	1 aa Del	68–114	Fig. S6
Dephospho-CoA kinase	WP_188636162	1 aa Ins	17–52	Fig. S7
SpoVR family protein	WP_188636806	1 aa Ins	05–50	Fig. S8
Serine/threonine protein kinase	WP_090272375	3 aa Ins	142–193	Fig. S9
Recombination-associated protein RdgC	WP_090272025	1 aa Del	34–79	Fig. S10
OprD family porin	SDR85642	1 aa Ins	125–163	Fig. S11
Bifunctional diguanylate cyclase/phosphodiesterase	SDS89372	1 aa Ins	176–225	Fig. S12
NADPH-dependent 7-cyano-7-deazaguanine reductase QueF	WP_090272302	1 aa Ins	213–257	Fig. S13
YcgN family cysteine cluster protein	WP_090272373	1 aa Ins	35–76	Fig. S14
SprT family zinc-dependent metalloprotease	WP_090272415	1 aa Del	104–149	Fig. S15
Universal Stress Protein	WP_090272538	1 aa Del	61–107	Fig. S16
Glyceraldehyde-3-phosphate dehydrogenase	WP_090273031	1 aa Del	110–160	Fig. S17
DUF4892 domain-containing protein	WP_090273091	1 aa Ins	159–206	Fig. S18
Symmetrical bis (5′-nucleosyl)-tetraphosphatase	WP_090274322	1 aa Del	193–244	Fig. S19
YaiI/YqxD family protein	WP_090274602	1 aa Ins	105–143	Fig. S20
Diaminopimelate decarboxylase	WP_090274749	1 aa Ins	306–357	Fig. S21
YifB family Mg chelatase-like AAA ATPase	WP_090274765	1aa Ins	46–91	Fig. S22
Potassium transporter	WP_090275387	1aa Del	273–320	Fig. S23
LEA type 2 family protein	WP_157718669	1aa Del	91–137	Fig. S24

*Most of these CSIs specific for this clade are also present in the following strains of members of the genus *Pseudomonas:* 16W4-4-3, MYb185, Y22, FME51, OIL-1, gcc21, 5Ae-yellow, WN033 and S-6–2, which branch with the members of the Pertucinogena clade.

Azomonas, Entomomonas, Oblitimonas and *Thiopseudomonas*. The GTDB taxonomy also places this species into a separate clade *Pseudomonas*_L [29]. Our comparative genomic analysis has identified 22 CSIs in different proteins that are uniquely found in this species but absent in all other organisms. One example of a CSI that is specific for *P. hussainii* is presented in Fig. 3b. This figure shows partial sequence alignment of the protein glutathione-disulfide reductase in which a four amino acid insert (boxed) is present only in *P. hussainii* but not in any other member of the genus *Pseudomonas* or other bacteria. Sequence information for 21 other CSIs that are also specific for *P. hussainii* is presented in Figs S30 –S50 and some of their characteristics are summarized in Table 3. It is important to note that the CSIs that are specific for *P. hussainii* are not shared by any other member of the genus *Pseudomonas* or uncharacterized species, indicating the distinctness of this species from all others.

Three other species of the genus *Pseudomonas* namely *P. cissicola, P. geniculata* and '*P. acidophila*' branch very deeply in the tree. Results from earlier studies indicate that these species are more closely related to other prokaryotic genera [17, 30, 31, 59] and they are misclassified as species of the genus *Pseudomonas*. However, these species are still listed as members of the genus *Pseudomonas* in the NCBI genome sequence database and also in the LPSN [4] and Names for Life servers [6]. In BLASTp searches with proteins from *P. cissicola, P. geniculata* and '*P. acidophila*', the top 50 BLAST hits observed were from the species from genera *Xanthomonas, Stenotrophomonas* and *Paraburkholderia* respectively (Figs S51–S53). In a phylogenetic tree based on concatenated sequences for five conserved proteins

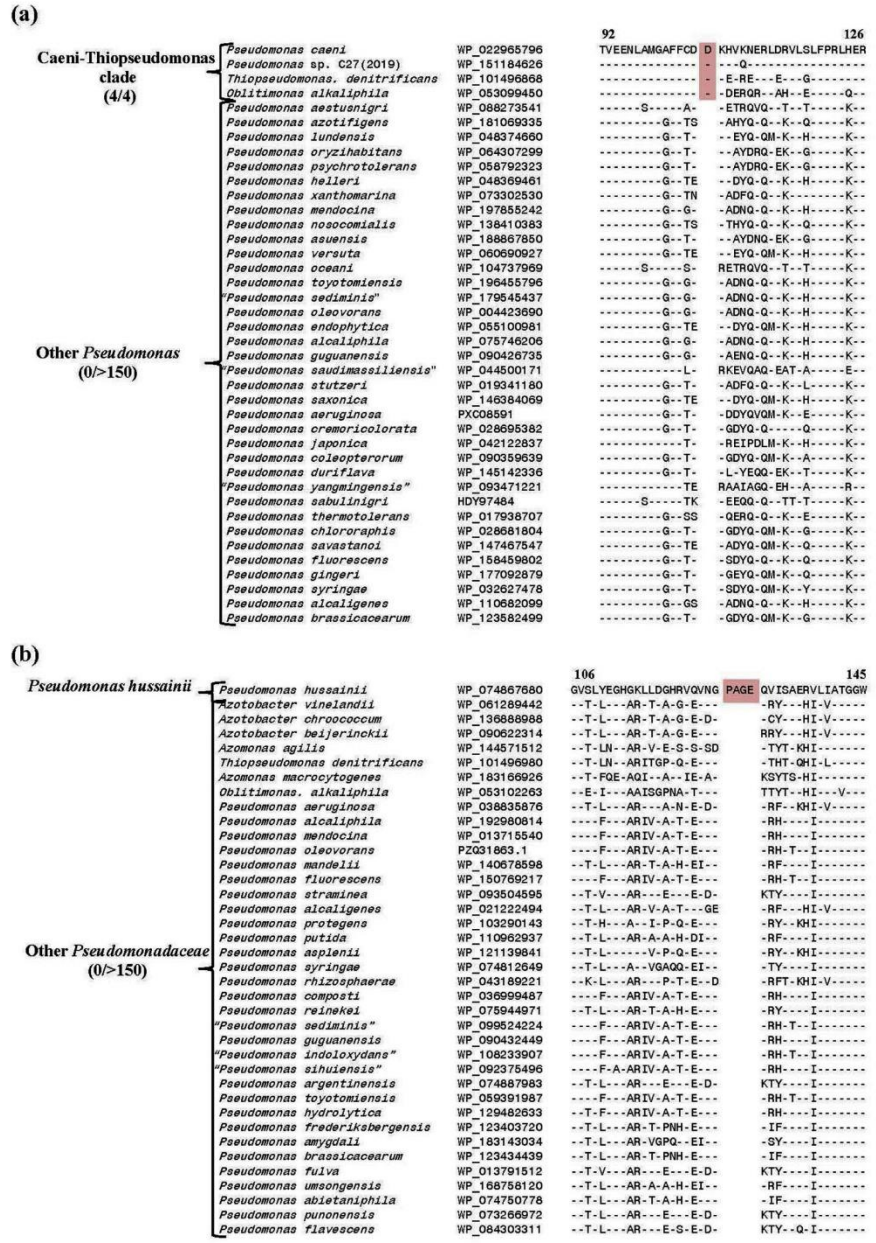

Fig. 3. (a) Partial sequence alignment of a an ABC transporter ATP-binding protein showing a one amino acid CSI (boxed) that is commonly and exclusively shared by, *P. caeni*, *O. alkaliphila*, *T. denitrificans* and an unnamed member of the genus *Pseudomonas*, C27 (2019), demarcating this group of species in molecular terms. Information for five other CSIs specific for this clade is provided in Figs S25–S29 and summarized in Table 2. (b) Excerpts from the sequence alignment of glutathione-disulfide reductase protein showing a four amino acid insert in a conserved region that is exclusively present in *P. hussainii*. Sequence and information for 21 other CSIs specific for *P. hussainii* is presented in Figs S30–S50 and some of their characteristics are summarized in Table 3. The dashes (-) in all sequence alignment indicate identity with the amino acids on the top line.

Table 3. Summary of CSIs specific for members of the Caeni –Thiopseudomonas clade and *Pseudomonas hussainii*

Protein name	Accession number	Indel size	Indel position	Figure number	Specificity
ABC transporter ATP-binding protein	WP_022965796	1 aa Ins	92–126	Fig. 3(a)	**Caeni–Thiopseudomonas clade**
Si-specific NAD(P)(+) transhydrogenase	WP_022967548	1 aa Del	117–156	S25	
Esterase	WP_022965206	1 aa Ins	18–54	S26	
Response regulator transcription factor	WP_022965478	1aa Ins	20–64	S27	
Phosphoribosylamine glycine ligase	WP_022965180	1 aa Ins	42–82	S28	
C40 family peptidase	WP_051145790	1 aa Ins	157–200	S29	
Glutathione-disulfide reductase	WP_074867680	4 aa Ins	106–145	Fig. 3(b)	***Pseudomonas hussainii***
Dihydrolipoyl dehydrogenase	WP_071870339	1 aa Ins	55–96	S30	
UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	WP_071870788	1 aa Ins	37–80	S31	
Phosphotransferase	WP_074867300	1 aa Del	153–192	S32	
Citrate (Si)-synthase	WP_071870346	2 aa Ins	187–232	S33	
HugZ family protein	WP_074867266	1 aa Del	166–204	S34	
ATP-binding cassette domain-containing protein	WP_074867615	2 aa Del	92–135	S35	
DNA polymerase III subunit beta	WP_071871523	2 aa Ins	200–242	S36	
Cupin domain-containing protein	WP_071872405	1aa Ins	185–224	S37	
L,D-transpeptidase family protein	WP_071872639	2aa Ins	161–200	S38	
MBL fold metallo-hydrolase	WP_083432004	1aa Ins	257–296	S39	
Phosphate signalling complex protein PhoU	WP_074864309	1aa Ins	106–149	S40	
Arginine/ornithine succinyltransferase subunit alpha	WP_074864650	2aa Ins	113–166	S41	
Excinuclease ABC subunit A	SEL53496	1aa Del	635–672	S42	
Sulfurtransferase complex subunit TusC	WP_074869676	1aa Del	30–68	S43	
Potassium/proton antiporter	WP_074865112	1aa Del	156–196	S44	
Metalloprotease PmbA	WP_071872802	4aa Ins	171–214	S45	
Hsp20 family protein	WP_074865611	1aa Del	9–51	S46	
Serine hydrolase	WP_074870128	1aa Del	299–334	S47	
Recombination-associated protein RdgC	WP_074870192	1aa Del	36–72	S48	
Sulfite exporter TauE/SafE family protein	WP_074866252	1aa Del	51–95	S49	
2-oxoglutarate dehydrogenase E1	WP_074866854	1aa Ins	273–313	S50	

(RpoA, RpoB, RpoC, GyrA and GyrB), which included representative species from different clades of the genus *Pseudomonas* clades as well as limited numbers of species from specific genera (Fig. 4), these three species of the genus *Pseudomonas* branched reliably with members of the genera *Xanthomonas, Stenotrophomonas* and *Paraburkholderia*, respectively. The observed results confirm that these species are more closely related to these other genera instead of the genus *Pseudomonas*. On the basis of our earlier work and the results of additional analysis conducted in this work, we have also identified some CSIs that are specific for members of the genera *Xanthomonas, Stenotrophomonas* and *Paraburkholderia* [44, 54, 59]. One example each of a CSI specific for each of these three genera is shown in Fig. 5. As seen from this figure, the species *P. cissicola, P. geniculata* and '*P. acidophila*' are found to share CSIs that are specific for the genera *Xanthomonas, Stenotrophomonas* and *Paraburkholderia*, respectively. These results provide further evidence indicating that these species are specifically related to these genera and should be reclassified accordingly.

Predictive abilities of the CSIs for identification and classification of uncharacterized species

Extensive earlier work on CSIs provides evidence that these molecular characteristics are highly specific for a given group of organisms and they exhibit strong predictive ability to be present in other members of a given group [34, 35, 45, 49]. As noted earlier, the CSIs that are specific

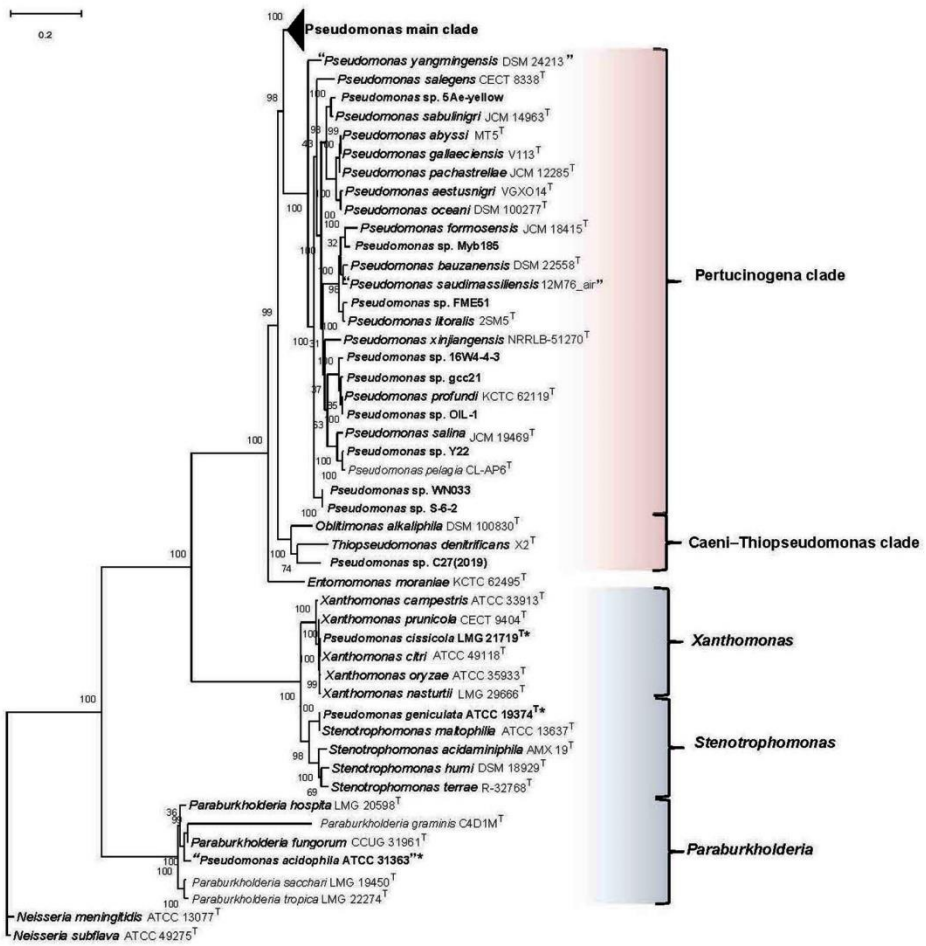

Fig. 4. Maximum likelihood phylogenetic tree based on concatenated sequences for the RpoA, RpoB, RpoC, GyrA and GyrB proteins. This tree includes all of the species from the Pertucinogena and Caeni-Thiopseudomonas clades, along with the 10 uncharacterized members of the genus *Pseudomonas* (names shown in bold type) sharing the CSIs specific for these clades. The tree also includes representative species from all of the main clades of the genus *Pseudomonas* . Additionally, this tree also includes sequences for three species misclassified as members of the genus *Pseudomonas* (*P. cissicola*, *P. geniculata* and '*P. acidophila*' names shown in bold type) and representative species from the genera *Xanthomonas*, *Stenotrophomonas* and *Paraburkholderia*. The names of the species are followed by the strain number of the species whose sequences were used in this study.

for the Pertucinogena clade are also shared by nine other uncharacterized members of the genus *Pseudomonas* (strains 16W4-4-3, MYb185, Y22, FME51, OIL-1, gcc21, 5Ae-yellow, WN033, S-6–2). Likewise, the CSI that is specific for the Caeni–Thiopseudomonas clade is also shared by one uncharacterized member of the genus *Pseudomonas* i.e. C27 (2019). Thus, it was important to determine whether the shared presence of these CSIs in these members of the genus *Pseudomonas* (strains) is fortuitous or these uncharacterized strains are part of these specific clades and that is why they are sharing these group-specific CSIs. To examine this, we have reconstructed another phylogenetic tree shown in Fig. 4 (based on RpoA, RpoB, RpoC, GyrA and GyrB proteins) which included all of the species from these two clades, along with the 10 uncharacterized members of the genus *Pseudomonas* sharing the CSIs for these two clades, as well as representatives from all of the main clades of species of the genus *Pseudomonas*. In addition, as noted previously, this tree also included sequences for *P. cissicola, P. geniculata* and '*P. acidophila*'

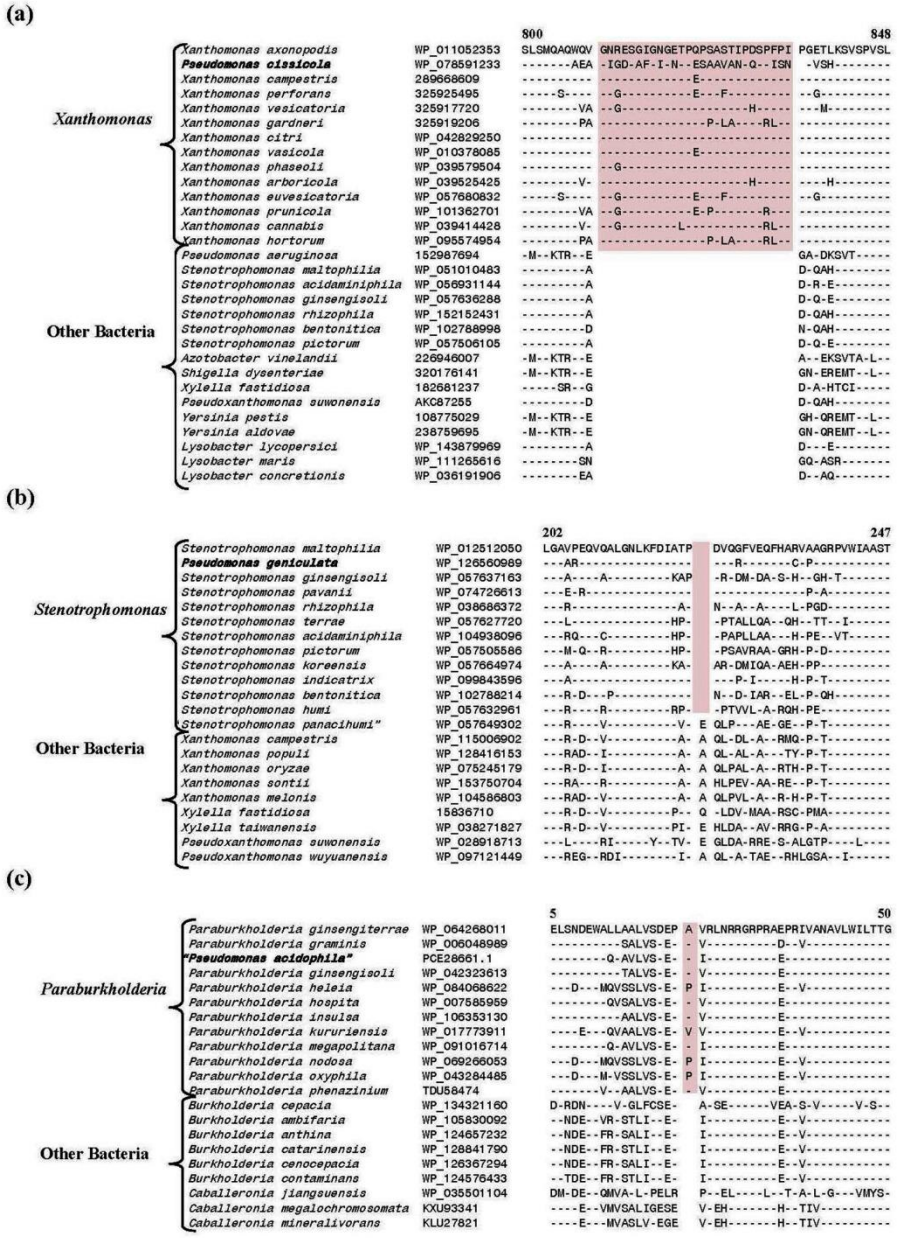

Fig. 5. (a) Partial sequence alignment of the protein phosphoribosyl formylglycinamidine synthase showing a 27 amino acid insertion (boxed) which is specific for members of clades of the genus *Xanthomonas*, but which is also shared by *Pseudomonas cissicola*. (b) Partial sequence alignment of 3-deoxy-D-manno-octulosonic-acid transferase showing a one amino acid deletion (boxed) which is commonly shared by all members of the genus *Stenotrophomonas* and also *Pseudomonas geniculata*. (c) Excerpts from sequence alignment of the protein undecaprenyl-phosphate glucose phosphotransferase showing a one amino acid insertion (boxed) that is commonly shared by '*P. acidophila*' and members of the genus *Paraburkholderia*. The dashes (-) in all sequence alignments indicate identity with the amino acids on the top line.

and representative species from the genera *Xanthomonas, Stenotrophomonas* and *Paraburkholderia*. As seen from tree in Fig. 4, all nine members of the genus *Pseudomonas* which shared the CSIs specific for the Pertucinogena clade branched reliably within this clade with other members of the clade. Likewise, *Pseudomonas* C27 (2019), which shared the CSI specific for the Caeni–Thiopseudomonas clade also grouped with other members of this clade. These results provide strong evidence that the shared presence of CSIs specific for a given group in an uncharacterized species or strain is strongly indicative that the particular species or strain sharing the CSI is evolutionary related to the group for which the CSI is specific.

DISCUSSION

The genus *Pseudomonas*, which is a part of the family *Pseudomonadaceae,* constitutes one of the largest genera within the prokaryotes, harboring more than 240 species with validly published names. The members of this genus span enormous genetic and metabolic diversity and the number of species in this genus is growing at a rapid rate. In 2020 alone, more than 30 descriptions of novel species of the genus *Pseudomonas* were published and indexed in PubMed [14]. However, it is well known that the genus *Pseudomonas* as currently known is not monophyletic and it exhibits polyphyly with species from other genera of the family *Pseudomonadaceae* [8, 16, 17, 24–26, 29]. As the genus *Pseudomonas* contains many medically and agriculturally important species, as well others members producing a variety of biologically active metabolites that are useful for diverse applications including plant growth promotion, biocontrol and bioremediation agents etc. [2, 8, 9, 11, 12, 60, 61], it is of much interest to develop a reliable understanding of the evolutionary relationships among species of the genus *Pseudomonas* and family *Pseudomonadaceae.*

Genome sequences are now available for 210 of the 240 species of the genus *Pseudomonas* with validly published names as well as 33 other members of the genus *Pseudomonas*, for which the names are not validly published. In addition, genome sequences are also available for a majority of the species from other genera within the family *Pseudomonadaceae*. This breadth of genomic data affords excellent coverage of the genetic diversity that exists within the genus *Pseudomonas* and the family *Pseudomonadaceae* and provides an exclusive, previously unavailable, resource for examining and clarifying the evolutionary relationships amongst the species of the genus *Pseudomonas* and the family *Pseudomonadaceae*. Using these genome sequences, we have reconstructed a robust phylogenomic tree for species of the family *Pseudomonadaceae* based on concatenated sequences for 118 conserved proteins for members of the class *Gammaproteobacteria* [33]. In this tree, similar to the results from earlier studies [8, 15, 16, 22, 24], species of the genus *Pseudomonas* form a number of different groups or clusters (Figs 1 and S1). The observed grouping of species of the genus *Pseudomonas* into different clades and subclades in our tree (Figs 1 and S1) is very similar to that observed by Hesse *et al.* [24] based on their analysis of 163 species of the genus *Pseudomonas*. However, as the present study includes many additional species of the genus *Pseudomonas*, it reveals the existence of a number of smaller species clades (Alcaligenes, Flexibilis, Kuykendallii, Massiliensis and Rhizosphaerae) not seen in earlier work. Most of the main groups and clusters of species of the genus *Pseudomonas* seen in these trees, with minor differences (see [16]), are also observed in the MLSA trees reconstructed by various investigators based on 16S rDNA, *gyrB, rpoB* and *rpoD* genes [8, 15, 16, 22, 23]. Additionally, most of the species clades seen in our tree are also observed in GTDB taxonomy [26] as indicated in Fig. 1, Table 1.

An important aspect of the present work is that here we have examined in detail the evolutionary relationships of species of the genus *Pseudomonas* to members of other genera of the family *Pseudomonadaceae*. Hence, the results from this study more clearly depict the evolutionary relationships among different species of the family *Pseudomonadaceae* and the extensive polyphyletic nature of the genus *Pseudomonas*. In the tree shown in Fig. 1, a majority of the species of the genus *Pseudomonas* are part of one large and genetically highly diverse cluster (highlighted in blue). This cluster also includes species from the genera *Azotobacter* and *Azomonas*, indicating that these genera are currently inseparable from the genus *Pseudomonas* [23, 25–27]. However, in addition to this main cluster of species of the genera *Pseudomonas–Azotobacter–Azomonas*, a number of other species or species clades of the genus *Pseudomonas* are observed, which exhibit deeper branching in the tree and their closest relatives are species from other genera of the family *Pseudomonadaceae*. These species include a large clade of species of the genus *Pseudomonas* referred to as the Pertucinogena clade and several isolated species (*P. caeni, P. hussainii*, '*P. acidophila*', *P. cissicola* and *P. geniculata*), which either branched as separate lineages in the tree or grouped with species from other genera. On the basis of their branching in the tree, these species are clearly distinct from the main cluster of the genus *Pseudomonas*. In this work, we have also conducted comprehensive analyses on the protein sequences from these species to identify multiple molecular markers that are specific for these species. The results obtained and presented here provide strong and independent evidence that these species, or clusters of species, are distinct from the main cluster of species of the genus *Pseudomonas* and that they should be reclassified into either novel genera or other known genera.

Of the species of the genus *Pseudomonas* that are not part of the main cluster, the Pertucinogena clade comprises the largest group consisting of 19 named species, of which 16 names are validly published. These species form a strongly supported clade in Fig. 1 as well as in earlier phylogenetic studies [8, 15, 16, 22, 24]. In the GTDB webserver this clade is referred to as the Pseudomonas_D taxon [29]. Although a deep branching of Pertucinogena clade has been observed in earlier work [8, 15, 16, 22, 24], the present

study shows clearly for the first time, to our knowledge, that the members of this clade are more distantly related to the main cluster of species of the genus *Pseudomonas* than the members of several other genera of the family *Pseudomonadaceae* including *Oblitimonas, Thiopseudomonas* and *Entomomonas*. Thus, the members of this clade should be recognized as representing a novel genus within the family *Pseudomonadaceae*. In the present work, we have identified 24 CSIs in diverse proteins that are exclusively shared by all members of this clade. It should be mentioned that the proteins in which these CSIs are found were not part of the 118 proteins used for the reconstruction of our phylogenetic tree (Fig. 1). Hence, the specific grouping of these species in the phylogenetic tree is not influenced by the presence of these CSIs. Another interesting and distinguishing property of the Pertucinogena clade is that most of these species have been isolated from marine environments [62–67] and they either require high concentration of salt for growth or are able to grow in the presence of high salt concentrations. In view of the salt-tolerant property of this strongly supported clade, we are proposing transfer of species from the Pertucinogena clade into a novel genus *Halopseudomonas* gen. nov. within the family *Pseudomonadaceae*. Extensive earlier works on CSIs attest that these molecular features are highly specific characteristics of a given group of organisms with strong predictive ability to be found in other members of that group [35, 45–47]. In the present work, the CSIs specific for the Pertucinogena clade were also shared by nine unclassified strains of members of the genus *Pseudomonas*. The results of our phylogenetic analysis confirm that these nine uncharacterized isolates of the genus *Pseudomonas* branch reliably with other members of the Pertucinogena clade. Thus, it is expected that upon further characterization several of these members of the genus *Pseudomonas* will represent novel species within the genus *Halopseudomona*s. The shared presence of CSIs specific for a given genus in now increasingly used as a molecular characteristic for the assignment of novel species into specific genera [31, 68].

Pseudomonas caeni is another species that branches outside of the main cluster of species of the genus *Pseudomonas* [69]. In our phylogenetic tree, *P. caeni* groups reliably with members of the genera *Thiopseudomonas* and *Oblitimonas*. The GTDB webserver also assigns *P. caeni* into the genus *Thiopseudomonas* along with *Oblitimonas* [29]. In the present work, we have identified six CSIs that are commonly shared by all three of these species, along with an uncharacterized isolate *Pseudomonas* C27(2019), which branches within this group in phylogenetic tree (Fig. 4). Another unusual characteristic common to *P. caeni, O. alkaliphila* and *T. denitrificans* [24], is that they all have much smaller genome sizes (2.49 to 3.02 MB) and lower DNA G+C contents (47.4–59.0 mol%) in comparison with other species of the genus *Pseudomonas*. Thus, the results based on phylogenetic branching and identified molecular characteristics make a strong case for the combining of these three species into a single genus. According to the code governing the nomenclature of prokaryotic organisms [70], of the genera *Oblitimonas* and *Thiopseudomonas,* the genus *Thiopseudomonas* due to its earlier publication [57] has priority over the name *Oblitimonas* [28]. Hence, we are proposing to integrate *Oblitimonas* with *Thiopseudomonas* and transfer *P. caeni* into this genus as a new name combination. *P. hussainii* is another deep-branching species [58], which lies outside of the main cluster of species of the genus *Pseudomonas*. This species is assigned into Pseudomonas_L clade by the GTDB webserver [29]. The distinctness of *P. hussainii* from all other members of the family *Pseudomonadaceae* is strongly supported by 22 identified CSIs that are exclusively present in this species. On the basis of these observations, we are proposing transfer of *P. hussainii* into a novel genus, *Atopomonas* gen. nov., within the family *Pseudomonadaceae*.

The present study also clarifies that the species *P. cissicola*, *P. geniculata* and '*P. acidophila*', which are currently listed as members of the genus *Pseudomonas* in various databases [4, 6, 14], do not belong to this genus. Results of earlier work have indicated that *P. cissicola* and *P. geniculata* are related to the genera *Xanthomonas* and *Stenotrophomonas,* respectively [17, 30–32]. In the present work, we provide strong evidence based on the phylogenetic branching of these species and the shared presence of CSIs specific for the genera *Xanthomonas*, *Stenotrophomonas* and *Paraburkholderia* by *P. cissicola*, *P. geniculata* and '*P. acidophila*', respectively that these species are members of these other genera and should thus be accordingly reclassified.

In the present work, although we are reclassifying some of the deep-branching and misclassified species of the genus *Pseudomonas* into other genera, the genus *Pseudomonas* still remains a polyphyletic and genetically highly diverse assemblage of microorganisms. In the phylogenetic tree reconstructed in this work (Fig. 1) and earlier studies [8, 15, 16, 22, 24], the remaining species of the genus *Pseudomonas* are comprised of minimally 13–18 main groups (the exact number is difficult to ascertain as it is difficult to reliably determine the boundaries of these clades), in addition to the genera *Azotobacter* and *Azomonas* and multiple isolated species of the genus *Pseudomonas* branching in between these clades (see Fig. 1, Table 1). Of these clades, according to the code governing prokaryotic nomenclature [70], the genus name *Pseudomonas* should be limited only to the members of the Aeruginosa clade (Fig. 1), which contains the type species of this genus, *P. aeruginosa* [1, 71]. However, before this can be accomplished, it will be necessary to reliably delineate the boundaries of the remaining 'Pseudomonas species clades' and transfer species from these clades into other novel genera. In phylogenetic trees, these 'Pseudomonas species clades' are separated by short branches (see Fig. 1), which indicates that it will be difficult to reliably demarcate these clades based solely on their branching in phylogenetic trees. In this context, it should also be noted that the genetic diversity of species of the genus *Pseudomonas* is far greater than indicated by the currently known species. Hesse *et al.* [24] found that of the

1224 uncharacterized genomes of members of the genus *Pseudomonas* from public databases which they examined, 394 were distinct from all other type strains, thus representing potential novel species. In view of this, any proposed classification scheme for the species of the genus *Pseudomonas* should be capable of reliably accommodating other novel species that will continue to be assigned to this genus. In this context, the approach described here, where different monophyletic clades are robustly demarcated on the basis of highly-specific molecular markers with high degree of predictive ability should prove very useful in developing a coherent and reliable classification of this important group of organisms.

We provide below the descriptions of the two novel genera, *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., proposed in this work as well as the emended description of the genus *Thiopseudomonas*. The new name combinations for species, which are part of these genera as well as some other misclassified species *P. cissicola* and *P. geniculata* are also presented below. Additionally, we are also describing '*P. acidophila*' as a novel species within the genus *Paraburkholderia*. Three other species of the genus *Pseudomonas*, which are indicated to be a part of the genus *Halopseudomonas*, '*P. jilinensis*' [72], '*P. saudimassiliensis*' [73] and '*P. yangmingensis*' [74], are also not validly published and type strains for them are not available in any culture collection. For two other species, *Pseudomonas abyssi* [75] and *Pseudomonas profundi* [76], type strains are not readily available from two culture collections. Hence, new name combinations for these species are not proposed here. However, once the strains for these species are deposited in additional culture collections, these species could be described as either novel species or new name combinations in the genus *Halopseudomonas*.

DESCRIPTION OF THE GENUS *HALOPSEUDOMONAS* GEN. NOV.

Halopseudomonas (Ha.lo.pseu.do.mo'nas. Gr. masc. n. *hals* (gen. *halos*), salt; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Halopseudomonas*, salt-loving or tolerating *Pseudomonas*).

Gram-negative, aerobic or facultatively anaerobic, non-spore forming rods. Most species are motile by means of a single polar flagellum. Isolated from a variety of marine sources as well as from desert sand, food waste, soil, air sample, aquatic plants and algae. Members have been reported to grow at temperatures ranging from 5 to 50 °C with optimal growth occurring in the range of 25–37 °C. Optimal pH for growth is between 6 and 10. Most species require between 1 and 10% (w/v) NaCl for optimal growth. Most species are able to utilize Tween 40, Tween 80, succinic acid, sebacic acid, and acetic acid as sole carbon sources. Colonies are circular, convex and white to greyish or pale yellow after 2–3 days of incubation in nutrient agar or other growth media. Most species are positive for alkaline phosphatase, acid phosphatase, esterase (C4), leucine arylamidase, naphthol-AS-BI-phosphohydrolase, catalase and oxidase. The predominant quinone system is ubiquinone-9 (Q-9). Members of this genus form a monophyletic clade in phylogenetic trees based on 16S rDNA and concatenated sequences for several large datasets of proteins. In addition, the members of this genus can be reliably distinguished from all other species of the family *Pseudomonadaceae* by 24 CSIs described in Table 2 found in the following proteins, flagellar protein FlgN, RND family transporter, tRNA cyclic N6 threonylcarbamoyladenosine(37) synthase TcdA, protein BatD, RNA polymerase-associated protein RapA, glyceraldehyde-3-phosphate dehydrogenase (two different CSIs), dephospho-CoA kinase, SpoVR family protein, serine/threonine protein kinase, recombination-associated protein RdgC, OprD family porin, bifunctional diguanylate cyclase/phosphodiesterase, NADPH-dependent 7-cyano-7-deazaguanine reductase QueF, YcgN family cysteine cluster protein, SprT family zinc-dependent metalloprotease, universal stress protein, DUF4892 domain-containing protein, symmetrical bis (5'-nucleosyl)-tetraphosphatase, YaiI/YqxD family protein, diaminopimelate decarboxylase, YifB family Mg chelatase-like AAA ATPase, potassium transporter and LEA type 2 family protein, which in most cases are exclusively shared by either all or most members of this genus. The DNA G+C content for known species is in the range of 57.5–63 mol% and genome sizes range from 3.0 to 4.6 Mbp.

The type species is *Halopseudomonas pertucinogena*.

DESCRIPTION OF *HALOPSEUDOMONAS PERTUCINOGENA* COMB. NOV.

(per.tu.ci.no'.ge.na. N.L. neut. n. *pertucinum*, pertucin, a bacteriocin active against *Bordetella pertussis*; L. suff. *genus -a -um*, producing; from L. v. *gigno*, to produce, give birth to, beget; N.L. fem. adj. *pertucinogena*, intended to mean pertucin producing)

Basonym: *Pseudomonas pertucinogena* Kawai and Yabuuchi 1975 (Approved Lists 1980)

The description of this species is as given by Kawai and Yabuuchi [77] for *Pseudomonas pertucinogena*.

The type strain is ATCC 190= CUG 7832=CIP 106696=DSM 18268=IFO 14163=JCM 11590=LMG 1874=NBRC 14163

DESCRIPTION OF *HALOPSEUDOMONAS LITORALIS* COMB. NOV.

(li.to.ra'lis. L. fem. adj. *litoralis* of or belonging to the seashore).

Basonym: *Pseudomonas litoralis* Pascual *et al.* 2012,

The description of this species is as given by Pascual *et al.* [64] for *Pseudomonas litoralis*.

The type strain is CECT 7670=DSM 26168=KCTC 23093=strain 2SM5

DESCRIPTION OF *HALOPSEUDOMONAS SALEGENS* COMB. NOV.

(sal.e'gens. L. masc. n. *sal* (gen. *salis*), salt; L. pres. part. *egens*, being in need; N.L. part. adj. *salegens*, being in need of salt.)

Basonym: *Pseudomonas salegens* Amoozegar *et al.* 2014.

The description of this species is as given by Amoozegar *et al.* [62] for *Pseudomonas salegens*.

Type strain: CECT 8338=GBPy5=IBRC-M 10762

DESCRIPTION OF *HALOPSEUDOMONAS FORMOSENSIS* COMB. NOV.

(for.mo.sen'sis. N.L. fem. adj. *formosensis* of or pertaining to Formosa (Taiwan), the beautiful island).

Basonym: *Pseudomonas formosensis* Lin *et al.* 2013.

The description of this species is as given by Lin *et al.* [78] for *Pseudomonas formosensis*.

The type strain is BCRC 80437;=CC-CY503;=JCM 18415

DESCRIPTION OF *HALOPSEUDOMONAS BAUZANENSIS* COMB. NOV.

(bau.za.nen'sis. N.L. fem. adj. *bauzanensis* of or belonging to Bauzanum medieval Latin name of Bozen/Bolzano, a city in South Tyrol, Italy, where the species was first isolated).

Basonym: *Pseudomonas bauzanensis* Zhang *et al.* 2011

The description of this species is as given by Zhang *et al.* [79] for *Pseudomonas bauzanensis*.

The type strain is BZ93=CGMCC 1.9095=DSM 22558=LMG 26048

DESCRIPTION OF *HALOPSEUDOMONAS SABULINIGRI* COMB. NOV.

(sa.bu.li.ni'gri. L. neut.n. *sabulum* sand; L. masc. adj. *niger* black; N.L. gen. n. *sabulinigri* of black sand).

Basonym: *Pseudomonas sabulinigri* Kim *et al.* 2009

The description of this species is as given by Kim *et al.* [80] for *Pseudomonas sabulinigri*.

The type strain is DSM 23971=J64=JCM 14963=KCTC 22137.

DESCRIPTION OF *HALOPSEUDOMONAS PACHASTRELLAE* COMB. NOV.

(pa.cha.strel'lae. N.L. gen. n. *pachastrellae*, of *Pachastrella*, the generic name of a sponge)

Basonym: *Pseudomonas pachastrellae* Romanenko *et al.* 2005

The description of this species is as given by Romanenko *et al.* [63] for *Pseudomonas pachastrellae*.

The type strain is CCUG 46540=DSM 17577=JCM 12285=KMM 330=NRIC 583.

DESCRIPTION OF *HALOPSEUDOMONAS GALLAECIENSIS* COMB. NOV.

(gal.lae.ci.en'sis. L. fem. adj. *gallaeciensis*, pertaining to Galicia, Spain, where the type strain was isolated)

Basonym: *Pseudomonas gallaeciensis* Mulet *et al.* 2018

The description of this species is as given by Mulet *et al.* [81] for *Pseudomonas gallaeciensis*.

The type strain is CCUG 67583=LMG 29038=V113.

DESCRIPTION OF *HALOPSEUDOMONAS OCEANI* COMB. NOV.

(o.ce.a'ni. L. gen. n. *oceani*, of the ocean)

Basonym: *Pseudomonas oceani* Wang and Sun 2016

The description of this species is as given by Wang and Sun [67] for *Pseudomonas oceani*.

The type strain is CGMCC 1.15195=DSM 100277=KX 20.

DESCRIPTION OF *HALOPSEUDOMONAS AESTUSNIGRI* COMB. NOV.

(a.es.tus.ni'gri. L. masc. n. *aestus*, tide; L. masc. adj. *niger* black; N.L. gen. n. *aestusnigri*, of black tide).

Basonym: *Pseudomonas aestusnigri* Sánchez *et al.* 2014.

The description of this species is as given by Sánchez *et al.* [82] for *Pseudomonas aestusnigri*.

The type strain is CCUG 64165=CECT 8317=VGXO14.

DESCRIPTION OF *HALOPSEUDOMONAS XINJIANGENSIS* COMB. NOV.

(xin.jiang.en'sis. N.L. fem. adj. *xinjiangensis* pertaining to Xinjiang, in north-west PR China, where the type strain was isolated).

Basonym: *Pseudomonas xinjiangensis* Liu *et al.* 2009

The description of this species is as given by Liu *et al.* [83] for *Pseudomonas xinjiangensis*

The type strain is CCTCC AB 207151=DSM 23391=NRRL B-51270=S3-3.

DESCRIPTION OF *HALOPSEUDOMONAS PELAGIA* COMB. NOV.

(pe.la'gi.a. L. fem. adj. *pelagia* of the sea).

Basonym: *Pseudomonas pelagia* Hwang *et al.* 2009.

The description of this species is as given by Hwang *et al.* [66] for *Pseudomonas pelagia*.

The type strain is DSM 25163=JCM 15562=KCCM 90073=strain CL-AP6.

DESCRIPTION OF *HALOPSEUDOMONAS SALINA* COMB. NOV.

(sa.li'na. N.L. fem. adj. *salina*, salty).

Basonym: *Pseudomonas salina* Zhong *et al.* 2015.

The description of this species is as given by Zhong *et al.* [65] for *Pseudomonas salina*.

The type strain is CGMCC 1.12482=JCM 19469=XCD-X85.

DESCRIPTION OF *HALOPSEUDOMONAS XIAMENENSIS* COMB. NOV.

(xia.men.en'sis N.L. fem. adj. *xiamenensis*, of Xiamen, a district in Fujian, PR China, where the type strain was isolated)

Basonym: *Pseudomonas xiamenensis* Lai and Shao 2008.

The description of this species is as given by Lai and Shao [84] for *Pseudomonas xiamenensis*.

Type strain: C10-2=CGMCC 1.6446=DSM 22326=JCM 13530=MCCC 1A00089

DESCRIPTION OF *ATOPOMONAS* GEN. NOV.

(A.to.po.mo'nas. Gr. masc. adj. *atopos*, having no place, strange; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Atopomonas*, strange monad)

Cells are Gram-stain negative, aerobic, motile by means of a monopolar flagellum, non-spore-forming, chemoheterotrophic and mesophilic [58]. Growth occurs at 15–40 °C (optimum, 30–37 °C), at pH 6.0–8.0 (optimum, pH 7.0) and on R2A agar supplemented with 0–5% NaCl (optimum, 1%). Positive for catalase and oxidase activities. Able to oxidize a variety of substrates including dextrin, glycogen, Tweens 40 and 80, α-D-glucose, maltose, pyruvic acid methyl ester, succinic acid monomethyl ester, acetic acid, α-hydroxybutyric acid, α-ketobutyric acid, DL-lactic acid and propionic acid. The predominant quinone is ubiquinone (Q-9). The other characteristics of this genus are the same as those described by Hameed *et al.* [58] for *Pseudomonas hussainii*. The members of this genus form a separate lineage in phylogenetic trees reconstructed based on the basis of different protein sequences and they are distinguished from all other genera of the family *Pseudomonadaceae* on the basis of the presence of distinctive conserved signature indels identified in the present work that are found in the following proteins (Table 3): glutathione-disulfide reductase, dihydrolipoyl dehydrogenase, UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase, phosphotransferase, citrate (Si)-synthase, HugZ family protein, ATP-binding cassette domain-containing protein, DNA polymerase III subunit beta, cupin domain-containing protein, L,D-transpeptidase family protein, MBL fold metallo-hydrolase, phosphate signalling complex protein PhoU, arginine/ornithine succinyltransferase subunit alpha, excinuclease ABC subunit A, sulfurtransferase complex subunit TusC, potassium/proton antiporter, metalloprotease PmbA, Hsp20 family protein, serine hydrolase, recombination-associated protein RdgC, sulfite exporter TauE/SafE family protein and 2-oxoglutarate dehydrogenase E1.

The type species is *Atopomonas hussainii*

DESCRIPTION OF *ATOPOMONAS HUSSAINII* COMB. NOV.

(hus.sai'ni.i. N.L. gen. masc. n. *hussainii*, named after S. A. Hussain, an Indian ornithologist and avian gut biologist)

Basonym: *Pseudomonas hussainii* Hameed *et al.* 2014 [58]

The description of this species is the same as described by Hameed *et al.* [58] for *Pseudomonas hussainii* and that described above for the genus *Atopomonas*. The DNA G+C content of the type strain is 58.8 mol%.

Type strain: BCRC 80696=CC-AMH-11=JCM 19513

EMENDED DESCRIPTION OF THE GENUS *THIOPSEUDOMONAS* TAN *ET AL.* 2015

(Thi.o.pseu.do.mo'nas Gr. neut. n. *theion*, sulfur (Latin transliteration *thium*); N.L. fem. n. *Pseudomonas*, false monad; from Gr. adj. *pseudo*, false; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Thiopseudomonas*, an organism with a false single unit and sulfur)

Cells are Gram-stain negative, rod-shaped and aerobic or facultatively anaerobic. Includes both motile and non-motile species. Most species require between 0.5 and 6% (w/v) NaCl for optimum growth. Growth occurs at 10–42 °C (optimum, 30–35 °C). Positive for catalase, oxidase, esterase lipase (C8) and leucine arylamidase. The type species of this genus is able to oxidize sulfide anaerobically with nitrate as an electron acceptor. Some species are able to utilize malate as sole carbon source. Indole is not produced. The genome size for the known species is in the range of 2.5–3.10 Mbp and DNA G+C contents range from 47.4 to 59.0 mol%. Members of this genus form a monophyletic clade in phylogenetic trees based on concatenated sequences for several large datasets of proteins. In addition, the members of this genus can be reliably distinguished from all other genera of the family *Pseudomonadaceae* by the shared presence of six CSIs identified in the present work (Table 3) in the following proteins: Si-specific NAD(P)(+) transhydrogenase, esterase, response regulator transcription factor, phosphoribosylamine glycine ligase, C40 family peptidase and ABC transporter ATP-binding protein, which in most

cases are exclusively shared by either all or most members of this genus.

Type species is *Thiopseudomonas denitrificans* Tan *et al.* 2015 [57]

DESCRIPTION OF *THIOPSEUDOMONAS ALKALIPHILA* COMB. NOV.

(al.ka.li'phi.la. N.L. neut. n. *alkali*, soda ash; N.L. adj. *philus -a -um*, friend, loving; from Gr. adj. *philos -ê -on*, loving; N.L. fem. adj. *alkaliphila*, alkaline-loving)

Basonym: *Oblitimonas alkaliphila* Drobish *et al.* 2016 [28]

The description of this species is as provided by Drobish *et al.* [28] for *Oblitimonas alkaliphila*.

Type strain: B4199=CCUG 67636=DSM 100830

DESCRIPTION OF *THIOPSEUDOMONAS CAENI* COMB. NOV.

(cae'ni L. gen. neut. n. *caeni*, of sludge)

Basonym: *Pseudomonas caeni* Xiao *et al.* 2009 [69]

The description of this species is as provided by Xiao *et al.* [69] for *Pseudomonas caeni*.

Type strain: CCTCC AB208156;=DSM 24390=CECT 7778=KCTC 22292=strain HY-14

DESCRIPTION OF *XANTHOMONAS CISSICOLA* COMB. NOV.

(N.L. fem. n. *Cissus*, generic name of a flowering plant; L. masc./fem. suff. *-cola*, dweller; from L. masc./fem. n. *incola*; N.L. n. *cissicola*, *Cissus* dweller)

Basonym: *Pseudomonas cissicola* (Takimoto 1939) Burkholder 1948 (Approved Lists 1980)

The description of this species is the same as given by Takimoto (1939) Burkholder [85] for *Pseudomonas cissicola*.

The type strain is ATCC 33616=CCUG 18839=CFBP 2432=CIP 106723=DSM 21306=JCM 13362=NCPPB 2982.

DESCRIPTION OF *STENOTROPHOMONAS GENICULATA* COMB. NOV.

(ge.ni.cu.la'ta. L. fem. adj. *geniculata*, jointed)

Basonym: *Pseudomonas geniculata* (Wright 1895) Chester 1901 (Approved Lists 1980).

The description of this species is the same as given by Chester [86] *Pseudomonas geniculata*.

The type strain is ATCC 19374=JCM 13324=LMG 2195=NCIB 9428=NCIMB 9428.

DESCRIPTION OF *PARABURKHOLDERIA ACIDICOLA* SP. NOV.

(a.ci.di'co.la. L. neut. adj. *acidum*, an acid; L. masc./fem. suff. *-cola*, dweller; from L. masc./fem. n. *incola*, an inhabitant; N.L. masc./fem. n. *acidicola*, an inhabitant of an acidic environment)

The description of this species is the same as given by Imada *et al.* [87] for '*Pseudomonas acidophila*'. Horsman *et al.* [32], based on their sequencing of the type strain of '*Pseudomonas acidophila*' (ATCC 31363), have suggested that this species should be reclassified as '*Paraburkholderia acidophila*'. However, as there is already a validly named species *Paraburkholderia acidiphila* [88], we have chosen to name this species as *Paraburkholderia acidicola*. Isolated from soil in Japan. Cells are Gram-negative and rod-shaped, strictly aerobic, motile with a polar flagellum or flagella and catalase-positive and oxidase-negative. Optimal growth is observed at 25–30 °C and pH 4.5–6.0. This species is best known for its production of beta lactam antibiotic sulfazecin and bulgecins [89].

The accession numbers for the 16S rRNA gene sequence and the draft genome sequence for the type strain G-6302 of *Paraburkholderia acidicola* are MW628182 and ASM236231v1 (GCF_002362315.1), respectively.

Type strain: G-6302=ATCC 31363=BCRC 13035=IFO: 13774=FERM-P No. 4344.

Funding information
This work was supported by a Discovery research grant (RGPIN-2019-06397) and an Alliance grant from the Natural Science and Engineering Research Council of Canada and research funding from the Ontario Ministry of Research Innovation and Science.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Migula W**. Uber Ein Neues system der Bakterien. Arb Bakt Inst Kar1sruhe. 1894:1.
2. **Palleroni NJ**. *Pseudomonas*. John Wiley and Sons in Association with Bergey's Manual Trust, 2015, pp. 1–105.
3. **Peix A, Ramírez-Bahena M-H, Velázquez E**. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect Genet Evol* 2009;9:1132–1147.
4. **Parte AC**. LPSN - the list of prokaryotic names with standing in nomenclature. *Int J Syst Evol Microbiol* 2018;68:1825–1829.
5. **Garrity GM**. NAMESFORLIFE browser tool takes expertise out the database and puts it right in the browser. *Microbiology Today* 2010;2:9.
6. **Names for Life Server**. 2021. https://www.namesforlife.com/10.1601/tx.2551
7. **Palleroni NJ**. Genus I. *Pseudomonas* Migula 1894. **Brenner D, Krieg N, Staley J** and **Garrity G** (eds). In: *Bergey's Manual of Systematic Bacteriology (The Proteobacteria), part B (The Gammaproteobacteria)*, 2nd edn. New York: Springer; 2005. pp. 323–379.
8. **Peix A, Ramírez-Bahena M-H, Velázquez E**. The current status on the taxonomy of *Pseudomonas* revisited: An update. *Infect Genet Evol* 2018;57:106–116.

9. **Rossi E**, **La Rosa R**, **Bartell JA**, **Marvig RL**, **Haagensen JAJ**, ***et al***. *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat Rev Microbiol* 2021;19:331–342.
10. **Winsor GL**, **Griffiths EJ**, **Lo R**, **Dhillon BK**, **Shay JA**, ***et al***. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 2016;44:D653–D646.
11. **Lund-Palau H**, **Turnbull AR**, **Bush A**, **Bardin E**, **Cameron L**, ***et al***. *Pseudomonas aeruginosa* infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches. *Expert Rev Respir Med* 2016;10:685–697.
12. **Xin X-F**, **Kvitko B**, **He SY**. *Pseudomonas syringae*: what it takes to be a pathogen. *Nat Rev Microbiol* 2018;16:316–328.
13. **Mulet M**, **David Z**, **Nogales B**, **Bosch R**, **Lalucat J**, ***et al***. *Pseudomonas* diversity in crude-oil-contaminated intertidal sand samples obtained after the Prestige oil spill. *Appl Environ Microbiol* 2011;77:1076–1085.
14. **Sayers EW**, **Agarwala R**, **Bolton EE**, **Brister JR**, **Canese K**, ***et al***. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2019;47:D23–D28.
15. **García-Valdés E**, **Lalucat J**. *Pseudomonas*: Molecular phylogeny and current taxonomy. In: *Pseudomonas: Molecular and Applied Biology*. pp. 1–23.
16. **Lalucat J**, **Mulet M**, **Gomila M**, **García-Valdés E**. Genomics in bacterial taxonomy: Impact on the genus *Pseudomonas*. *Genes (Basel)* 2020;11:139.
17. **Anzai Y**, **Kim H**, **Park JY**, **Wakabayashi H**, **Oyaizu H**. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int J Syst Evol Microbiol* 2000;50 Pt 4:1563–1589.
18. **Palleroni NJ**. The *Pseudomonas* story. *Environ Microbiol* 2010;12:1377–1383.
19. **Sneath PH**, **Stevens M**, **Sackin MJ**. Numerical taxonomy of *Pseudomonas* based on published records of substrate utilization. *Antonie van Leeuwenhoek* 1981;47:423–448.
20. **Gomila M**, **Peña A**, **Mulet M**, **Lalucat J**, **García-Valdés E**. Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* 2015;6:214.
21. **Mulet M**, **Lalucat J**, **García-Valdés E**. DNA sequence-based analysis of the *Pseudomonas* species. *Environ Microbiol* 2010;12:1513–1530.
22. **Garrido-Sanz D**, **Meier-Kolthoff JP**, **Göker M**, **Martín M**, **Rivilla R**, ***et al***. Correction: Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PLoS One* 2016;11:e0153733.
23. **Bueno-Gonzalez V**, **Brady C**, **Denman S**, **Allainguillaume J**, **Arnold D**. *Pseudomonas kirkiae* sp. nov., a novel species isolated from oak in the United Kingdom, and phylogenetic considerations of the genera *Pseudomonas*, *Azotobacter* and *Azomonas*. *Int J Syst Evol Microbiol* 2020;70:2426–2434.
24. **Hesse C**, **Schulz F**, **Bull CT**, **Shaffer BT**, **Yan Q**, ***et al***. Genome-based evolutionary history of *Pseudomonas* spp. *Environ Microbiol* 2018;20:2142–2159.
25. **Özen AI**, **Ussery DW**. Defining the *Pseudomonas* genus: where do we draw the line with *Azotobacter*? *Microb Ecol* 2012;63:239–248.
26. **Young JM**, **Park DC**. Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. *Int J Syst Evol Microbiol* 2007;57:2894–2901.
27. **Rediers H**, **Vanderleyden J**, **De Mot R**. *Azotobacter vinelandii*: a *Pseudomonas* in disguise? *Microbiology (Reading)* 2004;150:1117–1119.
28. **Drobish AM**, **Emery BD**, **Whitney AM**, **Lauer AC**, **Metcalfe MG**, ***et al***. *Oblitimonas alkaliphila* gen. nov., sp. nov., in the family *Pseudomonadaceae*, recovered from a historical collection of previously unidentified clinical strains. *Int J Syst Evol Microbiol* 2016;66:3063–3070.
29. **Parks DH**, **Chuvochina M**, **Waite DW**, **Rinke C**, **Skarshewski A**, ***et al***. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;39:996–1004.
30. **Hu FP**, **Young JM**, **Stead DE**, **Goto M**. Transfer of *Pseudomonas cissicola* (Takimoto 1939) Burkholder 1948 to the genus *Xanthomonas*. *Int J Syst Evol Microbiol* 1997;47:228–230.
31. **Cutino-Jimenez AM**, **Menck CFM**, **Cambas YT**, **Diaz-Perez JC**. Protein signatures to identify the different genera within the *Xanthomonadaceae* family. *Braz J Microbiol* 2020;51:1515–1526.
32. **Horsman ME**, **Marous DR**, **Li R**, **Oliver RA**, **Byun B**, ***et al***. Whole-genome shotgun sequencing of two beta-proteobacterial species in search of the bulgecin biosynthetic cluster. *ACS Chem Biol* 2017;12:2552–2557.
33. **Wang Z**, **Wu M**. A phylum-level bacterial phylogenetic marker database. *Mol Biol Evol* 2013;30:1258–1262.
34. **Gupta RS**, **Patel S**, **Saini N**, **Chen S**. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int J Syst Evol Microbiol* 2020;70:5753–5798.
35. **Adeolu M**, **Alnajar S**, **Naushad S**, **Gupta S**. 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. *Int J Syst Evol Microbiol* 2016;66:5575–5599.
36. **Eddy SR**. Accelerated Profile HMM Searches. *PLoS Comput Biol* 2011;7:e1002195.
37. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, ***et al***. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
38. **Capella-Gutiérrez S**, **Silla-Martínez JM**, **Gabaldón T**. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
39. **Whelan S**, **Goldman N**. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001;18:691–699.
40. **Price MN**, **Dehal PS**, **Arkin AP**. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490.
41. **Guindon S**, **Dufayard JF**, **Lefort V**, **Anisimova M**, **Hordijk W**, ***et al***. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
42. **Stamatakis A**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
43. **Kumar S**, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K**. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 2018;35:1547–1549.
44. **Naushad HS**, **Gupta RS**. Phylogenomics and molecular signatures for species from the plant pathogen-containing order *Xanthomonadales*. *PLoS One* 2013;8:e55216.
45. **Patel S**, **Gupta RS**. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 2020;70:406–438.
46. **Gupta RS**. Identification of conserved indels that are useful for classification and evolutionary studies. *Methods in Microbiology* 2014;41:153–182.
47. **Alnajar S**, **Gupta RS**. Phylogenomics and comparative genomic studies delineate six main clades within the family *Enterobacteriaceae* and support the reclassification of several polyphyletic members of the family. *Infect Genet Evol* 2017;54:108–127.
48. **Jeanmougin F**, **Thompson JD**, **Gouy M**, **Higgins DG**, **Gibson TJ**. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998;23:403–405.
49. **Gupta RS**. Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol Rev* 2016;40:520–553.

50. **Parker CT**, **Lyons CM**, **Roston GR**, **Garrity GM**. Systems and Methods for Automatically Identifying and Linking Names in Digital Resources (Continuation). patent U.S. Patent Grant No. 10 204 168. 2019.
51. **Felsenstein J**. *Inferring Phylogenies*. Sunderland, Mass: Sinauer Associates, Inc; 2004.
52. **Gupta RS**. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 1998;62:1435–1491.
53. **Baldauf SL**. Phylogeny for the faint of heart: a tutorial. *Trends Genet* 2003;19:345–351.
54. **Sawana A**, **Adeolu M**, **Gupta RS**. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet* 2014;5:429.
55. **Rokas A**, **Holland PW**. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* 2000;15:454–459.
56. **Aldridge P**, **Karlinsey J**, **Hughes KT**. The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Mol Microbiol* 2003;49:1333–1345.
57. **Tan WB**, **Jiang Z**, **Chen C**, **Yuan Y**, **Gao LF**, ***et al***. *Thiopseudomonas denitrificans* gen. nov., sp. nov., isolated from anaerobic activated sludge. *Int J Syst Evol Microbiol* 2015;65:225–229.
58. **Hameed A**, **Shahina M**, **Lin S-Y**, **Liu Y-C**, **Young C-C**. *Pseudomonas hussainii* sp. nov., isolated from droppings of a seashore bird, and emended descriptions of *Pseudomonas pohangensis*, *Pseudomonas benzenivorans* and *Pseudomonas segetis*. *Int J Syst Evol Microbiol* 2014;64:2330–2337.
59. **Naushad S**, **Adeolu M**, **Wong S**, **Sohail M**, **Schellhorn HE**, ***et al***. A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie van Leeuwenhoek* 2015;107:467–485.
60. **Desnoues N**, **Lin M**, **Guo X**, **Ma L**, **Lopez R**, ***et al***. Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology (Reading)* 2003;149:2251–2262.
61. **Silby MW**, **Winstanley C**, **Godfrey SA**, **Levy SB**, **Jackson RW**. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 2011;35:652–680.
62. **Amoozegar MA**, **Shahinpei A**, **Sepahy AA**, **Makhdoumi-Kakhki A**, **Seyedmahdi SS**, ***et al***. *Pseudomonas salegens* sp. nov., a halophilic member of the genus *Pseudomonas* isolated from a wetland. *Int J Syst Evol Microbiol* 2014;64:3565–3570.
63. **Romanenko LA**, **Uchino M**, **Falsen E**, **Frolova GM**, **Zhukova NV**, ***et al***. *Pseudomonas pachastrellae* sp. nov., isolated from a marine sponge. *Int J Syst Evol Microbiol* 2005;55:919–924.
64. **Pascual J**, **Lucena T**, **Ruvira MA**, **Giordano A**, **Gambacorta A**, ***et al***. *Pseudomonas litoralis* sp. nov., isolated from Mediterranean seawater. *Int J Syst Evol Microbiol* 2012;62:438–444.
65. **Zhong ZP**, **Liu Y**, **Hou TT**, **Liu HC**, **Zhou YG**, ***et al***. *Pseudomonas salina* sp. nov., isolated from a salt lake. *Int J Syst Evol Microbiol* 2015;65:2846–2851.
66. **Hwang CY**, **Zhang GI**, **Kang SH**, **Kim HJ**, **Cho BC**. *Pseudomonas pelagia* sp. nov., isolated from a culture of the Antarctic green alga *Pyramimonas gelidicola*. *Int J Syst Evol Microbiol* 2009;59:3019–3024.
67. **Wang MQ**, **Sun L**. *Pseudomonas oceani* sp. nov., isolated from deep seawater. *Int J Syst Evol Microbiol* 2016;66:4250–4255.
68. **Dobritsa AP**, **Samadpour M**. Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *Int J Syst Evol Microbiol* 2019;69:2057–2063.
69. **Xiao Y-P**, **Hui W**, **Wang Q**, **Roh SW**, **Shi X-Q**, ***et al***. *Pseudomonas caeni* sp. nov., a denitrifying bacterium isolated from the sludge of an anaerobic ammonium-oxidizing bioreactor. *Int J Syst Evol Microbiol* 2009;59:2594–2598.
70. **Parker CT**, **Tindall BJ**, **Garrity GM**. International Code of Nomenclature of prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1-S111.
71. **Skerman VBD**, **McGowan V**, **Sneath PHA**. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980;30:225–420.
72. **Wang JW**, **Cai M**, **Nie Y**, **Hu B**, **Yang Y**, ***et al***. *Pseudomonas jilinensis* sp. nov., isolated from oil production water of Jilin Oilfield in China. *Curr Microbiol* 2020;77:688–694.
73. **Azhar EI**, **Papadioti A**, **Bibi F**, **Ashshi AM**, **Raoult D**, ***et al***. *Pseudomonas saudimassiliensis* sp. nov. a new bacterial species isolated from air samples in the urban environment of Makkah, Saudi Arabia. *New Microbes New Infect* 2017;16:43–44.
74. **Wong BT**, **Lee DJ**. *Pseudomonas yangmingensis* sp. nov., an alkaliphilic denitrifying species isolated from a hot spring. *J Biosci Bioeng* 2014;117:71–74.
75. **Wei Y**, **Mao H**, **Xu Y**, **Zou W**, **Fang J**, ***et al***. *Pseudomonas abyssi* sp. nov., isolated from the abyssopelagic water of the Mariana Trench. *Int J Syst Evol Microbiol* 2018;68:2462–2467.
76. **Sun J**, **Wang W**, **Ying Y**, **Zhu X**, **Liu J**, ***et al***. *Pseudomonas profundi* sp. nov., isolated from deep-sea water. *Int J Syst Evol Microbiol* 2018;68:1776–1780.
77. **Kawai Y**, **Yabuuchi E**. *Pseudomonas pertucinogena* sp. nov., an organism previously misidentified as *Bordetella pertussis*. *Int J Syst Bacteriol* 1975;25:317–323.
78. **Lin SY**, **Hameed A**, **Liu YC**, **Hsu YH**, **Lai WA**, ***et al***. *Pseudomonas formosensis* sp. nov., a gamma-proteobacteria isolated from food-waste compost in Taiwan. *Int J Syst Evol Microbiol* 2013;63:3168–3174.
79. **Zhang DC**, **Liu HC**, **Zhou YG**, **Schinner F**, **Margesin R**. *Pseudomonas bauzanensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* 2011;61:2333–2337.
80. **Kim K-H**, **Roh SW**, **Chang H-W**, **Nam Y-D**, **Yoon J-H**, ***et al***. *Pseudomonas sabulinigri* sp. nov., isolated from black beach sand. *Int J Syst Evol Microbiol* 2009;59:38–41.
81. **Mulet M**, **Sánchez D**, **Rodríguez AC**, **Nogales B**, **Bosch R**, ***et al***. *Pseudomonas gallaeciensis* sp. nov., isolated from crude-oil-contaminated intertidal sand samples after the Prestige oil spill. *Syst Appl Microbiol* 2018;41:340–347.
82. **Sánchez D**, **Mulet M**, **Rodríguez AC**, **David Z**, **Lalucat J**, ***et al***. *Pseudomonas aestusnigri* sp. nov., isolated from crude oil-contaminated intertidal sand samples after the Prestige oil spill. *Syst Appl Microbiol* 2014;37:89–94.
83. **Liu M**, **Luo X**, **Zhang L**, **Dai J**, **Wang Y**, ***et al***. *Pseudomonas xinjiangensis* sp. nov., a moderately thermotolerant bacterium isolated from desert sand. *Int J Syst Evol Microbiol* 2009;59:1286–1289.
84. **Lai Q**, **Shao Z**. *Pseudomonas xiamenensis* sp. nov., a denitrifying bacterium isolated from activated sludge. *Int J Syst Evol Microbiol* 2008;58:1911–1915.
85. **Burkholder WH**. Bacteria as plant pathogens. *Annu Rev Microbiol* 1948;2 (1 vol.):389–412.
86. **Chester FD**. *A Manual of Determinative Bacteriology*. New York: The MacMillan Co; 1901.
87. **Imada I**, **Nishinomiya KK**, **Sakai AM**, **Asaik MT**. Antibiotic G-6302. Patent US Patent 4,229,436. 1978. 1978.
88. **Gao ZH**, **Zhang QM**, **Lv YY**, **Wang YQ**, **Zhao BN**, ***et al***. *Paraburkholderia acidiphila* sp. nov., *Paraburkholderia acidisoli* sp. nov. and *Burkholderia guangdongensis* sp. nov., isolated from forest soil, and reclassification of *Burkholderia ultramafica* as *Paraburkholderia ultramafica* comb. nov. *Int J Syst Evol Microbiol* 2021;71.
89. **Imada A**, **Kitano K**, **Kintaka K**, **Muroi M**, **Asai M**. Sulfazecin and isosulfazecin, novel beta-lactam antibiotics of bacterial origin. *Nature* 1981;289:590–591.

CHAPTER 3

Phylogenomics Studies and Molecular Markers Reliably Demarcate Genus *Pseudomonas sensu stricto* and Twelve other *Pseudomonadaceae* Species Clades Representing Novel and Emended Genera.

This chapter describes comprehensive phylogenomic studies, consistently identifying 13 major clades/groups containing different *Pseudomonas* species within the Aeruginosa lineage. To support the distinctiveness of these observed clades, this study presents the identification of 98 CSIs that are highly specific to these different clades. These CSIs serve as unique molecular markers distinguishing different clades and provide strong independent evidence for the genetic cohesiveness of these clades. Based on the clade-specific CSIs, robust phylogenetic analysis, and other genomic similarity indices (AAI and POCP), this study reclassifies the distinct species clades into seven novel genera and five emended genera. Additionally, the findings support restricting the genus *Pseudomonas* only to the species within the Aeruginosa clade containing the type species *P. aeruginosa*. My contributions to this chapter include constructing phylogenetic trees, identifying CSIs, conducting genomic similarity analyses, drafting the manuscript, and producing all main and supplemental figures and tables.

Due to space constraints, supplementary figures and tables are not included in this chapter but are available alongside the entire manuscript at:

Rudra, B., & Gupta, R. S. (2024). Front. Microbiol., 14, 1273665.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 05 January 2024
DOI 10.3389/fmicb.2023.1273665

OPEN ACCESS

EDITED BY
Iain Sutcliffe,
Northumbria University, United Kingdom

REVIEWED BY
Margarita Gomila,
University of the Balearic Islands, Spain
Carrie Brady,
University of the West of England,
United Kingdom
Aharon Oren,
Hebrew University of Jerusalem, Israel

*CORRESPONDENCE
Radhey S. Gupta
✉ gupta@mcmaster.ca

RECEIVED 07 August 2023
ACCEPTED 17 November 2023
PUBLISHED 05 January 2024

CITATION
Rudra B and Gupta RS (2024) Phylogenomics studies and molecular markers reliably demarcate genus *Pseudomonas sensu stricto* and twelve other *Pseudomonadaceae* species clades representing novel and emended genera.
Front. Microbiol. 14:1273665.
doi: 10.3389/fmicb.2023.1273665

Phylogenomics studies and molecular markers reliably demarcate genus *Pseudomonas sensu stricto* and twelve other *Pseudomonadaceae* species clades representing novel and emended genera

Bashudev Rudra and Radhey S. Gupta*

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada

Genus *Pseudomonas* is a large assemblage of diverse microorganisms, not sharing a common evolutionary history. To clarify their evolutionary relationships and classification, we have conducted comprehensive phylogenomic and comparative analyses on 388 *Pseudomonadaceae* genomes. In phylogenomic trees, *Pseudomonas* species formed 12 main clusters, apart from the "Aeruginosa clade" containing its type species, *P. aeruginosa*. In parallel, our detailed analyses on protein sequences from *Pseudomonadaceae* genomes have identified 98 novel conserved signature indels (CSIs), which are uniquely shared by the species from different observed clades/groups. Six CSIs, which are exclusively shared by species from the "Aeruginosa clade," provide reliable demarcation of this clade corresponding to the genus *Pseudomonas sensu stricto* in molecular terms. The remaining 92 identified CSIs are specific for nine other *Pseudomonas* species clades and the genera *Azomonas* and *Azotobacter* which branch in between them. The identified CSIs provide strong independent evidence of the genetic cohesiveness of these species clades and offer reliable means for their demarcation/circumscription. Based on the robust phylogenetic and molecular evidence presented here supporting the distinctness of the observed *Pseudomonas* species clades, we are proposing the transfer of species from the following clades into the indicated novel genera: Alcaligenes clade – *Aquipseudomonas* gen. nov.; Fluvialis clade – *Caenipseudomonas* gen. nov.; Linyingensis clade – *Geopseudomonas* gen. nov.; Oleovorans clade – *Ectopseudomonas* gen. nov.; Resinovorans clade – *Metapseudomonas* gen. nov.; Straminea clade – *Phytopseudomonas* gen. nov.; and Thermotolerans clade – *Zestomonas* gen. nov. In addition, descriptions of the genera *Azomonas*, *Azotobacter*, *Chryseomonas*, *Serpens*, and *Stutzerimonas* are emended to include information for the CSIs specific for them. The results presented here should aid in the development of a more reliable classification scheme for *Pseudomonas* species.

KEYWORDS

***Pseudomonas* classification, phylogenomic and comparative genomic analyses, conserved signature indels (CSIs), molecular markers specific for *Pseudomonas* species clades/groups, proposals for reclassification of *Pseudomonas* species into novel genera**

Introduction

Genus *Pseudomonas* (Migula, 1894) is a large assemblage of motile, rod-shaped, aerobic, non-spore forming, Gram-negative bacteria, generally containing one or more polar flagella that assist in their movement (Palleroni, 2005, 2015). The members of this genus presently contain >300 species with validly published names (Parte et al., 2020), and they span enormous genetic and metabolic diversity, inhabiting diverse niches and environments including soil, water, plants and animal tissues (Peix et al., 2009; Palleroni, 2015). Its members include species which are opportunistic pathogens of humans, animals, and plants, and other species of economic and ecological significance (Palleroni, 2005; Lund-Palau et al., 2016; Winsor et al., 2016; Xin et al., 2018; Rossi et al., 2021). The best studied species from this genus, which is also its nomenclature type (Migula, 1894; Skerman et al., 1980), is *Pseudomonas aeruginosa*, which is an opportunistic human pathogen capable of causing a wide array of life-threatening acute and chronic diseases (Stover et al., 2000; Planquette et al., 2013). Despite the clinical and environmental importance of *Pseudomonas* species, evolutionary relationships among the members of this genus are not clearly understood (Anzai et al., 2000; Peix et al., 2009; Palleroni, 2015; García-Valdés and Lalucat, 2016; Jun et al., 2016; Passarelli-Araujo et al., 2022). In different phylogenetic and genomic studies on *Pseudomonas* species, members of this genus consistently form multiple clades, which are unrelated to each other (i.e., not evolved from a common ancestor) (Peix et al., 2009; Gomila et al., 2015; Jun et al., 2016; Hesse et al., 2018; Peix et al., 2018; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021; Lalucat et al., 2022). Additionally, in these trees, species from several genera including *Azomonas*, *Azotobacter* and *Chryseomonas* branch in between *Pseudomonas* species, making this genus polyphyletic (Jun et al., 2016; Hesse et al., 2018; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021; Lalucat et al., 2022). In recent work, a large number of *Pseudomonas* species, which generally branched outside the main cluster of *Pseudomonas* species, have been reclassified into several novel genera (*viz. Atopomonas, Halopseudomonas* and *Stutzerimonas*) (Rudra and Gupta, 2021; Lalucat et al., 2022), or in other existing genera (*viz. Chryseomonas*, *Stenotrophomonas*, *Thiopseudomonas* and *Xanthomonas*) (Holmes et al., 1987; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021).

Importantly, in all constructed phylogenomic trees, the type species *P. aeruginosa*, along with a limited number of other species, forms a distinct clade referred to as the "Aeruginosa clade" (Jun et al., 2016; Hesse et al., 2018; Peix et al., 2018; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). The remainder (>95%) of the *Pseudomonas* species group into 12–18 main clusters, some of which are referred to as the Alcaligenes, Anguilliseptica, Flexibilis, Fluorescens, Kuykendallii, Linyingensis, Lutea, Massiliensis, Oleovorans, Oryzihabitans, Pertucinogena, Putida, Resinovorans, Rhizosphaerae, Straminea, Stutzeri and Syringae clades, named after one of the species from each of these clusters (Palleroni, 2015; Hesse et al., 2018; Peix et al., 2018; Girard et al., 2021; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021; Lalucat et al., 2022). Species from the Pertucinogena and Stutzeri clusters were recently reclassified into the genera *Halopseudomonas* and *Stutzerimonas*, respectively (Rudra and Gupta, 2021; Lalucat et al., 2022). Of these species' clades, according to the Code governing the nomenclature of Prokaryotes (Oren et al., 2023), the "Aeruginosa clade," which contains the type species *P. aeruginosa*, constitute the genus *Pseudomonas sensu stricto*. It is generally recognized that the species from clades other than the "Aeruginosa clade," should be reclassified into novel genera (Hesse et al., 2018; Peix et al., 2018; Girard et al., 2021; Rudra and Gupta, 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). This task requires that the boundaries of different *Pseudomonas* species clades, including the "Aeruginosa clade," are reliably demarcated so that any proposed reclassification is stable. Different *Pseudomonas* species clades are presently identified primarily based on the clustering of species in phylogenetic trees. However, the numbers of observed species clusters as well as the species grouping within them often vary in different phylogenetic studies (Hesse et al., 2018; Girard et al., 2021; Rudra and Gupta, 2021; Lalucat et al., 2022; Rudra et al., 2022), which makes it difficult to reliably demarcate the boundaries of these clades.

The availability of whole genome sequences is enabling construction of more reliable phylogenetic trees based on large dataset of genes/proteins (Parks et al., 2018). Additionally, the genome sequences also provide an important resource for identification of novel molecular markers, such as conserved signature indels (CSIs), which are uniquely shared characteristics of different monophyletic clades of organisms. Due to their clade specificities, these novel molecular synapomorphies are providing robust means for the demarcation of different observed species clades/taxa in molecular terms (Gupta et al., 2013; Gupta, 2014; Adeolu et al., 2016; Gupta et al., 2020). The use of these markers in conjunction with phylogenomic analyses has recently led to the development of a reliable classification scheme for members of the highly polyphyletic genus *Bacillus* (Gupta et al., 2020). Genome sequences are now available for >300 *Pseudomonas* species in the NCBI genome database[1] (Sayers et al., 2019). With the objective of clarifying evolutionary relationships and classification of *Pseudomonas* species, we have conducted comprehensive phylogenomic and molecular marker-based studies on their genome sequences. In two genome scale phylogenetic trees constructed in this study, *Pseudomonas* species formed approximately 13 main clades, like those seen in earlier work (Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). In parallel, our detailed studies on protein sequences from *Pseudomonas* genomes have identified 98 novel CSIs which are unique characteristics of the species from different observed clades. Based on these CSIs, species from the "Aeruginosa clade" (i.e., genus *Pseudomonas sensu stricto*), 10 other *Pseudomonas* species clades, and the genera *Azomonas* and *Azotobacter*, can now be reliably demarcated based on multiple uniquely shared molecular characteristics. Based on the strong evidence obtained from our phylogenomic studies and identified molecular markers, we are proposing the reclassification of *Pseudomonas* species from the following clades, *viz.* Alcaligenes, Fluvialis, Linyingensis, Oleovorans, Resinovorans, Straminea, and Thermotolerans, into seven novel genera. In addition, we are also emending the descriptions of the genera *Azomonas, Azotobacter, Chryseomonas*, *Serpens* and *Stutzerimonas* to include information for the diagnostic CSIs for these genera.

1 https://www.ncbi.nlm.nih.gov/genome/

Methods

Construction of phylogenetic trees

Genome sequences were downloaded from the NCBI for 342 named *Pseudomonas* species and 46 sequences from other *Pseudomonadaceae* genera available as of December 16, 2022, in the database. Each species is represented in the tree by a single genomic sequence, which is generally of the type strain, when available. Based on these genome sequences, a rooted phylogenetic tree was constructed based on concatenated sequences of 118 conserved proteins that are a part of the phyloeco set for the class *Gammaproteobacteria* (Wang and Wu, 2013) (listed in Supplementary Table S1). Genome sequences for *Moraxella bovoculi* and *M. bovis* were included in this dataset for rooting purposes. Another comprehensive phylogenetic tree was constructed based on the core proteins from the genomes of *Pseudomonadaceae* species. This latter tree was based on genome sequences for 174 species, which included most of the species from the other main clades of *Pseudomonas* species, but only 41 divergent species from the Fluorescens superclade (lineage). Trees were constructed using an internally developed pipeline described in earlier work (Adeolu et al., 2016; Gupta et al., 2020; Rudra and Gupta, 2021; Saini and Gupta, 2021). Briefly, the CD-HIT program (Li and Godzik, 2006; Fu et al., 2012) was used to identify protein families (or homologs of different proteins) where the proteins were present in at least 80% of the genomes in the dataset and they shared at least 50% of sequence length and identity. The Clustal Omega program (Sievers et al., 2011) was then used to generate multiple sequence alignments (MSA) of the proteins. These MSAs were converted into profile Hidden Markov Models (HMMs) using HMMer 3-1b2 (Eddy, 2011), which were then used to search for other members of the protein families in the input genomes. These analyses identified 1,503 protein families meeting the stated criteria (also listed in Supplementary Table S1). The sequence alignments of these proteins were trimmed using TrimAl program (Capella-Gutiérrez et al., 2009) to remove poorly aligned sections prior to their concatenation. The concatenated sequence alignment for the phyloeco set of proteins for *Gammaproteobacteria* was created similarly using the published profile HMMs for these proteins (Wang and Wu, 2013). The concatenated sequence alignments used for the construction of phyloeco and the core genome trees consisted of 42,362 and 494,143 amino acid (aa) positions, respectively. Using these alignments, maximum likelihood (ML) trees were initially constructed using FastTree 2 (Price et al., 2010) with the Whelan and Goldman (2001) model of protein sequence evolution. The resulting trees were optimized with RAxML 8 (Stamatakis, 2014) and to obtain the Shimodaira-Hasegawa (SH) statistical support values, which are similar to the bootstrap scores, for different nodes. The trees were labeled and formatted using MEGA X (Kumar et al., 2018). The percentage of conserved proteins (POCP) and average amino acid identity (AAI) for different pairs of genomes were calculated as described by Thompson et al. (2013) and Qin et al. (2014).

Identification of conserved signature indels

Identification of CSIs was carried out by similar procedures as described in earlier work (Gupta, 2014, 2016; Gupta et al., 2020). Briefly, local BLASTp searches were carried out on protein sequences from the genomes of several *Pseudomonas* species representing different clades of interest and other outgroup species. Based on these BLAST searches, sequences of high scoring homologs (E value <1e-20) of different proteins were retrieved for several species (generally between 4 to 12) from the group of interest, and 10–15 species from other *Pseudomonas* clades or other *Pseudomonadaceae* genera. Multiple sequence alignments for the proteins were created using Clustal X 2.1 program (Jeanmougin et al., 1998). Alignments were visually examined for insertions or deletions of fixed length that were present in conserved regions (i.e., flanked on both sides by minimally 5–6 conserved aa residues in the neighboring 40–50 aa), and which were only found in the *Pseudomonas* species from the clade of interest. The indels which were not present in conserved regions were not further considered. The query sequences consisting of the conserved indels and their flanking 30–40 aa on each side were subjected to a second BLASTp search against the NCBI nr database and the top 250–500 hits were evaluated to determine the group specificities of the CSIs. Based on these results, indels which were specific for different clades of *Pseudomonas* were formatted using the SIG_CREATE and SIG_STYLE programs (Gupta, 2014, 2016). Due to space constraints, sequence information is shown for only a limited number of species in the main figures. However, unless otherwise indicated the CSIs reported here are specifically found in different named *Pseudomonas* species from the indicated groups. More detailed information for different CSIs is provided in the Supplemental Data files.

Results

Phylogenomic analyses of *Pseudomonas* and related species

To understand the interspecies relationships among different *Pseudomonadaceae* species whose genomes were available in the NCBI as of December 16, 2022, two genome-scale phylogenetic trees were constructed. The first of these trees shown in Figure 1 (Supplementary Figure S1), which will be referred to as the phyloeco tree, is based on concatenated sequences for 118 conserved proteins, which comprise the phyloeco set for the class *Gammaproteobacteria* (Wang and Wu, 2013). Another comprehensive tree constructed is a core genome (protein) tree based on 1,503 proteins which are shared by at least 80% of the input *Pseudomonadaceae* species. This latter tree included only representative species (41) from the Fluorescens superclade (lineage), which is not the focus of this study. In both constructed trees, most observed nodes are supported with 100% SH values (like bootstrap scores) indicating that the observed evolutionary relationships are reliable.

The overall branching and grouping of *Pseudomonadaceae* species in different clusters in both the phyloeco (Figure 1) and the core protein tree (Supplementary Figure S2) is nearly identical, and it is similar to that observed in our earlier work (Rudra and Gupta, 2021), and other phylogenetic studies (Gomila et al., 2015; Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020; Girard et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). In both these trees, *Pseudomonas* species formed several distinct clades/groups, and species from the genera *Azomonas* and *Azotobacter* consistently branched between

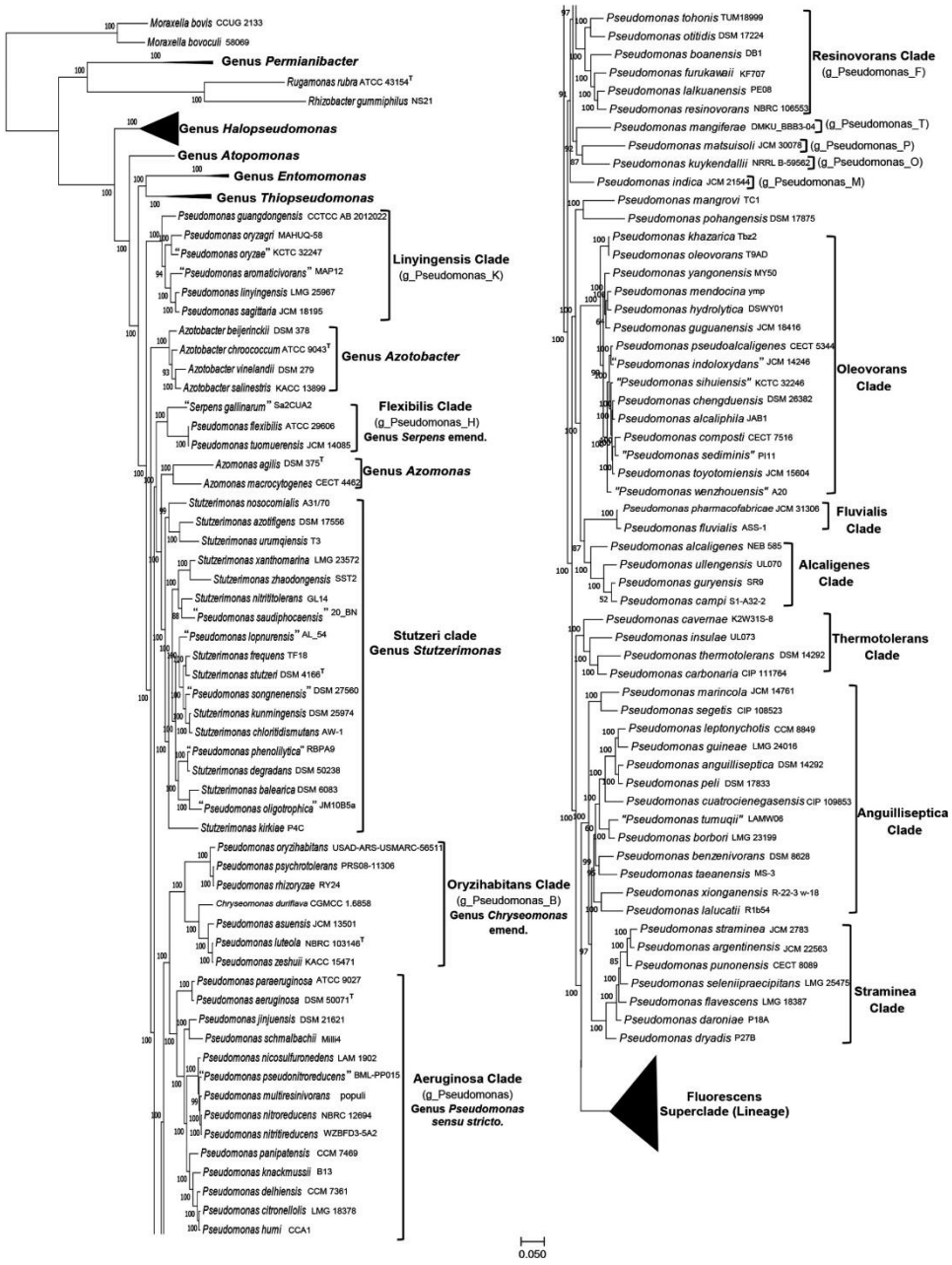

FIGURE 1
A maximum-likelihood tree for 388 genome-sequenced *Pseudomonadaceae* species based on concatenated sequences for 118 conserved proteins. The tree is shown into two halves, and species from the Fluorescens superclade (lineage) are compressed, so that the species compositions of other clades of interest can be seen. The species clades of interest are demarcated and labeled with the commonly used names and in some cases with the GTDB taxon assignment for the clade.

them (Hesse et al., 2018; Rudra and Gupta, 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). Additionally, species from the two recently proposed genera *Stutzerimonas* and *Chryseomonas* also branched within other *Pseudomonas* species, thus further contributing to the polyphyly of this genus. We have labeled different *Pseudomonas* species clades in Figure 1 and Supplementary Figure S2 by their commonly used clade/group names (Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022). One distinct clade observed in all constructed trees is the "Aeruginosa clade," which contains the type species *P. aeruginosa* and 13 other *Pseudomonas* species. As this clade contains the type species of the genus *Pseudomonas*, we have labeled it as the "Genus *Pseudomonas sensu stricto*." Other species' clades observed and labeled in Figure 1 (Supplementary Figure S2) include: the Alcaligenes, Anguilliseptica, Azomonas, Azotobacter, Flexibilis, Fluvialis, Linyingensis, Oleovorans, Oryzihabitans, Resinovorans, Straminea, Stutzeri (*Stutzerimonas*), Thermotolerans, and Fluorescens superclade (lineage).The Genome Taxonomy Database (GTDB),[2] based on phylogenetic analysis of 120 ubiquitously conserved proteins, now provides an important resource for taxonomic inferences (Parks et al., 2018). The GTDB refers to the "Aeruginosa clade" as the genus *Pseudomonas* whereas most of the other observed species clades are referred to as distinct genera denoted by designations such as g_*Pseudomonas*_B, g_*Pseudomonas*_K, etc., which are also indicated in the tree in Figure 1.

Of these observed clades, the Fluorescens superclade (lineage) is the largest harboring 245 *Pseudomonas* species. It is separated from all other *Pseudomonas* species by a long branch in both constructed trees (Figure 1; Supplementary Figure S2). Due to the large number of species present in this clade, it is shown in a compressed form in Figure 1. However, detailed information for species comprising this clade is provided in Supplementary Figure S1. The Fluorescens superclade (lineage) is made up of multiple distinct clades and subclades (see Supplementary Figure S1) (Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020; Rudra and Gupta, 2021; Lalucat et al., 2022). However, all species grouping within the Fluorescens superclade (lineage) are part of the GTDB taxon "g_*Pseudomonas*_E." Although the *Pseudomonas*_E cluster in GTDB also encompasses the Alcaligenes, Anguilliseptica, Oleovorans and Thermotolerans clades, these clades in our phylogenomic trees (Figure 1; Supplementary Figure S1), and in several other published studies (Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022), branch separately from the Fluorescens superclade. This discrepancy in the branching positions of the Alcaligenes, Anguilliseptica, Oleovorans and Thermotolerans clades between the GTDB taxonomy and other phylogenomic trees, was also noted by Lalucat et al. (2022). However, in the present work, we will not be examining the evolutionary relationships of different species within the Fluorescens superclade. Besides the "Aeruginosa clade" and the Fluorescens superclade (lineage), the other clades marked in Figure 1 (Supplementary Figure S2) contain between 2–18 species. Except for the Anguilliseptica clade, which shows poor resolution and weak statistical support, all other clades in our phylogenetic trees are statistically strongly supported. Besides these species' clades, a limited number of *Pseudomonas* species (*viz. P. indica, P. kuykendallii, P. mangiferae, P. mangrovi, P. matsuisoli* and *P. pohangensis*) are not part of any of the observed clades.

The analyzed genome sequences were also used for determination of percentage of conserved proteins (POCP) and average amino acid identity (AAI) between different pairs of genomes. The results of pairwise AAI and POCP values, for different *Pseudomonadaceae* genomes are presented in Supplementary Tables S2 and S3, respectively. Genome pairs exhibiting higher AAI or POCP values are shown by a darker shade of green/red, and different clades observed in our phylogenetic trees (Figure 1; Supplementary Figure S2) are outlined. In Table 1, we present a summary of the ranges of the AAI and POCP values for different *Pseudomonas* species clades for the ingroup and outgroup species. Based on the results in Table 1, the AAI and POCP values for species within different clades are higher (AAI values range: 0.70–1.00; POCP values range: 0.66–1.00) in comparison to these values for species from the other clades (AAI values range: 0.67–0.81; POCP values range: 0.42–0.77), which is an expected result. However, based on the AAI and POCP values (Table 1), only species from the Alcaligenes, Azotobacter, Flexibilis, Fluvialis, Lingyingensis, Oleovorans and Thermotolerans clades show no overlap with species from the other clades. In contrast, these values for several other clades (*viz.* "Aeruginosa." Anguiliiseptica, Azomonas, Oryzihabitans, Resinovorans, Straminea, Stutzeri) either show significant overlap or are very close to those from the outgroup species. Thus, based on these genome similarity indices, species from different observed *Pseudomonadaceae* clades cannot be reliably demarcated. In Table 1, the highest overlap in the AAI and POCP values between the ingroup versus outgroup species is observed for the species from Anguilliseptica clade, which also shows poor resolution and weak statistical support in the phylogenetic trees.

Identification of molecular markers demarcating/distinguishing different *Pseudomonas* species clades

Although *Pseudomonadaceae* species form similar clades in different genome scale trees (Hesse et al., 2018; Parks et al., 2018; Girard et al., 2021; Lalucat et al., 2022; Figure 1; Supplementary Figure S2), branching of species in phylogenetic trees is influenced by large numbers of variables (Gupta, 1998; Baldauf, 2003; Felsenstein, 2004). Moreover, in phylogenetic trees for *Pseudomonas*, species from several clades are separated from each other by short branches (Figure 1; Supplementary Figure S2), which makes it difficult to reliably determine their boundaries. The POCP and AAI values for several clades also overlap or are very close to the other species (Table 1), thus they do not permit reliable determination of the boundaries of these clades. Hence, it was important to discover other reliable means for the demarcation of these clades. Molecular synapomorphies consisting of CSIs in genes/proteins sequences, which are uniquely shared characteristics of species from different clades, provide important means for the demarcation of taxa of different ranks in molecular terms (Gupta, 2014; Adeolu et al., 2016; Gupta et al., 2020; Patel and Gupta, 2020; Rudra and Gupta, 2021). Hence, detailed studies were conducted on protein sequences from *Pseudomonadaceae* species to identify CSIs which are specific for different observed clades. These analyses have identified 98 novel CSIs which are specific for different *Pseudomonadaceae* clades, providing

2 http://gtdb.ecogenomic.org/

TABLE 1 Range of AAI and POCP values among different *Pseudomonadaceae* species clades.

Clades	AAI values		POCP values	
	Ingroup	Outgroup	Ingroup	Outgroup
"Aeruginosa clade" (*Pseudomonas sensu stricto*)	0.75–1.00	0.67–0.75	0.66–1.00	0.42–0.73
Alcaligenes clade (*Aquipseudomonas* gen. nov.)	0.83–1.00	0.69–0.79	0.79–1.00	0.49–0.75
Anguilliseptica clade	0.77–1.00	0.68–0.81	0.68–1.00	0.45–0.75
Genus *Azomonas*	0.73–1.00	0.68–0.74	0.68–1.00	0.42–0.67
Genus *Azotobacter*	0.86–1.00	0.68–0.76	0.80–1.00	0.49–0.67
Flexibilis clade (Genus *Serpens* emend.)	0.79–1.00	0.69–0.76	0.83–1.00	0.51–0.69
Fluvialis clade (*Caenipseudomonas* gen. nov.)	1.00	0.70–0.77	1.00	0.48–0.71
Linyingensis clade (*Geopseudomonas* gen. nov.)	0.82–1.00	0.69–0.75	0.69–1.00	0.49–0.67
Oleovorans clade (*Ectopseudomonas* gen. nov.)	0.88–1.00	0.67–0.81	0.75–1.00	0.43–0.77
Oryzihabitans clade (Genus *Chryseomonas* emend.)	0.71–1.00	0.67–0.72	0.70–1.00	0.47–0.67
Resinovorans clade (*Metapseudomonas* gen. nov.)	0.79–1.00	0.68–0.77	0.70–1.00	0.44–0.74
Straminea clade (*Phytopseudomonas* gen. nov.)	0.76–1.00	0.67–0.81	0.69–1.00	0.47–0.76
Stutzeri clade (Genus *Stutzerimonas*)	0.77–1.00	0.68–0.76	0.72–1.00	0.49–0.66
Thermotolerans clade (*Zestomonas* gen. nov.)	0.81–1.00	0.70–0.79	0.75–1.00	0.48–0.74

Detailed information regarding the pairwise AAI and POCP values for species from different clades is provided in Supplementary Tables S2 and S3.

independent evidence for the genetic distinctness of these clades and affording reliable means for their demarcation. Brief descriptions of the characteristics of these CSIs are given below.

CSIs specific for the "Aeruginosa clade"

The "Aeruginosa clade" representing the genus *Pseudomonas sensu stricto*, encompasses 14 named species (*viz.*, *P. aeruginosa, P. paraeruginosa, P. citronellolis, P. delhiensis, P. humi, P. jinjuensis, P. knackmussii, P. multiresinivorans, P. nicosulfuronedens, P. nitritireducens, P. nitroreducens, P. panipatensis, "P. pseudonitroreducens"* and *P. schmalbachii*) (Figure 1). Our analyses have identified six CSIs in proteins involved in different functions (Table 2), which are commonly and, in most cases, uniquely shared by different species from the "Aeruginosa clade." Sequence information for one of these is presented in Figure 2. In the example shown, a two aa insertion (highlighted) in a conserved region of the HugZ family protein is commonly shared by all 14 species from the "Aeruginosa clade" but absent in all other *Pseudomonadaceae* species. Sequence information is shown in Figure 2 for only a limited number of species. However, more detailed information for this CSI is presented in Supplementary Figure S3. Like the CSI shown in Figure 2, we have identified five additional CSIs in other proteins which, except for an isolated occurrence, are uniquely shared by different species from the "Aeruginosa clade." Sequence information for these CSIs is provided in Supplementary Figures S4–S8 and some of their characteristics are summarized in Table 2. Due to their unique shared presence in species from the "Aeruginosa clade," genetic changes responsible for these CSIs likely occurred in a common ancestor of this clade and subsequently inherited by all members. Due to their specificities for the species from the "Aeruginosa clade," these molecular synapomorphies provide robust means for the demarcation of this clade in molecular terms.

CSIs specific for the Alcaligenes clade

P. alcaligenes was indicated to branch separately from other clades in earlier studies (Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022). In our phylogenetic trees (Figure 1; Supplementary Figure S2), three recently identified species (*viz.*, *P. campi, P. guryensis, P. ullengensis*) also reliably grouped with *P. alcaligenes*. Our analysis has identified six novel CSIs, which in most cases are exclusively shared by all four species from the Alcaligenes clade. Sequence information for one of these CSIs is presented in Figure 3A, where a two aa insertion in the protein ferric iron uptake transcriptional regulator is exclusively present in all four species from the Alcaligenes clade. Five additional CSIs in other proteins are also generally specific for the species from this clade. Detailed sequence information for these six CSIs is provided in Supplementary Figures S9–S14, and some of their characteristics are listed in Table 2. The identified CSIs provide reliable means for the demarcation of species from the Alcaligenes clade in molecular terms and we are proposing their transfer into *Aquipseudomonas* gen. nov.

CSIs specific for the Oleovorans clade

Oleovorans clade is a strongly supported clade consisting of 15 *Pseudomonas* species (*viz.*, *P. alcaliphila, P. chengduensis, P. composti, P. guguanensis, P. hydrolytica, "P. indoloxydans," P. khazarica, P. mendocina, P. oleovorans, P. pseudoalcaligenes, "P. sediminis," "P. sihuiensis," P. toyotomiensis, "P. wenzhouensis," P. yangonensis*), which reliably group together in the constructed phylogenetic trees (Figure 1; Supplementary Figure S2). The genetic distinctness of this clade is also independently supported by five novel identified CSIs which, excepting an isolated occurrence, are uniquely shared by all species from this clade. Sequence information for one of these CSIs is provided in Figure 3B, where a one aa

TABLE 2 Summary of CSIs specific for the "Aeruginosa," Alcaligenes, and Oleovorans clades.

Protein name	Accession no	Figure number	Indel size	Indel location	Specificity
HugZ family protein	WP_058144759	Figure 2; **Supplementary Figure S3**	2 aa Ins	126–156	"Aeruginosa clade" (*Pseudomonas sensu stricto*)
TetR family transcriptional regulator	WP_162953821	**Supplementary Figure S4**	1 aa Ins	68–104	
Transglutaminase family protein[#]	WP_089389603	**Supplementary Figure S5**	1aa Ins	39–83	
Multidrug efflux RND transporter permease subunit	WP_038803172	**Supplementary Figure S6**	2 aa Ins	233–269	
Alginate O-acetyltransferase[#]	PXC05278	**Supplementary Figure S7**	1 aa Del	24–61	
23S rRNA (cytidine(2498)-2'-O)-methyltransferase RlmM[#]	OVZ41066	**Supplementary Figure S8**	1 aa Ins	54–98	
Ferric iron uptake transcriptional regulator	WP_110680887	Figure 3A; **Supplementary Figure S9**	2 aa Ins	6–52	Alcaligenes clade (*Aquipseudomonas* gen. nov.)
DUF1853 family protein	WP_061903990	**Supplementary Figure S10**	1 aa Del	55–93	
SCP2 sterol-binding domain-containing protein	WP_076424264	**Supplementary Figure S11**	1 aa Del	55–98	
Hypothetical protein[$]	GIZ66354	**Supplementary Figure S12**	4 aa Del	125–167	
Zinc ABC transporter substrate-binding protein	WP_061902889	**Supplementary Figure S13**	4 aa Del	261–297	
Hybrid sensor histidine kinase/response regulator	WP_203791762	**Supplementary Figure S14**	2 aa Del	130–170	
Cysteine synthase A	WP_150609166	Figure 3B; **Supplementary Figure S15**	1 aa Del	119–160	Oleovorans clade (*Ectopseudomonas* gen. nov.)
Lipopolysaccharide export system permease protein[$]	NYF64131	**Supplementary Figure S16**	1 aa Ins	19–61	
Succinylglutamate desuccinylase[$]	WP_125875007	**Supplementary Figure S17**	1 aa Ins	121–164	
Fe2 + −dependent dioxygenase	WP_206407640	**Supplementary Figure S18**	4 aa Del	124–155	
Osmoprotectant NAGGN system M42 family peptidase[#]	WP_206408901	**Supplementary Figure S19**	3 aa Ins	46–85	

[#]The CSIs listed here are specific for the indicated clades of bacteria, apart from an isolated exception present in some CSIs (#; see Supplementary Figures for details). [$]The protein homologs were not found in some species.

deletion (highlighted), within a conserved region of the protein cysteine synthase A, is exclusively shared by all species from the Oleovorans clade. More detailed sequence information for this CSI and four additional CSIs specific for the Oleovorans clade is provided in Supplementary Figures S15–S19 and some of their characteristics are listed in Table 2. Based on the strong evidence presented here demonstrating the distinctness of species from the Oleovorans clade, we are proposing the transfer of these species into *Ectopseudomonas* gen. nov.

In addition to the species with validly published names, Oleovorans clade also encompasses four species [*viz.*, *"P. indoloxydans"* (Manickam et al., 2008), *"P. sediminis"* (Behera et al., 2018), *"P. sihuiensis"* (Wu et al., 2014) and *"P. wenzhouensis"* (Zhang et al., 2021)], whose names have not been validly published. Because of their non-validly published status, new name combinations for these species are not proposed. However, in view of their reliable grouping with the Oleovorans clade, it is suggested that these species should also be recognized as members of the genus *Ectopseudomonas* with the names *"E. indoloxydans," "E. sediminis," "E. sihuiensis"* and *"E. wenzhouensis,"* respectively.

CSIs specific for the Straminea clade

The Straminea clade is a strongly supported cluster encompassing seven *Pseudomonas* species (*P. argentinensis, P. daroniae, P. dryadis, P. flavescens, P. punonensis, P. seleniipraecipitans, P. straminea*) (Figure 1; Supplementary Figure S2). Species from this clade have also been found to group together in earlier studies (Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). The members of this clade can be reliably distinguished from all other *Pseudomonadaceae* species by 12 novel CSIs identified in this study, which in most cases are exclusively shared by the species from this clade. Sequence information for one of these CSIs consisting of a three aa insertion in the protein Di-trans, poly-cis-decaprenylcistransferase is presented in Figure 3C. Detailed sequence information for this CSI and the 11 other CSIs specific for this clade are presented in Supplementary Figures S20–S31 and some of their characteristics are listed in Table 3. Based on the presented results showing the distinctness of this clade, we are proposing the transfer of species from this clade into *Phytopseudomonas* gen. nov.

CSIs specific for the genus *Stutzerimonas*

The genus *Stutzerimonas* was recently described by Lalucat et al. (2022) by the transfer of several *Pseudomonas* species which branched distinctly in their phylogenetic tree. The clade labeled as *Stutzerimonas* in our phylogenetic tree (Figure 1) encompasses all 13 named *Stutzerimonas* species, whose genome sequences were available in the NCBI database at the time of analysis, as well as five non-validly published *Pseudomonas* species. Apart from their

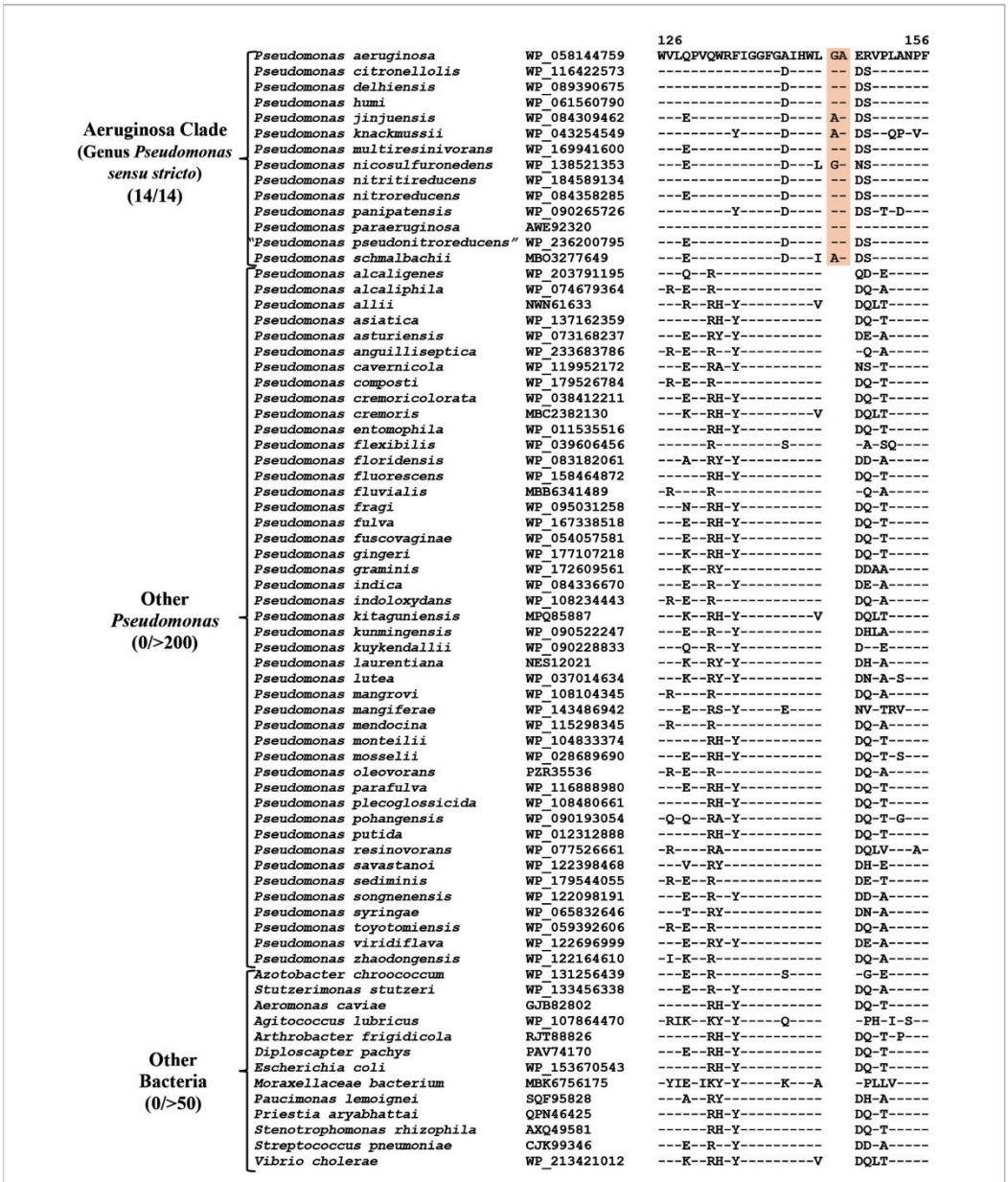

FIGURE 2
Partial sequence alignment of the HugZ family protein showing a two aa insertion (highlighted) that is exclusively present in all members of the "Aeruginosa clade." The dashes (–) in this and all other sequence alignments indicate identity with the amino acids on the top line. Accession numbers for different sequences are indicated in the second column and the numbers at the top indicate the position of this sequence in the protein sequences. Detailed sequence information for this CSI and five other CSIs specific for this clade is provided in Supplementary Figures S3–S8.

clustering in phylogenetic trees, there is no known reliable characteristic which is specific for the members of this genus. Our analyses have identified seven CSIs in different proteins, which in most cases are uniquely shared by all/most species from this clade. Sequence information for one of these CSIs is shown in Figure 4A. In this instance, a one aa insertion in a conserved region of the PAS

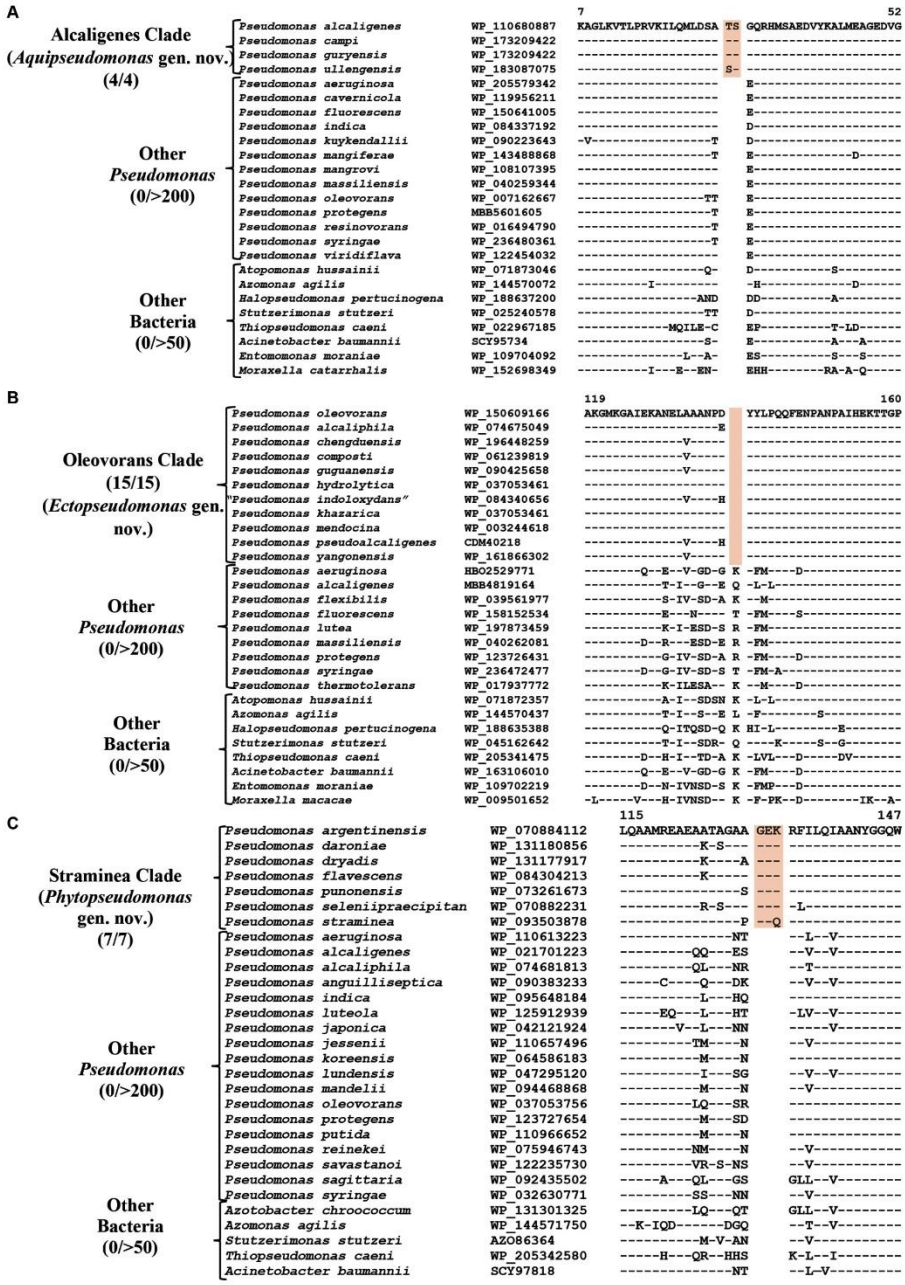

FIGURE 3
Partial sequence alignments of **(A)** Ferric iron uptake protein showing a two aa insertion within a conserved region that is a distinctive characteristic of all members of the Alcaligenes clade. **(B)** A one aa deletion in a conserved region of the protein Cysteine synthase A which is specific for the species from Oleovorans clade. **(C)** A three aa insertion within a conserved region in the protein Di-trans, poly-cis-decaprenylcistransferase, specific for the species from the Straminea clade. Detailed sequence information for these CSIs along with other CSIs specific for these clades are provided in Supplementary Figures S9–S31.

TABLE 3 Summary of CSIs specific for members of the Straminea, Stutzeri, and Linyingensis clades.

Protein name	Accession no	Figure number	Indel size	Indel location	Specificity
Di-trans, poly-cis-decaprenylcistransferase	WP_070884112	Figure 3C; **Supplementary Figure S20**	3 aa Ins	110–150	Straminea clade (*Phytopseudomonas* gen. nov.)
Efflux RND transporter periplasmic adaptor subunit	WP_074886159	**Supplementary Figure S21**	2 aa Del	203–245	
Beta-ketoacyl-ACP synthase III	WP_093501944	**Supplementary Figure S22**	1 aa Ins	233–273	
Sugar ABC transporter ATPase[&]	WP_093502557	**Supplementary Figure S23**	2 aa Del	26–65	
DNA polymerase III subunit alpha[&]	WP_093503860	**Supplementary Figure S24**	4 aa Ins	818–855	
Polyprenyl diphosphate synthase[&]	WP_093503878	**Supplementary Figure S25**	3 aa Ins	110–153	
Ubiquinol-cytochrome c[&] reductase cytochrome b subunit	SFD97069	**Supplementary Figure S26**	5 aa Ins	65–102	
GTP diphosphokinase[&]	WP_093502677	**Supplementary Figure S27**	1 aa Ins	108–150	
tRNA (adenosine(37)-N6)-dimethylallyltransferase MiaA[&].	WP_093506440	**Supplementary Figure S28**	5 aa Del	167–203	
Transporter substrate-binding domain-containing protein[#]	WP_093500877	**Supplementary Figure S29**	1 aa Ins	112–152	
YIP1 family protein[#]	WP_074882567	**Supplementary Figure S30**	1 aa Del	48–87	
Methyltransferase[#]	WP_074882425	**Supplementary Figure S31**	1 aa Ins	55–85	
PAS domain-containing methyl-accepting chemotaxis protein	WP_084903134	Figure 4A; **Supplementary Figure S32**	1 aa Ins	83–127	Stutzeri clade (Genus *Stutzerimonas*)
DUF1329 domain-containing protein	WP_049338638	**Supplementary Figure S33**	1 aa Del	115–121	
Autotransporter assembly complex protein TamA[$]	WP_084904442	**Supplementary Figure S34**	1 aa Del	112–147	
2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase[#]	WP_014818653	**Supplementary Figure S35**	1 aa Ins	105–149	
Rhomboid family intramembrane serine protease[#]	WP_218422476	**Supplementary Figure S36**	2 aa Ins	237–265	
RnfABCDGE type electron transport complex subunit D[#]	WP_106442915	**Supplementary Figure S37**	1 aa Del	165–212	
16S rRNA (uracil(1498)-N(3))-methyltransferase[$]	WP_221292728	**Supplementary Figure S38**	1 aa Del	142–170	
UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase	WP_090305970	Figure 4B; **Supplementary Figure S39**	5 aa Ins	372–421	Linyingensis clade (*Geopseudomonas* gen. nov.)
Septal ring lytic transglycosylase RlpA family protein	WP_090305376	**Supplementary Figure S40**	1 aa Ins	272–311	
Dephospho-CoA kinase	WP_090305710	**Supplementary Figure S41**	1 aa Ins	107–142	
ATP-dependent zinc metalloprotease FtsH	WP_090308457	**Supplementary Figure S42**	1 aa Del	413–445	
Penicillin-binding protein 1A	WP_090307056	**Supplementary Figure S43**	1 aa Ins	232–282	
bifunctional [glutamate--ammonia ligase]-adenylyl-L-tyrosine phosphorylase/[glutamate--ammonia-ligase] adenylyltransferase	WP_090307131	**Supplementary Figure S44**	1 aa Ins	672–718	
Repressor LexA	WP_090307764	**Supplementary Figure S45**	2 aa Ins	166–201	
Malate dehydrogenase	WP_090312804	**Supplementary Figure S46**	1 aa Ins	131–162	
Uridylyltransferase	WP_090313706	**Supplementary Figure S47**	1 aa Ins	629–676	
CHAD domain-containing protein[$]	WP_090307991	**Supplementary Figure S48**	3 aa Del	166–203	
Protocatechuate 3,4-dioxygenase subunit alpha[$]	WP_090309801	**Supplementary Figure S49**	4 aa Del; 1 aa Del	109–141	
Secretin	WP_090310373	**Supplementary Figure S50**	1 aa Del	194–231	
CDP-6-deoxy-delta-3,4-glucoseen reductase	WP_090312664	**Supplementary Figure S51**	3 aa Ins	236–276	
YkgJ family cysteine cluster protein	WP_090306967	**Supplementary Figure S52**	1 aa Ins	9–45	
tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA[#]	WP_090305582	**Supplementary Figure S53**	2 aa Ins	145–182	

[#]Isolated exception present in some CSIs (#; see Supplementary Figures for details). [$]The protein homologs were not found in some species. [&]CSI is not found in *P. dryadis*, which is the deepest branching member of the clade.

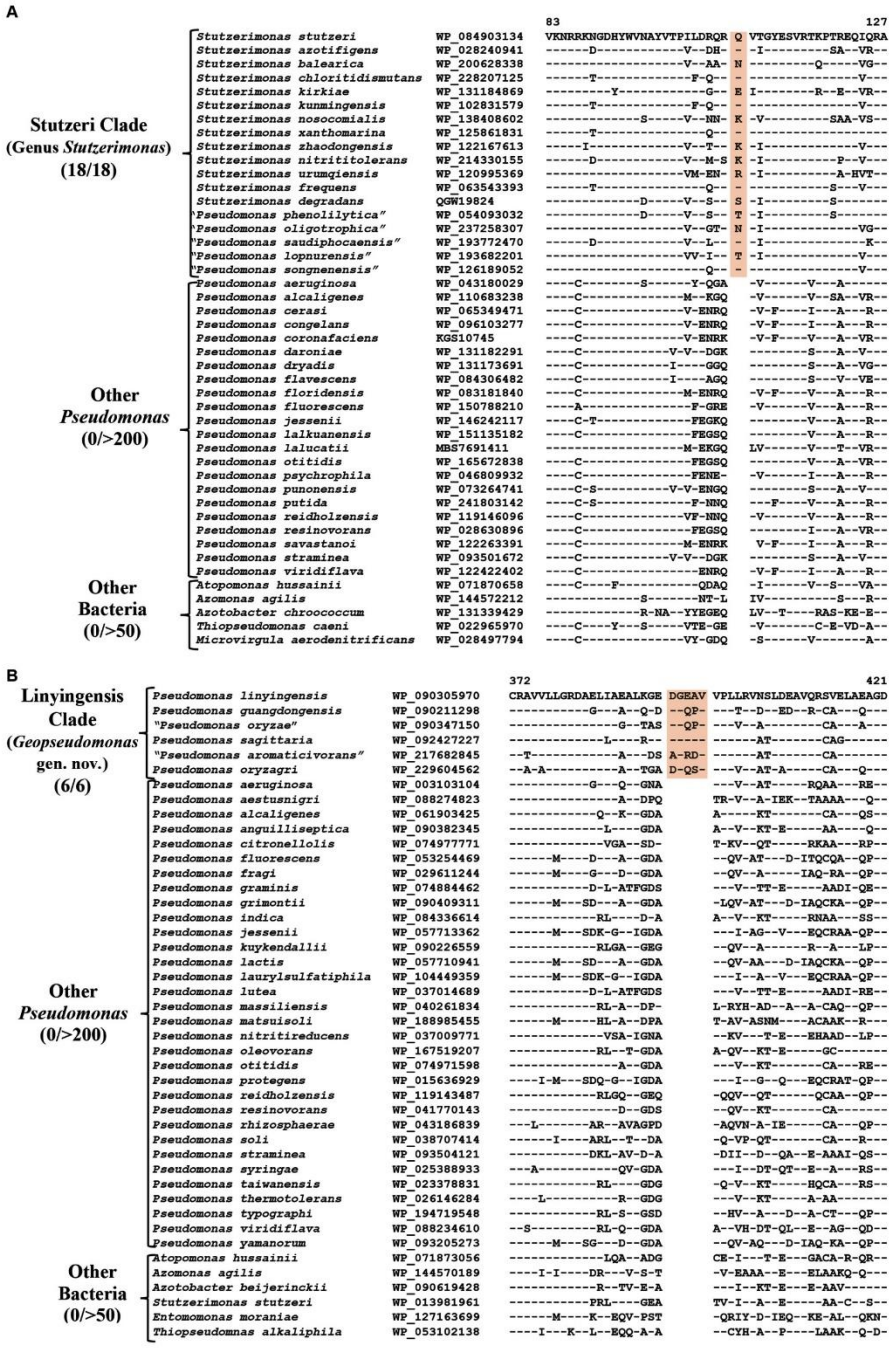

FIGURE 4
Partial sequence alignment of the protein **(A)** PAS domain containing Methyl-accepting chemotaxis protein showing a one aa insertion within a conserved region (highlighted) that is uniquely present in all members of the Stutzeri clade. **(B)** A five aa insertion within a conserved region of the protein UDP-N-acetylmuramoyl-L-alanine, which is specific for the species from Linyingensis clade. Detailed sequence information for these CSIs and other CSIs specific for Stutzeri and Linyingensis clades are provided in Supplementary Figures S32–S53.

domain-containing methyl-accepting chemotaxis protein is uniquely shared by all species from the *Stutzerimonas* clade. Detailed sequence information for this CSI and the six other CSIs specific for this clade/genus is provided in Supplementary Figures S32–S38 and some of their characteristics are summarized in Table 3. The identified CSIs provide reliable means for distinguishing *Stutzerimonas* species from all other *Pseudomonadaceae* species. Hence, we are emending the description of this genus to include these diagnostic characteristics.

Five species with non-validly published names [*viz.* "*P. lopnurensis*" (Mamtimin et al., 2021), "*P. phenolilytica*" (Kujur and Das, 2022), "*P. oligotrophica*" (Zhang et al., 2022), "*P. saudiphocaensis*" (Azhar et al., 2017) and "*P. songnenensis*" (Zhang et al., 2015)], also group reliably within the *Stutzerimonas* clade and share CSIs specific for this clade. These species should also be recognized as members of this genus with the names "*S. lopnurensis*," "*S. phenolilytica*," "*S. oligotrophica*," "*S. saudiphocaensis*" and "*S. songnenensis*" respectively.

CSIs specific for the Linyingensis clade

The Linyingensis clade consists of six *Pseudomonas* species *viz.*, *P. aromaticivorans, P. guangdongensis, P. linyingensis, P. oryzagri,* "*P. oryzae*" and *P. sagittaria*, which form a strongly supported clade in our phylogenetic trees (Figure 1; Supplementary Figure S2). This clade is also denoted as g_Pseudomonas_K in the GTDB taxonomy (Parks et al., 2018). A specific evolutionary relationship among these species is supported by 15 CSIs (Table 3), which in most cases are uniquely shared by all species from this clade. In Figure 4B, we present one example of a CSI specific for this clade, where a five aa insertion in UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase protein is uniquely shared by all members of this clade. Detailed sequence information for this CSI and 14 other CSIs specific for this clade is presented in Supplementary Figures S39–S53. Based on these results, which robustly demarcate this species clade, we are proposing the transfer of these species into *Geopseudomonas* gen. nov.

CSIs specific for the Resinovorans clade

The Resinovorans clade (Figure 1; Supplementary Figure S2), which is denoted as the taxon g_*Pseudomonas*_F in GTDB taxonomy (Parks et al., 2018), consists of six species *viz. P. boanensis, P. furukawaii, P. lalkuanensis, P. otitidis, P. resinovorans* and *P. tohonis*. Species from this clade also formed a distinct clade in earlier studies (Girard et al., 2021; Lalucat et al., 2022; Passarelli-Araujo et al., 2022). The members of this clade can be reliably distinguished from all other *Pseudomonadaceae* species by five identified CSIs, which in most cases are exclusively shared by all/most species from this clade. One example of a CSI specific for this clade is presented in Figure 5A, where in the Murein L, D-transpeptidase catalytic domain family protein, a two aa insertion is exclusively present in all species from the Resinovorans clade. Detailed sequence information for this CSI and four other identified CSIs, specific for this clade, is presented in Supplementary Figures S54–S58 and some of their characteristics are listed in Table 4. Based on these results, we are proposing the transfer of species from Resinovorans clade into *Metapseudomonas* gen. nov.

CSIs specific for the Oryzihabitans clade (genus *Chryseomonas*)

Oryzihabitans clade (denoted as the taxon g_*Pseudomonas*_B in GTDB taxonomy) consists of seven named *Pseudomonas* species *viz. P. asuensis, P. duriflava, P. luteola, P. oryzihabitans, P. psychrotolerans, P. rhizoryzae* and *P. zeshuii*, which form a strongly supported clade in our phylogenetic trees (Figure 1; Supplementary Figure S2). These species also formed a distinct clade in earlier phylogenetic studies (Hesse et al., 2018; Girard et al., 2021; Saati-Santamaría et al., 2021; Passarelli-Araujo et al., 2022). The best-studied species from this clade is *P. luteola*, which was originally a member of the genus *Chryseomonas* (Holmes et al., 1986). However, in 1997, based on 16S rRNA gene sequence similarity, this species was transferred into the genus *Pseudomonas* (Anzai et al., 1997). More recently, based on genomic studies, this species along with two other *Pseudomonas* species (*P. asuensis* and *P. duriflava*) were transferred into the genus *Chryseomonas*. It should be noted that *C. luteola* is a synonym of *C. polytricha* (Holmes et al., 1986), which is the type species of genus *Chryseomonas* (Parte et al., 2020). The genetic distinctness of the clade formed by these seven species is strongly supported by 11 novel identified CSIs which are uniquely shared by these species. One example of a CSIs specific for this clade is shown in Figure 5B. In this case, a one aa insertion in the protein cytochrome d ubiquinol oxidase subunit II is exclusively shared by all members of this clade. Detailed sequence information for this CSI and 10 other CSIs specific for this clade are presented in Supplementary Figures S59–S69 and some of their characteristics are listed in Table 4. In addition to the three species which are presently assigned to the genus *Chryseomonas*, four additional *Pseudomonas* species *viz. P. oryzihabitans, P. psychrotolerans, P. rhizoryzae* and *P. zeshuii* reliably group within this clade and share different CSIs specific for this genus. Hence, we are proposing new name combinations of these species to transfer them into the genus *Chryseomonas*.

CSIs specific for the Thermotolerans clade

The Thermotolerans clade includes the species *P. carbonaria, P. cavernae*, *P. insulae* and *P. thermotolerans,* which form a distinct clade in our phylogenomic trees (Figure 1; Supplementary Figure S2). Species from this clade also formed a distinct cluster in earlier studies (Girard et al., 2021; Lalucat et al., 2022). A specific evolutionary relationship among these species is strongly supported by five CSIs, which are exclusively shared by all members of this clade. One example of a CSI specific for this clade is shown in Figure 6A, where a six aa insertion in the TerB family tellurite resistance protein is exclusively found in all four species from this clade. Detailed sequence information for the five CSIs specific for this clade are presented in Supplementary Figures S70–S74 and some of their characteristics are listed in Table 4. Based on these results, we are proposing the transfer of species from this clade into *Zestomonas* gen. nov.

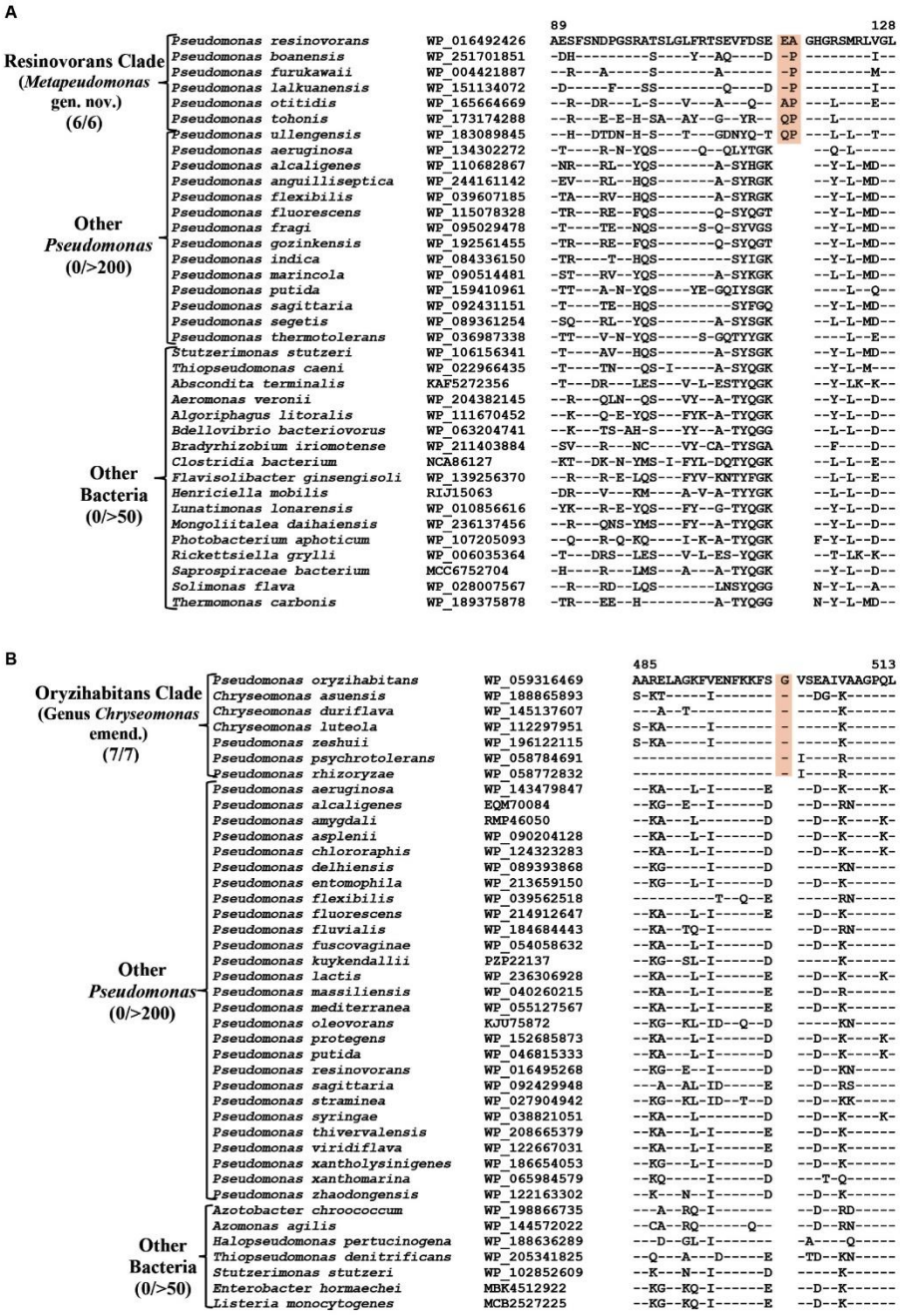

FIGURE 5
Partial sequence alignment of the protein **(A)** Murein L, D-transpeptidase catalytic domain family protein showing a two aa insertion within a conserved region that is commonly shared by all members of the Resinovorans clade. **(B)** A one aa insertion in the protein Cytochrome d ubiquinol oxidase subunit II which is specific for the species from the Oryzihabitans clade. Detailed sequence information for these CSIs and other CSIs specific for Resinovorans and Oryzihabitans clades are provided in Supplementary Figures S54–S69.

TABLE 4 Summary of CSIs specific for members of the Resinovorans, Oryzihabitans, Thermotolerans, and Flexibilis clades.

Protein name	Accession no	Figure number	Indel size	Indel location	Specificity
Murein L, D-transpeptidase catalytic domain family protein[#]	WP_016492426	Figure 5A; **Supplementary Figure S54**	2 aa Ins	89–128	Resinovorans clade (*Metapseudomonas* gen. nov.)
Leucine--tRNA ligase[#]	WP_016490742	**Supplementary Figure S55**	5 aa Ins	260–304	
Alginate biosynthesis protein Alg44	WP_028628607	**Supplementary Figure S56**	1 aa Del	17–49	
YggL family protein	WP_051246415	**Supplementary Figure S57**	1 aa Ins	61–93	
Glycine--tRNA ligase subunit beta	WP_016489954	**Supplementary Figure S58**	3 aa Del	597–641	
Cytochrome d ubiquinol oxidase subunit II	WP_241809250	Figure 5B; **Supplementary Figure S59**	1 aa Ins	236–279	Oryzihabitans clade (Genus *Chryseomonas*)
Phosphoenolpyruvate carboxykinase	WP_059316469	**Supplementary Figure S60**	1 aa Ins	485–513	
GTPase HflX	WP_059316391	**Supplementary Figure S61**	1 aa Ins	317–385	
ATP-binding protein	WP_059313194	**Supplementary Figure S62**	1 aa Ins	192–230	
16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase RsmA	WP_059313310	**Supplementary Figure S63**	1 aa Del	77–115	
PTS fructose transporter subunit IIBC	HJE68896	**Supplementary Figure S64**	1 aa Del	36–75	
Glucokinase	WP_007158679	**Supplementary Figure S65**	1 aa Ins	164–201	
Dienelactone hydrolase family protein	WP_160922865	**Supplementary Figure S66**	1 aa Ins	40–77	
Bifunctional D-glycero-beta-D-manno-heptose-7-phosphate kinase/D-glycero-beta-D-manno-heptose 1-phosphate adenylyltransferase HldE	WP_059313726	**Supplementary Figure S67**	1 aa Ins	415–457	
Zinc transporter ZntB	WP_197850824	**Supplementary Figure S68**	1 aa Ins	209–245	
NADH-dependent 7-cyano-7-deazaguanine	WP_208691271	**Supplementary Figure S69**	1 aa Ins	180–220	
TerB family tellurite resistance protein	WP_017939833	Figure 6A; **Supplementary Figure S70**	6 aa Ins	27–75	Thermotolerans clade (*Zestomonas* gen. nov.)
TIGR02099 family protein	WP_119894903	**Supplementary Figure S71**	1 aa Del	175–206	
HAMP domain-containing histidine kinase	WP_187671317	**Supplementary Figure S72**	1 aa Ins	359–390	
23S rRNA (adenine(2030)-N(6))-methyltransferase RlmJ	WP_119895222	**Supplementary Figure S73**	1 aa Del	47–87	
Esterase-like activity of phytase family protein	WP_119895183	**Supplementary Figure S74**	1 aa Ins	261–299	
GTP diphosphokinase	WP_039562945	Figure 6B; **Supplementary Figure S75**	1 aa ins	464–500	Flexibilis clade (Genus *Serpens* emend.)
Zinc ABC transporter permease subunit ZnuB	WP_039607122	**Supplementary Figure S76**	1 aa Del	85–120	
LutB/LldF family L-lactate oxidation iron–sulfur protein	WP_039560866	**Supplementary Figure S77**	1 aa Del	433–469	

[#]Isolated exception present in some CSIs (#; see Supplementary Figures for details).

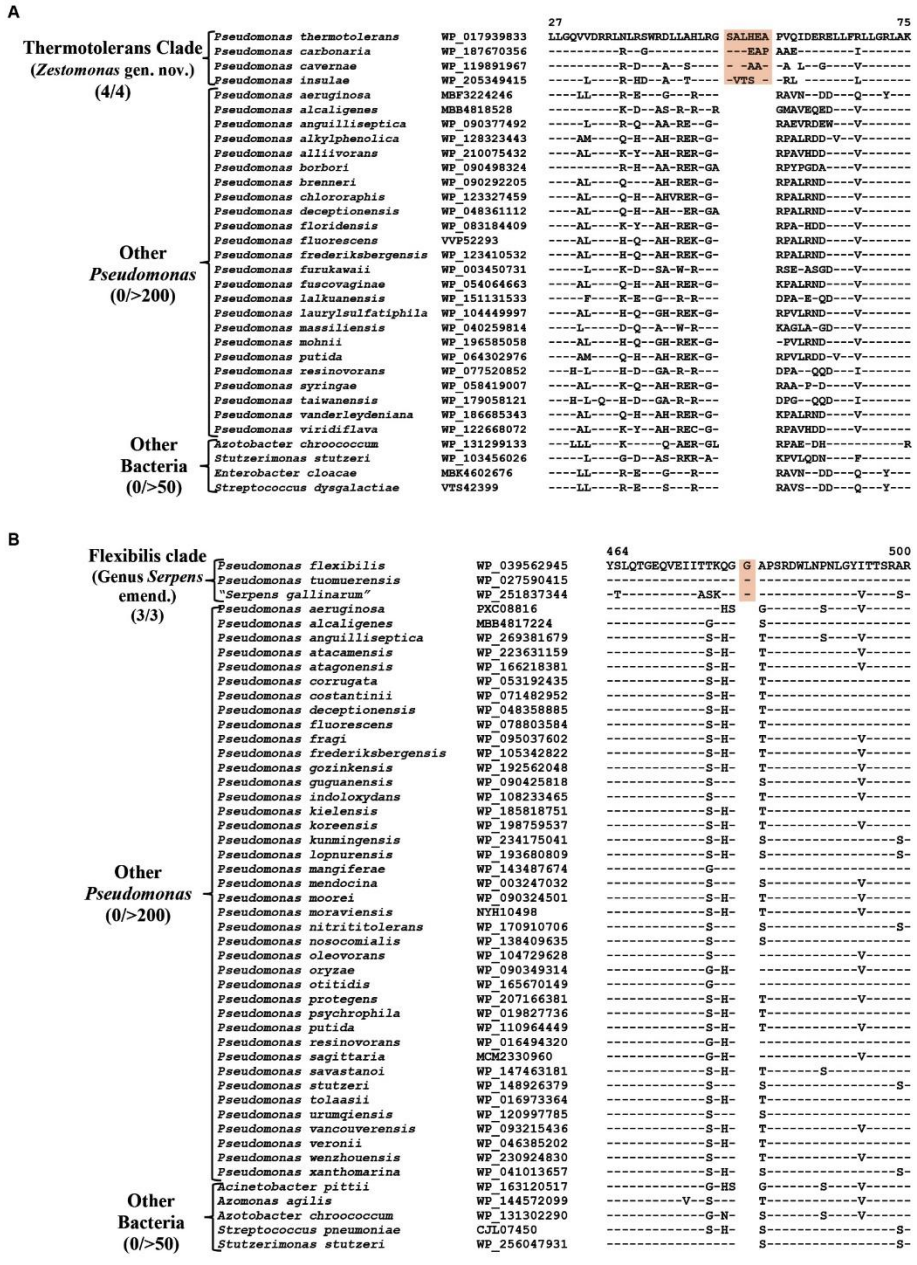

FIGURE 6
Partial sequence alignment of the protein **(A)** TerB family tellurite resistance protein showing a six aa insertion within a conserved region (highlighted) that is uniquely shared by members of the Thermotolerans clade. **(B)** A one aa insertion in a conserved region of the protein GTP diphosphokinase which is specific for the species from Flexibilis clade. Detailed sequence information for these CSIs and other CSIs specific for the Thermotolerans and Flexibilis clades are provided in Supplementary Figures S70–S77.

CSIs specific for the Flexibilis clade (genus *Serpens*)

Pseudomonas flexibilis, formerly known as *Serpens flexibilis* (Hespell, 1977) was recently transferred into the genus *Pseudomonas* based on 16S rRNA similarity with *P. pseudoalcaligenes* (Shin et al., 2015). In our phylogenomic tree (Figure 1), this species branches separately from other *Pseudomonas* species and forms a distinct clade together with a newly described non-validly published species "*Serpens gallinarum*" (Gilroy et al., 2021) and another species *P. tuomuerensis*, which according to Shin et al. (2015) is a heterotypic synonym of *P. flexibilis*. This clade is identified as the taxon g_*Pseudomonas*_H in the GTDB taxonomy (Parks et al., 2018). A close and specific relationship of *P. flexibilis* (*P. tuomuerensis*) to "*S. gallinarum*" is independently supported by three CSIs identified in this study, which are exclusively shared by these species. One example of a CSI specific for this clade is shown in Figure 6B, where a one aa insertion in the protein GTP diphosphokinase is specifically shared by these three species. Detailed sequence information for this CSI and the two other CSIs specific for this clade is presented in Supplementary Figures S75–S77 and some of their characteristics are summarized in Table 4. Based on these results we are presenting an emended description of the genus *Serpens* with *S. flexibilis* as its type species.

CSIs specific for the Fluvialis clade

The *Fluvialis* clade consists of the species *P. fluvialis* and *P. pharmacofabricae*, which formed a strongly supported clade in different phylogenetic trees (Figure 1; Supplementary Figure S2). Our analyses have identified eight CSIs in different proteins that are uniquely shared by these two species. Figure 7A depicts an example of a CSI, consisting of a seven aa deletion within a conserved region of an ATP binding protein, which is exclusively shared by these two species. Detailed sequence information for this and the six other CSIs specific for the Fluvalis clade is presented in Supplementary Figures S78–S85 and a summary of some of their sequence characteristics is presented in Table 5. Based on the results presented here, we are proposing the transfer of species from this clade into *Caenipseudomonas* gen. nov.

Identification of CSIs specific for the *Azotobacter* and *Azomonas* genera

The genus *Azotobacter* was described by Beijerinck (1901) and its members are known to branch in between *Pseudomonas* species (Young and Park, 2007; Özen and Ussery, 2012; Lalucat et al., 2022). Four *Azotobacter* species whose genome sequences were analyzed in this study (*viz. A. beijerinckii, A. chroococcum, A. salinestris,* and *A. vinelandii*), formed a distinct clade branching in the proximity of Stutzeri and Linyingensis clades (Figure 1; Supplementary Figure S2). Similar branching of *Azotobacter* species has been reported in earlier work (Jun et al., 2016; Hesse et al., 2018; Lalucat et al., 2022). Our analyses have identified 10 CSIs which are exclusively found in all four *Azotobacter* species providing reliable means for the demarcation of this clade. Partial sequence information for one of the CSIs specific for this genus, found in the alginate export family protein, is shown in Figure 7B. Detailed sequence information for this CSI and nine other CSIs specific for this genus is provided in Supplementary Figures S86–S95, and some of their sequence characteristics are listed in Table 5.

Azomonas is another genus whose members branch in between *Pseudomonas* species (Figure 1; Supplementary Figure S2; Young and Park, 2007; Kennedy and Rudnick, 2015; Rudra and Gupta, 2021; Lalucat et al., 2022). The two *Azomonas* species included in our analyses (*viz.*, *A. agilis* and *A. macrocytogenes*) formed a distinct cluster in our phylogenomic trees (Figure 1; Supplementary Figure S2). The distinctness of this clade is also supported by five CSIs identified in this work, which are exclusively shared by these two species. Sequence information for one of these CSIs, containing a five aa insertion within the protein succinate dehydrogenase flavoprotein, is shown in Figure 7C. Detailed sequence information for this CSI and the other four CSIs specific for this genus are provided in the Supplementary Figures S96–S100, and a summary of some of their sequence characteristics is listed in Table 5.

Discussion

The genus *Pseudomonas* is one of the earliest known and largest prokaryotic genera encompassing a large assemblage of organisms exhibiting enormous genetic and metabolic diversity (Palleroni, 2005;Peix et al., 2009; Silby et al., 2011; Palleroni, 2015). The nomenclature type of this genus, *P. aeruginosa*, is an important human pathogen capable of causing a wide array of life-threatening acute and chronic diseases (Lund-Palau et al., 2016; Rossi et al., 2021). However, this genus also includes some animals and plant pathogenic species, as well as other economically and ecologically significant species (Desnoues et al., 2003; Silby et al., 2011; Xin et al., 2018). According to the LPSN (Parte et al., 2020), the genus *Pseudomonas* presently contains ≈310 species with validly published names. However, this number is increasing at a rapid pace (Girard et al., 2021), and in 2022 alone, more than 50 novel *Pseudomonas* species were listed in the LPSN server (Parte et al., 2020). As indicated in the introduction, and reviewed by others (Palleroni, 2010; Peix et al., 2018; Lalucat et al., 2022), evolutionary studies on the genus *Pseudomonas* have consistently shown that these species form multiple distinct clusters/clades, which are not specifically related to each other (Gomila et al., 2015; Hesse et al., 2018; Girard et al., 2021; Rudra and Gupta, 2021; Saati-Santamaría et al., 2021). Furthermore, it is generally recognized that of these species' clades, circumscription of the genus *Pseudomonas* should be limited to the "Aeruginosa clade" harboring its type species, whereas species from the other observed clades should be reclassified into either novel or existing genera. In recent years, although several *Pseudomonas* species from deep branching clusters have been reclassified into novel genera (*viz. Atopomonas, Chryseomonas, Halopseudomonas* and *Stutzerimonas*) (Rudra and Gupta, 2021; Saati-Santamaría et al., 2021; Lalucat et al., 2022), the task of reliably reclassifying majority (>90%) of the *Pseudomonas* species into well-demarcated genera has proven challenging.

With the aim of reliably demarcating some of the observed *Pseudomonas* species clades, we have conducted here comprehensive phylogenomic and comparative analyses on the genome sequences of *Pseudomonadaceae* species. In our phylogenomic trees, *Pseudomonas* species formed multiple distinct clades (Figure 1; Supplementary Figure S2), which are similar to those reported in earlier studies (Gomila et al., 2015; Peix et al., 2018; Girard et al., 2021;

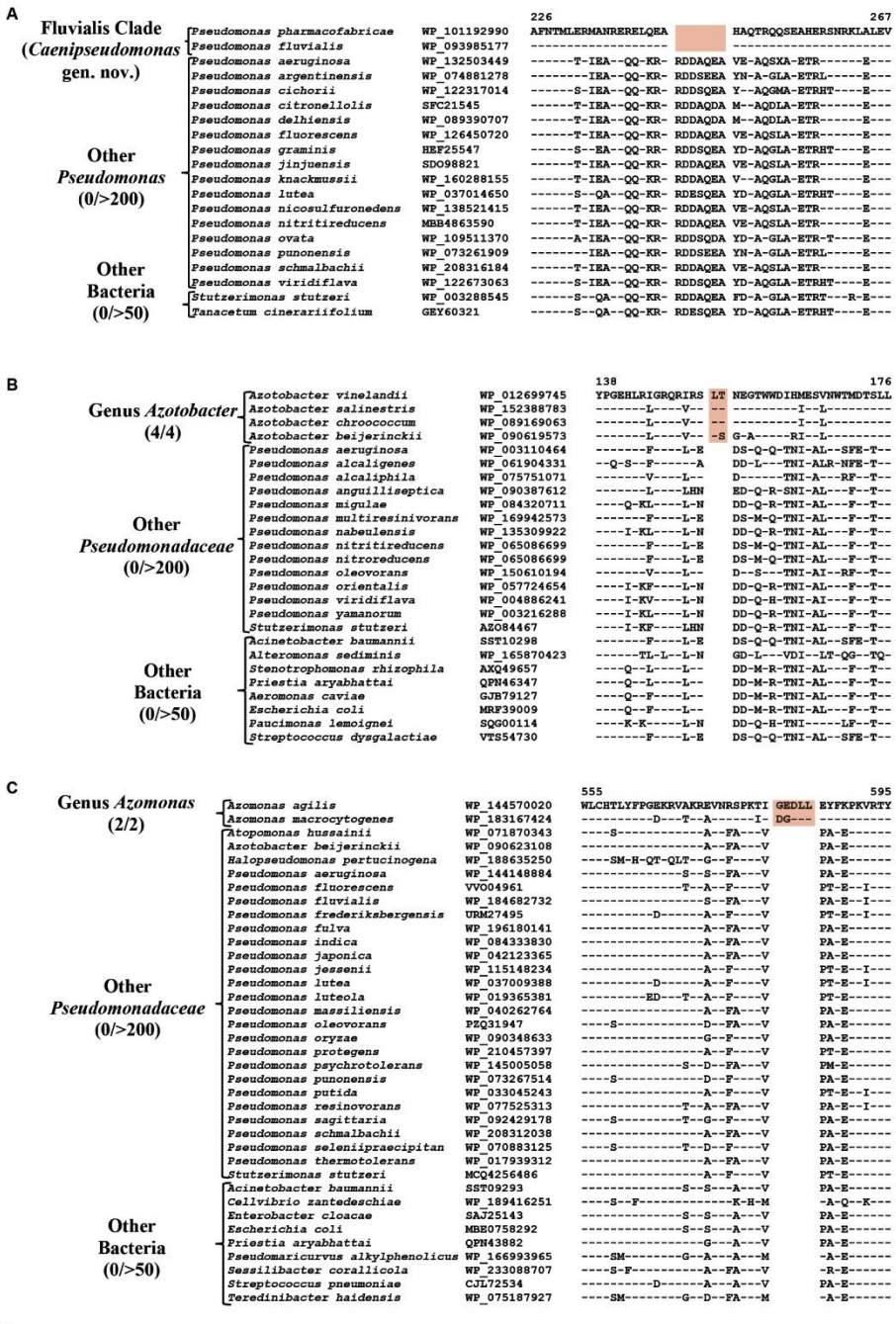

FIGURE 7
Partial sequence alignment of **(A)** ATP binding protein showing seven aa deletion within a conserved region (highlighted) that is uniquely shared by species from the Fluvialis clade. **(B)** A two aa insertion in a conserved region of the Alginate export family protein showing that is exclusively shared by species from the genus *Azotobacter*. **(C)** A five aa insertion in the protein Succinate dehydrogenase flavoprotein subunit which is specific for the species from genus *Azomonas*. Detailed sequence information for these CSIs and other CSIs specific for the Fluvialis clade and the *Azotobacter* and *Azomonas* genera are provided in Supplementary Figures S78–S100.

TABLE 5 Summary of CSIs specific for members of the Fluvialis clade, and the genera *Azotobacter* and *Azomonas*.

Protein name	Accession no	Figure number	Indel size	Indel location	Specificity
ATP-binding protein	WP_101192990	Figure 7A; **Supplementary Figure S78**	7 aa Del	226–267	Fluvialis clade (*Caenipseudomonas* gen. nov.)
Hypothetical protein	WP_101193738	**Supplementary Figure S79**	5 aa Del, 1 aa Del	146–197	
DUF2868 domain-containing protein	WP_101193981	**Supplementary Figure S80**	5 aa Del	415–452	
Hypothetical protein	WP_093984635	**Supplementary Figure S81**	2 aa Del	101–143	
Putative chorismate pyruvate-lyase	GGH90722	**Supplementary Figure S82**	2 aa Del	72–119	
Bifunctional aminoglycoside phosphotransferase/ATP-binding protein	WP_093984289	**Supplementary Figure S83**	2 aa Ins	77–117	
RDD family protein	WP_101192354	**Supplementary Figure S84**	1 aa Ins	160–200	
Translocation/assembly module TamB	WP_093986880	**Supplementary Figure S85**	2 aa Ins	493–539	
Alginate export family protein	WP_012699745	Figure 7B; **Supplementary Figure S86**	2 aa Ins	138–176	Genus *Azotobacter*
DNA polymerase III subunit alpha	WP_012702399	**Supplementary Figure S87**	1 aa Ins	88–132	
Pyrroloquinoline quinone biosynthesis protein	WP_152387189	**Supplementary Figure S88**	1 aa Del	238–276	
Protein-export chaperone SecB	WP_012699152	**Supplementary Figure S89**	1 aa Ins	33–71	
Protein Ion transporter	WP_012701585	**Supplementary Figure S90**	1 aa Del	25–70	
Cysteine synthase A	WP_012701826	**Supplementary Figure S91**	1 aa Ins	268–309	
DUF2066 domain-containing protein	WP_012702209	**Supplementary Figure S92**	2 aa Del	156–195	
GGDEF domain-containing phosphodiesterase	WP_012702302	**Supplementary Figure S93**	1 aa Del	389–431	
Flagellar hook-associated protein FlgL	WP_012700992	**Supplementary Figure S94**	1 aa Ins	131–167	
LLM class flavin-dependent oxidoreductase	WP_012699059	**Supplementary Figure S95**	3 aa Del	87–130	
Succinate dehydrogenase flavoprotein subunit	WP_144570020	Figure 7C; **Supplementary Figure S96**	5 aa Ins	555–595	Genus *Azomonas*
Mechanosensitive channel MscK	WP_183165886	**Supplementary Figure S97**	4 aa Del	790–819	
SPOR domain-containing protein	WP_144571310	**Supplementary Figure S98**	2 aa del	73–110	
Bifunctional [glutamate--ammonia ligase]-adenylyl-L-tyrosine phosphorylase adenylyltransferase	WP_183165719	**Supplementary Figure S99**	1 aa Del	153–185	
Alkyl hydroperoxide reductase subunit F	WP_144571471	**Supplementary Figure S100**	1 aa Del	366–398	

Lalucat et al., 2022) excepting some differences resulting from the inclusion of several new species in our analysis. However, while similar species clusters are observed in different studies, based on their branching in phylogenetic trees (see Figure 1; Supplementary Figure S2), which is dynamic in nature and influenced by multiple variables including addition of new species (Gupta, 1998; Baldauf, 2003; Felsenstein, 2004), it is difficult to reliably demarcate the boundaries of different clades. Thus, a major focus of this study was to identify robust molecular markers, which independent of phylogenetic analyses, can confirm the existence of observed species clades and can provide reliable means for their demarcation.

Although genome sequence based indices such as average nucleotide identity (ANIb) and genome to genome DNA hybridization (GGDC) are now widely used for the delimitation of species level taxa (Goris et al., 2007; Kim et al., 2014; Yarza et al., 2014), such methods including AAI (Konstantinidis and Tiedje, 2007) or POCP (Qin et al., 2014) have shown limited usefulness for the delineation of genus level taxa (Parks et al., 2018; Gupta, 2019;

Gupta and Kanter-Eivin, 2023). In the present work, while based on POCP and AAI values, some *Pseudomonas* species clades appear to be distinct (Table 1 and Supplementary Tables S2 and S3), for most of the observed clades these values generally show some overlap between the ingroup and outgroup species. Thus, based on these indices, it is difficult to reliably demarcate the boundaries of most of the clades. However, genome sequences are also enabling identification of highly specific molecular markers such as CSIs which are uniquely shared by different groups of organisms and provide dependable means for taxonomic and diagnostic studies (Gupta, 2014; Adeolu et al., 2016; Gupta, 2016; Gupta et al., 2020). As the CSIs in genes/proteins sequences result from rare genetic changes, their presence or absence in different species is generally not affected by most factors which can confound inferences from phylogenetic analyses (Baldauf and Palmer, 1993; Gupta, 1998; Rokas and Holland, 2000; Gupta, 2014, 2016). Furthermore, as the CSIs in different genes/proteins result from unrelated genetic changes, each of them provides independent evidence of a close and specific evolutionary relationship among a given group of species. In the present work, detailed analyses conducted on protein sequences from *Pseudomonadaceae* species, have identified 98 CSIs, which are specific for the species from 13 different *Pseudomonadaceae* species clades including the genera *Azomonas* and *Azotobacter*. Table 6 shows a summary of the CSIs that were identified for different *Pseudomonadaceae* clades along with the species that currently comprise these clades.

The results presented in Table 6 show that most of the *Pseudomonas* species clades, which are observed in our phylogenomic trees (Figure 1; Supplementary Figure S2), can now be robustly demarcated based on multiple identified CSIs, which are exclusively shared by the species from these clades. The genetic relatedness of the species from several of these clades is also supported by the results from AAI and POCP indices (Table 1). However, one clade for which CSIs were not identified is the Anguilliseptica clade. Species from this clade do not also form a well-resolved and strongly supported lineage in our phylogenetic trees (Figure 1; Supplementary Figure S2), and in earlier studies (Hesse et al., 2018; Busquets et al., 2021; Lalucat et al., 2022). In some phylogenetic trees [Supplementary Figure S2, unpublished results, and (Hesse et al., 2018)], one or more species from this clade (*viz. P. cuatrocienegasensis*) branch outside this clade. The results from AAI and POCP analyses (Table 1) also do not support the distinctness of this clade. All these observations indicate that the Anguilliseptica clade is not a trustworthy lineage and the cladistic relationships of species from this clade need to be further investigated. Of the CSIs identified by our analysis, six are uniquely shared by different species from the "Aeruginosa clade," providing reliable molecular means for the demarcation/circumscription of this clade representing the genus *Pseudomonas sensu stricto*. Our analyses have also identified multiple CSIs reliably demarcating the species from Alcaligenes, Fluvialis, Linyingensis, Oleovorans, Resinovorans, Straminea, and Thermotolerans clades. Based on the strong and consistent evidence provided by phylogenomic analyses and

TABLE 6 Summary of different *Pseudomonadaceae* species clades reliably demarcated based on phylogenomic analyses and identified CSIs specific for these clades.

Clade name (Genus name)	Number of CSIs	Species composition of the clades
"Aeruginosoa clade" (*Pseudomonas sensu stricto*)	6	*P. aeruginosa, P. citronellolis, P. delhiensis, P. humi, P. jinjuensis, P. knackmussii, P. multiresinivorans, P. nicosulfuronedens, P. nitritireducens, P. nitroreducens, P. paraeruginosa, P. panipatensis, "P. pseudonitroreducens," P. schmalbachii.*
Alcaligenes clade (*Aquipseudomonas* gen. nov.)	6	*P. alcaligenes, P. campi, P. guryensis, P. ullengensis*
Genus *Azomonas*	5	*A. agilis A. macrocytogenes*
Genus *Azotobacter*	10	*A. chroococcum, A. beijerinckii, A. salinestris, A. vinelandii.*
Flexibilis clade (Genus *Serpens* emend.)	3	*P. flexibilis, "Serpens gallinarum," P. tuomuerensis.*
Fluvialis clade (*Caenipseudomonas* gen. nov.)	8	*P. fluvialis, P. pharmacofabricae*
Linyingensis clade (*Geopseudomonas* gen. nov.)	15	*P. guangdongensis, P. aromaticivorans, P. linyingensis, "P. oryzae," P. oryzagri, P. sagittaria*
Oleovorans clade (*Ectopseudomonas* gen. nov.)	5	*P. alcaliphila, P. chengduensis, P. composti, P. guguanensis, P. hydrolytica, "P. indoloxydans," P. khazarica, P. mendocina, P. oleovorans, P. pseudoalcaligenes, "P. sediminis," "P. sihuiensis," P. toyotomiensis, "P. wenzhouensis," P. yangonensis*
Oryzihabitans clade (Genus *Chryseomonas* emend.)	11	*C. asuensis, C. duriflava, C. luteola, P. oryzihabitans, P. psychrotolerans, P. rhizoryzae, P. zeshuii*
Resinovorans clade (*Metapseudomonas* gen. nov.)	5	*P. boanensis, P. furukawaii, P. lalkuanensis, P. otitidis, P. resinovorans, P. tohonis*
Straminea clade (*Phytopseudomonas* gen. nov.)	12	*P. argentinensis, P. daroniae, P. dryadis, P. flavescens, P. punonensis, P. seleniipraecipitans, P. straminea.*
Stutzeri clade (Genus *Stutzerimonas*)	7	*S. azotifigens, S. balearica, S. chloritidismutans, S. decontaminans, S. degradans, S. frequens, S. kirkiae, S. kunmingensis, S. nitritotolerans, S. nosocomialis, S. perfectomarina, S. stutzeri, S. tarimensis, S. xanthomarina, S. zhaodongensis, S. urumqiensis, "P. lopnurensis," "P. phenolilytica," "P. oligotrophica," "P. saudiphocaensis," "P. songnenensis."*
Thermotolerans clade (*Zestomonas* gen. nov.)	5	*P. carbonaria, P. cavernae, P. insulae, P. thermotolerans*

identified molecular signatures supporting the distinctness of these clades, we are proposing that the species from the above noted clades should be reclassified into the following novel genera *Aquipseudomonas* gen. nov., *Caenipseudomonas* gen. nov., *Geopseudomonas* gen. nov., *Ectopseudomonas* gen. nov., *Metapseudomonas* gen. nov., *Phytopseudomonas* gen. nov., and *Zestomonas* gen. nov., respectively (Table 6). Our work has also identified 11 CSIs which are shared by all species from the Oryzihabitans clade providing robust means for the demarcation of species from this clade. Previously, only three species, which form a subclade of the Oryzihabitans clade, were reclassified into the genus *Chryseomonas* (Saati-Santamaría et al., 2021). Based on the results presented, we are proposing that the other species from this clade should also be transferred into the emended genus *Chryseomonas*. Species from the Flexibilis clade containing *P. flexibilis* are also transferred into the emended genus *Serpens*. Seven identified CSIs are commonly shared by all 13 species from the *Stutzerimonas* clade (Lalucat et al., 2022) providing robust molecular means for the demarcation of this genus. Lastly, multiple CSIs identified by our analyses are specific for the genera *Azomonas* and *Azotobacter* providing trustworthy means for the demarcation of these genera in molecular terms. As the identified CSIs provide important diagnostic characteristics of the above noted genera, we are also providing emended descriptions of these genera to include this information.

Although the present work represents a significant step toward clarifying the evolutionary relationships and classification scheme for *Pseudomonas* species, a vast majority of *Pseudomonas* species representing more than two thirds of the known species (see Supplementary Figure S1), are part of the Fluorescens superclade. As seen from Supplementary Figure S1, this large lineage is comprised of multiple clades and subclades (Palleroni, 2015; Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020; Girard et al., 2021). To develop a reliable classification scheme for all *Pseudomonas* species, it will be necessary to reliably distinguish and demarcate different species clades within the Fluorescens superclade and reclassify them appropriately. In view of this consideration, despite our reliable demarcation of the genus *Pseudomonas sensu stricto*, an emended description of this genus is not proposed, until most other *Pseudomonas* species are reliably classified.

All newly proposed genera and other studied genera/clades in this work have been circumscribed based on their harboring multiple uniquely shared CSIs. One notable characteristic of the CSIs, which is of much importance for classification purposes, is that these markers exhibit high degree of predictive ability to be found in other (uncharacterized or unidentified) members of a given group/taxon (Bhandari et al., 2013; Gupta, 2014, 2016; Dobritsa and Samadpour, 2019; Patel and Gupta, 2020; Montecillo and Bae, 2022). Thus, the CSIs specific for the genus *Halopseudomonas* identified in our earlier work (Rudra and Gupta, 2021) are also present in all newly described species from this genus (Supplementary Figure S2). Similarly, the CSIs specific for the genus *Atopomonas* were also present in a newly described species from this genus (Li et al., 2023). Due to the demonstrated predictive abilities of the CSIs to be found in other members of specific taxa, we have recently developed a web-based tool/server,[3] which can predict taxonomic affiliation based on the presence of known taxon-specific CSIs in a genome sequence (Gupta and Kanter-Eivin, 2023). Therefore, upon the addition of information for these newly identified CSIs to the AppIndels server, it should greatly facilitate the classification of both cultured and uncultured isolates related to the described taxa (Gupta and Patel, 2019). The CSIs specific for different taxa also provide useful means for the development of sensitive and specific diagnostic tests using *in silico* and experimental methods (Ahmod et al., 2011; Wong et al., 2014). Lastly, the earlier work on CSIs show that these molecular characteristics are functionally important for the group of organisms for which they are specific (Singh and Gupta, 2009; Khadka et al., 2020). Hence, genetic, and biochemical studies on the identified CSIs could lead to the discovery of novel biochemical and/or other characteristics of different groups of organisms.

The descriptions of different novel genera proposed and other emended genera are given below. The new name combinations for different species resulting from the proposed taxonomic changes are provided in Tables 7, 8. The names for the newly proposed genera are generally based on some characteristics of the proposed group of species.

Description of the genus *Aquipseudomonas* gen. nov.

Aquipseudomonas (A.qui.pseu.do.mo'nas. L. fem. n. *aqua*, water; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Aquipseudomonas, Pseudomonas*-like species isolated from water).

Cells are Gram-stain negative, motile and rod shaped. The species are aerobic in respiration and have been isolated from soil and swimming pool water. Optimum temperature for growth ranges from 30 – 37°C with <2% (w/v) NaCl and pH range from 4–10. Genome sizes for the species vary from 4.3 Mb to 4.6 Mb and the GC content ranges from 63.3 to 65.5%. Of the species from this genus, the type species *A. alcaligenes* can degrade polycyclic aromatic hydrocarbons and has been proven useful for bioremediation of oil pollution, pesticide substances, and certain chemical substances. Species from this genus form a strongly supported clade in phylogenomic tree based on large datasets of concatenated proteins. Additionally, species from this genus can be reliably distinguished from all other *Pseudomonadaceae* genera based on six CSIs (Table 2) which are exclusively found in the species from this genus. New name combinations for the species that are part of this genus are provided in Table 7.

The type species of this genus is *Aquipseudomonas alcaligenes.*

Description of the genus *Caenipseudomonas* gen. nov.

Caenipseudomonas (Cae.ni.pseu.do.mo'nas. L. neut. n. *caenum*, mud; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Caenipseudomonas*, *Pseudomonas*-like organism(s) isolated from river sediments).

Cells are strictly aerobic, Gram-stain-negative, non-fluorescent and occur mostly as short rods. Cells are motile and contain a single polar flagellum. Chemoorganotrophic growth. Species have been isolated from river sediment, and wastewater sample from a pharmaceutical company. Growth occurs in the temperature range from 4-22°C with optimum growth occurring between 25-35° C at

3 AppIndels.com

TABLE 7 Descriptions of the new name combinations for different proposed genera.

New name combination and etymology	Basonym	Description	Type strain
Aquipseudomonas gen. nov.			
Aquipseudomonas alcaligenes comb. nov. (type species) (al.*ca*.li'ge.nes. N.L. n. *alcali*, alkali; from Arabic article *al*, the; from Arabic masc. n. *qaly*, ashes of saltwort; Gr. suff. *-genes*, producing; from Gr. ind. v. *gennaô*, to produce; N.L. part. adj. *alcaligenes*, alkali-producing)	*Pseudomonas alcaligenes* Monias, 1928 (Approved Lists 1980).	The description of this species is the same as provided by Monias (1928).	ATCC 14909; CCUG 1425; CCUG 1425 A; CFBP 2437; CIP 101034; DSM 50342; IFO 14159; JCM 5967; LMG 1224; NBRC 14159; NCCB 76044; NCTC 10367; VKM B-2171.
Aquipseudomonas campi comb. nov. (cam'pi. L. gen. n. *campi*, of a field, of grassland)	*Pseudomonas campi* Timsy et al., 2021	The description of this species is the same as provided by Timsy et al. (2021).	31,521; DSM 110222; LMG 31521; S1-A32-2
Aquipseudomonas guryensis comb. nov. (gu.ryen'sis. N.L. fem. adj. *guryensis*, pertaining to Gurye, a geographic location where the type strain was isolated)	*Pseudomonas guryensis* Kim et al., 2021.	The description of this species is the same as provided by Kim et al. (2021).	JCM 34509; KCTC 82228; SR9.
Aquipseudomonas ullengensis comb. nov. (ull.eng.en'sis. N.L. fem. adj. *ullengensis*, pertaining to Ulleng Island, a geographic location where the type strain was isolated)	*Pseudomonas ullengensis* Kim et al., 2021.	The description of this species is the same as provided by Kim et al. (2021).	JCM 34510; KCTC 82229; UL070.
Caenipseudomonas gen. nov.			
Caenipseudomonas fluvialis comb. nov. (type species) (flu.vi.a'lis. L. fem. adj. *fluvialis*, belonging to a river, the source of the isolate)	*Pseudomonas fluvialis* Sudan et al., 2018.	The description of this species is the same as provided by Sudan et al. (2018).	ASS-1; CCM 8778; KCTC 52437.
Caenipseudomonas pharmacofabricae comb. nov. (phar.ma.co.fa'bri.cae. N.L gen. n. pharmacofabricae from a pharmaceutical factory)	*Pseudomonas pharmafabricae* Yu et al., 2018.	The description of this species is the same as provided by Yu et al. (2018).	CGMCC 1.15498; JCM 31306; ZYSR67-Z.
Ectopseudomonas gen. nov.			
Ectopseudomonas oleovorans comb. nov. (type species) (o.le.o.vo.rans. L. neut. n. *oleum*, oil; L. pres. part. *vorans*, eating, devouring; N.L. part. adj. *oleovorans*, oil devouring)	*Pseudomonas oleovorans* Lee and Chandler, 1941 (Approved Lists 1980).	The description of this species is the same as provided by Lee and Chandler (1941).	ATCC 8062; CCUG 2087; CFBP 5589; CIP 59.11; DSM 1045; IFO 13583; JCM 11598; LMG 2229; NBRC 13583; NCIB 6576; NCIMB 6576; NCTC 10692; NRRL B-778; VKM B-1522.
Ectopseudomonas alcaliphila comb. nov. (al.*ca*.li.phi'la. N.L. n. *alcali*, alcali (from Arabic article al, the; Arabic n. qaliy, ashes of saltwort); N.L. fem. adj. suff. *-phila*, friend, loving; from Gr. fem. adj. *philê*, loving; N.L. fem. adj. *alcaliphila*, liking alkaline environments)	*Pseudomonas alcaliphila* Yumoto et al., 2001.	The description of this species is the same as provided by Yumoto et al. (2001).	AL15-21; DSM 17744; IAM 14884; JCM 10630; NBRC 102411.
Ectopseudomonas chengduensis comb. nov. (cheng.du.en'sis. N.L. fem. adj. *chengduensis*, pertaining to Chengdu, where the type strain was isolated)	*Pseudomonas chengduensis* Tao et al., 2014.	The description of this species is the same as provided by Tao et al. (2014).	CGMCC 2318; DSM 26382; MBR.
Ectopseudomonas composti comb. nov. (com.pos'ti. N.L. gen. n. *composti*, of compost, from which strains were first isolated)	*Pseudomonas composti* Gibello et al., 2011.	The description of this species is the same as provided by Gibello et al. (2011).	C2; CCUG 59231; CECT 7516; DSM 25648.
Ectopseudomonas guguanensis comb. nov. (gu.guan.en'sis. N.L. fem. adj. *guguanensis*, of or pertaining to Guguan, the location of a favorite hot spring attraction in Taiwan)	*Pseudomonas guguanensis* Liu et al., 2013.	The description of this species is the same as provided by Liu et al. (2013).	BCRC 80438; CC-G9A; JCM 18416.
Ectopseudomonas hydrolytica comb. nov. (hy.dro.ly'ti.*ca*. Gr. neut. n. *hydôr*, water; Gr. masc. adj. *lytikos*, dissolving, splitting; N.L. fem. adj. *hydrolytica*, splitting with water, referring to the hydrolytic enzymatic activity of the bacterium).	*Pseudomonas hydrolytica* Zhou et al., 2020.	The description of this species is the same as provided by Zhou et al. (2020).	CCTCC AB 2018053; DSM 106702; DSWY01.

(Continued)

TABLE 7 (Continued)

New name combination and etymology	Basonym	Description	Type strain
Ectopseudomonas khazarica comb. nov. (kha.za'ri.*ca*. N.L. fem. adj. *khazarica*, pertaining to Khazar, a lake in the north of Iran as the largest lake in the world, from where the organism was isolated)	*Pseudomonas khazarica* Tarhriz et al., 2020.	The description of this species is the same as provided by Tarhriz et al. (2020).	KCTC 52410; LMG 29674; Tbz2.
Ectopseudomonas mendocina comb. nov. (men.do.ci.na. N.L. fem. adj. *mendocina*, pertaining to Mendoza (Argentina))	*Pseudomonas mendocina* Palleroni et al., 1970 (Approved Lists 1980).	The description of this species is the same as provided by Palleroni et al. (1970).	ATCC 25411; CCUG 1781; CFBP 2434; CIP 75.21; DSM 50017; IFO 14162; JCM 5966; LMG 1223; NBRC 14162; NCCB 76043; NCTC 10897; VKM B-972.
Ectopseudomonas pseudoalcaligenes comb. nov. (pseu.do.al.*ca*.li'ge.nes. Gr. masc. adj. *pseudes*, false; N.L. n. *alcali*, alkali; from Arabic article *al*, the; from Arabic masc. n. *qaly*, ashes of saltwort; Gr. suff. *-genes*, producing; from Gr. ind. v. *gennaô*, to produce; N.L. pres. part. *alcaligenes*, alkali-producing; N.L. part. adj. *pseudoalcaligenes*, false alkali producing)	*Pseudomonas pseudoalcaligenes* Stanier et al., 1966 (Approved Lists 1980).	The description of this species is the same as provided by Stanier et al. (1966).	ATCC 17440; CCUG 51525; CFBP 2435; CIP 66.14; DSM 50188; IFO 14167; JCM 5968; LMG 1225; NBRC 14167; NCCB 76045; NCTC 10860.
Ectopseudomonas toyotomiensis comb. nov. (to.yo.to.mi.en'sis. N.L. fem. adj. *toyotomiensis*, pertaining to Toyotomi, where the type strain was isolated)	*Pseudomonas toyotomiensis* Hirota et al., 2011.	The description of this species is the same as provided by Hirota et al. (2011).	DSM 26169; HT-3; JCM 15604; NCIMB 14511.
Ectopseudomonas yangonensis comb. nov. (yan.gon.en'sis. N.L. fem. adj. *yangonensis*, from or originating from Yangon, Myanmar, where the type strain was isolated)	*Pseudomonas yangonensis* Tohya et al., 2020.	The description of this species is the same as provided by Tohya et al. (2020).	JCM 33396; LMG 31602; MY50.
***Geopseudomonas* gen. nov.**			
Geopseudomonas sagittaria comb. nov. (type species) (sa.git.ta'ria. L. fem. adj. *sagittaria*, pertaining to the constellation Sagittarius as the novel species was isolated during the month of November, the birthday of first author (Shih-Yao Lin) of the paper describing this species; from L. masc. adj. *sagittarius*, the constellation Sagittarius)	*Pseudomonas sagittaria* Lin et al., 2013.	The description of this species is the same as provided by Lin et al. (2013).	BCRC 80399; CC-OPY-1; DSM 27945; JCM 18195.
Geopseudomonas aromaticivorans comb. nov. (a.ro.ma.ti.ci.vo'rans. L. masc. adj. *aromaticus*, fragrant; L. pres. part. *vorans*, devouring; N.L. part. adj. *aromaticivorans*, devouring aromatic compounds)	*Pseudomonas aromaticivorans* Banerjee et al., 2022.	The description of this species is the same as provided by Banerjee et al. (2022).	LMG 32466; MAP12; NCAIM B.02668.
Geopseudomonas linyingensis comb. nov. (lin.ying.en'sis. N.L. fem. adj. *linyingensis*, pertaining to Linying, in Henan province, China, where the type strain was isolated).	*Pseudomonas linyingensis* He et al., 2012.	The description of this species is the same as provided by He et al. (2012).	CGMCC 1.10701; LMG 25967; LYBRD3-7
Geopseudomonas guangdongensis comb. nov. (guang.dong.en'sis. N.L. fem. adj. *guangdongensis*, of or pertaining to Guangdong, a province in south-east China, from where the type strain was isolated).	*Pseudomonas guangdongensis* Yang et al., 2013.	The description of this species is the same as provided by Yang et al. (2013).	CCTCC AB 2012022; DSM 100318; KACC 16606; SgZ-6.
Geopseudomonas oryzagri comb. nov. (o.ryz.a'gri. L. fem. n. *oryza*, rice; L. n. *ager*, a field; N.L. gen. n. *oryzagri*, of a rice field)	*Pseudomonas oryzagri* Huq et al., 2022.	The description of this species is the same as provided by Huq et al. (2022).	CGMCC 1.18518; KACC 22005; MAHUQ-58

pH between 7–8 in presence of 0–2% (w/v) NaCl concentration. Genome size range is from 3.3–3.4 Mb and the GC content is 62.6%. Species from this genus form a distinct lineage in phylogenomic trees based on large datasets of proteins, as well as in trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB*, *rpoB*, and *rpoD* genes. In addition, species from this genus can be reliably distinguished based on eight exclusively shared CSIs listed in Table 5. The new name combinations for species from this genus are provided in Table 7.

The type species is *Caenipseudomonas fluvialis*.

TABLE 8 Descriptions of the new name combinations for different proposed and emended genera.

New name combination and etymology	Basonym	Description	Type strain
Metapseudomonas gen. nov.			
Metapseudomonas resinovorans comb. nov. (type species) (re.si.no.vo'rans. L. fem. n. *resina*, resin; L. pres. part. *vorans*, eating, devouring; N.L. part. adj. *resinovorans*, resin devouring)	*Pseudomonas resinovorans* Delaporte et al., 1961 (Approved Lists 1980).	The description of this species is the same as provided by Delaporte et al. (1961).	ATCC 14235; CCUG 2473; CCUG 4439; CFBP 5590; CIP 61.9; DSM 21078; LMG 2274; NRRL B-2649.
Metapseudomonas boanensis comb. nov. (bo.a.nen'sis. N.L. fem. adj. *boanensis*, pertaining to the Boane District in Mozambique)	*Pseudomonas boanensis* Nicklasson et al., 2022.	The description of this species is the same as provided by Nicklasson et al. (2022).	CCUG 62977; CECT 30359; DB1.
Metapseudomonas furukawaii comb. nov. (fu.ru.ka.wa'i.i. N.L. gen. masc. n. *furukawaii*, of Furukawa named after Kensuke Furukawa, a Japanese microbiologist who notably contributed to the understanding of microbial and molecular biological mechanisms involved in biphenyl/PCB degradation)	*Pseudomonas furukawaii* Kimura et al., 2018.	The description of this species is the same as provided by Kimura et al. (2018).	DSM 10086; KF707; NBRC 110670.
Metapseudomonas lalkuanensis comb. nov. (lal.ku.an.en'sis. N.L. fem. adj. *lalkuanensis*, pertaining to Lalkuan, a town in the Nainital district of Uttarakhand, India, where the type strain was isolated)	*Pseudomonas lalkuanensis* Thorat et al., 2020.	The description of this species is the same as provided by Thorat et al. (2020).	CCUG 73691; KCTC 72454; MCC 3792; PE08.
Metapseudomonas otitidis comb. nov. (o.ti'ti.dis. Gr. neut. n. *oûs* (gen. *ôtos*), ear; N.L. suff. *-itis -idis*, used in names of inflammations; N.L. gen. Fem. n. *otitidis*, of inflammation of the ear)	*Pseudomonas otitidis* Clark et al., 2006.	The description of this species is the same as provided by Clark et al. (2006).	ATCC BAA-1130; DSM 17224; MCC 10330.
Metapseudomonas tohonis comb. nov. (to.ho'nis. N.L. gen. n. *tohonis*, of Toho University, where the type strain was first isolated and analyzed)	*Pseudomonas tohonis* Yamada et al., 2021	The description of this species is the same as provided by Yamada et al. (2021)	GTC 22698; NCTC 14580; TUM18999
Phytopseudomonas gen. nov.			
Phytopseudomonas straminea comb. nov. (type species) (stra.mi.ne'a. L. fem. adj. *straminea*, made of straw)	*Pseudomonas straminea* corrig. Iizuka and Komagata, 1963 (Approved Lists 1980).	The description of this species is the same as provided by Iizuka and Komagata (1963).	ATCC 33636; CCUG 12539; CIP 106745; DSM 17727; IAM 1598; IFO 16665; JCM 2783; NBRC 16665; NRIC 164.
Phytopseudomonas argentinensis comb. nov. (ar.gen.ti.nen'sis. N.L. fem. adj. *argentinensis*, pertaining to the Argentine, of the Argentine)	*Pseudomonas argentinensis* Peix et al., 2005.	The description of this species is the same as provided by Peix et al. (2005).	CECT 7010; CH01; DSM 17259; LMG 22563.
Phytopseudomonas daroniae comb. nov. (da.ron.i'ae. N.L. gen. fem. n. *daroniae*, from Daron, the Celtic goddess of oak).	*Pseudomonas daroniae* Bueno-Gonzalez et al., 2019.	The description of this species is the same as provided by Bueno-Gonzalez et al. (2019).	FRB228; LMG 31088; NCPPB 4672.
Phytopseudomonas dryadis comb. nov. (dry.a'dis. L. gen. fem. n. *dryadis*, of a Dryad, of an oak tree nymph)	*Pseudomonas dryadis* Bueno-Gonzalez et al., 2019.	The description of this species is the same as provided by Bueno-Gonzalez et al. (2019).	FRB230; LMG 31087; NCPPB 4673.
Phytopseudomonas flavescens comb. nov. (fla.ves'cens. L. part. adj. *flavescens*, becoming golden yellow)	*Pseudomonas flavescens* Hildebrand et al., 1994.	The description of this species is the same as provided by Hildebrand et al. (1994).	ATCC 51555; B62; CCUG 49622; CFBP 5586; CIP 104204; DSM 12071; JCM 21586; LMG 18387; NBRC 103044; NCPPB 3063.
Phytopseudomonas gununensis comb. nov. (pu.no.nen'sis. N.L. fem. adj. *gununensis*, of or belonging to Puno, a region of Peru where the type strain was isolated)	*Pseudomonas gununensis* Ramos et al., 2013.	The description of this species is the same as provided by Ramos et al. (2013).	CECT 8089; DSM 27507; LMG 26839; LMT03.
Phytopseudomonas seleniipraecipitans comb. nov. (se.le.ni.i.prae.ci'pi.tans. N.L. neut. n. *selenium*, selenium; L. part. adj. *praecipitans*, precipitating; N.L. part. adj. *seleniipraecipitans*, selenium precipitating, referring to the organism's ability to remove the selenium oxyanion selenite from aqueous solution)	*Pseudomonas seleniipraecipitans* corrig. Hunter and Manter, 2011.	The description of this species is the same as provided by Hunter and Manter (2011).	CA5; DSM 25106; LMG 25475; NRRL B-51283.

(Continued)

TABLE 8 (Continued)

New name combination and etymology	Basonym	Description	Type strain
Zestomonas gen nov.			
Zestomonas thermotolerans comb. nov. (type species) (ther.mo.to'le.rans. Gr. masc. adj. *thermos*, hot; N.L. part. adj. *thermotolerans*, able to tolerate high temperatures)	*Pseudomonas thermotolerans* Manaia and Moore, 2002.	The description of this species is the same as provided by Manaia and Moore (2002).	CM3; DSM 14292; LMG 21284.
Zestomonas carbonaria comb. nov. (car.bo.na'ri.a. L. fem. adj. *carbonaria*, of or relating to charcoal, the source of isolation)	*Pseudomonas carbonaria* Kämpfer et al., 2021.	The description of this species is the same as provided by Kämpfer et al. (2021).	CCM 9017; CIP 111764; DSM 110367; Wesi-4.
Zestomonas insulae comb. nov. (in'su.lae. L. gen. fem. n. *insulae*, of an island, referring to the source of isolation of the type strain)	*Pseudomonas insulae* Lee et al., 2022.	The description of this species is the same as provided by Lee et al. (2022).	JCM 34511; KCTC 82407; UL073.
Zestomonas cavernae comb. nov. (*ca*.ver'nae. L. gen. n. *cavernae*, of a cave)	*Pseudomonas cavernae* Zhu et al., 2021	The description of this species is the same as provided by Zhu et al. (2021)	CGMCC 1.13586; K2W31S-8; KCTC 82191
Genus *Chryseomonas*			
Chryseomonas oryzihabitans comb. nov. (o.ry.zi.ha'bi.tans. L. fem. n. *oryza*, rice; L. pres. part. *habitans*, inhabiting; N.L. part. adj. *oryzihabitans*, rice inhabiting)	*Pseudomonas oryzihabitans* Kodama et al., 1985.	The description of this species is the same as provided by Kodama et al. (1985).	AJ 2197; ATCC 43272; CCUG 12540; CIP 102996; DSM 6835; IAM 1568; JCM 2952; KS0036; I.-1; LMG 7040; NBRC 102199.
Chryseomonas psychrotolerans comb. nov. (psy.chro.to'le.rans. Gr. masc. adj. *psychros*, cold; L. pres. part. *tolerans*, tolerating; N.L. part. adj. *psychrotolerans*, cold-tolerating)	*Pseudomonas psychrotolerans* Hauser et al., 2004	The description of this species is the same as provided by Hauser et al. (2004).	C36; CCUG 51516; DSM 15758; LMG 21977.
Chryseomonas rhizoryzae comb. nov. (rhiz.o.ry'zae. Gr. fem. n. *rhiza*, root; L. fem. n. *oryza*, rice; N.L. gen. n. *rhizoryzae*, of rice root).	*Pseudomonas rhizoryzae* Wang et al., 2020.	The description of this species is the same as provided by Wang et al. (2020)	ACCC 61555; JCM 33201; RY24.
Chryseomonas zeshuii comb. nov. (ze.shu'i.i. N.L. gen. masc. n. *zeshuii*, of Ze-Shu, in honor of Ze-Shu Qian, a respected microbiologist, for his enormous contributions to promoting the development of soil microbiology in China)	*Pseudomonas zeshuii* Feng et al., 2012	The description of this species is the same as provided by Feng et al. (2012).	ACCC 5688; BY; BY-1; DSM 27927; KACC 15471.
Genus *Serpens*			
Serpens flexibilis comb. nov. (type species) (fle.xi'bi.lis. L. fem. adj. *flexibilis*, flexible, pliant)	*Pseudomonas flexibilis* Hespell, 1977; Shin et al., 2015.	The description of this species is the same as provided by Shin et al. (2015).	ATCC 29606; LMG 29034.
Serpens tuomuerensis comb. nov. (tuo.muer.en'sis. N.L. fem. adj. *tuomuerensis*, pertaining to the region of Tuomuer Peak of Tianshan Mountain, where the type strain was isolated)	*Pseudomonas tuomuerensis* Xin et al., 2009.	The description of this species is the same as provided by Xin et al. (2009).	78–123; CGMCC 1.1365; DSM 25351; JCM 14085.
Genus *Stutzerimonas*			
Stutzerimonas marianensis comb. nov. (ma.ri.an.en'sis. N.L. fem. adj. *marianensis*, pertaining to the Mariana Trench, the source of the type strain)	*Pseudomonas marianensis* Yang et al., 2022	The description of this species is the same as provided by Yang et al. (2022)	DSM 112238; MCCC 1 K05112; P S1

Description of the genus *Ectopseudomonas* gen. nov.

Ectopseudomonas (Ec.to.pseu.do.mo'nas. Gr. prep. *Ecto*, outside; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Ectopseudomonas*, a genus outside of *Pseudomonas*).

Cells are Gram-stain negative, motile and rod shaped. Excepting *E. chengduensis*, all other species from this genus are motile due to the presence of a polar flagellum. Species have been isolated from diverse sources including sea water, soil, hot spring, compost, and lake sediments, etc. Chemoorganotrophic life cycle. Most species grow aerobically; however, some are indicated to be facultatively anerobic. Colonies are generally brownish yellow. Growth can occur from 4°-42°C with optimum growth temperature between 30–37°C, with or without NaCl, in the pH range from 3.0–10.5 (optimum between pH 6–8). Genome sizes for known species vary from 4.5 Mb to 5.6 Mb and

the GC content ranges from 62.2 to 65.0%. Of the species from this genus, *E. mendocina* can degrade toluene and it is indicated to cause opportunistic nosocomial infections. Members of this genus form a monophyletic clade in phylogenetic trees based on concatenated sequences of several large datasets of core genome proteins. Additionally, species from this genus also generally cluster together in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, rpoB,* and *rpoD* genes. In addition of their distinct branching in phylogenetic trees, members of this genus can be reliably distinguished from other *Pseudomonadaceae* species based on five CSIs (Table 2) which in most cases are exclusively shared by the members of this genus. The new name combinations for species that are part of this genus are provided in Table 7.

The type species of this genus is *Ectopseudomonas oleovorans.*

Description of the genus *Geopseudomonas* gen. nov.

Geopseudomonas (Ge.o.pseu.do.mo'nas. Gr. fem. n. *gê*, the Earth; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Geopseudomonas*, *Pseudomonas* like organisms isolated from soil).

Strictly aerobic to facultatively anaerobic, rod-shaped bacteria. Motile due to the presence of one or more polar or peritrichous flagella. Chemoorganotrophs, with cells exhibiting Gram-stain negative staining response. Cells generally do not produce fluorescent pigments. Members have been isolated from diverse sources including paddy soil, electroactive biofilm, herbicide applied wheat field and oil contaminated soil. Optimum growth occurs in the range of 30–37°C, between pH 7–8, in medium containing 1–2% NaCl (w/v). Genome lengths of the species vary from 3.2 to 4.7 Mb, and GC contents vary from 66.4 to 68.3%. Members of this genus form a monophyletic clade in phylogenetic tree based on concatenated sequences for several large datasets of proteins. Species from this genus also cluster together in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, ropB,* and *rpoD* genes. In addition, the members of this genus can be reliably distinguished from all other *Pseudomonadaceae* genera by the 15 CSIs described in Table 3, which in most cases are exclusively shared by either all or most species from this genus. The new name combinations for species which are part of this genus are provided in Table 7.

The type species is *Geopseudomonas sagittaria.*

Description of the genus *Metapseudomonas* gen. nov.

Metapseudomonas (Me.ta.pseu.do.mo'nas. Gr. adv. *Meta*, besides; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Metapseudomonas*, a genus beside *Pseudomonas*).

Species of this genus are Gram-negative, motile, aerobic and rod shaped. Chemoorganotrophic growth, cells do not produce fluorescent pigments. Members have been isolated from different sources such as clinical samples, soil or oil of wood mills and biphenyl contaminated soil. Optimum growth temperature is in the range of 30-37°C. Genome sizes for known species are in the range of 6.1 Mb to 6.8 Mb and GC content varies from 64.2 to 66.80%. Species from this genus form a strongly supported clade in phylogenomic trees based on large datasets of proteins. In addition, most of the species from this genus also cluster together in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, ropB,* and *rpoD* genes. Importantly, the species from this genus can also be reliably distinguished from all other *Pseudomonadaceae* genera by the shared presence of five CSIs listed in Table 4. The new name combinations for the species of this genus are provided in Table 8.

The type species of this genus is *Metapseudomonas resinovorans.*

Description of the genus *Phytopseudomonas* gen. nov.

Phytopseudomonas (Phy.to.pseu.do.mo'nas. Gr. neut. n. *phyton*, plant; N.L. fem. n. *Pseudomonas*, a bacterial genus; N.L. fem. n. *Phytopseudomonas*, *Pseudomonas*-like species isolated from plants).

Cells are Gram-stain negative, motile due to the presence of a polar flagellum, aerobic, and rod shaped. Chemoorganotrophs. Most species have been isolated from different plant sources such as *Quercus robur* stem tissues, straw grass, rice paddy, walnut blight cankers etc. All species produce a diffusible fluorescent pigment. Optimum temperature for growth is between 25-30°C, with <4% (w/v) or without NaCl in the pH range from 6–8. Genome sizes for the species vary from 4.5 Mb to 5.9 Mb and the GC content ranges from 61.5 to 65.0%. Members of this genus form a monophyletic clade in phylogenetic trees based on concatenated sequences of several large datasets of core genome proteins. Additionally, species from this genus also generally cluster together in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, ropB*, and *rpoD* genes. Additionally, members of this genus can be reliably distinguished from other *Pseudomonadaceae* genera based on the presence of 12 CSIs summarized in Table 3. which in most cases are exclusively present in the species from this genus. The new name combinations for species that are part of this genus are provided in Table 8.

The type species of this genus is *Phytopseudomonas straminea.*

Description of the genus *Zestomonas* gen. nov.

Zestomonas (Zes.to.mo'nas. Gr. masc. Adj. *zestos*, hot, boiling; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Zestomonas*, a monad that can grow at high temperature).

Aerobic, motile rods exhibiting Gram-negative staining response. Chemoorganotrophs. Species have been cultivated from different sources such as cooking water, forest soil, charcoal, and cave sediment. Temperature range for growth for species from this genus differs considerably. While the optimum growth of the type species *Zestomonas thermotolerans* occurs at 47°C (growth range 25–56°C), other species from this genus grow optimally at 28–30°C. Genome length ranges from 3.8 to 5.5 Mb and the GC content varies from 64.5 to 66.8%. Members of this genus form a monophyletic clade in phylogenomic tree based on concatenated sequences for several large datasets of proteins. In addition, members of this genus can be reliably distinguished from other *Pseudomonadaceae* genera by their uniquely sharing five CSIs listed in Table 4. New name combinations for the species from this genus are provided in Table 8.

The type species is *Zestomonas thermotolerans.*

Emended description of the genus *Azomonas* Winogradsky, 1938 (Approved lists 1980)

Azomonas (A.zo.mo.nas. N.L. pref. *Azo-*, pertaining to nitrogen; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Azomonas*, nitrogen monad).

Description of this genus is in large part based on that provided by Kennedy and Rudnick (2015) in the Bergey's Manual of Systematics of Archaea and Bacteria. Cells are Gram-stain variable or sometimes Gram-stain negative depending on the culture age, aerobic, ellipsoidal to rod shaped. Species are motile with peritrichous or lophotrichous polar flagella. Cells may occur singly, in pairs, or in clumps. All species fix atmospheric nitrogen under aerobic conditions. Alternative nitrogenases containing vanadium (nitrogenase-2) or iron (nitrogenase-3) may only be synthesized in Mo-deficient media. Cultures can grow both aerobically and microaerobically. Chemoorganotrophic. Sugars, alcohols, and organic acids are used as carbon sources. Ammonium salts and sometimes nitrate (*A. insignis* only) are used as nitrogen sources; amino acids are not used. Water-soluble and fluorescent pigments are produced by nearly all strains. Species are catalase positive. The optimum pH for nitrogen fixation is close to neutrality, but certain strains can also fix nitrogen at a pH of 4.6–4.8. Species isolated from water or soil. The G+C content of DNA from known species varies from 52.0–58.6% and their genome size ranges from 3.3 to 4.1 MB. Species belonging to this genus form a distinct clade in phylogenomic trees based on concatenated sequences of large number of proteins and in the tree based on 16S rRNA gene sequences. In addition, members of this genus can be reliably distinguished from *Azotobacter* as well as all other *Pseudomonadaceae* genera based on their exclusive sharing five CSIs described in this work (Table 5).

Type species is *Azomonas agilis* (Beijerinck, 1901) Winogradsky, 1938 (Approved Lists 1980).

Emended description of the genus *Azotobacter* Beijerinck, 1901 (Approved lists 1980)

Azotobacter (A.zo.to.bac.ter. N.L. neut. n. *azotum*, nitrogen; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Azotobacter*, a nitrogen rod).

Description of this genus is in large part based on that provided by Kennedy et al. (2015) in the 2015 Bergey's Manual of Systematics of Archaea and Bacteria. Cells range from straight rods with rounded ends to more ellipsoidal or coccoid. Motile with peritrichous flagella or nonmotile. Aerobic, having a strictly respiratory type of metabolism with oxygen as the terminal electron acceptor. Nitrogen is fixed under microaerobic conditions (2% oxygen), under full aerobiosis, or after adaptation in hyperbaric oxygen. N_2 fixation uses Mo-, V-, or Fe-containing nitrogenase enzymes, depending on the environmental metal supply. Water-soluble and water-insoluble pigments are produced by some strains. Growth is heterotrophic; sugars, alcohols, and salts of organic acids are used as carbon sources. Ammonium salts, nitrate, and urea are used as sources of fixed nitrogen. The pH range for growth is from 4.8 to 8.5, with optimum pH for diazotrophic growth between 7.0–7.5. Most isolates are from soil, but a few are from water. The GC content of the DNA varies from 65.5–67.5%. Genome size ranges from 4.9–5.4 Mb. Species belonging to this genus group together in phylogenetic trees based on 16S rRNA gene sequences, and in phylogenomic trees based on concatenated sequences of large number of proteins. In addition, members of this genus can be reliably distinguished from all other *Pseudomonadaceae* genera by 10 uniquely shared CSIs listed in Table 5.

Type species is *Azotobacter chroococcum* Beijerinck, 1901 (Approved Lists 1980).

Emended description of the genus *Chryseomonas* Holmes et al., 1986

Chryseomonas (Chry.se.o.mo'nas. Gr. masc. Adj. *chryseos*, golden; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Chryseomonas*, a yellow unit).

The description of this genus is partially based on that given by Holmes et al. (1986) for the type species (*C. polytricha*) of this genus. The cells are rod-shaped, Gram-negative, aerobic, and exhibit chemoorganotrophic growth. Except for *C. duriflava* (and its synonym *C. zeshuii*), which do not exhibit motility, cells from the other species are motile by either a single or several polar or trichous flagella. Known species have been isolated from diverse sources including rice seeds and paddy, desert soil, herbicide-contaminated soil, grass rhizosphere, clinical specimens, and medical clinic for small animals. *C. oryzihabitans* has been reported as pathogenic to plants and animals. Some species (*C. luteola*) can reduce nitrate. Growth can occur in the temperature range from 4–42°C with optimum growth occurring between 30 to 37°C at pH 7.0 (pH range 6–8) in medium supplemented with 1–2% (w/v) NaCl. The cells are catalase positive but oxidase negative. The GC content of species varies from 53.6 to 66.2% and their genome lengths range from 4.3 to 5.4 Mb. Species from this genus form a distinct clade in the phylogenomic trees based on a large number of proteins. Additionally, these species also cluster together in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, ropB,* and *rpoD* genes. Apart from their grouping together in phylogenetic trees, species from this genus can be reliably distinguished from all other *Pseudomonadaceae* genera by their 11 CSIs listed in Table 4, which in most cases are exclusively present in the species from this genus. New name combinations for four *Pseudomonas* species, which are transferred to this genus, are provided in Table 8.

Type species of this genus is *Chryseomonas polytrichia* (Holmes et al., 1986).

Emended description of the genus *Serpens* Hespell, 1977 (Approved lists 1980)

Description of this genus is modified from that given by Hespell (1977). Gram-negative, aerobic, rod-shaped, non-spore forming, bacterial cells. Cells from the type species, *S. flexibilis*, are very flexible, and motile due to containing a flagellum, and exhibit serpentine-like movement in agar gels. Metabolism is respiratory, and molecular oxygen serves as the terminal electron acceptor. *S. flexibilis* mainly

uses lactate as the energy and carbon source. Catalase and oxidase are produced. Temperature range for optimal growth is from 28 to 37°C. The G + C content of DNA ranges from 61.0–65.8 mol% and genome size varies from 3.8–3.9 Mb. Species from this genus form a monophyletic clade in the phylogenetic tree based on large dataset of proteins. The type species also forms a distinct lineage in phylogenetic trees based on *rpoD* gene, or concatenated partial sequences for the 16S rDNA, *gyrB, ropB,* and *rpoD* genes. Additionally, species from this genus can be reliably distinguished from other *Pseudomonadaceae* genera by the presence of three exclusively shared CSIs (Table 4). New name combinations for the two species which are part of this genus are provided in Table 8.

Type species of this genus is *Serpens flexibilis* Hespell, 1977 (Approved lists).

Emended description of the genus *Stutzerimonas* Lalucat et al., 2022

Stutzerimonas (Stut.ze.ri.mo'nas. L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Stutzerimonas*, monad of Stutzer, named in honor of Albert Stutzer, who in 1895 described the bacterium today known).

The description of this genus, especially in terms of its morphological, chemotaxonomic and growth characteristics, remains the same as provided by Lalucat et al. (2022). In addition to the genomic characteristics described by Lalucat et al. (2022), members of this genus can be reliably distinguished from other *Pseudomonadaceae* genera by seven novel CSIs identified in this study (listed in Table 3), which in most cases are exclusively found in the species from this genus. New name combination for *P. marianensis* (Table 8) is based on its branching in the 16S rRNA gene tree (Yang et al., 2022).

The type species is *Stutzerimonas stutzeri* (Lehmann and Neumann 1896) Lalucat et al. 2022.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

BR: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Validation. RG: Conceptualization, Funding acquisition, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a research grant (RGPIN-2019-06397) from the Natural Science and Engineering Research Council of Canada.

Acknowledgments

We thank Dr. Aharon Oren for checking the etymology of the proposed genera.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1273665/full#supplementary-material

References

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, S. (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. *Int. J. Syst. Evol. Microbiol.* 66, 5575–5599. doi: 10.1099/ijsem.0.001485

Ahmod, N. Z., Gupta, R. S., and Shah, H. N. (2011). Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* 87, 278–285. doi: 10.1016/j.mimet.2011.08.015

Anzai, Y., Kim, H., Park, J. Y., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int. J. Syst. Evol. Microbiol.* 50, 1563–1589. doi: 10.1099/00207713-50-4-1563

Anzai, Y., Kudo, Y., and Oyaizu, H. (1997). The phylogeny of the genera *Chryseomonas*, *Flavimonas*, and *Pseudomonas* supports synonymy of these three genera. *Int. J. Syst. Bacteriol.* 47, 249–251. doi: 10.1099/00207713-47-2-249

Azhar, E. I., Papadioti, A., Bibi, F., Ashshi, A. M., Raoult, D., and Angelakis, E. (2017). '*Pseudomonas saudimassiliensis*' sp. nov. a new bacterial species isolated from air samples in the urban environment of Makkah, Saudi Arabia. *New Microbes New Infect.* 16, 43–44. doi: 10.1016/j.nmni.2016.12.021

Baldauf, S. L. (2003). Phylogeny for the faint of heart: a tutorial. *Trends Genet.* 19, 345–351. doi: 10.1016/S0168-9525(03)00112-4

Baldauf, S. L., and Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U. S. A.* 90, 11558–11562. doi: 10.1073/pnas.90.24.11558

Banerjee, S., Bedics, A., Tóth, E., Kriszt, B., Soares, A. R., Bóka, K., et al. (2022). Isolation of *Pseudomonas aromaticivorans* sp. nov from a hydrocarbon-contaminated groundwater capable of degrading benzene-, toluene-, m- and p-xylene under microaerobic conditions. *Front. Microbiol.* 13:929128. doi: 10.3389/fmicb.2022.929128

Behera, P., Mahapatra, M., Seuylemezian, A., Vaishampayan, P., Ramana, V. V., Joseph, N., et al. (2018). Taxonomic description and draft genome of *Pseudomonas*

sediminis sp. nov., isolated from the rhizospheric sediment of *Phragmites karka*. *J. Microbiol.* 56, 458–466. doi: 10.1007/s12275-018-7549-x

Beijerinck, M. (1901). Über oligonitrophile Mikroben. *Zentralbl. Bakterol. Parasitenkd. Infektionskr. Hyg. Abt. II* 7, 561–582.

Bhandari, V., Ahmod, N. Z., Shah, H. N., and Gupta, R. S. (2013). Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *Int. J. Syst. Evol. Microbiol.* 63, 2712–2726. doi: 10.1099/ijs.0.048488-0

Bueno-Gonzalez, V., Brady, C., Denman, S., Plummer, S., Allainguillaume, J., and Arnold, D. (2019). *Pseudomonas daroniae* sp. nov. and *Pseudomonas dryadis* sp. nov., isolated from pedunculate oak affected by acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* 69, 3368–3376. doi: 10.1099/ijsem.0.003615

Busquets, A., Mulet, M., Gomila, M., and García-Valdés, E. (2021). *Pseudomonas lalucatii* sp. nov. isolated from Vallgornera, a karstic cave in Mallorca, Western Mediterranean. *Syst. Appl. Microbiol.* 44:126205. doi: 10.1016/j.syapm.2021.126205

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Clark, L. L., Dajcs, J. J., Mclean, C. H., Bartell, J. G., and Stroman, D. W. (2006). *Pseudomonas otitidis* sp. nov., isolated from patients with otic infections. *Int. J. Syst. Evol. Microbiol.* 56, 709–714. doi: 10.1099/ijs.0.63753-0

Delaporte, B., Raynaud, M., and Daste, P. (1961). Une bactérie du sol capable d'utiliser, comme source de carbone, la fraction fixe de certaines oléorésines, *Pseudomonas resinovorans* n. sp. *Compte Rendu des Séances de l'Académie des Sci.* 252, 1073–1075.

Desnoues, N., Lin, M., Guo, X., Ma, L., Carreño-Lopez, R., and Elmerich, C. (2003). Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology (Reading)* 149, 2251–2262. doi: 10.1099/mic.0.26270-0

Dobritsa, A. P., and Samadpour, M. (2019). Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *Int. J. Syst. Evol. Microbiol.* 69, 2057–2063. doi: 10.1099/ijsem.0.003431

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Associates, Inc.

Feng, Z., Zhang, J., Huang, X., Zhang, J., Chen, M., and Li, S. (2012). *Pseudomonas zeshuii* sp. nov., isolated from herbicide-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 62, 2608–2612. doi: 10.1099/ijs.0.037796-0

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

García-Valdés, E., and Lalucat, J. (2016). "*Pseudomonas*: molecular phylogeny and current taxonomy" in *Pseudomonas: Molecular and applied biology*. ed. R. S. Kahlon (New York: Springer), 1–23.

Gibello, A., Vela, A. I., Martín, M., Mengs, G., Alonso, P. Z., Garbi, C., et al. (2011). *Pseudomonas composti* sp. nov., isolated from compost samples. *Int. J. Syst. Evol. Microbiol.* 61, 2962–2966. doi: 10.1099/ijs.0.027086-0

Gilroy, R., Ravi, A., Getino, M., Pursley, I., Horton, D. L., Alikhan, N. F., et al. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ* 9:e10941. doi: 10.7717/peerj.10941

Girard, L., Lood, C., Höfte, M., Vandamme, P., Rokni-Zadeh, H., Van Noort, V., et al. (2021). The ever-expanding *Pseudomonas* genus: description of 43 new species and partition of the *Pseudomonas putida* group. *Microorganisms* 9:1766. doi: 10.3390/microorganisms9081766

Gomila, M., Peña, A., Mulet, M., Lalucat, J., and García-Valdés, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Front. Microbiol.* 6:214. doi: 10.3389/fmicb.2015.00214

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491. doi: 10.1128/MMBR.62.4.1435-1491.1998

Gupta, R. S. (2014). Identification of conserved indels that are useful for classification and evolutionary studies. *Methods Microbiol.* 41, 153–182. doi: 10.1016/bs.mim.2014.05.003

Gupta, R. S. (2016). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol. Rev.* 40, 520–553. doi: 10.1093/femsre/fuw011

Gupta, R. S. (2019). Distinction between *Borrelia* and *Borreliella* is more robustly supported by molecular and phenotypic characteristics than all other neighbouring prokaryotic genera: response to Margos' et al. "the genus *Borrelia* reloaded" (PLoS One 13: e0208432). *PLoS One* 14:e0221397. doi: 10.1371/journal.pone.0221397

Gupta, R. S., Chander, P., and George, S. (2013). Phylogenetic framework and molecular signatures for the class *Chloroflexi* and its different clades; proposal for division of the class *Chloroflexia* class. nov. [corrected] into the suborder *Chloroflexineae* subord. nov., consisting of the emended family *Oscillochloridaceae* and the family *Chloroflexaceae* fam. nov., and the suborder *Roseiflexineae* subord. nov., containing the family *Roseiflexaceae* fam. nov. *Antonie Van Leeuwenhoek* 103, 99–119. doi: 10.1007/s10482-012-9790-3

Gupta, R. S., and Kanter-Eivin, D. A. (2023). AppIndels.com server: a web based tool for the identification of known taxon-specific conserved signature indels in genome sequences: validation of its usefulness by predicting the taxonomic affiliation of >700 unclassified strains of *Bacillus* species. *Int. J. Syst. Evol. Microbiol.* 73:005844. doi: 10.1099/ijsem.0.005844

Gupta, R. S., and Patel, S. (2019). Robust demarcation of the family *Caryophanaceae* (*Planococcaceae*) and its different genera including three novel genera based on phylogenomics and highly specific molecular signatures. *Front. Microbiol.* 10:2821. doi: 10.3389/fmicb.2019.02821

Gupta, R. S., Patel, S., Saini, N., and Chen, S. (2020). Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the subtilis and Cereus clades of species. *Int. J. Syst. Evol. Microbiol.* 70, 5753–5798. doi: 10.1099/ijsem.0.004475

Hauser, E., Kämpfer, P., and Busse, H. J. (2004). *Pseudomonas psychrotolerans* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54, 1633–1637. doi: 10.1099/ijs.0.03024-0

He, W.-H., Wang, Y.-N., Du, X., Zhou, Y., Jia, B., Bian, J., et al. (2012). *Pseudomonas linyingensis* sp. nov.: a novel bacterium isolated from wheat soil subjected to long-term herbicides application. *Curr. Microbiol.* 65, 595–600. doi: 10.1007/s00284-012-0187-3

Hespell, R. B. (1977). *Serpens flexibilis* gen. nov., sp. nov., an unusually flexible, lactate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* 27, 371–381. doi: 10.1099/00207713-27-4-371

Hesse, C., Schulz, F., Bull, C. T., Shaffer, B. T., Yan, Q., Shapiro, N., et al. (2018). Genome-based evolutionary history of *Pseudomonas* spp. *Environ. Microbiol.* 20, 2142–2159. doi: 10.1111/1462-2920.14130

Hildebrand, D., Palleroni, N., Hendson, M., Toth, J., and Johnson, J. (1994). *Pseudomonas flavescens* sp. nov., isolated from walnut blight cankers. *Int. J. Syst. Evol. Microbiol.* 44, 410–415. doi: 10.1099/00207713-44-3-410

Hirota, K., Yamahira, K., Nakajima, K., Nodasaka, Y., Okuyama, H., and Yumoto, I. (2011). *Pseudomonas toyotomiensis* sp. nov., a psychrotolerant facultative alkaliphile that utilizes hydrocarbons. *Int. J. Syst. Evol. Microbiol.* 61, 1842–1848. doi: 10.1099/ijs.0.024612-0

Holmes, B., Steigerwalt, A., Weaver, R., and Brenner, D. J. (1986). *Chryseomonas polytricha* gen. nov., sp. nov., a *Pseudomonas*-like organism from human clinical specimens and formerly known as group Ve-1. *Int. J. Syst. Evol. Microbiol.* 36, 161–165. doi: 10.1099/00207713-36-2-161

Holmes, B., Steigerwalt, A., Weaver, R., and Brenner, D. J. (1987). *Chryseomonas luteola* comb. nov. and *Flavimonas oryzihabitans* gen. nov., comb. nov., *Pseudomonas*-like species from human clinical specimens and formerly known, respectively, as groups Ve-1 and Ve-2. *Int. J. Syst. Evol. Microbiol.* 37, 245–250. doi: 10.1099/00207713-37-3-245

Hunter, W. J., and Manter, D. K. (2011). *Pseudomonas seleniipraecipitatus* sp. nov.: a selenite reducing gamma-proteobacteria isolated from soil. *Curr. Microbiol.* 62, 565–569. doi: 10.1007/s00284-010-9745-8

Huq, M. A., Lee, S. Y., Ma, J., Rahman, M. M., Rahman, M. S., Park, J. H., et al. (2022). *Pseudomonas oryzagri* sp. nov., isolated from a rice field soil. *Int. J. Syst. Evol. Microbiol.* 72:005534. doi: 10.1099/ijsem.0.005534

Iizuka, H., and Komagata, K. (1963). New species of *Pseudomonas* belonging to fluorescent group. (studies on the microorganisms of cereal grains. Part V). *Nippon Nogeikagaku Kaishi* 37, 137–141. doi: 10.1271/nogeikagaku1924.37.137

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403–405. doi: 10.1016/S0968-0004(98)01285-7

Jun, S. R., Wassenaar, T. M., Nookaew, I., Hauser, L., Wanchai, V., Land, M., et al. (2016). Diversity of *Pseudomonas* genomes, including populus-associated isolates, as revealed by comparative genome analysis. *Appl. Environ. Microbiol.* 82, 375–383. doi: 10.1128/AEM.02612-15

Kämpfer, P., Glaeser, S., Mcinroy, J. A., Clermont, D., Criscuolo, A., and Busse, H.-J. (2021). *Pseudomonas carbonaria* sp. nov., isolated from charcoal. *Syst. Appl. Microbiol.* 71:004750. doi: 10.1099/ijsem.0.004750

Kennedy, C. R. P., Macdonald, M. L., and Melton, T. (2015). "Azotobacter" in *Bergey's Manual of Systematics of Archaea and Bacteria* (Hoboken, NJ: John Wiley and Sons)

Kennedy, C., and Rudnick, P. (2015). "Azomonas" in *Bergey's manual of systematics of Archaea and Bacteria* (Hoboken, NJ: John Wiley and Sons)

Khadka, B., Persaud, D., and Gupta, R. S. (2020). Novel sequence feature of SecA translocase protein unique to the thermophilic bacteria: bioinformatics analyses to investigate their potential roles. *Microorganisms* 8:59. doi: 10.3390/microorganisms8010059

Kim, C. M., Jeong, J. W., Lee, D. H., and Kim, S. B. (2021). *Pseudomonas guryensis* sp. nov. and *Pseudomonas ullengensis* sp. nov., isolated from soil. *Int. J. Syst. Evol. Microbiol.* 71:005082. doi: 10.1099/ijsem.0.005082

Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Kimura, N., Watanabe, T., Suenaga, H., Fujihara, H., Futagami, T., Goto, M., et al. (2018). *Pseudomonas furukawaii* sp. nov., a polychlorinated biphenyl-degrading bacterium isolated from biphenyl-contaminated soil in Japan. *Int. J. Syst. Evol. Microbiol.* 68, 1429–1435. doi: 10.1099/ijsem.0.002670

Kodama, K. K., Kimura, N., and Komagata, K. (1985). Two new species of *Pseudomonas*: *P. oryzihabitans* isolated from rice paddy and clinical specimens and *P. luteola* isolated from clinical specimens. *Int. J. Syst. Evol. Microbiol.* 35, 467–474. doi: 10.1099/00207713-35-4-467

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006

Kujur, R. R. A., and Das, S. K. (2022). *Pseudomonas phenolilytica* sp. nov., a novel phenol-degrading bacterium. *Arch. Microbiol.* 204:320. doi: 10.1007/s00203-022-02912-y

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lalucat, J., Gomila, M., Mulet, M., Zaruma, A., and García-Valdés, E. (2022). Past, present and future of the boundaries of the *Pseudomonas* genus: proposal of *Stutzerimonas* gen. nov. *Syst. Appl. Microbiol.* 45:126289. doi: 10.1016/j.syapm.2021.126289

Lalucat, J., Mulet, M., Gomila, M., and García-Valdés, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes (Basel)* 11:139. doi: 10.3390/genes11020139

Lee, M., and Chandler, A. C. (1941). A study of the nature, growth and control of bacteria in cutting compounds. *J. Bacteriol.* 41, 373–386. doi: 10.1128/jb.41.3.373-386.1941

Lee, D. H., Kim, C. M., and Kim, S. B. (2022). *Pseudomonas insulae* sp. nov., isolated from island soil. *Int. J. Syst. Evol. Microbiol.* 72:005363. doi: 10.1099/ijsem.0.005363

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, K., Ye, Z., Chen, G., Zheng, K., Yin, J., Debnath, S. C., et al. (2023). *Atopomonas sediminilitoris* sp. nov., isolated from beach sediment of Zhairuo Island, China. *Antonie Van Leeuwenhoek* 116, 97–107. doi: 10.1007/s10482-022-01780-2

Lin, S.-Y., Hameed, A., Liu, Y.-C., Hsu, Y.-H., Lai, W.-A., Chen, W.-M., et al. (2013). *Pseudomonas sagittaria* sp. nov., a siderophore-producing bacterium isolated from oil-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 63, 2410–2417. doi: 10.1099/ijs.0.045567-0

Liu, Y.-C., Young, L.-S., Lin, S.-Y., Hameed, A., Hsu, Y.-H., Lai, W.-A., et al. (2013). *Pseudomonas guguanensis* sp. nov., a gammaproteobacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 63, 4591–4598. doi: 10.1099/ijs.0.047712-0

Lund-Palau, H., Turnbull, A. R., Bush, A., Bardin, E., Cameron, L., Soren, O., et al. (2016). *Pseudomonas aeruginosa* infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches. *Expert Rev. Respir. Med.* 10, 685–697. doi: 10.1080/17476348.2016.1177460

Mamtimin, T., Anwar, N., Abdurahman, M., Kurban, M., Rozahon, M., Mamtimin, H., et al. (2021). *Pseudomonas lopnurensis* sp. nov., an endophytic bacterium isolated from *Populus euphratica* at the ancient Ugan river. *Antonie Van Leeuwenhoek* 114, 399–410. doi: 10.1007/s10482-021-01524-8

Manaia, C. M., and Moore, E. R. (2002). *Pseudomonas thermotolerans* sp. nov., a thermotolerant species of the genus *Pseudomonas sensu stricto*. *Int. J. Syst. Evol. Microbiol.* 52, 2203–2209. doi: 10.1099/00207713-52-6-2203

Manickam, N., Ghosh, A., Jain, R. K., and Mayilraj, S. (2008). Description of a novel indole-oxidizing bacterium *Pseudomonas indoloxydans* sp. nov., isolated from a pesticide-contaminated site. *Syst. Appl. Microbiol.* 31, 101–107. doi: 10.1016/j.syapm.2008.02.002

Migula, W. (1894). Uber ein neues System der Bakterien. *Arb. Bakt. Inst.* 1:238.

Monias, B. L. (1928). Classification of bacterium alcaligenes, pyocyaneum, and fluorescens. *J. Infect. Dis.* 43, 330–334. doi: 10.1093/infdis/43.4.330

Montecillo, J. A. V., and Bae, H. (2022). Reclassification of *Brevibacterium frigoritolerans* as *Peribacillus frigoritolerans* comb. nov. based on phylogenomics and multiple molecular synapomorphies. *Int. J. Syst. Evol. Microbiol.* 72:005389. doi: 10.1099/ijsem.0.005389

Nicklasson, M., Martín-Rodríguez, A. J., Thorell, K., Higdon, S. M., Neves, L., Mussagy, A., et al. (2022). *Pseudomonas boanensis* sp. nov., a bacterium isolated from river water used for household purposes in Boane District, Mozambique. *Int. J. Syst. Evol. Microbiol.* 72:005461. doi: 10.1099/ijsem.0.005461

Oren, A., Arahal, D. R., Göker, M., Moore, E. R. B., Rossello-Mora, R., and Sutcliffe, I. C. (2023). International code of nomenclature of prokaryotes. Prokaryotic code (2022 revision). *Int. J. Syst. Evol. Microbiol.* 73:005585. doi: 10.1099/ijsem.0.005782

Özen, A. I., and Ussery, D. W. (2012). Defining the *Pseudomonas* genus: where do we draw the line with *Azotobacter*? *Microb. Ecol.* 63, 239–248. doi: 10.1007/s00248-011-9914-8

Palleroni, N. J. (2005). "Genus I. *Pseudomonas* Migula 1894" in *Bergey's manual of systematic bacteriology (the Proteobacteria), part B (the Gammaproteobacteria)*. eds. D. J. Brenner, N. R. Krieg, J. T. Staley and G. M. Garrity, vol. *2. 2nd* ed (New York: Springer), 323–379.

Palleroni, N. J. (2010). The *Pseudomonas* story. *Environ. Microbiol.* 12, 1377–1383. doi: 10.1111/j.1462-2920.2009.02041.x

Palleroni, N. J. (2015). "Pseudomonas" in *Bergey's Manual of Systematics of Archaea and Bacteria* (Hoboken, NJ: John Wiley and Sons)

Palleroni, N. J., Doudoroff, M., Stanier, R. Y., Solanes, R., and Mandel, M. (1970). Taxonomy of the aerobic pseudomonads: the properties of the *Pseudomonas stutzeri* group. *Microbiology* 60, 215–231. doi: 10.1099/00221287-60-2-215

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Passarelli-Araujo, H., Franco, G. R., and Venancio, T. M. (2022). Network analysis of ten thousand genomes shed light on *Pseudomonas* diversity and classification. *Microbiol. Res.* 254:126919. doi: 10.1016/j.micres.2021.126919

Patel, S., and Gupta, R. S. (2020). A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* 70, 406–438. doi: 10.1099/ijsem.0.003775

Peix, A., Berge, O., Rivas, R., Abril, A., and Velázquez, E. (2005). *Pseudomonas argentinensis* sp. nov., a novel yellow pigment-producing bacterial species, isolated from rhizospheric soil in Cordoba, Argentina. *Syst. Appl. Microbiol.* 55, 1107–1112. doi: 10.1099/ijs.0.63445-0

Peix, A., Ramírez-Bahena, M.-H., and Velázquez, E. (2009). Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect. Genet. Evol.* 9, 1132–1147. doi: 10.1016/j.meegid.2009.08.001

Peix, A., Ramirez-Bahena, M. H., and Velázquez, E. (2018). The current status on the taxonomy of *Pseudomonas* revisited: an update. *Infect. Genet. Evol.* 57, 106–116. doi: 10.1016/j.meegid.2017.10.026

Planquette, B., Timsit, J.-F., Misset, B. Y., Schwebel, C., Azoulay, E., Adrie, C., et al. (2013). *Pseudomonas aeruginosa* ventilator-associated pneumonia. Predictive factors of treatment failure. *Am. J. Respir. Crit. Care Med.* 188, 69–76. doi: 10.1164/rccm.201210-1897OC

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Qin, Q. L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215. doi: 10.1128/JB.01688-14

Ramos, E., Ramírez-Bahena, M.-H., Valverde, A., Velázquez, E., Zúñiga, D., Velezmoro, C., et al. (2013). *Pseudomonas punonensis* sp. nov., isolated from straw. *Int. J. Syst. Evol. Microbiol.* 63, 1834–1839. doi: 10.1099/ijs.0.042119-0

Rokas, A., and Holland, P. W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459. doi: 10.1016/S0169-5347(00)01967-4

Rossi, E., La Rosa, R., Bartell, J. A., Marvig, R. L., Haagensen, J. A. J., Sommer, L. M., et al. (2021). *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat. Rev. Microbiol.* 19, 331–342. doi: 10.1038/s41579-020-00477-5

Rudra, B., Duncan, L., Shah, A. J., Shah, H. N., and Gupta, R. S. (2022). Phylogenomic and comparative genomic studies robustly demarcate two distinct clades of *Pseudomonas aeruginosa* strains: proposal to transfer the strains from an outlier clade to a novel species *Pseudomonas paraeruginosa* sp. nov. *Int. J. Syst. Evol. Microbiol.* 72:005542. doi: 10.1099/ijsem.0.005542

Rudra, B., and Gupta, R. S. (2021). Phylogenomic and comparative genomic analyses of species of the family *Pseudomonadaceae*: proposals for the genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., merger of the genus *Oblitimonas* with the genus *Thiopseudomonas*, and transfer of some misclassified species of the genus *Pseudomonas* into other genera. *Int. J. Syst. Evol. Microbiol.* 71:005011. doi: 10.1099/ijsem.0.005011

Saati-Santamaría, Z., Peral-Aranega, E., Velázquez, E., Rivas, R., and García-Fraile, P. (2021). Phylogenomic analyses of the genus *Pseudomonas* lead to the rearrangement of several species and the definition of new genera. *Biology (Basel)* 10:782. doi: 10.3390/biology10080782

Saini, N., and Gupta, R. S. (2021). A robust phylogenetic framework for members of the order *Legionellales* and its main genera (*Legionella*, *Aquicella*, *Coxiella* and

Rickettsiella) based on phylogenomic analyses and identification of molecular markers demarcating different clades. *Antonie Van Leeuwenhoek* 114, 957–982. doi: 10.1007/s10482-021-01569-9

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., et al. (2019). Database resources of the National Center for biotechnology information. *Nucleic Acids Res.* 47, D23–D28. doi: 10.1093/nar/gky1069

Shin, S. K., Hwang, C. Y., Cho, Y. J., Yi, H., and Yi, H. (2015). Reclassification of *Serpens flexibilis* Hespell 1977 as *Pseudomonas flexibilis* comb. nov., with *Pseudomonas tuomuerensis* Xin et al. 2009 as a later heterotypic synonym. *Syst. Appl. Microbiol.* 38, 563–566. doi: 10.1016/j.syapm.2015.09.007

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Silby, M. W., Winstanley, C., Godfrey, S. A., Levy, S. B., and Jackson, R. W. (2011). *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol. Rev.* 35, 652–680. doi: 10.1111/j.1574-6976.2011.00269.x

Singh, B., and Gupta, R. S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Gen. Genomics.* 281, 361–373. doi: 10.1007/s00438-008-0417-3

Skerman, V. B. D., Mcgowan, V., and Sneath, P. H. A. (1980). Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* 30, 225–420. doi: 10.1099/00207713-30-1-225

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

Stanier, R. Y., Palleroni, N. J., and Doudoroff, M. (1966). The aerobic pseudomonads a taxonomic study. *Microbiology* 43, 159–271. doi: 10.1099/00221287-43-2-159

Stover, C. K., Pham, X. Q., Erwin, A., Mizoguchi, S., Warrener, P., Hickey, M., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959–964. doi: 10.1038/35023079

Sudan, S. K., Pal, D., Bisht, B., Kumar, N., Chaudhry, V., Patil, P., et al. (2018). *Pseudomonas fluvialis* sp. nov., a novel member of the genus *Pseudomonas* isolated from the river Ganges, India. *Int. J. Syst. Evol. Microbiol.* 68, 402–408. doi: 10.1099/ijsem.0.002520

Tao, Y., Zhou, Y., He, X., Hu, X., and Li, D. (2014). *Pseudomonas chengduensis* sp. nov., isolated from landfill leachate. *Int. J. Syst. Evol. Microbiol.* 64, 95–100. doi: 10.1099/ijs.0.050294-0

Tarhriz, V., Nouioui, I., Spröer, C., Verbarg, S., Ebrahimi, V., Cortés-Albayay, C., et al. (2020). *Pseudomonas khazarica* sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium isolated from Khazar Sea sediments. *Antonie Van Leeuwenhoek* 113, 521–532. doi: 10.1007/s10482-019-01361-w

Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., and Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC Genomics* 14:913. doi: 10.1186/1471-2164-14-913

Thorat, V., Kirdat, K., Tiwarekar, B., Dacosta, E., Debbarma, P., Shouche, Y., et al. (2020). *Pseudomonas lalkuanensis* sp. nov., isolated from a bacterial consortia of contaminated soil enriched for the remediation of e-waste. *Int. J. Syst. Evol. Microbiol.* 70, 6468–6475. doi: 10.1099/ijsem.0.004559

Timsy, T., Ulrich, A., Kublik, S., Foesel, B. U., Kolb, S., Horn, M. A., et al. (2021). *Pseudomonas campi* sp. nov., a nitrate-reducing bacterium isolated from grassland soil. *Int. J. Syst. Evol. Microbiol.* 71:4799. doi: 10.1099/ijsem.0.004799

Tohya, M., Watanabe, S., Teramoto, K., Tada, T., Kuwahara-Arai, K., Mya, S., et al. (2020). *Pseudomonas yangonensis* sp. nov., isolated from wound samples of patients in a hospital in Myanmar. *Int. J. Syst. Evol. Microbiol.* 70, 3597–3605. doi: 10.1099/ijsem.0.004181

Wang, X., He, S.-W., Guo, H.-B., Thin, K. K., Gao, J.-S., Wang, Y., et al. (2020). *Pseudomonas rhizoryzae* sp. nov., isolated from rice. *Int. J. Syst. Evol. Microbiol.* 70, 944–950. doi: 10.1099/ijsem.0.003852

Wang, Z., and Wu, M. (2013). A phylum-level bacterial phylogenetic marker database. *Mol. Biol. Evol.* 30, 1258–1262. doi: 10.1093/molbev/mst059

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851

Winogradsky, S. (1938). Sur la morphologie et l' oecologie des *Azotobacter*. *Ann. Inst. Pasteur* 55, 351–400.

Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., and Brinkman, F. S. (2016). Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* 44, D646–D653. doi: 10.1093/nar/gkv1227

Wong, S. Y., Paschos, A., Gupta, R. S., and Schellhorn, H. E. (2014). Insertion/deletion-based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments. *Environ. Sci. Technol.* 48, 11462–11470. doi: 10.1021/es502794h

Wu, M., Wen, J., Chang, M., Yang, G., and Zhou, S. (2014). *Pseudomonas sihuiensis* sp. nov., isolated from a forest soil in South China. *Antonie Van Leeuwenhoek* 105, 781–790. doi: 10.1007/s10482-014-0134-3

Xin, X. F., Kvitko, B., and He, S. Y. (2018). *Pseudomonas syringae*: what it takes to be a pathogen. *Nat. Rev. Microbiol.* 16, 316–328. doi: 10.1038/nrmicro.2018.17

Xin, Y. H., Zhang, D. C., Liu, H. C., Zhou, H. L., and Zhou, Y. G. (2009). *Pseudomonas tuomuerensis* sp. nov., isolated from a bird's nest. *Int. J. Syst. Evol. Microbiol.* 59, 139–143. doi: 10.1099/ijs.0.000547-0

Yamada, K., Sasaki, M., Aoki, K., Nagasawa, T., Murakami, H., Ishii, M., et al. (2021). *Pseudomonas tohonis* sp. nov., isolated from the skin of a patient with burn wounds in Japan. *Int. J. Syst. Evol. Microbiol.* 71:005115. doi: 10.1099/ijsem.0.005115

Yang, Y., Gao, Y., Liu, Y., Liu, B., Wang, D., Xu, Y., et al. (2022). *Pseudomonas marianensis* sp. nov., a marine bacterium isolated from deep-sea sediments of the Mariana trench. *Arch. Microbiol.* 204:638. doi: 10.1007/s00203-022-03250-9

Yang, G., Han, L., Wen, J., and Zhou, S. (2013). *Pseudomonas guangdongensis* sp. nov., isolated from an electroactive biofilm, and emended description of the genus *Pseudomonas* Migula 1894. *Int. J. Syst. Evol. Microbiol.* 63, 4599–4605. doi: 10.1099/ijs.0.054676-0

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330

Young, J. M., and Park, D. C. (2007). Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. *Int. J. Syst. Evol. Microbiol.* 57, 2894–2901. doi: 10.1099/ijs.0.64969-0

Yu, X.-Y., Zhai, J.-Y., Wu, C., Zhang, C.-Y., Shi, J.-Y., Ding, L.-X., et al. (2018). *Pseudomonas pharmafabricae* sp. nov., isolated from pharmaceutical wastewater. *Curr. Microbiol.* 75, 1119–1125. doi: 10.1007/s00284-018-1495-z

Yumoto, I., Yamazaki, K., Hishinuma, M., Nodasaka, Y., Suemori, A., Nakajima, K., et al. (2001). *Pseudomonas alcaliphila* sp. nov., a novel facultatively psychrophilic alkaliphile isolated from seawater. *Int. J. Syst. Evol. Microbiol.* 51, 349–355. doi: 10.1099/00207713-51-2-349

Zhang, P., Dong, X., Zhou, K., Zhu, T., Liang, J., Shi, W., et al. (2021). Characterization of a novel chromosome-encoded AmpC β-lactamase gene, Bla PRC–1, in an isolate of a newly classified *Pseudomonas* species, *Pseudomonas wenzhouensis* A20, from animal farm sewage. *Front. Microbiol.* 12:732932. doi: 10.3389/fmicb.2021.732932

Zhang, M., Li, A., Yao, Q., Xiao, B., and Zhu, H. (2022). *Pseudomonas oligotrophica* sp. nov., a novel denitrifying bacterium possessing nitrogen removal capability under low carbon–nitrogen ratio condition. *Front. Microbiol.* 13:882890. doi: 10.3389/fmicb.2022.882890

Zhang, L., Pan, Y., Wang, K., Zhang, X., Zhang, S., Fu, X., et al. (2015). *Pseudomonas songnenensis* sp. nov., isolated from saline and alkaline soils in Songnen plain, China. *Antonie Van Leeuwenhoek* 107, 711–721. doi: 10.1007/s10482-014-0365-3

Zhou, S., Wang, Y., Xia, H., Liu, D., Chen, S., and Li, F. (2020). *Pseudomonas hydrolytica* sp. nov., multiple polymer-degrading bacteria isolated from soil in China. *Int. J. Syst. Evol. Microbiol.* 70, 3049–3054. doi: 10.1099/ijsem.0.004129

Zhu, H.-Z., Zhang, Z.-F., Zhou, N., Jiang, C.-Y., Wang, B.-J., Cai, L., et al. (2021). Bacteria and metabolic potential in karst caves revealed by intensive bacterial cultivation and genome assembly. *Appl. Environ. Microbiol.* 87, e02440–e02420. doi: 10.1128/AEM.02440-20

CHAPTER 4

Phylogenomic and Comparative Genomic Studies Robustly Demarcate Two Distinct Clades of *Pseudomonas aeruginosa* Strains: Proposal to Transfer the Strains from an Outlier Clade to a Novel Species *Pseudomonas paraeruginosa* sp. nov.

This chapter presents a comprehensive phylogenomic analysis of different *Pseudomonas aeruginosa* strains, revealing the existence of two clades (Classical and Outlier). The distinctness of these clades is supported by identifying molecular markers (CSIs and CSPs) that are highly specific for the strain of these two clades. Using CSIs and CSPs, phylogenetic analysis, species demarcation criteria (16S similarity, ANI, AAI, and dDDH), and additional phenotypic tests, this study concludes with a proposal to reclassify the strains from the Outlier clade as a new species, *Pseudomonas paraeruginosa*. My contributions to this chapter include identifying CSIs, conducting genomic similarity analyses, drafting the manuscript, and finalizing all main and supplemental figures and tables. This was a collaborative project with King's College London, UK, where the co-first author, Louise Duncan, conducted all biochemical assays, including growth, morphology, motility, carbon utilization, enzyme production, chemical tolerance, and antibiotic resistance, on the type strains or representative strains from both clades/species.

Due to space limitations, supplementary materials for this study are not included in this chapter but are available along with the entire manuscript at:

Rudra, B., et. al., (2022). Int. J. Syst. Evol. Microbiol., 72(11), 005542.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION
Rudra *et al.*, *Int. J. Syst. Evol. Microbiol.* 2022;72:005542
DOI 10.1099/ijsem.0.005542

MICROBIOLOGY SOCIETY
OPEN ACCESS

Phylogenomic and comparative genomic studies robustly demarcate two distinct clades of *Pseudomonas aeruginosa* strains: proposal to transfer the strains from an outlier clade to a novel species *Pseudomonas paraeruginosa* sp. nov.

Bashudev Rudra[1]†, Louise Duncan[2]†, Ajit J. Shah[3], Haroun N. Shah[3] and Radhey S. Gupta[1,*]

Abstract

The strains of *Pseudomonas aeruginosa* exhibit considerable differences in their genotypic and pathogenic properties. To clarify their evolutionary/taxonomic relationships, comprehensive phylogenomic and comparative genomic studies were conducted on the genome sequences of 212 *P. aeruginosa* strains covering their genetic diversity. In a phylogenomic tree based on 118 conserved proteins, the analysed strains formed two distinct clades. One of these clades, Clade-1, encompassing >70% of the strains including the type strain DSM 50071^T, represents the species *P. aeruginosa sensu stricto*. Clade-2, referred to in earlier work as the outlier group, with NCTC 13628^T as its type strain, constitutes a novel species level lineage. The average nucleotide identity, average amino acid identity and digital DNA–DNA hybridization values between the strains from Clade-1 and Clade-2 are in the range of 93.4–93.7, 95.1–95.3 and 52–53%, respectively. The 16S rRNA gene of *P. aeruginosa* DSM 50071^T also shows 98.3% similarity to that of NCTC 13628^T. These values are lower than the suggested cut-off values for species distinction, indicating that the Clade-2 strains (NCTC 13628^T) constitute a new species. We also report the identification of 12 conserved signature indels in different proteins and 24 conserved signature proteins that are exclusively found in either Clade-1 or Clade-2, providing a reliable means for distinguishing these clades. Additionally, in contrast to swimming motility, twitching motility is only present in Clade-1 strains. Based on earlier work, the strains from these two clades also differ in their pathogenic mechanisms (presence/absence of Type III secretion system), production of biosurfactants, phenazines and siderophores, and several other genomic characteristics. Based on the evidence from different studies, we propose that the Clade-2 strains constitute a novel species for which the name *Pseudomonas paraeruginosa* is proposed. The type strain is NCTC 13628^T (=PA7^T=ATCC 9027^T). The description of *Pseudomonas aeruginosa* is also emended to include information for different molecular markers specific for this species.

DATA SUMMARY

All supporting data have been provided within the article or through supplementary data files, Figshare - 10.6084/m9.figshare.20277480 [1]. The GenBank accession number for the 16S rRNA gene sequence of strain NCTC 13628^T is ON359917 and the accession number for its genome sequence is GCA_900706985 [1].

Author affiliations: [1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton L8N 3Z5, Canada; [2]School of Cancer and Pharmaceutical Sciences, King's College London, London SE1 9NH, UK; [3]Department of Natural Sciences, Middlesex University, London NW4 4BT, UK.
***Correspondence:** Radhey S. Gupta, gupta@mcmaster.ca
Keywords: classical clade; conserved signature indels (CSIs); conserved signature proteins (CSPs); outlier clade; *Pseudomonas aeruginosa*; *Pseudomonas paraeruginosa*; systematics; T3SS system.
Abbreviations: AAI, average amino acid identity; ANI, average nucleotide identity; CSI, conserved signature indel; CSP, conserved signature protein; dDDH, digital DNA–DNA hybridization; LB, Luria–Bertani; MFS, major facilitator superfamily; ML, maximum-likelihood.
The GenBank accession number for the 16S rRNA gene sequence of strain NCTC 13628^T is ON359917 and the accession number for its genome sequence is GCA_900706985. The accession numbers for the 16S rRNA gene sequence and the genome sequence for the type strain DSM 50071^T of *Pseudomonas aeruginosa* are NR_026078 and GCA_001045685, respectively.
†These authors contributed equally to this work
One supplementary table and 15 supplementary figures are available with the online version of this article.

INTRODUCTION

Pseudomonas aeruginosa is a Gram-negative bacterium inhabiting a wide range of niches from plants to animals [2, 3]. This bacterium, originally described in 1872 as 'Bacterium aeruginosum' by Schroter *et al.* [4], constitutes one of the earliest known micro-organisms, which in 1894 was reclassified by Migula [5, 6] as *Pseudomonas aeruginosa. P. aeruginosa* is an opportunistic human pathogen capable of causing a wide array of life-threatening acute and chronic diseases including nosocomial infections (ventilator-associated pneumonia), cystic fibrosis and various sepsis syndromes [7, 8]. This bacterium is intrinsically resistant to many drugs, which makes it difficult to treat with available antibiotics [9].

The strains of *P. aeruginosa* with variable genotypic and phenotypic properties are increasing rapidly worldwide [10, 11] and whole genome sequences are now available for large numbers of these strains in public databases [12, 13]. In the NCBI Genome sequence database [13], as of 20 March 2021, genome sequences of 6593 *P. aeruginosa* strains were available. Several studies based on genomic and biochemical properties of *P. aeruginosa* strains indicate that these strains form multiple clades in phylogenetic analysis [14–22]. Among these clades, one major clade which includes the type strain of *P. aeruginosa* DSM 50071 (represented in several studies by the strain PAO1) is commonly referred to as the 'classical' clade [20, 21, 23, 24]. Besides this main clade, another clade of *P. aeruginosa* observed in most studies includes strains NCTC 13628, ATCC 9027, PA7 and CR1, and it is often referred to as the 'outlier' clade or group [20, 21, 25, 26]. Although the virulence pattern and pathogenicity of *P. aeruginosa* is multifactorial [14, 27, 28], several studies indicate that the strains from these two clades differ in terms of how they exert their cytolytic activity on human cells [14, 28].

The degree of pathogenicity of the majority of strains from the classical clade depends on the presence of type III secretion system (T3SS), which enables the injection of four main effector proteins (ExoS, ExoT, ExoY and ExoU) and several virulence factors (*viz.* proteases, exotoxin A, pili, flagella, quorum-sensing proteins) directly into the host cell cytoplasm [29–33]. In contrast, the strains from the outlier clade not only lack the entire T3SS-encoding locus but also genes encoding type III secreted exoenzymes [24, 28]. In addition, Basso *et al.* [16] have identified a novel virulence mechanism using exolysin toxins A and B (*viz.* ExlA and ExlB), which is present in strains from the outlier clade that lack the T3SS. Although differences in the pathogenic properties and some genomic characteristics between the studied strains from these two clades have been indicated in several studies [18–21], due to difficulties in reliably distinguishing between the members of these two clades, currently there is no distinction made between different strains of *P. aeruginosa*. Hence, it is of great interest to reliably establish the existence of two distinct lineages of *P. aeruginosa* strains and develop robust means for clearly demarcating and distinguishing members of these two lineages from each other.

We report here detailed phylogenomic and comparative genomic analyses on *P. aeruginosa* strains to clearly elucidate their evolutionary relationship and our work on identifying novel molecular markers that can reliably demarcate/distinguish the strains from its two main clades. Using genome sequences for 212 strains of *P. aeruginosa* covering their genetic diversity, we have reconstructed a robust phylogenetic tree using concatenated genome sequences of 118 conserved proteins. In this tree, the analysed *P. aeruginosa* strains formed two distinct clades referred to as Clade-1 and Clade-2. Clade-1 harbours the type strain DSM 50071^{T}, and the well-studied strain PAO1, and it corresponds to the classical clade, which represents *P. aeruginosa sensu stricto*. Clade-2, which comprises the outlier clade, harbours the well-studied strains ATCC 9027, PA7 and NCTC 13628 [20, 21, 25, 34]. Comparisons of the average nucleotide identity (ANI), average amino acid identity (AAI) and digital DNA–DNA hybridization (dDDH) values, as well as the 16S rRNA gene sequence similarity, for the reference strains from Clade-1 and Clade-2 show that these values for the two clades are lower than the suggested cut-off values for species boundaries. Some differences are also observed in the morphological characteristics of the reference strains from the two clades. Lastly, and most importantly, our comparative analyses of protein sequences from the genomes of these strains have identified multiple highly specific molecular signatures consisting of conserved signature indels (CSIs) in proteins and conserved signature proteins (CSPs), which are exclusive characteristics of the members of either Clade-1 or Clade-2. These molecular markers provide a novel and reliable means for the demarcation of the strains from these two clades and distinguishing them from each other and other bacteria. Based on the compelling evidence presented here, we propose that the *P. aeruginosa* strains from Clade-2 should be recognized as a novel species for which the name *Pseudomonas paraeruginosa* sp. nov. is proposed.

METHODS

Phylogenomic analysis of the *P. aeruginosa* strains

Genome sequences were available for >6500 *P. aeruginosa* strains in the NCBI database (www.ncbi.nlm.nih.gov/genome/; accessed on 20 March 2021) [13]. Based on earlier studies on these genomes [15, 20, 22, 35], we downloaded genome sequences for 212 *P. aeruginosa* strains covering their phylogenetic/genetic diversity. The downloaded genomes included all 39 strains that the Genome Taxonomy Database (GTDB) [35] server places into a separate group called 'Pseudomonas aeruginosa_A', as well as >100 strains from the main *P. aeruginosa* clade, containing all other sequenced strains [35]. The 'Pseudomonas aeruginosa_A' strains encompass all strains referred to as the outlier strains by Sood *et al.* [20]. In addition, Ozer *et al.* [22] based on their analysis of 739 *P. aeruginosa* strains from diverse sources have grouped these strains into three lineages A, B and C. The downloaded genomes

include several genomes representatives of each of these three lineages [22]. Besides these, genomes of several *Pseudomonas* species (*P. citronellolis*, *P. delhiensis*, *P. knackmussii*, *P. humi*, *P. jinjuensis*, *P. multiresinivorans*, *P. nitroreducens* and *P. panipatensis*) belonging to the 'Aeruginosa clade' [36, 37] were included to serve as an outgroup. Initial phylogenetic analysis indicated that about two-thirds of the downloaded/analysed strains were identical and hence they were removed from the dataset and further analysis was carried out on 57 strains covering the genetic diversity of available strains. Based on these genome sequences, a rooted phylogenetic tree was reconstructed based on concatenated sequences of 118 conserved proteins. The proteins used for tree reconstruction represent the phyloeco set for the class *Gammaproteobacteria* [38]. The tree was reconstructed using an internally developed pipeline described in our earlier work [36, 39, 40]. Briefly, using the profile hidden Markov models [41] of different proteins from the phyloeco set, the members of these protein families were identified in the input genomes using HMMer 3.1 [42]. Based on these results, only those protein families where the proteins shared a minimum of 50% in sequence identity and sequence length, and where the protein was found in at least 80% of the input genomes, were retained for phylogenetic analysis. Multiple sequence alignments of these protein families were generated using the Clustal Omega algorithm [43], and TrimAl [44] was used to remove poorly aligned regions before their concatenation into a single file. The final concatenated sequence alignment used for phylogenetic analysis contained a total of 39224 aligned positions. A maximum-likelihood (ML) tree based on this was reconstructed using the Whelan and Goldman model [45] of protein sequence evolution in FastTree 2 [46]. Optimization of the robustness of the trees was completed by conducting SH tests [47] in RAxML 8 [48] and the tree was drawn using MEGA X [49].

Genome sequences for the type and some well-studied strains from the two clades were used to calculate the pairwise average amino acid identity (AAI) and pairwise average nucleotide identity (ANI) between different strains. The AAI values were calculated using the AAI calculator available online (http://enve-omics.ce.gatech.edu/aai/) [50, 51], whereas OrthoANI values were determined using the EzBioCloud.net webserver [52, 53]. The digital DNA–DNA hybridization values (dDDH) values for the same genomes were determined using the Genome to Genome Distance Calculator, available at https://ggdc.dsmz.de/home.php [54]. The 16S rRNA gene sequences for the type and some other strains of *P. aeruginosa* strains were retrieved from SILVA ribosomal RNA database and the NCBI database. Pairwise similarity between the 16S rRNA gene sequences was determined using the 'Align two sequences option' from the BLASTn program [55].

Identification of CSIs specific for *P. aeruginosa* clades

Identification of CSIs was carried out by similar procedures as described in our earlier work [36, 56, 57]. A local database was created for the downloaded *P. aeruginosa* genomes and genomes for >500 other bacteria that included various *Pseudomonadaceae* species. Local BLASTp searches [55] were carried out on this database using different protein sequences from the genomes of *P. aeruginosa* DSM 50071. Based on these BLASTp searches, sequences for 6–8 strains each from Clade-1 and Clade-2, and 8–10 other species, were retrieved for each protein [56]. Multiple sequence alignments of different proteins were created using ClustalX [58] and these alignments were visually inspected for inserts or deletions of fixed lengths which were flanked on both sides by minimally 4–5 conserved amino acids (aa) in the neighbouring 40–50 aa, and which were specific for either the Clade-1 or Clade-2 *P. aeruginosa* strains. Query sequences encompassing the potential indels and flanking regions (60–100 aa long) were collected and more detailed BLASTp searches were performed on them against the entire local database to determine the group specificities of the observed indels. Signature files for all CSIs of interest were formatted using the SIG_CREATE and SIG_STYLE programs from the GLEANS software package described in our earlier workne [57, 59].

Identification of CSPs specific for the two clades

Identification of CSPs was carried out as described in earlier work [60, 61]. To identify CSPs specific for either Clade-1 or Clade-2 strains, local BLASTp searches were carried out on different proteins from the genomes of *P. aeruginosa* NCTC 13628^T and DSM 50071^T. Results of BLASTp searches were examined to identify those proteins where all significant hits (i.e. E value <10–3 was used as the cutoff) were limited to either the Clade-1 or Clade-2 *P. aeruginosa* strains. Additional BLASTp searches on the sequences of these proteins were carried out against the NCBI nr database to confirm that the identified proteins are specific for the indicated clades.

Bacterial strains and morphological, biochemical and physiological tests

P. aeruginosa strains PAO1 and DSM 50071^T, which are representative strains for Clade-1, were purchased from the German Culture Collection of Microorganisms (DSMZ). Strains NCTC 13628^T and NCTC 12924, representatives for Clade-2, were purchased from the National Collection of Type Cultures, Public Health England. Colony morphotype was assessed on Luria–Bertani (LB) agar and compared with the morphology descriptions by Phillips [62]. Pyocyanin and pyoverdine pigment production was visually observed on Kings A and B agar (Sigma-Aldrich), respectively. Fluorescent pyoverdine production was also confirmed under UV light. Swimming assays performed on tryptone swim plates (1% tryptone (w/v), 0.5% NaCl (w/v), 0.3% agarose (w/v)) dried overnight as described by Rashid and Kornberg [63]. Twitching assays were performed on LB agar with 1% granulated agar (w/v). Plates were incubated at 37 °C for 24 h and the halo diameter was measured. Twitching diameter was confirmed by removing the agar and staining the halo with 1% crystal violet on the surface of the plate.

Elastase activity was assessed on cells grown in LB broth by the elastase Congo-red assay described by Pearson *et al.* [64]. Gelatin hydrolysis assay was carried out as described by Cruz and Torres [65]. Casein hydrolytic activity was measured by plating 100 µl of 1×10^3 c.f.u. ml^{-1} of cells on LB agar plated with 1% (v/v) skimmed milk and measuring the zones of clearance after 24 h incubation at 37 °C. Biolog GEN III MicroPlates were used to assess single carbon source use and chemical tolerance profiles. Susceptibilities of the strains to antibiotics were tested by using the Kirby–Bauer disc diffusion assay by measuring the zone of inhibition in presence of antibiotics.

RESULTS

Phylogenetic analysis and comparative studies on *P. aeruginosa* strains

P. aeruginosa strains in earlier phylogenetic studies have been reported to form several clades [20, 22, 35, 66]. With the aim of reliably delineating the branching pattern of *P. aeruginosa* strains, a phylogenomic tree was reconstructed for 212 *P. aeruginosa* strains covering the genetic diversity of available strains. Of these strains >150 corresponded to the two clades referred to as *Pseudomonas aeruginosa* and 'Pseudomonas aeruginosa_A' clades in the GTDB database [35], and >50 strains covering the four lineages A, B and C1 and C2, described by Ozer *et al.* based on their analysis of 739 *P. aeruginosa* strains [22]. A maximum-likelihood phylogenetic tree for these strains was initially reconstructed based on concatenated sequences of 118 conserved proteins comprising the phyloeco set for the class *Gammaproteobacteria* [38]. The resulting tree, which is presented in Fig. S1 (available in the online version of this article) [1], shows the grouping of strains into two main clades. In this tree, the strains from the *P. aeruginosa* main clade and the '*P. aeruginosa*_A' clade from the GTDB are marked by (1) and (2), whereas the numbers (A), (B) and (C1) and (C2) after the strain numbers denote the strains from these specific lineages from the work of Ozer *et al.* [22]. However, the initial analysis of genome sequences in our dataset indicated that the protein sequences for many of these strains were identical, and hence these genomes were omitted from further analysis. A phylogenetic tree based on a smaller subset of 57 *P. aeruginosa* strains covering their genetic diversity is presented in Fig. 1. The analysed *P. aeruginosa* strains in this tree formed two distinct clades. The first clade containing >70% of the analysed strains including strain DSM 50071^T is labelled as Clade-1 (or classical clade). This clade encompasses all strains identified as *P. aeruginosa* (*sensu stricto*) in the GTDB taxonomy [35], as well as all examined strains from lineages (A), (B) and (C1) of Ozer *et al.* [22]. On the other hand, Clade-2 consists solely of different *P. aeruginosa* strains that are assigned to the group '*P. aeruginosa*_A' in the GTDB taxonomy [35], or in the lineage C2 by Ozer *et al.* [22]. Clade-2 includes the well-studied strains ATCC 9027, PA7 and NCTC 13628^T [20, 21, 25, 34]. We have designated strain NCTC 13628^T (=ATCC 9027=PA7) as the type strain of this clade. Phylogenetic distinctness of the strains from these two clades is also supported by earlier studies [16, 17, 20, 21, 28, 67]. We also reconstructed a phylogenetic tree based on 16S rRNA gene sequences for selected *P. aeruginosa* strains. However, in contrast to the phylogenetic trees based on concatenated sequences for core proteins, the strains from these two clades are not resolved in the tree based on 16S rRNA gene sequences (Fig. S2).

In view of the grouping of *P. aeruginosa* strains into two distinct clades, further studies were carried out on genome sequences of selected strains from these clades to determine similarities/differences based on whole genome sequence-based criteria. The ANI, AAI and dDDH values provide three genome-sequence based criteria with established threshold values for bacteria species delineation [52, 54, 68–70]. The results for pairwise ANI, AAI and dDDH similarities between the reference and some other well-studied strains from Clade-1 and Clade-2 are summarized in Table 1. As can be seen from this table, the intragroup OrthoANI and AAI values for the strains from these two clades are >98.8%, while the intergroup OrthoANI and AAI values for these two clades are in the range of 93.4–93.7% and 95.0–95.3%, respectively. Previously, Sood *et al.* [20] reported an ANI matrix for multiple strains from Clade-1 (classical strains) and Clade-2 (outlier clade). Their results also showed that the ANI values within the clades were in the range of 98–99%, whereas interclade ANI values were around 94% [20]. In addition to the information presented in Table 1, we have also constructed a pairwise AAI comparison matrix for several other strains from Clade-1 and -2. Based on the results presented in Table 1, Fig. S3 and by Sood *et al.* [20], the ANI and AAI values between the Clade -1 and Clade -2 strains are consistently lower than the threshold values of <95.0% (ANI) and <95.5% (AAI) for species boundaries [52, 68, 69]. Similarly, while the intragroup dDDH similarity between the strains from these two clades is very high (>89.7%), the interclade dDDH values for all studied strains were in the range of 52–53%, which is again much lower than the threshold value of 70% for species demarcation [54]. We have also determined sequence similarity between the 16S rRNA gene sequences from the reference strains and some other strains (Table 1). As can be seen from Table 1, the 16S rRNA gene from the type strain of *P. aeruginosa* DSM 50071^T exhibited <98.3% sequence similarity to the reference strain NCTC 13628^T of Clade-2. However, the sequence similarity of another strain PAO1 from Clade-1 to the strains from Clade-2 was high, indicating that the 16S rRNA similarity does not provide a reliable means for distinguishing the strains from these clades, which is in accordance with the tree reconstructed based on 16S rRNA gene sequences (Fig. S2).

We have also examined reference strains from Clade-1 and Clade-2 for differences in morphological, biochemical and physiological characteristics (Table 2). The strains from the two clades are very similar in terms of colony morphology, pigmentation and enzyme production. They also do not show any consistent differences in their ability to utilize different carbon sources or other properties such as chemical tolerance or antibiotic resistance profile (Table S1). However, one important difference seen between the strains from these two clades is in their motility pattern. While members of both clades exhibited swimming motility,

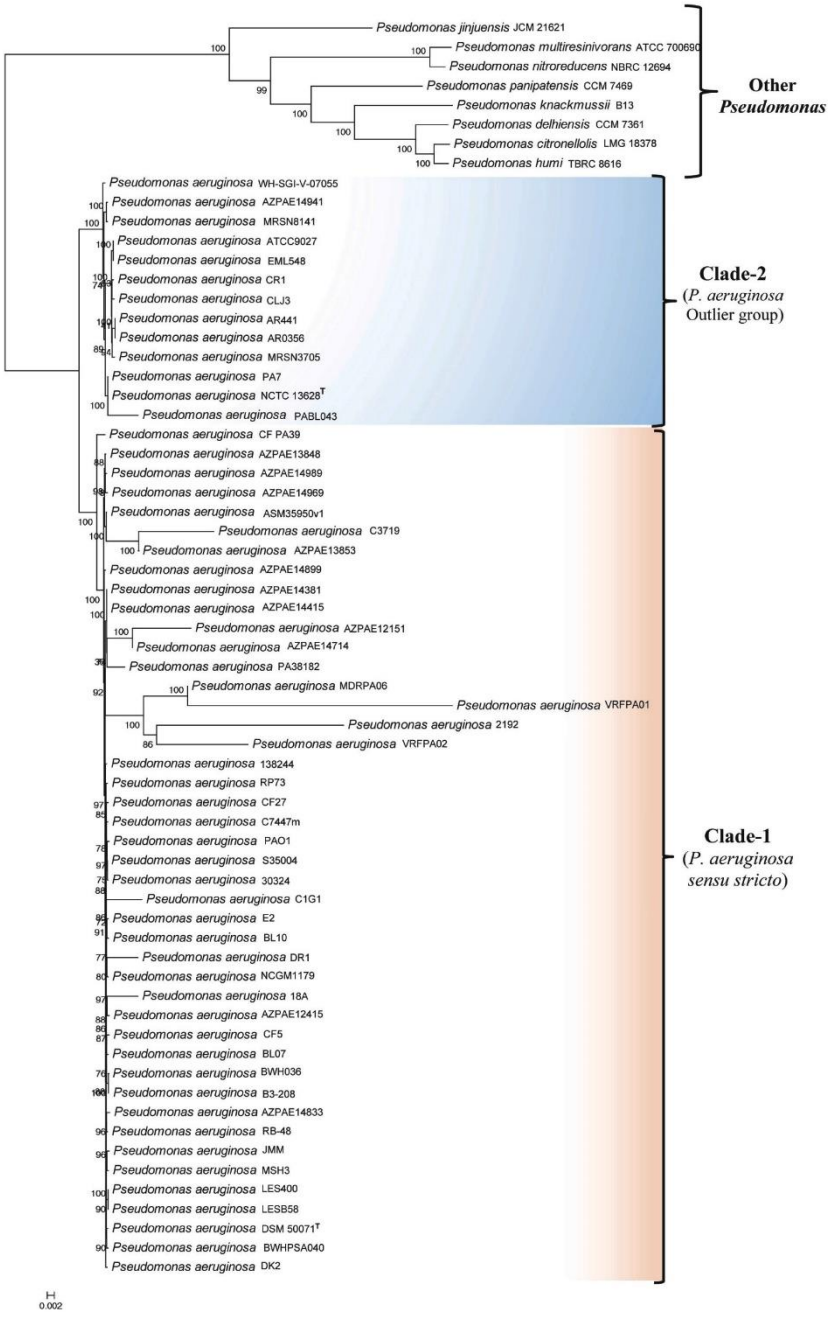

Fig. 1. A bootstrapped maximum likelihood tree for 57 genome-sequenced *P. aeruginosa* strains covering the genetic diversity of available strains, based on concatenated sequences for 118 conserved proteins. A more detailed tree for 212 *P. aeruginosa* strains is presented in Fig. S1. The two main clades of *P. aeruginosa* strains observed in this tree are marked. Clade-1, which contains the type strain DSM 50071^T, corresponds to *P. aeruginosa sensu stricto*. Clade-2 labelled as the *P. aeruginosa* outlier group, represents a new species level lineage. These two clades have also been referred to in earlier work as the 'classical' and 'outlier' groups.

Table 1. Comparison of *P. aeruginosa* strains from Clade-1 and Clade-2 based on average nucleotide identity (ANI), average amino acid identity (AAI), digital DNA–DNA hybridization (dDDH) and 16S rRNA gene sequence similarity

			Clade-1 (*P. aeruginosa sensu stricto*)			Clade-2 (*P. aeruginosa* outlier group)		
Properties		**Strain no.**	**DSM 50071T**	**PAO1**	**Strain B**	**PA7**	**CR1**	**NCTC 13628T**
OrthoANI values (%)*		DSM 50071^T	100					
	Clade-1	PAO1	99.40	100				
		Strain B	98.82	98.81	100			
		PA7	93.70	93.60	93.62	100		
	Clade-2	CR1	93.49	93.44	93.41	99.10	100	
		NCTC 13628^T	93.50	93.40	93.49	98.98	99.48	100
AAI values (%)*		DSM 50071^T	100					
	Clade-1	PAO1	99.20	100				
		Strain B	99.14	99.16	100			
		PA7	95.20	95.30	95.10	100		
	Clade-2	CR1	95.13	95.09	95.08	99.36	100	
		NCTC 13628^T	95.10	95.20	95.03	98.80	99.47	100
dDDH values (%)*		DSM 50071^T	100					
	Clade-1	PAO1	94.90	100				
		Strain B	89.90	89.70	100			
		PA7	53.10	52.60	52.90	100		
	Clade-2	CR1	52.40	52.00	52.00	92.20	100%	
		NCTC 13628^T	52.50	52.00	52.30	91.40	95.90	100
16S rRNA gene sequence similarity (%)†		DSM 50071^T	100					
	Clade-1	PAO1	98.50	100				
		Strain B	99.47	98.14	100			
		PA7	98.20	99.73	98.19	100		
	Clade-2	CR1	98.31	99.79	97.87	99.99	100	
		NCTC 13628^T	98.17	99.66	97.82	99.80	99.93	100

*The accession numbers of genome sequence used for these comparisons are as follows: DSM 50071^T, GCA_001045685; PAO1, GCA_000006765; strain B, GCA_900148065; PA7, GCA_000017205; CR1, GCA_003025345.2; NCTC 13628^T, GCA_900706985.
†The accession numbers of 16S rRNA gene sequences used for these studies are: DSM 50071^T, NR_026078; PAO1, DQ777865; strain B, MW190086; PA7, CP000744; CR1, KC522362; NCTC 13628^T, ON359917.

twitching motility, which plays an important role in bacterial pathogenesis [71], was only seen for the Clade-1 strains and not observed in the strains from Clade-2.

Identification of molecular markers distinguishing the two *P. aeruginosa* clades

To further investigate the differences between the strains from Clade-1 and Clade-2, we performed detailed comparative studies on protein sequences from their genomes to identify reliable molecular markers that are specific for the members of these two clades. Based on genome sequences, one important class of molecular markers which have proven very useful for evolutionary and taxonomic studies consists of conserved signature indels (inserts or deletions) referred to as CSIs in gene/protein sequences, which are specifically shared by the members of a given clade [57, 59, 72, 73]. Due to the exclusive presence of these molecular signatures in the members of a given clade, these synapomorphic characteristics provide reliable means for the demarcation of prokaryotic taxa in molecular terms [39, 74]. We have recently used CSIs to reclassify >20 *Pseudomonas* species into two novel genera (*viz*. *Halopseudomonas* and *Atopomonas*) and some other genera [36]. Another important category of molecular marker consists of CSPs, which are uniquely found in a specific group of organisms [61, 75–77]. Hence, detailed studies were conducted

Table 2. Selected morphological and biochemical properties for representative Clade-1 and Clade-2 strains

Strains: 1, *P. aeruginosa* PAO1; 2, *P. aeruginosa* DSM 50071^T; 3, *P. aeruginosa* NCTC 13628^T; 4, *P. aeruginosa* NCTC 12924. +, Positive; –, negative; ND, no data.

	Clade-1 (*P. aeruginosa sensu stricto*)		**Clade-2 (*P. aeruginosa* outlier group)**	
Characteristics	**1**	**2**	**3**	**4**
Isolation source	Wound	Hospital in Japan	Outer ear infection	ND
Motility:				
Swimming	+	+	+	+
Twitching	+	+	–	–
Colony morphology	Type 1 typical	Type 1 typical	Type 1 typical	Type 1 typical
Pigments:				
Pyoverdine	+	+	+	+
Pyocyanin	+	+	+	+
Protease and enzyme production:				
Elastase	+	+	+	+
Gelatinase	+	+	+	+
Casein hydrolysis	+	+	+	+
Arginine dihydrolase	+	+	+	+
Carbon utilization sources:				
Trehalose	–	+	–	–
Turanose	–	+	–	–
Lactose	–	+	–	–
D-Arabitol	+	+	+	+
D-Glucose-6-PO4	-	+	+	–
Gelatin	+	+	+	–
Glycyl-L-proline	+	+	+	+
Pectin	–	+	–	–
Mucic acid	–	+	+	–
D-Lactic acid methyl ester	–	+	–	–
D-Malic acid	–	+	+	–
Chemical tolerance:				
Sodium bromate	+	+	+	+
8% NaCl	+	+	+	–

Information for antibiotic resistance profiles of the strains is presented in Table S1.

on genome sequences of *P. aeruginosa* strains to identify CSIs and CSPs that are distinctive characteristics of their two clades. These studies have identified multiple CSIs that are exclusively shared by either the members of Clade-1 or Clade-2.

In Fig. 2, we show partial sequence information for two CSIs that are specifically shared by different strains from Clade-1. In the example shown in Fig. 2a two amino acid (aa) insertion (highlighted in pink) is present in a conserved region of the fimbrial biogenesis outer membrane usher protein that is commonly shared by all strains from clade one but is not found in strains from Clade-2 and other *Pseudomonas* species. Likewise, in Fig. 2b in the partial sequence alignment of the Type II secretion system F family protein, a one aa deletion is present which is specific for the Clade-1 strains. In Fig. 2A and B (and also in Fig. 3), sequence

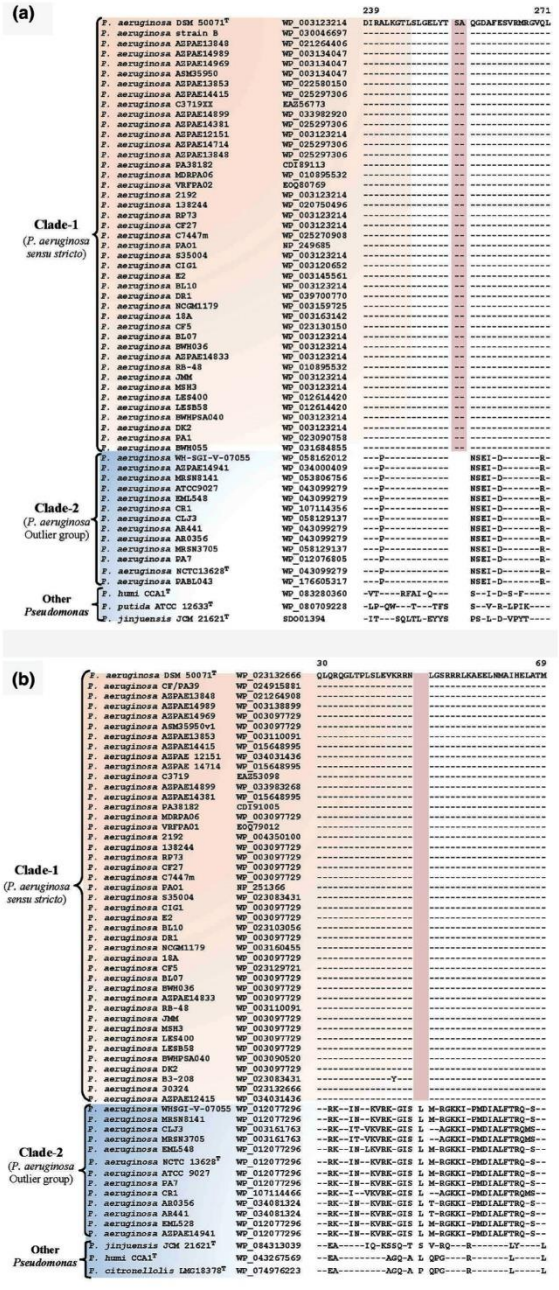

Fig. 2. Partial sequence alignments of (a) fimbrial biogenesis outer membrane usher protein containing a two aa insertion (boxed), and (b) Type II secretion system F family protein containing a one aa deletion (highlighted), which are commonly and exclusively shared by different strains which are a part of the P. aeruginosa *sensu stricto* clade (Clade-1) but not found in any of the strains from Clade-2 or the outgroup species. More detailed sequence information for these two CSIs and five other CSIs specific for Clade-1 is provided in Figs S4–S10 and some of their characteristics are summarized in Table 3. The dashes (-) in this and other sequence alignments indicate identity with the amino acids on the top line. Accession numbers for different sequences are indicated in the second column and the numbers on the top of the sequence indicate the position of this sequence fragment within the indicated protein.

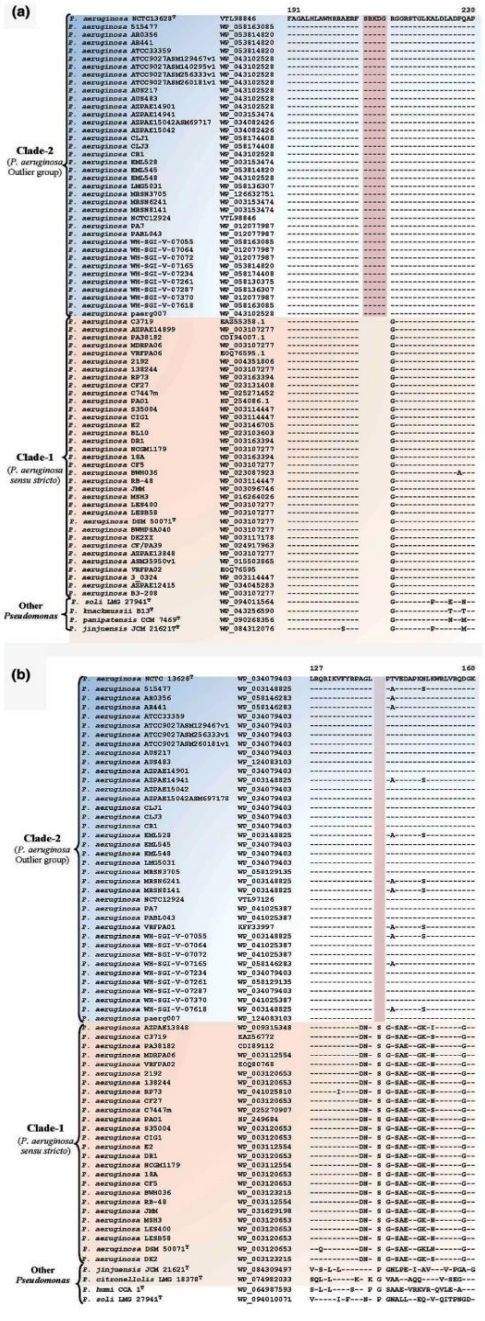

Fig. 3. Partial sequence alignments of (a) dimethylglycine demethylation protein (DgcB) containing a five aa insertion (boxed), and (b) chaperone protein FocC containing a one aa deletion (highlighted), which are commonly and exclusively shared by different *P. aeruginosa* strains from Clade-2, but not found in strains from Clade-1 or outgroup species. More detailed sequence information for these two CSIs and three other CSIs specific for the Clade-2 strains is provided in Figs S11–S15 and some of their characteristics are summarized in Table 3. The dashes (-) in these sequence alignments indicate identity with the amino acids on the top line.

Table 3. Conserved signature indels (CSIs) specific for *P. aeruginosa* strains from Clade-1 and Clade-2

Protein name	Accession no:	Figure no.	Indel size (aa)	Indel location	Specificity
Fimbrial biogenesis outer membrane usher protein	WP_003123214.1	Fig. 2a; Fig. S4	2 aa Ins	239–271	Clade-1 (*P. aeruginosa sensu stricto*)
Type II secretion system F family protein	WP_023132666.1	Fig. 2b;Fig. S5	1 aa Del	30–74	
Major facilitator superfamily (MFS) transporter protein	WP_003106206.1	Fig. S6	1 aa Ins	155–212	
LysR family transcriptional regulator	WP_048520948.1	Fig. S7	1 aa Ins	251–293	
Fe(3+)-pyochelin receptor FptA protein	WP_003118950 .1	Fig. S8	1 aa Ins	493–518	
MFS family transporter	WP_017002609	Fig. S9	1 aa Del	216–257	
Putative oxidoreductase	EOQ77226	Fig. S10	2 aa Del	109–149	
Dimethylglycine demethylation protein DgcB	VTL98846.1	Fig. 3a; Fig. S11	5 aa Ins	191–230	Clade-2 (*P. aeruginosa* outlier group)
Molecular chaperon	WP_034079403.1	Fig. 3b; Fig. S12	1 aa Del	127–160	
Adenylate cyclase	WP_053817914.1	Fig. S13	7 aa Del	163–191	
TolC family protein	WP_053816642.1	Fig. S14	1 aa Del	251–283	
Type one fimbrial protein	WP_079384945 .1	Fig. S15	1 aa Ins	96–144	

information is presented for only a limited number of representative *P. aeruginosa* strains. However, more detailed information for the presence/absence of these CSIs in different *P. aeruginosa* that are a part of our study is presented in Figs S4 and S5. As both these CSIs are absent in all Clade-2 strains and other *Pseudomonas* species, they represent genetic changes which were likely introduced in a common ancestor of the Clade-1 strains. In addition to these two CSIs, our analyses have identified five CSIs in other proteins, which are also exclusively shared by the strains from Clade-1. Detailed sequence information for these five CSIs is presented in Figs S6–S10 and some of their characteristics are summarized in Table 3. Besides the CSIs that are specific for the Clade-1 strains, our analyses have also identified five CSIs that are only found in the strains from Clade-2. Fig. 3 shows partial sequence information for two CSIs those are specific for the Clade-2 strains. In the first of these CSIs (Fig. 3a), a five aa insertion is present in the dimethylglycine demethylation protein (DgcB), which is specifically present in all strains from Clade-2, but not found in any of the strains from Clade-1. Similarly, in the CSI shown in Fig. 3b one aa deletion is present in the chaperon protein FocC, which is exclusively shared by the Clade-2 strains. More detailed sequence information for these two CSIs and three CSIs in other proteins, which are also specific for the Clade-2 strains, is presented in Figs S11–S15 and some of their characteristics are summarized in Table 3. Because of the clade specificity and highly specific nature of the genetic changes that are responsible for these molecular markers, the identified CSIs provide novel and reliable means for the demarcation of these clades in molecular terms and distinguishing them from each and all other bacteria.

In addition to these CSIs, our comparative genomic studies have also identified multiple CSPs that are exclusively found in the members of these two clades. In contrast to the CSIs, where the homologues of the proteins containing the identified CSIs are present in different strains, the homologues of the identified CSPs are only found in a given clade/group of organisms [61, 75–77]. Our analyses have identified nine CSPs for which all significant BLASTP hits are limited to the Clade-1 strains, and 15 CSPs for whom homologues showing significantly sequence similarity are only found in Clade-2 strains. In Table 4, we have summarized information regarding the accession numbers and some other characteristics of the identified CSPs for the Clade-1 and Clade-2 strains. Most of the identified CSPs represent proteins whose functions are not known (annotated as hypothetical proteins). However, two of the CSPs specific for Clade-1 are related to type III secretion system (*viz*. Acr1 family type III secretion system gatekeeper subunit Pcr1 and Type III secretion system needle filament protein PscF), which is only found in Clade-1 strains. The identified CSIs and CSPs, due to their specificity and exclusive presence in *P. aeruginosa* strains from either Clade-1 or Clade-2, again provide strong evidence for the genetic distinctness of these two clades and provide novel and reliable means for distinguishing/identifying them from each other.

DISCUSSION

Results presented here provide compelling evidence that the existing *P. aeruginosa* strains form two phylogenetic distinct clades differing from each other in numerous regards. In addition to their distinct branching in a phylogenomic tree based on core genomic proteins, the grouping of these strains into clades similar to those observed here, has been demonstrated in several earlier studies [16, 17, 20, 21, 28, 35, 67]. Of the two observed clades, Clade-1, which encompasses >99% of the sequenced strains including the type strain DSM 50071^T (based on GTDB grouping/classification) [35], corresponds to the species *P. aeruginosa*

Table 4. Some characteristics of the conserved signature proteins (CSPs) that are uniquely found in the Clade-1 and Clade-2 strains*

Protein name	Accession no.	Length (aa)	Specificity
Hypothetical protein	WP_003091936.1	39	Clade-1 (*P. aeruginosa sensu stricto*)
Hypothetical protein	WP_003093484.1	99	
Hypothetical protein	WP_003083536.1	131	
Hypothetical protein	WP_003082890.1	120	
Acr1 family type III secretion system gatekeeper subunit Pcr1	WP_003087693.1	92	
Type III secretion system needle filament protein PscF	WP_003087729.1	85	
Alpha/beta hydrolase	WP_003082501.1	211	
TauD/TfdA family dioxygenase	WP_003082503.1	319	
phytanoyl-CoA dioxygenase family protein	WP_003082507.1	292	
Hypothetical protein	WP_033996971.1	542	Clade-2 (*P. aeruginosa* outlier group)
Hypothetical protein	WP_034080176.1	105	
DUF3277 family protein	WP_034080180.1	143	
Hypothetical protein	WP_034080540.1	370	
Hypothetical protein	WP_034080541.1	269	
NUDIX hydrolase	WP_034080754.1	191	
Hypothetical protein	WP_034080756.1	222	
Hypothetical protein	WP_034081197.1	107	
Hypothetical protein	WP_034081220.1	217	
Hypothetical protein	WP_043099652.1	251	
Hypothetical protein	WP_043101353.1	92	
Hypothetical protein	WP_043103086.1	305	
Esterase-like activity of phytase family protein	WP_052151135.1	318	
Glycosyltransferase family 39 protein	WP_053814840.1	559	
Fatty acid desaturase family protein	WP_053816651.1	339	

*For the proteins listed in this table, homologues showing significant sequence similarity are only limited to the members of the indicated clades.

sensu stricto, whereas the strains from Clade-2, which is referred to in earlier work as the outlier group [20, 21, 25, 26], or as 'Pseudomonas aeruginosa_A' in the GTDB classification [35], constitutes a novel species level lineage. Results presented here for OrthoANI, AAI and dDDH similarity values, which are established genome sequence-based criteria for species demarcation, show that these values between the members of Clade-1 and Clade-2 are lower than the accepted cut-off values for species boundaries, thus indicating that the strains from Clade-2 constitute a novel species. Although the 16S rRNA sequence similarity values vary for different strains from these two clades and often fails to distinguish among closely related species [37, 78, 79], the type strain DSM 50071^T of *P. aeruginosa* also shows <98.3% sequence similarity to the reference (type strain) NCTC 13628^T from Clade-2, supporting the inference that the Clade-2 constitutes a novel species.

In addition to the distinctness of these two clades based on phylogenetic and whole-genome sequence-based criteria, the present work has identified multiple molecular markers consisting of CSIs and CSPs that are uniquely shared characteristics of either the members of Clade-1 or Clade-2. These molecular markers provide strong independent evidence that the species/strains from these two clades are distinct from each other. The identified CSIs and CSPs, which are specific for these two clades, provide very useful and reliable means for the circumscription of these two clades in molecular terms and for distinguishing the members of this clade from each other and all other bacteria based upon the presence or absence of the identified molecular characteristics.

Based on earlier work, these molecular markers possess high degree of predictive ability to be found in other members of these clades [36, 80, 81]. Thus, based on BLASTp searches with the sequences of these molecular markers, the strains belonging to Clade-1 and Clade-2 can be readily identified and distinguished from each other. Furthermore, based on earlier work on CSIs and CSPs, these molecular markers are predicted to play functionally important roles in the organisms for which these are specific [82, 83]. It is of interest in this regard that some of the identified CSIs and CSPs are present in proteins that are related to the Type III secretion system (Tables 3 and 4). Thus, functional studies on the identified CSIs and CSPs could provide useful information regarding novel genetic, biochemical or pathogenic properties of these two groups of species/strains, of which the *P. aeruginosa sensu stricto* (Clade-1) represents an important pathogen [7, 8, 84].

Besides the differences in phylogenetic, genomic and molecular characteristics differentiating the Clade-1 or Clade-2 strains, the strains from these two clades also differ from each other in their pathogenic mechanisms [15, 20, 25], in their ability to produce biosurfactants, phenazines and siderophores (mainly limited to strains from clade 2) [23, 34], presence or absence of gene clusters required for survival in stress conditions [20], and the differences in major protein–protein interaction hubs between strains from these two groups [18, 20, 26]. Additionally, the results presented here show that in contrast to the swimming motility, which is found in strains from both Clade-1 and Clade-2, the twitching motility is only present in Clade-1 strains. Results from different lines of evidence discussed and presented here, make a strong case that the strains from Clade-2, which is distinct from the *P. aeruginosa sensu stricto* clade (Clade-1), constitute a novel species for which we are proposing the name *Pseudomonas paraeruginosa* with the strain NCTC 13628^T, as its type strain.

DESCRIPTION OF *PSEUDOMONAS PARAERUGINOSA* SP. NOV.

Pseudomonas paraeruginosa (par.ae.ru.gi.no'sa. Gr. pref. *para*, beside, alongside, near, like; L. fem. adj. *aeruginosa*, copper-rust coloured, specific epithet of a *Pseudomonas* species. N.L. fem. adj. *paraeruginosa*, next to or near *aeruginosa*).

Gram-reaction-negative, aerobic, rod-shaped bacteria. Exhibit swimming motility but lack twitching motility. Growth occurs in medium containing 0–6.5% (w/v) NaCl at 15–42 °C, whereas no growth occurs at 4 or 45 °C. Produce the pigments pyocyanin and pyoverdine and are positive in casein hydrolysis, arginine dihydrolase, elastase and gelatinase tests. Strains differ in their ability but can utilize D-glucose-6-PO_4, gelatin, D-arabitol, glycyl-L-proline, mucic acid and D-malic acid as a carbon source. Strains lack type III secretion system (T3SS) but contain exolysin (exlBA). Generally, less virulent than the *P. aeruginosa* strains. Currently sensitive to tobramycin, piperacillin and meropenem, but show intermediate resistance to several other antibiotics (ceftazidime, aztreonam, ciprofloxacin, levofloxacin, ticarcillin and cefepime). Several strains produce siderophores, phenazines, biofilm-inhibiting pyocyanin and a biosurfactant (mono-rhamnolipids). Contain fused *phzA1* with *phzB1* genes. The strains from this species form a monophyletic clade distinct from *P. aeruginosa* strains in phylogenetic trees based on concatenated sequences for large datasets of conserved proteins. Can also be differentiated from *P. aeruginosa* strains based on ANI, AAI and dDDH analyses. The strains of this species can be reliably distinguished from *P. aeruginosa* strains and other *Pseudomonas* species based upon the presence of five CSIs (Table 3), which are uniquely shared by the strains from this species. The proteins containing these CSIs are: dimethylglycine demethylation protein DgcB, Chaperone protein FocC, Adenylate cyclase, TolC family protein and Type 1 fimbrial protein. Additionally, homologues of 15 CSPs described in Table 4 are also uniquely found in the strains of this species and provide reliable means for their identification.

The accession numbers for the 16S rRNA gene sequence and genome sequence for the type strain NCTC 13628^T of *Pseudomonas paraeruginosa* are ON359917 and GCA_900706985, respectively.

The type strain is NCTC 13628^T (=PA7^T=ATCC 9027^T).

EMENDED DESCRIPTION OF THE SPECIES *PSEUDOMONAS AERUGINOSA* (SCHROETER 1872) MIGULA 1900 (APPROVED LISTS 1980)

The description of this species is modified from that given by Doudoroff andPalleroni in *Bergey's Manual of* Determinative Bacteriology [85], and also by Palleroni[86], and Diggle and Whiteley [87]]. The cells are rod-shaped, about 1–5 µm long and 0.5–1.0 µm wide. Gram-stain-negative. Generally, strictly aerobic, chemoorganotrophs using respiratory metabolism with molecular oxygen as the electron acceptor. Exhibits both swimming and twitching mobility. Growth occurs in medium containing 0–6.5% (w/v) NaCl at 15–42 °C, whereas no growth occurs at 4 or 45 °C. Strains of this species generally possess a type III secretion system (T3SS), which enables the injection of four main *P. aeruginosa* effectors (ExoS, ExoT, ExoY and ExoU) directly into the host cell cytoplasm. Opportunistic pathogen, associated with diseases such as cystic fibrosis, ventilator-associated pneumonia and various sepsis syndromes. The type strain of this species can utilize different carbon sources including D-trehalose, D-turanose, D-lactose, D-glucose-6-PO_4, pectin, gelatin, D-arabitol, glycyl-L-Proline, mucic acid and D-malic acid. Shows positive results for casein hydrolysis, arginine dihydrolase, elastase and gelatinase assays. Strains are sensitive to tobramycin and meropenem but show intrinsic resistance to many antibiotics through multiple mechanisms, *viz.*, reduced membrane permeability, drug efflux systems and production of antibiotic-inactivating enzymes. Strains of this species form a monophyletic clade in phylogenetic

trees based on concatenated sequences for several large datasets of proteins. *P. aeruginosa* strains can also be differentiated from *P. paraeruginosa* based on ANI, AAI and dDDH analyses. In addition, these strains can be reliably distinguished from all other strains/species by the presence of seven CSIs described in Table 3, found in the following proteins: Fimbrial biogenesis outer membrane usher protein, Type II secretion system F family protein, major facilitator superfamily (MFS) transporter protein, LysR family transcriptional regulator, Fe(3+)-pyochelin receptor FptA protein, another CSI in an MFS family transporter protein, and a CSI in putative oxidoreductase. Additionally, homologs of nine CSPs described in Table 4 are also uniquely found in the strains of species and provide reliable means for their identification.

The accession numbers for the 16S rRNA gene sequence and genome sequence for the type strain DSM 50071^T of *Pseudomonas aeruginosa* are NR_026078 and GCA_001045685, respectively.

The type strain is DSM 50071^T (=ATCC 10145^T=ATCC 10145 U^T=CCEB 481^T=CCUG 28447^T=CCUG 29297^T=CCUG 551^T=CFBP 2466^T=CIP 100.720^T=DSM 50071^T=IBCS 277^T=IFO 12689^T=JCM 5962^T=LMG 1242^T=NBRC 12689^T=NCCB 76039^T=NCIB 8295^T=NCIMB 8295^T=NCTC 10332^T=NRRL B-771^T=RH 815^T=VKM B-588^T).

Funding information
This work was supported in part by a Discovery research grant (RGPIN 2019-06397) from the Natural Sciences and Engineering Research Council of Canada, and a grant from the Ontario Research Fund to R.S.G. L.D. is supported by the Biotechnology and Biological Sciences Research Council (grant BB/M009513/1).

Acknowledgements
We thank Alvin Han for assistance in this study.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Rudra BD L, Duncan L, Shah AJ, Shah HN, Gupta RS.** Phylogenomic and comparative genomic studies robustly demarcate two distinct clades of *Pseudomonas aeruginosa* strains: proposal to transfer the strains from an outlier clade to a novel species *Pseudomonas paraeruginosa* sp. nov. *Figshare* 2022. DOI: 10.6084/m9.figshare.20277480.
2. **Peix A, Ramírez-Bahena M-H, Velázquez E.** Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect Genet Evol* 2009;9:1132–1147.
3. **Whitman WB, Rainey F, Kämpfer P, Trujillo M, Chun J, *et al.*** *Pseudomonas*. John Wiley and Sons in Association with Bergey's Manual Trust, 2015. pp. 1–105.
4. **Schroeter J.** Über einige durch Bacterien gebildete Pigmente. In: **Cohn F** (eds). *Beiträge Zur Biologie Der Pflanzen*. Breslau: Max Müller; 1872. pp. 109–126.
5. **Migula W.** Uber ein neues system der bakterien. *Arb Bakt Inst Kar1sruhe* 1894:1.
6. **Sneath PHA, McGowan V, Skerman VBD.** Approved lists of bacterial ames. *Int J Syst Bacteriol* 1980;30:225–420.
7. **Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, *et al.*** Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 2000;406:959–964.
8. **Planquette B, Timsit J-F, Misset BY, Schwebel C, Azoulay E, *et al.*** *Pseudomonas aeruginosa* ventilator-associated *Pneumonia*. predictive factors of treatment failure. *Am J Respir Crit Care Med* 2013;188:69–76.
9. **Castañeda-Montes FJ, Avitia M, Sepúlveda-Robles O, Cruz-Sánchez V, Kameyama L, *et al.*** Population structure of *Pseudomonas aeruginosa* through a MLST approach and antibiotic resistance profiling of a Mexican clinical collection. *Infect Genet Evol* 2018;65:43–54.
10. **Gad GF, El-Domany RA, Zaki S, Ashour HM.** Characterization of *Pseudomonas aeruginosa* isolated from clinical and environmental samples in Minia, Egypt: prevalence, antibiogram and resistance mechanisms. *J Antimicrob Chemother* 2007;60:1010–1017.
11. **Kouda S, Ohara M, Onodera M, Fujiue Y, Sasaki M, *et al.*** Increased prevalence and clonal dissemination of multidrug-resistant *Pseudomonas aeruginosa* with the blaIMP-1 gene cassette in Hiroshima. *J Antimicrob Chemother* 2009;64:46–51.
12. **Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, *et al.*** Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 2016;44:D646–53.
13. **Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, *et al.*** Database resources of the national center for biotechnology nformation. *Nucleic Acids Res* 2019;47:D23–D28.
14. **Toska J, Sun Y, Carbonell DA, Foster AN-S, Jacobs MR, *et al.*** Diversity of virulence phenotypes among type III secretion negative *Pseudomonas aeruginosa* clinical isolates. *PLoS One* 2014;9:e86829.
15. **Freschi L, Bertelli C, Jeukens J, Moore MP, Kukavica-Ibrulj I, *et al.*** Genomic characterisation of an international *Pseudomonas aeruginosa* reference panel indicates that the two major groups draw upon distinct mobile gene pools. *FEMS Microbiol Lett* 2018;365:14.
16. **Basso P, Ragno M, Elsen S, Reboud E, Golovkine G, *et al.*** *Pseudomonas aeruginosa* pore-forming exolysin and Type IV Pili cooperate to induce host cell lysis. *mBio* 2017;8:e02250-16.
17. **Brzozowski M, Jursa-Kulesza JB, Kosik-Bogacka D.** Virulence analysis of 81 of *Pseudomonas aeruginosa* genomes available in public sequence databases. *In Review*. 2019. DOI: 10.21203/rs.2.10146/v1.
18. **García-Reyes S, Soto-Aceves MP, Cocotl-Yañez M, González-Valdez A, Servín-González L, *et al.*** The outlier *Pseudomonas aeruginosa* strain ATCC 9027 harbors a defective LasR quorum-sensing transcriptional regulator. *FEMS Microbiol Lett* 2020;367:16.
19. **García-Reyes S, Soberón-Chávez G, Cocotl-Yanez M.** The third quorum-sensing system of *Pseudomonas aeruginosa*: *Pseudomonas* quinolone signal and the enigmatic PqsE protein. *J Med Microbiol* 2020;69:25–34.
20. **Sood U, Hira P, Kumar R, Bajaj A, Rao DLN, *et al.*** Comparative genomic analyses reveal core-genome-wide genes under positive selection and major regulatory hubs in outlier strains of *Pseudomonas aeruginosa*. *Front Microbiol* 2019;10:53.
21. **Zhang Z, Zhang X.** Evolution of subfamily I.1 lipases in *Pseudomonas aeruginosa*. *Curr Microbiol* 2021;78:3494–3504.
22. **Ozer EA, Nnah E, Didelot X, Whitaker RJ, Hauser AR.** The population structure of *Pseudomonas aeruginosa* Is characterized by genetic isolation of exoU+ and exoS+ Lineages. *Genome Biol Evol* 2019;11:1780–1796.

23. **Sood U**, **Singh DN**, **Hira P**, **Lee J-K**, **Kalia VC**, ***et al***. Rapid and solitary production of mono-rhamnolipid biosurfactant and biofilm inhibiting pyocyanin by a taxonomic outlier *Pseudomonas aeruginosa* strain CR1. *J Biotechnol* 2020;307:98–106.
24. **Elsen S**, **Huber P**, **Bouillot S**, **Couté Y**, **Fournier P**, ***et al***. A type III secretion negative clinical strain of *Pseudomonas aeruginosa* employs a two-partner secreted exolysin to induce hemorrhagic *Pneumonia*. *Cell Host Microbe* 2014;15:164–176.
25. **Roy PH**, **Tetu SG**, **Larouche A**, **Elbourne L**, **Tremblay S**, ***et al***. Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 2010;5:e8842.
26. **García-Ulloa M**, **Ponce-Soto G-Y**, **González-Valdez A**, **González-Pedrajo B**, **Díaz-Guerrero M**, ***et al***. Two *Pseudomonas aeruginosa* clonal groups belonging to the PA14 clade are indigenous to the Churince system in Cuatro Ciénegas Coahuila, México. *Environ Microbiol* 2019;21:2964–2976.
27. **Ruiz-Roldán L**, **Rojo-Bezares B**, **de Toro M**, **López M**, **Toledano P**, ***et al***. Antimicrobial resistance and virulence of *Pseudomonas* spp. among healthy animals: concern about exolysin ExlA detection. *Sci Rep* 2020;10:1–11.
28. **Huber P**, **Basso P**, **Reboud E**, **Attrée I**. *Pseudomonas aeruginosa* renews its virulence factors. *Environ Microbiol Rep* 2016;8:564–571.
29. **Hauser AR**. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Microbiol* 2009;7:654–665.
30. **Engel J**, **Balachandran P**. Role of *Pseudomonas aeruginosa* type III effectors in disease. *Curr Opin Microbiol* 2009;12:61–66.
31. **Shaver CM**, **Hauser AR**. Relative contributions of *Pseudomonas aeruginosa* ExoU, ExoS, and ExoT to virulence in the lung. *Infect Immun* 2004;72:6969–6977.
32. **Vance RE**, **Rietsch A**, **Mekalanos JJ**. Role of the type III secreted exoenzymes S, T, and Y in systemic spread of *Pseudomonas aeruginosa* PAO1 *in vivo*. *Infect Immun* 2005;73:1706–1713.
33. **Lyczak JB**, **Cannon CL**, **Pier GB**. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect* 2000;2:1051–1060.
34. **Zhang Y**, **Miller RM**. Enhanced octadecane dispersion and biodegradation by a *Pseudomonas rhamnolipid* surfactant (biosurfactant). *Appl Environ Microbiol* 1992;58:3276–3282.
35. **Parks DH**, **Chuvochina M**, **Waite DW**, **Rinke C**, **Skarshewski A**, ***et al***. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
36. **Rudra B**, **Gupta RS**. Phylogenomic and comparative genomic analyses of species of the family *Pseudomonadaceae*: proposals for the genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., merger of the genus *Oblitimonas* with the genus *Thiopseudomonas*, and transfer of some misclassified species of the genus *Pseudomonas* into other genera. *Int J Syst Evol Microbiol* 2021;71.
37. **Lalucat J**, **Mulet M**, **Gomila M**, **García-Valdés E**. Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes* 2020;11:E139.
38. **Wang Z**, **Wu M**. A phylum-level bacterial phylogenetic marker database. *Mol Biol Evol* 2013;30:1258–1262.
39. **Adeolu M**, **Alnajar S**, **Naushad S**, **S Gupta R**. 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. *Int J Syst Evol Microbiol* 2016;66:5575–5599.
40. **Gupta RS**, **Patel S**, **Saini N**, **Chen S**. Erratum: robust demarcation of seventeen distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for emended genus *Bacillus* limiting it only to the members of the subtilis and cereus clades of species. *Int J Syst Evol Microbiol* 2020;70:6531–6533.
41. **Eddy SR**. Profile hidden Markov models. *Bioinformatics* 1998;14:755–763.
42. **Eddy SR**. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:10.
43. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, ***et al***. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
44. **Capella-Gutiérrez S**, **Silla-Martínez JM**, **Gabaldón T**. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
45. **Whelan S**, **Goldman N**. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001;18:691–699.
46. **Price MN**, **Dehal PS**, **Arkin AP**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490.
47. **Guindon S**, **Dufayard J-F**, **Lefort V**, **Anisimova M**, **Hordijk W**, ***et al***. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
48. **Stamatakis A**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
49. **Kumar S**, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K**. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
50. **Rodriguez-R LM**, **Konstantinidis KT**. Bypassing cultivation to identify bacterial species. *Microbe Magazine* 2014;9:111–118.
51. **Thompson CC**, **Chimetto L**, **Edwards RA**, **Swings J**, **Stackebrandt E**, ***et al***. Microbial genomic taxonomy. *BMC Genomics* 2013;14:913.
52. **Lee I**, **Ouk Kim Y**, **Park SC**, **Chun J**. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
53. **Varghese NJ**, **Mukherjee S**, **Ivanova N**, **Konstantinidis KT**, **Mavrommatis K**, ***et al***. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 2015;43:6761–6771.
54. **Meier-Kolthoff JP**, **Carbasse JS**, **Peinado-Olarte RL**, **Göker M**. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807.
55. **Altschul SF**, **Madden TL**, **Schäffer AA**, **Zhang J**, **Zhang Z**, ***et al***. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
56. **Gupta RS**. Identification of conserved indels that are useful for classification and evolutionary studies. In: *Methods in Microbiology*, vol. 41. 2014. pp. 153–182.
57. **Gupta RS**. Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol Rev* 2016;40:520–553.
58. **Jeanmougin F**, **Thompson JD**, **Gouy M**, **Higgins DG**, **Gibson TJ**. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998;23:403–405.
59. **Gupta RS**. Identification of conserved indels that are useful for classification and evolutionary studies. In: *Methods in Microbiology*. Oxford: Elsevier, 2014. pp. 153–182.
60. **Gupta RS**. Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (Campylobacterales). *BMC Genomics* 2006;7:167.
61. **Naushad HS**, **Lee B**, **Gupta RS**. Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya*, *Pectobacterium* and *Brenneria*. *Int J Syst Evol Microbiol* 2014;64:366–383.
62. **Phillips I**. Identification of *Pseudomonas aeruginosa* in the clinical laboratory. *J Med Microbiol* 1969;2:9–16.
63. **Rashid MH**, **Kornberg A**. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 2000;97:4885–4890.

64. **Pearson JP**, **Pesci EC**, **Iglewski BH**. Roles of *Pseudomonas aeruginosa* las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *J Bacteriol* 1997;179:5756–5767.
65. **de la Cruz TEE**, **Torres JMO**. Gelatin hydrolysis test protocol. American Society for Microbiology; 2012. https://asm.org/Protocols/Gelatin-Hydrolysis-Test-Protocol
66. **Subedi D**, **Vijay AK**, **Kohli GS**, **Rice SA**, **Willcox M**. Comparative genomics of clinical strains of *Pseudomonas aeruginosa* strains isolated from different geographic sites. *Sci Rep* 2018;8:15668.
67. **Reboud E**, **Basso P**, **Maillard AP**, **Huber P**, **Attrée I**. Exolysin shapes the virulence of *Pseudomonas aeruginosa* clonal outliers. *Toxins* 2017;9:E364.
68. **Kim M**, **Oh H-S**, **Park S-C**, **Chun J**. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.
69. **Konstantinidis KT**, **Tiedje JM**. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 2005;187:6258–6264.
70. **Girard L**, **Lood C**, **Höfte M**, **Vandamme P**, **Rokni-Zadeh H**, ***et al***. The ever-expanding *Pseudomonas* genus: description of 43 new species and partition of the *Pseudomonas putida* group. *Microorganisms* 2021;9:1766.
71. **Mattick JS**. Type IV pili and twitching motility. *Annu Rev Microbiol* 2002;56:289–314.
72. **Gupta RS**. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 1998;62:1435–1491.
73. **Gupta RS**. Microbial taxonomy: how and why name changes occur and their significance for (clinical) microbiology. *Clin Chem* 2021;68:134–137.
74. **Patel S**, **Gupta RS**. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 2020;70:406–438.
75. **Griffiths E**, **Gupta RS**. Protein signatures distinctive of chlamydial species: horizontal transfers of cell wall biosynthesis genes *glmU* from *Archaea* to *Chlamydiae* and *murA* between *Chlamydiae* and *Streptomyces*. *Microbiology* 2002;148:2541–2549.
76. **Gao B**, **Paramanathan R**, **Gupta RS**. Signature proteins that are distinctive characteristics of *Actinobacteria* and their subgroups. *Antonie van Leeuwenhoek* 2006;90:69–91.
77. **Dutilh BE**, **Snel B**, **Ettema TJG**, **Huynen MA**. Signature genes as a phylogenomic tool. *Mol Biol Evol* 2008;25:1659–1667.
78. **Peix A**, **Ramírez-Bahena M-H**, **Velázquez E**. The current status on the taxonomy of *Pseudomonas revisited*: an update. *Infect Genet Evol* 2018;57:106–116.
79. **Garrido-Sanz D**, **Meier-Kolthoff JP**, **Göker M**, **Martín M**, **Rivilla R**, ***et al***. Genomic and genetic diversity within the *Pseudomonas fluorescens* omplex. *PLoS One* 2016;11:e0150183.
80. **Singh B**, **Gupta RS**. Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol Genet Genomics* 2009;281:361–373.
81. **Chen S**, **Rudra B**, **Gupta RS**. Phylogenomics and molecular signatures support division of the order *Neisseriales* into emended families *Neisseriaceae* and *Chromobacteriaceae* and three new families *Aquaspirillaceae* fam. nov., *Chitinibacteraceae* fam. nov., and *Leeiaceae* fam. nov. *Syst Appl Microbiol* 2021;44:126251.
82. **Gao B**, **Sugiman-Marangos S**, **Junop MS**, **Gupta RS**. Structural and phylogenetic analysis of a conserved actinobacteria-specific protein (ASP1; SCO1997) from *Streptomyces coelicolor*. *BMC Struct Biol* 2009;9:40.
83. **Lorenzini E**, **Singer A**, **Singh B**, **Lam R**, **Skarina T**, ***et al***. Structure and protein-protein interaction studies on *Chlamydia trachomatis* protein CT670 (YscO Homolog). *J Bacteriol* 2010;192:2746–2756.
84. **Rossi E**, **La Rosa R**, **Bartell JA**, **Marvig RL**, **Haagensen JAJ**, ***et al***. *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat Rev Microbiol* 2021;19:331–342.
85. **Doudoroff M**, **Palleroni NJ**, **Genus I**. Pseudomonas Migula 1894, 237 Nom. cons. Opin. 5, Jud. Comm. 1952, 121. In: **Buchanan RE**, **Gibbons NE**, **Cowan ST**, **Holt JG** and **Liston J**, et al (eds). *Bergey's Manual of Determinative Bacteriology*, 8th edn. Baltimore, USA: The Williams & Wilkins Company; 1977. pp. 217–243.
86. **Palleroni NJ**. *Genus I.* Pseudomonas Migula 1894 237 Nom. Cons. Opin. 5. Jud. Comm. 1952, 121 . In: *Bergey's Manual of Systematic Bacteriology*. 2005. pp. 323-379
87. **Diggle SP**, **Whiteley M**. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. *Microbiology* 2020;166:30–33.

CHAPTER 5

Molecular Markers Specific for the *Pseudomonadaceae* Genera Provide Novel and Reliable Means for the Identification of Other *Pseudomonas* Strains/spp. Related to These Genera

This chapter highlights the application of CSIs for predicting the taxonomic affiliation of uncharacterized *Pseudomonas* species or strains. Using CSI information specific to different *Pseudomonadaceae* genera, this study assigns ~300 uncharacterized *Pseudomonas* strains to 14 *Pseudomonadaceae* genera. The proposed taxonomic placements are further validated through robust phylogenetic analysis. My contributions to this chapter include analyzing results from the CSI-based server (Appindels.com), conducting phylogenetic analyses to confirm taxonomic assignments, formatting and finalizing all main and supplemental figures and tables, and revising the manuscript.

Due to space limitations, supplementary materials for this study are not included in this chapter but are available along with the entire manuscript at:

Rudra, B., & Gupta, R. S. (2025). Genes, 16(2), 183.

genes

MDPI

Article

Molecular Markers Specific for the *Pseudomonadaceae* Genera Provide Novel and Reliable Means for the Identification of Other *Pseudomonas* Strains/spp. Related to These Genera

Bashudev Rudra and Radhey S. Gupta *

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada; rudrab@mcmaster.ca

* Correspondence: gupta@mcmaster.ca

Abstract: Background/Objectives: Taxon-specific conserved signature indels (CSIs) exhibit a strong predictive ability of being found in other members of specific taxa/genera. Recently, multiple exclusively shared CSIs were identified for several newly described *Pseudomonadaceae* genera (viz. *Aquipseudomonas*, *Atopomonas*, *Caenipseudomonas*, *Chryseomonas Ectopseudomonas*, *Geopseudomonas*, *Halopseudomonas*, *Metapseudomonas*, *Phytopseudomonas*, *Serpens*, *Stutzerimonas*, *Thiopseudomonas*, and *Zestomonas*). This study examines the potential applications of these CSIs for identifying other *Pseudomonas* spp. (strains) related to these genera. **Methods:** This work utilized the AppIndels.com server, which uses information regarding the presence of known taxon-specific CSIs in a genome for predicting its taxonomic affiliation. For this purpose, sequence information for different CSIs specific for the *Pseudomonadaceae* species/genera were added to the server's database. **Results:** The AppIndels server was used to predict the taxonomic affiliation of 1972 genomes of unclassified *Pseudomonas* spp. (strains/isolates). Based upon finding a significant number of CSIs matching a specific taxon, the AppIndels server made positive predictions regarding the taxonomic affiliation of 299 examined genomes into the following clades/genera: *Pseudomonas sensu stricto* clade (46), *Pseudomonas aeruginosa* (64), *Ectopseudomonas* (46), *Chryseomonas* (32), *Stutzerimonas* (31), *Metapseudomonas* (22), *Aquipseudomonas* (21), *Phytopseudomonas* (17), *Halopseudomonas* (9), *Geopseudomonas* (4), *Thiopseudomonas* (3), *Serpens* (2), and *Caenipseudomonas* and *Zestomonas* (1 each). Phylogenetic studies confirmed that the taxonomic predictions by the server were 100% accurate. **Conclusions:** Our results demonstrate that the CSIs specific for *Pseudomonadaceae* species/genera, in conjunction with the AppIndels server, provides a novel and useful tool for identifying other species/strains affiliated with these species/genera. Phylogenetic studies suggest that many examined *Pseudomonas* strains constitute novel species in the indicated genera.

Keywords: taxon-specific molecular markers; genomic sequences; phylogenetic analysis; Appindels.com server: prediction of taxonomic affiliation; unclassified *Pseudomonas* spp./strains; *Pseudomonadaceae genera*; *Pseudomonas aeruginosa*

Academic Editor: Jose María Requena

Received: 31 December 2024
Revised: 28 January 2025
Accepted: 31 January 2025
Published: 2 February 2025

Citation: Rudra, B.; Gupta, R.S. Molecular Markers Specific for the *Pseudomonadaceae* Genera Provide Novel and Reliable Means for the Identification of Other *Pseudomonas* Strains/spp. Related to These Genera. *Genes* **2025**, *16*, 183. https://doi.org/10.3390/genes16020183

1. Introduction

The family *Pseudomonadaceae* harbors several genera of which the genus *Pseudomonas* is one of the largest and earliest known prokaryotic genera [1,2]. The genus *Pseudomonas* encompasses >300 species representing more than 2/3rd of the *Pseudomonadaceae* species. Extensive earlier work on *Pseudomonas* species, using phylogenetic trees constructed based

on multiple different sets of genes/proteins, including core genomic proteins, has reliably established that the species from this genus do not form a monophyletic lineage. In phylogenetic trees, *Pseudomonas* species generally form three main groupings or lineages, referred to as the Pertucinogena, the Aeruginosa, and the Fluorescens lineages [3–10]. Additionally, species from both the Aeruginosa and Fluorescens lineages form multiple distinct genus-level clades, which are not specifically (i.e., evolutionarily) related to each other [5,9,11]. Species from other genera, including *Azomonas*, *Azotobacter*, *Chryseomonas*, *Entomomonas*, and *Thiopseudomonas*, branch in between these clades/lineages, demonstrating the polyphyletic nature of *Pseudomonas* species [3–7]. It is widely recognized that in accordance with the code governing the nomenclature of Prokaryotes [12], of the observed *Pseudomonas* species clades, only the species from the "Aeruginosa clade", which contains the type species *P. aeruginosa* of the genus *Pseudomonas*, should be recognized as the genus *Pseudomonas* [4,6,9,11,13–16].

It is important to note that the nomenclature type of the genus *Pseudomonas*, *P. aeruginosa*, is an important human pathogen capable of causing a wide array of life-threatening acute and chronic diseases [17,18]. However, this genus also includes some animals and plant pathogenic species, as well as other economically and ecologically significant species [19–23]. Additionally, species from this genus also produce several medically and agriculturally important compounds and a multitude of biologically active secondary metabolites [24–26]. Thus, it is of much importance to develop a reliable and informative classification scheme for *Pseudomonas* species, where different monophyletic groups of organisms are reliably demarcated and suitably named to distinguish them from each other. Naming different groups of organisms by distinct names indicates (implies) that all species bearing a specific genus name are more closely related to each other, and they commonly share several genotypic, phenotypic, and other properties (e.g., pathogenicity profile or potential), which differentiate them from other genera [15,27–30]. Distinct genus names also convey useful information about organisms, including their involvement in disease causation (i.e., risk group category), outbreaks, and diagnostic and treatment strategies [29,30]. Thus, taxonomy provides the central framework regarding our current understanding of microorganisms.

With the availability of genome sequences, extensive work has been carried out in the past few years to clarify the evolutionary relationships and classification of *Pseudomonas* species using multiple genome sequence-based approaches. The approaches used include the construction of phylogenetic trees based upon different large datasets of core genomic proteins [4–7,11,13,16,31] and an assessment of the overall relatedness of species from different clades based on genomic similarity matrices such as average nucleotide identity (ANIb) [4,16], average amino acid identity (AAI) [4,11], and the percentage of conserved proteins (POCP) [4,11,31]. In addition, analyses of genome sequences have also proven instrumental in the identification of highly specific molecular markers, such as conserved signature indels (CSIs) in genes/proteins, which are uniquely shared characteristics of species from different clades and afford unambiguous means for both distinguishing and the demarcation of different specific clades [5,11,32–34]. Based upon the consistent evidence acquired using different genomic approaches, most of the *Pseudomonas* species from the Pertucinogena and Aeruginosa lineages have now been reclassified into several novel genera (viz. *Aquipseudomonas*, *Atopomonas*, *Caenipseudomonas*, *Ectopseudomonas*, *Geopseudomonas*, *Halopseudomonas*, *Metapseudomonas*, *Phytopseudomonas*, *Stutzerimonas*, and *Zestomonas*) [4–6,11] and some preexisting genera (*Chryseomonas*, *Paraburkholderia*, *Serpens*, *Stenotrophomonas*, *Thiopseudomonas*, and *Xanthomonas*) [35,36]. Importantly, these studies have led to the identification of multiple highly specific molecular markers (i.e., CSIs) that are uniquely shared characteristics of the species noted above. Additionally, several

molecular markers have also been identified, which are exclusively found in the species from the genus *Pseudomonas sensu stricto*, *Azotobacter*, and *Azomonas* and for the species *P. aeruginosa*.

Due to the presence of *Pseudomonas*-related species in diverse niches and environments, including soil, water, and plant and animal tissues [10,37], and as its type species, *P. aeruginosa*, is an important human pathogen [17,18], species related to this genus are subjects of extensive studies and novel species and strains related to this genus are continually being discovered at a rapid pace [38]. Since 2022 alone, more than 100 novel species related to *Pseudomonas* are listed in the List of Prokaryotic Names with Standing in Nomenclature (LPSN) server [38]. However, in addition to the species with validly published names, the NCBI server holds genome sequences for >2000 uncharacterized *Pseudomonas* spp. (strains or isolates). Several of these uncharacterized strains/isolates will likely be identified as novel species. However, there is no information available at present regarding their taxonomic affiliation. In our earlier work on *Bacillus* related and other genera we have provided convincing evidence that the CSIs specific for different genera exhibit a high degree of predictive ability to be found in other members of these genera, and the presence of known taxon-specific CSIs in a genome sequence can be used to predict its taxonomic affiliation. The predictive abilities of the CSIs to be found in other related species form the basis of the recently developed AppIndels.com server, which based upon the presence of known taxon-specific CSIs in a submitted genome sequence can predict its taxonomic affiliation [39].

In this study, we have used the AppIndels.com server to determine whether based upon the information for the CSIs specific to different *Pseudomonadaceae* genera it can predict the phylogenetic/taxonomic affiliations of several of the unclassified *Pseudomonas* spp. (strains). The results of these studies presented here show that based upon the information for identified *Pseudomonadaceae* CSIs, the server was able to predict the taxonomic affiliation of 299 unclassified *Pseudomonas* strains/isolates into 14 *Pseudomonadaceae* clades/genera. Phylogenetic studies conducted on these strains show that the predictions made by the server regarding the taxonomic affiliations of these 299 strains were 100% accurate. Thus, the identified CSIs specific for the *Pseudomonadaceae* genera provide a novel and useful means for the identification of other novel or unclassified *Pseudomonas* species/strains related to these genera.

2. Materials and Methods

Analysis of Pseudomonas spp. Using the AppIndels Server

Sequence information for the CSIs specific to different *Pseudomonadaceae* clades/genera was added to the database of the AppIndels server (https://appindels.com/, accessed on 11 March 2024) [39]. Genome sequences for 2000 unclassified strains/isolates of *Pseudomonas* spp. were downloaded (in .faa format) from the NCBI Genome Database (https://www.ncbi.nlm.nih.gov/datasets/genome/, accessed on 1 June 2024) [39]) [40]. Details of these downloaded genomes, including their strain numbers, accession numbers, GC content, and genome sizes, are given in Supplementary Tables S1–S3. Of these genomes, some genomes that contained either <100 Kb sequence information or were indicated as contaminated were excluded from analyses. The remaining 1972 genomes were analyzed using the AppIndels server one at a time as indicated in earlier work and on the server's main page. The predictions made by the server regarding the taxonomic affiliation of the submitted sequence and the number of CSIs identified in it specific for the predicted genus were recorded.

A maximum-likelihood phylogenetic tree for the *Pseudomonas* spp. strains for which taxonomic assignments were made by the server, along with sequences of representative

species from different examined *Pseudomonadaceae* genera, was constructed based on the concatenated sequences for 118 conserved proteins comprising the phyloeco set for the class Gammaproteobacteria [41]. The tree was constructed using an internally developed pipeline, as described in our recent work [5,34]. The tree was labeled and formatted using MEGA X [42].

3. Results

3.1. Predictive Ability of a CSI Specific for the Genus Halopseudomonas

Earlier work on CSIs specific for multiple prokaryotic taxa provides compelling evidence that these molecular characteristics exhibit a high degree of predictive ability to be found in other species related to a specific taxon. To illustrate, in Figure 1, we show the results for a CSI specific for the genus *Halopseudomonas* [5]. This genus was created in 2021 by the reclassification of *Pseudomonas* species, which corresponded to the Pertucinogena lineage. More than 20 CSIs specific to the genus *Halopseudomonas* were identified in this earlier study, and the example depicted in Figure 1 shows the results for one of these CSIs, where a 2 aa insert in a conserved region of the flagellar protein FlgN was present exclusively in all 19 *Pseudomonas* species that corresponded to the genus *Halopseudomonas*.

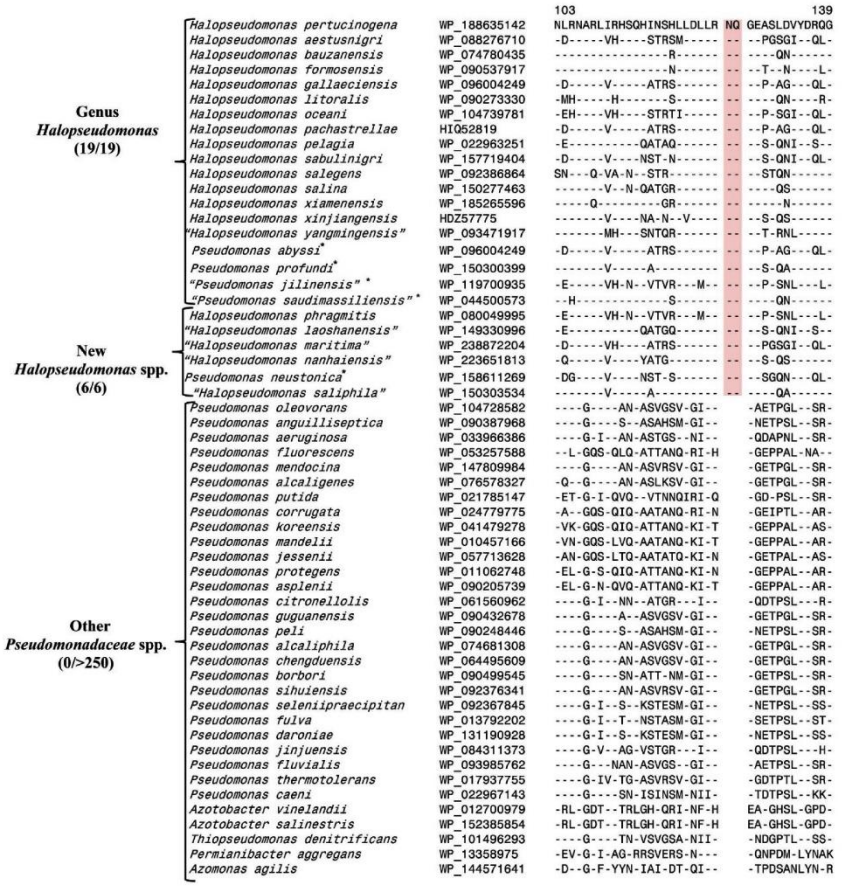

Figure 1. Partial sequence alignment showing a two amino acid insertion (CSI) in the flagellar FlgN protein (highlighted in pink color) described in our earlier work [5], which is specific for the genus *Halopseudomonas*. Sequences for six new *Halopseudomonas*-related species have since become available,

and all of them share this CSI, demonstrating the predictive ability of this CSI. The species marked with the symbol * have not yet been reclassified as *Halopseudomonas* due to the lack of availability of type strains in two different culture collections, or some species are listed in the LPSN under the genus *Neopseudomonas*, which is a synonym of *Halopseudomonas* [38]. Quotation marks " " surrounding a species name indicates that this name is not yet validly published. The dashes (-) in the alignment indicate identity with the amino acids on the top line. Accession numbers for different sequences are indicated in the second column, and the numbers at the top indicate the position of this sequence fragment within the protein sequences.

Two of these species in this figure are listed as "*Pseudomonas*" as they have not yet been reclassified as *Halopseudomonas* due to the lack of availability of the type strains in two different culture collections. Since the publication of this work, six other species related to *Halopseudomonas* have been described [4,6,43,44]. Some of these species presently are either not validly published (indicated by their placement within " ") or they are misclassified into the genus *Neopseudomonas* [6], which is a homotypic synonym of *Halopseudomonas* [38]. Nonetheless, as shown in Figure 1, the 2 aa CSI specific for the *Halopseudomonas* is commonly and uniquely shared by all six newly described species related to *Halopseudomonas*, but it is not found in any other *Pseudomonadaceae* species. In a phylogenetic tree that we have constructed, all species share this CSI group reliably within a clade corresponding to the genus *Halopseudomonas* (Figure S1). These results provide further evidence supporting the predictive ability of taxon-specific CSIs to be found in other species/strains that are related to them.

3.2. Examining the Usefulness of the CSIs Specific for the Pseudomonadaceae Genera for Determining the Taxonomic Affiliation of Unclassified Pseudomonas spp. Using the AppIndels.com Server

As indicated earlier, in addition to the genomes for >300 *Pseudomonas* species with validly published names, the NCBI database also holds genome sequences for >2000 unclassified strains/isolates of *Pseudomonas* spp. An earlier study by Hess et al. [7] provides evidence that these unclassified strains encompass enormous genetic diversity, which remains to be understood. Thus, it is important to develop novel means or tools by which the genetic diversity and taxonomic affiliation of these unclassified strains could be assessed. In this work, we have investigated whether the identified CSIs specific to several *Pseudomonadaceae* genera can be used for identifying unclassified *Pseudomonas* spp./strains that are related to these genera. These analyses were carried out using the AppIndels.com server, which has been specifically created to take advantage of the predictive abilities of the known taxon-specific CSIs, to identify other species/strains related to them. The working of the AppIndels.com server has been described in detail elsewhere [39], but it is briefly explained below.

The AppIndels.com server is a web-based tool that uses sequence information for validated CSIs specific for known prokaryotic taxa for determining the presence of these molecular characteristics in any input genome sequence [39]. Based on the taxon-specificities of the CSIs in the server's database and their predictive ability to be found in other members of these taxa, if the server identifies that the input genome sequence contains significant numbers of CSIs matching a specific taxon, it predicts that the analyzed genome (strain/species) is affiliated with that taxon. The AppIndels server database presently contains sequence information for >1000 previously identified CSIs specific to different (>100) prokaryotic genera [39]. We have now added to this database the sequence information for different identified CSIs specific to the *Pseudomonadaceae* genera [5,11]. In Table 1, we have provided information regarding the *Pseudomonadaceae* genera/taxa for which the CSIs have been identified and the numbers of CSIs, which are specific for each of these genera or

taxa. This list also includes several CSIs that are specific for the species *P. aeruginosa* and *P. paraeruginosa* [45].

Table 1. List of *Pseudomonadaceae* Genera for which CSIs have been identified.

Genera/Species Name	No. of Identified CSIs	Weight Value of Each CSI
Aquipseudomonas	6	0.4
Atopomonas	22	0.2
Azomonas	5	0.4
Azotobacter	10	0.4
Caenipseudomonas	8	0.4
Chryseomonas	11	0.3
Ectopseudomonas	5	0.4
Geopseudomonas	15	0.3
Halopseudomonas	24	0.2
Metapseudomonas	5	0.4
Phytopseudomonas	12	0.3
Pseudomonas sensu stricto	6	0.4
Serpens	3	0.5
Stutzerimonas	7	0.4
Thiopseudomonas	6	0.3
Zestomonas	5	0.4
P. aeruginosa	7	0.3
P. paraeruginosa	5	0.4

The last column in Table 1 indicates the weight values given to individual CSIs from different taxa. The rationale of giving weight values to different CSIs is discussed in detail in earlier work [39]. However, its main purpose is to increase the specificity of taxon prediction by the AppIndels server by requiring that multiple CSIs specific to a given taxon be present before a positive identification is made. When a genome sequence is uploaded or submitted to the AppIndels.com server, it conducts BLASTp searches on the submitted genome against the sequences of all CSIs in its database. Based on these searches, the server identifies matching sequences in the submitted genome where the indels of specific lengths are present in protein sequences in the exact location as present in the protein sequences in the CSI database. The server then gathers information regarding the taxon specificities of different matching CSIs. If the combined weight of all CSIs matching a specific taxon exceeds the threshold value of 1.0, the server makes a positive prediction that the submitted genome is affiliated with the indicated taxon. As all CSIs specific for the *Pseudomonadaceae* genera/clades have a weight value of 0.4 or less (Table 1), the server will make a positive identification for any *Pseudomonadaceae* genus/clade only when three or more CSIs matching that taxon are found in the submitted genome. As all described CSIs for the *Pseudomonadaceae* species/genera exhibit a high degree of specificity for the indicated taxon (barring an isolated exception) [5,11,45], the possibility of finding three CSIs matching a specific genus/taxon in the genome of an unrelated species/strain is considered highly unlikely.

To test the usefulness of identified CSIs using the AppIndels server, genome sequences were downloaded for 2000 stains/isolates of *Pseudomonas* spp. from the NCBI genome database. Of these, 28 genomes, where the genome sequence consisted of <100 Kb, or was indicated as contaminated, were not further analyzed. Of the remaining 1972 genomes, 266 genomes were chromosomes or complete (Table S1), 1197 consisted of contigs (Table S2), and 509 were scaffolds (Table S3). Some information regarding these genomes, including their strain numbers, accession numbers, assembly stage, G-C content (mol%), and genome sizes, is provided in the Supplementary Material (Tables S1–S3). The analyses on these

genomes were conducted using the AppIndels server by uploading the sequences of these genomes, one at a time, onto the server. The server checks the uploaded genome sequence for the presence of CSIs matching different taxa in its database. If the server identifies significant numbers of CSIs matching any specific taxon, then the result from the server shows a positive match to that taxon. In such cases, the server also provides information regarding the number of CSIs matching the predicted taxon. However, if the submitted genome corresponds to a taxon/genus for which no CSIs are present in the server or if the total weight of the identified CSIs is less than the threshold value of 1.0, then the server shows a negative "None" result.

In Figure 2, we show the results obtained from the server for two *Pseudomonas* strains/isolates. The server indicates that the strain ZM24 is related to the *Pseudomonas sensu stricto* clade, and its genome contained five CSIs specific for this clade (Figure 2A). On the other hand, the server predicted that the genome of strain ABC1 is related to the genus *Stutzerimonas*, and its genome contained six CSIs specific for this genus (Figure 2B). In addition to indicating the numbers of CSIs specific to the predicted taxon, the server also provides sequence information for all matching CSIs, which can be viewed upon clicking the down arrow beside the number of CSIs.

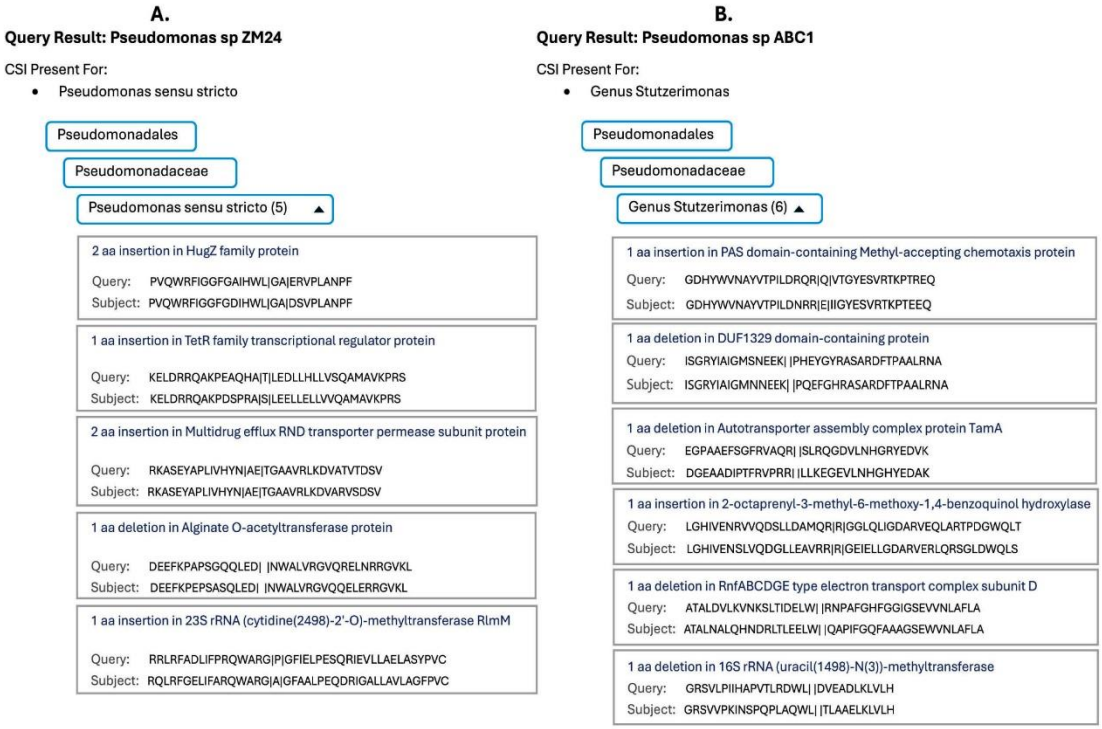

Figure 2. The results from the AppIndels server for the genome sequences of two representative unclassified *Pseudomonas* spp./strains. (**A**) The genome of *Pseudomonas* strain ZM24 is predicted by the server as affiliated with the *Pseudomonas sensu stricto* clade, and it contained five CSIs specific for this clade. (**B**) The *Pseudomonas* strain ABC1 was identified by the server as belonging to the genus *Stutzerimonas*, and it contained six CSIs specific to this genus.

Based on the analysis of genome sequences for 1972 examined *Pseudomonas* strains/isolates, the server made specific predictions regarding the taxonomic affiliations of 299 of the examined genomes to specific *Pseudomonadaceae* genera. The results from the server

for the genomes of all 299 *Pseudomonas* strains/isolates for which specific predictions were made are shown in Table S4, and a summary of these results is presented in Table 2. In Table 2, we have organized the results from the server for different strains according to their predicted affiliation for the *Pseudomonadaceae* genera. Table 2 also shows the numbers of CSIs (range) specific for the indicated genus/species, which were identified in the analyzed genomes.

Table 2. The results from the AppIndels Server regarding the taxonomic affiliations of the genome sequences of 299 unclassified *Pseudomonas* spp.

Genera/Species	No. of Strains	Range of CSIs	*Pseudomonas* spp. Strain Nos.
Pseudomonas sensu stricto	46	5–6	21, 273, 30_B, AAC, ADPe, ATCC 13867, AU11447, AU12215, BJa5, EGD-AKN5, GCEP-101, GD03691, GD03903, GD04087, HMSC75E02, HS-18, LA21, M1, NBRC 111135, NBRC100443, PDM17, PDM18, PDM19, PDM20, PDM21, PDM22, PDM23, PDM33, PDNC002, PI1, PSE14, R3.Fl, RW407, SCB32, UMA601, UMA603, UMA643, UMC3103, UMC3106, UMC3129, UMC631, UMC76, UME83, ZM23, ZM24, ZM25.
P. aeruginosa	64	5–7	203-8, 17023526, 17023671, 17033095, 17053182, 17053418, 17053703, 17063399, 17072548, 17073326, 17102422, 17103552, 17104299, 18073667, 18082547, 18081308, 18082551, 18082574, 18083194, 18083202, 18083259, 18083286, 18084127, 18092229, 18093371, 18101001-2, 18102011, 18103014, 18113298, 19062259, 19064969, 19072337-2, 19082381, 2VD, 3PA37B6, AF1, AFW1, AK6U, B111, BDPW, BIS, BIS1, CP-1, FDAARGOS_761, HMSC057H01, HMSC072F09, HMSC16B01, HMSC076A11, HMSC060F12, HMSC065H01, HMSC066A08, HMSC065H02, HMSC067G02, HMSC063H08, HMSC058C05, P179, P20, P22, PAH14, Pseudomonas_assembly, PS1(2021), RGIG3665, S33, S68.
Aquipseudomonas	21	4–6	8AS, BLCC-B13, BMS12, F(2018), GD03869, GD03875, GD03985, GD04015, GD04019, GD04042, GD04045, GOM6, J452, L-22-4S-12, ML96, PDM15, PDM16, R-28-1W-6, UBA6718, SO81,WS 5013.
Caenipseudomonas	1	7	Go_SlPrim_bin_81
Chryseomonas	32	6–11	313, AS2.8, BAV 2493, BAV 4579, GM_Psu_1, GM_Psu_2, HUK17, LTJR-52, MAG002Y, PS02302, RIT 411, S1C77_SP397, S2C3242, SP152, SP29, SP3, SP403, SP421, WAC2, HPB0071, Snoq117.2, MS15, JUb52, EpSL25, PLB05, HR1, CBMAI 2609,UBA6549, UBA7233, UBA3149, UBA4102.
Ectopseudomonas	46	3–5	297, 07-Jan, 905_Psudmo1, AA-38, ALS1131, ALS1279, AOB-7, B11D7D, BMW13, DS1.001, EGD-AK9, EggHat1, GD03721, GD03722, GD03919, GD04158, GOM7, GV_Bin_12, Gw_UH_bin_155, HS-2, KB-10, KHPS1, LPH1, Leaf83, MDMC17, MDMC216, MDMC224, MSPm1, Marseille-Q0931, NCCP-436, NFACC19-2, NFPP33, o96, OA3, P818, 8O, 8Z, REST10, RGIG627, THAF187a, THAF42, WS 5019, YY-1, Z8(2022), ZH-FAD, phDV1.
Geopseudomonas	4	4–15	A-1, OF001, R2F_R2FSRR_metabat.60, Gw_Prim_bin_4.
Halopseudomonas	9	20–24	5Ae-yellow, FME51, MYb185, NORP239, NORP330, OIL-1, SSM44, WN033, gcc21.

Table 2. *Cont.*

Genera/Species	No. of Strains	Range of CSIs	*Pseudomonas* spp. Strain Nos.
Metapseudomonas	22	3–5	57B-090624, 1D4, A46, BN102, BN411, BN414, BN415, BN417, BN515, BN606, D(2018), DY-1, ENNP23, FeS53a, JG-B, JM0905a, LFM046, PDM13, Pc102, Q1-7, SLBN-26, TCU-HL1.
Phytopseudomonas	17	9–12	AG1028, Bi70, BIGb0408, CrR14, CNPSo 3701, MEJ086, MM211, PDM11, PDM12, S2C11432_SP223, S2C78296_SP133, sia0905, SP200_1_metabat2_genome_mining.44, SP236_1_metabat2_genome_mining.8, PA1, PA15, PA27.
Serpens	2	3	N24CT, RL.
Stutzerimonas	31	4–7	10B238, 9Ag, A192_concoct.bin.7, ABC1, ALOHA_A2.5_105, BAY1663, BRH_c35, C42_metabat.bin.8, Choline-3u-10, DF_1_3.23, DNDY-54, IC_126, JI-2, KSR10, M30B71, MCMED-G45, MT-1, MT4, MTM4, N17CT, NP21570, Q2-TVG4-2, RS261_metabat.bin.8, S5(2021), SCT, SST3, TTU2014-066ASC, TTU2014-096BSC, TTU2014-105ASC, WS 5018, s199.
Thiopseudomonas	3	4–5	AS08sgBPME_395, C27(2019), SO_2017_LW2 bin 68.
Zestomonas	1	3	LS44

As seen from Table 2, in all cases, the predicted affiliation of any genome to a specific *Pseudomonadaceae* species/genera is based on the shared presence of a minimum of three CSIs specific to that taxon. The numbers of CSIs identified for different genera (or species) in the analyzed genomes varied from a low of 3 for the genus *Serpens* to more than 20 for *Halopseudomonas*. This variation is solely due to the differences in the number of CSIs that have been identified for different genera (see Table 1) [5,11,45]. The results presented in Table 2 show that of the genomes for which the server made specific predictions, about 20% corresponded to the species *P. aeruginosa*. Other *Pseudomonadaceae* genera to which large numbers of analyzed strains (genomes) belonged included the *Pseudomonas sensu stricto* clade (46 strains), *Ectopseudomonas* (46 strains), *Chryseomonas* (32 strains), *Stutzerimonas* (31 strains), *Metapseudomonas* (22 strains), *Aquipseudomonas* (21 strains), *Phytopseudomonas* (17 strains), *Halopseudomonas* (9 strains), and *Geopseudomonas* (4 strains). The server also predicted that a limited number of strains are affiliated with the genera *Caenipseudomonas*, *Serpens*, *Thiopseudomonas*, and *Zestomonas*, which consist of only a few species [5,11].

We have examined the reliability of taxon predictions by the server by constructing a phylogenomic tree based on genome sequences of different *Pseudomonas* strains for which the server made taxonomic predictions. This tree was constructed based on concatenated sequences of 118 conserved proteins (corresponding to the phyloeco set for the class *Gammaproteobacteria*), and it also included the sequences of representative species from relevant *Pseudomonadaceae* genera. We show the results from this tree in Figure 3. Due to the considerable number of strains in this tree, we have compressed the clades for some *Pseudomonadaceae* genera in Figure 3. However, the uncompressed results for these clades are presented in Figure 4. In the phylogenetic trees shown in Figures 3 and 4, all *Pseudomonas* strains/isolates for which the server made taxonomic predictions grouped reliably (100% concordance) with the other species from the indicated genera (Figures 3, 4 and S2). Based upon the branching of different *Pseudomonas* strains/isolates in Figures 3 and 4, while many unclassified strains are closely related to the known species, several other strains branched

distinctly from the known species. Thus, many of these strains may constitute novel species within the indicated genera.

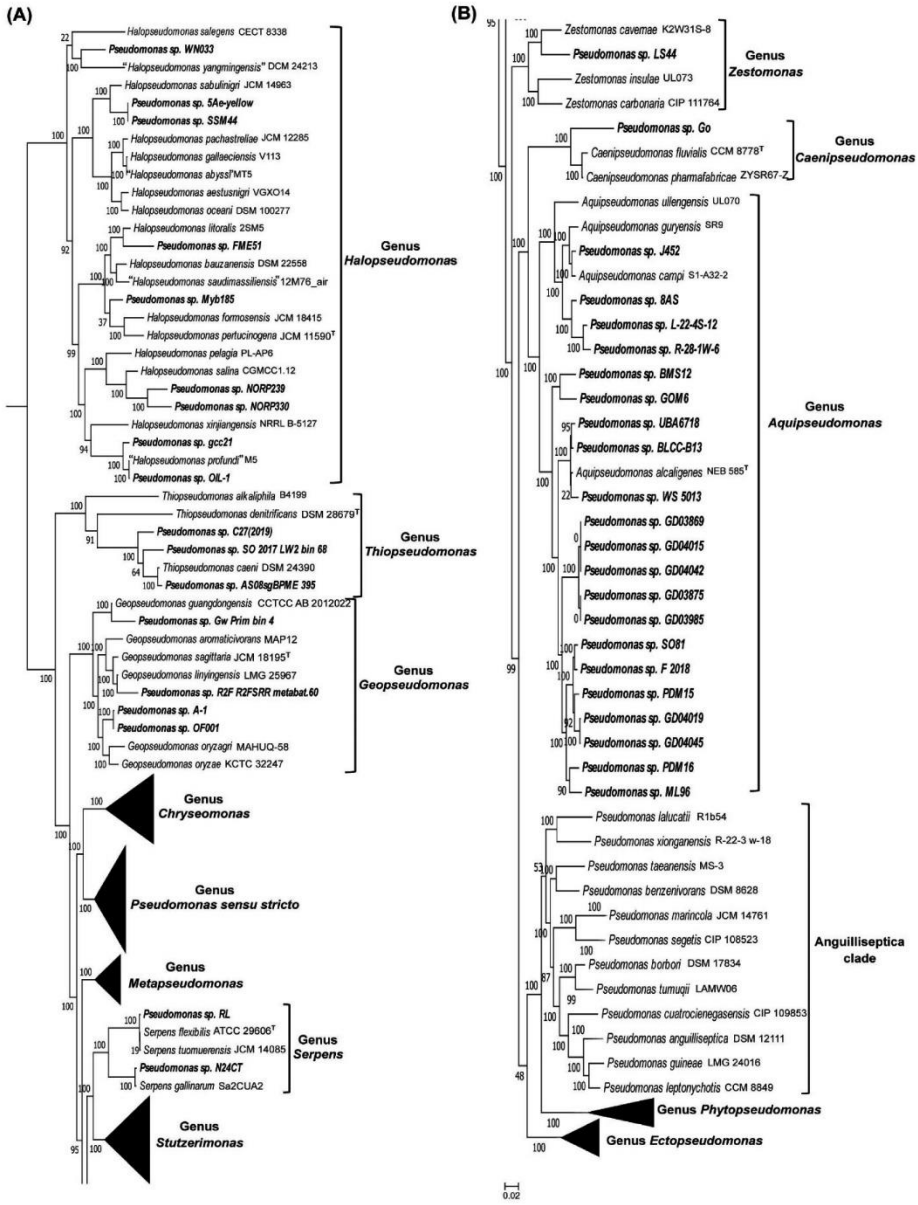

Figure 3. A phylogenetic tree based on genome sequences for the representative species, including type species of different *Pseudomonadaceae* genera and genomes of different *Pseudomonas* spp. (strains/isolates) for which positive predictions were made by the server regarding affiliation with specific clades/genera (Tables 2 and S4). For the ease of visualization of information for different strains, the clades for some genera, viz. *Chryseomonas*, *Ectopseudomonas*, *Metapseudomonas*, *Phytopseudomonas*, *Pseudomonas sensu stricto*, and *Stutzerimonas*, are compressed in this figure. The figure is shown in two parts (**A**,**B**), and part B is a continuation of (**A**). All *Pseudomonas* strains for which the server made taxonomic predictions branched with 100% accuracy with the indicated genera in these trees.

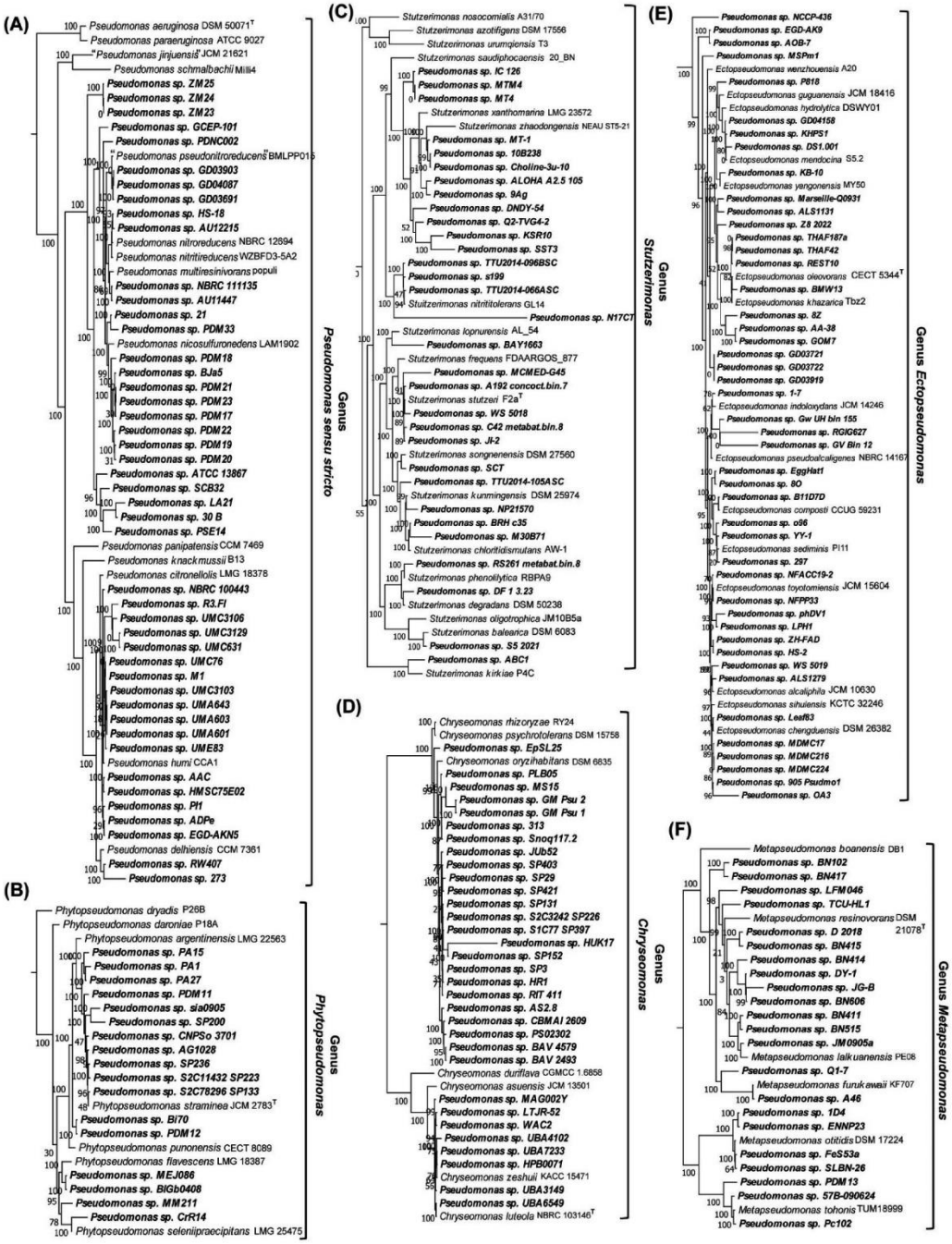

Figure 4. Phylogenetic branching of *Pseudomonas* spp. (strains/isolates), which based upon the results obtained from AppIndels server (Tables 2 and S4) were predicted to be related to the genera (**A**) *Pseudomonas sensu stricto*, (**B**) *Phytopseudomonas*, (**C**) *Stutzerimonas*, (**D**) *Chryseomonas*, (**E**) *Ectopseudomonas*, and (**F**) *Metapseudomonas*. All strains for which the server made taxonomic predictions branched with 100% accuracy with the indicated genera in this tree.

4. Discussion

Members of the genus *Pseudomonas*, which are genetically and evolutionarily highly diverse, are widely distributed in different environments. This group includes species that are opportunistic pathogens of humans, animals, and plants and other species of economic and ecological significance [17–19,46,47]. For example, the type species of this genus, *P. aeruginosa*, which is one of the most researched species, is an opportunistic multidrug-resistant human pathogen capable of infecting multiple tissues, especially in individuals with weakened immune systems, and is often responsible for serious illnesses, such as ventilator-associated pneumonia and several sepsis syndromes [48–50]. *P. aeruginosa* infections in patients with cystic fibrosis cause significant economic burden in the health care industry [18]. Due to its resistance to different antibiotics, the World Health Organization recognizes *P. aeruginosa* as one of the six important pathogens posing greatest threats to humans in terms of antibiotic resistance [50,51]. On the other hand, some *Pseudomonas* species, such as *P. syringae*, are pathogenic to plants [19], whereas other species, such as *P. fluorescens*, are beneficial to plants and have been used in the agriculture industry for sustainable plant growth as well as disease management [24]. Several other *Pseudomonas* species have found significant roles as biocontrol agents, as bioremediation agents, as detectors of food spoilage agents in milk [25], and in the degradation of anthropogenic pollutants [26].

In view of the importance of these species from clinical and other perspectives, this group of species is extensively studied, and it constitutes one of the fastest growing groups of bacteria [4]. In recent years, extensive work using genomic approaches has been carried out to more reliably delineate the evolutionary relationships and classification scheme for *Pseudomonas* and related species. These studies have led to the reclassification of >150 *Pseudomonas* species into 14 novel genera [5,11]. Members of all these newly described genera can be reliably distinguished from each other based upon multiple highly specific molecular markers (i.e., CSIs) that are uniquely shared characteristics of the species from these genera. Similarly, the clade corresponding to the genus *Pseudomonas sensu stricto*, which harbors *P. aeruginosa*, can also be reliably distinguished from all other *Pseudomonas* based on multiple exclusively shared CSIs [11]. However, the genetic diversity of *Pseudomonas* extends far beyond the known species (>300) with validly published names. The NCBI [40] harbors genomes for >2000 uncharacterized strains/isolates of *Pseudomonas* species, for which no information is available regarding their phylogenetic affiliation. As these uncharacterized strains are likely to harbor many novel species related to both the known *Pseudomonadaceae* genera as well as other novel taxa related to these bacteria [6], it is important to characterize them. However, there is no easy-to-use methods available for dependably identifying strains that are related to the existing *Pseudomonadaceae* genera.

Therefore, the objective of this study was to determine whether the CSIs specific for different *Pseudomonadaceae* genera, due to their known predictive ability to be found in other group members, can be used to identify other unclassified *Pseudomonas* (spp.) strains that are related to these genera. These investigations were greatly facilitated by the recent development of AppIndels.com server, which based upon the presence of known taxon-specific CSIs in a genome sequence, can predict its taxonomic affiliation [39]. In this work, the AppIndels server was used after supplementing its database with the sequence information for different CSIs specific to the *Pseudomonadaceae* genera for predicting the taxonomic affiliations of genome sequences for 1972 unclassified *Pseudomonas* strains/isolates. The results presented here show that based upon the identified CSIs for the *Pseudomonadaceae* genera, the server was able to predict the taxonomic affiliation of 299 of these unclassified *Pseudomonas* strains into 14 distinct clades of *Pseudomonadaceae* species/genera. The genera or species groups into which these unclassi-

fied *Pseudomonas* strains/isolates were assigned included the *Pseudomonas sensu stricto* clade (46 strains), *Ectopseudomonas* (46 strains), *Chryseomonas* (32 strains), *Stutzerimonas* (31 strains), *Metapseudomonas* (22 strains), *Aquipseudomonas* (21 strains), *Phytopseudomonas* (17 strains), *Halopseudomonas* (9 strains), *Geopseudomonas* (4 strains), *Thiopseudomonas* (3 strains), *Serpens* (2 strains), *Caenipseudomonas* (1 strain), and *Zestomonas* (1 strain). In addition, 64 *Pseudomonas* strains/isolates were identified as *P. aeruginosa*. In all cases, the assignment of *Pseudomonas* strains to different *Pseudomonadaceae* genera (or to *P. aeruginosa*) was based on the shared presence of multiple (minimum 3) CSIs, which are exclusive characteristics of the indicated genera. The results of phylogenetic studies conducted here confirm that the taxonomic predictions made by the server were 100% in agreement with the branching of these strains with the species from the indicated genera. These results provide further strong evidence regarding (i) the predictive abilities of the taxon-specific CSIs to be found in other (unclassified) members of these taxa and (ii) the conclusion that the use of these molecular markers provides a novel and trustworthy means for the identification of other species/strains related to these genera [39].

Although the AppIndels server accurately predicted the taxonomic affiliations of 299 *Pseudomonas* strains, it provided no results for the remainder of the strains. This is not surprising because the AppIndels server can make taxonomic predictions for only those strains that are related to the taxa for which CSIs are known and present in the server's database [39]. As noted previously, the genus *Pseudomonas* is a very large and diverse grouping of microorganisms, harboring >300 validly named species that form multiple distinct clades/lineages [4,6–9,11,16]. Thus far, CSIs have been identified for only a limited number of these groupings, consisting mainly of the genus *Halopseudomonas* and some clades/genera within the Aeruginosa lineage. However, a vast majority of the *Pseudomonas* species, representing more than two-thirds of named species, are part of the Fluorescens lineage, which is composed of multiple distinct genus-level clades and subclades [7,9,13,16,31,37]. No CSIs are known at present for the species from different clades and subclades of the Fluorescens lineage. In addition, no CSIs have been identified for the Anguilliseptica clade of species and many other species within the Aeruginosa lineage (viz. *P. benzenivorans*, *P. cuatrocienegasensis*, *P. indica*, *P. kuykendallii*, *P. lalucatii*, *P. mangiferae*, *P. mangrovi*, *P. matsuisoli*, and *P. pohangensis*), which branch distinctly from the described clades/genera. In view of the paucity of CSIs for these other groups/clades of *Pseudomonas* species, if an examined strain (genome) is affiliated with these species clades/genera, the server will not be able to make any taxonomic predictions for those strains. Therefore, as indicated on the server's website, while the absence of any taxonomic prediction by the server is not very informative, a specific prediction by the server regarding taxonomic affiliation is a highly trustworthy result.

It should be noted that the genomes of *Pseudomonas* spp./strains for which the server was able to make correct taxonomic predictions consisted of different assembly stages ranging from chromosome and complete to contigs and scaffolds (see Tables S1–S3). Previously, we have also shown that based on genome sequence information, the server can also predict the taxonomic affiliation of uncultured strains/isolates [52]. These results and observations indicate that the AppIndels server provides a valuable and easy-to-use tool for the identification and taxonomic characterization of cultured and uncultured strains/isolates for the species/genera for which CSIs are known. Based upon the phylogenetic branching of *Pseudomonas* spp./strains for which taxonomic predictions were made by the server, several of these strains branched distinctly from the other known species within these genera (Figures 3 and 4). Thus, upon further characterization, a number of these strains would likely constitute novel species within these genera. This should lead to a considerable increase in

the genetic diversity of species within these genera advancing our understanding of the *Pseudomonas*-related species/genera.

It should be noted that the AppIndels server, in addition to its demonstrated utility for predicting the taxonomic affiliation of any genome-sequenced strains/isolate, also provides a novel and useful diagnostic tool. Amongst the *Pseudomonas* species, *P. aeruginosa* is of particular significance, as it can cause numerous life-threatening diseases in humans [17,18]. Hence, an accurate identification of this species from other closely related species is of considerable importance in clinical settings. The results presented here show that based upon the identified CSIs, the server can reliably distinguish *P. aeruginosa* from all other *Pseudomonas*-related species, including other species from the *Pseudomonas sensu stricto* clade. Of particular importance is the fact that the server can also reliably distinguish *P. aeruginosa* from *P. paraeruginosa*. The latter species was recently created from *P. aeruginosa* by the transfer of several strains, which lacked the Type III secretion system (i.e., differing in terms of pathogenicity) and produced various biosurfactants [53], into this new species [45]. However, *P. paraeruginosa* is genetically closely related to *P. aeruginosa*, and it is difficult to distinguish between these two species with the most available diagnostic methods [45]. However, the AppIndels server provides a rapid and easy-to-use method to reliably detect the presence of *P. aeruginosa* based on genome sequence information. Additionally, based on genome sequence information, the server can also rapidly and reliably detect the presence of any species/strains related to the *Pseudomonas sensu stricto* clade, which due to being part of this monophyletic clade may share its pathogenicity traits for humans [29]. Similarly, based upon the CSIs specific for other species/genera, the server can also reliably detect the presence of other related species based on genome sequence information.

Lastly, based upon earlier work on CSIs in genes/proteins sequences, these molecular characteristics, in addition to their specificity and predictive abilities for reliable identification of species from different clades, also play important/essential functions in the group of organisms for which they are specific [11,33,54–57]. Hence, genetic, biochemical, and functional studies on the CSIs specific for different genera provide means for the identification of novel biochemical and other characteristics that are specific to these organisms.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/genes16020183/s1, Table S1. List of 266 downloaded *Pseudomonas* spp. genomes (chromosome and complete) for analysis in this study. Table S2. List of 1197 downloaded *Pseudomonas* spp. genomes (contigs) to analyze in this study. Table S3. List of 510 downloaded *Pseudomonas* spp. genomes (scaffold) to analyze in this study. Table S4. Information on the genome sequences of 299 uncharacterized *Pseudomonas* spp. whose taxonomic affiliations were predicted by the AppIndels web server. Figure S1. A maximum-likelihood tree, constructed using concatenated sequences of 118 conserved proteins, depicts the branching of all newly identified species or species with new name combinations that share the CSIs specific to the genus *Halopseudomonas*. Newly described species are highlighted in bold, and non-validly published species are shown within " ". Figure S2. A maximum-likelihood tree based on concatenated sequences of 118 conserved proteins showing all strains from the Aeruginosa clade (Genus *Pseudomonas sensu stricto*).

Author Contributions: B.R. and R.S.G. carried out analysis using the AppIndels server; B.R. constructed phylogenetic trees; R.S.G. planned and supervised the work, obtained funding for the project, and wrote and finalized the manuscript; B.R. updated the sequence information for the CSIs and checked and formatted different figures and tables, R.S.G. and B.R. wrote and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a research grant (RGPIN-2019-06397) from the Natural Science and Engineering Research Council of Canada awarded to Radhey S. Gupta.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Genome sequences for different *Pseudomonas* spp./strains, whose accession numbers are given in Supplementary Tables S1–S3, were downloaded from the NCBI Genome Database (https://www.ncbi.nlm.nih.gov/datasets/genome/, accessed on 30 January 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Migula, W. Uber ein neues System der Bakterien. *Arb. Bakt. Inst. Kar1sruhe* **1894**, *1*, 238.
2. Skerman, V.B.D.; McGowan, V.; Sneath, P.H.A. Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* **1980**, *30*, 225–420. [CrossRef]
3. Jun, S.R.; Wassenaar, T.M.; Nookaew, I.; Hauser, L.; Wanchai, V.; Land, M.; Timm, C.M.; Lu, T.Y.; Schadt, C.W.; Doktycz, M.J.; et al. Diversity of *Pseudomonas* Genomes, Including Populus-Associated Isolates, as Revealed by Comparative Genome Analysis. *Appl. Environ. Microbiol.* **2016**, *82*, 375–383. [CrossRef] [PubMed]
4. Lalucat, J.; Gomila, M.; Mulet, M.; Zaruma, A.; Garcia-Valdes, E. Past, present and future of the boundaries of the *Pseudomonas* genus: Proposal of *Stutzerimonas* gen. Nov. *Syst. Appl. Microbiol.* **2022**, *45*, 126289. [CrossRef] [PubMed]
5. Rudra, B.; Gupta, R.S. Phylogenomic and comparative genomic analyses of species of the family *Pseudomonadaceae*: Proposals for the genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., merger of the genus *Oblitimonas* with the genus *Thiopseudomonas*, and transfer of some misclassified species of the genus *Pseudomonas* into other genera. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 005011. [CrossRef]
6. Saati-Santamaria, Z.; Peral-Aranega, E.; Velazquez, E.; Rivas, R.; Garcia-Fraile, P. Phylogenomic Analyses of the Genus *Pseudomonas* Lead to the Rearrangement of Several Species and the Definition of New Genera. *Biology* **2021**, *10*, 782. [CrossRef] [PubMed]
7. Hesse, C.; Schulz, F.; Bull, C.T.; Shaffer, B.T.; Yan, Q.; Shapiro, N.; Hassan, K.A.; Varghese, N.; Elbourne, L.D.H.; Paulsen, I.T.; et al. Genome-based evolutionary history of *Pseudomonas* spp. *Environ. Microbiol.* **2018**, *20*, 2142–2159. [CrossRef] [PubMed]
8. Gomila, M.; Pena, A.; Mulet, M.; Lalucat, J.; Garcia-Valdes, E. Phylogenomics and systematics in *Pseudomonas*. *Front. Microbiol.* **2016**, *6*, 214. [CrossRef] [PubMed]
9. Peix, A.; Ramirez-Bahena, M.H.; Velazquez, E. The current status on the taxonomy of *Pseudomonas* revisited: An update. *Infect. Genet. Evol.* **2018**, *57*, 106–116. [CrossRef] [PubMed]
10. Peix, A.; Ramírez-Bahena, M.-H.; Velázquez, E. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect. Genet. Evol.* **2009**, *9*, 1132–1147. [CrossRef]
11. Rudra, B.; Gupta, R.S. Phylogenomics studies and molecular markers reliably demarcate genus *Pseudomonas sensu stricto* and twelve other *Pseudomonadaceae* species clades representing novel and emended genera. *Front. Microbiol.* **2024**, *14*, 1273665. [CrossRef] [PubMed]
12. Oren, A.; Arahal, D.R.; Goker, M.; Moore, E.R.B.; Rossello-Mora, R.; Sutcliffe, I.C. International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). *Int. J. Syst. Evol. Microbiol.* **2023**, *73*, 005585. [CrossRef]
13. Lalucat, J.; Mulet, M.; Gomila, M.; Garcia-Valdes, E. Genomics in Bacterial Taxonomy: Impact on the Genus *Pseudomonas*. *Genes* **2020**, *11*, 139. [CrossRef] [PubMed]
14. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **2018**, *39*, 996–1004. [CrossRef] [PubMed]
15. Hugenholtz, P.; Chuvochina, M.; Oren, A.; Parks, D.H.; Soo, R.M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **2021**, *15*, 1879–1892. [CrossRef] [PubMed]
16. Girard, L.; Lood, C.; Hofte, M.; Vandamme, P.; Rokni-Zadeh, H.; van Noort, V.; Lavigne, R.; De Mot, R. The Ever-Expanding *Pseudomonas* Genus: Description of 43 New Species and Partition of the *Pseudomonas putida* Group. *Microorganisms* **2021**, *9*, 1766. [CrossRef] [PubMed]
17. Rossi, E.; La Rosa, R.; Bartell, J.A.; Marvig, R.L.; Haagensen, J.A.J.; Sommer, L.M.; Molin, S.; Johansen, H.K. *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat. Rev. Microbiol.* **2021**, *19*, 331–342. [CrossRef] [PubMed]
18. Lund-Palau, H.; Turnbull, A.R.; Bush, A.; Bardin, E.; Cameron, L.; Soren, O.; Wierre-Gore, N.; Alton, E.W.; Bundy, J.G.; Connett, G.; et al. *Pseudomonas aeruginosa* infection in cystic fibrosis: Pathophysiological mechanisms and therapeutic approaches. *Expert Rev. Respir. Med.* **2016**, *10*, 685–697. [CrossRef]
19. Xin, X.F.; Kvitko, B.; He, S.Y. *Pseudomonas syringae*: What it takes to be a pathogen. *Nat. Rev. Microbiol.* **2018**, *16*, 316–328. [CrossRef] [PubMed]
20. Silby, M.W.; Winstanley, C.; Godfrey, S.A.; Levy, S.B.; Jackson, R.W. *Pseudomonas* genomes: Diverse and adaptable. *FEMS Microbiol. Rev.* **2011**, *35*, 652–680. [CrossRef] [PubMed]

21. Desnoues, N.; Lin, M.; Guo, X.; Ma, L.; Carreno-Lopez, R.; Elmerich, C. Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology* **2003**, *149*, 2251–2262. [CrossRef]
22. Duman, M.; Mulet, M.; Altun, S.; Saticioglu, I.B.; Gomila, M.; Lalucat, J.; Garcia-Valdes, E. *Pseudomonas piscium* sp. nov., *Pseudomonas psciculturae* sp. nov., *Pseudomonas mucoides* sp. nov. and *Pseudomonas neuropathica* sp. nov. isolated from rainbow trout. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004714. [CrossRef] [PubMed]
23. Oueslati, M.; Mulet, M.; Gomila, M.; Berge, O.; Hajlaoui, M.R.; Lalucat, J.; Sadfi-Zouaoui, N.; García-Valdés, E. New species of pathogenic *Pseudomonas* isolated from citrus in Tunisia: Proposal of *Pseudomonas kairouanensis* sp. nov. and *Pseudomonas nabeulensis* sp. nov. *Syst. Appl. Microbiol.* **2019**, *42*, 348–359. [CrossRef] [PubMed]
24. Mehmood, N.; Saeed, M.; Zafarullah, S.; Hyder, S.; Rizvi, Z.F.; Gondal, A.S.; Jamil, N.; Iqbal, R.; Ali, B.; Ercisli, S.; et al. Multifaceted Impacts of Plant-Beneficial *Pseudomonas* spp. in Managing Various Plant Diseases and Crop Yield Improvement. *ACS Omega* **2023**, *8*, 22296–22315. [CrossRef] [PubMed]
25. Weimer, A.; Kohlstedt, M.; Volke, D.C.; Nikel, P.I.; Wittmann, C. Industrial biotechnology of *Pseudomonas putida*: Advances and prospects. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 7745–7766. [CrossRef] [PubMed]
26. Kruse, L.; Loeschcke, A.; de Witt, J.; Wierckx, N.; Jaeger, K.E.; Thies, S. *Halopseudomonas* species: Cultivation and molecular genetic tools. *Microb. Biotechnol.* **2024**, *17*, e14369. [CrossRef] [PubMed]
27. Moore, E.R.; Mihaylova, S.A.; Vandamme, P.; Krichevsky, M.I.; Dijkshoorn, L. Microbial systematics and taxonomy: Relevance for a microbial commons. *Res. Microbiol.* **2010**, *161*, 430–438. [CrossRef]
28. Garrity, G.M. A New Genomics-Driven Taxonomy of Bacteria and Archaea: Are We There Yet? *J. Clin. Microbiol.* **2016**, *54*, 1956–1963. [CrossRef]
29. Gupta, R.S. Microbial Taxonomy: How and Why Name Changes Occur and Their Significance for (Clinical) Microbiology. *Clin. Chem.* **2021**, *68*, 134–137. [CrossRef] [PubMed]
30. Goodfellow, M. Microbial Systematics: Background and Uses. In *Applied Microbial Systematics*; Priest, F.G.G.M., Ed.; Kluwer Academic Publishers: Dordrrecht, The Netherlands, 2000; pp. 1–18.
31. Passarelli-Araujo, H.; Franco, G.R.; Venancio, T.M. Network analysis of ten thousand genomes shed light on *Pseudomonas* diversity and classification. *Microbiol. Res.* **2022**, *254*, 126919. [CrossRef]
32. Gupta, R.S. Identification of conserved indels that are useful for classification and evolutionary studies. *Methods Microbiol.* **2014**, *41*, 153–182.
33. Gupta, R.S. Impact of genomics on the understanding of microbial evolution and classification: The importance of Darwin's views on classification. *FEMS Microbiol. Rev.* **2016**, *40*, 520–553. [CrossRef] [PubMed]
34. Gupta, R.S.; Patel, S.; Saini, N.; Chen, S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: Description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5753–5798. [PubMed]
35. Holmes, B.; Steigerwalt, A.; Weaver, R.; Brenner, D.J. *Chryseomonas polytricha* gen. nov., sp. nov., a *Pseudomonas*-like organism from human clinical specimens and formerly known as group Ve-1. *Int. J. Syst. Evol. Microbiol.* **1986**, *36*, 161–165. [CrossRef]
36. Hespell, R.B. *Serpens flexibilis* gen. nov., sp. nov., an unusually flexible, lactate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* **1977**, *27*, 371–381. [CrossRef]
37. Palleroni, N.J. Pseudomonas. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley and Sons: Hoboken, NJ, USA, 2015; pp. 1–105.
38. Parte, A.C.; Sarda Carbasse, J.; Meier-Kolthoff, J.P.; Reimer, L.C.; Goker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5607–5612. [CrossRef]
39. Gupta, R.S.; Kanter-Eivin, D.A. AppIndels.com Server: A Web Based Tool for the Identification of Known Taxon-Specific Conserved Signature Indels in Genome Sequences: Validation of Its Usefulness by Predicting the Taxonomic Affiliation of >700 Unclassified strains of *Bacillus* Species. *Int. J. Syst. Evol. Microbiol.* **2023**, *73*, 005844. [CrossRef]
40. Sayers, E.W.; Agarwala, R.; Bolton, E.E.; Brister, J.R.; Canese, K.; Clark, K.; Connor, R.; Fiorini, N.; Funk, K.; Hefferon, T.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2019**, *47*, D23–D28. [CrossRef]
41. Wang, Z.; Wu, M. A phylum-level bacterial phylogenetic marker database. *Mol. Biol. Evol.* **2013**, *30*, 1258–1262. [CrossRef] [PubMed]
42. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
43. Girard, L.; Lood, C.; De Mot, R.; van Noort, V.; Baudart, J. Genomic diversity and metabolic potential of marine *Pseudomonadaceae*. *Front. Microbiol.* **2023**, *14*, 1071039. [CrossRef]
44. Kujur, R.R.A.; Ghosh, M.; Basak, S.; Das, S.K. Phylogeny and structural insights of lipase from *Halopseudomonas maritima* sp. nov., isolated from sea sand. *Int. Microbiol.* **2023**, *26*, 1021–1031. [CrossRef] [PubMed]

45. Rudra, B.; Duncan, L.; Shah, A.J.; Shah, H.N.; Gupta, R.S. Phylogenomic and comparative genomic studies robustly demarcate two distinct clades of *Pseudomonas aeruginosa* strains: Proposal to transfer the strains from an outlier clade to a novel species *Pseudomonas paraeruginosa* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2022**, *72*, 005542. [CrossRef]
46. Winsor, G.L.; Griffiths, E.J.; Lo, R.; Dhillon, B.K.; Shay, J.A.; Brinkman, F.S. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* **2016**, *44*, D646–D653. [CrossRef] [PubMed]
47. Palleroni, N.J.; Genus, I. Pseudomonas Migula 1894. In *Bergey's Manual of Systematic Bacteriology (The Proteobacteria), Part B (The Gammaproteobacteria)*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Eds.; Springer: New York, NY, USA, 2005; Volume 2, pp. 323–379.
48. Stover, C.K.; Pham, X.Q.; Erwin, A.; Mizoguchi, S.; Warrener, P.; Hickey, M.; Brinkman, F.; Hufnagle, W.; Kowalik, D.; Lagrou, M. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **2000**, *406*, 959–964. [CrossRef] [PubMed]
49. Planquette, B.; Timsit, J.-F.; Misset, B.Y.; Schwebel, C.; Azoulay, E.; Adrie, C.; Vesin, A.; Jamali, S.; Zahar, J.-R.; Allaouchiche, B. *Pseudomonas aeruginosa* ventilator-associated pneumonia. predictive factors of treatment failure. *Am. J. Respir. Crit. Care Med.* **2013**, *188*, 69–76. [CrossRef]
50. Spagnolo, A.M.; Sartini, M.; Cristina, M.L. *Pseudomonas aeruginosa* in the healthcare facility setting. *Rev. Res. Med. Microbiol.* **2021**, *32*, 169–175. [CrossRef]
51. Mulani, M.S.; Kamble, E.E.; Kumkar, S.N.; Tawre, M.S.; Pardesi, K.R. Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.* **2019**, *10*, 539. [CrossRef] [PubMed]
52. Malhotra, M.; Bello, S.; Gupta, R.S. Phylogenomic and molecular markers based studies on clarifying the evolutionary relationships among *Peptoniphilus* species. Identification of several Genus-Level clades of *Peptoniphilus* species and transfer of some *Peptoniphilus* species to the genus *Aedoeadaptatus*. *Syst. Appl. Microbiol.* **2024**, *47*, 126499. [CrossRef]
53. Sood, U.; Singh, D.N.; Hira, P.; Lee, J.K.; Kalia, V.C.; Lal, R.; Shakarad, M. Rapid and solitary production of mono-rhamnolipid biosurfactant and biofilm inhibiting pyocyanin by a taxonomic outlier *Pseudomonas aeruginosa* strain CR1. *J. Biotechnol.* **2020**, *307*, 98–106. [CrossRef] [PubMed]
54. Khadka, B.; Persaud, D.; Gupta, R.S. Novel Sequence Feature of SecA Translocase Protein Unique to the Thermophilic Bacteria: Bioinformatics Analyses to Investigate Their Potential Roles. *Microorganisms* **2020**, *8*, 59. [CrossRef]
55. Singh, B.; Gupta, R.S. Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genom.* **2009**, *281*, 361–373. [CrossRef] [PubMed]
56. Hashimoto, K.; Panchenko, A.R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20352–20357. [CrossRef] [PubMed]
57. Miton, C.M.; Tokuriki, N. Insertions and deletions (indels): A missing piece of the protein engineering jigsaw. *Biochemistry* **2022**, *62*, 148–157. [CrossRef] [PubMed]

CHAPTER 6

Phylogenomics and Molecular Marker-Based Studies to Clarify Evolutionary Relationships Among Species from The Fluorescens Lineage.

This chapter provides a comprehensive phylogenetic analysis of *Pseudomonas* species within the Fluorescens lineage, revealing 13 distinct genus-level clades. This chapter also highlights the use of CSIs for reliably distinguishing some of these clades and outlines a future objective to identify CSIs for the remaining clades. My contributions to this work include constructing the phylogenetic trees, performing comparative genomic analyses to identify CSIs, and contributing to the writing of this chapter.

This Chapter is an altered manuscript that will be submitted for publication in the coming months.

Abstract

The genus *Pseudomonas* includes large assemblages of bacteria with diverse properties. In addition to infecting humans, animals, insects, and plants, species play vital roles in biocontrol, bioremediation, and plant-microbe interactions. Phylogenetic studies have shown that the genus *Pseudomonas* is polyphyletic, and species form three major lineages: Aeruginosa, Fluorescens, and Pertucinogena. While recent taxonomic revisions have reassigned distinct clades of species from the Aeruginosa and Pertucinogena lineages into several novel and emended genera, the classification of the Fluorescens lineage, accounting for ~70% of *Pseudomonas* species, remains unresolved. To address this, we conducted a comprehensive phylogenomic and comparative genomic analysis on the Fluorescens lineage of species. Our genome-scale phylogenetic tree identified 13 distinct genus-level clades, such as Asplenii, Chlororaphis, Corrugata, Fluorescens, Fragi, Jessenii, Koreensis, Lutea, Mandelii, Massiliensis, Putida, Rhizosphaerae, and Syringae. To confirm their distinctiveness and reliable demarcation, we identified multiple Conserved Signature Indels (CSIs) uniquely shared by the species from the Massiliensis, Putida, and Rhizosphaerae clades. Future studies will identify CSIs for the remaining clades serving as independent molecular markers for their reliable demarcation.

Introduction

The genus *Pseudomonas* comprises over 350 species with validly published names (LPSN, accessed on February 2025) (Parte, 2018), exhibiting immense genetic and metabolic diversity (Palleroni, 2005; Peix et al., 2009; Lund-Palau et al., 2016; Winsor et al., 2016; Pang et al., 2021; Rossi et al., 2021). Due to their clinical, economic, and ecological importance, *Pseudomonas* species have been extensively studied (Hesse et al., 2018; Peix et al., 2018; Rudra and Gupta, 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022). Phylogenetic analyses reported in these studies establish that this genus is highly polyphyletic, and the large assemblages of species that are currently part of the genus *Pseudomonas* form three main lineages in phylogenetic trees: Aeruginosa, Fluorescens, and Pertucinogena. Of these three lineages, the Aeruginosa and Fluorescens lineages consist of multiple distinct clades, which are phenotypically and genotypically distinct from a specific species clade, *viz.* the Aeruginosa clade, which contains the type species (*P. aeruginosa*) of this genus. Additionally, species from other genera have been found to cluster in between different *Pseudomonas* species, making this genus highly polyphyletic (Hesse et al., 2018; Peix et al., 2018; Rudra and Gupta, 2021). To address these taxonomic inconsistencies and establish a reliable classification framework for the genus *Pseudomonas*, several taxonomic revisions were carried out using comparative genomics-based approaches on the Pertucinogena and Aeruginosa lineages of species (Rudra and Gupta, 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022; Rudra and Gupta, 2024). These studies have led to the reclassification of over 150 *Pseudomonas* species into several novel and emended genera, including *Aquipseudomonas, Atopomonas, Caenipseudomonas, Chryseomonas, Ectopseudomonas, Geopseudomonas, Halopseudomonas, Metapseudomonas,*

Paraburkholderia, Phytopseudomonas, Serpens, Stenotrophomonas, Stutzerimonas, Thiopseudomonas, *Xanthomonas*, and *Zestomonas*. Moreover, according to the Code governing the nomenclature of Prokaryotes (Oren et al., 2023), only species from the Aeruginosa clade should constitute the genus *Pseudomonas*. Despite these advancements, the taxonomic status of ~70% of *Pseudomonas* species forming the Fluorescens lineage remains unresolved.

Species within the Fluorescens lineage show remarkable diversity and have been isolated from a wide range of environments, including water (Miranda and Zemelman, 2002), soil (Andersen et al., 2000), plant tissues (Brown et al., 2012), fungi (Rainey et al., 1993), animals (Vela et al., 2006), and humans (Scales et al., 2015). These bacteria play essential agricultural and ecological roles, contributing to biocontrol, siderophore production, denitrification, toxin synthesis, bioremediation, and plant-microbe interactions (Silby et al., 2011; Scales et al., 2014; Garrido-Sanz et al., 2016). Among them, *P. fluorescens, P. corrugata, P. chlororaphis*, and *P. protegens* are extensively studied for their effectiveness as biocontrol agents. They employ diverse mechanisms, such as competitive colonization in plant tissues, antibiosis, siderophore production, and secretion of lytic enzymes (Silby et al., 2011; Scales et al., 2014; Raio, 2024). In contrast, *P. syringae* is recognized as one of the most studied plant pathogens, ranked among the top 10 plant-pathogenic bacteria (Mansfield et al., 2012). It primarily infects the phyllosphere, where it exists as an epiphyte on plant surfaces (Xin et al., 2018). Another notable species, *P. putida*, is a metabolically versatile soil bacterium capable of degrading a wide range of organic compounds, including xenobiotics, making it a key player in bioremediation efforts (Iyer and Damania, 2016; Papadopoulou et al.,

2018). *P. jessenii* is also suited for bioremediation and rhizoremediation applications (Garrido-Sanz et al., 2016; Raio, 2024).

Given the ecological and agricultural impact of the species from the Fluorescens lineage, several phylogenomic studies have explored the evolutionary relationships among these species (Mulet et al., 2010; Beiki et al., 2016; Garrido-Sanz et al., 2016; Garrido-Sanz et al., 2017; Girard et al., 2020; Lalucat et al., 2020; Passarelli-Araujo et al., 2022). These studies have grouped the species from the Fluorescens lineages into five phylogenetic groups: Asplenii*,* Fluorescens, Lutea, Syringae, and Putida. Of these groups, the Fluorescens group is further subdivided into nine subgroups. However, the species composition of these subgroups often varies in different studies due to the inclusion of new species over time (Gomila et al., 2015; Garrido-Sanz et al., 2016; Gomila et al., 2017; Lalucat et al., 2020; Mulet et al., 2024). These different phylogroups exhibit pronounced phenotypic and genetic diversity, making it difficult to establish a phylogeny that accurately reflects their genetic cohesiveness and shared evolutionary history (Mulet et al., 2010;Garrido-Sanz et al., 2017). This complexity is further compounded by the frequent discovery of new species and the inclusion of unrelated strains within this lineage (Lalucat et al., 2020; Girard et al., 2021; Lalucat et al., 2022). Whole-genome-based studies reveal that strains of several species, including *P. fluorescens, P. putida,* etc., are often misclassified, resulting in their dispersed placement within the Fluorescens lineage (Nikolaidis et al., 2020; Passarelli-Araujo et al., 2022). Since the genus *Pseudomonas* has been proposed and needs to be restricted to the species within the Aeruginosa clade, and most of the species from the Pertucinogena and Aeruginosa lineages have already been reliably reclassified into distinct monophyletic genera (Saati-Santamaria et al., 2021; Lalucat et al., 2022; Rudra

and Gupta, 2024), it is of great interest to conduct a comprehensive study on the species from the Fluorescens lineage to identify distinct monophyletic groups/subgroups that are robustly supported using multiple genome sequence-based approaches.

With the aim of establishing a reliable phylogenetic framework for the classification of the Fluorescens lineage of species, in the present work, we have conducted detailed phylogenomic and comparative genomic studies on the genome sequences of 245 species from the Fluorescens lineage. Based on their genome sequences, we have constructed a robust genome-scale phylogenomic tree based on large datasets of conserved proteins from these species. In the phylogenomic tree that we have constructed, the Fluorescens lineage of species forms 13 distinct clades/subclades (*viz.*, Asplenii, Chlororaphis, Corrugata, Fluorescens, Fragi, Jessenii, Koreensis, Lutea, Mandelii, Massiliensis, Putida, Rhizosphaerae and Syringae), similar to those reported in earlier studies (Garrido-Sanz et al., 2017; Hesse et al., 2018; Lalucat et al., 2020; Passarelli-Araujo et al., 2022). In addition to the phylogenetic studies showing the distinctness of these species clades, we also report here the results of detailed comparative genomic studies on protein sequences from these species to identify novel molecular markers, consisting of Conserved Signature Indels (CSIs) in protein sequences, which are exclusively found in the species from several clades including Massiliensis, Putida, and Rhizosphaerae. The molecular markers that are uniquely found in different monophyletic groups of species provide strong evidence supporting the distinctness of different species clades independently of the phylogenetic trees and provide robust means for their demarcation in molecular terms (Gupta, 1998; Sawana et al., 2014; Hu et al., 2019; Bello et al., 2022b). More CSIs will be identified, exclusively shared by the species from the remaining 10 clades, in the coming months.

Using these CSIs, all 13 main clades within the Fluorescens lineage can be reliably demarcated and distinguished from each other based upon multiple exclusively shared molecular characteristics.

Methods

Construction of Phylogenetic Tree

Genome sequences for 388 named *Pseudomonadaceae* species were obtained from the NCBI database, using type strains where available. A phylogenetic tree for these species was constructed based on concatenated sequences of 118 conserved proteins, similar to our recent study (Rudra and Gupta, 2024). *Moraxella bovoculi* and *M. bovis* were used as outgroup species to root the tree. A maximum-likelihood (ML) tree was generated using an internally developed pipeline, as described in other studies (Adeolu et al., 2016; Rudra and Gupta, 2021; Rudra and Gupta, 2024). Protein families with at least 50% sequence identity and present in 80% of the input genomes were identified using the CD-HIT program (Li and Godzik, 2006; Fu et al., 2012), and multiple sequence alignments were performed with Clustal Omega (Sievers et al., 2011). TrimAl (Capella-Gutiérrez et al., 2009) was used to remove poorly aligned regions before concatenation. The alignment contained 42,362 amino acid positions. FastTree 2 (Price et al., 2010) was used for initial tree construction, using the Whelan and Goldman model (Whelan and Goldman, 2001), followed by refinement in RAxML 8 (Stamatakis, 2014) with the Le and Gascuel model (Le and Gascuel, 2008). The resulting phylogenetic tree was formatted and labeled using MEGA X (Kumar et al., 2018).

Identification of Conserved Signature Indels (CSIs)

Identification of CSIs was carried out using procedures similar to those described in our earlier work (Gupta, 2014; Rudra and Gupta, 2021; Rudra and Gupta, 2024). Briefly, local BLASTp searches were carried out on protein sequences from the genomes of several *Pseudomonas* species representing diverse clades/lineages of interest, as well as other outgroup species. Based on these BLAST searches, sequences of high-scoring homologs (E value <1e-20) of different proteins were retrieved for several species (generally between 2 to 10) from the group of interest and 10-15 species from other *Pseudomonas* clades or other *Pseudomonadaceae* genera. Multiple sequence alignments were created using the ClustalX 2.1 program (Jeanmougin et al., 1998). The alignments were visually analyzed for fixed-length insertions or deletions found in conserved regions (i.e., flanked on both sides by minimally 5-6 conserved aa residues in the neighboring 40-50 aa) and which were only found in the *Pseudomonas* species from the clade of interest. The indels, not present in conserved regions, were not further considered. The query sequences consisting of the conserved indels and their flanking 30-40 aa on each side (generally beginning and ending with a stretch of conserved amino acids) were selected for a second BLASTp search. This latter BLASTp search was conducted against the NCBI nr database, and the resultant 250-500 hits were evaluated to determine the group specificities of the observed indels (CSIs). Based on these BLASTp results, the indels that were only present in different species from a specific clade of *Pseudomonas* were further formatted using the SIG_CREATE and SIG_STYLE programs (Gupta, 2014;2016).

Results

Phylogenetic Studies on the Species from the Fluorescens Lineage

To understand the cladistic relationships among species within the Fluorescens lineage, we have reconstructed a phylogenomic tree for all named *Pseudomonadaceae* species (Sayers et al., 2019; Parte et al., 2020), similar to our recent work (Rudra and Gupta, 2024). The tree, shown in Fig. 1, is based on concatenated sequences of 118 conserved proteins comprising the phyloeco set for the class *Gammaproteobacteria* (Wang and Wu, 2013). *Moraxella* species (*M. bovis* and *M. bovoculi*) from the family *Moraxellaceae* were used to root the tree. This tree will be referred to as the "phyloeco tree" in this study. In this tree, nearly all nodes were supported with 100% bootstrap values (SH scores), confirming the robustness of the inferred evolutionary relationships among different *Pseudomonadaceae* species clades.

The overall clustering of *Pseudomonadaceae* species in the tree shown in Fig. 1A is consistent with our previous studies (Rudra and Gupta, 2021; Rudra and Gupta, 2024). In this tree, all *Pseudomonas* species are grouped into three major lineages (compressed and labeled as Pertucinogena, Aeruginosa, and Fluorescens). Within the *Aeruginosa* and *Fluorescens* lineages, multiple distinct genus-level clades are observed. The species from most of the clades within the *Pertucinogena* and *Aeruginosa* lineages have now been reclassified into other genera. These clades are shown in the tree in compressed form and labeled with their new genera names. However, the present study focuses on the Fluorescens lineage of species, which is compressed in Fig. 1A, and the uncompressed form of this large species lineage, showing its different clades, is presented in Fig. 1 (B) and 2 (A-E). As seen in these Figs., species from the *Fluorescens*

lineage group into 13 distinct clades, each labeled according to commonly used clade/subclade names based on representative species from them. These clades include Asplenii*, Chlororaphis, Corrugata, Fluorescens, Fragi, Jessenii, Koreensis, Lutea, Mandelii, Massiliensis, Putida, Rhizosphaerae,* and *Syringae. Two species, P. frederiksbergensis* and *P. akappageensis,* do not group with any observed clades and form separate lineages. These different clade structures, branching, and naming are consistent with other studies on Fluorescens lineage, *except for some variations in species composition within some clades/subclades* (Garrido-Sanz et al., 2017; Hesse et al., 2018; Nikolaidis et al., 2020; Girard et al., 2021; Lalucat et al., 2022). To facilitate the visualization of species composition, the uncompressed forms of Asplenii, Chlororaphis, Corrugata, Fragi, Jessenii, Lutea, Massiliensis, and Rhizosphaerae clades are shown in Fig. 1B, while those of the species from the Koreensis, Mandelii, Putida, Fluorescens, and Syringae clades are depicted in Fig. 2A-E, respectively.

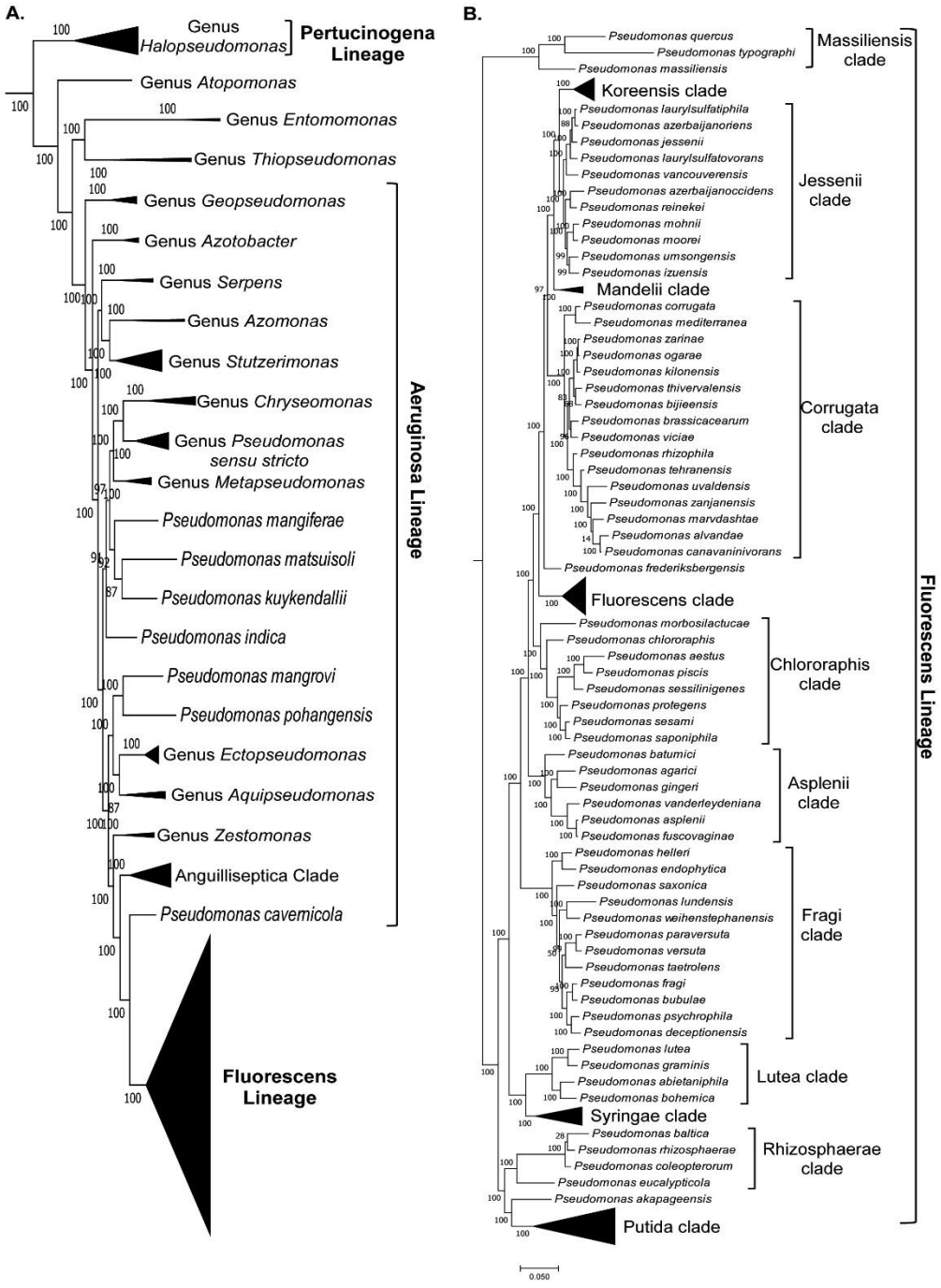

Figure 1. (A) Maximum-likelihood tree of 388 genome-sequenced *Pseudomonadaceae* species, constructed using concatenated sequences of 118 conserved proteins. All clades/ genera within the Pertucinogena and Aeruginosa lineages have been compressed and labeled with the recently described new and existing genera names. (B) Expanded view of all clades within the Fluorescens lineage. For clarity and ease of presentation, some clades/subclades, including Koreensis, Mandelii, Putida, Fluorescens, and Syringae, are compressed, with their uncompressed versions shown in Fig. 2 (A-E).

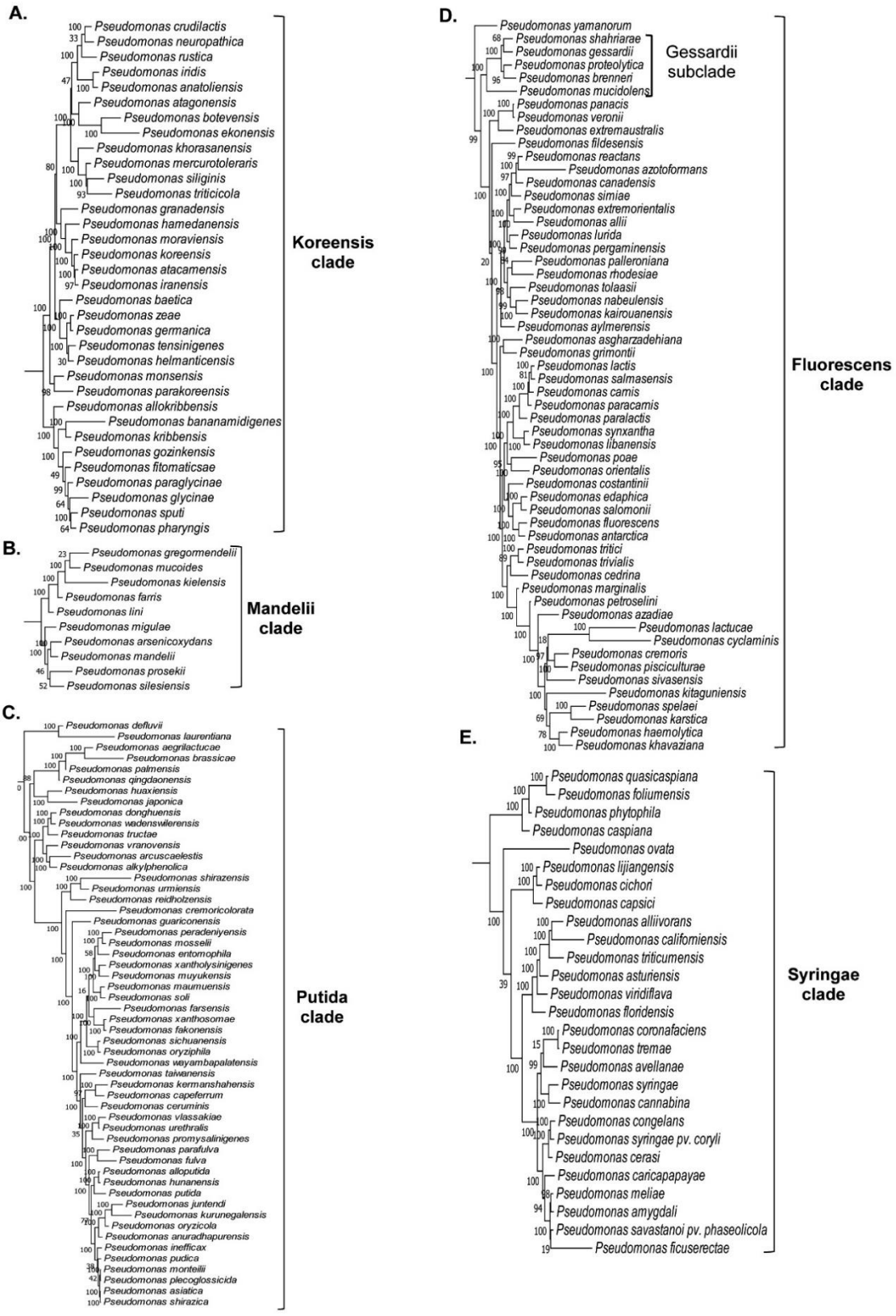

Figure 2. Expanded view of the clades **(A)** Koreensis, **(B)** Mandelii, **(C)** Putida, **(D)** Fluorescens, and **(E)** Syringae.

As seen from the tree shown in Fig. 1B, the Massiliensis clade of species forms a distinct, deepest-branching clade within the *Fluorescens* species lineage. Of the remaining groups/clades, species belonging to eight distinct genus-level clades, *viz.*, Koreensis, Jessenii, Mandelii, Corrugata, Fluorescens, Chlororaphis, Asplenii, and Fragi, are commonly referred to as the subclades of the Fluorescens clade (Garrido-Sanz et al., 2016; Garrido-Sanz et al., 2017; Hesse et al., 2018; Girard et al., 2021; Lalucat et al., 2022), where in some of these studies, these clades were classified as part of the *P. fluorescens* complex, constituting one of the most diverse groups within the genus *Pseudomonas (Garrido-Sanz et al., 2016; Garrido-Sanz et al., 2017)*. Based on their distinctness in phylogenetic analysis, our study labeled all these groups as distinct clades. In earlier studies, two additional subgroups, i.e., *P. gessardii* and *P. protegens* subclades, are also indicated as distinct lineages within the *P. fluorescens* complex. However, based on short branches in our phylogenomic tree shown in Figs. 1 and 2, and in earlier studies (Garrido-Sanz et al., 2017; Lalucat et al., 2020; Lalucat et al., 2022), as well as molecular evidence (will be discussed later), these subgroups cannot be reliably distinguished from the species in neighboring clades. Hence, in the present study, we consider the Gessardii and Protegens subclades as part of the Fluorescens clade (shown in Fig. 2D) and the Chlororaphis clade (Fig. 1B), respectively. Further evidence supporting the grouping of these subclades with the Fluorescens and Chlororaphis clades will be discussed later. The distinctness of the 13 distinct species clades observed in the tree shown in Figs. 1 and 2 are also strongly supported by the results of our comparative genomics analyses presented below.

Identification of Molecular Markers for Different Clades Within the Fluorescens Lineage:

As previously mentioned, *Pseudomonas* species from the Fluorescens lineage consistently form distinct genus-level clades and subclades across different genome-scale phylogenies. However, the branching patterns in these trees are influenced by multiple factors such as the number and type of chosen gene sets, the algorithm used for sequences alignment and tree construction, selection of outgroups, long branch attraction effects, horizontal gene transfer (HGT) events, etc. (Gupta, 1998; Gupta and Griffiths, 2002; Baldauf, 2003; Felsenstein, 2004). Moreover, the branching of different clades/subclades in the phylogenetic trees forms a continuum, with some nodes being separated by short branch lengths, making it difficult to define their boundaries precisely. Given these challenges, it is important to identify other definitive markers for distinguishing these clades/subclades. Molecular synapomorphies in the form of CSIs in gene and protein sequences serve as unique, clade-specific features, and they have proven useful for taxonomic classification at different levels (Gao and Gupta, 2005; Griffiths and Gupta, 2006b; Gupta, 2014; Adeolu et al., 2016). In our earlier work, based on the presence or absence of CSIs, we clarified the taxonomic position of a large number of *Pseudomonas* species clades from both the Aeruginosa and Pertucinogena lineages by transferring them into different novel genera (Rudra and Gupta, 2021; Rudra and Gupta, 2024). Hence, as a continuation of my earlier work on the genus *Pseudomonas*, I have also identified several CSIs, which are specific to different clades within the Fluorescens lineage. These clade-specific CSIs provide strong genetic evidence for the distinctiveness of these species' clades and offer a more reliable method for their taxonomic demarcation. We provide below some examples of the identified

CSIs, which are specific for some of the Fluorescens lineage of species clades noted above. More detailed and comprehensive information regarding the CSIs specific to different observed clades will be provided in a manuscript that will be submitted in the coming months.

Identification of CSIs for the Massiliensis clade of species:

The Massiliensis clade consists of three species: "*P. massiliensis*"*, P. quercus,* and *P. typographi*. Species from this clade have been isolated from diverse sources, including fecal flora (stool) (Bardet et al., 2018), leaf spot disease (Li et al., 2021b), and bark beetles (Peral-Aranega et al., 2020). In the phylogenetic tree, they form a well-supported, distinct clade that is separated from other clades within the Fluorescens lineage by a long branch. Our comparative genomic analysis identified six CSIs that are exclusively shared by all three species within the Massiliensis clade but are absent in all other *Pseudomonadaceae* species or other bacteria. One of these CSIs, shown in Fig. 3A, is a four aa insertion in the mannitol dehydrogenase family protein, uniquely present in all three species. This is an important metabolic enzyme that facilitates the reversible oxidation of D-mannitol to D-fructose, D-arabinitol to D-xylulose, and D-sorbitol to L-sorbose by transferring the C2 hydride to the pro-S position on the nicotinamide (Kavanagh et al., 2002). Thus, based on the molecular markers (CSIs), the Massiliensis clade of species can be reliably demarcated from the rest of the *Pseudomonasaceae* species clades. Due to space constraints, sequence information for limited *Pseudomonadaceae* species (outgroups) is presented here. Additionally, as the results for the CSIs identified in this study have not yet been published, I am presenting only limited information here for the CSIs specific to different clades.

Identification of CSIs for the Rhizosphaerae and Putida clades of species:

Another distinct clade, the Rhizosphaerae clade, consists of four species: *P. baltica, P. coleopterorum,* "*P. eucalypticola*"*,* and *P. rhizosphaerae*. These species were isolated from diverse sources, including grass (Peix et al., 2003), insects (Menendez et al., 2015), plant leaves (Liu et al., 2021b), and raw milk (Gieschler et al., 2021). Notably, this clade consists of species known to produce antifungal agents (Liu et al., 2021b), whereas the neighboring Putida clade is primarily recognized for its role in bioremediation (Papadopoulou et al., 2018). In the phylogenetic tree, species from the Rhizosphaerae clade form a well-supported clade distinct from the neighboring Putida clade and other *Pseudomonadaceae* species clades. Our comparative genomic analysis identified four CSIs that are exclusively present in all species of this clade but absent in other *Pseudomonadaceae* species or other bacteria. One example, shown in Fig. 3B, highlights a three-amino-acid deletion within the thioesterase family protein, which is shared explicitly by species of the Rhizosphaerae clade but not found in other bacteria. Thus, based on the phylogenomic distinctiveness and presence of shared CSIs, species from this clade (Rhizosphaerae) should be reclassified as a distinct genus.

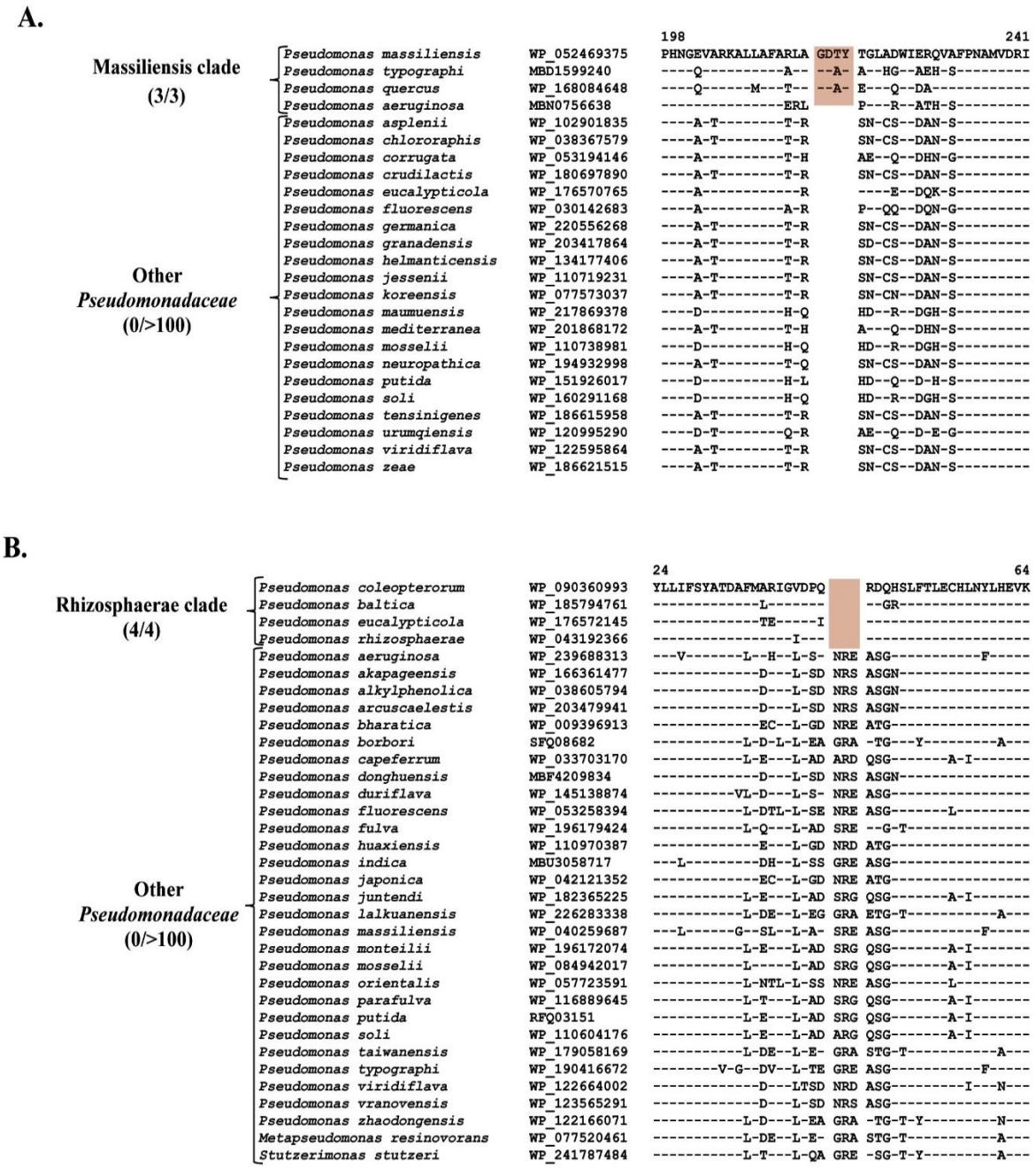

Figure 3: Partial sequence alignment of **(A)**-the mannitol dehydrogenase family protein showing a four aa insertion (highlighted) that is exclusively present in all members of the Massiliensis clade; **(B)** thioesterase family protein showing a three aa deletion (highlighted) that is shared explicitly by the Rhizosphaerae clade of species but not shared by any other species. The dashes (−) in sequence alignments indicate identity with the amino acids on the top line. Accession numbers for different sequences are indicated in the second column, and the numbers at the top indicate the position of this sequence in the protein sequences.

Within the Fluorescens lineage, the Putida clade forms the second-largest clade, encompassing 54 species (Fig. 1E). These species are widely studied for their biotechnological potential, particularly in environmental and industrial applications (Poblete-Castro et al., 2012; Keshavarz-Tohid et al., 2019). Strains within this group play a crucial role in carbon cycling due to their remarkable metabolic adaptability and ability to degrade a broad range of natural organic compounds and xenobiotics, including plastics, pesticides, lubricants, and other industrial pollutants (Udaondo et al., 2024). In our phylogenetic analysis (Figs. 1 and 2), this group of bacteria forms a strongly supported clade. However, some smaller subgroups are observed within the Putida clade, which are separated by short branches, indicating their close evolutionary relationships to the Putida clade. Similar observations are also reported in other studies (Girard et al., 2021). The results of our comparative genomic analyses have identified nine CSIs, which are exclusively shared by all 54 species from the Putida clade, providing a reliable means for the demarcation of this species clade. One example of a CSI specific to the Putida clade of species is presented in Fig. 4, where a one aa insertion in the protein Leucyl aminopeptidase is found explicitly in all species from this clade but not in any other *Pseudomonadaceae* species or other bacteria, providing strong evidence for the distinctiveness of this clade.

Like the examples of CSIs shown in Figs. 3 and 4, which are specific to the species from the Massiliensis, Rhizosphaerae, and Putida clades, in the coming months, our future work aims to identify CSIs specific to the remaining species clades within the *Fluorescens* lineage, as shown in Fig. 5. Using these CSIs, different species clades within the Fluorescens lineage can be reliably demarcated based on multiple exclusively shared molecular characteristics.

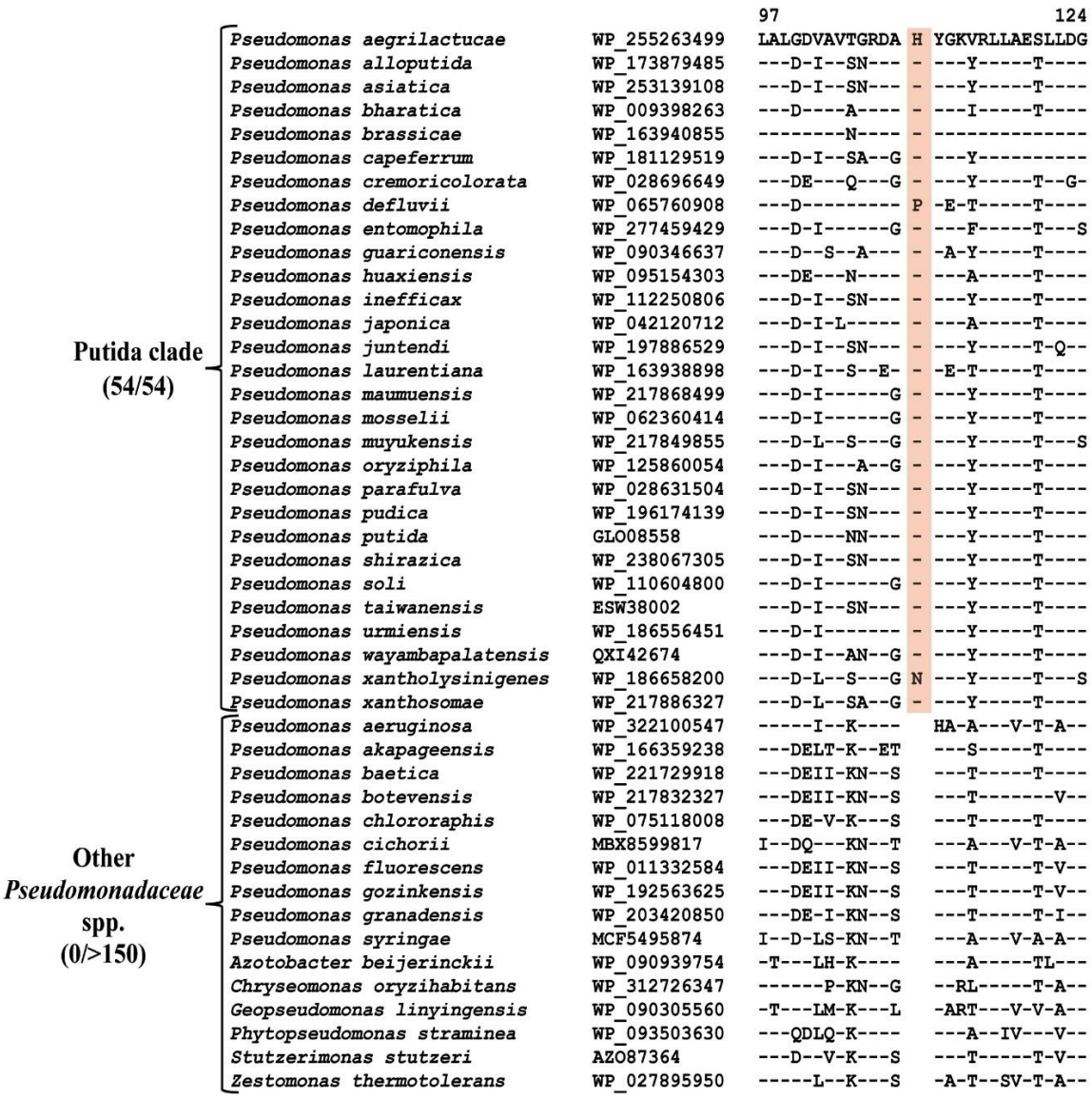

Figure 4: Partial sequence alignment of the protein Leucyl aminopeptidase showing a one aa insertion (highlighted) that is exclusively shared by all members of the Putida clade but not shared by any other *Pseudomonadaceae* species. Due to space constraints, this figure displays a limited number of ingroup and outgroup species.

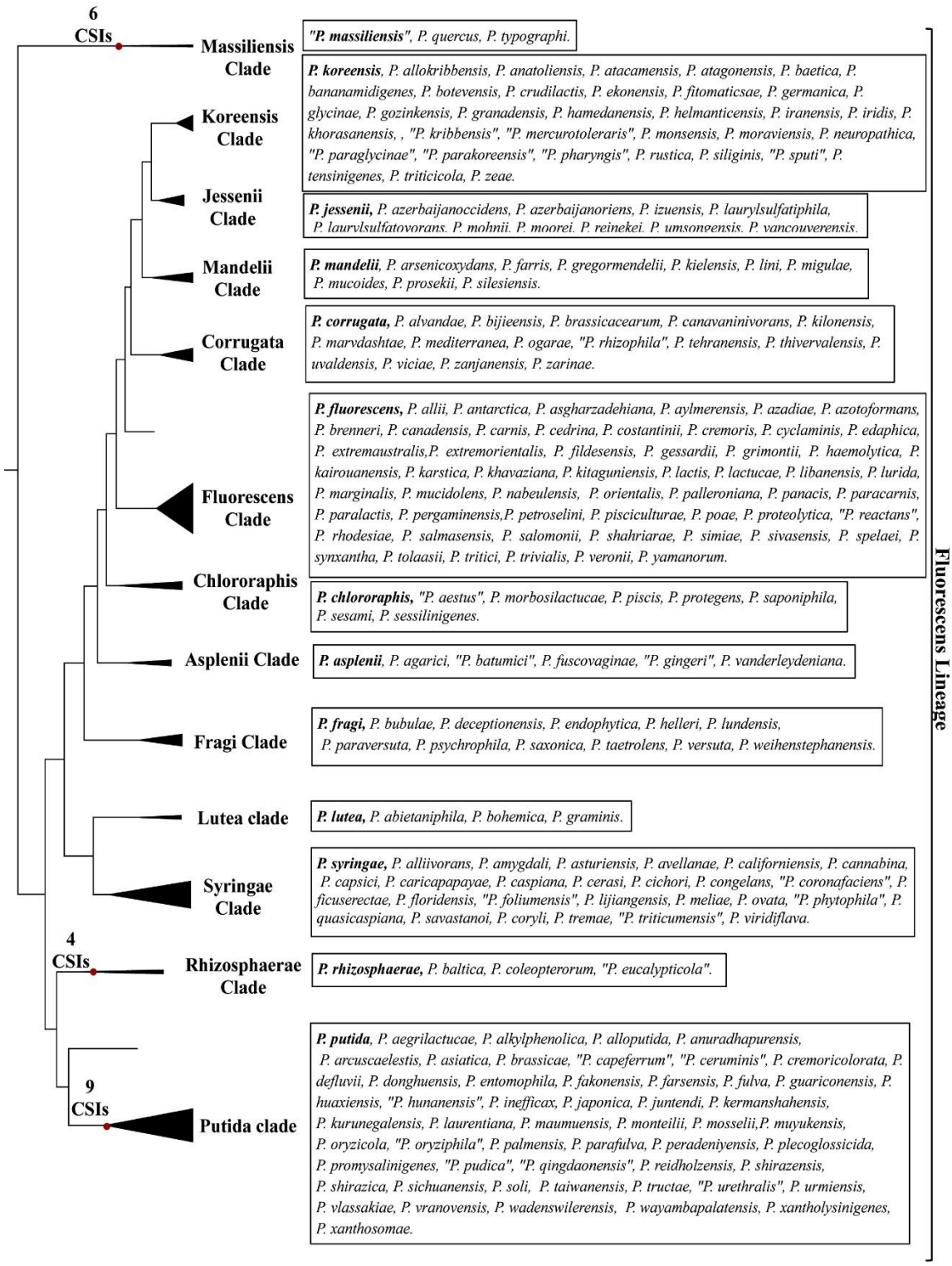

Figure 5: A conceptual diagram illustrating the branching pattern of all clades, highlighting representative species and showing the number and positions of CSIs specific to different clades within the *Fluorescens* lineage.

Discussion

Establishing a reliable and informative taxonomic framework for the genus *Pseudomonas* has been a long-standing challenge in the field of microbial taxonomy (Anzai et al., 2000; Peix et al., 2009; Peix et al., 2018; Lalucat et al., 2022). As indicated in the Introduction, this genus consists of species that are highly diverse and polyphyletic. Phylogenetic analyses reveal that *Pseudomonas* species cluster into three major lineages/groups: Pertucinogena, Aeruginosa, and Fluorescens. Earlier studies clarified the evolutionary relationships among species from the Aeruginosa and Pertucinogena lineages using phylogenomic and comparative genomics-based polyphasic approaches (Rudra and Gupta, 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022; Rudra and Gupta, 2024). However, these efforts have not reliably resolved the evolutionary relationships among species from the Fluorescens lineage, which comprises approximately two-thirds of *Pseudomonas* species. These large assemblages of diverse species are ecologically important, with valuable agricultural and environmental roles and different biotechnological applications (Vanparys et al., 2006; Silby et al., 2011; Scales et al., 2014; Raio, 2024).

With the aim of clarifying evolutionary relationships among species from the Fluorescens lineage, several phylogenomic-based studies were conducted. These studies demonstrated that the Fluorescens lineage consists of multiple genus-level clades/subclades in the phylogenetic trees with distinct phenotypic features (Mulet et al., 2010; Gomila et al., 2015; Garrido-Sanz et al., 2016; Garrido-Sanz et al., 2017). However, the distinguishing properties identified for several clades/subclades often overlap, making them unreliable for taxonomic classification. For example,

bioremediation or rhizoremediation traits are observed in both the Jesseni and Putida clades (Furmanczyk et al., 2018; Papadopoulou et al., 2018). Similarly, species with biocontrol or insecticidal properties are dispersed across the *Fluorescens*, *Corrugata*, and *Chlororaphis* clades (Scales et al., 2014; Garrido-Sanz et al., 2017; Raio, 2024). While the *Syringae* clade predominantly comprises plant pathogens (Gomila et al., 2017; Dutta et al., 2018; Mulet et al., 2024), pathogenic species also exist in the *Koreensis* clade (Kwon et al., 2003; Garrido-Sanz et al., 2016). A group of microbiologists who conducted extensive studies on the genus *Pseudomonas* highlighted the challenge of identifying clade- or subclade-specific traits, stating, "…*we attempted to find specific phenotypic traits that could be characteristic of and differentiate between groups, but our attempt was not successful…*" (Mulet et al., 2010). Therefore, it is of great interest to identify clade-specific biochemical or molecular taxonomic markers that can support the taxonomic reorganization of the *Fluorescens* lineage of species to be consistent with taxonomic principles, as species within the same genus should share similar biochemical or molecular properties (Gupta, 2021; Hugenholtz et al., 2021; Saati-Santamaria et al., 2021; Lalucat et al., 2022; Malhotra et al., 2024).

To address this gap, we conducted comprehensive phylogenomic and comparative genomic analyses on all named *Pseudomonas* species from LPSN (Parte et al., 2020) whose genome sequences are available in the NCBI genome database (https://www.ncbi.nlm.nih.gov/datasets/genome/). These analyses aim to revisit the phylogenetic relationships among different clades within the *Fluorescens* lineage and identify reliable molecular markers for defining distinct groups. The use of phylogenomic trees alongside conserved traits is instrumental in defining genera and

higher taxonomic ranks, helps to resolve ambiguities in poorly classified taxa, and promotes a more standardized and balanced classification across different phyla (Chun et al., 2018).

As our first attempt, we conducted a comprehensive phylogenetic analysis based on 118 conserved proteins, providing a robust overview of the cladistic relationships among different Fluorescens lineage of species. The phylogenomic trees constructed based on core genes or a large number of genes yield more accurate and robust results than a single gene (e.g., 16S rRNA) or a small group of genes (e.g., MLSA) (Rokas et al., 2003; Wu et al., 2009; Gao and Gupta, 2012b). In our phylogenetic analysis, 13 distinct clades/subclades (*viz.*, Asplenii, Chlororaphis, Corrugata, Fluorescens, Fragi, Jessenii, Koreensis, Lutea, Mandelii, Massiliensis, Putida, Rhizosphaerae and Syringae) were observed within the Fluorescens lineage, which are consistent with the findings from the earlier studies (Gomila et al., 2015; Garrido-Sanz et al., 2016; Hesse et al., 2018; Girard et al., 2021). Since these clades are distinct, their uniqueness is further supported by a large number of clade-specific molecular markers in the form of CSIs, with our major focus of this study being to define the boundaries of different clades using stable and reliable molecular markers. Our comparative genomics studies identified several CSIs, which are highly specific for different clades, including Massiliensis, Putida, and Rhizosphaerae. However, we will also identify CSIs for the remaining clades within the Fluorescens lineage in the coming months for their reliable demarcation by molecular means. The cladistic relationships of different clades, their species composition, along with their position and number of identified clade-specific CSIs, are shown in the conceptual Fig. 5. CSIs, found in gene/protein sequences and uniquely shared by different groups of organisms, offer reliable tools for taxonomic and

diagnostic studies (Gupta, 1998; Ahmod et al., 2011; Gupta and Kanter-Eivin, 2023). Since CSIs arise from rare genetic changes, their presence or absence in species is generally unaffected by most factors that could influence phylogenetic analyses (Baldauf and Palmer, 1993; Gupta, 1998; Rokas and Holland, 2000; Gupta, 2016). Additionally, because CSIs in different genes/proteins result from independent genetic events, each provides distinct evidence of a close and specific evolutionary relationship within a species group (Gao et al., 2006; Bhandari et al., 2013; Gupta, 2016; Hu et al., 2018). Thus, based on the shared molecular markers and phylogenomic distinctiveness, the species from the Fluorescens lineage could be reliably demarcated into 13 distinct genus-level clades, shown in Fig. 5.

One of the key features of this CSI-based classification is its predictive ability. Highly specific CSIs for different clades or genera can be used to determine the taxonomic affiliation of uncharacterized species or strains (Barbour et al., 2017; Gupta et al., 2020; Gupta and Kanter-Eivin, 2023). To harness this potential, the web server AppIndels.com was developed to predict the taxonomic placement of unclassified strains and species (Gupta and Kanter-Eivin, 2023). Our recent analysis of approximately 2,000 strains reassigned ~300 uncharacterized *Pseudomonas* species to their respective genera using the AppIndels server (Rudra and Gupta, 2025). However, a substantial number of *Pseudomonas* strains remain uncharacterized in the NCBI database, many of which are misclassified. Previous studies have reported that 25.65% of *Pseudomonas* genomes are misclassified (Passarelli-Araujo et al., 2022). By identifying CSIs specific to distinct clades within the Fluorescens lineage, we aim to provide a valuable resource for accurately classifying uncharacterized strains, thereby improving the overall accuracy of *Pseudomonas* taxonomy.

CHAPTER 7

DISCUSSION AND CONCLUSIONS

Discussion

Microbial systematics and taxonomy provide the foundational framework for organizing and understanding the diverse microbial world (Buchanan, 1955; Woese et al., 1990; Gupta, 1998; Garrity, 2016; Hugenholtz et al., 2021). It systematically organizes microbial strains into taxonomic ranks (species to phyla) based on shared characteristics while reconstructing their phylogenetic and evolutionary history (Woese et al., 1990; Gupta, 1998; Sutcliffe et al., 2012; Yarza et al., 2014; Parks et al., 2018; Gupta et al., 2020). Beyond evolutionary and systematic studies, a stable and informative taxonomic framework is essential across numerous scientific and applied fields, including medicine, biotechnology, environmental science, food production, and agriculture (Ward, 2002; Gevers et al., 2006; Bourdichon et al., 2012; Fan and Smith, 2021; Li et al., 2021a). Accurate taxonomy is crucial in medicine for identifying pathogens, selecting appropriate antibiotics, and analyzing antimicrobial resistance (Qin et al., 2022; Miller and Arias, 2024). In biotechnology, it facilitates strain selection for novel enzyme discovery and bioprospecting for antibiotics and other bioactive compounds (Saati-Santamaría et al., 2018; Nouioui and Sangal, 2022; Kruse et al., 2024). It also plays a key role in food production by supporting the selection of fermentation starter cultures and identifying spoilage organisms to ensure food safety (Bourdichon et al., 2012; Zheng et al., 2020). In agriculture, it aids in identifying nitrogen-fixing bacteria such as *Rhizobia*, characterizing plant growth-promoting bacteria, assessing soil microbial communities, and developing biofertilizers to enhance crop productivity (Majeed et al., 2015; David et al., 2018; Fan and Smith, 2021; Mora et al., 2022). Overall,

microbial taxonomy serves as a cornerstone across multiple disciplines, advancing microbial research and supporting critical applications in health, industry, and environmental sustainability.

The genomic era has revolutionized prokaryotic taxonomy, offering more profound insights into microbial evolution, diversity, and relationships. As discussed in Chapter 1 (Introduction), advances in whole-genome sequencing have revealed a hidden world of microbial life, challenged traditional classification methods, and driven the need for new tools and approaches to better understand microbial evolution (Woese et al., 1990; Gupta, 1998; Konstantinidis and Tiedje, 2005b; Thompson et al., 2013; Gupta, 2016; Parks et al., 2018; Barco et al., 2020). The availability of genomic data has enabled the identification of two important classes of novel molecular markers: Conserved Signature Indels (CSIs) and Conserved Signature Proteins (CSPs), uniquely shared by evolutionarily related groups of organisms (Gupta, 1998; Griffiths and Gupta, 2001; Gao et al., 2006; Bhandari et al., 2013; Gupta, 2016; Hu et al., 2019). Building on these advancements, my PhD research applied phylogenomic and molecular marker-based polyphasic approaches to clarify evolutionary relationships among *Pseudomonas* species, a longstanding challenge in prokaryotic taxonomy.

The genus *Pseudomonas*, belonging to the family *Pseudomonadaceae* within the class *Gammaproteobacteria*, is estimated to have evolved from Hydrobacteria approximately 1.75 billion years ago (Battistuzzi and Hedges, 2009). With over 350 recognized species listed in LPSN (accessed on March 1, 2025) (Parte et al., 2020), *Pseudomonas* is among the most diverse bacterial genera. While *P. aeruginosa* is one of the most extensively studied pathogenic species, known to cause

different diseases in humans and animals (Burrows, 2012; Huber et al., 2016; Garcia-Reyes et al., 2020; Miller and Arias, 2024), other species thrive in diverse environments, including the soil (Weller et al., 2012), water (Bollinger et al., 2020) plant surfaces (Hirano and Upper, 2000), and insect guts (Vodovar et al., 2005). Beyond its ecological diversity, *Pseudomonas* spp. have important applications in biotechnology, plant growth promotion, bioremediation, and biological control (Kwon et al., 2003; Hultberg et al., 2010b; Girard et al., 2021). However, despite the extensive phenotypic and genotypic diversity, all *Pseudomonas* species are classified within a single genus, a classification that does not accurately reflect their evolutionary relationships.

To explore the evolutionary relationships among different *Pseudomonas* species, extensive studies were conducted (Gomila et al., 2015; Garrido-Sanz et al., 2016; Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020; Girard et al., 2021; Lalucat et al., 2022). Phylogenetic analyses from these studies consistently demonstrate that *Pseudomonas* is highly polyphyletic, comprising unrelated bacterial groups clustering into three evolutionary lineages: Pertucinogena, Aeruginosa, and Fluorescens. The Aeruginosa and Fluorescens lineages contain multiple genus-level clades and subclades, with species from other genera (e.g., *Azomonas, Azotobacter*) interspersed among them. Additionally, several species (e.g., *P. cissicola, P. geniculate,* etc.) have been misclassified within the genus *Pseudomonas* (Hu et al., 1997; Anzai et al., 2000; Cutino-Jimenez et al., 2020). Based on the observed taxonomic inconsistencies, the prevailing perspective is that only members of the Aeruginosa clade (which includes the type species *P. aeruginosa*) should remain within the genus *Pseudomonas*, while other clades should be reassigned to distinct

genera (Hesse et al., 2018; Lalucat et al., 2022). This perspective is also aligned with the Code governing the Nomenclature of Prokaryotes (Parker et al., 2019) and the GTDB taxonomy (Parks et al., 2022). However, due to a lack of definitive evidence for reclassification, these diverse species remain within the genus *Pseudomonas* (Parte et al., 2020). A major challenge in the reclassification effort is precisely defining the boundaries between different clades, as phylogenetic analyses reveal a continuum of species branching in the trees. Also, the branching of species in the phylogenetic trees is constrained by several factors, such as the choice of tree-building algorithms, alignment quality, evolutionary rate models, species composition, and outgroup selection (Gupta, 1998; Gupta and Griffiths, 2002; Ludwig, 2005; Philippe et al., 2005; Gupta, 2016). Furthermore, traditional taxonomic methods have failed to provide reliable markers/properties for distinguishing closely related species or clades.

Therefore, intending to demarcate the observed *Pseudomonas* species clades reliably, we conducted comprehensive phylogenomic and comparative analyses on the available genome sequences of *Pseudomonadaceae* species from the NCBI genome database (https://www.ncbi.nlm.nih.gov/datasets/genome/). As our first approach, we constructed comprehensive phylogenetic trees using large sets of conserved genes and proteins, where *Pseudomonas* species consistently clustered into three major lineages: Pertucinogena, Aeruginosa, and Fluorescens. Multiple genus-level distinct clades are observed within the Aeruginosa and Fluorescens lineages, similar to those reported in other studies (Gomila et al., 2015; Garrido-Sanz et al., 2016; Hesse et al., 2018; Peix et al., 2018; Lalucat et al., 2020), excepting some differences resulting from including several new species in our analysis. To define the boundaries of different observed clades, our major focus (second approach) was to

conduct comparative genomics analysis on available genomic sequences of different *Pseudomonas* species for the identification of CSIs and CSPs (Gupta, 1998; Gupta, 2006; Naushad et al., 2014). Independent of the phylogenetic analyses, these markers can confirm the existence of observed clades and provide reliable means for their demarcation. As discussed in different chapters of this thesis, CSIs and CSPs originate from rare genetic changes that arose in common ancestors and were subsequently inherited by all descendants. Due to their clade-specific nature, these markers provide strong evidence of evolutionary relationships and species relatedness within a clade (Bhandari et al., 2012; Adeolu and Gupta, 2013; Naushad and Gupta, 2013; Gupta, 2014; Hu et al., 2019). Their exclusive presence within specific groups has made them a reliable tool for taxonomic classification and evolutionary studies (Gupta, 1998; Griffiths and Gupta, 2002; Gao and Gupta, 2012b; Adeolu et al., 2016; Hu et al., 2018). In addition to evolutionary and systematics studies, these signatures can also be used for biochemical and functional studies of the groups of organisms in which they are specific. Previous studies on CSIs have demonstrated their crucial roles in the organisms where they are present (Singh and Gupta, 2009). Research also indicates that CSIs are located in surface loops of the proteins, key regions involved in protein-protein or protein-ligand interactions. These interactions are often essential for the survival and functionality of CSI-containing organisms (Khadka and Gupta, 2017; Hassan and Gupta, 2018; Khadka et al., 2020).

Hence, to identify molecular markers specific to observed distinct clades within the Pertucinogena, Aeruginosa, and Fluorescens lineages, we conducted detailed comparative genomics analyses on the protein sequences from different *Pseudomonadaceae* species, which led to the identification of a large number of CSIs.

We first focused on the Pertucinogena lineage of species (presented in Chapter 2), which is distinct from all other *Pseudomonas* species groups. Our analysis identified 24 CSIs exclusively shared by all species in this clade. Based on their distinct phylogenetic positioning, shared CSIs, and salt tolerance properties, we reclassified the Pertucinogena clade as a novel genus, *Halopseudomonas*. Additionally, we identified 22 CSIs specific to *P. hussainii*, which formed a separate lineage in the phylogenetic trees. These findings (phylogenomic distinctness and shared CSIs) supported its reclassification as a new genus, *Atopomonas*. The reliability of this classification was further supported when *Atopomonas sediminilitoris* (Li et al., 2023) was described, sharing all *Atopomonas*-specific CSIs. Furthermore, based on the evidence from our studies, we reassigned several misclassified species, *P. acidophila, P. caeni, P. cissicola*, and *P. geniculate,* to the genera *Paraburkholderia, Thiopseudomonas, Xanthomonas*, and *Stenotrophomonas*, respectively. This study, presented in Chapter 2, provides a foundation for classifying *Pseudomonas* clades from the Aeruginosa lineage (presented in Chapter 3) and the Fluorescens lineage (described in Chapter 6).

In the study discussed in Chapter 3, I explored the evolutionary relationships within the Aeruginosa lineage of species. Phylogenomic analysis revealed the existence of 12 distinct clades within the Aeruginosa lineage: Aeruginosa, Alcaligenes, Anguilliseptica, Flexibilis, Fluvialis, Linyingensis, Oleovorans, Oryzihabitans, Resinovorans, Straminea, Stutzeri, and Thermotolerans clades, along with the genera *Azotobacter* and *Azomonas*, branching in-between different *Pseudomonas* species clades. These distinct clades are widely recognized as requiring reclassification into separate genera, as demonstrated by the reclassification of species

from the Oryzihabitans and Stutzeri clades into the genera *Chryseomonas* (Saati-Santamaria et al., 2021) and *Stutzerimonas* (Lalucat et al., 2022), respectively. Hence, to reclassify the rest of the clades and define their boundaries, we identified 98 CSIs uniquely shared by different clades within the *Aeruginosa* lineage, including the genera *Azomonas* and *Azotobacter*. Of the CSIs, six are identified by our analysis, uniquely shared by all species from the Aeruginosa clade, providing reliable molecular means for the demarcation/circumscription of this clade, representing the genus *Pseudomonas sensu stricto*. The rest of the CSIs are specific for distinct clades, which provide a reliable basis for delineating these clades into novel genera. We also conducted genomic similarity studies (AAI and POCP) to assess genomic relatedness, but in most cases, these values overlapped between ingroup and outgroup species, making them unreliable for genus-level demarcation. Similar limitations of AAI and POCP for genus-level demarcation have been noted in other studies (Gupta, 2019; Barco et al., 2020). Thus, based on the strong and consistent evidence provided by phylogenomic analyses and identified molecular signatures, we reclassified the species from the Alcaligenes, Fluvialis, Linyingensis, Oleovorans, Resinovorans, Straminea, and Thermotolerans clades into following novel genera *Aquipseudomonas* gen. nov., *Caenipseudomonas* gen. nov., *Geopseudomonas* gen. nov., *Ectopseudomonas* gen. nov., *Metapseudomonas* gen. nov., *Phytopseudomonas* gen. nov., and *Zestomonas* gen. nov., respectively. In addition, we also identified CSIs for the emended genera *Chryseomonas* (Oryzihabitans clade), *Serpens* (Flexibilis clade), *Stutzerimonas* (Stutzeri clade), *Azotobacter,* and *Azomonas,* providing robust molecular means for the demarcation of these genera. Notably, we did not reclassify the species from the Anguilliseptica clade for which no CSIs were identified. The

Anguilliseptica clade of species does not form a single cohesive group but instead consists of multiple distinct lineages, and further work is needed to clarify their evolutionary relationship. While reclassifying the Aeruginosa lineage greatly improves our understanding of the evolution and taxonomy of different *Pseudomonas* spp., a fully comprehensive classification system requires precise delineation of species clades within the Fluorescens lineage, as presented in Chapter 6.

The *Fluorescens* lineage includes over two-thirds of known *Pseudomonas* species. Phylogenomic analyses consistently reveal 13 distinct genus-level clades within this lineage: *Asplenii, Chlororaphis, Corrugata, Fluorescens, Fragi, Jessenii, Koreensis, Lutea, Mandelii, Massiliensis, Putida, Rhizosphaerae,* and *Syringae.* Classifying these clades as distinct genera has proven difficult due to a lack of well-defined, clade-specific characteristics (Mulet et al., 2010). Therefore, similar to other studies on Pertucinogena and Aeruginosa lineages, we identified 19 CSIs uniquely shared by species of three clades (*viz., Massiliensis, Putida,* and *Rhizosphaerae*). Our ongoing comparative genomics-based studies will identify more CSIs for demarcating the remaining 10 clades. This large number of molecular markers will offer a reliable and independent means for distinguishing and delineating these clades, supporting their potential reclassification as distinct novel genera. Besides, some single-branching species within the Aeruginosa and Fluorescens lineages (*viz., P. cavernicola, P. indica, P. kuykendallii, P. mangiferae, P. mangrovi, P. matsuisoli,* and *P. pohangensis*) will temporarily be classified as "*Pseudomonadaceae incertae sedis*". Future reclassification into novel genera will occur as more related strains or species are discovered.

While our reclassification efforts primarily focused on the genus level, as outlined in Chapters 2, 3, and 6, our findings revealed that molecular markers are also specific at the species or strain level, as discussed in Chapter 4. In this chapter, I presented the phylogenomic and comparative genomic analyses of numerous *P. aeruginosa* strains. The phylogenetic analysis identified two distinct clades: the Classical clade, which contains the Type III secretion system, and the Outlier clade, which lacks this system (Sood et al., 2019). We identified CSIs and CSPs that were exclusively shared by the strains from these two clades. Based on the markers or synapomorphies uniquely shared by these clades, widely accepted species demarcation criteria, including 16S rRNA similarity, dDDH, ANI, and key phenotypic traits, we reclassified the subset of *P. aeruginosa* strains (Outlier clade) as a new species, *P. paraeruginosa*, distinct from *P. aeruginosa* (Classical clade).

This reclassification of the genus *Pseudomonas* (into >25 distinct genera, along with one new species, *P. paraeruginosa*) aligns with other taxonomists' expectations, as Palleroni described phylogenomic-based taxonomy as a transformative “big bang” that extensively reorganizes existing genera and facilitates the reclassification into numerous new genera (Lalucat et al., 2020). Lalucat and colleagues, a leading research group on *Pseudomonas*, emphasized the importance of reclassification of the genus *Pseudomonas* into numerous novel genera by asserting that the reclassification of this genus will serve as a model in modern bacterial taxonomy, aiding in the clarification and reorganization of other bacterial genera (Lalucat et al., 2020).

A key aspect of prokaryotic classification is the preference for a stable, broadly applicable system with strong predictive capabilities (Vandamme et al., 1996;

Rosselló-Mora and Amann, 2001; Barbour et al., 2017). Based on molecular markers, our proposed classification framework meets these criteria by providing a reliable and stable taxonomic structure. The CSIs used to define different *Pseudomonas* spp. clades/genera demonstrate high predictive accuracy, consistently appearing in newly identified or sequenced members of the same clade (Dobritsa and Samadpour, 2019; Gupta et al., 2020; Gupta and Kanter-Eivin, 2023). Recently, a CSI-based tool, AppIndels.com, was developed to predict the taxonomic affiliation of uncharacterized bacterial strains (Gupta and Kanter-Eivin, 2023). Using the AppIndels web server, we explored the application of CSIs for predicting the taxonomic affiliation of ~300 uncharacterized *Pseudomonas* strains, successfully reassigning them to their respective genera. This study on the predictive ability of CSIs is presented in Chapter 5 of my thesis. One limitation of this server is that it can only determine the taxonomic affiliation of strains or species for which CSIs have been identified and uploaded in the AppIndels.com database (Gupta and Kanter-Eivin, 2023). Since we have not yet uploaded CSI information specific to different clades/genera within the Fluorescens lineage, the server cannot assign taxonomic affiliation to uncharacterized species or strains related to those groups. However, once CSIs are identified and published, we will update the server with the relevant CSI information, enabling it to predict their taxonomic placement.

The molecular signature-based classification framework we established for *Pseudomonas* provides greater reliability and stability for several key points. First, it not only accurately identifies and classifies species and strains, which is critical for clinical diagnostics and environmental monitoring, but it also effectively illustrates evolutionary relationships among different *Pseudomonas* species. Modern bacterial

taxonomy is best determined by evolutionary lineage, as phenotypic traits often fail to reflect common ancestry (Stanier and van Niel, 1962; Woese, 1992; Gupta, 1998; Sapp, 2006). Second, the molecular markers we use to distinguish different clades/genera are highly reliable, informative, and valuable in taxonomic, biochemical, and evolutionary studies (Adeolu et al., 2016; Gupta, 2016; Hu et al., 2018; Patel and Gupta, 2020; Bello et al., 2022a). Third, defining species and genus boundaries is fundamental to prokaryotic classification and systematics (Kauffmann, 1963; Konstantinidis and Tiedje, 2005a; Richter and Rossello-Mora, 2009; Qin et al., 2014; Barbour et al., 2017; Barco et al., 2020; Patel and Gupta, 2020). While species-level classification is established relying on criteria such as 16S rRNA gene similarity (98.65%), ANI (95–96%), and dDDH (70%) (Stackebrandt and Goebel, 1994; Goris et al., 2007; Kim et al., 2014), the genus-level classification remains challenging due to the lack of universally reliable methods. The limitations of existing genus demarcation approaches, such as AAI (Konstantinidis and Tiedje, 2005b) and POCP (Qin et al., 2014), have been discussed in chapters 1 and 3. In contrast, the CSI-based approach provides a more reliable and precise framework for demarcating the boundaries based on shared markers at the genus level and other higher taxonomic ranks (Gao et al., 2009a; Hu et al., 2019; Gupta et al., 2020; Patel and Gupta, 2020; Chen et al., 2021). Finally, the molecular markers we have identified offer high predictive power for classifying uncharacterized strains into their appropriate genera, further enhancing their accuracy and applicability (Gupta and Kanter-Eivin, 2023). Thus, the classification scheme developed in my work, utilizing molecular markers, provides a highly reliable, informative, and stable classification for the genus *Pseudomonas* and other *Pseudomonadaceae* genera. Figure 7.1 (Markmap) provides

a summary of the *Pseudomonas* reclassification, highlighting its overall importance as discussed in the different chapters of this thesis.

While our genome-based taxonomic revisions provide a more accurate reflection of evolutionary relationships, the implementation can have some implications across several fields, including clinical microbiology, epidemiology, laboratory diagnostics, and education (Baron and Allen, 1993; Janda, 2018; Gajdács and Urbán, 2019; Munson and Carroll, 2019; Hugenholtz et al., 2021). However, the taxonomic reclassification proposed in my study does not impact clinical microbiologists, as no taxonomic changes were made to clinically important species such as *P. aeruginosa*, the type species of the genus. Instead, we reclassified unrelated species into distinct genera based on phenotypic and genomic distinctions. This revision provides clarity, facilitating the identification of pathogenic species, species with biotechnological potential, plant pathogenicity, plant growth-promoting properties, or bioremediation capabilities. It is important to note that while validly published names hold nomenclatural standing, their adoption depends on acceptance by the scientific community. Although the taxonomic changes we made may present short-term challenges, they ultimately enhance taxonomic accuracy and improve microbiological research and applications in the long run (Gupta, 2021).

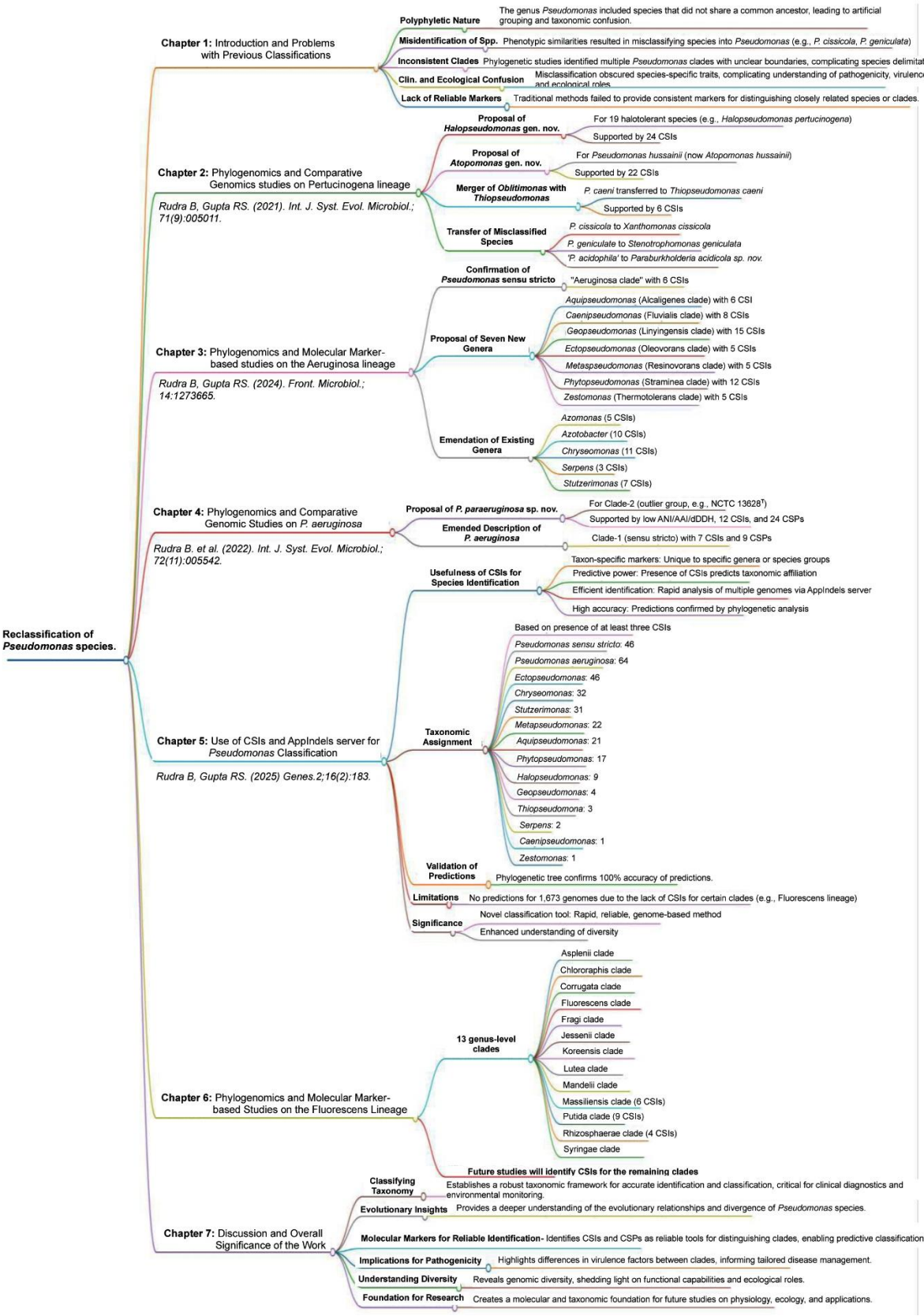

Fig. 1: Markmap summarizing the thesis chapters (Chapters 1-7).

Future Directions

My thesis presents a comprehensive approach to the taxonomic revision of the genus *Pseudomonas* using phylogenomic, molecular signatures-based polyphasic approaches. A substantial number of molecular markers in the form of CSIs and CSPs were identified, which are shared explicitly by different *Pseudomonas* species clades/genera. These taxonomic synapomorphies can be efficiently utilized through the CSI-based web server AppIndels.com to classify uncharacterized strains/species.

Beyond their role in evolutionary and systematic studies, these markers have valuable diagnostic applications. Their high sequence conservation enables the design of polymerase chain reaction (PCR) primers, allowing for precise and reliable amplification of CSI and CSP-containing DNA regions (Griffiths and Gupta, 2002; Gao and Gupta, 2005). Previously, molecular assays utilizing these markers have been successfully used to differentiate *Bacillus anthracis* from *Bacillus cereus* (Ahmod et al., 2011) and to identify enterohemorrhagic *E. coli* O157:H7 (Wong et al., 2014). Additionally, a CSPs-based assay was developed to improve the monitoring of recreational water quality, providing enhanced detection of *E. coli* strains that are otherwise challenging to distinguish using conventional methods (Saleem et al., 2024). In addition to taxonomic and diagnostic applications, these markers represent promising targets for functional studies. Previous research has demonstrated that CSIs play essential roles in protein function within the bacterial groups in which they are found. Disruptions or deletions in these regions have been shown to impair protein activity, leading to loss of cellular function (Singh and Gupta, 2009). Structural analyses further indicate that CSIs are predominantly located within surface loops of proteins, away

from active sites, where they are likely involved in mediating specific interactions with other proteins (Akiva et al., 2008; Gupta et al., 2017; Hassan and Gupta, 2018; Khadka et al., 2020; Miton and Tokuriki, 2022). Investigating the functional relevance of CSIs and CSPs identified in this study could lead to the discovery of novel biological mechanisms and provide deeper insights into bacterial physiology, evolution, and adaptation strategies.

Conclusions

The rapid expansion of genomic sequencing data has revolutionized our understanding of evolutionary relationships among organisms. My graduate research focuses on analyzing the available genome sequences of *Pseudomonas* species to robustly elucidate their evolutionary relationships using multiple independent approaches such as phylogenetic analysis based on several large data sets of conserved proteins, overall genomic similarity studies using AAI and POCP matrices and the identification of molecular markers such as CSIs and CSPs specific for different *Pseudomonas* species clades supported by other methods. Using these approaches, a robust phylogenetic framework for *Pseudomonas* species has been established. Under this framework, the genus *Pseudomonas* is now proposed to restrict only to species within the *Aeruginosa* clade, while other *Pseudomonas* species have been reclassified into >25 distinct genera, accurately reflecting their evolutionary relationships. Furthermore, the molecular markers used in this reclassification serve as predictive tools for determining the taxonomic placement of uncharacterized strains or species. Beyond their significance in evolutionary studies, these conserved signatures hold great potential as diagnostic markers for identifying specific groups of organisms. Further exploration of these molecular features may uncover novel biological functions and adaptations, providing deeper insights into microbial physiology and evolutionary history.

BIBLIOGRAPHY

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, S. (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. *Int J Syst Evol Microbiol* 66**,** 5575-5599.

Adeolu, M., and Gupta, R.S. (2013). Phylogenomics and molecular signatures for the order *Neisseriales*: proposal for division of the order *Neisseriales* into the emended family *Neisseriaceae* and *Chromobacteriaceae* fam. nov. *Antonie van Leeuwenhoek* 104**,** 1-24.

Ahmod, N.Z., Gupta, R.S., and Shah, H.N. (2011). Identification of a *Bacillus anthracis* specific indel in the yeaC gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* 87**,** 278-285.

Akiva, E., Itzhaki, Z., and Margalit, H. (2008). Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. *Proc. Natl. Acad. Sci. U. S. A* 105**,** 13292-13297.

Andersen, S.M., Johnsen, K., Sørensen, J., Nielsen, P., and Jacobsen, C.S. (2000). *Pseudomonas frederiksbergensis* sp. nov., isolated from soil at a coal gasification site. *Int J Syst Evol Microbiol* 50**,** 1957-1964.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55**,** 539-552.

Ankenbrand, M.J., and Keller, A. (2016). bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome* 59**,** 783-791.

Anzai, Y., Kim, H., Park, J.Y., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int J Syst Evol Microbiol* 50 Pt 4**,** 1563-1589.

Anzai, Y., Kudo, Y., and Oyaizu, H. (1997). The phylogeny of the genera *Chryseomonas*, *Flavimonas*, and *Pseudomonas* supports synonymy of these three genera. *Int J Syst Bacteriol* 47**,** 249-251.

Baldauf, S.L. (2003). Phylogeny for the faint of heart: a tutorial. *Trends Genet* 19**,** 345-351.

Baldauf, S.L., and Palmer, J.D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* 90**,** 11558-11562.

Banerjee, S., Bedics, A., Toth, E., Kriszt, B., Soares, A.R., Boka, K., and Tancsics, A. (2022). Isolation of *Pseudomonas aromaticivorans* sp. nov from a hydrocarbon-contaminated groundwater capable of degrading benzene-, toluene-, m- and p-xylene under microaerobic conditions. *Front Microbiol* 13**,** 929128.

Bapteste, E., O'malley, M.A., Beiko, R.G., Ereshefsky, M., Gogarten, J.P., Franklin-Hall, L., Lapointe, F.-J., Dupré, J., Dagan, T., and Boucher, Y. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct.* 4**,** 1-20.

Barbour, A.G., Adeolu, M., and Gupta, R.S. (2017). Division of the genus *Borrelia* into two genera (corresponding to Lyme disease and relapsing fever groups) reflects their genetic and phenotypic distinctiveness and will lead to a better understanding of these two groups of microbes (Margos et al.(2016) There is inadequate evidence to support the division of the genus *Borrelia*. Int. J. Syst. Evol. Microbiol. doi: 10.1099/ijsem. 0.001717). *Int J Syst Evol Microbiol* 67**,** 2058-2067.

Barco, R.A., Garrity, G.M., Scott, J.J., Amend, J.P., Nealson, K.H., and Emerson, D. (2020). A Genus Definition for Bacteria and Archaea Based on a Standard Genome Relatedness Index. *mBio* 11.

Bardet, L., Cimmino, T., Buffet, C., Michelle, C., Rathored, J., Tandina, F., Lagier, J.-C., Khelaifia, S., Abrahão, J., and Raoult, D. (2018). Microbial Culturomics Application for Global Health: Noncontiguous Finished Genome Sequence and Description of *Pseudomonas massiliensis* Strain CB-1T sp. nov. in Brazil. *OMICS: A Journal of Integrative Biology* 22**,** 164-175.

Baron, E.J., and Allen, S.D. (1993). Should clinical laboratories adopt new taxonomic changes? If so, when? *Clin. Infect. Dis.* 16**,** S449-S450.

Basso, P., Ragno, M., Elsen, S., Reboud, E., Golovkine, G., Bouillot, S., Huber, P., Lory, S., Faudry, E., and Attree, I. (2017). *Pseudomonas aeruginosa* Pore-Forming Exolysin and Type IV Pili Cooperate To Induce Host Cell Lysis. *mBio* 8.

Battistuzzi, F.U., and Hedges, S.B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* 26**,** 335-343.

Beiki, F., Busquets, A., Gomila, M., Rahimian, H., Lalucat, J., and Garcia-Valdes, E. (2016). New *Pseudomonas* spp. Are Pathogenic to Citrus. *PLoS One* 11**,** e0148796.

Bello, S., Howard-Azzeh, M., Schellhorn, H.E., and Gupta, R.S. (2022a). Phylogenomic Analyses and Molecular Signatures Elucidating the Evolutionary Relationships amongst the *Chlorobia* and *Ignavibacteria* Species: Robust Demarcation of Two Family-Level Clades within the Order *Chlorobiales* and Proposal for the Family *Chloroherpetonaceae* fam. nov. *Microorganisms* 10.

Bello, S., Rudra, B., and Gupta, R.S. (2022b). Phylogenomic and comparative genomic analyses of *Leuconostocaceae* species: identification of molecular signatures specific for the genera *Leuconostoc, Fructobacillus* and *Oenococcus* and proposal for a novel genus *Periweissella* gen. nov. *Int J Syst Evol Microbiol* 72**,** 005284.

Bergey Dh, H.F., Breed Rs, Hammer Bw, Huntoon Fm. (1923). Bergey's Manual of Determinative Bacteriology Baltimore, MD. *Williams & Wilkins*.

Bhandari, V., Ahmod, N.Z., Shah, H.N., and Gupta, R.S. (2013). Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *Int J Syst Evol Microbiol* 63**,** 2712-2726.

Bhandari, V., Naushad, H.S., and Gupta, R.S. (2012). Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution. *Front. Cell. Infect. Microbiol.*. 2**,** 98.

Bininda-Emonds, O.R. (2004). *Phylogenetic supertrees: combining information to reveal the tree of life.* Springer Science & Business Media.

Bollinger, A., Thies, S., Katzke, N., and Jaeger, K.E. (2020). The biotechnological potential of marine bacteria in the novel lineage of *Pseudomonas pertucinogena*. *Microb. Biotechnol.* 13**,** 19-31.

Boucher, Y., Douady, C.J., Sharma, A.K., Kamekura, M., and Doolittle, W.F. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186**,** 3980-3990.

Bourdichon, F., Casaregola, S., Farrokh, C., Frisvad, J.C., Gerds, M.L., Hammes, W.P., Harnett, J., Huys, G., Laulund, S., and Ouwehand, A. (2012). Food fermentations: microorganisms with technological beneficial use. *Int. J. Food Microbiol.* 154**,** 87-97.

Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E., and Stanhope, M.J. (2001). Universal trees based on large combined protein sequence data sets. *Nat. Genet.* 28**,** 281-285.

Brown, S.D., Utturkar, S.M., Klingeman, D.M., Johnson, C.M., Martin, S.L., Land, M.L., Lu, T.-Y.S., Schadt, C.W., Doktycz, M.J., and Pelletier, D.A. (2012). Twenty-One Genome Sequences from Pseudomonas Species and 19 Genome Sequences from Diverse Bacteria Isolated from the Rhizosphere and Endosphere of Populus deltoides. *J. Bacteriol.* 194**,** 5991-5993.

Buchanan, R.E. (1925). *General systematic bacteriology: history, nomenclature, groups of bacteria.* Williams & Wilkins.

Buchanan, R.E. (1955). Taxonomy. *Annu Rev Microbiol* 9**,** 1-20.

Buchanan, R.E., and Gibbons, N.E. (1974). *Bergey's manual of determinative bacteriology.* Baltimore: Williams & Wilkins.

Burrows, L.L. (2012). *Pseudomonas aeruginosa* twitching motility: type IV pili in action. *Annu Rev Microbiol* 66**,** 493-520.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinform.* 25**,** 1972-1973.

Chen, S., Rudra, B., and Gupta, R.S. (2021). Phylogenomics and molecular signatures support division of the order *Neisseriales* into emended families *Neisseriaceae* and *Chromobacteriacea*e and three new families *Aquaspirillaceae* fam. nov., *Chitinibacteraceae* fam. nov., and *Leeiaceae* fam. nov. *Syst Appl Microbiol* 44**,** 126251.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31**,** 3497-3500.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., Da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.-W., and De Meyer, S. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 68**,** 461-466.

Ciccarelli, F.D., Doerks, T., Von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311**,** 1283-1287.

Cody, A.J., Bennett, J.S., and Maiden, M.C. (2014). "Multi-locus sequence typing and the gene-by-gene approach to bacterial classification and analysis of population variation," in *Methods in microbiology*. Elsevier), 201-219.

Coenye, T., Gevers, D., Van De Peer, Y., Vandamme, P., and Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* 29**,** 147-167.

Cohn, F. (1872). Grundzuge einer neuen naturlichen Anordnung der Kryptogamischen Pflanzen. *Jahrb. Schles. Ges. Vaterl. Kult.* 49**,** 83-89.

Cohn, F. (1875). Untersuchungen über Bakterien. II. *Beitr. Biol. Pflanz* 3**,** 141-208.

Cowan, S., and Liston, J. (1974). The mechanism of identification. *by Buchanan, RE and Gibbons, NE, Williams & Wilkins Co., Baltimore, MD***,** 10-13.

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227**,** 561-563.

Crick, F.H. (1958). On protein synthesis. *Symp Soc Exp Biol* 12**,** 138-163.

Cutino-Jimenez, A.M., Menck, C.F.M., Cambas, Y.T., and Diaz-Perez, J.C. (2020). Protein signatures to identify the different genera within the *Xanthomonadaceae* family. *Braz. J Microbiol* 51**,** 1515-1526.

D'agata, E. (2015). "*Pseudomonas aeruginosa* and other *Pseudomonas* species," in *Mandell, Douglas, and Bennett's principles and practice of infectious diseases*. Elsevier), 2518-2531. e2513.

Danchin, A. (2003). Genomes and evolution. *Curr Issues Mol Biol* 5**,** 37-42.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray.

Daubin, V., Gouy, M., and Perriere, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12**,** 1080-1090.

David, B.V., Chandrasehar, G., and Selvam, P.N. (2018). "*Pseudomonas fluorescens*: a plant-growth-promoting rhizobacterium (PGPR) with potential role in biocontrol of pests of crops," in *Crop improvement through microbial biotechnology*. Elsevier), 221-243.

Dayhoff, M., Barker, W., and Mclaughlin, P. (Year). "Inferences from protein and nucleic acid sequences: early molecular evolution, divergence of kingdoms and rates of change", in: *Cosmochemical Evolution and the Origins of Life: Proceedings of the Fourth International Conference on the Origin of Life and the First Meeting of the International Society for the Study of the Origin of Life, Barcelona, June 25–28, 1973, Volume I: Invited Papers and Volume II: Contributed Papers*: Springer), 311-330.

Deloger, M., El Karoui, M., and Petit, M.A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J Bacteriol* 191**,** 91-99.

Den Dooren De Jong, L.E. (1926). Bijdrage tot de kennis van het mineralisatieproces.

Dobritsa, A.P., and Samadpour, M. (2019). Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *Int J Syst Evol Microbiol* 69**,** 2057-2063.

Doolittle, W.F. (1999). Phylogenetic classification and the universal tree. *Science* 284**,** 2124-2128.

Doudoroff, M. (1974). Gram-negative aerobic rods and cocci. Family I: *Pseudomonaceae*. Genus I: *Pseudomonas*. *Bergey's manual of determinative microbiology*.

Dutilh, B.E., He, Y., Hekkelman, M.L., and Huynen, M.A. (2008a). Signature, a web server for taxonomic characterization of sequence samples using signature genes. *Nucleic Acids Res* 36**,** W470-W474.

Dutilh, B.E., Snel, B., Ettema, T.J., and Huynen, M.A. (2008b). Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* 25**,** 1659-1667.

Dutta, B., Gitaitis, R., Agarwal, G., Coutinho, T., and Langston, D. (2018). *Pseudomonas coronafaciens* sp. nov., a new phytobacterial species diverse from *Pseudomonas syringae*. *PLoS One* 13**,** e0208271.

Eck, R.V., and Dayhoff, M.O. (1966). Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. *Science* 152**,** 363-366.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5**,** 113.

Efron, B. (1992). "Bootstrap methods: another look at the jackknife," in *Breakthroughs in statistics: Methodology and distribution*. Springer), 569-593.

Ehrenberg, C.G. (1838). Die Infusionsthierchen als vollkommene Organismen. Ein Blick in das tiefere organische Leben der Natur. *(No Title)*.

El-Solh, A.A., Hattemer, A., Hauser, A.R., Alhajhusain, A., and Vora, H. (2012). Clinical outcomes of type III *Pseudomonas aeruginosa* bacteremia. *Crit Care Med* 40**,** 1157-1163.

Elsen, S., Huber, P., Bouillot, S., Coute, Y., Fournier, P., Dubois, Y., Timsit, J.F., Maurin, M., and Attree, I. (2014). A type III secretion negative clinical strain of *Pseudomonas aeruginosa* employs a two-partner secreted exolysin to induce hemorrhagic pneumonia. *Cell Host Microbe* 15**,** 164-176.

Engel, J., and Balachandran, P. (2009). Role of *Pseudomonas aeruginosa* type III effectors in disease. *Curr Opin Microbiol* 12**,** 61-66.

Fan, D., and Smith, D.L. (2021). Characterization of Selected Plant Growth-Promoting Rhizobacteria and Their Non-Host Growth Promotion Effects. *Microbiol Spectr* 9**,** e0027921.

Fang, G., Rocha, E.P., and Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. *BMC genomics* 9**,** 1-14.

Farías-Rico, J.A., and Mourra-Díaz, C.M. (2022). A short tale of the origin of proteins and ribosome evolution. *Microorganisms* 10**,** 2115.

Felsenstein, J. (1981). Evolutionary Trees from Gene Frequencies and Quantitative Characters: Finding Maximum Likelihood Estimates. *Evol.* 35**,** 1229-1242.

Felsenstein, J. (2004). *Inferring phylogenies.* Sunderland, Mass.: Sinauer Associates, Inc.

Filloux, A. (2011). Protein secretion systems in Pseudomonas aeruginosa: an essay on diversity, evolution, and function. *Front. microbiol.* 2**,** 155.

Fitch, W.M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol.* 20**,** 406-416.

Fitch, W.M., and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* 155**,** 279-284.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., and Merrick, J.M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science* 269**,** 496-512.

Fox, G.E., Pechman, K.R., and Woese, C.R. (1977). Comparative cataloging of 16 S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. *Int J Syst Evol Microbiol.* 27**,** 44-57.

Fox, G.E., Wisotzkey, J.D., and Jurtshuk, P., Jr. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 42**,** 166-170.

Freschi, L., Bertelli, C., Jeukens, J., Moore, M.P., Kukavica-Ibrulj, I., Emond-Rheault, J.G., Hamel, J., Fothergill, J.L., Tucker, N.P., Mcclean, S., Klockgether, J., De Soyza, A., Brinkman, F.S.L., Levesque, R.C., and Winstanley, C. (2018). Genomic characterisation of an international *Pseudomonas aeruginosa* reference panel indicates that the two major groups draw upon distinct mobile gene pools. *FEMS Microbiol Lett* 365.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinform.* 28**,** 3150-3152.

Furmanczyk, E.M., Lipinski, L., Dziembowski, A., and Sobczak, A. (2018). Genomic and Functional Characterization of Environmental Strains of SDS-Degrading *Pseudomonas* spp., Providing a Source of New Sulfatases. *Front Microbiol* 9**,** 1795.

Gadagkar, S.R., Rosenberg, M.S., and Kumar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304**,** 64-74.

Gajdács, M., and Urbán, E. (2019). Resistance trends and epidemiology of *citrobacter-enterobacter-serratia* in urinary tract infections of inpatients and outpatients (RECESUTI): a 10-year survey. *Medicina* 55**,** 285.

Gao, B., and Gupta, R.S. (2005). Conserved indels in protein sequences that are characteristic of the phylum *Actinobacteria*. *Int J Syst Evol Microbiol* 55**,** 2401-2412.

Gao, B., and Gupta, R.S. (2012a). Microbial systematics in the post-genomics era. *Antonie van Leeuwenhoek* 101**,** 45-54.

Gao, B., and Gupta, R.S. (2012b). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. rev.* 76**,** 66-112.

Gao, B., Mohan, R., and Gupta, R.S. (2009a). Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria*. *Int J Syst Evol Microbiol* 59**,** 234-247.

Gao, B., Paramanathan, R., and Gupta, R.S. (2006). Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie Van Leeuwenhoek* 90**,** 69-91.

Gao, B., Sugiman-Marangos, S., Junop, M.S., and Gupta, R.S. (2009b). Structural and phylogenetic analysis of a conserved actinobacteria-specific protein (ASP1; SCO1997) from Streptomyces coelicolor. *BMC Struct. Biol.* 9**,** 1-13.

Garcia-Reyes, S., Soberon-Chavez, G., and Cocotl-Yanez, M. (2020). The third quorum-sensing system of *Pseudomonas aeruginosa*: *Pseudomonas* quinolone signal and the enigmatic PqsE protein. *J Med Microbiol* 69**,** 25-34.

Gardan, L., Bella, P., Meyer, J.-M., Christen, R., Rott, P., Achouak, W., and Samson, R. (2002). *Pseudomonas salomonii* sp. nov., pathogenic on garlic, and *Pseudomonas palleroniana* sp. nov., isolated from rice. *Int J Syst Evol Microbiol* 52**,** 2065-2074.

Gardan, L., Bollet, C., Ghorrah, M.A., Grimont, F., and Grimont, P. (1992). DNA Relatedness among the Pathovar Strains of *Pseudomonas syringae* subsp. savastanoi Janse (1982) and Proposal of *Pseudomonas savastanoi* sp. nov. *Int J Syst Evol Microbiol* 42**,** 606-612.

Garrido-Sanz, D., Arrebola, E., Martinez-Granero, F., Garcia-Mendez, S., Muriel, C., Blanco-Romero, E., Martin, M., Rivilla, R., and Redondo-Nieto, M. (2017). Classification of Isolates from the *Pseudomonas fluorescens* Complex into Phylogenomic Groups Based in Group-Specific Markers. *Front Microbiol* 8**,** 413.

Garrido-Sanz, D., Meier-Kolthoff, J.P., Goker, M., Martin, M., Rivilla, R., and Redondo-Nieto, M. (2016). Genomic and Genetic Diversity within the *Pseudomonas fluorescens* Complex. *PLoS. One* 11**,** e0150183.

Garrido-Sanz, D., Redondo-Nieto, M., Martin, M., and Rivilla, R. (2021). Comparative genomics of the *Pseudomonas corrugata* subgroup reveals high species diversity and allows the description of *Pseudomonas ogarae* sp. nov. *Microb Genom* 7.

Garrity, G.M. (2010). Browser Tool takes expertise out the database and puts it right in the browser. *Microbiol Today* 2**,** 9.

Garrity, G.M. (2016). A new genomics-driven taxonomy of bacteria and archaea: are we there yet? *J. Clin. Microbiol.* 54**,** 1956-1963.

Garrity, G.M., Bell, J.A., and Lilburn, T. (2005). "The revised road map to the manual," in *Bergey's Manual® of Systematic Bacteriology*. Springer), 159-187.

Gevers, D., Dawyndt, P., Vandamme, P., Willems, A., Vancanneyt, M., Swings, J., and De Vos, P. (2006). Stepping stones towards a new prokaryotic taxonomy. *Philos Trans R Soc Lond B Biol Sci* 361**,** 1911-1916.

Gieschler, S., Fiedler, G., Böhnlein, C., Grimmler, C., Franz, C.M., and Kabisch, J. (2021). *Pseudomonas kielensis* sp. nov. and *Pseudomonas baltica* sp. nov., isolated from raw milk in Germany. *Int J Syst Evol Microbiol* 71**,** 004717.

Girard, L., Lood, C., Hofte, M., Vandamme, P., Rokni-Zadeh, H., Van Noort, V., Lavigne, R., and De Mot, R. (2021). The Ever-Expanding *Pseudomonas* Genus: Description of 43 New Species and Partition of the *Pseudomonas putida* Group. *Microorganisms* 9.

Girard, L., Lood, C., Rokni-Zadeh, H., Van Noort, V., Lavigne, R., and De Mot, R. (2020). Reliable identification of environmental *Pseudomonas* isolates using the *rpoD* gene. *Microorganisms* 8**,** 1166.

Glaeser, S.P., and Kampfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* 38**,** 237-245.

Gomila, M., Busquets, A., Mulet, M., García-Valdés, E., and Lalucat, J. (2017). Clarification of taxonomic status within the *Pseudomonas syringae* species group based on a phylogenomic analysis. *Front. microbiol.* 8**,** 2422.

Gomila, M., Pena, A., Mulet, M., Lalucat, J., and Garcia-Valdes, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* 6**,** 214.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol* 57**,** 81-91.

Gram, C. (1884). Ueber die isolirte Farbung der Schizomyceten in Schnitt-und Trockenpraparaten. *Fortschritte der Medicin* 2**,** 185-189.

Gray, M.W., Sankoff, D., and Cedergren, R.J. (1984). On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Res.* 12**,** 5837-5852.

Griffiths, E., and Gupta, R.S. (2001). The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that *Fibrobacter* diverged at a similar time to *Chlamydia* and the Cytophaga–Flavobacterium–Bacteroides division. *Microbiol.* 147**,** 2611-2622.

Griffiths, E., and Gupta, R.S. (2002). Protein signatures distinctive of chlamydial species: Horizontal transfer of cell wall biosynthesis genes *glmU* from *Archaebacteria* to *Chlamydiae*, and *murA* between *Chlamydiae* and *Streptomyces*. *Microbiology* 148**,** 2541-2549.

Griffiths, E., and Gupta, R.S. (2004). Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the Deinococcus-Thermus phylum. *J. Bacteriol.* 186**,** 3097-3107.

Griffiths, E., and Gupta, R.S. (2006a). Lateral transfers of serine hydroxymethyl transferase (*glyA*) and UDP-N-acetylglucosamine enolpyruvyl transferase (*murA*) genes from free-living *Actinobacteria* to the parasitic chlamydiae. *J. Mol. Evol* 63**,** 283-296.

Griffiths, E., and Gupta, R.S. (2006b). Molecular signatures in protein sequences that are characteristics of the phylum *Aquificae*. *Int J Syst Evol Microbiol* 56**,** 99-107.

Griffiths, E., Ventresca, M.S., and Gupta, R.S. (2006). BLAST screening of chlamydial genomes to identify signature proteins that are unique for the *Chlamydiales, Chlamydiaceae*, Chlamydophila and Chlamydia groups of species. *BMC genomics* 7**,** 1-20.

Grimont, P.A., Popoff, M.Y., Grimont, F., Coynault, C., and Lemelin, M. (1980). Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr. Microbiol.* 4**,** 325-330.

Gupta, R.S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev* 62**,** 1435-1491.

Gupta, R.S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Rev.* 24**,** 367-402.

Gupta, R.S. (2006). Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (Campylobacterales). *BMC Genomics* 7**,** 167.

Gupta, R.S. (2014). "Identification of Conserved Indels that are Useful for Classification and Evolutionary Studies," in *Bacterial Taxonomy, Methods in Microbiology Volume 41,* eds. M. Goodfellow, I.C. Sutcliffe & J. Chun. (London: Elsevier), 153-182.

Gupta, R.S. (2016). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol Rev* 40**,** 520-553.

Gupta, R.S. (2019). Distinction between *Borrelia* and *Borreliella* is more robustly supported by molecular and phenotypic characteristics than all other neighbouring prokaryotic genera: Response to Margos' et al. "The genus *Borrelia* reloaded" (PLoS ONE 13(12): e0208432). *PLoS One* 14**,** e0221397.

Gupta, R.S. (2021). Microbial Taxonomy: How and Why Name Changes Occur and Their Significance for (Clinical) Microbiology. *Clin. Chem.* 68**,** 134-137.

Gupta, R.S., and Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor. Popul. Biol.* 61**,** 423-434.

Gupta, R.S., and Griffiths, E. (2006). *Chlamydiae*-specific proteins and indels: novel tools for studies. *Trends Microbiol* 14**,** 527-535.

Gupta, R.S., and Kanter-Eivin, D.A. (2023). AppIndels.com Server: A Web Based Tool for the Identification of Known Taxon-Specific Conserved Signature Indels in Genome Sequences: Validation of Its Usefulness by Predicting the Taxonomic Affiliation of >700 Unclassified strains of *Bacillus* Species. *Int J Syst and Evol Microbiol* 73:**,** 005844.

Gupta, R.S., Lo, B., and Son, J. (2018). Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium i*nto an Emended Genus *Mycobacterium* and Four Novel Genera. *Front Microbiol* 9**,** 67.

Gupta, R.S., and Mathews, D.W. (2010). Signature proteins for the major clades of *Cyanobacteria*. *BMC Evol. Biol.* 10**,** 1-20.

Gupta, R.S., Nanda, A., and Khadka, B. (2017). Novel molecular, structural and evolutionary characteristics of the phosphoketolases from bifidobacteria and Coriobacteriales. *PLoS One* 12**,** e0172176.

Gupta, R.S., Naushad, S., and Baker, S. (2015a). Phylogenomic analyses and molecular signatures for the class *Halobacteria* and its two major clades: a proposal for division of the class *Halobacteria* into an emended order *Halobacteriales* and two new orders, *Haloferacales* ord. nov. and *Natrialbales* ord. nov., containing the novel families *Haloferacaceae* fam. nov. and *Natrialbaceae* fam. nov. *Int J Syst Evol Microbiol* 65**,** 1050-1069.

Gupta, R.S., Naushad, S., Chokshi, C., Griffiths, E., and Adeolu, M. (2015b). A phylogenomic and molecular markers based analysis of the phylum *Chlamydiae*: proposal to divide the class *Chlamydiia* into two orders, *Chlamydiales* and *Parachlamydiales* ord. nov., and emended description of the class *Chlamydiia*. *Antonie Van Leeuwenhoek* 108**,** 765-781.

Gupta, R.S., Patel, S., Saini, N., and Chen, S. (2020). Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus

Bacillus limiting it only to the members of the Subtilis and Cereus clades of species. *Int J Syst Evol Microbiol* 70**,** 5753-5798.

Haeckel, E. (1866). *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformierte Descendenz-Theorie. Band 1: Allgemeine Anatomie. Band 2: Allgemeine Entwicklungsgeschichte.* de Gruyter.

Haenni, M., Hocquet, D., Ponsin, C., Cholley, P., Guyeux, C., Madec, J.Y., and Bertrand, X. (2015). Population structure and antimicrobial susceptibility of *Pseudomonas aeruginosa* from animal infections in France. *BMC Vet Res* 11**,** 9.

Hall, B.D., and Spiegelman, S. (1961). Sequence complementarity of T2-DNA and T2-specific RNA. *Proc Natl Acad Sci U S A* 47**,** 137-146.

Hall, D., Cammack, R., and Rao, K. (1973). Ferredoxins in the evolution of photosynthetic systems from anaerobic bacteria to higher plants. *Space life sciences* 4**,** 455-468.

Harada, K., Arima, S., Niina, A., Kataoka, Y., and Takahashi, T. (2012). Characterization of *Pseudomonas aeruginosa* isolates from dogs and cats in Japan: current status of antimicrobial resistance and prevailing resistance mechanisms. *Microbiol. Immunol.* 56**,** 123-127.

Hasegawa, M., and Hashimoto, T. (1993). Ribosomal RNA trees misleading? *Nature* 361**,** 23-23.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K.-I., and Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, Giardia lamblia. *Mol. Biol. Evol.* 11**,** 65-71.

Hassan, F.N., and Gupta, R.S. (2018). Novel sequence features of DNA repair genes/proteins from *Deinococcus* species implicated in protection from oxidatively generated damage. *Genes* 9**,** 149.

Hassen, W., Neifar, M., Cherif, H., Najjari, A., Chouchane, H., Driouich, R.C., Salah, A., Naili, F., Mosbah, A., and Souissi, Y. (2018). *Pseudomonas rhizophila* S211, a new plant growth-promoting rhizobacterium with potential in pesticide-bioremediation. *Front Microbiol* 9**,** 34.

Hauser, A.R. (2009). The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Microbiol* 7**,** 654-665.

Hemm, M.R., Paul, B.J., Schneider, T.D., Storz, G., and Rudd, K.E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.* 70**,** 1487-1501.

Henz, S.R., Huson, D.H., Auch, A.F., Nieselt-Struwe, K., and Schuster, S.C. (2005). Whole-genome prokaryotic phylogeny. *Bioinform.* 21**,** 2329-2335.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36**,** 39-56.

Hesse, C., Schulz, F., Bull, C.T., Shaffer, B.T., Yan, Q., Shapiro, N., Hassan, K.A., Varghese, N., Elbourne, L.D.H., Paulsen, I.T., Kyrpides, N., Woyke, T., and Loper, J.E. (2018). Genome-based evolutionary history of *Pseudomonas* spp. *Environ Microbiol* 20**,** 2142-2159.

Hilario, E., Buckley, T.R., and Young, J.M. (2004). Improved resolution on the phylogenetic relationships among *Pseudomonas* by the combined analysis of atp*D, carA, recA* and 16S rDNA. *Antonie Van Leeuwenhoek* 86**,** 51-64.

Hirano, S.S., and Upper, C.D. (2000). Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*—a pathogen, ice nucleus, and epiphyte. *Microbiol. Mol. Biol. Rev* 64**,** 624-653.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., and Plemel, R.L. (2010). A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell host & microbe* 7**,** 25-37.

Hu, D., Cha, G., and Gao, B. (2018). A phylogenomic and molecular markers based analysis of the class *Acidimicrobiia*. *Front Microbiol* 9**,** 987.

Hu, D., Zang, Y., Mao, Y., and Gao, B. (2019). Identification of molecular markers that are specific to the class *Thermoleophilia*. *Front. microbiol.* 10**,** 1185.

Hu, F.P., Young, J.M., Stead, D.E., and Goto, M. (1997). Transfer of *Pseudomonas cissicola* (Takimoto 1939) Burkholder 1948 to the genus *Xanthomonas*. *Int J Syst Evol Microbiol* 47**,** 228-230.

Huber, P., Basso, P., Reboud, E., and Attree, I. (2016). *Pseudomonas aeruginosa* renews its virulence factors. *Environ Microbiol Rep* 8**,** 564-571.

Huddleson, I. (1947). Proceedings of 4th International Congress on Microbiology.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., and Ise, K. (2016). A new view of the tree of life. *Nat. Microbiol* 1**,** 1-6.

Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D.H., and Soo, R.M. (2021). Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.*

Hughes, J., Loy, R., Asbury, A., and Burd, H.E. (1966). The occurrence of *Pseudomonas* in the reproductive tract of mares and its effect on fertility. *Cornell Vet.* 56**,** 595-610.

Hultberg, M., Bengtsson, T., and Liljeroth, E. (2010b). Late blight on potato is suppressed by the biosurfactant-producing strain *Pseudomonas koreensis* 2.74 and its biosurfactant. *BioControl* 55**,** 543-550.

Huss, V.A., Festl, H., and Schleifer, K.H. (1983). Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Syst Appl Microbiol.* 4**,** 184-192.

Iyer, R., and Damania, A. (2016). Draft genome sequence of *Pseudomonas putida* CBF10-2, a soil isolate with bioremediation potential in agricultural and industrial environmental settings. *Genome Announc.* 4**,** 10.1128/genomea. 00670-00616.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., and Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36**,** 338-345.

Janda, J.M. (2018). Clinical decisions: how relevant is modern bacterial taxonomy for clinical microbiologists? *Clin. Microbiol. Newsl.* 40**,** 51-57.

Janda, J.M., and Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin. Microbiol* 45**,** 2761-2764.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23**,** 403-405.

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *TRENDS in Genetics* 22**,** 225-231.

Johnsson, L., Hökeberg, M., and Gerhardson, B. (1998). Performance of the *Pseudomonas chlororaphis* biocontrol agent MA 342 against cereal seed-borne diseases in field experiments. *Eur. J. Plant Pathol.* 104**,** 701-711.

Jolley, K.A., Chan, M.S., and Maiden, M.C. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinform.* 5**,** 86.

Jun, S.R., Wassenaar, T.M., Nookaew, I., Hauser, L., Wanchai, V., Land, M., Timm, C.M., Lu, T.Y., Schadt, C.W., Doktycz, M.J., Pelletier, D.A., and Ussery, D.W. (2016). Diversity of *Pseudomonas* Genomes, Including Populus-Associated Isolates, as Revealed by Comparative Genome Analysis. *Appl Environ Microbiol* 82**,** 375-383.

Jyot, J., Balloy, V., Jouvion, G., Verma, A., Touqui, L., Huerre, M., Chignard, M., and Ramphal, R. (2011). Type II secretion system of Pseudomonas aeruginosa: in vivo evidence of a significant role in death due to lung infection. *J Infect Dis* 203**,** 1369-1377.

Kämpfer, P. (2012). Systematics of prokaryotes: the state of the art. *Antonie Van Leeuwenhoek* 101**,** 3-11.

Kampfer, P., and Glaeser, S.P. (2011). Prokaryotic taxonomy in the sequencing era and the role of MLSA in classification. *Microbiol. Aust.* 32**,** 66-70.

Karlin, S., Weinstock, G.M., and Brendel, V. (1995). Bacterial classifications derived from recA protein sequence comparisons. *J Bacteriol* 177**,** 6881-6893.

Kasting, J.F. (1993). Earth's early atmosphere. *Science* 259**,** 920-926.

Katoh, K., Kuma, K., Miyata, T., and Toh, H. (2005). Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform* 16**,** 22-33.

Kauffmann, F. (1963). On the species-definition. *Int J Syst Evol Microbiol* 13**,** 181-186.

Kavanagh, K.L., Klimacek, M., Nidetzky, B., and Wilson, D.K. (2002). Crystal Structure of *Pseudomonas fluorescens* Mannitol 2-Dehydrogenase Binary and Ternary Complexes: SPECIFICITY AND CATALYTIC MECHANISM*. *J. Biol. Chem.* 277**,** 43433-43442.

Ke, R., Mignardi, M., Hauling, T., and Nilsson, M. (2016). Fourth Generation of Next-Generation Sequencing Technologies: Promise and Consequences. *Hum Mutat* 37**,** 1363-1367.

Keshavarz-Tohid, V., Vacheron, J., Dubost, A., Prigent-Combaret, C., Taheri, P., Tarighi, S., Taghavi, S.M., Moenne-Loccoz, Y., and Muller, D. (2019). Genomic, phylogenetic and catabolic re-assessment of the *Pseudomonas putida* clade supports the delineation of *Pseudomonas alloputida* sp. nov., *Pseudomonas inefficax* sp. nov., *Pseudomonas persica* sp. nov., and *Pseudomonas shirazica* sp. nov. *Syst Appl Microbiol* 42**,** 468-480.

Khadka, B., and Gupta, R.S. (2017). Identification of a conserved 8 aa insert in the PIP5K protein in the *Saccharomycetaceae* family of fungi and the molecular dynamics simulations and structural analysis to investigate its potential functional role. *Proteins: Structure, Function, and Bioinformatics* 85**,** 1454-1467.

Khadka, B., Persaud, D., and Gupta, R.S. (2020). Novel Sequence Feature of SecA Translocase Protein Unique to the Thermophilic Bacteria: Bioinformatics Analyses to Investigate Their Potential Roles. *Microorganisms* 8**,** 59.

Kidd, T.J., Gibson, J.S., Moss, S., Greer, R.M., Cobbold, R.N., Wright, J.D., Ramsay, K.A., Grimwood, K., and Bell, S.C. (2011). Clonal complex *Pseudomonas aeruginosa* in horses. *Vet Microbiol* 149**,** 508-512.

Kim, M., Oh, H.S., Park, S.C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 64**,** 346-351.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., and Schmidt, T.M. (2001). rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res.* 29**,** 181-184.

Klenk, H.P., and Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol* 33**,** 175-182.

Klingner, A.E., Palleroni, N.J., and Pontis, R.E. (1976). Isolation of *Pseudomonas syringae* from lesions on Vitis vinifera. *J. Phytopathol.* 86**,** 107-116.

Kluyver, A.J., and Van Niel, C.B. (1936). Prospects for a natural system of classification of bacteria. *Zentbl Bakteriolog* 94**,** 369-403.

Knopp, M., Gudmundsdottir, J.S., Nilsson, T., König, F., Warsi, O., Rajer, F., Ädelroth, P., and Andersson, D.I. (2019). De novo emergence of peptides that confer antibiotic resistance. *MBio* 10**,** 10.1128/mbio. 00837-00819.

Konstantinidis, K.T., and Tiedje, J.M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A* 102**,** 2567-2572.

Konstantinidis, K.T., and Tiedje, J.M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 187**,** 6258-6264.

Kruse, L., Loeschcke, A., De Witt, J., Wierckx, N., Jaeger, K.E., and Thies, S. (2024). *Halopseudomonas* species: Cultivation and molecular genetic tools. *Microb Biotechnol* 17**,** e14369.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35**,** 1547-1549.

Kuo, C.-H., and Ochman, H. (2008). The fate of new bacterial genes. *FEMS microbiology reviews* 33**,** 38-43.

Kwon, S.W., Kim, J.S., Park, I.C., Yoon, S.H., Park, D.H., Lim, C.K., and Go, S.J. (2003). *Pseudomonas koreensis* sp. nov., *Pseudomonas umsongensis* sp. nov. and *Pseudomonas jinjuensis* sp. nov., novel species from farm soils in Korea. *Int J Syst Evol Microbiol* 53**,** 21-27.

Lalucat, J., Gomila, M., Mulet, M., Zaruma, A., and Garcia-Valdes, E. (2022). Past, present and future of the boundaries of the *Pseudomonas* genus: Proposal of *Stutzerimonas* gen. Nov. *Syst Appl Microbiol* 45**,** 126289.

Lalucat, J., Mulet, M., Gomila, M., and Garcia-Valdes, E. (2020). Genomics in Bacterial Taxonomy: Impact on the Genus *Pseudomonas*. *Genes (Basel)* 11**,** 139.

Lang, J.M., Darling, A.E., and Eisen, J.A. (2013). Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PloS one* 8**,** e62510.

Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol* 25**,** 1307-1320.

Ledizet, M., Murray, T.S., Puttagunta, S., Slade, M.D., Quagliarello, V.J., and Kazmierczak, B.I. (2012). The ability of virulence factor expression by *Pseudomonas aeruginosa* to predict clinical disease in hospitalized patients. *PLoS One* 7**,** e49578.

Lee, D.H., Kim, C.M., and Kim, S.B. (2022). *Pseudomonas insulae* sp. nov., isolated from island soil. *Int J Syst Evol Microbiol* 72.

Lelliott, R.A., Billing, E., and Hayward, A.C. (1966). A determinative scheme for the fluorescent plant pathogenic pseudomonads. *J Appl Bacteriol* 29**,** 470-489.

Li, K., Ye, Z., Chen, G., Zheng, K., Yin, J., Debnath, S.C., Yan, C., Hong, Y., Guo, Z., Wang, P., Wu, F., and Zheng, D. (2023). *Atopomonas sediminilitoris* sp. nov., isolated from beach sediment of Zhairuo Island, China. *Antonie Van Leeuwenhoek* 116**,** 97-107.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22**,** 1658-1659.

Li, Y., Wang, M., Sun, Z.-Z., and Xie, B.-B. (2021a). Comparative genomic insights into the taxonomic classification, diversity, and secondary metabolic potentials of *Kitasatospora*, a genus closely related to *Streptomyces*. *Front Microbiol* 12**,** 683814.

Li, Y., Wang, S., Bian, D., Sun, S., and Ma, T. (2021b). *Pseudomonas quercus* sp. nov, associated with leaf spot disease of *Quercus mongolica*. *Int J Syst Evol Microbiol* 71**,** 004800.

Linnaeus, C. (1735). *Systema naturae, sive Regna tria naturae systematice proposita per classes, ordines, genera, et species.* Theod. Haak.

Linnaeus, C. (1789). *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species; cum characteribus, differentiis, synonymis, locis.* apud JB Delamolliere.

Lister, P.D., Wolter, D.J., and Hanson, N.D. (2009). Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin Microbiol Rev* 22**,** 582-610.

Liu, F., Yang, S., Xu, F., Zhang, Z., Lu, Y., Zhang, J., and Wang, G. (2022). Characteristics of biological control and mechanisms of *Pseudomonas chlororaphis* zm-1 against peanut stem rot. *BMC Microbiology* 22**,** 9.

Liu, X., Xiang, L., Yin, Y., Li, H., Ma, D., and Qu, Y. (2021a). Pneumonia caused by *Pseudomonas fluorescens*: a case report. *BMC pulmonary medicine* 21**,** 1-6.

Liu, Y., Song, Z., Zeng, H., Lu, M., Zhu, W., Wang, X., Lian, X., and Zhang, Q. (2021b). *Pseudomonas eucalypticola* sp. nov., a producer of antifungal agents isolated from *Eucalyptus dunnii* leaves. *Sci. Rep.* 11**,** 1-10.

Loeschcke, A., and Thies, S. (2015). *Pseudomonas putida*—a versatile host for the production of natural products. *Appl. Microbiol. Biotechnol.* 99**,** 6197-6214.

Lopez, J.R., Dieguez, A.L., Doce, A., De La Roca, E., De La Herran, R., Navas, J.I., Toranzo, A.E., and Romalde, J.L. (2012). *Pseudomonas baetica* sp. nov., a fish pathogen isolated from wedge sole, *Dicologlossa cuneata* (Moreau). *Int J Syst Evol Microbiol* 62**,** 874-882.

Ludwig, W., and Klenk, H.-P. (2005). Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. *Bergey's manual® of systematic bacteriology***,** 49-66.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M., and Schleifer, K.H. (1998). Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* 19**,** 554-568.

Ludwig, W., Viver, T., Westram, R., Francisco Gago, J., Bustos-Caparros, E., Knittel, K., Amann, R., and Rossello-Mora, R. (2021). Release LTP_12_2020, featuring a new ARB alignment and improved 16S rRNA tree for prokaryotic type strains. *Syst Appl Microbiol* 44**,** 126218.

Ludwig, W.a.K., H.P. (2005). Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. *Bergey's manual® of systematic bacteriology***,** 49-66.

Lund-Palau, H., Turnbull, A.R., Bush, A., Bardin, E., Cameron, L., Soren, O., Wierre-Gore, N., Alton, E.W., Bundy, J.G., Connett, G., Faust, S.N., Filloux, A., Freemont, P., Jones, A., Khoo, V., Morales, S., Murphy, R., Pabary, R., Simbo, A., Schelenz, S., Takats, Z., Webb, J., Williams, H.D., and Davies, J.C. (2016).

Pseudomonas aeruginosa infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches. *Expert. Rev. Respir. Med* 10**,** 685-697.

Lyczak, J.B., Cannon, C.L., and Pier, G.B. (2000). Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect* 2**,** 1051-1060.

Madhaiyan, M., Poonguzhali, S., Saravanan, V.S., Selvapravin, K., Duraipandiyan, V., and Al-Dhabi, N.A. (2017). *Pseudomonas sesami* sp. nov., a plant growth-promoting *Gammaproteobacteria* isolated from the rhizosphere of *Sesamum indicum* L. *Antonie van Leeuwenhoek* 110**,** 843-852.

Maiden, M.C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 60**,** 561-588.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95**,** 3140-3145.

Maiden, M.C., Jansen Van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., and Mccarthy, N.D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11**,** 728-736.

Majeed, A., Abbasi, M.K., Hameed, S., Imran, A., and Rahim, N. (2015). Isolation and characterization of plant growth-promoting rhizobacteria from wheat rhizosphere and their effect on plant growth promotion. *Front. microbiol.* 6**,** 198.

Malhotra, M., Bello, S., and Gupta, R.S. (2024). Phylogenomic and molecular markers based studies on clarifying the evolutionary relationships among *Peptoniphilus* species. Identification of several Genus-Level clades of *Peptoniphilus* species and transfer of some *Peptoniphilus* species to the genus *Aedoeadaptatus*. *Syst Appl Microbiol* 47**,** 126499.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S.V., and Machado, M.A. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13**,** 614-629.

Mccarthy, B.J., and Bolton, E.T. (1963). An approach to the measurement of genetic relatedness among organisms. *Proc Natl Acad Sci U S A* 50**,** 156-164.

Méheust, R., Castelle, C.J., Jaffe, A.L., and Banfield, J.F. (2022). Conserved and lineage-specific hypothetical proteins may have played a central role in the rise and diversification of major archaeal groups. *BMC biology* 20**,** 154.

Mehmood, N., Saeed, M., Zafarullah, S., Hyder, S., Rizvi, Z.F., Gondal, A.S., Jamil, N., Iqbal, R., Ali, B., and Ercisli, S. (2023). Multifaceted impacts of plant-beneficial *Pseudomonas* spp. in managing various plant diseases and crop yield improvement. *ACS omega* 8**,** 22296-22315.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics* 14**,** 1-14.

Menendez, E., Ramirez-Bahena, M.H., Fabryova, A., Igual, J.M., Benada, O., Mateos, P.F., Peix, A., Kolarik, M., and Garcia-Fraile, P. (2015). *Pseudomonas coleopterorum* sp. nov., a cellulase-producing bacterium isolated from the bark beetle *Hylesinus fraxini*. *Int J Syst Evol Microbiol* 65**,** 2852-2858.

Mignard, S., and Flandrois, J.P. (2006). 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *J Microbiol Methods* 67**,** 574-581.

Migula, W. (1894). Uber ein neues System der Bakterien. *Arb. Bakt. Inst. Kar1sruhe* 1**,** 238.

Migula, W. (1900). *System der bakterien: Handbuch der morphologie, entwicklungsgeschichte und systematik der bakterien.* Fischer.

Miller, W.R., and Arias, C.A. (2024). ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. *Nat. Rev. Microbiol.* 22**,** 598-616.

Miranda, C.D., and Zemelman, R. (2002). Antimicrobial multiresistance in bacteria isolated from freshwater Chilean salmon farms. *Sci. Total Environ.* 293**,** 207-218.

Miton, C.M., and Tokuriki, N. (2022). Insertions and deletions (indels): a missing piece of the protein engineering jigsaw. *Biochem.* 62**,** 148-157.

Montecillo, J.a.V. (2022). Phylogenomics and comparative genomics analyses support the creation of the novel family *Ignatzschineriaceae* fam. nov. comprising the genera *Ignatzschineria* and *Wohlfahrtiimonas* within the order *Cardiobacteriales*. *Res Microbiol***,** 103988.

Moore, E.R., Mihaylova, S.A., Vandamme, P., Krichevsky, M.I., and Dijkshoorn, L. (2010). Microbial systematics and taxonomy: relevance for a microbial commons. *Res. Microbiol.* 161**,** 430-438.

Mora, M.R., Pastrana, V.M.F., Reguero, D.G., Oliva, L.L.G., Lobo, A.P., and Gómez, P.a.J. (2022). Oxidative stress protection and growth promotion activity of *Pseudomonas mercuritolerans* sp. nov., in forage plants under mercury abiotic stress conditions. *Front Microbiol* 13.

Moradali, M.F., Ghods, S., and Rehm, B.H. (2017). *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Front Cell Infect Microbiol* 7**,** 39.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffiere, A., Yan, S., Dominguez, H., and Thompson, B.M. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J* 2**,** 321-334.

Mulet, M., Gomila, M., Busquets, A., Sánchez, D., Lalucat, J., and García-Valdés, E. (2024). Genome-Based Taxonomy of Species in the *Pseudomonas syringae* and *Pseudomonas lutea* Phylogenetic Groups and Proposal of *Pseudomonas maioricensis* sp. nov., Isolated from Agricultural Soil. *Microorganisms* 12**,** 460.

Mulet, M., Lalucat, J., and Garcia-Valdes, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol* 12**,** 1513-1530.

Müller, O.F., and Fabricius, O. (1786). *Animalcula infusoria fluviatilia et marina.* Typis N. Mölleri.

Munson, E., and Carroll, K.C. (2019). An update on the novel genera and species and revised taxonomic status of bacterial organisms described in 2016 and 2017. *J. Clin. Microbiol.* 57**,** 10.1128/jcm. 01181-01118.

Na, S.-I., Kim, Y.O., Yoon, S.-H., Ha, S.-M., Baek, I., and Chun, J. (2018). UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol. Methods* 56**,** 280-285.

Nadal Jimenez, P., Koch, G., Thompson, J.A., Xavier, K.B., Cool, R.H., and Quax, W.J. (2012). The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. *Microbiology and Molecular Biology Reviews* 76**,** 46-65.

Narra, H.P., Cordes, M.H., and Ochman, H. (2008). Structural features and the persistence of acquired proteins. *Proteomics* 8**,** 4772-4781.

Naushad, H.S., and Gupta, R.S. (2013). Phylogenomics and molecular signatures for species from the plant pathogen-containing order *Xanthomonadales*. *PLoS. One* 8**,** e55216.

Naushad, H.S., Lee, B., and Gupta, R.S. (2014). Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya, Pectobacterium* and *Brenneria*. *Int J Syst Evol Microbiol* 64**,** 366-383.

Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H.E., and Gupta, R.S. (2015). A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie Van Leeuwenhoek* 107**,** 467-485.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., and Ketchum, K.A. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399**,** 323-329.

Nikel, P.I., and De Lorenzo, V. (2018). *Pseudomonas putida* as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metab Eng* 50**,** 142-155.

Nikolaidis, M., Mossialos, D., Oliver, S.G., and Amoutzias, G.D. (2020). Comparative analysis of the core proteomes among the *Pseudomonas* major evolutionary groups reveals species-specific adaptations for *Pseudomonas aeruginosa* and *Pseudomonas chlororaphis*. *Diversity* 12**,** 289.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302**,** 205-217.

Nouioui, I., and Sangal, V. (2022). Advanced prokaryotic systematics: the modern face of an ancient science. *New Microbes and New Infections* 49**,** 101036.

Olsen, G.J., and Woese, C.R. (1993). Ribosomal RNA: a key to phylogeny. *The FASEB journal* 7**,** 113-123.

Olsen, G.J., Woese, C.R., and Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol* 176**,** 1-6.

Oren, A. (2010). "Concepts about Phylogeny of Microorganisms - an Historical Overview," in *Molecular Phylogeny of Microorganisms,* eds. A. Oren & R.T. Papke. (Norfolk: Caister Academic Press), 1-21.

Oren, A., Arahal, D.R., Goker, M., Moore, E.R.B., Rossello-Mora, R., and Sutcliffe, I.C. (2023). International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). *Int J Syst Evol Microbiol* 73.

Oren, A., and Garrity, G.M. (2014). Then and now: a systematic review of the systematics of prokaryotes in the last 80 years 13. *Antonie van Leeuwenhoek* 106**,** 43-56.

Oren, A., and Garrity, G.M. (2022). Notification that new names of prokaryotes, new combinations, and new taxonomic opinions have appeared in volume 72, part 7 of the IJSEM. *Int J Syst Evol Microbiol* 72.

Orla-Jensen, S. (1909). Die Hauptlinien de natürlichen Baketeriensystems nebst einer Uebersicht der Gärungsphenomene. *Zentr. Bakt. Parasitenk. II* 22**,** 305-346.

Osborne, A., Armstrong, K., Catrysse, N., Butler, G., and Versavel, L. (1981). An outbreak of *Pseudomonas* mastitis in dairy cows. *Can Vet J.* 22**,** 215.

Palleroni, N. (1984). Genus I. Pseudomonas Migula 1894, p. 141-199. In N. R. Krieg and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 1. The Williams & Wilkins Co., Baltimore, Md..

Palleroni, N.J. (2005). Genus I. *Pseudomonas* Migula 1894. In Bergey's Manual of Systematic Bacteriology (The Proteobacteria), part B (The Gammaproteobacteria), 2nd edn,. *Edited by D. J. Brenner, N. R. Krieg, James T. Staley & G. M. Garrity. New York: Springer* 2**,** 323-379.

Palleroni, N.J. (2010). The *Pseudomonas* story. *Environ. Microbiol* 12**,** 1377-1383.

Palleroni, N.J. (2015). "*Pseudomonas*." John Wiley and Sons in Association with Bergey's Manual Trust), 1-105.

Pang, Y., Zhang, Y., Chen, M., Lu, W., Chen, M., Yan, Y., Lin, M., Zhang, W., and Zhou, Z. (2021). *Pseudomonas nanhaiensis* sp. nov., a lipase-producing bacterium

isolated from deep-sea sediment of the South China Sea. *Antonie Van Leeuwenhoek* 114**,** 1791-1804.

Pang, Z., Raudonis, R., Glick, B.R., Lin, T.-J., and Cheng, Z. (2019). Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies. *Biotechnology Advances* 37**,** 177-192.

Panpatte, D.G., Jhala, Y.K., Shelat, H.N., and Vyas, R.V. (2016). *Pseudomonas fluorescens*: a promising biocontrol agent and PGPR for sustainable agriculture. *Microbial inoculants in sustainable agricultural productivity: Vol. 1: Research perspectives***,** 257-270.

Papadopoulou, E.S., Perruchon, C., Vasileiadis, S., Rousidou, C., Tanou, G., Samiotaki, M., Molassiotis, A., and Karpouzas, D.G. (2018). Metabolic and evolutionary insights in the transformation of diphenylamine by a *Pseudomonas putida* strain unravelled by genomic, proteomic, and transcription analysis. *Front. microbiol.* 9**,** 676.

Parker, C.T., Tindall, B.J., and Garrity, G.M. (2019). International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 69**,** S1-S111.

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (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**,** D785-D794.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol* 39**,** 996-1004.

Parte, A.C. (2018). LPSN - The List of Prokaryotic Names with Standing in Nomenclature. *Int J Syst Evol Microbiol* 68**,** 1825-1829.

Parte, A.C., Sarda Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C., and Goker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 70**,** 5607-5612.

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

Passarelli-Araujo, H., Jacobs, S.H., Franco, G.R., and Venancio, T.M. (2021). Phylogenetic analysis and population structure of *Pseudomonas alloputida*. *Genomics* 113**,** 3762-3773.

Patel, J.B. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol. Diagn* 6**,** 313-321.

Patel, S., and Gupta, R.S. (2020). A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 70**,** 406-438.

Peix, A., M.H., R.-B., and Velázquez, E. (2009). Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect Genet. Evol* 9**,** 1132-1147.

Peix, A., Ramirez-Bahena, M.H., and Velazquez, E. (2018). The current status on the taxonomy of *Pseudomonas* revisited: An update. *Infect Genet. Evol* 57**,** 106-116.

Peix, A., Rivas, R., Mateos, P.F., Martínez-Molina, E., Rodríguez-Barrueco, C., and Velazquez, E. (2003). *Pseudomonas rhizosphaerae* sp. nov., a novel species that actively solubilizes phosphate in vitro. *Int J Syst Evol Microbiol* 53**,** 2067-2072.

Pendleton, J.N., Gorman, S.P., and Gilmore, B.F. (2013). Clinical relevance of the ESKAPE pathogens. *Expert Rev. Anti-infect. Ther.* 11**,** 297-308.

Peral-Aranega, E., Saati-Santamaría, Z., Kolařik, M., Rivas, R., and García-Fraile, P. (2020). Bacteria belonging to *Pseudomonas typographi* sp. nov. from the bark beetle Ips typographus have genomic potential to aid in the host ecology. *Insects* 11**,** 593.

Philippe, H., and Douady, C.J. (2003). Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* 6**,** 498-505.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., and Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.* 5**,** 1-8.

Planquette, B., Timsit, J.F., Misset, B.Y., Schwebel, C., Azoulay, E., Adrie, C., Vesin, A., Jamali, S., Zahar, J.R., Allaouchiche, B., Souweine, B., Darmon, M., Dumenil, A.S., Goldgran-Toledano, D., Mourvillier, B.H., Bedos, J.P., and Group, O.S. (2013). *Pseudomonas aeruginosa* ventilator-associated pneumonia. predictive factors of treatment failure. *Am J Respir Crit Care Med* 188**,** 69-76.

Poblete-Castro, I., Becker, J., Dohnt, K., Dos Santos, V.M., and Wittmann, C. (2012). Industrial biotechnology of *Pseudomonas putida* and related species. *Appl. Microbiol. Biotechnol.* 93**,** 2279-2290.

Preston, G.M. (2000). *Pseudomonas syringae* pv. tomato: the right pathogen, of the right plant, at the right time. *Mol. Plant Pathol.* 1**,** 263-275.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5**,** e9490.

Pringsheim, E. (1923). Zur kritik der Bakteriensystematik. *Lotos* 71**,** 357-377.

Puigbo, P., Wolf, Y.I., and Koonin, E.V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol* 8**,** 59.

Qin, Q.L., Xie, B.B., Zhang, X.Y., Chen, X.L., Zhou, B.C., Zhou, J., Oren, A., and Zhang, Y.Z. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 196**,** 2210-2215.

Qin, S., Xiao, W., Zhou, C., Pu, Q., Deng, X., Lan, L., Liang, H., Song, X., and Wu, M. (2022). *Pseudomonas aeruginosa*: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. *Signal Transduct Target Ther* 7**,** 199.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41**,** D590-D596.

Quenouille, M.H. (Year). "Approximate tests of correlation in time-series 3", in: *Mathematical Proceedings of the Cambridge Philosophical Society*: Cambridge University Press), 483-484.

Rainey, P.B., Brodey, C.L., and Johnstone, K. (1993). Identification of a gene cluster encoding three high-molecular-weight proteins, which is required for synthesis of tolaasin by the mushroom pathogen *Pseudomonas tolaasii*. *Mol. Microbiol.* 8**,** 643-652.

Raio, A. (2024). Diverse roles played by “Pseudomonas fluorescens complex” volatile compounds in their interaction with phytopathogenic microrganims, pests and plants. *World J. Microbiol. Biotechnol* 40**,** 80.

Ramsey, D.M., Baynham, P.J., and Wozniak, D.J. (2005). Binding of *Pseudomonas aeruginosa* AlgZ to sites upstream of the algZ promoter leads to repression of transcription. *J Bacteriol* 187**,** 4430-4443.

Rannala, B., and Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* 43**,** 304-311.

Reller, L.B., Weinstein, M.P., and Petti, C.A. (2007). Detection and identification of microorganisms by gene amplification and sequencing. *Clin. Infect. Dis.* 44**,** 1108-1114.

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106**,** 19126-19131.

Rokas, A., and Holland, P.W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol* 15**,** 454-459.

Rokas, A., Williams, B.L., King, N., and Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425**,** 798-804.

Rosselló-Mora, R. (2006). "DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation," in *Molecular Identification, Systematics, and Population Structure of Prokaryotes,* ed. E. Stackebrandt. (Berlin, Heidelberg: Springer Berlin Heidelberg), 23-50.

Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS microbiology reviews* 25**,** 39-67.

Rossi, E., La Rosa, R., Bartell, J.A., Marvig, R.L., Haagensen, J.a.J., Sommer, L.M., Molin, S., and Johansen, H.K. (2021). *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat Rev Microbiol* 19**,** 331-342.

Roy, P.H., Tetu, S.G., Larouche, A., Elbourne, L., Tremblay, S., Ren, Q., Dodson, R., Harkins, D., Shay, R., Watkins, K., Mahamoud, Y., and Paulsen, I.T. (2010). Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 5**,** e8842.

Rudra, B., and Gupta, R.S. (2021). Phylogenomic and comparative genomic analyses of species of the family *Pseudomonadaceae*: Proposals for the genera *Halopseudomonas* gen. nov. and *Atopomonas* gen. nov., merger of the genus *Oblitimonas* with the genus *Thiopseudomonas*, and transfer of some

misclassified species of the genus *Pseudomonas* into other genera. *Int J Syst Evol Microbiol* 71.

Rudra, B., Duncan, L., Shah, A. J., Shah, H. N., & Gupta, R. S. (2022). Phylogenomic and comparative genomic studies robustly demarcate two distinct clades of *Pseudomonas aeruginosa* strains: proposal to transfer the strains from an outlier clade to a novel species *Pseudomonas paraeruginosa* sp. nov. *Int J Syst Evol Microbiol.*, *72*(11), 005542.

Rudra, B., and Gupta, R.S. (2024). Phylogenomics studies and molecular markers reliably demarcate genus *Pseudomonas* sensu stricto and twelve other *Pseudomonadaceae* species clades representing novel and emended genera. *Front. microbiol.* 14**,** 1273665.

Rudra, B., and Gupta, R.S. (2025). Molecular Markers Specific for the *Pseudomonadaceae* Genera Provide Novel and Reliable Means for the Identification of Other *Pseudomonas* Strains/spp. Related to These Genera. *Genes* 16**,** 183.

Rutherford, S.T., and Bassler, B.L. (2012). Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med* 2.

Saati-Santamaría, Z., López-Mondéjar, R., Jiménez-Gómez, A., Díez-Méndez, A., Větrovský, T., Igual, J.M., Velázquez, E., Kolarik, M., Rivas, R., and García-Fraile, P. (2018). Discovery of Phloeophagus Beetles as a Source of *Pseudomonas* Strains That Produce Potentially New Bioactive Substances and Description of *Pseudomonas bohemica* sp. nov. *Front Microbiol* 9**,** 913.

Saati-Santamaria, Z., Peral-Aranega, E., Velazquez, E., Rivas, R., and Garcia-Fraile, P. (2021). Phylogenomic Analyses of the Genus *Pseudomonas* Lead to the Rearrangement of Several Species and the Definition of New Genera. *Biology (Basel)* 10.

Saeki, K., Yao, Y., Wakabayashi, S., Shen, G.-J., Zeikus, J.G., and Atsubara, H. (1989). Ferredoxin and rubredoxin from *Butyribacterium methylotrophicum*: complete primary structures and construction of phylogenetic trees. *J. Biol. Chem.* 106**,** 656-662.

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* 14**,** 225-IN226.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4**,** 406-425.

Saleem, F., Li, E., Tran, K.L., Rudra, B., Edge, T.A., Schellhorn, H.E., and Gupta, R.S. (2024). Utilizing novel *Escherichia coli*-specific conserved signature proteins for enhanced monitoring of recreational water quality. *MicrobiologyOpen* 13**,** e1410.

Sanger, F. (1959). Chemistry of insulin: determination of the structure of insulin opens the way to greater understanding of life processes. *Science* 129**,** 1340-1344.

Sapp, J. (2006). "The bacterium's place in nature.," in *Microbial Phylogeny and Evolution: Concepts and Controversies,* ed. J. Sapp. (New York: Oxford University Press), 3-52.

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

Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., Holmes, J.B., Kim, S., Kimchi, A., Kitts, P.A., Lathrop, S., Lu, Z., Madden, T.L., Marchler-Bauer, A., Phan, L., Schneider, V.A., Schoch, C.L., Pruitt, K.D., and Ostell, J. (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 47**,** D23-D28.

Scales, B.S., Dickson, R.P., Lipuma, J.J., and Huffnagle, G.B. (2014). Microbiology, genomics, and clinical significance of the *Pseudomonas fluorescens* species complex, an unappreciated colonizer of humans. *Clin. Microbiol. Rev.* 27**,** 927-948.

Scales, B.S., Erb-Downward, J.R., Lipuma, J.J., and Huffnagle, G.B. (2015). Draft genome sequences of five *Pseudomonas fluorescens* subclade I and II strains, isolated from human respiratory samples. *Genome Announc.* 3**,** 10.1128/genomea. 00837-00815.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum Mol Genet* 19**,** R227-240.

Schildkraut, C.L., Marmur, J., and Doty, P. (1961). The formation of hybrid DNA molecules and their use in studies of DNA homologies. *J. Mol. Biol.* 3**,** 595-IN516.

Schleifer, K.H. (2009). Classification of Bacteria and Archaea: past, present and future. *Syst Appl Microbiol.* 32**,** 533-542.

Schopf, J., and Packer, B. (1987). Early Archean (3.3 to 3.5 Ga-old) fossil microorganisms from the Warrawoona Group, Western Australia. *Science* 237**,** 70-73.

Schopf, J.W. (1978). The evolution of the earliest cells. *Scientific American* 239**,** 110-120.

Schroeter, J. (1872). "Über einige durch Bacterien gebildete Pigmente," in *Beiträge zur Biologie der Pflanzen,* ed. C. F. (Breslau: Max Müller), 109-126.

Schroth, M.N., Hildebrand, D.C., and Panopoulos, N. (2006). Phytopathogenic pseudomonads and related plant-associated pseudomonads. *Prokaryotes* 6**,** 714-740.

Schwartz, R.M., and Dayhoff, M.O. (1978). Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts: A perspective is derived from protein and nucleic acid sequence data. *Science* 199**,** 395-403.

Segata, N., Bornigen, D., Morgan, X.C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun* 4**,** 2304.

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16**,** 1114.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst Biol* 7**,** 539.

Silby, M.W., Winstanley, C., Godfrey, S.A., Levy, S.B., and Jackson, R.W. (2011). *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 35**,** 652-680.

Singh, B., and Gupta, R.S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genomics* 281**,** 361-373.

Skerman, V.B.D., Mcgowan, V., and Sneath, P.H.A. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* 30**,** 225-420.

Sneath, P.H., Mair, N.S., Sharpe, M.E., and Holt, J.G. (1986). *Bergey's manual of systematic bacteriology. Volume 2.*

Snel, B., Huynen, M.A., and Dutilh, B.E. (2005). Genome trees and the nature of genome evolution. *Annu. Rev. Microbiol.* 59**,** 191-209.

Sokal, R.R., and Sneath, P.H. (1963). Principles of numerical taxonomy.

Sood, U., Hira, P., Kumar, R., Bajaj, A., Rao, D.L.N., Lal, R., and Shakarad, M. (2019). Comparative genomic analyses reveal core-genome-wide genes under positive selection and major regulatory hubs in outlier strains of *Pseudomonas aeruginosa*. *Front Microbiol* 10**,** 53.

Sood, U., Singh, D.N., Hira, P., Lee, J.K., Kalia, V.C., Lal, R., and Shakarad, M. (2020). Rapid and solitary production of mono-rhamnolipid biosurfactant and biofilm inhibiting pyocyanin by a taxonomic outlier *Pseudomonas aeruginosa* strain CR1. *J Biotechnol* 307**,** 98-106.

Spagnolo, A.M., Sartini, M., and Cristina, M.L. (2021). *Pseudomonas aeruginosa* in the healthcare facility setting. *Reviews and Research in Medical Microbiology* 32**,** 169-175.

Stackebrandt, E. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbial Today* 33**,** 152.

Stackebrandt, E., and Goebel, B.M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44**,** 846-849.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinform.* 30**,** 1312-1313.

Stanier, R.Y., Adelberg, E.A., and Ingraham, J.L. (1976). *The Microbial World.* Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Stanier, R.Y., Palleroni, N.J., and Doudoroff, M. (1966). The aerobic pseudomonads a taxonomic study. *Microbiol.* 43**,** 159-271.

Stanier, R.Y., and Van Niel, C.B. (1941). The Main Outlines of Bacterial Classification. *J Bacteriol* 42**,** 437-466.

Stanier, R.Y., and Van Niel, C.B. (1962). The concept of a bacterium. *Archiv für Mikrobiologie* 42**,** 17-35.

Steel, M., Lockhart, P., and Penny, D. (1993). Confidence in evolutionary trees from biological sequence data. *Nature* 364**,** 440-442.

Stover, C.K., Pham, X.Q., Erwin, A., Mizoguchi, S., Warrener, P., Hickey, M., Brinkman, F., Hufnagle, W., Kowalik, D., and Lagrou, M. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406**,** 959-964.

Sun, D.L., Jiang, X., Wu, Q.L., and Zhou, N.Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microbiol* 79**,** 5962-5969.

Susko, E., and Roger, A.J. (2021). Long branch attraction biases in phylogenetics. *Syst. Biol.* 70**,** 838-843.

Sutcliffe, I.C., Trujillo, M.E., and Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie Van Leeuwenhoek* 101**,** 13-20.

Tang, Y.W., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E., and Persing, D.H. (1998). Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. *J Clin Microbiol* 36**,** 3674-3679.

Tateno, Y., Takezaki, N., and Nei, M. (1994). Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.*. 11**,** 261-277.

Thompson, C.C., Chimetto, L., Edwards, R.A., Swings, J., Stackebrandt, E., and Thompson, F.L. (2013). Microbial genomic taxonomy. *BMC. Genomics* 14**,** 913.

Tindall, B.J., Rossello-Mora, R., Busse, H.J., Ludwig, W., and Kampfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 60**,** 249-266.

Toska, J., Sun, Y., Carbonell, D.A., Foster, A.N., Jacobs, M.R., Pearlman, E., and Rietsch, A. (2014). Diversity of virulence phenotypes among type III secretion negative *Pseudomonas aeruginosa* clinical isolates. *PLoS One* 9**,** e86829.

Udaondo, Z., Ramos, J.L., and Abram, K. (2024). Unraveling the genomic diversity of the *Pseudomonas putida* group: exploring taxonomy, core pangenome, and antibiotic resistance mechanisms. *FEMS Microbiol. Rev.* 48**,** fuae025.

Van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet* 34**,** 666-681.

Van Niel, C.B. (Year). "The classification and natural relationships of bacteria", in: *Cold Spring Harbor Symposia on Quantitative Biology*: Cold Spring Harbor Laboratory Press), 285-301.

Vandamme, P., Pot, B., Gills, M., De Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60**,** 407-438.

Vandamme, P.A. (2011). Taxonomy and classification of bacteria. *Manual of clinical microbiology***,** 210-227.

Vanparys, B., Heylen, K., Lebbe, L., and De Vos, P. (2006). *Pseudomonas peli* sp. nov. and *Pseudomonas borbori* sp. nov., isolated from a nitrifying inoculum. *Int J Syst Evol Microbiol* 56**,** 1875-1881.

Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C., and Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 43**,** 6761-6771.

Vasconcellos, R.L.F., Santos, S.N., Zucchi, T.D., Silva, F.S.P., Souza, D.T., and Melo, I.S. (2017). *Pseudomonas aestus* sp. nov., a plant growth-promoting bacterium isolated from mangrove sediments. *Arch Microbiol* 199**,** 1223-1229.

Vela, A.I., Gutiérrez, M.C., Falsen, E., Rollan, E., Simarro, I., García, P., Domínguez, L., Ventosa, A., and Fernandez-Garayzabal, J.F. (2006). *Pseudomonas simiae* sp. nov., isolated from clinical specimens from monkeys (Callithrix geoffroyi). *Int J Syst Evol Microbiol* 56**,** 2671-2676.

Vodovar, N., Vinals, M., Liehl, P., Basset, A., Degrouard, J., Spellman, P., Boccard, F., and Lemaitre, B. (2005). *Drosophila* host defense after oral infection by an

entomopathogenic *Pseudomonas* species. *Proc Natl Acad Sci U. S A* 102**,** 11414-11419.

Wang, Z., and Wu, M. (2013). A phylum-level bacterial phylogenetic marker database. *Mol. Biol. Evol* 30**,** 1258-1262.

Ward, B.B. (2002). How many species of prokaryotes are there? *Proc Natl Acad Sci USA* 99**,** 10234-10236.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171**,** 737-738.

Wayne, L.G., Brenner, D., Colwell, R., Grimont, P., Kandler, O., Krichevsky, M., Moore, L., Moore, W., Murray, R., and Stackebrandt, E. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Evol Microbiol.* 37**,** 463-464.

Weimer, A., Kohlstedt, M., Volke, D.C., Nikel, P.I., and Wittmann, C. (2020). Industrial biotechnology of *Pseudomonas putida*: advances and prospects. *Appl. Microbiol. Biotechnol.* 104**,** 7745-7766.

Weller, D.M., Mavrodi, D.V., Van Pelt, J.A., Pieterse, C.M., Van Loon, L.C., and Bakker, P.A. (2012). Induced systemic resistance in *Arabidopsis thaliana* against *Pseudomonas syringae* pv. tomato by 2, 4-diacetylphloroglucinol-producing *Pseudomonas fluorescens*. *Phytopathol.* 102**,** 403-412.

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol* 18**,** 691-699.

Whitman, W.B., Rainey, F., Kämpfer, P., Trujillo, M., Chun, J., and Devos, P. (2015). *Bergey's manual of systematics of Archaea and Bacteria.* Wiley Online Library.

Whittaker, R.H. (1969). New Concepts of Kingdoms of Organisms: Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science* 163**,** 150-160.

Wiener-Kronish, J.P., and Pittet, J.F. (2011). Therapies against virulence products of *Staphylococcus aureus* and *Pseudomonas aeruginosa*. *Semin Respir Crit Care Med* 32**,** 228-235.

Winslow, C.-E., Broadhurst, J., Buchanan, R., Krumwiede Jr, C., Rogers, L., and Smith, G. (1920). The families and genera of the bacteria final report of the Committee

of the Society of American bacteriologists on characterization and classification of bacterial types. *J Bacteriol* 5**,** 191-229.

Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., and Brinkman, F.S. (2016). Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 44**,** D646-D653.

Woese, C. (1991). The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria. *Evolution at the molecular level***,** 1-24.

Woese, C., Stackebrandt, E., Weisburg, W., Paster, B., Madigan, M., Fowler, V., Hahn, C., Blanz, P., Gupta, R., and Nealson, K. (1984). The phylogeny of purple bacteria: the alpha subdivision. *Syst Appl Microbiol* 5**,** 315-326.

Woese, C.R. (1987). Bacterial evolution. *Microbiol. Rev* 51**,** 221-271.

Woese, C.R. (1992). "Prokaryote systematics: the evolution of a science," in *The Prokayotes,* eds. Balows A & D.M.E.A. Tru¨ Per Hg. (New York: Springer), 3-18.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* 74**,** 5088-5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87**,** 4576-4579.

Wolfgang, M.C., Kulasekara, B.R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C.G., and Lory, S. (2003). Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 100**,** 8484-8489.

Wong, S.Y., Paschos, A., Gupta, R.S., and Schellhorn, H.E. (2014). Insertion/deletion-based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments. *Environ. Sci. Technol* 48**,** 11462-11470.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., Hooper, S.D., Pati, A., Lykidis, A., Spring, S., Anderson, I.J., D'haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J.F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E.M., Kyrpides, N.C., Klenk,

H.P., and Eisen, J.A. (2009). A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. *Nature* 462**,** 1056-1060.

Wu, N., and Jing, C. (2024). Bloodstream infection caused by *Pseudomonas fluorescens*: a case report. *ESS Open Archive eprints* 361**,** 36148585.

Xin, X.F., Kvitko, B., and He, S.Y. (2018). *Pseudomonas syringae*: what it takes to be a pathogen. *Nat. Rev. Microbiol* 16**,** 316-328.

Yahr, T.L., Goranson, J., and Frank, D.W. (1996). Exoenzyme S of *Pseudomonas aeruginosa* is secreted by a type III pathway. *Mol. Microbiol.* 22**,** 991-1003.

Yamamoto, S., Kasai, H., Arnold, D.L., Jackson, R.W., Vivian, A., and Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiol.* 146 (Pt 10)**,** 2385-2394.

Yang, Z., and Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nat. Rev. Genet* 13**,** 303-314.

Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F.O., and Rosselló-Móra, R. (2010). Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst. Appl. Microbiol.* 33**,** 291-299.

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K.H., Ludwig, W., Glockner, F.O., and Rossello-Mora, R. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol* 31**,** 241-250.

Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzeby, J., Amann, R., and Rossello-Mora, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences 1. *Nat. Rev. Microbiol* 12**,** 635-645.

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glockner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 42**,** D643-D648.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C.M., Harris, H.M., Mattarelli, P., O'toole, P.W., Pot, B., Vandamme, P., and Walter, J. (2020). A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the

genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 70**,** 2782-2858.

Zhi, X.Y., Zhao, W., Li, W.J., and Zhao, G.P. (2012). Prokaryotic systematics in the genomics era. *Antonie van Leeuwenhoek* 101**,** 21-34.

Zotin, A., Ozerniuk, N., Zotin, A., and Konoplev, V. (1975). Possible route for the origin of prokaryotes. *Zhurnal Obshchei Biologii* 36**,** 163-172.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as Documents of Evolutionary History. *J. Theor. Biol.* 8**,** 357-366.

Appendix A

1. Bello, S., Mcquay, S., **Rudra, B.**, and Gupta, R.S. (2024). Robust demarcation of the family *Peptostreptococcaceae* and its main genera based on phylogenomic studies and taxon-specific molecular markers. *Int J Syst Evol Microbiol.* 74**,** 006247.

2. Bello, S.*, Mudassir, S.H.*, **Rudra, B*.**, and Gupta, R.S. (2023). Phylogenomic and molecular markers based studies on *Staphylococcaceae* and *Gemella* species. Proposals for an emended family *Staphylococcaceae* and three new families (*Abyssicoccaceae* fam. nov., *Salinicoccaceae* fam. nov. and *Gemellaceae* fam. nov.) harboring four new genera, *Lacicoccus* gen. nov., *Macrococcoides* gen. nov., *Gemelliphila* gen. nov., and *Phocicoccus* gen. nov. *Antonie van Leeuwenhoek* 116**,** 937-973. (* indicates the equal contribution)

3. Bello, S.*****, **Rudra, B.***, and Gupta, R.S. (2022). Phylogenomic and comparative genomic analyses of *Leuconostocaceae* species: identification of molecular signatures specific for the genera *Leuconostoc, Fructobacillus* and *Oenococcus* and proposal for a novel genus *Periweissella* gen. nov. *Int J Syst Evol Microbiol* 72**,** 005284. (* indicates the equal contribution)

4. Chen, S.*****, **Rudra, B.***, and Gupta, R.S. (2021). Phylogenomics and molecular signatures support division of the order *Neisseriales* into emended families *Neisseriaceae* and *Chromobacteriacea*e and three new families *Aquaspirillaceae* fam. nov., *Chitinibacteraceae* fam. nov., and *Leeiaceae* fam. nov. *Syst Appl Microbiol* 44**,** 126251. (* indicates the equal contribution)

5. Saleem, F., Li, E., Tran, K.L., **Rudra, B.**, Edge, T.A., Schellhorn, H.E., and Gupta, R.S. (2024). Utilizing novel *Escherichia coli*-specific conserved signature proteins for enhanced monitoring of recreational water quality. *MicrobiologyOpen* 13**,** e1410.

6. Zhang, B., Tan, T., Smith, D.D., Duceppe, M.-O., **Rudra, B.**, Gupta, R.S., and Ogunremi, D. (2024). Complete genome sequence of multidrug-resistant *Raoultella terrigena* strain RT01-5M1 isolated from farmed salmon in Canada. *Microbiol. Resour. Announc.* 13**,** e00153-00124.

Book Chapter:

Duncan, L., Shah, A.J., Ward, M., Gupta, R.S., **Rudra, B.**, Han, A., Bruce, K., and Shah, H.N. (2023). "Diversity, Transmission and Selective Pressure on the Proteome of *Pseudomonas aeruginosa*," in *Microbiological Identification using MALDI-TOF and Tandem Mass Spectrometry* (eds H.N. Shah, S.E. Gharbia, A.J. Shah, E.Y. Tranfield and K.C. Thompson). 183-209.