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,

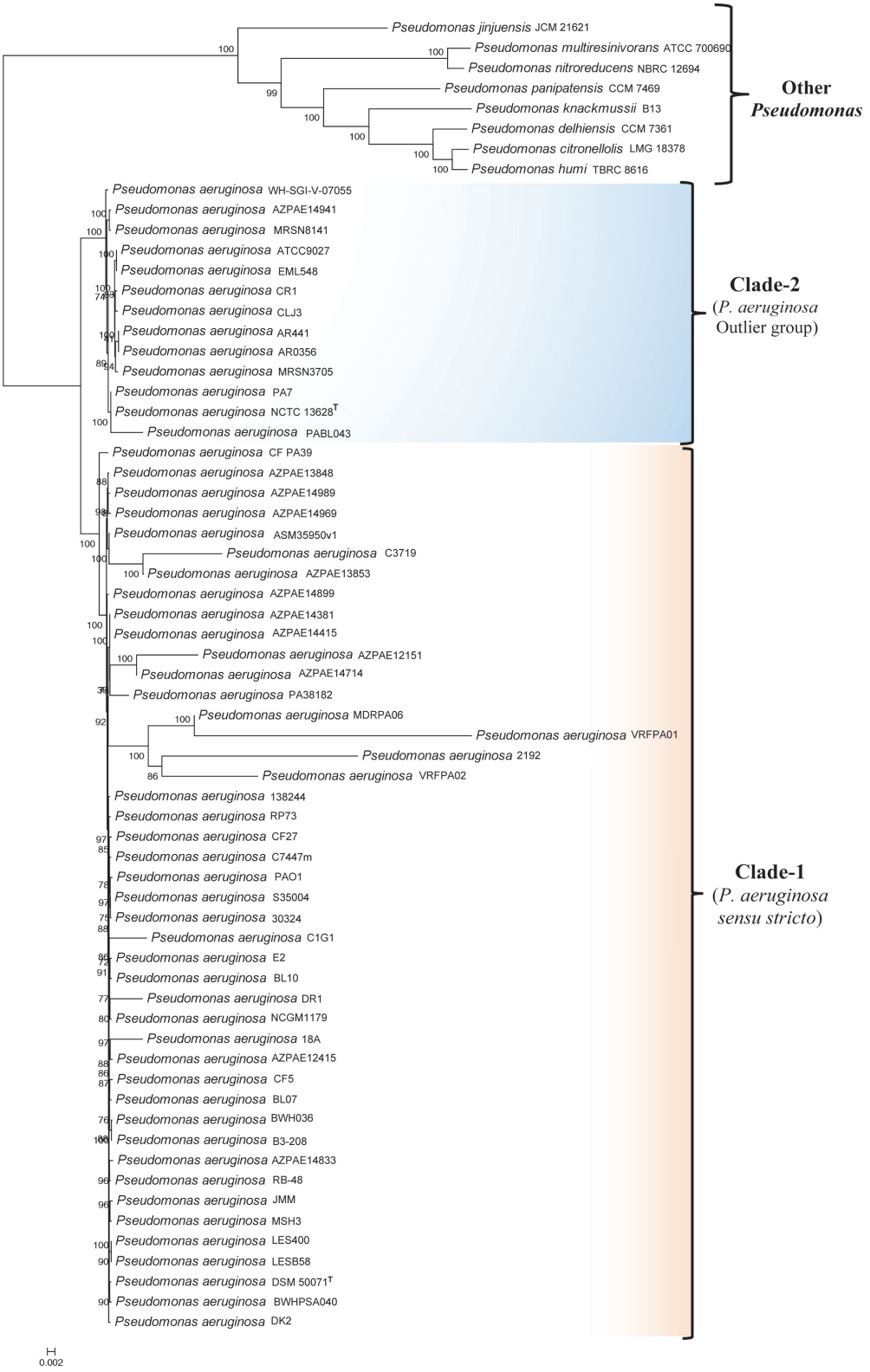

Fig. 1. A bootstrapped maximum likelihood tree for 57 genome-sequenced *P. aeruginosa* strains covering the genetic diversity of available strains, based on concatenated sequences for 118 conserved proteins. A more detailed tree for 212 *P. aeruginosa* strains is presented in Fig. S1. The two main clades 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

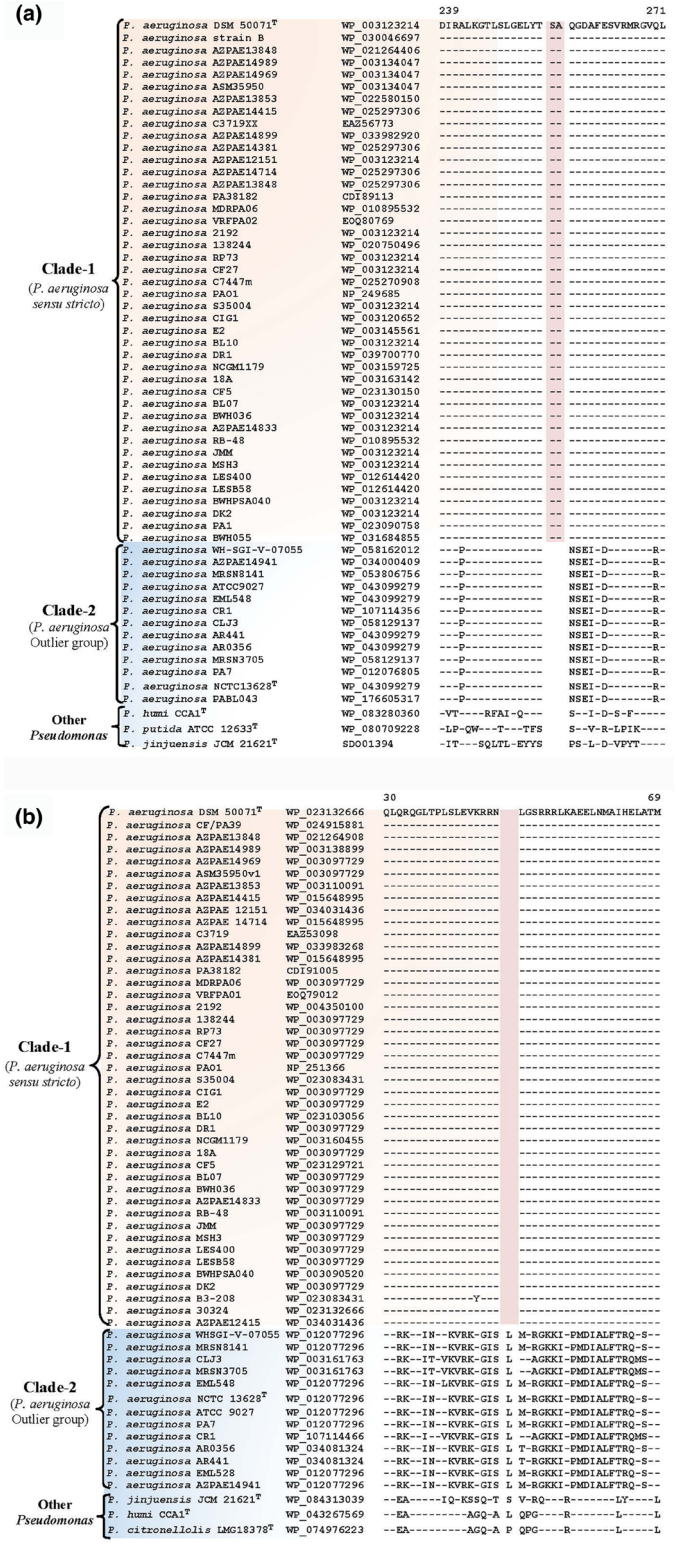

Fig. 2. Partial sequence alignments of (a) fimbrial biogenesis outer membrane usher protein containing a two aa insertion (boxed), and (b) Type II secretion system F family protein containing a one aa deletion (highlighted), which are commonly and exclusively shared by different strains which are a part 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.

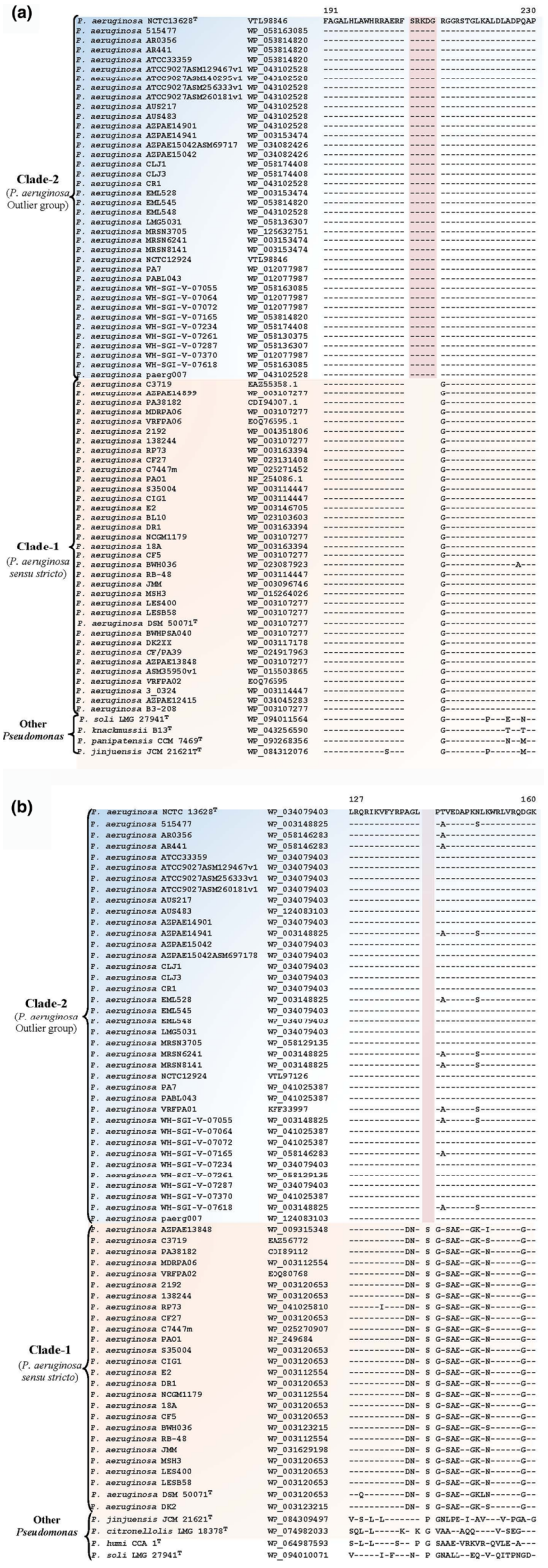

Fig. 3. Partial sequence alignments of (a) dimethylglycine demethylation protein (DgcB) containing a five aa insertion (boxed), and (b) chaperone protein FocC containing a one aa deletion (highlighted), which are commonly and exclusively shared by different *P. aeruginosa* strains from Clade-2, but not 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* PA01, 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., *Hafniacea*e 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.