MINIREVIEWS

Defining the *Pseudomonas* Genus: Where Do We Draw the Line with *Azotobacter*?

Asli I. Özen · David W. Ussery

Abstract The genus *Pseudomonas* has gone through many taxonomic revisions over the past 100 years, going from a very large and diverse group of bacteria to a smaller, more refined and ordered list having specific properties. The relationship of the *Pseudomonas* genus to *Azotobacter vinelandii* is examined using three genomic sequence-based methods. First, using 16S rRNA trees, it is shown that *A. vinelandii* groups within the *Pseudomonas* close to *Pseudomonas aeruginosa*. Genomes from other related organisms (*Acinetobacter*, *Psychrobacter*, and *Cellvibrio*) are outside the *Pseudomonas* cluster. Second, pan genome family trees based on conserved gene families also show *A. vinelandii* to be more closely related to *Pseudomonas* than other related organisms. Third, exhaustive BLAST comparisons demonstrate that the fraction of shared genes between *A. vinelandii* and *Pseudomonas* genomes is similar to that of *Pseudomonas* species with each other. The results of these different methods point to a high similarity between *A. vinelandii* and the *Pseudomonas* genus, suggesting that *Azotobacter* might actually be a *Pseudomonas*.

Introduction

Pseudomonas bacteria are naturally widespread in the environment. For example, the plant pathogen, *Pseudomo-*

A. I. Özen · D. W. Ussery (✉)
Center for Biological Sequence Analysis,
Department of Systems Biology,
The Technical University of Denmark,
Lyngby, Denmark
e-mail: dave@cbs.dtu.dk

nas syringae has been linked to the environmental cycle of water as an ice nucleus in the clouds and is found in rain, snow, lakes, and plants [31]. Because of its abundance in the environment, the *Pseudomonas* genus was first characterized long ago, and over the past hundred years, it has gone through many taxonomic revisions. The number of organisms placed in the *Pseudomonas* group grew steadily over a period of 60 years. However, through refinement of defining criteria, many bacteria were moved to other genera over the next 50 [24, 36, 42, 47].

Early studies based on rRNA–DNA hybridization postulated five RNA subdivisions in the genus, where rRNA group I, including the type species *Pseudomonas aeruginosa*, was named after the genus as *Pseudomonas* [34]. Studies on the determination and comparison of 16S rRNA sequences of *Pseudomonas* species resulted in the clustering of *Pseudomonas* into two groups: *P. aeruginosa* and *Pseudomonas fluorescens* [32]. Later on, the extensive study of Anzai and collaborators on more than 100 *Pseudomonas* species based on 16S rRNA sequence comparison suggests seven clusters from the group of species of *Pseudomonas* sensu stricto, which also agreed in some parts with Palleroni's report in 1973 [3]. Although it is still a widely accepted method, debates on the poor resolution of the phylogeny analysis with *rrs* gene sequences lead to the idea of using other marker genes to characterize and classify *Pseudomonas*, such as *gryB*, *rpoD*, *oprI*, *oprF*, and *rpoB* sequences [2, 8, 13, 57]. In another study, ten housekeeping genes were used to assess the phylogeny of 2,4-diacetylphloroglucinol-producing fluorescent *Pseudomonas* spp. [16]. Other phenotypic methods, such as siderotyping, were also suggested for the classification of plant-associated *Pseudomonas* [30]. *Pseudomonas* sensu stricto (rRNA similarity group I) could be further divided into subgroups due to its

considerable heterogeneity based on pathogenicity or pigment production [35].

The current status of the *Pseudomonas* genus today shows 202 species assigned to *Pseudomonas* on the Approved Lists of Bacterial Names, where the classification method depends on a combination of 16S rRNA, the analysis of the cellular fatty acids, and differentiating classical physiological and biochemical tests [52]. The genus consists of a group of medically and biotechnologically important bacteria that are inhabitants of a wide range of niches including soil and water environments, in addition to plant and animal associations. Hence, they are well known for having enormous metabolic versatility [17, 18, 47]. They are non-sporulating, aerobic Gram-negative rods that are found in biofilms or in planktonic forms. Most of the pathogenic members are related to plants, whereas several strains are pathogenic to animals [35].

Azotobacter vinelandii, in this context, is interesting because of its common metabolic characteristics with *Pseudomonas*. A nitrogen-fixing member of Gammaproteobacteria, *A. vinelandii* is found mostly in soil environments where its nitrogen and energy metabolism is significant to agriculture. Many years ago, this organism was often used in biochemistry experiments for isolating enzymes during the kinetics studies which resulted in surprising yields and qualities [29]. It is a free-living obligate aerobe known for having the highest respiratory characteristics, but it can still fix atmospheric nitrogen using a respiratory protection mechanism [23]. It also has distinct properties, such as dramatic increase in chromosome numbers when reached at a stationary phase, formation of cysts under carbon depletion that helps the bacteria to resist dehydration [41], where alginate is a structural component, and accumulation of poly-beta-hydroxybutyrate at the end of the exponential growth as a carbon and energy source storage [48]. Although the *Azotobacter* genus has been studied over 100 years in various experiments, currently, there is only one complete genome sequence available on NCBI GenBank database— *A. vinelandii DJ* [43]. There are no further ongoing projects listed for this genus, out of several thousand bacterial genome projects.

Azotobacter and *Pseudomonas* are members of the Pseudomonadaceae family. They both have a significant genomic diversity and genetic adaptability in a wide range of niches. However, numerous studies show that they share many biochemical metabolic pathways such as nitrogen fixation, alginate production, and respiratory mechanisms, and they are found in similar environments [11, 58]. It was long thought that *Pseudomonas* species (sensu stricto) do not have nitrogen fixation abilities; however, recently, it has been demonstrated that some *Pseudomonas* strains can fix nitrogen and that their genes related to this machinery closely resemble that of *A. vinelandii* [39, 58, 59]. Another similarity is the alginate production in *A. vinelandii*, which is also a by-product in pathogenic *P. aeruginosa* infections in the lungs of cystic fibrosis patients [20]. However, other phenotypic characteristics of *Pseudomonas* have been shown to be different from *Azotobacter* species, such as cell morphology and motility [35]. This suggests that the diversity in some phenotypic characteristics might be the outcome of their adaptive properties since the two genera share the same set of core housekeeping genes or other conserved genes [59]. In this context, analysis of 16S rRNA gene sequences by Rediers and collaborators showed that *A. vinelandii* was in the *P. aeruginosa* clade, sharing 96% identity with *P. aeruginosa* PAO1 strain. Owing to the low resolution in 16S rRNA sequences for these genomes, they conducted a phylogenetic analysis of 25 protein-coding genes, some of which were housekeeping genes. Phylogenetic trees generated with their dataset again revealed that *A. vinelandii* homologues are clustered within or close to the *Pseudomonas* group. The consensus tree out of these 25 topologies showed *A. vinelandii* phylogeny being closest to *P. aeruginosa* PAO1, concluding *A. vinelandii* to belong to the *Pseudomonas* genus [39]. Young and Park [59] use a broader approach to this idea, taking into account all the morphological differences, concluding that *Azotobacter* species can be transferred to *Pseudomonas* along with a change in the criteria used for classification.

In this article, we analyze the evolutionary relationships of the *Pseudomonas* genus to *A. vinelandii*, discussing whether or not this species is actually a *Pseudomonas*, using comparative genomic methods such as phylogeny trees, pan–core genome analysis, and protein BLAST across the whole genomes. Genomes from related genera— *Acinetobacter*, *Psychrobacter*, and *Cellvibrio*—were also used for the analysis to provide a better resolution. All genomes are therefore members of Pseudomonadales order, where *Pseudomonas*, *Azotobacter*, and *Cellvibrio* belong to the Pseudomonadaceae family and *Acinetobacter* and *Psychrobacter* are members of Moraxellaceae in the same order.

Materials and Methods

Gathering Genomes and Gene Annotation

All of the 29 genomes used in the analysis are complete sequences downloaded from GenBank [6]. The list consists of *A. vinelandii*, 17 *Pseudomonas* species (including *P. aeruginosa*, *Pseudomonas putida*, *Pseudomonas entomophila*, *Pseudomonas mendocina*, *P. fluorescens*, *P. syringae*,

and *Pseudomonas stutzeri*), 7 *Acinetobacter*, 3 *Psychrobacter*, and one *Cellvibrio* genome.

Gene annotations for the sequences were extracted from the GenBank files of each genome. Annotations from the GenBank files for the full genomes (including plasmids when present) were used. The properties of all the strains are listed in Table 1, and the color coding in this table is used throughout the paper.

Phylogenetic Analysis—16S rRNA and Pan Genome Family Trees

Phylogenetic analysis was performed in two stages, one using 16S rRNA sequences and the other using the pan genome protein families. For the first method, 16S rRNA sequences were predicted from the whole genome sequence using RNAmmer [26]. Single and full sequences that have the highest scores from the RNAmmer predictions for each

genome were aligned in CLUSTALW [27] program, and with this alignment, the 16S rRNA tree was generated using a neighbor-joining method with 1,000 bootstrap resamplings and viewed with MEGA4 [50] (Fig. 1). An additional tree was generated with partial 16S rRNA sequences of *Azotobacter* strains partial sequences of *Azotobacter* strains—*A. vinelandii* IAM 15004 (type strain), *Azotobacter tropicalis* KBS, *A. tropicalis* RBS, *Azotobacter nigricans* IAM 15005 (type strain), *Azotobacter armeniacus* DSM 2284 (type strain), *Azotobacter salinestris* ATCC 49674(type strain), *A. vinelandii* ISSDS-384, *Azotobacter beijerinckii* ICMP 8673 (type strain), *A. beijerinckii* ICMP 4032, *Azotobacter chroococcum* IAM 12666(type strain), *A. chroococcum* ICMP 4031, *Azomonas macrocytogenes* ICMP 8674, *Azorhizophilus paspali* ATCC 23833, and *Rhizobacter dauci* H6—which are retrieved from the Ribosomal Database Project (RDP) [12], where the same procedure as above was performed, aligned with

Table 1 List of genomes used in the comparative analysis (the colors of the groups are the same throughout all the figures)

Row	Organism	Project ID	Total size (bp)	No. of genes	16S rRNA	% GC	Reference
1	*Acinetobacter baumannii* AYE	28921	4,048,735	3,789	6	39.3	[54]
2	*Acinetobacter baumannii* AB307-0294	30993	3,760,981	3,458	6	39	[1]
3	*Acinetobacter baumannii* AB0057	21111	4,059,242	3,801	6	39.2	[1]
4	*Acinetobacter baumannii* ACICU	17827	3,996,761	3,759	6	38.9	[19]
5	*Acinetobacter baumannii* ATCC 17978	17477	4,001,457	3,368	5	38.9	[46]
6	*Acinetobacter* sp. ADP1	12352	3,598,621	3,325	7	40.4	[5]
7	*Acinetobacter baumannii* SDF	13001	3,477,996	3,577	5	39.1	[54]
8	*Psychrobacter arcticus* 273-4	9633	2,650,701	2,147	4	42.8	GenBank CP000082
9	*Psychrobacter cryohalolentis* K5	13920	3,101,097	2,511	4	42.2	[4]
10	*Psychrobacter* sp. PRwf-1	15759	2,995,049	2,385	5	44.8	GenBank CP000713
11	*Cellvibrio japonicus* Ueda107	28329	4,576,573	3,754	3	52	[14]
12	*Pseudomonas mendocina* ymp	17457	5,072,807	4,594	4	64.7	[10, 22]
13	*Pseudomonas stutzeri* A1501	16817	4,567,418	4,128	4	63.9	[58]
14	*Pseudomonas fluorescens* Pf-5	327	7,074,893	6,138	5	63.3	[37]
15	*Pseudomonas fluorescens* Pf0-1	12	6,438,405	5,736	6	60.5	[44]
16	*Pseudomonas fluorescens* SBW25	31229	7,147,633	6,487	5	60.1	[44]
17	*Pseudomonas putida* KT2440	267	6,181,863	5,350	7	61.5	[33]
18	*Pseudomonas putida* F1	13909	5,959,964	5,252	6	61.9	[38]
19	*Pseudomonas putida* GB-1	17629	6,078,430	5,409	7	61.9	GenBank CP000926
20	*Pseudomonas putida* W619	17053	5,774,330	5,182	7	61.4	GenBank CP000949
21	*Pseudomonas entomophila* L48	16800	5,888,780	5,168	7	64.2	[55]
22	*Pseudomonas aeruginosa* PAO1	331	6,264,404	5,568	4	66.6	[49]
23	*Pseudomonas aeruginosa* LESB58	31101	6,601,757	5,965	4	66.3	[56]
24	*Pseudomonas aeruginosa* UCBPP-PA14	386	6,537,648	5,892	4	66.3	[28]
25	*Pseudomonas aeruginosa* PA7	16720	6,588,339	6,286	4	66.4	[40]
26	*P. syringae* pv. tomato str. DC3000	359	6,538,260	5,614	5	58.3	[9]
27	*P. syringae* pv. syringae B728a	323	6,093,698	5,136	5	59.2	[15]
28	*P. syringae* pv. phaseolicola 1448A	12416	6,112,448	5,171	5	57.9	[21]
29	*Azotobacter vinelandii* DJ	16	5,365,318	5,051	6	65.7	[43]
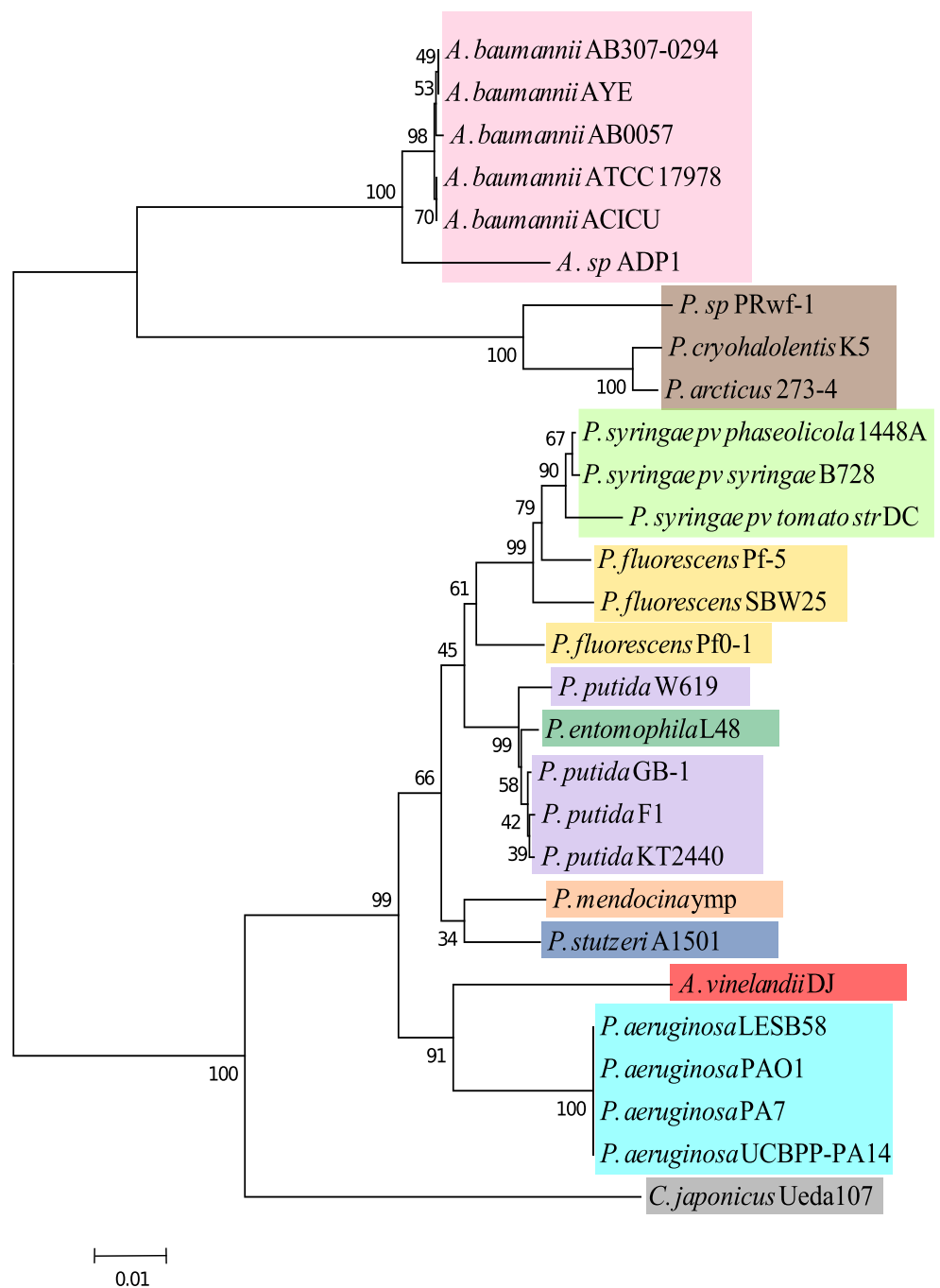
CLUSTALW, and viewed with MEGA4 [50] (see Electronic supplementary material (ESM) Fig. 1).

For the second method, the pan genome family tree (Fig. 2) of all the genomes was generated using BLASTP similarity between each proteome, as described in Snipen and Ussery's method [45]. The "50/50 rule" was used to define the homology, meaning that a protein is assumed to be in the same family if 50% of its length shows 50% sequence identity with the reference protein [51]. According to this criterion, genes that have a significant hit to each other are considered to be in one gene family. In order to see the relations between the different gene families, a matrix is constructed containing the gene families in columns and the genomes in rows, assigning 1 for the presence of that gene family in the corresponding genome and 0 otherwise. Manhattan distances are calculated from this matrix and hierarchical clustering is made. This tree shows the similarities based on the shared gene families, excluding the gene families that are represented only in one genome (ORFans) [45].
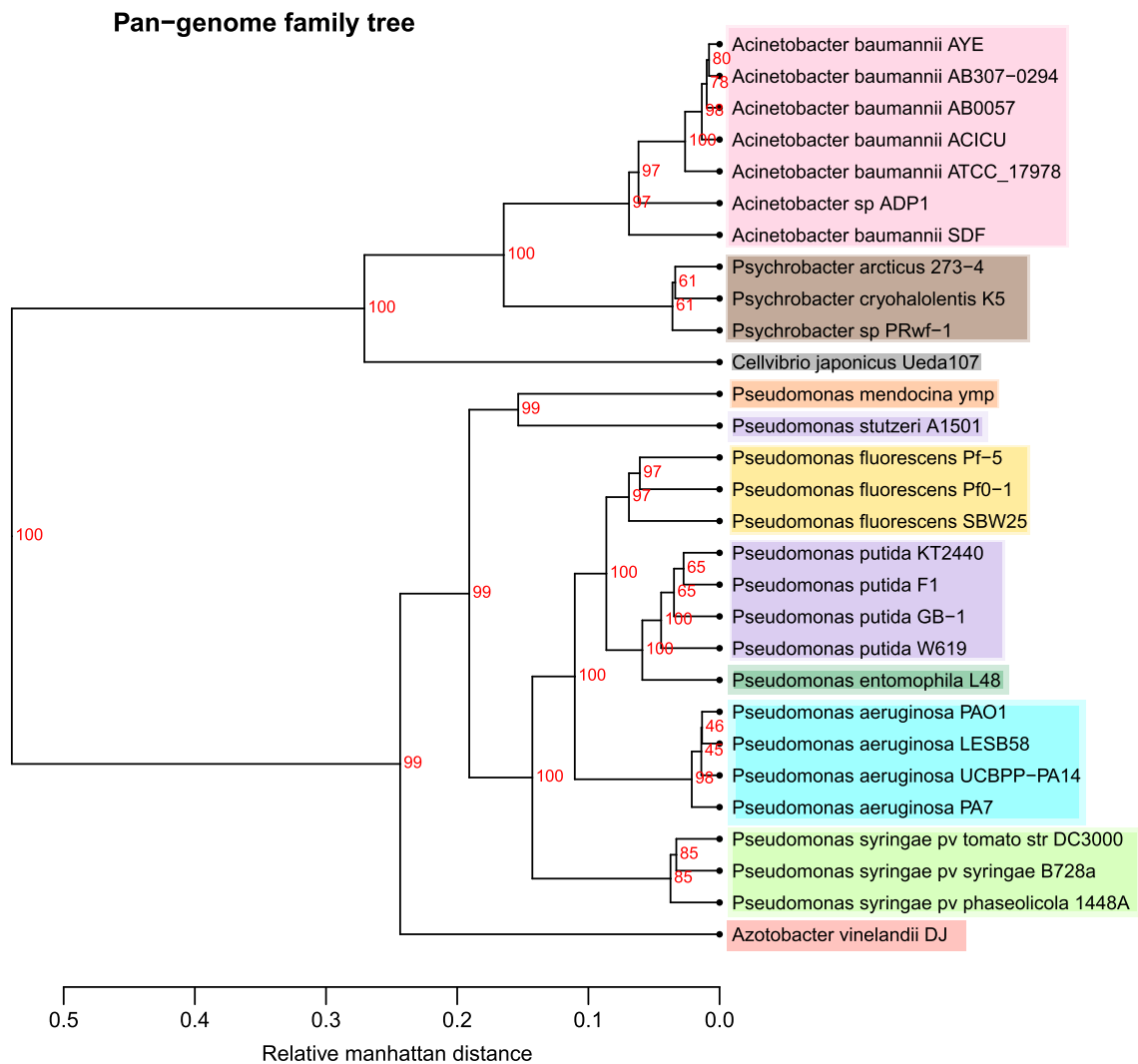
Pan−genome family tree

Figure 2 Pan genome family tree. The tree shows the phylogenetic relationships based on the gene families found in the pan genome, excluding the families found in only one genome. Color coding is again based on Table 1

Core and Pan Genome Analysis

In order to make a pan–core genome analysis, BLASTP similarity analysis was used and a pan–core genome plot was generated based on the clustering from the pan genome family tree (Fig. 3). *Pseudomonas* genomes were plotted first, followed by the other genera. *Azotobacter* was added after *P. stutzeri* and then *Acinetobacter*, *Psychrobacter*, and, finally, *Cellvibrio*. The plot goes from the first to the last column in an information accumulative manner. Each column shows the BLAST hit results against all the previous ones in terms of new genes, new gene families, and pan and core genome sizes. The accumulative number of all gene families found (blue line in Fig. 3) increases as new genomes are added, leading to the pan genome in total, whereas the number of common gene families found in all genomes (red line in Fig. 3) decreases with genome addition, leading to the core genome at the end [53].

BLASTMatrix

In this method, the protein BLAST identities were used to compare the proteome of each genome against all the other proteomes, pictured by a BLASTMatrix (Fig. 4), showing the fraction of genes shared between different genomes in the green cells [7]. The percentages in the cells are calculated by the number of gene families shared in two genomes over the union of the gene families found in both. Similarity criterion was again the 50/50 rule for the selection of significant hits and for the gene family assignment as in the pan genome family tree method. The term "homology" is also used to indicate the similarity based on shared gene families between two genomes. Protein BLAST results of a genome against itself (excluding self-hits) were used to define the protein homology within the same genome, as seen in the red cells.
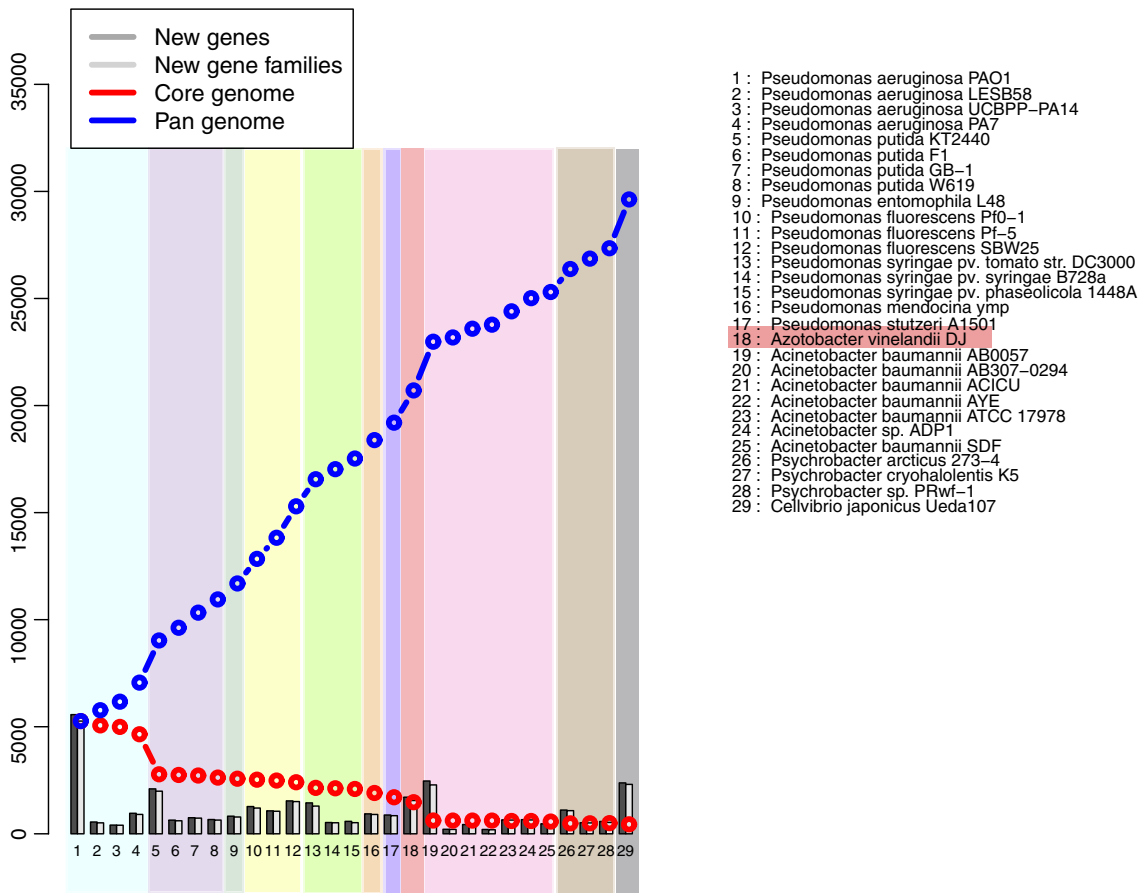
Figure 3 Core and pan genome plot of the *Pseudomonas* genomes. *Azotobacter* is *highlighted*. The number of pan genome for all strains used is 29,696 (*dashed line on top*) and the core genome is reduced to 443 (*dashed line below*). The *light purple color* indicates where *P. putida* is added, and the *light red column* indicates where *A. vinelandii* is added
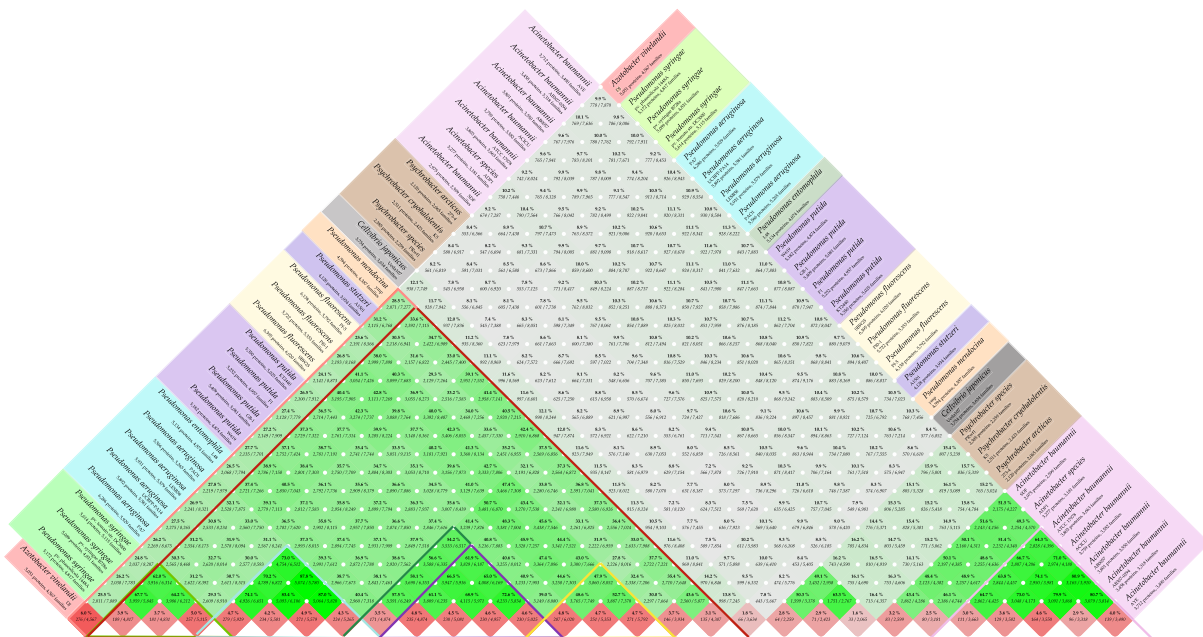

Figure 4 BLASTMatrix shows the homology between the pairs of genomes compared (*green*) and homology within each genome itself (*red*) based on the 50/50 rule

Results

16S rRNA Tree

According to the 16S rRNA phylogenetic tree in Fig. 1, *A. vinelandii* has a close relationship with *Pseudomonas*. The tree shows that the 16S rRNA sequence of this organism is very similar to *P. aeruginosa* 16S rRNA sequences, even more similar than the other *Pseudomonas* species as mentioned before in Rediers et al. [39]. There are clear clusters among the different strains of the same species, with few exceptions, and as expected, genomes from the genus *Pseudomonas* cluster together when compared with the other *Pseudomonadales*, *Acinetobacter*, and *Psychrobacter*, which are more distantly related. In general, the evolutionary relationships indicated in this tree are in agreement with the known biology of these organisms. Furthermore, alignment of sequences including more 16S rRNA sequences obtained from RDP also shows the same clustering results for *Azotobacter* (see ESM Fig. 1). All the sequences from *Azotobacter*, *Azorhizophilus*, and *Azomonas* genera cluster close to the *P. aeruginosa* group, while other *Pseudomonas* species are more distant and *Rhizobacter* is an outgroup (note that it was not specifically selected as an outgroup during the methods).

Pan Genome Family Tree

The pan genome family tree (Fig. 2) compares the presence of gene families in each genome and measures the distances for the tree depending on the common gene families found between genomes, resulting in groups which share more gene families clustering together. As expected, the *Pseudomonas* genus groups together and has clear clusters for each species, this time with often 100% bootstrap values in the main nodes for each species. Some positions of the genomes are changed, such as *Cellvibrio japonicus* and *P. fluorescens*, but, most importantly for this work, the *A. vinelandii* now is in a different position on this tree. In this figure, *A. vinelandii* does not group with any individual *Pseudomonas* species but with all of the *Pseudomonas* clusters, still indicating that it shows a larger fraction of conserved protein families with Pseudomonads than the other Gammaproteobacteria.

Core and Pan Genome Analysis

The core and pan genome analysis is another method that uses the proteomes of the genomes (Fig. 3). The plot shows the change in the number of gene families that are common to the compared genomes, the core genome, and the pan genome [51]. The bars indicate the new genes and gene families compared with a BLASTP against the genomes

that were previously added to the list. Hence, every new gene family is accounted for in the pan genome, which increases with the addition of each new genome (blue line in Fig. 3), whereas the size of the core genome is reduced (red line). The order of the genomes is related to the evolutionary distance seen on the pan genome tree.

The overall result of the core and pan genome plot shows that there are only 443 conserved core gene families found in 29 genomes. The core genome size for just the *Pseudomonas* genomes is 1,706, and after *A. vinelandii* is added, it is reduced by 231 gene families, leading to 1,475 core gene families for the first 18 genomes. The pan genome size for all the strains adds up to 29,626 gene families. The increase in pan genome size after the addition of *A. vinelandii* is 1,506 gene families, and there are roughly 1,700 genes that are designated as new. It is also seen that after the addition of *Pseudomonas putida* strains, the core genome for Pseudomonads has a steep drop with 1,870 gene families and the pan genome has a sharp increase of 1,969 more gene families. There is also a big jump of 2,275 gene families in the pan genome after the addition of *Acinetobacter* species, where the core genome drops by 855 gene families.

BLASTMatrix

The BLASTMatrix shows the shared gene families between and within the compared genomes (Fig. 4). There is a high fraction of shared genes within the same species, as denoted in darker green colors on the bottom of the matrix. The genus *Pseudomonas* can readily be distinguished, and it is highlighted with the dark red triangle. The results closely resemble the relations in the pan genome tree where *Psychrobacter* and *Acinetobacter* have a very low homology with *Pseudomonas* species. *Azotobacter* on the other hand shares as many gene families with *Pseudomonas* as some of the other members of *Pseudomonadaceae*. More specifically, it shares between 24% and 31% of its protein families with *Pseudomonas*; the maximum homology is seen with *P. stutzeri* with 31.2%, where *P. stutzeri* is homologous with other *Pseudomonas* between 28% and 34%. Also seen from the matrix is the homology within the species which is on average 79% for *P. aeruginosa*, 66% for *P. putida*, 49% for *P. fluorescens*, 65% for *P. syringae*, and 72% for *Acinetobacter baumannii*. In contrast, homology across various species within *Pseudomonas* indicates a level between 30% and 50%.

Discussion

The results of using different comparative genomic methods to understand the phylogeny of *Pseudomonas* indicate a

considerably close evolutionary relationship with *A. vinelandii*. Both in the 16S rRNA tree and pan genome family tree, *A. vinelandii* is clustered together or within the *Pseudomonas* species. It is not as clear whether or not *A. vinelandii* should be classified as being closest to *P. aeruginosa*. Although they are clustered together in the 16S rRNA tree, the low resolution of the 16S rRNA phylogenetic analysis has been noted by others [39, 59]. It does show, however, that they have a common ancestral *rrs* gene that is closer to each other than to other species. The outcome of the 16S rRNA tree also shows that members of the Moraxellaceae family (*Acinetobacter* and *Psychrobacter*) are clustered together and Pseudomonodaceae (*Pseudomonas*, *Azotobacter*, and *Cellvibrio*) are on another clade. In the pan genome family tree, on the other hand, clear clusters of each species can be seen. For example, *P. fluorescens* strains are in a group rather than separated as in the rRNA tree. Since this tree shows the distances of each group according to how many gene families they share, it is clear that *A. vinelandii* shares more gene families with the *Pseudomonas* group than it does with the other Pseudomonadales members used in this comparison, especially when compared with *Cellvibrio*, which is also another genus in the same family. Since the results are restricted to only one *Azotobacter* genome, other supportive results are crucial in order to understand the true relationship.

The core and pan genome analysis reveals the set of conserved gene families ("core") and the total number of gene families ("pan genome") for the set of sequenced genomes compared. The core genome refers to the idea of a backbone genome for organisms in the same genus; as more species from the same genus are added, the core is expected to approach a stable plateau in the same genus. The pan genome is, however, very flexible in size and, for some bacteria, can be quite large [53]. For instance, the pan genome of *Pseudomonas* is more than ten times larger than its core genome [25]. Figure 3 shows that the *Pseudomonas* core genome size does not have a dramatic change with the addition of *A. vinelandii*. Although the pan genome has a slight increase, such a small change in the core genome size is not expected from an organism that is in a different genus. The sharp decrease in the core genome when *P. putida* is added is due to its position on the list, being right after *P. aeruginosa*, which is clearly in another group in the phylogeny trees. However, after 18 genomes, the addition of a new genus (in this case, an *Acinetobacter* species) creates a big difference in the plot, reducing the core more than half in size and expanding the pan genome clearly by 10%, which is an obvious result of the addition of a different genus on the list.

The BLASTMatrix also strongly agrees with the pan genome family tree and pan–core genome plot as they are all comparing the gene families across the genomes.

Among them, the BLASTMatrix provides more quantitative results on the similarities such as percentages or number of gene families. According to this, *Azotobacter* has a high fraction of shared protein families with *Pseudomonas*, mostly with *P. stutzer*i having 31%. They also share a similar fraction of protein family levels with other *Pseudomonas*, with *A. vinelandii* having on average 25% homology and *P. stutzeri* with on average 30%. The relation between these two organisms might be because of the similar functions that they have in their environment as they are both free-living, root-associated, nitrogen-fixing bacteria [58]. However, phylogenetic analysis, based on 16S rRNA in this paper and in the other works also based on housekeeping genes, suggests that *A. vinelandii* could have a closer evolutionary relationship with *P. aeruginosa*. Taking into account that *P. aeruginosa* is the type species for *Pseudomonas* and that it shares a common evolutionary history on the conserved genes with *A. vinelandii*, all of these results strongly suggests that *A. vinelandii* has a *Pseudomonas*-like backbone including the conserved genes, while the diversity among them is caused by adaptive strategies during their evolution which comes from the transfer of genetic material.

It should be noted that complete genomes were chosen in the dataset for the reliability of the methods. For many of the incomplete genomes in the database, 16S rRNA sequences are missing or partial, and so as the many protein sequences. In a study where comparative analysis mostly relies on the BLAST comparison on the proteomes, partial or poorly annotated sequences cause artifacts in the search, which leads to unreliable conclusions. Another aspect from the taxonomical point of view is that in theory, a good classification method should be able to show the relations of organisms regardless of the number of data. Hence, having more *Pseuodmonas* genomes should not be the effecting factor in the discrimination of the organisms, especially at the genus level. On the other hand, having more *Azotobacter* genomes would be more reliable in terms of reducing the effect of sequencing, assembly, and annotation errors.

Concluding remarks

The increase in the availability of genome sequences of different organisms makes comparative genomic analysis a fundamental part of research on evolutionary relations between organisms and the genetic basis of their diversity [53]. This applies as well to *Pseudomonas*. Looking at the comparative genomic analysis of nucleotide and the coded protein sequences of *Pseudomonas*, we can easily see the distinction between *Pseudomonas* species. This distinction is supported by the physical and biochemical classification that has been established over the last 100 years.

Although there is only one *Azotobacter* genome available, general assumptions from these comparative analyses that rely on the extensively studied algorithms and tools are worth examining. Support from similar analysis on the comparison of these genera is also taken into account. In conclusion, we suggest that *A. vinelandii* has a *Pseudomonas*-like backbone genome where the core functions of these groups are same. There are various lines of evidence which lead to this: the finding that *A. vinelandii* has approximately a third of the same protein families, clustering with the whole *Pseudomonas* clade on the pan genome family tree rather than being in the *P. aeruginosa* clade, and not causing a big drop on the core genome size. All three of these observations are consistent with the idea of having the same backbone, perhaps the same origins, but different adaptations throughout their evolution. This leads us to the question of the boundaries of new member assignments in the *Pseudomonas* genus. Where do we draw the line? We propose that based on whole-genome analysis, it is possible to better differentiate members of a genus by looking at their core genome properties. The standards for classification should be set using comparative methods on both DNA and protein levels. If this is the case, the *Azotobacter* can be assigned to the *Pseudomonas* genus. Perhaps, for future investigations, a detailed analysis of the core genes can be made and functionally categorized to see the background of their similarity.

References

1. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ et al (2008) Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. J Bacteriol 190:8053–8064

2. Ait Tayeb L, Ageron E, Grimont F, Grimont PAD (2006) Molecular phylogeny of the genus *Pseudomonas* based on rpoB sequences and application for the identification of isolates. Res Microbiol 156:763–773

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

4. Bakermans C, Tsapin AI, Souza-Egipsy V, Gilichinsky DA, Nealson KH (2003) Reproduction and metabolism at −10°C of bacteria isolated from Siberian permafrost. Environ Microbiol 5:321–326

5. Barbe V, Vallenet D, Fonknechten N, Kreimeyer A, Oztas S, Labarre L et al (2004) Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res 32:5766–5779

6. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Res 36:D25–D30

7. Binnewies TT, Hallin PF, Staerfeldt HH, Ussery DW (2005) Genome Update: proteome comparisons. Microbiology 151:1–4

8. Bodilis J, Hedde M, Orange N, Barray S (2006) OprF polymorphism as a marker of ecological niche in *Pseudomonas*. Environ Microbiol 8:1544–1551

9. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML et al (2003) The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc Natl Acad Sci USA 100:10181–10186

10. Chi CY, Lai CH, Fung CP, Wang JH (2005) *Pseudomonas mendocina spondylodiscitis*: a case report and literature review. Scand J Infect Dis 37:950–953

11. Clementi F (1997) Alginate production by *Azotobacter vinelandii*. Crit Rev Biotechnol 17:327–361

12. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141–D145

13. De Vos D, Bouton C, Sarniguet A, De Vos P, Vauterin M, Cornelis P (1998) Sequence diversity of the *oprI* gene, coding for major outer membrane lipoprotein I, among rRNA group I pseudomonads. J Bacteriol 180:6551–6556

14. DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB et al (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. J Bacteriol 190:5455–5463

15. Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A et al (2005) Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc Natl Acad Sci USA 102:11064–11069

16. Frapolli M, Défago G, Moënne-Loccoz Y (2007) Multilocus sequence analysis of biocontrol fluorescent *Pseudomonas* spp. producing the antifungal compound 2,4-diacetylphloroglucinol. Environ Microbiol 9:1939–1955

17. He J, Baldini RL, Déziel E, Saucier M, Zhang Q, Liberati NT et al (2004) The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. Proc Natl Acad Sci USA 101:2530–2535

18. Høiby N, Ciofu O, Bjarnsholt T (2010) *Pseudomonas aeruginosa* biofilms in cystic fibrosis. Future Microbiol 5:1663–1674

19. Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJP et al (2008) Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. Antimicrob Agents Chemother 52:2616–2625

20. Jelsbak L, Johansen HK, Frost AL, Thogersen R, Thomsen LE, Ciofu O et al (2007) Molecular epidemiology and dynamics of *Pseudomonas aeruginosa* populations in lungs of cystic fibrosis patients. Infect Immun 75:2214–2224

21. Joardar V, Lindeberg M, Jackson RW, Selengut J, Dodson R, Brinkac LM et al (2005) Whole-genome sequence analysis of

Pseudomonas syringae pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. J Bacteriol 187:6488–6498

22. Kao CM, Liu JK, Chen YL, Chai CT, Chen SC (2005) Factors affecting the biodegradation of PCP by *Pseudomonas mendocina* NSYSU. J Hazard Mater 124:68–73

23. Kennedy C, Bishop P (2005) Genetics of nitrogen fixation and related aspects of metabolism in species of *Azotobacter*: history and current status. In: Klipp W, Masepohl B, Gallon JR, Newton WE (eds) Genetics and regulation of nitrogen fixation in free-living bacteria. Kluwer, Dordrecht, pp 27–52

24. Kersters K, Ludwig W, Vancanneyt M, Devos P, Gillis M, Schleifer KH (1996) Recent changes in the classification of the pseudomonads: an overview. Syst Appl Microbiol 19:465–477

25. Kiil K, Binnewies TT, Willenbrock H, Hansen SK, Yang L, Jelsbak L et al (2008) Comparative genomics of *Pseudomonas*. In: Rehm BHA (ed) *Pseudomonas*. Wiley-VCH, Weinheim, pp 1–24

26. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes Tr, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108

27. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

28. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S et al (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol 7:R90

29. Lineweaver H, Burk D (1934) The determination of enzyme dissociation constants. J Am Chem Soc 56:658–666

30. Meyer JM, Gruffaz C, Fischer-LeSaux M (2008) Siderotyping, a straightforward tool to identify soil and plant-related pseudomonads. In: Nautiyal CS, Patrice D (eds) Molecular mechanisms of plant and microbe coexistence. Springer, Berlin, pp 369–382

31. Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, Buffière A, Yan S, Dominguez H, Thompson BM (2008) The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. ISME J 2:321–334

32. Moore ERB, Mau M, Arnscheidt A, Boettger EC, Hutson RA, Collins MD, Van de Peer Y, De Wachter R, Timmis KN (1996) The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationship. Syst Appl Microbiol 19:478–492

33. Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VAP et al (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4:799–808

34. Palleroni NJ, Kunisawa R, Contopoulou R, Doudoroff M (1973) Nucleic acid homologies in the genus *Pseudomonas*. Int J Syst Bacteriol 23:333–339

35. Palleroni NJ (2005) Genus I. *Pseudomonas* Migula 1894, 237^AL. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, Boone DR, Vos P et al (eds) Bergey's manual of systematic bacteriology. Springer, Boston

36. Palleroni NJ (2008) The road to the taxonomy of *Pseudomonas*. In: Cornelis P (ed) *Pseudomonas*. Genomics and molecular biology. CaisterAcademic, Norfolk, pp 1–18

37. Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV et al (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. Nat Biotechnol 23:873–878

38. Phoenix P, Keane A, Patel A, Bergeron H, Ghoshal S, Lau PCK (2003) Characterization of a new solvent-responsive gene locus in *Pseudomonas putida* F1 and its functionalization as a versatile biosensor. Environ Microbiol 5:1309–1327

39. Rediers H, Vanderleyden J, de Mot R (2004) *Azotobacter vinelandii*: a *Pseudomonas* in disguise? Microbiology 150:1117–1119

40. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, Dodson R, Harkins D, Shay R, Watkins K, Mahamoud Y, Paulsen IT (2010) Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. PLoS One 5(1):e8842

41. Sadoff HL (1975) Encystment and germination in *Azotobacter vinelandii*. Bacteriol Rev 39:516–539

42. Schleifer KH (2009) Classification of *Bacteria* and *Archaea*: past, present and future. Syst Appl Microbiol 32:533–542

43. Setubal JC, dos Santos P, Goldman BS, Ertesvåg H, Espin G, Rubio LM et al (2009) Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J Bacteriol 191:4534–4545

44. Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM et al (2009) Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. Genome Biol 10:R51

45. Snipen L, Ussery D (2010) Pan-genome trees. Stand Genomic Sci 2:135–141

46. Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M (2007) New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev 21:601–614

47. Stanier RY, Palleroni NJ, Doudoroff M (1966) The aerobic pseudomonads: a taxonomic study. J Gen Microbiol 43:159–271

48. Stevenson LH, Socolofsky MD (1966) Cyst formation and poly-{beta}-hydroxybutyric acid accumulation in *Azotobacter*. J Bacteriol 91:304–310

49. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ et al (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature 406:959–964

50. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

51. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D et al (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102:13,950–13,955

52. Tindall BJ, Kampfer P, Euzeby JP, Oren A (2006) Valid publication of names of prokaryotes according to the rules of nomenclature: past history and current practice. Int J Syst Evol Microbiol 56:2715–2720

53. Ussery DW, Wassenaar TM, Borini S (2009) Computing for comparative microbial genomics. Springer, London

54. Vallenet D, Nordmann P, Barbe Vr, Poirel L, Mangenot S, Bataille E et al (2008) Comparative analysis of *Acinetobacters*: three genomes for three lifestyles. PLoS One 3:e1805

55. Vodovar N, Vallenet D, Cruveiller Sp, Rouy Z, Barbe Vr, Acosta C et al (2006) Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. Nat Biotechnol 24:673–679

56. Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F et al (2009) Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool epidemic strain of *Pseudomonas aeruginosa*. Genome Res 19:12–23

57. Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S (2000) Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. Microbiology 146:2385–2394

58. Yan Y, Yang J, Dou Y, Chen M, Ping S, Peng J et al (2008) Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. Proc Natl Acad Sci USA 105:7564–7569

59. Young JM, Park DC (2007) Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. Int J Syst Evol Microbiol 57:2894–2901