



Article

Reliable Identification of Environmental *Pseudomonas* Isolates Using the *rpoD* Gene

Léa Girard ¹, Cédric Lood ^{1,2}, Hassan Rokni-Zadeh ³, Vera van Noort ^{1,4}, Rob Lavigne ² and René De Mot ^{1,*}

¹ Centre of Microbial and Plant Genetics, KU Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium; lea.girard@kuleuven.be (L.G.); cedric.lood@kuleuven.be (C.L.); vera.vannoort@kuleuven.be (V.v.N.)

² Department of Biosystems, Laboratory of Gene Technology, KU Leuven, Kasteelpark Arenberg 21, 3001 Leuven, Belgium; rob.lavigne@kuleuven.be

³ Zanjan Pharmaceutical Biotechnology Research Center, Zanjan University of Medical Sciences, 45139-56184 Zanjan, Iran; hassan.roknizadeh@gmail.com

⁴ Institute of Biology, Leiden University, Sylviusweg 72, 2333 Leiden, The Netherlands

* Correspondence: rene.demot@kuleuven.be; Tel.: +32-16329681

Received: 23 June 2020; Accepted: 28 July 2020; Published: 31 July 2020



Abstract: The taxonomic affiliation of *Pseudomonas* isolates is currently assessed by using the 16S rRNA gene, MultiLocus Sequence Analysis (MLSA), or whole genome sequencing. Therefore, microbiologists are facing an arduous choice, either using the universal marker, knowing that these affiliations could be inaccurate, or engaging in more laborious and costly approaches. The *rpoD* gene, like the 16S rRNA gene, is included in most MLSA procedures and has already been suggested for the rapid identification of certain groups of *Pseudomonas*. However, a comprehensive overview of the *rpoD*-based phylogenetic relationships within the *Pseudomonas* genus is lacking. In this study, we present the *rpoD*-based phylogeny of 217 type strains of *Pseudomonas* and defined a cutoff value of 98% nucleotide identity to differentiate strains at the species level. To validate this approach, we sequenced the *rpoD* of 145 environmental isolates and complemented this analysis with whole genome sequencing. The *rpoD* sequence allowed us to accurately assign *Pseudomonas* isolates to 20 known species and represents an excellent first diagnostic tool to identify new *Pseudomonas* species. Finally, *rpoD* amplicon sequencing appears as a reliable and low-cost alternative, particularly in the case of large environmental studies with hundreds or thousands of isolates.

Keywords: species identification; taxonomy; phylogeny; diversity; genome sequencing

1. Introduction

Pseudomonas species are ubiquitous bacteria present in terrestrial, aquatic, and marine environments [1,2]. While several species are a threat to human health or food industry, with detrimental impacts on crops and aquaculture, the large majority of the species are commensals [3–5]. In fact, several species have been described as potential biocontrol agents to fight against diverse plant pathogens but also for bioaccumulation or biodegradation of pollutants [6,7]. Furthermore, *Pseudomonas* species are known producers of a wide diversity of bioactive secondary metabolites with potentially high added value [8]. Consequently, it is of great interest to be able to identify *Pseudomonas* isolates, in a fast and inexpensive way, to monitor their occurrence and diversity in the environment.

To date, the identification of bacteria in microbiology is still based on the 16S rRNA gene sequence, although its discriminative power is often limited to the delineation of groups or clades within a particular genus [9]. For *Pseudomonas*, the 16S rRNA gene allows the differentiation with sister genera in the *Pseudomonadaceae* family (*Cellvibrio*, *Oblitimonas*, *Thiopseudomonas*, and *Ventosimonas*) and

the delineation of the three main lineages (*P. aeruginosa*, *P. fluorescens*, and *P. pertucinogena*) [10–12]. However, intra-genomic heterogeneities originate from multiple copies of the 16S rRNA gene in *Pseudomonas* genomes, and, based on previously established cutoff values (between 98.2 to 99% of similarity), it is not possible to differentiate environmental isolates at the species level [9,12–14]. Therefore, MultiLocus Sequence Analysis (MLSA), based on the concatenation of four housekeeping genes (16S rRNA, *gyrB*, *rpoB*, and *rpoD*), has been proposed for the identification of *Pseudomonas* isolates [15,16]. Nevertheless, the expansion of genomics in bacterial taxonomy, and attractive prices, have led to the genome sequencing of many environmental isolates of *Pseudomonas* [17,18]. Recently, the *Pseudomonas* phylogeny based on the 16S rRNA gene and the concatenation of 4, 100, and 120 genes was compared to whole genome analyses [12]. As a result, it seems that the MLSA based on four housekeeping genes has the best price–performance ratio, although the use of only two housekeeping genes (*rpoB* and *rpoD*) can also be found in the literature [19,20]. However, these methodologies are time-consuming, demanding, and/or expensive, particularly in the case of large environmental studies with hundreds or thousands of isolates. Consequently, the use of the 16S rRNA gene to identify *Pseudomonas* isolates remains relevant despite being prone to misidentifications, as observed frequently in public databases [21,22].

The *rpoD* gene is included in most MLSA procedures and was previously proposed to identify *Pseudomonas* species in environmental samples [23,24] or for the rapid identification of isolates belonging to the *Pseudomonas syringae* complex [25]. However, since these studies included only a limited number of *Pseudomonas* species, a genus-wide comprehensive overview of the *rpoD* phylogeny is presently missing. In this study, we present the *rpoD*-based *Pseudomonas* phylogeny, including a total of 217 type strains for which genomes were available in public databases. Secondly, we adapted the methodology described by Mulet et al. (2011) to sequence the *rpoD* gene and identify 145 environmental *Pseudomonas* isolates. To test our *rpoD*-based taxonomic affiliations, we sequenced one-third of these isolates with Illumina and used whole genome analysis (Average Nucleotide Identity and digital DNA–DNA hybridization) for comparison with the established taxonomy. We show here that the *rpoD* locus streamlines the identification of *Pseudomonas* isolates from a labor and cost perspective and provides, in comparison with the 16 rRNA gene, an excellent tool to accurately affiliate isolates.

2. Materials and Methods

All type strains used in this study, together with NCBI accession numbers, are listed in Table S1. The environmental isolates of *Pseudomonas*, their origin, and their NCBI accession number (*rpoD* and/or genome) are listed in Table S2.

To avoid DNA extractions, PCR was performed on cell lysates. One colony of each strain was suspended in 50 μ L of Milli-Q water in 96-well PCR microplates, and plates were immersed three times, alternating between liquid nitrogen and water bath (+70 $^{\circ}$ C). The lysates were then stored at -20 $^{\circ}$ C until their use as PCR templates. PCR amplifications of the *rpoD* gene were performed using previously designed primers (PsEG30F and PsEG790R; [23]) and KAPA2G Fast HotStart ReadyMix (Sigma–Aldrich, Saint-Louis, Missouri, USA). Cycling conditions were as follows: initial denaturation at 95 $^{\circ}$ C for 5 min followed by 30 cycles of annealing at 60 $^{\circ}$ C for 30 s, extension at 72 $^{\circ}$ C for 30 s and denaturation at 95 $^{\circ}$ C for 15 s, and reactions were completed at 72 $^{\circ}$ C for 2 min. PCR products were then purified using the GenElute PCR Clean-Up kit (Sigma–Aldrich Saint-Louis, Missouri, USA). Sequencing was performed using the reverse primer PsEG790R first, and, in case of failed sequencing, the forward primer PsEG30F was used. Purified PCR products were then sequenced using Sanger sequencing (Macrogen Europe, Amsterdam, The Netherlands) to obtain a final fragment of approximately 650 bp. The *rpoD* sequences of 145 environmental isolates were aligned to those from the 217 type strains of *Pseudomonas* to generate a similarity matrix based on a ~650 bp fragment (Table S3). A first phylogenetic tree containing all type strains of *Pseudomonas* and members of the sister genera was constructed to confirm the relatedness between validly published species (Figure 1; Table S1). A second, restricted tree shows the phylogenetic relationship of our 145 *Pseudomonas* isolates with the known *Pseudomonas*

species (Figure 2; Table S2). MEGA-X was used to estimate the best evolutionary model. Both trees were constructed using the general time-reversible model (GTR+G+I), and bootstrap values were calculated based on 1000 replications [26]. iTOL was then used to annotate the trees and create high-quality figures [27].

To validate the *rpoD*-based taxonomic affiliation, the whole genomes of 55 environmental isolates were sequenced. Three strains, namely SWRI103, BW11P2, and BW11M1, were included in this analysis but were previously sequenced by our group [28–30]. Briefly, genomic DNA was extracted using the Genra Puregene Yeast/Bact. Kit (Qiagen, Hilden, Germany). A first subset of the genomes was sequenced by BASECLEAR (Leiden, The Netherlands) using the Nextera XT library preparation kit and the Illumina MiSeq sequencer. The second batch of strains was sequenced in-house using the Nextera Flex preparation kit and the Illumina MiniSeq device. All libraries were sequenced using a paired-end approach (2 × 150 bp), and the genome coverage was routinely above 40×. The quality of the Illumina reads was assessed using FastQC v. 0.11.9 and Trimmomatic v. 0.38 for adapter clipping, quality trimming (LEADING:3 TRAILING:3 SLIDINGWINDOW:4.15), and minimum length exclusion (>50 bp) [31]. De novo genome assembly was performed with the SPAdes assembler v. 3.13.0 [32].

Bioinformatics calculations, such as the Average Nucleotide Identity (ANI) and digital DNA–DNA Hybridization (dDDH), are now commonly used to describe new species or to affiliate strains to a specific taxon [20,33]. Threshold values for species delineation are, respectively, between 95 and 96.5 for ANI and 70 for dDDH [13,34–37]. In this study, we calculated ANIb values between a total of 275 *Pseudomonas* strains (217 type strains and 58 environmental strains) using pyani v0.2.10 with default parameters and the ANIb method. ANIb values in Tables 1–3 are the averages of the bidirectional comparisons. In the case of ambiguous ANIb values (between 95 and 96.5%), we supplemented the analysis with the calculation of dDDH using the online tool Genome-to-Genome Distance Calculator GGDC (<https://ggdc.dsmz.de/home.php>, June 2020) [34,35]. Finally, in order to evaluate reliability of *rpoD*-based taxonomic affiliations, we calculated the correlation *rpoD*/ANIb using the statistical functions included in the SciPy Python package [38].

3. Results and Discussion

3.1. *rpoD* Phylogeny

The *rpoD* phylogeny is based on 217 type strains of *Pseudomonas*, six representative strains from sister genera (*Cellvibrio*, *Oblitimonas*, *Thiopseudomonas*, and *Ventosimonas*), and two *Azotobacter* type strains (Figure 1, Table S3). The *rpoD* allows the differentiation from members of sister genera (nucleotide identities ranging from 42.51 to 65.68%) and the discrimination of the three *Pseudomonas* lineages (*P. pertucinogena*, *P. aeruginosa*, and *P. fluorescens*) (Figure 1 and Table S3). Within the three lineages, the *rpoD*-based grouping of type strains accurately reflects previously defined groups and subgroups, based on the concatenation of 4 or 100 genes [12,16,37]. The *P. aeruginosa* lineage is divided in eight phylogenetic groups (*P. oryzihabitans*, *P. stutzeri*, *P. oleovorans*, *P. aeruginosa*, *P. resinovorans*, *P. linyingensis*, *P. anguilliseptica*, and *P. straminea*). The *P. fluorescens* lineage is divided in five phylogenetic groups (*P. putida*, *P. asplenii*, *P. lutea*, *P. syringae*, and *P. fluorescens*), and overall these groups are supported by relatively high bootstrap values (Figures 1 and 2). The genus *Azotobacter* is included in the *P. aeruginosa* lineage with highest similarities with members of the *P. resinovorans* group (74.55 to 76.51%; (Table S3)). The *P. lutea* and *P. syringae* groups are included in the *P. fluorescens* group, and a total of 16 species are scattered across the three (i.e., *P. hussainii*, *P. caeni*, *P. indica*, *P. mangroviangrove*, *P. pohangensis*, *P. flexibilis*, *P. tuomuerensis*, *P. fluvialis*, *P. pharmacofabricae*, *P. thermotolerans*, *P. alcaligenes*, *P. kuykendallii*, *P. massiliensis*, *P. coleopterorum*, *P. rhizosphaerae*, and *P. turukhanskensis*). Conversely, *P. sichuanensis*, *P. guangdongensis*, and *P. cuatrocienegasensis* are integrated, respectively, in the *P. putida*, *P. linyingensis*, and *P. anguilliseptica* groups [12,16,37].

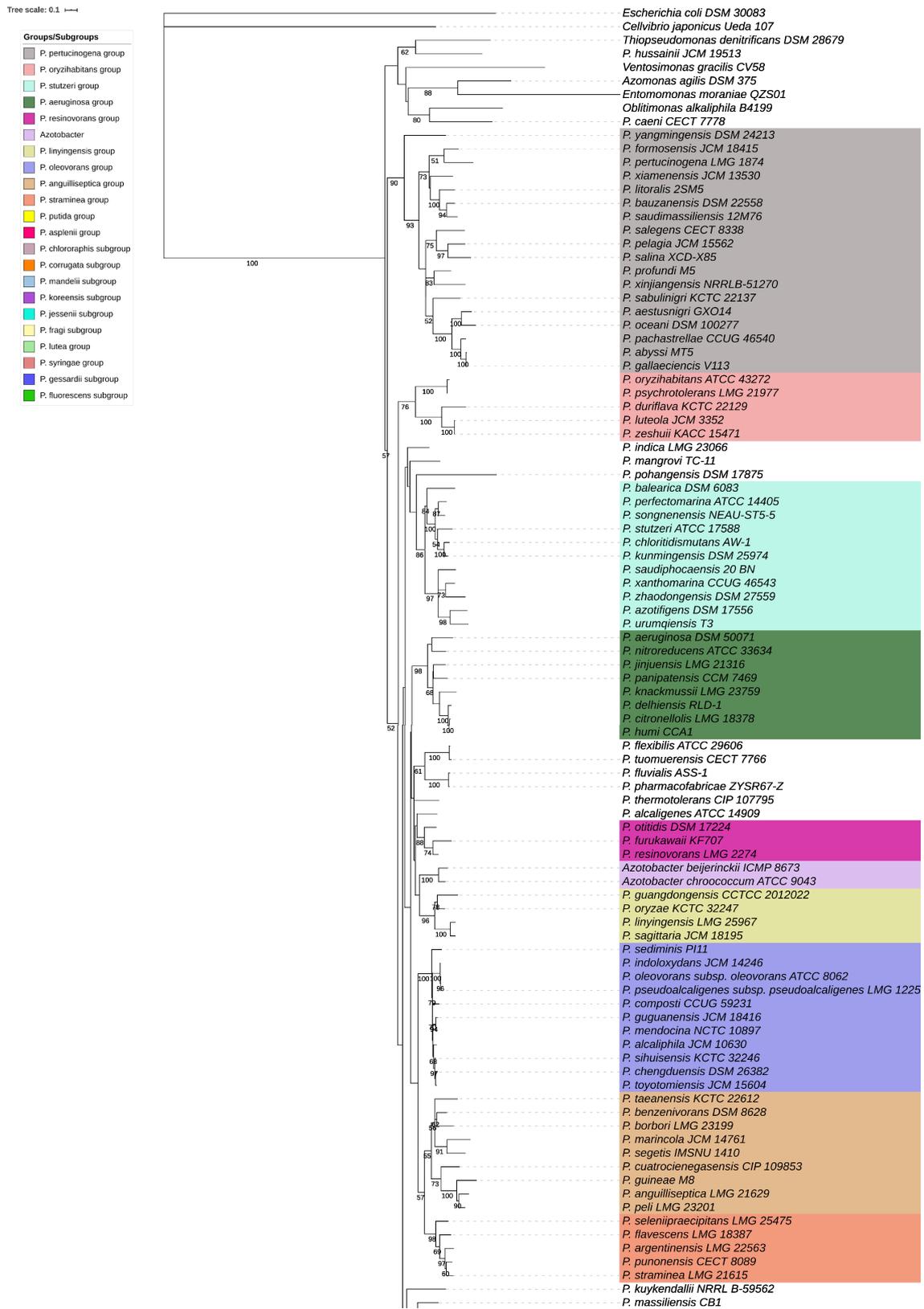


Figure 1. Cont.

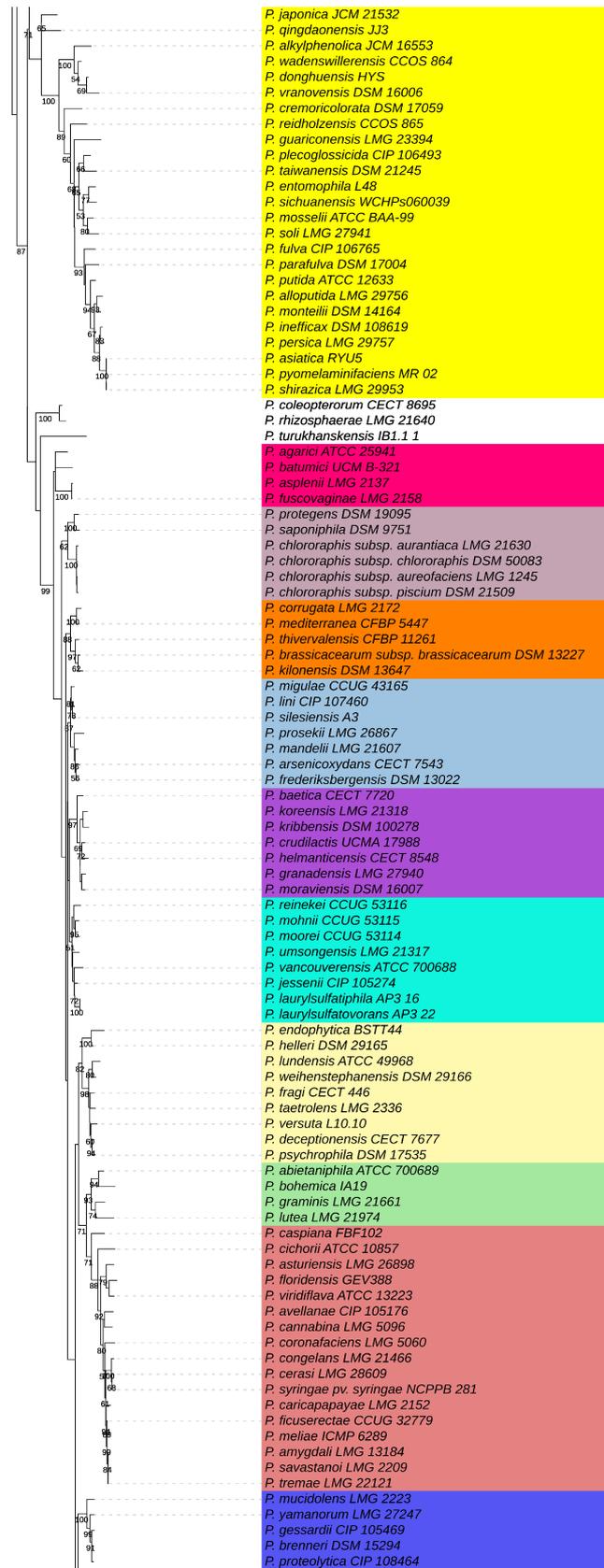


Figure 1. Cont.



Figure 1. Maximum likelihood phylogenetic tree based on a ~650 bp fragment of the *rpoD* gene for 226 type strains (217 *Pseudomonas*, six from sister genera, and two *Azotobacter*). The tree was constructed using the GTR+G+I model (MEGA-X), and only bootstrap values higher than 50% are indicated. The *rpoD* sequence of *Escherichia coli* was included as an outgroup.

3.2. Species Delineation

Within the *Pseudomonas* genus, the lowest nucleotide identity based on the *rpoD* locus among the 217 type strains included in this study was 51.57%. We observed that the lowest intra-group nucleotide identity ranged from 66.06 to 88.54% within, respectively, the *P. pertucinogena* and *P. oleovorans* group. Overall, all type strains were differentiated at a species level with a cutoff value of 98% identity (Table S3), which is supported by ANIb values <95% (Table S4). On the other hand, a certain number of *Pseudomonas* species were already described as synonymous, and we highlighted, in Table 1, ten cases where similarities based on *rpoD* (>98%) and ANIb values (>96.5%) were consistent with previous observations [12,16,33,37].

Table 1. *Pseudomonas* species considered as synonymous (Tables S3 and S4).

<i>Pseudomonas</i> Species		<i>rpoD</i>	ANIb	Reclassification
<i>P. gallaeciensis</i>	<i>P. abyssi</i>	99.24	97.56	<i>P. gallaeciensis</i>
<i>P. citronellolis</i>	<i>P. humi</i>	99.38	96.7	<i>P. citronellolis</i>
<i>P. flexibilis</i>	<i>P. tuomuerensis</i>	98.01	98.69	<i>P. flexibilis</i>
<i>P. fluvialis</i>	<i>P. pharmacofabricae</i>	98.44	98.61	<i>P. fluvialis</i>
<i>P. chengduensis</i>	<i>P. shuiensis</i>	98.79	96.24	<i>P. chengduensis</i>
<i>P. oleovorans</i> ^a	<i>P. pseudoalcaligenes</i> ^b	99.85	97.16	<i>P. oleovorans</i>
<i>P. luteola</i>	<i>P. zeshuui</i>	98.92	97.87	<i>P. luteola</i>
<i>P. asiatica</i>	<i>P. pyomelaninifaciens</i>	100	99.03	<i>P. asiatica</i>
	<i>P. shirazica</i>	99.85	99.17	
<i>P. amygdali</i>	<i>P. ficuserectae</i>	99.08	97.42	<i>P. amygdali</i>
	<i>P. meliae</i>	99.23	98.27	
	<i>P. savastanoi</i>	99.54	98.75	
<i>P. asplenii</i>	<i>P. fuscovaginae</i>	99.38	98.23	<i>P. asplenii</i>

^a *P. oleovorans* subsp. *oleovorans*, ^b *P. pseudoalcaligenes* subsp. *pseudoalcaligenes*.

However, we also observed limitations where similarities based on *rpoD* were not concordant with ANIb values (Table 2). MLSA was described as an efficient tool to identify *Pseudomonas* isolates,

and a threshold value of 97% was recently proposed to differentiate strains at the species level [12]. Consequently, we also performed MLSA (following previous instructions [15]) to determine if these misinterpretations could be avoided by the use of multiple loci (Table S5).

Table 2. Inconsistent species affiliations based on the comparison between *rpoD*, 4-genes MultiLocus Sequence Analysis (MLSA), and ANIb (Tables S3–S5). Species differentiation based on *rpoD* < 98%, MLSA < 97%, and ANIb < 96.5 were used.

<i>Pseudomonas</i> Species		<i>rpoD</i>	4-Genes MLSA	ANIb	Reclassification
<i>P. chengduensis</i>	<i>P. toyotomiensis</i>	98.76	97.74	94.48	-
<i>P. indoloxidans</i>	<i>P. oleovorans</i> ^a	98.61	97.51	95.78 ^c	-
	<i>P. pseudoalcaligenes</i> ^b	98.45	97.48	95.42 ^c	-
<i>P. chloritidismutans</i>	<i>P. kunmingensis</i>	94.79	98.24	96.47 ^d	<i>P. chloritidismutans</i>
<i>P. oryzihabitans</i>	<i>P. psychrotolerans</i>	97.4	98.60	98.22	<i>P. oryzihabitans</i>
<i>P. grimontii</i>	<i>P. marginalis</i>	99.23	98.41	93.46	-
	<i>P. panacis</i>	98.31	97.78	88.20	-
<i>P. veronii</i>	<i>P. panacis</i>	94.3	95.96	99.95	<i>P. veronii</i>
<i>P. tremae</i>	<i>P. coronafaciens</i>	92.45	93.86	98.73	<i>P. tremae</i>
<i>P. tremae</i>	<i>P. amygdali</i>	99.54	99.20	85.85	-
	<i>P. ficuserectae</i>	98.92	97.72	85.74	-
	<i>P. meliae</i>	99.08	98.40	85.99	-
	<i>P. savastanoi</i>	99.38	97.58	85.91	-
<i>P. libanensis</i>	<i>P. synxantha</i>	98.15	98.49	95.25 ^c	-
<i>P. guguaneensis</i>	<i>P. mendocina</i>	98.14	94.99	89.21	-

^a *P. oleovorans* subsp. *oleovorans*, ^b *P. pseudoalcaligenes* subsp. *pseudoalcaligenes*, ^c dDDH < 70%, ^d dDDH > 70%.

It is important to note that in 12 cases out of 15, MLSA and *rpoD* led to the same misinterpretations, while in three cases, incorrect affiliations were made by using the single locus. Overall, among the 23,436 pairwise comparisons, including the 217 type strains, only in 0.017% of the cases two strains belonging to the same species presented an identity < 98%, and in 0.047% of the cases, two strains belonging to different species presented an identity > 98%. In order to evaluate the overall reliability of these *rpoD*-based affiliations, we calculated the correlation between *rpoD* similarities (Table S3) and ANIb values (Table S4). We included 275 strains (217 type strains and 58 environmental strains) and obtained a Pearson correlation coefficient of 0.872 (*p*-Value < 10e-300).

3.3. Identification of Environmental Isolates

A total of 145 environmental *Pseudomonas* isolates were identified using the *rpoD* gene (taxonomic affiliations are summarized in Table S2). Briefly, *Pseudomonas* strains were isolated from banana plants (*n* = 7), rice (*n* = 33), maize (*n* = 16), or wheat (*n* = 89) and from diverse geographical locations in Belgium, Iran, and Sri Lanka [39,40]. Only two isolates clustered in the *P. resinovorans* group within the *P. aeruginosa* lineage while the majority, 143 isolates, were affiliated to the *P. fluorescens* lineage, respectively belonging to the *P. putida* (*n* = 63), *P. asplenii* (*n* = 1), and *P. fluorescens* (*n* = 79) groups (Figure 2). Within the *P. fluorescens* group, strains are distributed in five subgroups, *P. jessenii* (*n* = 3), *P. gessardii* (*n* = 4), *P. corrugata* (*n* = 22), *P. koreensis* (*n* = 23), and *P. fluorescens* (*n* = 27) (Figure 2). As previously observed with type strains of *Pseudomonas*, we identified a cutoff value of 98% nucleotide identity on the *rpoD* gene to differentiate strains at the species level. On this basis, 56 isolates were identified and affiliated to 20 known *Pseudomonas* species, with the most abundant ones being *P. asiatica* (*n* = 9), *P. simiae* (*n* = 8), and *P. alloputida* (*n* = 6), while the remaining 89 isolates were suspected to represent new species (Tables S2 and S3). These taxonomic affiliations were supported by ANIb calculations between our 58 genomes and all known *Pseudomonas* species included in this study (Table 3 and Table S4). Finally, the 89 unaffiliated isolates were confirmed to belong to 31 new *Pseudomonas* species (listed *Pseudomonas* sp. #1 to #31 in Table 3 and Table S2).

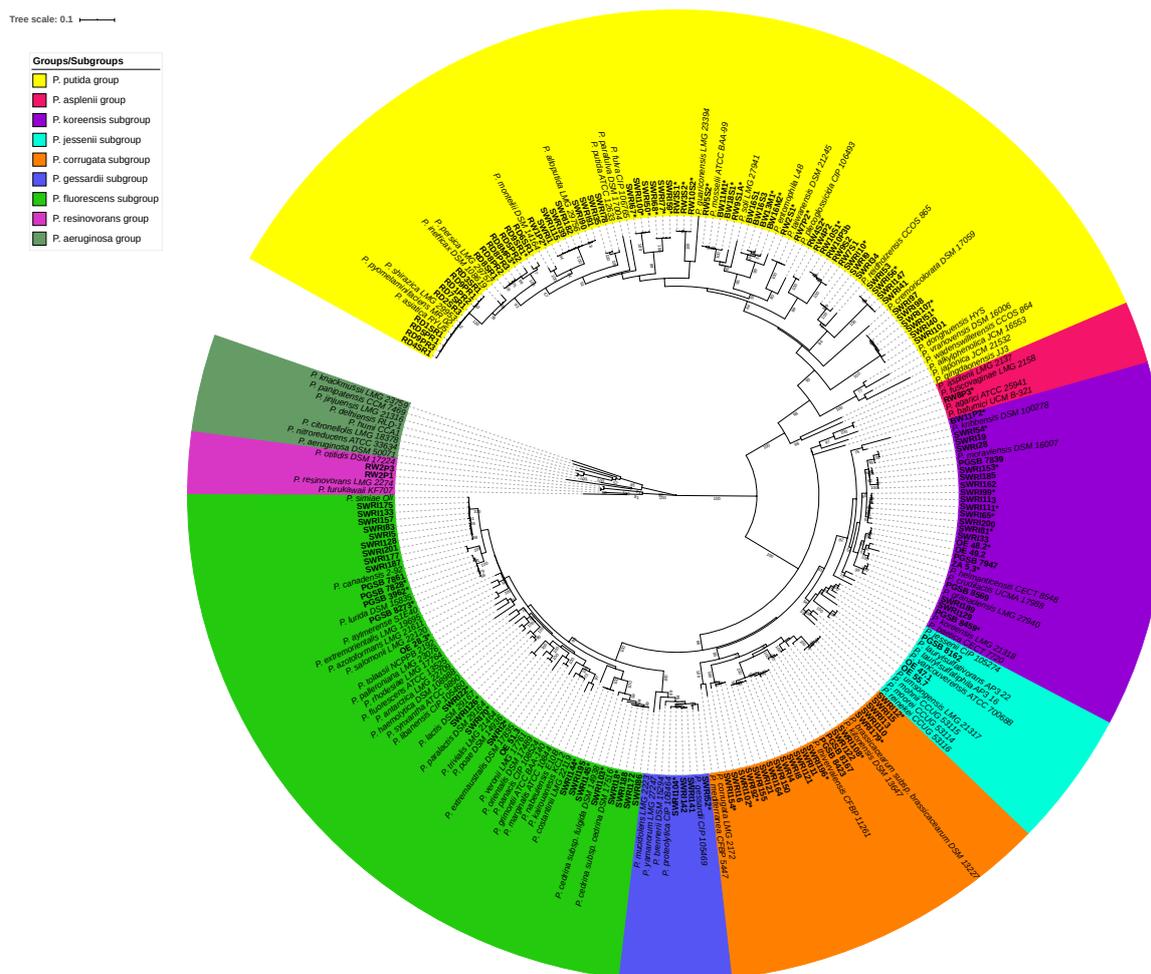


Figure 2. Maximum likelihood phylogenetic tree based on a ~650 bp fragment of the *rpoD* gene, including the 145 environmental isolates and related groups/subgroups. The tree was constructed using the GTR+G+I model (MEGA-X), and only bootstrap values higher than 50% are indicated. The *P. aeruginosa* group was used as the outgroup. Environmental isolates are highlighted in bold. * Indicates genome sequenced isolates.

The new species are distributed as follows: *P. putida* group ($n = 13$), *P. asplenii* group ($n = 1$), *P. gessardii* subgroup ($n = 1$), *P. koreensis* subgroup ($n = 7$), *P. corrugata* subgroup ($n = 4$), and *P. fluorescens* subgroup ($n = 5$). Their description and an update of the *Pseudomonas* phylogeny will be published elsewhere (Girard et al., in preparation). Overall, ~95% of taxonomic affiliations based on the *rpoD* gene that were confirmed by whole genome comparison were accurate. However, we also observed limitations (~5% of the cases) where the *rpoD* alone would have led to affiliation to known species (i.e., SWRI103, SWRI126, OE 28.3, Table 3), while based on ANIb/dDDH these strains represent new *Pseudomonas* species. These strains are also affiliated to known species when using MLSA (>97%, Table S5). Therefore, based on a total of 37,402 dual comparisons (217 type strains and 58 environmental strains), a threshold of 98% similarity led to 0.048% of misidentifications.

The 16S rRNA gene is still widely used in microbiology to identify bacteria [9]. However, the presence of multiple copies of this gene can lead to different ribotypes for a given *Pseudomonas* strain [14]. On the other hand, the *rpoD* is present in a single copy, offers a higher resolution than the 16S rRNA gene [12], and appears to be as efficient as MLSA to taxonomically assign environmental isolates (this study). Furthermore, the sequencing of four genes for MLSA or the 16S rRNA gene (i.e., sequencing of both ends to have the longest fragment) generates higher costs than *rpoD* (one end only). Consequently, the use of this locus for the identification of large batches of environmental isolates is advantageous in terms of

resolution but also from a practical and financial point of view. Moreover, the widespread distribution of the *rpoD* gene in public databases, especially through its use for MLSA, now makes it an interesting target for metagenomics studies heading to evaluate the diversity of *Pseudomonas* within environmental samples [23].

Table 3. Confirmation of *rpoD*-based taxonomic affiliations. Strains for which the similarities based on *rpoD* are not concordant with ANIb values are highlighted in bold.

Group/Subgroup	Strain	Closest-Related Strain	<i>rpoD</i>	ANIb	Taxonomic Affiliation
<i>P. putida</i> group	RD8MR3	RD9SR1	97.84	94.70	<i>Pseudomonas</i> sp. #1
	RD9SR1	RD8MR3	97.84	94.70	<i>Pseudomonas</i> sp. #2
	RW1P2	<i>P. monteilii</i>	98.00	95.77 *	<i>Pseudomonas</i> sp. #3
	SWRI67 SWRI100 SWRI50	SWRI68 SWRI77 SWRI59	93.03	<95	<i>Pseudomonas</i> sp. #4
	SWRI68 SWRI77 SWRI59	SWRI67 SWRI100 SWRI50	93.03	<95	<i>Pseudomonas</i> sp. #5
	RW3S1 RW3S2 RW10S2	BW13M1	84.07 84.07 83.92	85.36 85.32 85.44	<i>Pseudomonas</i> sp. #6
	RW5S2 BW11M1 BW18S1	<i>P. mosselii</i>	99.85 99.69 99.23	99.01 99.17 97.36	<i>P. mosselii</i>
	RW9S1A	BW18S1	92.94	89.14	<i>Pseudomonas</i> sp. #7
	BW13M1 BW16M2	BW18S1	91.06 90.91	94.81 94.62	<i>Pseudomonas</i> sp. #8
	RW2S1 RW7P2	<i>P. taiwanensis</i>	100 100	99.63 99.67	<i>P. taiwanensis</i>
	RW4S2	<i>P. plecoglossicida</i>	85.69	87.10	<i>Pseudomonas</i> sp. #9
	RW10S1	RW4S2	89.52	86.57	<i>Pseudomonas</i> sp. #10
	SWRI10	<i>P. reidholzensis</i>	92.91	86.65	<i>Pseudomonas</i> sp. #11
	SWRI56	<i>P. reidholzensis</i>	85.56	85.12	<i>Pseudomonas</i> sp. #12
	SWRI107 SWRI51	BW13M1	83.82 83.82	85.62 85.66	<i>Pseudomonas</i> sp. #13
<i>P. asplenii</i> group	RW8P3	<i>P. asplenii</i>	89.55	88.26	<i>Pseudomonas</i> sp. #14
<i>P. koreensis</i> subgroup	BW11P2	<i>P. kribbensis</i>	91.06	91.61	<i>Pseudomonas</i> sp. #15
	SWRI54	<i>P. moraviensis</i>	96.28	91.95	<i>Pseudomonas</i> sp. #16
	SWRI153	SWRI54	95.05	89	<i>Pseudomonas</i> sp. #17
	SWRI99 SWRI111 SWRI65 SWRI81	SWRI54	93.99 94.58 94.43 93.99	88.44 88.56 88.57 88.55	<i>Pseudomonas</i> sp. #18
	OE 48.2	<i>P. crudilactis</i>	96.74	91.60	<i>Pseudomonas</i> sp. #19
	ZA 5.3	<i>P. crudilactis</i>	96.89	93.47	<i>Pseudomonas</i> sp. #20
	PGSB 8459	<i>P. koreensis</i>	94.61	89.11	<i>Pseudomonas</i> sp. #21
	<i>P. corrugata</i> subgroup	SWRI12 SWRI179	<i>P. thivervalensis</i>	95.20 95.20	92.84 92.85
SWRI108		<i>P. kilonensis</i>	95.41	95.73*	<i>Pseudomonas</i> sp. #23
SWRI196		<i>P. kilonensis</i>	92.15	88.02	<i>Pseudomonas</i> sp. #24
SWRI92 SWRI102 SWRI154		<i>P. kilonensis</i>	91.69 91.69 90.92	87.92 87.89 87.66	<i>Pseudomonas</i> sp. #25

Table 3. Cont.

Group/Subgroup	Strain	Closest-Related Strain	<i>rpoD</i>	ANIb	Taxonomic Affiliation
<i>P. gessardii</i> subgroup	SWRI52	<i>P. gessardii</i>	96.92	92.50	<i>Pseudomonas</i> sp. #26
	SWRI104	<i>P. proteolytica</i>	99.38	98.35	<i>P. proteolytica</i>
<i>P. fluorescens</i> subgroup	SWRI18	<i>P. cedrina</i> subsp. <i>cedrina</i>	98.92	96.94	<i>P. cedrina</i> subsp. <i>cedrina</i>
	SWRI103	<i>P. cedrina</i> subsp. <i>cedrina</i>	98.61	94.19	<i>Pseudomonas</i> sp. #27
	SWRI145	<i>P. cedrina</i> subsp. <i>cedrina</i>	93.07	88.51	<i>Pseudomonas</i> sp. #28
	SWRI144		93.07	88.46	
	SWRI2	<i>P. poae</i>	99.54	98.85	<i>P. poae</i>
	SWRI70	<i>P. paralactis</i>	99.23	98.54	<i>P. paralactis</i>
	SWRI126	<i>P. lactis</i>	98.92	95.09*	<i>Pseudomonas</i> sp. #29
	SWRI22	<i>P. lactis</i>	97.99	94.45	<i>Pseudomonas</i> sp. #30
	OE 28.3	<i>P. salomonii</i>	98.31	94.10	<i>Pseudomonas</i> sp. #31
	PGSB 8273	<i>P. lurida</i>	99.85	99.19	<i>P. lurida</i>
PGSB 3962		99.69	99.26		
PGSB 7828		99.54	99.27		

* dDDH < 70%.

4. Conclusions

In this study, we used a ~650 bp *rpoD* locus to describe the phylogenetic structure of the genus *Pseudomonas*. We provide a genus-wide overview of the taxonomy and an integrative study making the link between single locus and whole genome analysis. The use of the *rpoD* allowed us to accurately assign strains to specific phylogenetic groups or subgroups and affiliate strains to known *Pseudomonas* species. Furthermore, it has proven to be efficient as a first diagnostic tool to identify 31 new species. The high correlation between *rpoD* similarities and ANIb values emphasize the high discriminative power of this gene. Importantly, we show that the *rpoD* is a powerful tool for microbiologists, superior to the 16S rRNA gene, for accurate identification of *Pseudomonas* isolates.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-2607/8/8/1166/s1>. Table S1: Species type strains used in this study. Table S2: List of environmental strains used in this study. Origin, NCBI accession number for the *rpoD* gene and/or genome, and taxonomic affiliations are specified. Table S3: *rpoD* similarity matrix, including all type and environmental strains. Table S4: ANIb similarity matrix, including all type and strains for which genomes were available. Table S5: MLSA similarity matrix only, including type and environmental strains included in the limitations of *rpoD*-based taxonomic affiliations.

Author Contributions: Conceptualization, methodology, investigation, writing—original draft and editing, L.G.; methodology, investigation, and writing—original draft, C.L.; resources and formal analysis, H.R.-Z.; writing—review & editing, R.L., V.v.N.; writing—review and editing and funding acquisition, R.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by the EOS grant 30650620. C.L. is supported by an SB PhD fellowship from FWO Vlaanderen (1S64720N).

Acknowledgments: We thank Marleen Voet and Alison Kerremans for their help with processing the samples for Illumina sequencing.

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

References

- Bollinger, A.; Thies, S.; Katzke, N.; Jaeger, K. The biotechnological potential of marine bacteria in the novel lineage of *Pseudomonas pertucinogena*. *Microb. Biotechnol.* **2020**, *13*, 19–31. [[CrossRef](#)] [[PubMed](#)]
- Silby, M.W.; Winstanley, C.; Godfrey, S.A.C.; Levy, S.B.; Jackson, R.W. *Pseudomonas* genomes: Diverse and adaptable. *FEMS Microbiol. Rev.* **2011**, *35*, 652–680. [[CrossRef](#)] [[PubMed](#)]
- Höfte, M.; De Vos, P. Plant pathogenic *Pseudomonas* species. In *Plant-Associated Bacteria*; Gnanamanickam, S.S., Ed.; Springer: Dordrecht, The Netherlands, 2007; pp. 507–533.

4. Wiklund, T. *Pseudomonas anguilliseptica* infection as a threat to wild and farmed fish in the Baltic Sea. *Microbiol. Aust.* **2016**, *37*, 135. [[CrossRef](#)]
5. Beaton, A.; Lood, C.; Cunningham-Oakes, E.; MacFadyen, A.; Mullins, A.J.; Bestawy, W.E.; Botelho, J.; Chevalier, S.; Coleman, S.; Dalzell, C.; et al. Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies. *FEMS Microbiol. Lett.* **2018**, *365*, fny069. [[CrossRef](#)] [[PubMed](#)]
6. Wasi, S.; Tabrez, S.; Ahmad, M. Use of *Pseudomonas* spp. for the bioremediation of environmental pollutants: A review. *Environ. Monit. Assess.* **2013**, *185*, 8147–8155. [[CrossRef](#)] [[PubMed](#)]
7. Weller, D.M. *Pseudomonas* biocontrol agents of soilborne pathogens: Looking back over 30 years. *Phytopathology* **2007**, *97*, 250–256. [[CrossRef](#)]
8. Gross, H.; Loper, J.E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep.* **2009**, *26*, 1408–1446. [[CrossRef](#)]
9. Mende, D.R.; Sunagawa, S.; Zeller, G.; Bork, P. Accurate and universal delineation of prokaryotic species. *Nat. Methods* **2013**, *10*, 881–884. [[CrossRef](#)]
10. Moore, E.R.B.; Mau, M.; Arnscheidt, A.; Böttger, E.C.; Hutson, R.A.; Collins, M.D.; Van De Peer, Y.; De Wachter, R.; Timmis, K.N. The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. *Syst. Appl. Microbiol.* **1996**, *19*, 478–492. [[CrossRef](#)]
11. Brosch, R.; Lefèvre, M.; Grimont, F.; Grimont, P.A.D. Taxonomic diversity of Pseudomonads revealed by computer-interpretation of ribotyping data. *Syst. Appl. Microbiol.* **1996**, *19*, 541–555. [[CrossRef](#)]
12. Lalucat, J.; Mulet, M.; Gomila, M.; García-Valdés, E. Genomics in bacterial taxonomy: Impact on the Genus *Pseudomonas*. *Genes* **2020**, *11*, 139. [[CrossRef](#)] [[PubMed](#)]
13. 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. [[CrossRef](#)] [[PubMed](#)]
14. Bodilis, J.; Nsique-Meilo, S.; Besaury, L.; Quillet, L. Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in *Pseudomonas*. *PLoS ONE* **2012**, *7*, e35647. [[CrossRef](#)] [[PubMed](#)]
15. Mulet, M.; Lalucat, J.; García-Valdés, E. DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol.* **2010**, *12*, 1513–1530. [[PubMed](#)]
16. Gomila, M.; Peña, A.; Mulet, M.; Lalucat, J.; García-Valdés, E. Phylogenomics and systematics in *Pseudomonas*. *Front. Microbiol.* **2015**, *6*, 214. [[CrossRef](#)] [[PubMed](#)]
17. Lopes, L.D.; Davis, E.W.; Pereira e Silva, M.d.C.; Weisberg, A.J.; Bresciani, L.; Chang, J.H.; Loper, J.E.; Andreote, F.D. Tropical soils are a reservoir for fluorescent *Pseudomonas* spp. biodiversity. *Environ. Microbiol.* **2018**, *20*, 62–74. [[CrossRef](#)]
18. Biessy, A.; Novinscak, A.; Blom, J.; Léger, G.; Thomashow, L.S.; Cazorla, F.M.; Josic, D.; Filion, M. Diversity of phytobeneficial traits revealed by whole-genome analysis of worldwide-isolated phenazine-producing *Pseudomonas* spp. *Environ. Microbiol.* **2019**, *21*, 437–455. [[CrossRef](#)]
19. Oni, F.E.; Geudens, N.; Omoboye, O.O.; Bertier, L.; Hua, H.G.K.; Adiobo, A.; Sinnaeve, D.; Martins, J.C.; Höfte, M. Fluorescent *Pseudomonas* and cyclic lipopeptide diversity in the rhizosphere of cocoyam (*Xanthosoma sagittifolium*). *Environ. Microbiol.* **2019**, *21*, 1019–1034. [[CrossRef](#)]
20. Keshavarz-Tohid, V.; Vacheron, J.; Dubost, A.; Prigent-Combaret, C.; Taheri, P.; Tarighi, S.; Taghavi, S.M.; Moëne-Loccoz, Y.; Muller, D. Genomic, phylogenetic and catabolic re-assessment of the *Pseudomonas putida* clade supports the delineation of *Pseudomonas alloputida* sp. nov., *Pseudomonas inefficax* sp. nov., *Pseudomonas persica* sp. nov., and *Pseudomonas shirazica* sp. nov. *Syst. Appl. Microbiol.* **2019**, *42*, 468–480. [[CrossRef](#)]
21. Camiade, M.; Bodilis, J.; Chaftar, N.; Riah-Anglet, W.; Gardères, J.; Buquet, S.; Ribeiro, A.F.; Pawlak, B. Antibiotic resistance patterns of *Pseudomonas* spp. isolated from faecal wastes in the environment and contaminated surface water. *FEMS Microbiol. Ecol.* **2020**, *96*, fiae008. [[CrossRef](#)]
22. Tran, P.N.; Savka, M.A.; Gan, H.M. In-silico taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus *Pseudomonas*. *Front. Microbiol.* **2017**, *8*, 1296. [[CrossRef](#)] [[PubMed](#)]
23. Mulet, M.; Bennasar, A.; Lalucat, J.; García-Valdés, E. An rpoD-based PCR procedure for the identification of *Pseudomonas* species and for their detection in environmental samples. *Mol. Cell. Probe.* **2009**, *23*, 140–147. [[CrossRef](#)] [[PubMed](#)]

24. Mulet, M.; David, Z.; Nogales, B.; Bosch, R.; Lalucat, J.; García-Valdés, E. *Pseudomonas* diversity in crude-oil-contaminated intertidal sand samples obtained after the prestige oil spill. *Appl. Environ. Microbiol.* **2011**, *77*, 1076–1085. [[CrossRef](#)] [[PubMed](#)]
25. Parkinson, N.; Bryant, R.; Bew, J.; Elphinstone, J. Rapid phylogenetic identification of members of the *Pseudomonas syringae* species complex using the *rpoD* locus. *Plant Pathol.* **2011**, *60*, 338–344. [[CrossRef](#)]
26. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]
27. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
28. Ghequire, M.G.K.; Swings, T.; Michiels, J.; Gross, H.; De Mot, R. Draft genome sequence of *Pseudomonas putida* BW11M1, a banana rhizosphere isolate with a diversified antimicrobial armamentarium. *Genome Announc.* **2016**, *4*. [[CrossRef](#)]
29. Nguyen, D.D.; Melnik, A.V.; Koyama, N.; Lu, X.; Schorn, M.; Fang, J.; Aguinaldo, K.; Lincecum, T.L.; Ghequire, M.G.K.; Carrion, V.J.; et al. Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat. Microbiol.* **2016**, *2*, 16197. [[CrossRef](#)]
30. Zarvandi, S.; Bahrami, T.; Pauwels, B.; Asgharzadeh, A.; Hosseini-Mazinani, M.; Salari, F.; Girard, L.; De Mot, R.; Rokni-Zadeh, H. Draft genome sequence of cyclic lipopeptide producer *Pseudomonas* sp. SWRI103 isolated from wheat rhizosphere. *Microbiol. Resour. Announc.* **2020**, *9*. [[CrossRef](#)]
31. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
32. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
33. Tohya, M.; Watanabe, S.; Tada, T.; Tin, H.H.; Kirikae, T. Genome analysis-based reclassification of *Pseudomonas fuscovaginae* and *Pseudomonas shirazica* as later heterotypic synonyms of *Pseudomonas asplenii* and *Pseudomonas asiatica*, respectively. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 3547–3552. [[CrossRef](#)] [[PubMed](#)]
34. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.-P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. [[CrossRef](#)] [[PubMed](#)]
35. Meier-Kolthoff, J.P.; Klenk, H.-P.; Göker, M. Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 352–356. [[CrossRef](#)] [[PubMed](#)]
36. Auch, A.F.; von Jan, M.; Klenk, H.-P.; Göker, M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* **2010**, *2*, 117–134. [[CrossRef](#)]
37. Hesse, C.; Schulz, F.; Bull, C.T.; Shaffer, B.T.; Yan, Q.; Shapiro, N.; Hassan, K.A.; Varghese, N.; Elbourne, L.D.H.; Paulsen, I.T.; et al. Genome-based evolutionary history of *Pseudomonas* spp. *Environ. Microbiol.* **2018**, *20*, 2142–2159. [[CrossRef](#)]
38. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
39. Vlassak, K.; Van Holm, L.; Duchateau, L.; Vanderleyden, J.; De Mot, R. Isolation and characterization of fluorescent *Pseudomonas* associated with the roots of rice and banana grown in Sri Lanka. *Plant Soil* **1992**, *145*, 51–63. [[CrossRef](#)]
40. Rasouli Sadaghiani, M.H.; Khavazi, K.; Rahimian, H.; Malakouti, M.J.; Asadi Rahmani, H. An evaluation of the potentials of indigenous fluorescent *Pseudomonads* of wheat rhizosphere for producing siderophore. *Iran. J. Soil Water Sci.* **2006**, *20*, 133–143.

