An rpoD gene sequence based evaluation of cultured Pseudomonas diversity on different growth media

Jonas Ghyselinck,1 An Coorevits,1,2 Anita Van Landschoot,2 Emly Samyn,1 Kim Heylen1 and Paul De Vos1,3

1Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, K.L. Ledeganckstraat 35, Gent B-9000, Belgium
2Faculty of Bioscience Engineering, Ghent University, Campus Schoonmeersen, Valentin Vaerwyckweg 1, Gent B-9000, Belgium
3BCCM/LMG Bacteria Collection, K.L. Ledeganckstraat 35, Gent B-9000, Belgium

The last decade has shown an increased interest in the utilization of bacteria for applications ranging from bioremediation to wastewater purification and promotion of plant growth. In order to extend the current number of micro-organism mediated applications, a continued quest for new agents is required. This study focused on the genus Pseudomonas, which is known to harbour strains with a very diverse set of interesting properties. The aim was to identify growth media that allow retrieval of a high Pseudomonas diversity, as such increasing the chance of isolating isolates with beneficial properties. Three cultivation media: trypticase soy agar (TSA), potato dextrose agar (PDA) and Pseudomonas isolation agar (PIA) were evaluated for their abilities to grow Pseudomonas strains. TSA and PDA were found to generate the largest Pseudomonas diversity. However, communities obtained with both media overlapped. Communities obtained with PIA, on the other hand, were unique. This indicated that the largest diversity is obtained by sampling from either PDA or TSA and from PIA in parallel. To evaluate biodiversity of the isolated Pseudomonas members on the media, an appropriate biomarker had to be identified. Hence, an introductory investigation of the taxonomic resolution of the 16S rRNA, rpoD, gyrB and rpoB genes was performed. The rpoD gene sequences not only had a high phylogenetic content and the highest taxonomic resolution amongst the genes investigated, it also had a gene phylogeny that related well with that of the 16S rRNA gene.

INTRODUCTION

The genus Pseudomonas historically developed as a kind of dumping ground for aerobic, motile Gram-negative rods, and thus lacked a profound classification of its members (Coorevits, 2011). Therefore, numerous efforts have been made to reclassify the genus. Members, which were originally distributed over the Alpha-, Beta-, and Gammaproteobacteria, were driven back to the Gammaproteobacteria. The genus Pseudomonas continuously harbours a collection of bacterial strains with very diverse characteristics. Numerous positive traits have been attributed to Pseudomonas strains, ranging from denitrification (Zhang et al., 2011) (e.g. in wastewater treatments) to the degradation of toxic components (Farhadian et al., 2008; Khan et al., 2009) and the promotion of plant growth (Chin-A-Woeng et al., 2003; Daayf et al., 2003; De Curtis et al., 2010; Ghyselinck et al., 2013; Haas & Défago, 2005; Kim & Jeun, 2006; Scherwinski et al., 2008; Yan et al., 2002). However, the genus Pseudomonas also has a negative reputation, as it includes opportunistic human pathogens (Gershman et al., 2008) and plant pathogens (Mansfield et al., 2012).

The genus Pseudomonas is composed of ten phylogenetic groups, each group being a collection of closely related species (Mulet et al., 2010). In addition, one of these groups, the Pseudomonas fluorescens group, consists of nine subgroups (Mulet et al., 2010). The taxonomic resolution...
of the 16S rRNA gene fails to differentiate *Pseudomonas* strains at the intrageneric level (Yamamoto et al., 2000); i.e. the gene does not allow species to be distinguished within a *Pseudomonas* group or subgroup. Hence, the 16S rRNA gene is not suited to adequately measure the diversity of *Pseudomonas* members. The MLSA scheme introduced by Mulet et al. (2010) allows a more in-depth identification of *Pseudomonas* strains. The scheme is based on the concatenated sequence of four genes, namely the 16S rRNA, *gyrB*, *rpoD* and *rpoB* genes.

To further advance the discovery of *Pseudomonas* isolates with beneficial properties, three different cultivation media were evaluated for their capacities to grow members of the genus *Pseudomonas*. We assumed that chances to obtain *Pseudomonas* isolates with beneficial properties increase with increased cultured *Pseudomonas* diversity. Hence, the medium that showed the largest diversity of *Pseudomonas* isolates was considered the most optimal medium. However, due to the complex taxonomy of the genus *Pseudomonas*, there is no straightforward protocol available for differentiating *Pseudomonas* isolates at deep taxonomic levels. Hence, it was unclear which taxonomic marker was best suited for the purpose of our study. To cope with this problem and select an appropriate biomarker, the taxonomic resolution of each of the four genes that were proposed in the identification scheme of Mulet et al. (2010) was compared. Results illustrated that the *rpoD* gene was preferred.

**METHODS**

**Sampling.** Potato rhizosphere and root samples were taken from three fields (E1, 02° 37’ 20.4” S 078° 56’ 04.7” W; E2, 079° 09’ 25.4” W 03° 20’ 15.9” S; and E3, 079° 13’ 32.6” W 03° 32’ 21.8” S) in the Central Andean Highlands of Ecuador. Ten plants were sampled per field. Per plant, 5 g of rhizosphere soil adhering to the potato roots was collected by brushing the roots. Rhizosphere soil samples obtained from each of the sampled plants were pooled per field. Five ml PBS (8 g l⁻¹ NaCl, 0.2 g l⁻¹ KCl, 1.44 g l⁻¹ NaHPO₄, and 0.24 g l⁻¹ KH₂PO₄ in water, adjusted to pH 7.4 by adding HCl) and 10 sterile glass beads (6 mm) were then added to 1 g pooled rhizosphere soil, and the obtained suspension was vortexed for 2 min.

Root fragments of each plant sampled were surface sterilized with 5 % sodium hypochlorite for 2 min, then rinsed in autoclaved water and cut into small pieces and pooled per field. Fifteen grams of pooled root fragments were then triturated in 50 ml PBS by using a sterile glass rod, and the resulting mixture was incubated for one hour at 28 °C with agitation (150 r.p.m.).

Serial dilutions were made (10⁻⁰–10⁻²) of the root and rhizosphere soil suspensions and 10⁻¹ and 10⁻² dilutions were plated (100 μl) on ten-fold diluted trypticase soy agar (TSA) (15 g l⁻¹ pancreatic digest of casein, 5 g l⁻¹ enzymic digest of soya bean, 3 g l⁻¹ sodium chloride and 15 g l⁻¹ agar), potato dextrose agar (PDA) (4 g l⁻¹ potato extract, 20 g l⁻¹ glucose, 15 g l⁻¹ agar) and *Pseudomonas* isolation agar (PIA; DifcoTM, BD) (20 g l⁻¹ peptone, 1.4 g l⁻¹ MgCl₂, 10 g l⁻¹ K₂SO₄, 25 mg l⁻¹ Irgasan® MD, 13.6 g l⁻¹ agar). TSA and PDA were supplemented with 0.005 % (w/v) cycloheximide to inhibit fungal growth. After 48 h of incubation at 28 °C, the bulk cultivable fraction was harvested per plate, and collected in Eppendorf tubes for DNA extraction.

**DNA extraction, PCR and clone libraries.** Two sample types (roots and rhizosphere soil), sampled in three fields (E1–E3), all plated in two dilutions on three different growth media makes 36 samples (cultivable fractions) from which DNA was extracted (Fig. S1, available in Microbiology Online). DNA was extracted according to Pitcher et al. (1989). *rpoD* gene amplification PCR was performed in triplicate on each DNA extract according to Mulet et al. (2009). PCR amplicons were purified with the NucleoFast® 96 PCR system (Millipore). The triplicate PCR products of the 10⁻³ and 10⁻² dilutions of a given sample (e.g. E1 rhizosphere TSA) were then pooled in equimolar amounts. Hence, 18 samples remained (Table S1) in which the *Pseudomonas* diversity was studied using clone libraries that were composed with the pGEM-T Vector System (Promega Benelux). Sequencing of clones was performed on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). Sequencing products were purified with a BigDye XTerminator® Purification kit (Applied Biosystems) and sequenced using a 3130xl Genetic Analyzer (Applied Biosystems). Sequence quality checking and trimming of primer sequences was performed manually in BioNumerics 5.1 (Applied Maths). Chimera detection was performed with Chimera Slayer (Haas et al., 2011). *rpoD* gene sequences were placed in the correct open reading frames by nucleotide to amino acid conversion, and subsequently checked by amino acid to protein translation using TransSeq (http://www.ebi.ac.uk/Tools/emboss/transeq/) and BLAST (Altschul et al., 1990), respectively.

**Selection of the *rpoD* gene to differentiate *Pseudomonas* isolates.** TaxonGap analysis (Slabbinck et al., 2008) was performed on type strains representing the different subgroups of the *Pseudomonas fluorescens* group. TaxonGap allows visual comparison of the discriminatory power of different biomarkers to differentiate a set of operational taxonomic units (OTUs), but also for their abilities to differentiate members constituting an OTU. OTUs in this study were the different subgroups of the *P. fluorescens* group. Hence, the analysis was performed to identify the gene that best differentiates strains within and between the different subgroups of the *P. fluorescens* group. The biomarkers used in this study were the 16S rRNA, *gyrB*, *rpoD* and *rpoB* genes (Mulet et al., 2010). All type strains, and corresponding 16S rRNA, *rpoB*, *gyrB* and *rpoD* genes that were included for the analysis are shown in Table S2. Species subgroups to which the species were assigned are also indicated in Table S2. Sequences were obtained from GenBank and the PseudoMLSA database (Mulet et al., 2010) (http://www.uib.es/microbiologiaBD/Welcome.php). The 16S rRNA gene sequences were aligned using the ARB software (Ludwig et al., 2004) with the integrated ARB aligner based on the secondary structures of the 16S rRNA gene. Aligned sequences were exported applying the position variability filter for bacteria (integrated in the software) and reimported in the Molecular Evolutionary Genetics Analysis (MEGA5) software (Tamura et al., 2011). Overhangs were trimmed resulting in a final alignment of 1337 positions. A maximum-likelihood (ML) tree was constructed applying the Jukes–Cantor substitution model, and bootstrap analysis was performed based on 1000 replications.

The three other genes were aligned based on aa sequences using the MEGA5 software. After reconversion into the original nucleotide sequences, overhangs were trimmed resulting in final alignments of 915 positions for *rpoB* gene sequences, 717 positions for *rpoD* gene sequences and 798 positions for *gyrB* gene sequences. ML trees were constructed applying the Jukes–Cantor substitution model (with complete deletion of gaps/missing data), and bootstrap analysis was performed based on 1000 replications. The obtained pairwise similarity matrices were used for TaxonGap analysis. Alignments of the four gene sequences were concatenated (3767 positions) using the

---

J. Ghyselinck and others

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11
On: Sat, 10 Nov 2018 07:41:12
Phylogeny of the rpoD gene

Construction of phylogenetic trees. An rpoD sequence library was constructed that contained all currently available rpoD sequences from Pseudomonas type strains. Sequences were obtained through query in Straininfo (Dawyndt et al., 2005) and the PseudoMLSA database (Mulet et al., 2010) (http://www.ubi.es/microbiologiaBD/Welcome.php) (Table S3). For sequence quality checking, nucleotide sequences were translated to aa sequences using Transeq (http://www.cbi.ac.uk/Tools/emboss/translate/). The functionality of aa sequences was confirmed with the BLAST tool of NCBI (Altschul et al., 1990). Sequence alignment was performed on the aa sequences using the MEGAS software (Tamura et al., 2011). After alignment, aa sequences were reconverted into the original nucleotide sequences. 16S rRNA gene sequences of the same Pseudomonas type strains were collected into a library. The 16S rRNA gene sequences were aligned using the ARB software (Ludwig et al., 2004) as mentioned above.

Sequences in both libraries were trimmed to obtain maximum overlap between the sequences. ML trees were then constructed from both the 16S rRNA and rpoD gene sequence libraries. The software used was RAxML v7.3.5. An ML search was performed under gamma in combination with rapid bootstrapping under CAT (Stamatakis, 2006). The substitution model used was General Time Reversible (GTR). Bootstrapping was performed with 1000 replicates. The command line used for the tree search was the following:raxmpHPC-PHTREADS-SSE3 -T<number of processors>-fT-mGTRGAMMA -N<replicates>-x<seed1>-P<seed2>-s<filename>-o<outputfile>.

Branch length based comparison of phylogenetic trees. To study the phylogenetic relation between two ML trees, the Pearson correlation (PC) was calculated between patristic distances between corresponding sequence pairs in the two trees. Patristic distances are defined as the length of the shortest path connecting two taxa in a phylogenetic tree. These patristic distances were extracted from the ML trees using a script that was kindly provided by Jeraldto et al. (2011).

A second method used to study the relation between a pair of phylogenetic trees was based on the vCEED approach that was developed by Choi & Gomez (2009). Distance matrices were generated from the ML trees using the PHYLOGOL software (Webb et al., 2008). These matrices were used as inputs for the vCEED script that was written in Matlab. The vCEED script maps taxa to a Euclidean space via metric multidimensional scaling (MDS), thus producing a multidimensional plot in which each point represents one sequence (or taxon) within the phylogenetic tree. This procedure was applied to both trees to be compared. Both embedded point patterns were then superimposed on one another and the degree of fit, which is expressed by the weighted root mean square deviation (wRMSD), was calculated. A low wRMSD indicated a high degree of fit, and thus a high similarity between trees.

Construction of correlation plots. To plot the correlation between 16S rRNA gene based ML trees and rpoD based ML trees graphically, corresponding patristic distances were transferred into a tuple, which formed the coordinate of a point in a plot. Distances were ordered for the 16S rRNA gene tree, and corresponding distances of the rpoD tree were rearranged accordingly. Subsequently, a binning step was performed by calculating the mean and standard deviations of corresponding patristic distances in both trees over each patristic distance interval of 0.001 in the 16S rRNA gene tree. Mean 16S rRNA gene sequence distances and corresponding mean rpoD sequence distances were then plotted in a graph, and the standard deviations on the mean rpoD distances were superimposed as error bars.

Robinson–Foulds distance calculations. The unweighted and weighted Robinson–Foulds (RF and WRF, respectively) distances (Robinson & Foulds, 1981) were calculated to gain insight in the topological differences between two phylogenetic trees. RF and WRF distances were calculated by importing the ML trees into RAxML v7.4.2Gui. The RF metric calculates the number of splits that are unique to one of both trees being compared, so it actually describes ancestral differences between trees. The higher the RF value, the lower the amount of shared ancestors. As such, phylogenetic trees are more similar as the RF values decrease. The WRF, however, takes into account the support values of the branches that are unique to one of the trees being compared instead of just counting the number of unique splits. Comparing RF and WRF distance values allows insight into the nature of differences between trees. If the WRF value approximates the RF value for a given tree comparison, this means that differences mainly occur on high supported branches. Conversely, if the WRF value is much smaller than the corresponding RF values, differences between the trees mainly occur on low supported branches.

Evaluation of Pseudomonas diversity on different media

Construction of rarefaction curves and taxonomic assignment based on rpoD sequences. The Pseudomonas diversity on the different media was evaluated by means of rarefaction curves that were calculated with the Mothur v1.27.0 software (Schloss et al., 2009). Sequences were assigned to OTUs using the cluster command and the average neighbour algorithm. The mean OTU numbers that were retrieved from 1000 iterations were used to construct rarefaction curves. For taxonomic assignment of sequences, the Bayesian classifier that is integrated in the Mothur v1.27.0 software was used. The same rpoD gene sequence library that was used for studying rpoD phylogeny (but containing one more sequence for which no 16S rRNA gene sequence was available Table S3) served as a reference database for the taxonomic assignment of environmental rpoD sequences. Assignments were performed with the classify.seqs() command on aligned rpoD sequences. The bootstrap cut-off for assigning a sequence to a specific taxon was set at 50 %.

Fast UniFrac analysis A total of 1500 rpoD sequences that were obtained from all samples and from the different media investigated were merged into one FASTA file. Nucleotide sequences were converted to aa sequences and the alignment was performed in MEGAS5 (Tamura et al., 2011) using the CLUSTAL W alignment tool. The alignment was checked and adjusted manually. After reconcatenating the aligned aa sequences back to the original nucleotide sequences, the end points of all sequences were trimmed to obtain maximum overlap between the sequences. An ML tree was constructed from the aligned sequences in RAxML v7.3.5 with the same parameters as mentioned above but with 500 bootstraps instead of 1000. The ML tree was then imported in the Fast UniFrac webtool that is available online (Hamady et al., 2010). Sample clustering was performed, and sample distance matrices were calculated.

RESULTS

Taxonomic resolution of the rpoD gene

In the present study, the TaxonGap software (Slabbinck et al., 2008) was used to evaluate each of the four biomarker genes for their power to differentiate strains belonging to the Pseudomonas fluorescens group on the one
hand, and for their ability to represent the phylogeny of members of the *P. fluorescens* group on the other hand. Fig. 1 shows that, within the *P. fluorescens* group species subgroup separability, in other words gene resolution (represented by dark grey bars), was highest for the *rpoD* gene followed by the *gyrB*, *rpoB* and 16S rRNA genes. Although species subgroup heterogeneity (indicated by the light grey bars) was generally large, thus indicating a high within *P. fluorescens* subgroup resolution, it mostly did not exceed species subgroup separability for the *rpoD* gene (Fig. 1). This finding indicates a high within *P. fluorescens* subgroup resolution that does not hamper differentiation of the different subgroups. Additionally, a good correlation was observed between *rpoD* gene based phylogeny and phylogeny based on all four genes of the MLSA scheme of Mulet et al. (2010) for members of the *Pseudomonas fluorescens* group. These findings supported the use of the *rpoD* biomarker for measuring the *Pseudomonas* diversity on the growth media. Hence, further studies were performed on the *rpoD* gene only.

**Phylogenetic content of *Pseudomonas rpoD* sequences**

**Patristic distance based comparison of phylogenetic trees.** The phylogenetic content refers to the amount of phylogenetic information contained within sequences. Sequences with a high phylogenetic content contain sufficient information to build robust phylogenetic trees. In order to study the phylogenetic content of *Pseudomonas rpoD* sequences, we calculated the PC between corresponding patristic distances in two ML trees that were constructed from independent tree searches on the same *rpoD* sequence library (Jeraldo et al., 2011). The sequence library contained all currently available *rpoD* sequences of *Pseudomonas* type strains and was not restricted to members of the *P. fluorescens* group. A high PC indicates a positive association between patristic distances in both trees being compared, implying that taxa positioned closely together in the first *rpoD* tree are also positioned closely together in the second *rpoD* tree. Conversely, a low PC indicates a very low association between corresponding patristic distances in both trees, which in turn means that pairs of taxa positioned closely together in the first *rpoD* tree are not necessarily positioned closely together in the second *rpoD* tree. A high PC thus indicates that the phylogenetic information within the sequences being studied is sufficient to calculate unequivocal patristic distances between taxa, which in turn points to a high phylogenetic content of the sequences.

Because ML trees are built using a heuristic method, there is no guarantee that the tree calculated is the best representation of the sequence data. As such, a high PC between a pair of *rpoD* trees generated from different tree searches on the same library may have been obtained by chance. To deal with this problem, we calculated four trees from the *rpoD* sequence data, and subsequently calculated the PCs between all possible pairs of trees (Table 1). The obtained PCs remained high for all tree comparisons. This indicated that *rpoD* based ML trees were very robust with respect to patristic distances between taxa, illustrating that *Pseudomonas rpoD* sequences contain enough phylogenetic information to construct phylogenetic trees unequivocally. However, the PC method to compare phylogenetic trees is known to have some weaknesses (Choi & Gomez, 2009),

![Fig. 1. The taxonomic resolution of four housekeeping genes and the concatenated sequence (Conc.). The different *Pseudomonas fluorescens* subgroups are represented on the left side of the graphic in a maximum-likelihood tree that was constructed with the MEGA5 software. Heterogeneity within subgroups is represented by grey bars; subgroup separability is represented by dark bars. The most closely related subgroup is written next to each bar. The scaled axes express the magnitude of the separability/heterogeneity. The vertical black line indicates the smallest separability recorded. The strain collection on which this study was conducted is given in Table S2. The trees were generated by TaxonGap2.4.1.](image-url)
which is why an additional study was performed by using the vCEED approach (Choi & Gomez, 2009). Results are given in Table 1 and are expressed in terms of degree of fit (wRMSD) between two trees. A very high correlation was obtained between results generated with the vCEED approach and results obtained with the PC method ($R^2 = 0.99$).

The PCs between trees obtained from different tree searches on a given sequence library, can be used as a measure to compare the phylogenetic content of different genes. It was striking that the PC between trees obtained from *Pseudomonas rpoD* sequences was systematically higher than the PC between trees constructed from 16S rRNA gene sequences. As can be observed from the superimposition plots given in Fig. 2, the vCEED method similarly showed that the degree of fit was higher for *rpoD* based trees than for 16S rRNA gene based trees. This implies that *rpoD* sequences allow the construction of phylogenetic trees that are more robust than trees built from 16S rRNA gene sequences. Therefore, these findings suggest that there is more phylogenetic information contained within *Pseudomonas rpoD* sequences than within 16S rRNA gene sequences.

**Topology based comparison of phylogenetic trees.** RF and WRF distances provide information on topological differences between trees, and the nature of those differences. Table 1 shows the RF and WRF values for the different tree comparisons. We observed that the RF and WRF distances were significantly smaller between trees generated from different tree searches on *rpoD* sequences than between trees generated from different tree searches on 16S rRNA sequences. Furthermore, differences between RF and WRF values for a given tree comparison were generally larger between trees generated from 16S rRNA gene sequences than between *rpoD* gene based trees. This indicates that differences in tree topology in 16S rRNA gene based trees mainly occur on branches with low bootstrap support values, while differences between *rpoD* gene based trees occur on branches with higher supporting values. In other words, topological differences between 16S rRNA gene based trees are mostly caused by inadequacies of sequences to validate the topology, suggesting that the differences are due to the lower phylogenetic content of 16S rRNA sequences to construct unequivocal tree topologies. As mentioned above, RF distances between trees obtained from different tree searches on the *rpoD* library were smaller, and the branches leading to differences were better supported. This again shows that the phylogenetic content of *rpoD* sequences is higher compared to that of the 16S rRNA gene sequences. Relative
to the mean RF distances between trees generated from different tree searches on 16S rRNA gene sequences, the mean RF distance between rpoD and 16S rRNA gene trees was only 1.6 times higher. This indicates that the differences between topologies in rpoD and 16S rRNA gene based trees did not significantly exceed topological differences between different trees that were generated from a given 16S rRNA gene sequence library.

Is rpoD based phylogeny in contradiction to 16S rRNA gene based phylogeny?

The phylogenetic information that is contained within rpoD sequences was found to be higher than the phylogenetic information contained within 16S rRNA gene sequences. However, the question still remains whether or not rpoD phylogeny is contradictory to 16S rRNA gene based phylogeny. Although microbiologists have deviated for a long time from the assumption that the 16S rRNA gene reflects true phylogenetic relationships between organisms, it is still regarded as the benchmark for reconstructing phylogenetic relationships amongst bacterial genera. Therefore, we considered this question very relevant for this study. To answer this question, the PC was calculated between patristic distances in an rpoD ML tree and the corresponding patristic distances in a 16S rRNA gene tree. To avoid the possibility that a high correlation would have been obtained by chance, PCs were calculated between all 16S rRNA and rpoD gene trees that were obtained from different tree searches on the 16S rRNA and rpoD gene libraries respectively. The patristic distance correlations ranged from 0.7777 to 0.8122 (Table 1) for the different tree comparisons. To visualize these numbers and to understand their meaning, a correlation plot was constructed for the rpoD(1) versus 16S(1) tree comparison (Fig. 3). Fig. 3 shows a clear positive relation between patristic distances in 16S rRNA trees and corresponding patristic distances in rpoD trees, and little deviation from a straight line relationship. This indicates that patristic distances in the rpoD gene tree follow corresponding patristic distances in the 16S rRNA gene tree, indicating that rpoD gene based phylogeny is, generally speaking, similar to 16S rRNA gene based phylogeny. Also in this case a high correlation was found between wRMSD values and corresponding PC values ($R = -0.85$), thus giving extra support for the positive association between patristic distances in rpoD gene trees and patristic distances in 16S rRNA gene trees.

To see how patristic distances in the rpoD tree related to patristic distances in the 16S rRNA gene tree, we calculated the slope of the line of best fit connecting the data points and the origin (Table 1). On average, patristic distances between rpoD sequences were 25.3 times higher than patristic distances between 16S rRNA gene sequences. This indicates a higher evolutionary rate of the rpoD gene, which
by definition implies that the organisms can be distinguished at a finer taxonomic level based on \textit{rpoD} sequences.

\textbf{Evaluation of \textit{Pseudomonas} diversity on different media}

\textbf{Construction of rarefaction curves.} \textit{Pseudomonas} diversity on all three growth media was expressed in terms of \textit{rpoD} sequence diversity. In theory, a one-base difference between a pair of sequences may indicate that both sequences originate from different bacterial strains. As we were interested in the \textit{Pseudomonas} diversity at the strain level, an OTU in this work had to be defined as a unique sequence (i.e. similarity cut-off of 100%). However, an OTU definition of 99\% similarity was used instead to avoid the effect of possible sequencing errors (Acinas \textit{et al.}, 2005). Since each of the clone libraries differed in size, OTU richness in the different samples was compared by rarefaction (Dunbar \textit{et al.}, 1999; Hughes \textit{et al.}, 2001; Moyer \textit{et al.}, 1998). Richness estimation based on curve extrapolation methods requires data from relatively well sampled communities (Hughes \textit{et al.}, 2001). However, this was not the case in this study. Therefore, we decided not to calculate richness estimators based on the obtained rarefaction curves, but use the rarefaction curves directly for data interpretation. To check whether or not the type of medium yielding the largest diversity varied with the sample being analysed, we constructed clone libraries for six different samples. As shown by the rarefaction curves in Fig. 4, the \textit{Pseudomonas} diversity was different for the three investigated media. The outcome also depended on the sample being analysed. In all but two samples (E2 root and E3 root), the \textit{Pseudomonas} specific medium (PIA) generated the lowest diversity. In each of the samples either PDA or TSA was found to pick up the largest diversity. Exceptions were the E2 and E3 root samples, where TSA was the least well-suited medium.

\textbf{Fast UniFrac analysis.} To check whether \textit{Pseudomonas} diversity obtained from the three media overlapped, we used the Fast UniFrac webtool (Hamady \textit{et al.}, 2010). Fast UniFrac allows the comparison of microbial communities based on phylogenetic information. All the analyses performed were unweighted, i.e. not taking into account sequence abundances. Due to biases that are inherent in working with clone libraries, such as PCR-bias and differences in efficiency of the \textit{E. coli} cells with respect to the uptake of amplicon sequences, the relative abundances obtained were considered not representative for the true abundances in the samples. Therefore, weighted UniFrac analyses were not applicable here. Fig. 5 shows that in most cases PIA samples clustered separate from TSA and PDA samples. With the exception of E1 PIA RH and E3 PIA RO, all PIA samples obtained from both rhizosphere and roots from the three locations clustered together, which indicated that the diversities overlapped. These findings

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig3.png}
\caption{The Pearson correlation plot obtained from corresponding patristic distances in a tree generated from 16S rRNA gene sequences (x-axis) and a tree generated from \textit{rpoD} gene sequences (y-axis). Patristic distances were subjected to a binning step prior to plotting. The binning step was performed by sorting patristic distances for the 16S rRNA gene tree, and subsequently calculating the mean patristic distances in the 16S rRNA and \textit{rpoD} trees over each patristic distance interval of 0.001 in the 16S rRNA gene tree. Mean 16S rRNA distances and corresponding mean \textit{rpoD} distances were then plotted in this graph, and the standard deviations on the mean \textit{rpoD} distances were represented by error bars on the chart.}
\end{figure}
illustrate that although the diversity obtained with PIA was generally lowest, the medium revealed a different diversity compared to TSA and PDA. Therefore, PIA is an interesting medium to use in parallel with either PDA or TSA. PDA and TSA on the other hand, appeared in the same clusters in most of the cases (i.e. E1 RH, E1 RO, E3 RH and E3 RO), illustrating that the obtained diversities overlapped. The UniFrac distances between the different samples investigated are shown in Table 2.

**Taxonomic assignment of rpoD sequences**

To obtain insight into the taxonomic diversity that was picked up from the media, the rpoD sequences were assigned using the Mothur v1.27.0 software. The reference dataset used was constructed from *Pseudomonas* type strain rpoD sequences. Whether or not rpoD sequences allow species identification is still not proven, regardless of the high taxonomic resolution and the phylogenetic congruence with 16S rRNA gene based phylogeny that was
Table 2. Diversity dissimilarity between the different samples analysed

Sample names correspond to names given in Table S1. Bold and italics indicate the dissimilarity between diversities obtained on the three different media for a given sample.


Cultured Andean potato-associated Pseudomonas diversity
observed for the rpoD gene in this study. Therefore, species names listed in Table S4 should be considered as tentative, rather than as exact identifications.

**DISCUSSION**

Many members of the genus *Pseudomonas* are known to harbour beneficial properties. For instance within the field of plant growth-promotion, a number of *Pseudomonas* strains have been commercialized to decrease or replace the application of chemical fertilizers and pesticides. To advance the discovery of *Pseudomonas* strains with beneficial properties, this study evaluated two generally used growth media – TSA and PDA – and one *Pseudomonas* specific medium – PIA – for their abilities to grow members of the genus *Pseudomonas*. Since chances of encountering beneficial *Pseudomonas* isolates increase with increasing *Pseudomonas* diversity on the media, the growth medium that yielded the largest diversity was considered to be the most interesting medium to conduct isolation campaigns with.

Of all media investigated, the *Pseudomonas* specific medium (PIA) resulted in the lowest diversity of *Pseudomonas* isolates, and was thus considered the least interesting of the three media tested to conduct isolation campaigns with. TSA on the contrary, outperformed the other two media in all but two samples for which PDA scored best. However, as compared to TSA, PDA never scored worst. Fast UniFrac (Hamady et al., 2010) analyses showed that in many cases the diversity picked up with PIA differed from the diversities obtained with either TSA or PDA. The latter two were found to show some extent of overlap in most of the samples investigated. Since no relation between choice of growth medium and yield of beneficial isolates has been established, our results suggest that the best results may be achieved from cultivations on either PDA or TSA and from PIA in parallel.

Since bacterial properties may be strain specific, a technique with a highly differentiating power was required to assess *Pseudomonas* diversity. The role of housekeeping genes in resolving the taxonomy of *Pseudomonas* has been established previously. Yamamoto et al. (2000) showed, based on combined gyrB and rpoD sequences, that the genus *Pseudomonas* diverges into two intragenic clusters IGCI and IGCCI, which could be further subdivided into a number of subclusters. Mulet et al. (2008) found a clear correlation between phylogenetic similarities based on concatenated sequences of the 16S rRNA, gyrB and rpoD genes on the one hand, and DNA–DNA relatedness values expressed as ATm on the other hand, for members of the *P. stutzeri* group. Later, the same authors (2010) proposed a MLSA scheme based on concatenated sequences of the 16S rRNA, rpoB, rpoD and gyrB genes, which allowed a thorough identification of *Pseudomonas* isolates at the *Pseudomonas* group or subgroup level. To select the most interesting gene for the purpose of our study, we used the TaxonGap software (Slabbinck et al., 2008). This software allowed calculating and visualizing the heterogeneity of gene sequences within and separability between *Pseudomonas* species subgroups that were part of the larger *Pseudomonas fluorescens* group based on the information obtained from each of the four genes. Results showed that the rpoD gene had the highest resolution within the *Pseudomonas fluorescens* group. Analyses on gene phylogeny showed similar results, as the slope of the best fitting line forced through the origin and connecting the data points in 16S rRNA gene versus rpoD patristic distance correlation plots also indicated a high taxonomic resolution of the rpoD gene. Our results confirm previous observations by Parkinson et al. (2011), who reported a high resolution of the rpoD gene for *Pseudomonas* species belonging to the *Pseudomonas syringae* complex. Similarly, Yamamoto et al. (2000) found that the phylogenetic distances between *Pseudomonas* rpoD sequences were generally larger than the phylogenetic distances between corresponding gyrB sequences, thus indicating its higher resolving power.

The PC, wRMSD, RF and WRF values that were obtained from comparisons of trees generated from different tree searches on the rpoD sequence library indicated a higher phylogenetic content of the rpoD compared to the 16S rRNA gene. Furthermore, it was shown that rpoD phylogeny was similar to 16S rRNA gene based phylogeny. This observation, however, is in contradiction with results obtained by Yamamoto & Harayama (1998), who showed that the genetic distances in the variable regions of the 16S rRNA gene correlated poorly with corresponding distances between the rpoD genes. However, contrary to their interesting approach, this study took into account the complete 16S rRNA gene instead of filtering out its conserved regions to focus on specific variable regions and vice versa. Furthermore, their findings were based on the comparison of pairwise distances, while in our opinion a comparison of phylogenetic distances gives a superior picture as they represent evolutionary history. The latter is necessary to compare gene phylogenies. Ultimately, we were able to base our analyses on a larger set of 129 strains of *Pseudomonas* species, while due to the limited amount of sequences available at that time, Yamamoto & Harayama (1998) had to restrict their analyses to 20 sequences.

Mulet et al. (2009) developed primers that allow specific targeting of the rpoD gene in a wide range of *Pseudomonas* species. Their primers were designed based on rpoD sequences of 35 species representing the different intragenic phylogenetic *Pseudomonas* clusters. Subsequent testing of these primers by amplifying the rpoD gene of 96 *Pseudomonas* type strains and a well characterized *Pseudomonas* collection of more than 100 strains indicated their universality within *Pseudomonas*. We considered that taxonomic assignment would have given additional insight into primer universality. However, from the 130 *Pseudomonas* type strain rpoD sequences that were available, from which our reference database was constructed, only 29 were
mentioned in the assignment report. Considering the low bootstrap values obtained in some identifications (Table S4), and because the number of currently known Pseudomonas species largely exceeds the 130 Pseudomonas species in our reference database, primer universality could not be investigated based on the results obtained. The media used may also have narrowed the number of Pseudomonas species. Sequence assignment indicated that a number of rpoD sequences showed only limited bootstrap support for assignment to rpoD sequences of already known Pseudomonas type strains. Considering the unexplored origin of the samples, we reason that this may be attributed to the fact that currently existing rpoD databases are underrepresented.

Our results illustrate that rpoD gene phylogeny is similar to 16S rRNA gene based phylogeny. In addition, it was found to have the highest taxonomic resolution within the P. fluorescens group amongst the four biomarker genes investigated. Still, we cannot exclude the possibility that different strains may share 100% rpoD gene sequence identity, which is why we acknowledge the fact that our measurement of Pseudomonas diversity on the agar plates may have been an underestimation. Still, as this underestimation occurred for all three media investigated, we do not believe that this weakness biased the results obtained in this study. Our results show that either TSA or PDA is recommended when isolation campaigns are performed from one single medium. However, the best choice may depend on the sample being analysed. A more optimal method would be to use either TSA or PDA in combination with PIA, considering the different communities obtained from each medium. We based this research on the hypothesis that an increased bacterial diversity increases chances of yielding beneficial strains.

ACKNOWLEDGEMENTS

The work was supported by the European Commission’s Seventh Framework Program FP7/2007-2013 under grant agreement no. 27552. K.H. is supported by the Flemish Fund for Scientific Research (FWO11/PDO/084).

REFERENCES


Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S. & other authors...


Edited by: W. Achouak