

## ***Azotobacter vinelandii*: a *Pseudomonas* in disguise?**

*Azotobacter vinelandii* is a widely distributed free-living soil bacterium. This Gram-negative, strictly aerobic bacterium has many interesting features, including the ability to grow on a wide variety of carbohydrates, alcohols and organic acids, alginate production and nitrogen fixation. *A. vinelandii* synthesizes three different types of nitrogenase enzymes. Unlike most diazotrophs, *A. vinelandii* is able to fix nitrogen in the presence of atmospheric oxygen concentrations. To protect its nitrogenase enzymes from oxygen inactivation, this bacterium is equipped with particular physiological mechanisms, such as a high respiration rate. Currently, *Azotobacter* is classified as a genus of the family *Pseudomonadaceae* (<http://www.azotobacter.org>).

Using *in vivo* expression technology (IVET), we screened a nitrogen-fixing *Pseudomonas stutzeri* strain for genes specifically expressed in rice rhizosphere (Rediers *et al.*, 2003). Sequence analysis of these genes revealed high sequence similarity not only to genes of several other *Pseudomonas* strains but also to *A. vinelandii* genes. These observations and the availability of a (draft) genome sequence of *A. vinelandii* and several *Pseudomonas* species urged us to inspect the phylogenetic relationship of *A. vinelandii* to *Pseudomonas sensu stricto*.

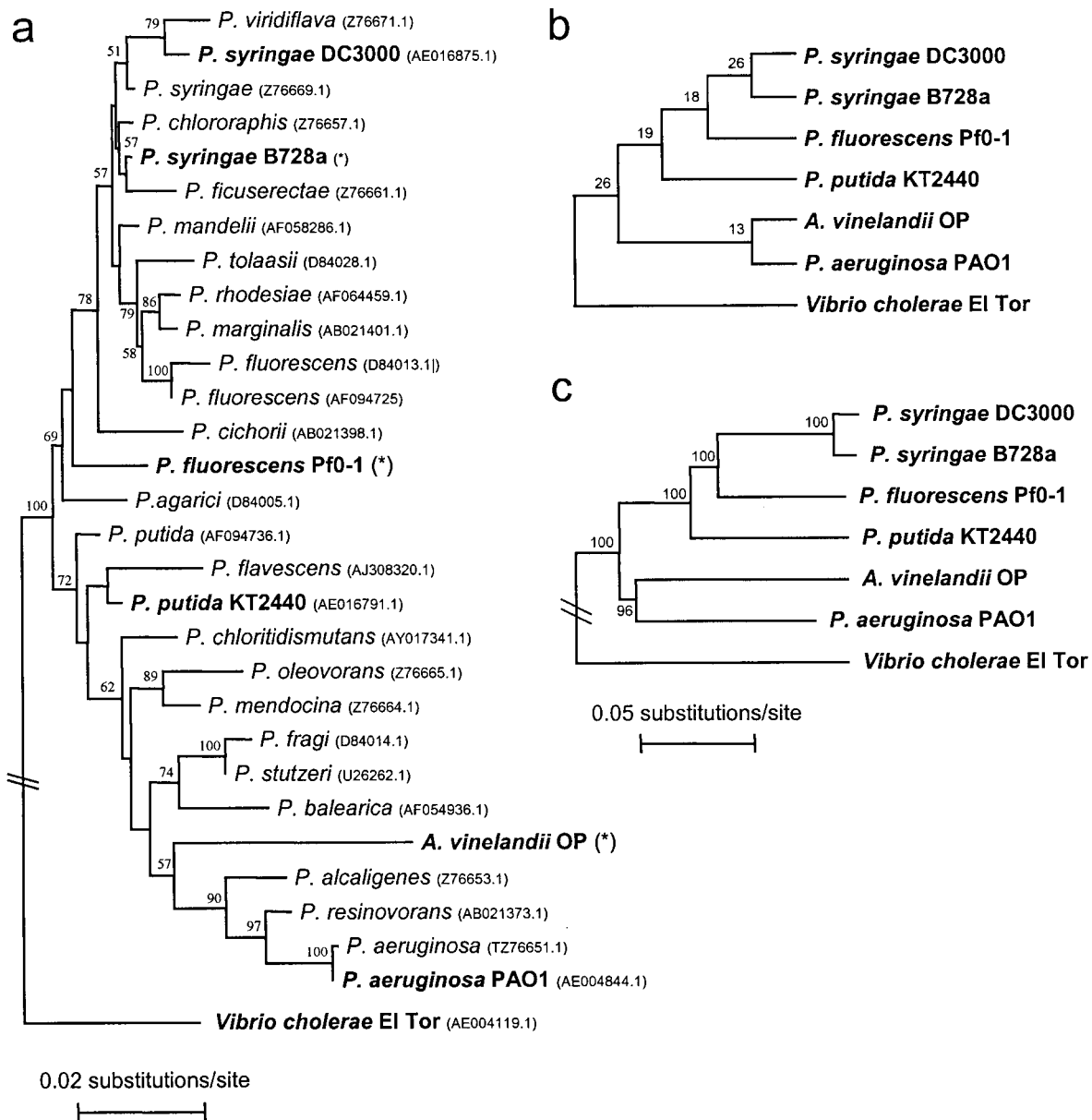
Analysis of 16S rRNA gene sequences is often used and generally accepted for analysing phylogenetic relationships. A phylogenetic tree based on partial 16S rRNA gene sequences (nucleotides 135–1475 of the *Pseudomonas aeruginosa* PAO1 16S rRNA gene) assigns *A. vinelandii* to the *P. aeruginosa* clade (Fig. 1a). The 16S rRNA gene sequences of *P. aeruginosa* PAO1 and *A. vinelandii* share higher identity with each other (96%) than they do with the corresponding sequences of other *Pseudomonas* strains. This is in line with the following observations. To detect *Pseudomonas sensu stricto* in different environments, a highly sensitive PCR protocol was developed to amplify 16S rRNA genes (Widmer *et al.*, 1998). Members of the genus *Pseudomonas*

display three distinct *Hae*III RFLP patterns of the amplified PCR products. We found by *in silico* analysis that *A. vinelandii* and *P. aeruginosa* PAO1 indeed share the same *Hae*III RFLP pattern. In another study, it was shown that the 16S rRNA gene sequences of *Pseudomonas* strains contain repeating elements which are highly conserved (Purohit *et al.*, 2003). We confirmed that these repeating elements were also found in *A. vinelandii*, while only five out of eight were found in *Escherichia coli*.

However, it is debatable whether a phylogenetic tree based on the comparison of a single gene is sufficient to accurately represent the evolution of bacterial species. Furthermore, since the 16S rRNA gene shows an extremely low rate of evolution, the degree of resolution obtained is not sufficient to infer intragenetic relationships. This can be circumvented by phylogenetic analysis of protein-encoding genes. In several studies, housekeeping genes have proven to be an appropriate target to assess phylogenetic relationships between bacteria, because these genes are usually highly expressed, highly conserved and evolve more rapidly than rRNA genes, resulting in a desirable resolution in evolutionary relationships between species (Coenye & Vandamme, 2003; Lerat *et al.*, 2003; Wertz *et al.*, 2003). For example, it has been shown that the highly conserved DNA recombinase (RecA) can be used to infer phylogeny (Eisen, 1995). RecA of *P. aeruginosa* PAO1 shows highest amino acid sequence identity (91%) with the *A. vinelandii* homologue. This observation was also reflected in a dendrogram based on the RecA amino acid sequences (not shown). Likewise, the sequences of both GyrB, encoding a DNA gyrase, and RpoD, encoding the  $\sigma^{70}$  sigma factor, have been used to reconstruct a phylogenetic tree of the genus *Pseudomonas* (Yamamoto *et al.*, 2000). Again, for both proteins as well as for the concatenated GyrB–RpoD amino acid sequences we noticed that *A. vinelandii* clusters with *P. aeruginosa* PAO1.

As the number of completely sequenced bacterial genomes has increased, there has been an increasing interest in the use of these genome sequence data to assess evolutionary relationships among bacterial

species. We used data retrieved from the complete genome sequences of *P. aeruginosa* PAO1, *Pseudomonas putida* KT2440 and *Pseudomonas syringae* pv. *tomato* str. DC3000 and from draft sequences of *Pseudomonas fluorescens* Pf0-1 and *P. syringae* pv. *syringae* B728a to scrutinize the taxonomic position of *A. vinelandii* strain OP (<http://www.azotobacter.org>). For phylogenetic tree construction, respective sequences of *Vibrio cholerae* El Tor N16961 were used as outgroup. We analysed the phylogenetic relationships between these seven species for several proteins/enzymes with housekeeping functions: AroK and DapB, involved in the synthesis of aromatic amino acids and lysine/peptidoglycan, respectively; CarB (large subunit of carbamoylphosphate synthetase), which is involved in arginine and pyrimidine nucleotide biosynthesis; protein synthesis (AlaS, a tRNA charger, and RplA, a ribosomal protein); peptidoglycan synthesis (MurD); secretion (SecA); stringent response (RelA, involved in ppGpp metabolism); intracellular ATP-dependent proteolysis (FtsH); co-enzyme biosynthesis (HemB and ThiE, involved in tetrapyrrole and thiamin biosynthesis, respectively); fatty acid synthesis (AccD, acetyl-CoA carboxylase subunit); glycolysis (FruK, phosphofructokinase); gluconeogenesis (PckA, phosphoenolpyruvate carboxykinase); heat-shock proteins (DnaK, GroEL and GrpE); purine nucleotide biosynthesis (PurF); regulation (GacA, response regulator); DNA modification (GyrB, RecA); DNA transcription (RpoN and RpoD, sigma factors); and energy metabolism (AtpA, a component of ATP synthase, and AceA, pyruvate dehydrogenase component E1). Generation of phylogenetic trees for all the proteins in our dataset revealed that the *A. vinelandii* homologues clustered within, or close to, the *Pseudomonas* clade (data not shown). In most cases the *A. vinelandii* proteins were most closely related to the *P. aeruginosa* PAO1 orthologues. Based on tree topologies of the 16S rDNA nucleotide and the 25 protein alignments, a consensus tree was built (Fig. 1b). This tree revealed an identical topology compared to the 16S rDNA-based dendrogram when only the seven mentioned species are considered. The concatenated sequence alignments of



**Fig. 1.** Phylogenetic relationships between *A. vinelandii* and *Pseudomonas* species (*sensu stricto*) based on comparison of 16S rRNA gene sequences (a) and the concatenated amino acid sequences of 25 proteins/enzymes with housekeeping functions (c). Nucleotide accession numbers are given in parentheses. Nucleotide sequences designated (\*) were retrieved from the DOE Joint Genome Institute (<http://www.jgi.doe.gov/>). Nucleotide and amino acid sequence alignments were carried out with the program CLUSTAL W 1.8. Dendrograms were generated with TREECON (Van de Peer & De Wachter, 1994) using the neighbour-joining algorithm with Poisson correction. *V. cholerae* El Tor N16961 orthologues were used as outgroups for phylogenetic tree reconstruction. The percentage of trees from 1000 bootstrap resamples supporting the topology is indicated when above 50. A consensus tree (b) was generated based on the separate phylogenetic trees of the 16S rRNA gene sequences and of the 25 proteins with ECONSENSE of the PHYLIP software package. The numbers at the forks indicate the number of times the group, consisting of the species which are to the right of that fork, occurred among the 26 trees.

all proteins in our dataset (consisting of 12 537–12 581 amino acid residues) yielded a phylogenetic tree with the same topology as the 16S rDNA and consensus trees (Fig. 1c).

Until now, the major outer-membrane protein OprF has been found only in *Pseudomonas sensu stricto* species and may be considered a diagnostic protein for *Pseudomonas* (De Mot *et al.*, 1994;

Vermeiren *et al.*, 1999). *A. vinelandii* has an OprF orthologue which showed high amino acid identity with the *P. aeruginosa* PAO1 OprF (68%), while identity between the *Pseudomonas fuscovaginae* and the

*P. aeruginosa* PAO1 OprF was only 49%. Construction of a phylogenetic tree based on OprF sequences confirmed the close relationship between *P. aeruginosa* PAO1 and *A. vinelandii* (not shown). Like OprF of *P. aeruginosa* PAO1, the *A. vinelandii* orthologue appears to be a 'cysteine-type' OprF, as it contains a central domain with four conserved cysteine residues, which is absent in several other OprF proteins (De Mot *et al.*, 1994). Alginate production is a common feature of *A. vinelandii* and *Pseudomonas* species (Rehm & Valla, 1997). Homology searches of *P. aeruginosa* PAO1 AlgX (a protein involved in alginate biosynthesis) revealed that AlgX orthologues are only present in other *Pseudomonas* strains and in *A. vinelandii*.

In conclusion, these data strongly suggest that *A. vinelandii* belongs to the genus *Pseudomonas sensu stricto*. Originally, nitrogen-fixing ability was considered a major physiological characteristic differentiating *A. vinelandii* from *Pseudomonas* species. The occurrence of nitrogen fixation in *Pseudomonas* species has been long debated, but in recent years several genuine *Pseudomonas* strains that can fix nitrogen have been identified (Kulakov *et al.*, 2002; Vermeiren *et al.*, 1999). Most likely, the nitrogen-fixing genes were acquired by lateral gene transfer. This hypothesis is supported by NifH phylogenies (Vermeiren *et al.*, 1999). Deduced amino acid sequences of the *P. stutzeri* A15 *nifHDK* operon revealed highest identity (87–91%) with the respective *A. vinelandii* homologues (Desnoues *et al.*, 2003). Moreover, it was shown that the gene organization in the *nifH* region of *A. vinelandii* was identical to the gene organization of *P. stutzeri* A15.

### Acknowledgements

H. R. is indebted to the 'Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen' for a predoctoral fellowship. Part of the data used in this study was generated by the Microbial Genomics Program of the DOE Joint Genome Institute.

**Hans Rediers, Jos Vanderleyden and René De Mot**

Centre of Microbial and Plant Genetics, Katholieke Universiteit Leuven, B-3001 Heverlee, Belgium

**Correspondence:** Hans Rediers (hans.rediers@agr.kuleuven.ac.be)

**Coenye, T. & Vandamme, P. (2003).**

Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case. *Microbiology* **149**, 3507–3517.

**De Mot, R., Schoofs, G., Roelandt, A., Declerck, P., Proost, P., Van Damme, J. & Vanderleyden, J. (1994).**

Molecular characterization of the major outer-membrane protein OprF from plant root-colonizing *Pseudomonas fluorescens*. *Microbiology* **140**, 1377–1387.

**Desnoues, N., Lin, M., Guo, X., Ma, L., Carreno-Lopez, R. & Elmerich, C. (2003).**

Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology* **149**, 2251–2262.

**Eisen, J. A. (1995).** The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41**, 1105–1123.

**Kulakov, L. A., McAlister, M. B., Ogden, K. L., Larkin, M. J. & O'Hanlon, J. F. (2002).** Analysis of bacteria contaminating ultrapure water in industrial systems. *Appl Environ Microbiol* **68**, 1548–1555.

**Lerat, E., Daubin, V. & Moran, N. A. (2003).** From gene trees to organismal phylogeny in prokaryotes: the case of the  $\gamma$ -Proteobacteria. *PLoS Biol* **1**, E19.

**Purohit, H. J., Raje, D. V. & Kapley, A. (2003).** Identification of signature and primers specific to genus *Pseudomonas* using mismatched patterns of 16S rDNA sequences. *BMC Bioinformatics* **4**, 19.

**Rediers, H., Bonnacarrère, V., Rainey, P. B., Hamonts, K., Vanderleyden, J. & De Mot, R. (2003).** Development and application of a *dapB*-based *in vivo* expression technology system to study colonization of rice by the endophytic nitrogen-fixing bacterium *Pseudomonas stutzeri* A15. *Appl Environ Microbiol* **69**, 6864–6874.

**Rehm, B. H. & Valla, S. (1997).** Bacterial alginates: biosynthesis and applications. *Appl Microbiol Biotechnol* **48**, 281–288.

**Van de Peer, Y. & De Wachter, R. (1994).** TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10**, 569–570.

**Vermeiren, H., Willems, A., Schoofs, G., de Mot, R., Keijers, V., Hai, W. & Vanderleyden, J. (1999).** The rice inoculant strain *Alcaligenes faecalis* A15 is a nitrogen-fixing *Pseudomonas stutzeri*. *Syst Appl Microbiol* **22**, 215–224.

**Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. (2003).** A molecular phylogeny of

enteric bacteria and implications for a bacterial species concept. *J Evol Biol* **16**, 1236–1248.

**Widmer, F., Seidler, R. J., Gillevet, P. M., Watrud, L. S. & Di Giovanni, G. D. (1998).** A highly selective PCR protocol for detecting 16S rRNA genes of the genus *Pseudomonas sensu stricto* in environmental samples. *Appl Environ Microbiol* **64**, 2545–2553.

**Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., 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.

DOI 10.1099/mic.0.27096-0