Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes

Satoshi Yamamoto and Shigeaki Harayama

Author for correspondence: Satoshi Yamamoto. Tel: +81 43 203 0226. Fax: +81 43 248 2109.
e-mail: yamamotost@nichirei.co.jp

Marine Biotechnology Institute, 3-75-1 Heita, Kamaishi, Iwate 026, Japan

Phylogenetic analysis of 20 *Pseudomonas* strains (*Pseudomonas putida*, *Pseudomonas fluorescens* and *Pseudomonas chlororaphis*) was conducted by using the nucleotide sequences of the genes for 16S rRNA, DNA gyrase B subunit (*gyrB*) and RNA polymerase σ70 factor (*rpoD*), which have been determined by the direct sequencing of PCR-amplified fragments. On the basis of *gyrB* and *rpoD* sequences, these strains were split into two major clusters: one including the type strain of *P. putida* and all biovar A strains and the other including all *P. putida* biovar B strains, *P. fluorescens* strains and the *P. chlororaphis* strain. In the phylogenetic tree reconstructed from the 16S rRNA sequences including variable regions, *P. putida* biovar A and B strains were not separated into two independent clusters, whereas in the phylogenetic tree reconstructed from the 16S rRNA sequences excluding the variable region sequences, these strains were separated into *P. putida* biovar A and biovar B clusters. The pairwise distances estimated from the variable regions of 16S rRNA correlated poorly with the synonymous distances estimated from the *gyrB* and *rpoD* genes. On the other hand, a highly significant correlation was observed between the pairwise distances estimated from the non-variable regions of 16S rRNA and the synonymous distances from *gyrB* and *rpoD* genes. Consequently, only the 16S rRNA sequences in the non-variable regions should be used for the phylogenetic analysis. The *gyrB* and *rpoD* analyses showed the necessity for the reclassification of *P. putida* biovar B strains.

Keywords: *gyrB*, *rpoD*, 16S rRNA, *Pseudomonas*, PCR

INTRODUCTION

One goal of systematics is the establishment of a phylogeny-based classification system (31), and the nucleotide sequences of various genes, especially those of small-subunit (SSU) rRNA, are widely used for such a purpose (27, 41). A comparison of the SSU rRNA sequences from various organisms has resulted in many new discoveries, including that of the *Archaea*, a group of prokaryotes that is separated from the *Bacteria*. More than 2000 bacterial SSU rRNA (16S rRNA) sequences are now available from the Ribosomal Database Project (RDP) (20), and a 16S rRNA-based phylogenetic tree of the *Bacteria* has been constructed (27). The current recommendation for bacterial taxonomy is the inclusion of 16S rRNA sequence data in the description of a new genus (35).

However, the results of the SSU rRNA-based analysis often do not correlate with the DNA reassociation values determined by DNA hybridization which has been used as the criterion for definition of bacterial species (12, 34). In these cases, it seems that the resolution of SSU rRNA-based analysis is low due to the small numbers of substitutions between compared SSU rRNA sequences. To resolve the phylogenetic relationships of closely related organisms, it would be necessary to use other gene sequences which provide a higher resolution than that of SSU rRNA. In this study, we examined the phylogenetic relationships of...
twenty Pseudomonas strains (P. putida, P. fluorescens and P. chlororaphis) on the basis of the nucleotide sequences of their genes for 16S rRNA, DNA gyrase B subunit (gyrB) and $\sigma^{70}$ factor (rpoD). DNA gyrase is the enzyme responsible for introducing negative supercoils into bacterial chromosomes and plays a crucial role in the replication of chromosomes (39), while the $\sigma^{70}$ factor is one of the sigma factors that confers promoter-specific transcription initiation on RNA polymerase (19). Both DNA gyrase and the $\sigma^{70}$ factor are essential for cell survival in bacteria, and horizontal transmission of these genes may be as rare as that of rRNA genes. Protein-encoding genes have been reported to evolve much faster than rRNAs (26); thus a phylogenetic analysis using the gyrB and rpoD sequences was expected to provide higher resolution than one using 16S rRNA sequences.

**METHODS**

**Bacterial strains.** P. putida strains IFO 14164$^T$ [ = ATCC 12633, biovar A (36)], IFO 14671, IFO 3738 [ = ATCC 8209, *Pseudomonas ovalis*]), ATCC 11172 [biovar A (21)], ATCC 17484 [biovar B (1)], ATCC 17522 [biovar B (36)], ATCC 23973 (9), JCM 6156 [ = ATCC 33015, mt-2 (40)], NCIMB 9816 (1), A10L (32), BH (37), EST1033 [ = Pseudomonas aeruginosa], were grown aerobically in nutrient broth at 30 °C, and was grown aerobically in nutrient broth at 30 °C, and was grown aerobically in nutrient broth at 30 °C.

**Preparation of chromosomal DNA.** Each bacterial sample was grown aerobically in nutrient broth at 30 °C, and was grown aerobically in nutrient broth at 30 °C, and was grown aerobically in nutrient broth at 30 °C.

**PCR amplification and direct sequencing of the gyrB and rpoD genes.** A set of PCR primers, 70F and 70R, were used to amplify rpoD gene fragments. The sequence of 70F [5'-AGGATGACACCGGGTACCGCATGTA(TC)ATG(CA)G(ATGC)GA(AG)ATG-3', mixed bases being indicated in parentheses], and that of 70R [5'-ATAGAAATAACCAGA CGTAAGTT-3'], mixed bases being indicated in parentheses, were designed from the amino acid sequences in the regions of RNA polymerase $\sigma^{70}$ factors which are conserved between Buchnera aphidicola, Escherichia coli K-12 (4), Salmonella typhimurium and Pseudomonas aeruginosa: namely, MYMREMGTV (positions 100–108 of *E. coli* $\sigma^{70}$) and KKEMVEAN (positions 376–383 of *E. coli* $\sigma^{70}$), respectively. The first 23 nucleotides from the 5' ends of the PCR primers are not degenerate, and these regions were used for priming the nucleotide sequencing as described below. PCR amplification was performed with a DNA thermal cycler 480 (Perkin-Elmer) by using a PCR reaction buffer (Perkin-Elmer) containing each of the deoxyribonucleoside triphosphates at a concentration of 200 mM, each of the primers at a concentration of 1 μM, 1 μg target DNA and 2.5 U Taq DNA polymerase in a total volume of 10 μl. A total of 30 amplification cycles were performed with template DNA: denaturation at 94 °C for 1 min, primer annealing at 59 °C for 45 s, and primer extension at 72 °C for 2 min. By using PCR primers 70F and 70R, DNA segments of about 800 bp were amplified from all the Pseudomonas strains examined. The amplified products were purified by preparative gel electrophoresis on 1.4% low-melting-temperature agarose (NuSieve GTG, FMC Bioproducts) according to the method described by Sambrook et al. (30) before the amplified segments were sequenced. The total nucleotide sequences of the amplified fragments were determined by using sequencing primers 70Fs (5'-ACGACCTGACCCGGTACCGATGTA-3'), and 70Rs (5'-ATAGAAATAACCAGA CGTAAGTT-3'), which correspond to the first 23 nucleotides of PCR primers 70F and 70R, respectively. The sequencing reaction was conducted by using a Taq DyeDeoxy Terminator cycle sequencing kit (Perkin-Elmer) according to the manufacturer's instructions, and the products were analysed with a 373A DNA sequencer (Perkin-Elmer).

**Data analysis.** The nucleotide sequences of 16S rRNA and the translated amino acid sequences of gyrB and rpoD were aligned by using the CLUSTAL W computer program (38). The gyrB and rpoD sequences were aligned based on the amino acid sequences of their products, while 16S rRNA sequences were aligned based on their secondary structures. The numbers of synonymous sites and synonymous substitutions between all possible pairs of the gyrB and rpoD genes were obtained by applying the method of Nei & Gojobori (23), using the MEGA computer program (17). The correction for multiple substitution was made by the Jukes–Cantor formula (16). Evolutionary trees were constructed with the PHYLP program package (10), using the neighbour-joining method (29) with genetic distances computed by using the Jukes–Cantor model (16). Two types of phylogenetic hypotheses were constructed from the 16S rRNA sequences: (i) the tree based on the 16S rRNA sequences including the sequences of variable regions V1–V8; (ii) the tree based on the 16S rRNA sequences excluding the variable regions V1–V8.

**RESULTS**

Comparison of the phylogenetic structures deduced from the nucleotide sequences of the gyrB, rpoD and 16S rRNA genes

The partial nucleotide sequences of the gyrB, rpoD and 16S rRNA genes from 20 Pseudomonas strains were determined, and phylogenetic trees based on these data were constructed by the neighbour-joining method (Fig. 1) using the genetic distances adjusted by the Jukes–Cantor formula for multiple substitutions...
Phylogeny of Pseudomonas strains

The topology of the gyrB- and rpoD-based trees was comparable. In these trees, Pseudomonas strains were grouped into two clusters, the first one containing 13 P. putida strains including all the biovar A strains, while the second one contained the remaining P. putida strains including all the biovar B strains, all the P. fluorescens strains, and the P. chlororaphis strain.

The topologies of the trees reconstructed from the 16S rRNA sequences including or excluding the sequences of variable regions were very different from each other (Fig. 1c, d). When the total 16S rRNA sequences were used for the analysis, Pseudomonas strain were divided into two major clusters, the cluster of P. putida strains and the cluster of the P. fluorescens and P. chlororaphis strains. Phylogenetic relationships between the P. putida strains were also largely different from those of the gyrB- and rpoD-based trees. On the other hand, when the sequences of non-variable regions were used for the analysis, P. putida strains were split into two independent clusters as in the gyrB- and rpoD-based trees. However, the detailed topology of the tree based on the 16S rRNA non-variable region was also different from that of the gyrB- and rpoD-based trees (e.g. position of P. putida IFO 3738 and EST1033). These topological differences between the gyrB- and rpoD-based trees and the 16S-rRNA-based trees were also observed when the trees were reconstructed by using the maximum-parsimony method (11) and unweighted pair group method with averages (UPGMA) (33).

Comparison of the genetic distances in the gyrB, rpoD and 16S rRNA genes

Most of the base substitutions in the gyrB and rpoD genes were synonymous, i.e. did not result in amino acid changes. For example, 185 nucleotide substitutions including 3 gaps were observed between the gyrB genes in strains NCIMB 9815 and K23-1. Of these, only 25 contributed to the amino acid substitution or to the introduction of a gap.

Fig. 1. Phylogenetic trees of 20 Pseudomonas strains based on the nucleotide sequences of the gyrB (a), rpoD (b) and 16S rRNA genes (c, d). Two types of phylogenetic trees were reconstructed from the 16S rRNA sequences: the tree based on the total 16S rRNA sequences including the sequences of variable regions (1329 bp) (c) and the tree based on the partial 16S rRNA sequences without the sequences of variable regions (1012 bp) (d). The trees were constructed by using the neighbour-joining method, and genetic distances were computed by the Jukes–Cantor model. The scale bar indicates the genetic distance of 0.1 or 0.01. The number shown next to each node indicates the percentage bootstrap value of 1000 replicates. The sequences from E. coli K-12 were treated as the outgroup. The topological characteristics of the gyrB- and rpoD-based phylogenetic trees produced by the neighbouring method, the maximum-parsimony method and the unweighted pair group method with averages (UPGMA) were almost identical (data not shown). However, the topological characteristics of the 16S RNA-based trees produced by these three methods were slightly different from each other.
The values on the upper right are the evolutionary distances between 16S rRNA sequences, corrected by the Jukes–Cantor formula (16) for multiple substitutions.

\[
\text{gyrB} = \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 0.10 & 0.09 & 0.09 & 0.09 & 0.09 \\
3 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
4 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
5 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
6 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
\end{array}
\]

The values on the lower left are base substitutions occurring in helix 6 of 16S rRNA sequences, corrected by the Jukes–Cantor formula (16) for multiple substitutions.

\[
\text{rpoD} = \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
3 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
4 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
5 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
6 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
\end{array}
\]

**Table 2. Evolutionary distances between 16S rRNA sequences, and numbers of base substitution in helix 6 and helix 18 of 16S rRNA**

The values on the upper right are the evolutionary distances between the 16S rRNA, corrected by the Jukes–Cantor formula (16) for multiple substitutions. The values on the lower left are base substitutions occurring in helix 6 (34 bp) and helix 18 (61 bp). ND, Base substitution numbers between \(E.\) coli K-12 and \(Pseudomonas\) strains were not determined because of ambiguous positions of gaps.

\[
\text{Strain} = \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
3 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
4 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
5 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
6 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
\end{array}
\]

**Table 3. Evolutionary distances between gyrB and rpoD sequences**

The values on the upper right and the lower left are the evolutionary distances between gyrB and rpoD sequences, respectively. Both values were corrected by the Jukes–Cantor formula (16) for multiple substitutions.

\[
\text{gyrB} = \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 0.10 & 0.09 & 0.09 & 0.09 & 0.09 \\
3 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
4 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
5 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
6 & 0.09 & 0.09 & 0.09 & 0.09 & 0.09 \\
\end{array}
\]

\[
\text{rpoD} = \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
3 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
4 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
5 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
6 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
\end{array}
\]

The synonymous distances calculated from the nucleotide substitution ratios at synonymous positions in the gyrB and rpoD genes were examined for all possible combinations of 20 \(Pseudomonas\) genes. A significant correlation between the synonymous distances in the gyrB genes and those in the rpoD genes was observed with a coefficient of correlation \(r=0.92\). This result was not unexpected, because it has been demonstrated that the synonymous substitution rate is constant for many chromosomal genes in many organisms, and can thus serve as a molecular clock of their evolution (18).

The base substitutions in the 16S rRNAs were not randomly distributed, but observed only at 71 positions among the 1329 bp of 16S rRNA sequences, mainly in so-called variable regions (53–71 = 746%). The substitutions were most frequently observed in helix 6 in variable region V1, and in helix 18 in variable region V2.
other variable regions. The base substitutions observed in the 34-bp-long helix 6 and in the 61 bp helix 18 are shown in Table 2. The relationships between the pairwise distances among helix 6 and 18 sequences and the synonymous distances in the gyrB and rpoD genes are illustrated in Fig. 2(a), while the relationships between the pairwise distances in the non-variable regions of the 16S rRNA and the synonymous distances in the gyrB and rpoD genes are shown in Fig. 2(b). There was a highly significant correlation between the genetic distances in the non-variable regions of the 16S rRNA and those in the gyrB and rpoD genes whereas the correlation between the genetic distances in the variable regions of the 16S rRNA and those in the gyrB and rpoD genes was much less significant. Therefore, it can be concluded that the base substitutions provoked in the variable regions of 16S rRNA are not reliable as molecular evolutionary clocks. Consequently, only the sequences of non-variable regions of 16S rRNA should be used for the phylogenetic analysis.

**DISCUSSION**

16S rRNA provides a scaffold for the assembly of ribosomal proteins into small subunits, and interacts with mRNA. Many base substitutions in 16S rRNA are located in a few helices called variable regions, indicating that regions other than the variable regions are crucial for ribosome functions. Generally, the multiple substitutions that are observed in the variable regions are mostly compensatory, thus keeping the helix structure unaltered (7, 14, 28). Presumably the secondary structures of 16S rRNA determined by the sequences in these variable regions are functionally important.

The genetic distances in the variable regions correlated poorly with the synonymous distances in the gyrB and rpoD genes (Fig. 2). This observation suggests that the base substitutions in these helices are not due to cumulative successive point mutations, but may be caused by single-event mutations introducing multiple substitutions. Hence, the mutations in the variable regions should not be included in the calculation of the genetic distance. The base substitution rates in the 16S rRNA outside the variable regions were very low between closely related bacteria, and a phylogenetic analysis based on the 16S rRNA sequences of the non-variable regions may have insufficient resolution to distinguish bacterial species.

The topologies of phylogenetic trees reconstructed from GyrB and RpoD amino-acid sequences (obtained by translating gyrB and rpoD sequences) were slightly different from those of gyrB and rpoD trees (data not shown). When the amino acid sequences were used for the analysis, the phylogenetic information from synonymous substitutions was lost. The synonymous distances estimated from the gyrB and rpoD genes were highly correlated, indicating that substitutions at the synonymous sites were not saturated in the present

![Image](image_url)
data set. Under such circumstances, a phylogenetic analysis which includes the synonymous substitutions may be more precise than that excluding them.

As already discussed, the phylogenetic distances calculated only by the synonymous substitution numbers in gyrB were almost identical to those calculated by the synonymous substitution numbers in rpoD. However, if the total substitution numbers were used for the calculation, the phylogenetic distances evaluated by the gyrB sequences were larger than those evaluated by the gyrB sequences (Table 1, Fig. 1). The larger nucleotide substitution rates in rpoD were due to larger non-synonymous substitution rates in rpoD: 13.9% of the amino acid residues (53/382) were changed among the 20 GyrB sequences, while 38.3% (103/269) were changed among the 20 RpoD sequences. It is known that different protein families have different amino acid substitution rates (6), and that such rates also vary in different regions within a single polypeptide. Therefore, it should be noted that the phylogenetic distance obtained from the total nucleotide substitutions is not an absolute measure of the divergence time.

Delineation of P. putida biovar B

In this study, we have shown that P. putida biovar B strains are phylogenetically separated from the main cluster of P. putida, including the type strain and biovar A strains. The following phenotypes are general characteristics of P. putida: Gram-negative rods with polar flagella, motile, produce fluorescein, unable to hydrolyse gelatin, give an egg-yolk reaction, do not denitrify, do not grow at 41°C and do not produce phenazine pigment. Strains of P. putida biovar B have been distinguished from biovar A mainly by their ability to grow on L-tryptophan and L-kyurenine (36). However, P. putida biovar B strains are more P. fluorescens-like in some respects. They can grow at 4°C and can utilize a broader range of sugars than biovar A strains (36). Thus, the closer phylogenetic positions of the biovar B strains to P. fluorescens strains than to biovar A strains deduced from the analyses using gyrB, rpoD and the non-variable region of 16S rRNA seem to be consistent with their phenotypic characteristics.

Unclustered P. putida NCIMB 9816 formed a cluster with biovar B strains in the trees based on gyrB, rpoD and the non-variable region of 16S rRNA (Fig. 1). Biochemical characteristics of strain NCIMB 9816 determined by the Biolog system (2) were more similar to those of biovar B strains ATCC 17484 and ATCC 17522 (80 and 82% similarity, respectively) than to those of biovar A strain P. putida IFO 14164T (61% similarity; our unpublished data). Furthermore, strain NCIMB 9816 can grow at 4°C (our unpublished data). These characteristics of strain NCIMB 9816 suggested that this strain belongs to the biovar B group. However, strain NCIMB 9816 could not grow on L-tryptophan and L-kyurenine although it could grow on glucose as a sole carbon source. Whether or not the utilization of these amino acids is a relevant marker to distinguish P. putida biosvars A and B should be re-examined using larger sets of biovar A and B strains. As an extension of such studies, the reclassification of P. putida biovar B strains is highly anticipated.

The strain IFO 3738 was grouped with neither P. putida biovar A strains nor biovar B strains in the trees based on gyrB, rpoD and non-variable region of 16S rRNA (Fig. 1). Thus, strain IFO 3738 may belong to an unascertained Pseudomonas species having phenotypical and morphological characteristics similar to P. putida. To clarify this point, comprehensive phylogenetic analysis of fluorescent group Pseudomonas species should be conducted.

ACKNOWLEDGEMENTS

We are grateful to Shigeto Miyachi and Kazuhide Yamasato for their encouragement, and to Ikuo Hiramatsu, Satoru Kanai and Hitoshi Ogawa for their technical assistance. This work was supported by Industrial Science and Technology Frontier Program from New Energy and Industrial Technology Development Organization.

REFERENCES


