

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 σ^{70} 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

† **Present address:** Food Science Laboratory, Research and Development Center, Nichirei Corporation, Shiminato 9, Mihama-Ku, Chiba 261, Japan.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are: 16S rRNA genes, D37923–D37925, D83788 and D85991–D86004; *gyrB*, D37926, D37297 and D86005–D86019; *rpoD*, D86020–D86036.

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 σ^{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 σ^{70} factor is one of the sigma factors that confers promoter-specific transcription initiation on RNA polymerase (19). Both DNA gyrase and the σ^{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 [a derivative of EST1001 (25)], KF715 (13), PpG7 [= ATCC 17485, biovar A (1, 44)], K23-1 (22) and PB4 (43), *P. fluorescens* strains IAM 12022^T [= ATCC 13525 (36)], DSM 50108 [= ATCC 17482 (36)] and NCIMB 9815, and *P. chlororaphis* strain IFO 3904^T [= ATCC 9446 (36)] were used. The ATCC, DSM, IAM, IFO, JCM and NCIMB strains were purchased respectively from American Type Culture Collection (Rockville, MD, USA), Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; Braunschweig, Germany), Institute of Applied Microbiology at University of Tokyo, (Tokyo, Japan), Institute for Fermentation, Osaka (Osaka, Japan), Japan Collection of Microorganisms at Riken (Saitama, Japan), and National Collections of Industrial and Marine Bacteria (Aberdeen, UK).

Preparation of chromosomal DNA. Each bacterial sample was grown aerobically in nutrient broth at 30 °C, and chromosomal DNA was prepared from the cells of overnight cultures, after being washed and resuspended in a TE buffer (10 mM Tris/HCl containing 1 mM EDTA, pH 7.5), by the method described by Johnson (15).

PCR amplification and direct sequencing of *rpoD*. A set of PCR primers, 70F and 70R, were used to amplify *rpoD* gene fragments. The sequence of 70F [5'-ACGACTGACCCGGTACGCATGTA(TC)ATG(CA)G(ATGC)GA(AG)ATGG(ATGC)AC(ATGC)GT-3', mixed bases being indicated in parentheses], and that of 70R [5'-ATAGAAATAAC-CAGACGTAAGTT(ATGC)GC(TC)TC(ATGC)ACCAT-(CT)TC(CT)TT(CT)TT-3'] were designed from the amino acid sequences in the regions of RNA polymerase σ^{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* σ^{70}) and KKEMVEAN (positions 376–383 of *E. coli* σ^{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 deoxynucleoside triphosphates at a concentration of 200 mM, each of the primers at a concentration of 1 μ M, 1 μ g target DNA and 2.5 U *Taq* DNA polymerase in a total volume of 10 μ l. 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 fragments were sequenced. The total nucleotide sequences of the amplified fragments were determined by using sequencing primers 70Fs (5'-ACGACTGACCCGGTACGCATGTA-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).

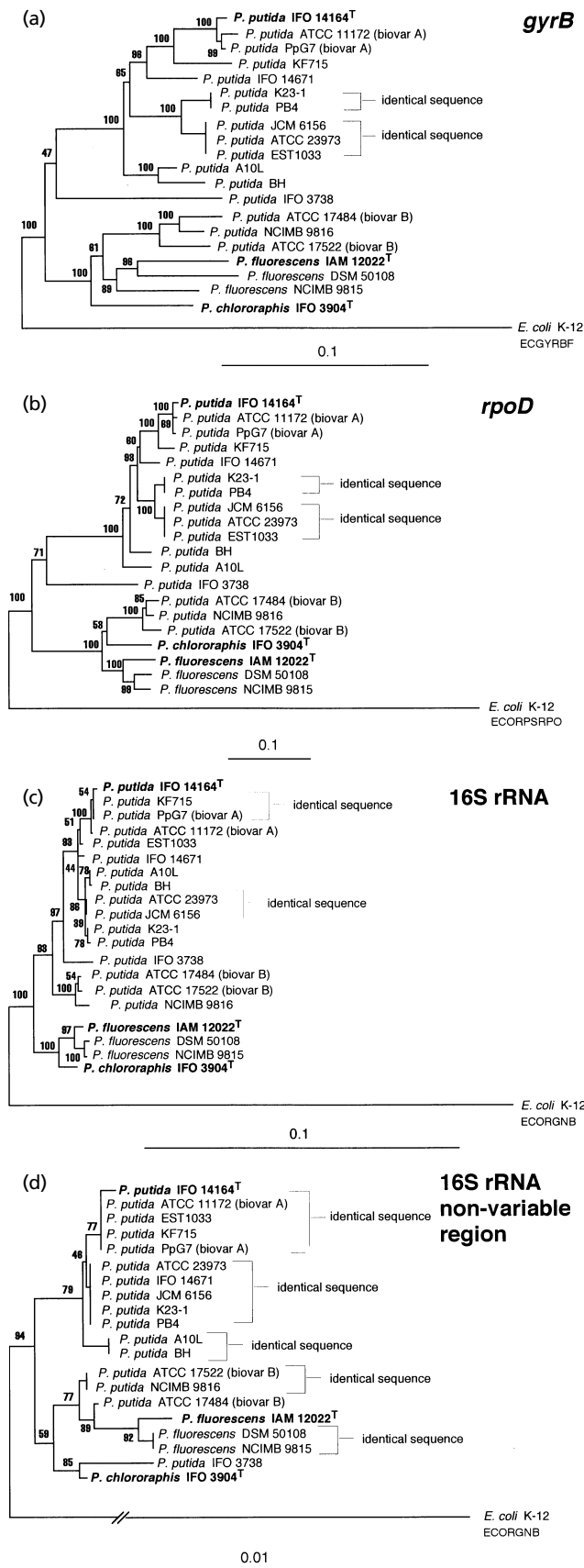
PCR amplification and direct sequencing of the *gyrB* and 16S rRNA genes. The *gyrB* fragments corresponding to positions 335–1482 of the *E. coli* K-12 *gyrB* sequence (42) were amplified from DNA of 20 *Pseudomonas* strains and sequenced as described previously (43). Partial nucleotide sequences of the 16S rRNA genes of 20 *Pseudomonas* strains corresponding to positions 37–1370 of the *E. coli* K-12 sequence (3) containing variable regions VI–V8 were also determined directly from the PCR fragments according to the method of Edwards *et al.* (8).

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 PHYLIP program package (10), using the neighbour-joining method (29) with genetic distances computed by using the Jukes–Cantor model (16). Two types of phylogenetic trees were constructed from the 16S rRNA sequences: (i) the tree based on the 16S rRNA sequences including the sequences of variable regions VI–V8; (ii) the tree based on the 16S rRNA sequences excluding the variable regions.

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



(Tables 1 and 2). 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* strains 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. The syn-

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 neighbour-joining method, the maximum-parsimony method and the unweighted pair group with averages (UPGMA) were almost identical (data not shown). However, the topological characteristics of the 16S rRNA-based trees produced by these three methods were slightly different from each other.

Table 1. Evolutionary distances between *gyrB* and *rpoD* sequences

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

Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 IFO 14164 ^T		0.014	0.009	0.062	0.153	0.084	0.084	0.084	0.090	0.090	0.076	0.097	0.053	0.171	0.163	0.163	0.168	0.151	0.167	0.152	0.315
2 ATCC 11172	0.010		0.009	0.065	0.154	0.090	0.090	0.090	0.095	0.095	0.078	0.101	0.060	0.176	0.166	0.168	0.172	0.158	0.172	0.157	0.327
3 PpG7	0.009	0.006		0.060	0.149	0.084	0.084	0.084	0.089	0.089	0.074	0.095	0.054	0.167	0.159	0.159	0.164	0.152	0.167	0.146	0.321
4 IFO 14671	0.076	0.074	0.073		0.149	0.069	0.069	0.069	0.070	0.070	0.063	0.074	0.067	0.166	0.156	0.153	0.164	0.143	0.167	0.152	0.317
5 IFO 3738	0.254	0.256	0.254	0.237		0.152	0.152	0.152	0.153	0.138	0.152	0.168	0.140	0.146	0.142	0.167	0.154	0.175	0.149	0.334	
6 JCM 6156	0.080	0.077	0.076	0.059	0.253		0.000	0.000	0.027	0.027	0.066	0.079	0.087	0.172	0.158	0.161	0.164	0.161	0.177	0.152	0.315
7 ATCC 23973	0.080	0.077	0.076	0.059	0.253	0.000		0.000	0.027	0.027	0.066	0.079	0.087	0.172	0.158	0.161	0.164	0.161	0.177	0.152	0.315
8 EST1033	0.080	0.077	0.076	0.059	0.253	0.000	0.000		0.027	0.027	0.066	0.079	0.087	0.172	0.158	0.161	0.164	0.161	0.177	0.152	0.315
9 K23-1	0.081	0.078	0.077	0.058	0.258	0.023	0.023	0.023		0.000	0.068	0.082	0.092	0.171	0.159	0.157	0.164	0.155	0.179	0.150	0.314
10 PB4	0.081	0.078	0.077	0.058	0.258	0.023	0.023	0.023	0.000		0.068	0.082	0.092	0.171	0.159	0.157	0.164	0.155	0.179	0.150	0.314
11 A10L	0.102	0.101	0.096	0.080	0.225	0.084	0.084	0.084	0.081	0.081		0.032	0.078	0.154	0.146	0.143	0.154	0.137	0.159	0.138	0.302
12 BH	0.078	0.077	0.076	0.062	0.234	0.065	0.065	0.065	0.063	0.063	0.059		0.088	0.168	0.163	0.159	0.168	0.147	0.172	0.156	0.303
13 KF715	0.043	0.041	0.039	0.058	0.244	0.078	0.078	0.078	0.077	0.077	0.098	0.077		0.174	0.171	0.165	0.170	0.157	0.176	0.161	0.322
14 ATCC 17484	0.308	0.304	0.302	0.312	0.293	0.310	0.310	0.310	0.308	0.308	0.297	0.319	0.317		0.055	0.032	0.109	0.102	0.111	0.097	0.326
15 ATCC 17522	0.315	0.314	0.310	0.321	0.293	0.319	0.319	0.319	0.317	0.317	0.300	0.325	0.325	0.041		0.044	0.123	0.104	0.118	0.102	0.310
16 NCIMB 9816	0.299	0.293	0.293	0.304	0.289	0.299	0.299	0.299	0.297	0.297	0.286	0.310	0.308	0.023	0.037		0.101	0.092	0.107	0.093	0.319
17 IAM 12022 ^T	0.310	0.312	0.305	0.316	0.275	0.306	0.306	0.306	0.299	0.299	0.294	0.316	0.322	0.126	0.118	0.129		0.097	0.092	0.121	0.351
18 DSM 50108	0.301	0.299	0.297	0.306	0.265	0.306	0.306	0.306	0.305	0.305	0.279	0.310	0.306	0.117	0.108	0.117	0.070		0.097	0.114	0.307
19 NCIMB 9815	0.288	0.288	0.284	0.308	0.281	0.308	0.308	0.308	0.301	0.301	0.281	0.303	0.297	0.113	0.103	0.108	0.072	0.038		0.130	0.340
20 IFO 3904 ^T	0.299	0.300	0.295	0.304	0.282	0.300	0.300	0.300	0.302	0.302	0.288	0.308	0.297	0.123	0.107	0.115	0.129	0.132	0.133		0.307
21 <i>E. coli</i> K-12	0.746	0.749	0.756	0.713	0.682	0.713	0.713	0.713	0.713	0.713	0.720	0.704	0.739	0.718	0.724	0.728	0.707	0.697	0.716	0.711	

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.

Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 IFO 14164 ^T		0.002	0.002	0.007	0.018	0.011	0.011	0.006	0.010	0.009	0.012	0.011	0.002	0.022	0.022	0.025	0.030	0.035	0.034	0.028	0.174
2 ATCC 11172	0/0		0.002	0.005	0.016	0.009	0.009	0.005	0.008	0.008	0.011	0.010	0.002	0.021	0.021	0.023	0.030	0.033	0.032	0.028	0.174
3 PpG7	0/0	0/0		0.005	0.016	0.009	0.009	0.005	0.010	0.009	0.011	0.010	0.000	0.022	0.022	0.025	0.030	0.033	0.032	0.028	0.173
4 IFO 14671	0/6	0/6	0/6		0.014	0.004	0.004	0.005	0.005	0.005	0.005	0.006	0.005	0.020	0.020	0.022	0.036	0.036	0.036	0.033	0.171
5 IFO 3738	0/6	0/6	0/6	0/2		0.018	0.018	0.016	0.018	0.018	0.016	0.016	0.016	0.021	0.022	0.025	0.033	0.034	0.033	0.029	0.176
6 JCM 6156	5/6	5/6	5/6	5/0	5/2		0.000	0.005	0.001	0.002	0.002	0.002	0.009	0.020	0.020	0.022	0.036	0.037	0.036	0.034	0.169
7 ATCC 23973	5/6	5/6	5/6	5/0	5/2	0/0		0.005	0.001	0.002	0.002	0.002	0.009	0.020	0.020	0.022	0.036	0.037	0.036	0.034	0.169
8 EST1033	3/3	3/3	3/3	3/3	3/3	2/3	2/3		0.005	0.005	0.006	0.005	0.005	0.021	0.021	0.023	0.033	0.034	0.033	0.031	0.174
9 K23-1	5/6	5/6	5/6	5/0	5/2	0/0	0/0	2/3		0.001	0.002	0.003	0.010	0.019	0.019	0.021	0.036	0.038	0.037	0.033	0.169
10 PB4	5/5	5/5	5/5	5/1	5/1	0/1	0/1	2/2	0/1		0.003	0.002	0.009	0.020	0.020	0.022	0.035	0.037	0.036	0.032	0.170
11 A10L	5/6	5/6	5/6	5/0	5/2	0/0	0/0	2/3	0/0	0/1		0.001	0.011	0.021	0.021	0.024	0.037	0.037	0.036	0.032	0.171
12 BH	5/5	5/5	5/5	5/1	5/1	0/1	0/1	2/2	0/1	0/0	0/1		0.010	0.022	0.022	0.025	0.036	0.036	0.036	0.032	0.172
13 KF715	0/0	0/0	0/0	0/6	0/6	5/6	5/6	3/3	5/6	5/5	5/6	5/5		0.022	0.022	0.025	0.030	0.033	0.032	0.028	0.173
14 ATCC 17484	4/6	4/6	4/6	4/2	4/4	4/2	4/4	4/2	4/4	4/2	4/3	4/2	4/3	4/6		0.002	0.008	0.026	0.028	0.027	0.171
15 ATCC 17522	4/7	4/7	4/7	4/3	4/5	4/3	4/3	4/5	4/3	4/4	4/3	4/4	4/7	1/1		0.005	0.028	0.029	0.029	0.028	0.172
16 NCIMB 9816	4/12	4/12	4/12	4/8	4/10	4/8	4/8	4/10	4/8	4/9	4/8	4/9	4/12	1/6	0/5		0.032	0.033	0.033	0.029	0.175
17 IAM 12022 ^T	11/0	11/0	11/0	11/6	11/6	12/6	12/6	12/3	12/6	12/5	12/6	12/5	11/0	15/6	15/7	15/12		0.008	0.007	0.014	0.167
18 DSM 50108	11/5	11/5	11/5	11/8	11/8	12/8	12/8	12/5	12/8	12/7	12/8	12/7	11/5	15/8	15/9	15/14	0/5		0.002	0.014	0.173
19 NCIMB 9815	11/4	11/4	11/4	11/7	11/7	12/7	12/7	12/4	12/7	12/6	12/4	12/6	11/4	15/7	15/8	15/13	0/4	0/1		0.014	0.173
20 IFO 3904 ^T	11/5	11/5	11/5	11/10	11/10	12/10	12/10	12/7	12/10	12/9	12/10	12/9	11/5	15/10	15/11	15/12	4/5	4/5	4/4		0.169
21 <i>E. coli</i> K-12	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND

onymous 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) of 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 = 74.6%). The substitutions were most frequently observed in helix 6 in variable region V1, and in helix 18 in variable

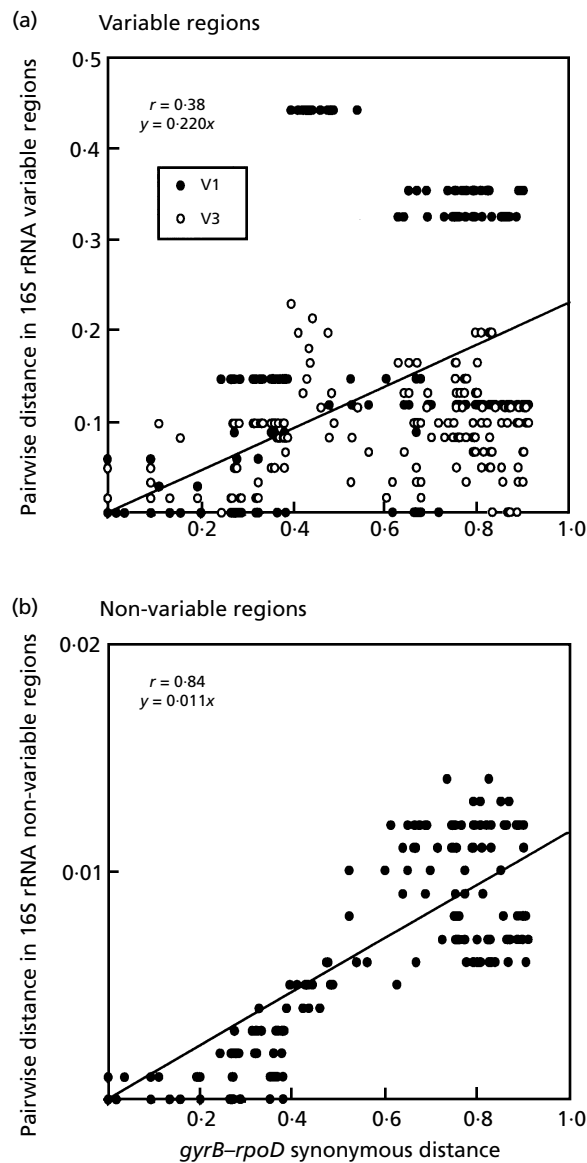


Fig. 2. Comparison between the genetic distances estimated from 16S rRNA and the synonymous distances estimated from the *gyrB* and *rpoD* genes. The synonymous distances were obtained from the combined nucleotide sequences of the *gyrB* and *rpoD* genes (total of 1956 bases) by applying the method of Nei & Gojobori (23). r indicates the coefficient of correlation. (a) Comparison between the pairwise distances in the variable regions of 16S rRNA and the synonymous distances calculated from the *gyrB* and *rpoD* genes. The nucleotide substitutions in helix 6 and helix 18, corresponding to positions 65–104 and 437–497, respectively, of *E. coli* 16S rRNA according to the numbering system of Noller (24), were counted (Table 2). (b) Comparison between the genetic distances calculated from the non-variable regions of 16S rRNA (1012 bp) and the synonymous distances calculated from the *gyrB* and *rpoD* genes.

region V3 [helix numbers conform to the numbering system of Dams *et al.* (5)]: among 20 compared sequences, seven variations were found in helix 6 and 11 variations in helix 18 in contrast to 2–5 variations in

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

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 *rpoD* 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-kynurenine (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.

Unclassified *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 14164^T (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-kynurenine 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* biovars 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 phenotypic 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 Ikuko 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

1. Barnsley, E. A. (1976). Role and regulation of the ortho and meta pathways of catechol metabolism in pseudomonads metabolizing naphthalene and salicylate. *J Bacteriol* **125**, 404–408.
2. Bchner, B. R. (1989). Sleuthing out bacterial identities. *Nature* **339**, 157–158.
3. Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc Natl Acad Sci USA* **75**, 4801–4805.
4. Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S. & Gross, C. A. (1981). The nucleotide sequence of the cloned *rpoD* gene for the RNA polymerase sigma subunit from *E. coli* K12. *Nucleic Acids Res* **9**, 2889–2903.
5. Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J. M., Smits, G., Vandembemt, I. & De Wachter, R. (1988). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res* **16** (suppl.), r87–r173.
6. Dickerson, R. E. (1971). The structures of cytochrome *c* and the rates of molecular evolution. *J Mol Evol* **1**, 26–45.
7. Dixon, M. T. & Hillis, D. M. (1993). Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol Biol Evol* **10**, 256–267.
8. Edwards, U., Rogall, T., Blocker, H., Emde, M. & Bottger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* **17**, 7843–7853.
9. Feist, C. F. & Hegeman, G. D. (1969). Regulation of the meta cleavage pathway for benzoate oxidation by *Pseudomonas putida*. *J Bacteriol* **100**, 1121–1123.
10. Felsenstein, J. (1989). PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.
11. Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool* **20**, 406–416.

12. Fox, G. E., Wisotzkey, J. D. & Jurtschuk, P. J. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.
13. Furukawa, K., Hayase, N., Taira, K. & Tomizuka, N. (1989). Molecular relationship of chromosomal genes encoding biphenyl/polychlorinated biphenyl catabolism: some soil bacteria possess a highly conserved bph operon. *J Bacteriol* **171**, 5467–5472.
14. Hancock, J. M., Tautz, D. & Dover, G. A. (1988). Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol Biol Evol* **5**, 393–414.
15. Johnson, J. L. (1981). Genetic characterization. In *Manual of Methods for General Bacteriology*, pp. 450–472. Edited by P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg & G. B. Philips. Washington, DC: American Society for Microbiology.
16. Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.
17. Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: molecular evolutionary genetics analysis, version 1.01. Pennsylvania State University, PA, USA.
18. Lawrence, J. G., Hartl, D. L. & Ochman, H. (1991). Molecular considerations in the evolution of bacteria genes. *J Mol Evol* **33**, 241–250.
19. Lonetto, M., Gribskov, M. & Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**, 3843–3849.
20. Maidak, B. L., Larsen, N., McCaughey, M. J., Overbeek, R., Olsen, G. J., Fogel, K., Blandy, J. & Woese, C. R. (1994). The ribosomal database project. *Nucleic Acid Res* **22**, 3485–3487.
21. Molin, G. & Nilsson, I. (1985). Degradation of phenol by *Pseudomonas putida* ATCC 11172 in continuous culture at different ratios of biofilm surface to culture volume. *Appl Environ Microbiol* **50**, 946–950.
22. Nagashima, H., Inoue, J., Sasaki, E., Yamamoto, S., Sasaki, Y., Yamauchi-Inomata, Y. & Harayama, S. (1996). Long-chain n-alkanol dehydrogenase from *Pseudomonas putida*. *J Ferment Bioeng* **82**, 328–333.
23. Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426.
24. Noller, H. F. (1984). Structure of ribosomal RNA. *Annu Rev Biochem* **53**, 119–162.
25. Nurk, A., Kasak, L. & Kivisaar, M. (1991). Sequence of the gene (*pheA*) encoding phenol monooxygenase from *Pseudomonas* sp. EST1001: expression in *Escherichia coli* and *Pseudomonas putida*. *Gene* **102**, 13–18.
26. Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**, 74–86.
27. Olsen, G. J., Woese, C. R. & Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**, 1–6.
28. Rousset, F., Pelandakis, M. & Solignac, M. (1991). Evolution of compensatory substitutions through G.U intermediate state in *Drosophila* rRNA. *Proc Natl Acad Sci USA* **88**, 10032–10036.
29. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
30. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
31. Schleifer, K.-H. & Ludwig, W. (1994). Molecular taxonomy: classification and identification. In *Bacterial Diversity and Systematics*, pp. 1–15. Edited by F. G. Priest, A. Ramos-Cormenzana & B. J. Tindall. New York: Plenum Press.
32. Shimao, M., Nakamura, T., Okuda, A., Abe, M. & Harayama, S. (1996). Characteristics of transposon insertion mutants of mandelic acid-utilizing *Pseudomonas putida* strain A10L. *Biosci Biotechnol Biochem* **60**, 1051–1055.
33. Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull* **28**, 1409–1438.
34. Stackebrandt, E. & Goebel, B. (1994). Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.
35. Stackebrandt, E. & Liesack, W. (1993). Nucleic acids and classification. In *Handbook of New Bacterial Systematics*, pp. 151–194. Edited by M. Goodfellow & A. G. O'Donnell. New York: Academic Press.
36. Stanier, R. Y., Palleroni, N. J. & Doudoroff, M. (1966). The aerobic pseudomonads: a taxonomic study. *J Gen Microbiol* **43**, 159–271.
37. Takeo, M., Maeda, Y., Okada, H., Miyama, K., Mori, K., Ike, M. & Fujita, M. (1995). Molecular cloning and sequencing of the phenol hydroxylase gene from *Pseudomonas putida* BH. *J Ferment Bioeng* **79**, 485–488.
38. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
39. Watt, P. M. & Hickson, I. D. (1994). Structure and function of type II DNA topoisomerases. *Biochem J* **303**, 681–695.
40. Williams, P. A. & Murray, K. (1974). Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. *J Bacteriol* **120**, 416–423.
41. Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.
42. Yamagishi, J., Yoshida, H., Yamayoshi, M. & Nakamura, S. (1986). Nalidixic acid-resistant mutations of the *gyrB* gene of *Escherichia coli*. *Mol Gen Genet* **204**, 367–373.
43. Yamamoto, S. & Harayama, S. (1995). PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**, 1104–1109.
44. Yen, K. M. & Gunsalus, I. C. (1985). Regulation of naphthalene catabolic genes of plasmid NAH7. *J Bacteriol* **162**, 1008–1013.