Molecular phylogeny based on enzoylformate decarboxylase of *Pseudomonas* strains

Choedchai Saehuan* [1], Palang Chotsiri, Theerasak Rojanarata*****

*Division of Medical Technology, Department of Science, Faculty of Science and Technology, Bansomdej Chaopraya Rajabhat University, Bangkok 10600, Thailand.

**Bioinformatics Program, School of Bioresources and Information Technology, King Monkut's University of Thonburi, Bangkok 10150, Thailand.

***Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Silpakorn University, Sanamchandra Palace Campus, Nakhon Pathom 73000, Thailand.

[1]corresponding author, E-mail address: choedchai@bansomdej.ac.th, Fax 02 473 7000 Ext. 3932

Abstract

We studied 10 protein sequences of the benzoylformate decarboxylase by the progressive alignment algorithm with CLUSTAL W in four *Pseudomonas* species: 5 *Pseudomonas aeruginosa*, 1 *Pseudomonas fluorescens*, 3*Pseudomonas putida*, and 1 *Pseudomonas stutzeri*. The phylogenetic trees developed were based on a MEGA4 package, using the neighbor-joining and minimum evolution method, and a bootstrap analysis to critically evaluate the validity of the analyses. The 10 benzoylformate decarboxylase from *Pseudomonas* strains separated into two main groups: all *Pseudomonas aeruginosa* forms with sub-grouping both *Pseudomonas fluorescens* Pf-5 and *Pseudomonas putida* ATCC 12633 allele B in conjunction with *Pseudomonas putida* ATCC 12633 allelle C: *Pseudomonas putida* ATCC 12633 allele A forms with *Pseudomonas stutzeri* ST-201. benzoylformate decarboxylase from *Pseudomonas putida* ATCC 12633 allele A and *Pseudomonas stutzeri* ST-201 coevoled in the lineage history and separated into an important role in the different metabolic pathway, mandelate pathway and novel D-phenylglycine pathway, respectively.

Introduction

Benzoylformate decarboxylase (BFDC, EC 4.1.1.7) is a thiamin pyrophosphate (TPP) - dependent enzyme that catalyzed the conversion of benzoylformate to benzaldehyde and carbon dioxide. BFDC has been increasingly important in pharmaceutical and medical technology for synthesis new chemical compounds. BFDC has been described in *Pseudomonas putida* ATCC 12633 (Gunsalus et al., 1953) with the most detailed studies by Tsou (Tsou et al.,1990) and other closely related microorganisms as part of a mandelate pathway that allow bacteria to use (*R*) - mandelate as a sole carbon source. More recently, a BFDC has been identified as a part of a novel D-phenylglycine degrading pathway in *Pseudomonas stutzeri* ST-201 (Saehuan et al., 2007).

The use of protein sequences for phylogenetic studies is advocated because they provide more information and reduce the possibility of convergent substitutions or chance alignment and problems related to substitution saturation and codon bias. Protein sequences are apparently less affected by differences in G+C content than 16S rRNA. Graur (Graur et al., 2000) and Li (Li et al., 2000) indicated that amino acids change less frequently during evolution than nucleotides and there are 20 amino acids and four nucleotides, hence the chance of two sites being identical is lower with amino acids. Also, considering the variance in codon usage among species and that our study encompasses bacteria in *Pseudomonas* strains we choose to use protein sequence as less problematic. Opinions supporting the substitution and advantage of proteins rather than nucleotides have been critically and extensively discussed by Opperdoes (Opperdoes et al., 2003) and many others.

The objective of this study was to determine the first phylogenetic relationship using benzoylformate decarboxylase protein among *Pseudomonas* strains. Although the phylogenetic tree of 16S rRNA of *Pseudomonas* group has been constructed to represent to the true evolution of *Pseudomonas* group by Yojiro (Yojiro et al., 2000), it did not include *Pseudomonas putida* ATCC 12633 and *Pseudomonas stutzeri* ST-201.

Materials and methods

BFDC sequences of all *Pseudomonas* strains and *Mycobacterium smegmatis* MC2-151 were obtained from GenBank databases (http://www.ncbi.nlm.nih.gov/). Accession numbers used in the analyses for all sequence were shown in Table 1.

Amino acid sequences were aligned using the progressive alignment algorithm (Feng et al., 1987), as implemented in CLUSTAL W (Thompson et al., 1994) based on a BLOSUM62 scoring matrix and rearranged further by visual inspection. Alignments regions with uncertain homology were discarded from the analysis. A

representative multiple alignment of BFDC sequences was also analyzed and constructed the phylogenetic trees using MEGA4 package (Tamura et al., 2007) based on the neighbor-joining (NJ) method (Saitou et al., 1987), and the minimum evolution (ME) method (Rzhetsky et al., 1992). Subsequent analysis for robustness of the resulting NJ and ME trees was carried out using bootstrap analysis (1,000 samplings from sequence data). Finally, the molecular phylogenies were edited and displayed by using MEGA4 package.

Results and discussion

1. A Multiple alignment of benzoylformate decarboxylase sequences

A representative multiple alignment of BFDC sequences is shown in Figure 1**.** There are three conserved domain of BFDC: N-terminal TPP binding domain (TPP_enzyme_N) shown in blue line, central TPP binding domain (TPP_enzyme_M) containing a 2-fold Rossman fold shown in pink line, and TPP-binding module (TPP_BFDC) composing of proteins similar to *Pseudomonas putida* ATCC 12633 benzoylformate decarboxylase shown in violet line.

Table1. Sequences used in this study and abbreviations of BFDC sequences shown in Figure 1-3.

Sequence	Accession No.	Bacterial strain	Abbreviation
PAE21	EAZ61646	*Pseudomonas aeruginosa* 2192	*P. aeruginosa* 2192
PAEC3	EAZ55823	*Pseudomonas aeruginosa* C3719	*P. aeruginosa* C3719
PAEPA	YP001350945	*Pseudomonas aeruginosa* PA7	*P. aeruginosa* PA7
PAEPO	NP253588	*Pseudomonas aeruginosa* PAO1	*P. aeruginosa* PAO1
PAEUC	YP793369	*Pseudomonas aeruginosa* UCBPP-PA142	*P. aeruginosa* UCBPP-PA142
PFLPF	YP260581	*Pseudomonas fluorescens* Pf-5	*P. fluorescens* Pf-5
PPUAA	AAC15502	*Pseudomonas putida* ATCC 12633 allele A[a]	*P. putida* A
PPUAB	CAK95976	*Pseudomonas putida* ATCC 12633 allele B[b]	*P. putida* B
PPUAC	CAK95977	*Pseudomonas putida* ATCC 12633 allele C[c]	*P. putida* C
PSTST	ABN80423	*Pseudomonas stutzeri* ST-201	*P. stutzeri*
MSMMC	YP885985	*Mycobacterium smegmatis* MC2-155	*M. smegmatis* MC2-155

[a] BFDC from a chromosome of *P. putida* ATCC 12633 (BfdA) (Tsou et al.,1990)

[b] BFDC from a chromosomal library of *P. putida* ATCC 12633(BfdB) (Henning et al., 2006)

[c] BFDC from a chromosomal library of *P.putida* ATCC 12633(BfdC) (Henning et al., 2006)

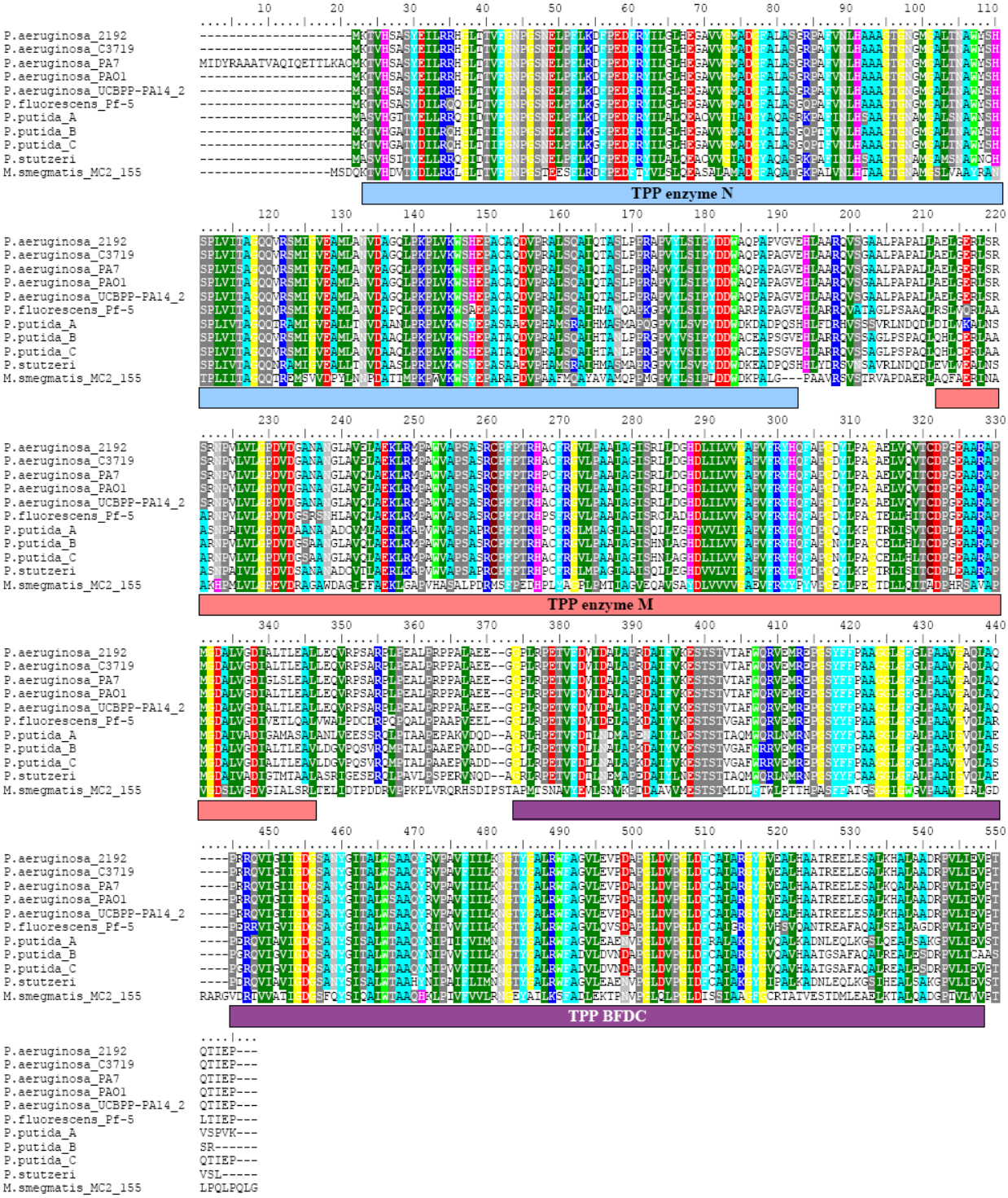

Figure 1. A Multiple alignments of 11 BFDC amino acid sequences based on Clustal W. The shading in the multiple alignment denotes the degree of match or mismatch. There are three conserved domain of BFDC: N-terminal TPP binding domain (TPP_enzyme_N) shown in blue line, central TPP binding domain (TPP_enzyme_M) shown in pink line, and TPP-binding module (TPP_BFDC) shown in violet line. Abbreviations of BFDC sequences were described in Table 1.

2. Phylogenetic analysis of benzoylformate decarboxylase protein among *Pseudomonas* strains

Phylogenies of BFDC based on distinct method by the neighbor-joining method and the minimum evolution method shown in Figure 2 and Figure 3, respectively reveal the parallel results with BFDC from *M. smegmatis* MC2-155 used as out group. BFDC from *Pseudomonas* strain*s* evolved from common ancestor and separated into two main groups. The first main group consists of BFDCs from all *Pseudomonas aeruginosa*, *Pseudomonas fluorescens* Pf-5 and *Pseudomonas putida* ATCC 12633 allele B (PPUAB) joined with *Pseudomonas putida* ATCC 12633 allele C (PPUAC). The second main group includes of BFDC from *Pseudomonas putida* ATCC 12633 allele A (PPUAA) and *Pseudomonas stutzeri* ST-201 (PSTST). PSTST is closer PPUAA than PPUAB and PPUAC while PPUAB and PPUAC, paralogs with PPUAA, are close to BFDC from *Pseudomonas fluorescens* Pf-5.

Phylogenetic analysis constructed by using single protein with benzoylformate decarboxylase might prove to be a useful tool for studying evolution in accordance with Cavalier (Cavalier 2001). The evolution of, e.g., BFDC or benzoylformate decarboxylas gene (*bfdc*) as revealed by single gene phylogenetic analyses may not only be related to the evolution of the genome it inhabits, but may also be related to the evolution of the metabolic process or pathway it belongs to and codes for, e.g., mandelate pathway and/or D-phenylglycine pathway according to the previous work proposed by Xiong with protein or gene of photosynthetic pathway (Xiong et al., 2000). We suggest that *bfdc* gene coding for benzoylformate decarboxylase in *Pseudomonas* group may be acquisition via horizontal gene transfer from other bacterial groups but the issue of ancestral benzoylformate decarboxylase gene or protein will be some hope of resolution in the future.

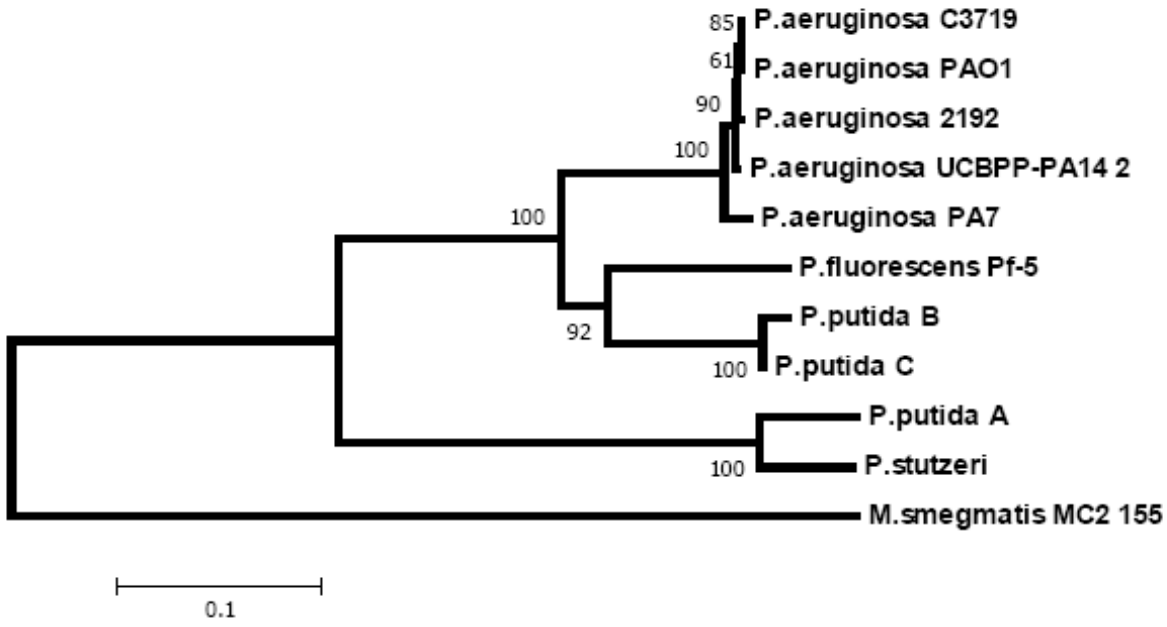

Figure 2. The evolutionary history was inferred using the neighbor-joining method. The optimal tree with the sum of branch length = 1.28015239 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the poisson correction method (Zuckerkandl, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 522 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. Abbreviations of BFDC sequences were described in Table 1.

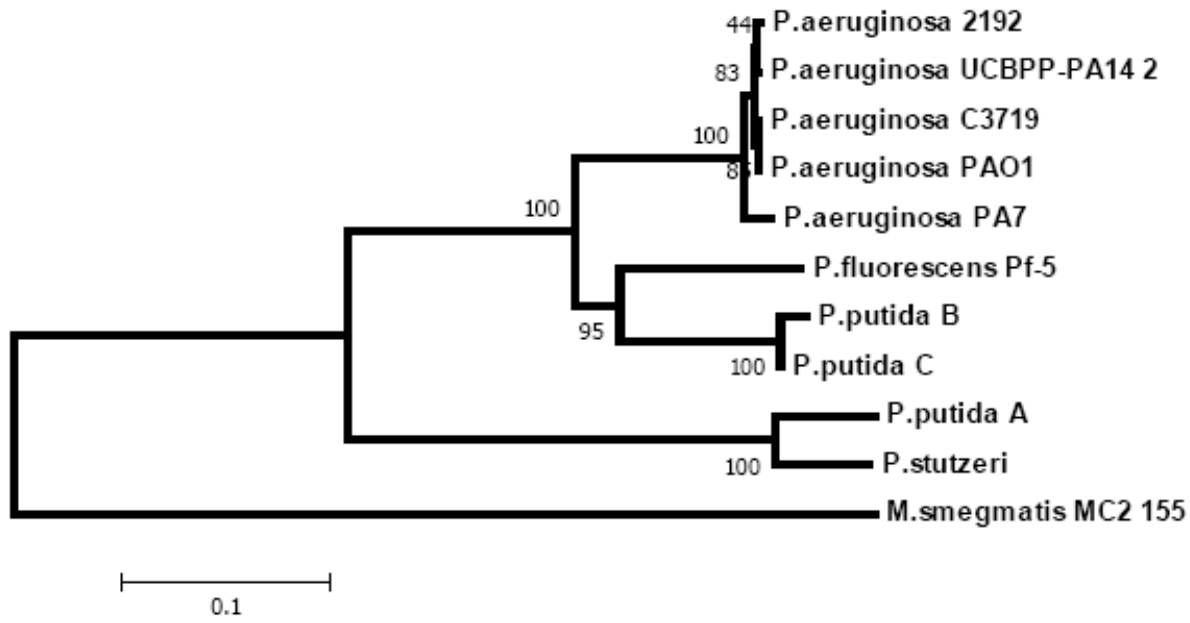

Figure 3. The evolutionary history was inferred using the minimum evolution method. The optimal tree with the sum of branch length = 1.28143541 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the close-neighbor-interchange algorithm (Nei et al., 2000) at a search level of 1. The neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 522 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. Abbreviations of BFDC sequences were described in Table 1.

Conclusion

This study has addressed the first report about BFDC's evolutionary history among *Pseudomonas* strains. The 10 benzoylformate decarboxylase from *Pseudomonas* strains separated into two major groups: the first group, one sub-grouping of all *Pseudomonas aeruginosa* forms with another sub-grouping both *Pseudomonas fluorescens* Pf-5 and *Pseudomonas putida* ATCC 12633 allele B together with *Pseudomonas putida* ATCC 12633 allelle C: the second group, *Pseudomonas putida* ATCC 12633 allele A forms with *Pseudomonas stutzeri* ST-201. Benzoylformate decarboxylase from *Pseudomonas putida* allele A and *Pseudomonas stutzeri* ST-201 coevolved along lineage history and separated into a major role in the different degradation pathway, previous reported mandelate pathway (Tsou et al.,

1990) and novel D-phenylglycine pathway (Saehuan et al., 2007), respectively. The evolutional origin of benzoylformate decarboxylase protein or benzoylformate decarboxylase gene spreaded throughout *Pseudomonas* strains via horizontal gene transfer will be hopefully elucidated in the future.

Reference

Cavalier-S. T. (2001). Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. **J. Mol. Evol**. 53(4-5): 555–595.

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. **Evolution** 39(4):783-791.

Feng D.-F., Doolittle R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. **J. Mol. Evol**. 25:351-360.

Graur D., Li W.-H. (2000). **Fundamentals of Molecular Evolution** (2nd ed). Massachusetts: Sinauer Association.

Gunsalus C.F., Stanier R.Y., Gunsalus I.C. (1953). The enzymatic conversion of mandelic acid to benzoic acid. III. Fractionation and properties of the soluble enzymes. **J. Bacteriol**. 66:548–553.

Henning H., Leggewie C., Pohl M., Muller M., Eggert T. and Jaeger K.E. (2006) Identification of novel benzoylformate decarboxylases by growth selection. **Appl. Environ. Microbiol.** 72(12): 7510-7517.

Li S., Pearl D.K., Doss H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. **J. Am. Statistical Assoc** 95: 493–508.

Nei M., Kumar S. (2000). **Molecular Evolution and Phylogenetics.** New York: Oxford University Press.

Opperdoes F.R. (2003). **The Phylogenetic Handbook. A Practical Approach to DNA and Protein Phylogeny.** Cambridge: Cambridge University Press.

Rzhetsky A., Nei M. (1992). A simple method for estimating and testing minimum evolution trees. **Mol. Biol. Evol.** 9:945-967.

Saehuan C., Rojanarata T., Wiyakrutta S., McLeish M.J., Meevootisom V. (2007). Isolation and characterization of a benzoylformate decarboxylase and a NAD^+/$NADP^+$-dependent benzaldehyde dehydrogenase involved in D-phenylglycine metabolism in *Pseudomonas stutzeri* ST-201. **Biochim. Biophys. Acta** 1770:1585–1592.

Saitou N., Nei M. (1987) .The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.** 4(4):406-425.

Tamura K, Dudley J., Nei M., Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol. Biol. Evol.** 24:1596-1599.

Thompson J.D., Higgins D.G., Gibson T.J. (1994). Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. **Nucl. Acids Res.** 22:4673-4680

Tsou A. Y, Ransom S. C, Gerlt J. A, Buechter D. D, Babbitt P. C, Kenyon G. L. (1990). Mandelate pathway of *Pseudomonas putida*: Sequence relationships involving mandelate racemase, (*S*)-mandelate dehydrogenase, benzoylformate decarboxylase and expression of benzoylformate decar boxylase in *Escherichia coli*. **Biochem.** 29:9856-9862.

Xiong J., Fischer W. M., Inoue K., Nakahara M., Bauer C.E. (2000). Molecular evidence for the early evolution of photosynthesis. **Science** 289: 1724–1730.

Yojiro A, Hongik K., Ju-Y. P., Hisatsugu W., Hiroshi O. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence**. Int. J. Syst. Evol. Microbiol.** 50: 1563-1589.

Zuckerkandl E., Pauling L. (1965). **Evolving Genes and Proteins.** New York: Academic Press.