Farnesyl diphosphate synthase gene of three phototrophic bacteria and its use as a phylogenetic marker

International Center for
Biotechnology, Osaka
University, 2-1 Yamada-oka,
Suita-shi, Osaka 565-0871,
Japan

Jose Jason L. Cantera, Hiroko Kawasaki and Tatsuji Seki

Author for correspondence: Hiroko Kawasaki. Tel: +81 6 6879 7455. Fax: +81 6 6879 7454.
e-mail: kawasaki@icb.osaka-u.ac.jp

Farnesyl diphosphate (FPP) synthase is essential not only for phototrophic bacteria in carotenoid biosynthesis, but also for non-phototrophic bacteria in the biosynthesis of physiologically important compounds. The gene encoding FPP synthase was assessed as a molecular marker to investigate the intermingled relationship between the phototropic and non-phototropic bacteria in the α -Proteobacteria based on 16S rRNA analysis. The FPP synthase amino acid sequences from three phototrophic bacteria, *Rhodobacter sphaeroides* ATCC 11167^T, *Rhodobacter capsulatus* ATCC 11166^T and *Rhodovulum sulfidophilum* W4^T, were determined and used in conjunction with sequences of other representative members of the α -, γ - and ϵ -Proteobacteria and the low-G+C Gram-positive bacteria for phylogenetic analyses by the neighbour-joining and maximum-likelihood methods. The overall topology of the FPP synthase gene tree is consistent with that of the 16S rRNA tree, producing a distinct cluster of the three phototrophic bacteria. A minor discordance between the two trees was observed in the cluster of the non-phototrophic *Bradyrhizobium japonicum* USDA 110 and *Mesorhizobium loti* MAFF 303099; the FPP synthase genes of these two rhizobial species are highly homologous as compared with their respective 16S rRNA. The results suggest that the FPP synthase and 16S rRNA genes have the same evolutionary pattern, evolving vertically from each common ancestral gene; the FPP synthase gene, therefore, could possibly be used for further study on the molecular systematics of photosynthetic bacteria.

Keywords: farnesyl diphosphate synthase, phylogeny, phototrophic bacteria, non-phototrophic bacteria, gene sequence

INTRODUCTION

Sequence analysis of 16S rRNA has revealed that phototrophic bacteria do not constitute a phylogenetically coherent group and are instead widely dispersed in five major phyla or divisions. The largest number of phototrophs is found in the phylum *Proteobacteria* (Stackebrandt *et al.*, 1996), wherein

they intermingle with non-phototrophic species (Kawasaki *et al.*, 1993). The intermingling of the phototrophic and non-phototrophic bacteria, as shown by 16S rRNA phylogenetic analysis, raises questions on the evolutionary history of their phototrophic character, since the origin of this character cannot be accurately inferred. Other sequences such as those of photosynthetic genes (Xiong *et al.*, 1998; Nagashima *et al.*, 1997) and other isoprenyl diphosphate synthases (Chen *et al.*, 1994), among others, have been used as tools to elucidate the evolution of photosynthesis. To date, however, these results provide little information on the diversity of the phototrophic bacteria in the *Proteobacteria* and, to our knowledge, no report has studied their relationship with their non-phototrophic relatives.

Published online ahead of print on 22 March 2002 as DOI 10.1099/ijs.0.02148-0.

Abbreviations: FARM, first aspartate-rich motif; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate.

The DDBJ accession numbers for the FPP synthase sequences reported in this paper are AB028046, AB028044 and AB028047.

Fig. 1. Biosynthesis of isoprenoid compounds involves the consecutive condensations of IPP by prenyl transferases. In the initial steps of the biosynthetic pathway, FPP synthase catalyses the sequential 1'-4 condensation of the 5-carbon isopentenyl diphosphate (IPP) with dimethylallyl diphosphate (DMAPP) to form the 10-carbon geranyl diphosphate (GPP); GPP is then condensed with another IPP molecule to form the 15-carbon FPP as the ultimate product. FPP is the common substrate for the synthesis of physiologically important compounds, such as sesquiterpenoids, the haem group of cytochromes, prenylated proteins, the prenyl side chain of ubiquinones and dolichols (Ogura *et al.*, 1997; Attucci *et al.*, 1995). Further condensation of FPP with IPP, catalysed by GGPP synthase, produces GGPP, which is involved in the production of other important compounds such as prenylated proteins in eukaryotes and ether-linked lipids in *Archaea* (Ohnuma *et al.*, 1994). GGPP is also the precursor of carotenoids in plants and in phototrophic and non-phototrophic, pigmented bacteria.

Farnesyl diphosphate (FPP) synthase (EC 2.5.1.10/EC 2.5.1.1), a type of isoprenyl diphosphate synthase, plays a key role in the biosynthesis of physiologically important isoprenoid compounds (Koyama *et al.*, 1993; Ogura *et al.*, 1997; Anderson *et al.*, 1989) (Fig. 1), including carotenoids in plants and in pigmented phototrophic and non-phototrophic bacteria. For phototrophic bacteria, carotenoids are essential components of the photosynthetic apparatus (Lang *et al.*, 1995; Armstrong *et al.*, 1993). As such, FPP synthase is universally present, an indispensable molecule for cell survival, and plays a crucial role not only in phototrophic bacteria but also in non-phototrophic bacteria. Thus, the FPP synthase gene has been chosen as a marker for investigating evolutionary relationships between phototrophic and non-phototrophic bacteria. However, no FPP synthase gene sequences from phototrophic bacteria in the α -*Proteobacteria* are available for systematic studies.

FPP synthase gene sequences from three phototrophic bacteria belonging to the *Proteobacteria* have been determined and characterized in this report and this gene has been introduced as a marker to infer the evolutionary history of phototrophic and non-phototrophic bacteria. Phylogeny of the FPP synthase

gene, in comparison with that of 16S rRNA, is presented.

METHODS

Bacterial strains and DNA preparation. *Rhodobacter capsulatus* ATCC 11166^T, *Rhodobacter sphaeroides* ATCC 11167^T and *Rhodovulum sulfidophilum* W4^T were cultivated as previously described (Kawasaki *et al.*, 1993). *Escherichia coli* strains for recombinant DNA manipulations were grown overnight in Luria-Bertani medium at 37 °C. Chromosomal DNA from the phototrophic strains was prepared by the method of Ausubel *et al.* (1995).

Primer design and PCR amplification. Table 1 lists the degenerate oligonucleotide primers that were designed based on well-conserved regions of FPP synthase sequences available in GenBank. Primers F-B3 and R-D2 were derived from conserved motifs in FPP synthases from *E. coli* (accession no. AB028046), *Bacillus stearothermophilus* (D13293), *Bacillus subtilis* (D84432), *Haemophilus influenzae* (U32822), *Bradyrhizobium japonicum* (U12678) and *Sinorhizobium* sp. NGR234 (AE000082, formerly *Rhizobium* sp. NGR234). These primers were used to amplify the partial FPP synthase of *Rhodobacter sphaeroides* ATCC 11167^T and *Rhodobacter capsulatus* ATCC 11166^T. α -F1 and α -R1 were designed from the sequences of *Rhodobacter sphaeroides* ATCC 11167^T and *Rhodobacter capsulatus* ATCC 11166^T and used to amplify the partial FPP synthase gene of *Rhodovulum sulfidophilum* W4^T. The primers used for sequencing the entire FPP synthase gene were derived from the exact FPP synthase nucleotide sequence of each strain and designed to amplify overlapping products to minimize errors attributed to sequencing of the entire FPP synthase.

The partial FPP synthase gene was amplified by PCR using 10–100 ng total DNA, 25 pmol each primer and 1.25 U Ex Taq polymerase (TaKaRa Shuzo) in a 25 μ l reaction mixture. The cycling conditions used were 30 cycles at 94 °C for 60 s, 55 °C for 60 s and 72 °C for 60 s. PCR reactions were carried out in a GeneAmp PCR system 9700 (PE Applied Biosystems).

Cloning and DNA sequencing. PCR products were cloned in pCR2.1 of the TA cloning kit (Invitrogen) according to the manufacturer's protocol. Either pUC18 or pUC19 was used for other routine DNA cloning procedures (Yanisch-Perron *et al.*, 1985) using the DNA ligation kit version 1 (TaKaRa Shuzo). Plasmid DNAs were purified using the QIAprep Spin Miniprep kit (QIAGEN) according to the manufacturer's protocol prior to cycle sequencing reactions. DNA sequences were determined by the dideoxy chain-termination method (Sanger *et al.*, 1977) with a BigDye Terminator cycle sequencing kit (PE Applied Biosystems) using an ABI PRISM 310 genetic analyser (Perkin Elmer). Sequence data were analysed by the ABI PRISM sequence analysis program and assembled using the ABI Auto Assembler (Perkin Elmer).

Southern and colony hybridizations. Southern and colony hybridizations were done as described by Sambrook *et al.* (1989): 2 μ g chromosomal DNA was digested with appropriate restriction enzymes (TaKaRa Shuzo), fractionated by electrophoresis on 0.8% agarose and transferred to Hybond-N+ membranes (Amersham Pharmacia Biotech) by capillary transfer. Hybridization probes from the PCR product were labelled using the DIG labelling kit (Roche Diagnostics) and hybridization signals were detected using

Table 1. Primer sequences used for PCR and sequencing the entire FPP synthase gene from the three phototrophic bacteria

Primer	Sequence (5' → 3')	Location*
PCR primers		
F-B3	CAYGAYGAYYTXCCXGXCATGGA	736–758
R-D2	RTCIARDATRTCRTC	1170–1156
α -F1	CAYGACGAYMTGCCCKSVATGGA	736–758
α -R1	CYKYSRCRTCSAGAAATGTCGTC	1177–1156
Sequencing primers		
Rs-1f	CCTGCATGGACGATGAC	748–766
Rs-1r	CGCAGGTCGTCATCGTC	773–758
Rs-2r	TGCATTCTGTCCTTCTG	814–797
Rs-4f	AAGACTGGCGCGCTGATCCG	1039–1061
Rs-5f	ACGGCCCTCGGCCTTG	1126–1142
Rs-5r	CATCGAGAATGTCGTC	1171–1156
Rcap-5f	GACGACATTCTGGATG	1156–1171
Rcap-2r	TCCCAGCACCGGATCGG	884–867
W4-1f	GTGCATGTGAAATGGG	790–805
W4-1r	ACCAGCGGATCAGCGC	1051–1066

* Positions are numbered relative to *E. coli ispA* (accession no. AB028046).

Table 2. Bacterial strains and GenBank/DDBJ/EMBL accession/protein ID numbers of FPP synthase (FPPS), GGPP synthase (GGPPS) and 16S rRNA sequences used in this study

Species	Phylogenetic group	Strain designation*	Accession/protein ID no.	
			FPPS/GGPPS	16S rRNA
<i>Rhodobacter capsulatus</i>	α -3 <i>Proteobacteria</i>	ATCC 11166 ^T	AB028046†	D13474
<i>Rhodobacter sphaeroides</i>	α -3 <i>Proteobacteria</i>	ATCC 11167 ^T = IFO 12203 ^T	AB028044†	D16425
<i>Rhodovulum sulfidophilum</i>	α -3 <i>Proteobacteria</i>	W4 ^T	AB028047†	D13475
<i>Bradyrhizobium japonicum</i>	α -2 <i>Proteobacteria</i>	USDA 110	U12678	D13430
<i>Mesorhizobium loti</i>	α -2 <i>Proteobacteria</i>	MAFF 303099	BAB52678	AP003001
<i>Campylobacter jejuni</i>	ϵ - <i>Proteobacteria</i>	NCTC 11168	CAB73632	AL139076
<i>Helicobacter pylori</i>	ϵ - <i>Proteobacteria</i>	26695	AE000602	AE000620
<i>Escherichia coli</i>	γ - <i>Proteobacteria</i>	K-12 ^T	D00694	AE000129
<i>Haemophilus influenzae</i>	γ - <i>Proteobacteria</i>	Rd	U32822	U32847
<i>Bacillus halodurans</i>	Low-G + C Gram-positive	C-125	BAB06500	AB013373
<i>Bacillus subtilis</i>	Low-G + C Gram-positive	168	CAB14359	Z99104
' <i>Micrococcus luteus</i> '	Low-G + C Gram-positive	B-P 26	AB003187	AB079788†
' <i>Aquifex aeolicus</i> '	<i>Aquificales</i>	VF-5	AE000705	AJ309733
<i>Thermotoga maritima</i>	<i>Thermotogales</i>	MSB8 ^T	G72410	AE001703
<i>Methanobacterium thermoautotrophicum</i>	Euryarchaeota	Δ H	AAB84557	AE000930
<i>Methanococcus jannaschii</i>	Euryarchaeota	DSM 2661 ^T	AAB98865	M59126
' <i>Pyrococcus abyssi</i> '	Euryarchaeota	Orsay	CAB49828	AJ248283
<i>Pyrococcus horikoshii</i>	Euryarchaeota	OT3	AP000004	D45214
<i>Archaeoglobus fulgidus</i>	Crenarchaeota	VC-16 ^T	AF120272	Y00275
<i>Sulfolobus solfataricus</i>	Crenarchaeota	P2	P95999	X03235

* The following abbreviations are used: ATCC, American Type Culture Collection, Manassas, VA, USA; USDA, United States Department of Agriculture, Washington DC, USA; DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany; MAFF, Ministry of Agriculture, Forestry and Fisheries, Ibaragi, Japan; IFO, Institute for Fermentation, Osaka, Japan; NCTC, National Collection of Type Cultures, Central Public Health Laboratory, London, UK.

† Determined in this study.

the DIG luminescent detection kit (Roche Diagnostics) according to the manufacturer's instructions.

Sequence alignments and phylogenetic trees. The FPP synthase nucleotide sequences were analysed with GENETIX-WIN (version 3.1). Database searches were performed using BLAST (Altschul *et al.*, 1997). Multiple sequence alignments of the translated nucleotide sequences with other bacterial FPP synthase and archaeal geranylgeranyl diphosphate (GGPP) synthase amino acid sequences available in the National Center for Biotechnology Information (NCBI) databases (<http://www.ncbi.nlm.nih.gov/>) were carried out using the program CLUSTAL X version 1.81 (Jeanmougin *et al.*, 1998; Thompson *et al.*, 1997). Gap-containing regions were excluded from the analyses. A total of 1217 nucleotide and 214 amino acid positions, excluding the gap positions, were used in the 16S rDNA and FPP synthase analyses, respectively. Phylogenetic trees were generated by the neighbour-joining method with CLUSTAL X (with 100 bootstrap replicates using default parameters) and by the maximum-likelihood method (100 replicates using the method of Jones, Taylor and Thornton) with MOLPHY version 2.3 (Adachi & Hasegawa, 1996). NJPLOT (Perrière & Gouy, 1996) and TREEVIEW (Page, 1996) were used to analyse the phylogenetic relationships of FPP synthase genes from various species. Table 2 shows the accession numbers of the FPP synthase genes and 16S rRNA sequences that were used in the alignment.

RESULTS AND DISCUSSION

Sequence analyses of the FPP synthase gene

Multiple sequence alignment with *ispA*, the gene encoding the *E. coli* FPP synthase, showed that the start codon in each of the FPP synthase genes from the three phototrophic strains is almost at the same position as in the *ispA* of *E. coli* at position 484 (Fujisaki *et al.*, 1990). The ORF of the 1373 bp *EcoRI* fragment containing the FPP synthase gene of *Rhodobacter capsulatus* ATCC 11166^T begins with an ATG codon at position 484 (*E. coli ispA* numbering system) and terminates with a TGA codon at position 1395. The ORF contains 863 nt that encode a protein with a predicted molecular mass of 29 793 Da. In *Rhodobacter sphaeroides* ATCC 11167^T, the ORF of the *BamHI* fragment containing the FPP synthase gene begins with a less commonly used GTG codon at position 481 and terminates with a TGA codon at position 1396. The GTG codon at that position was assigned as the translation initiation site because of its location, i.e. close to the possible ribosome-binding site, GGA, at position 475. The ORF encodes a protein that is one amino acid longer than that of *Rhodobacter capsulatus* ATCC 11166^T, with a molecular mass of 30 214 Da. In *Rhodovulum sulfidophilum* W4^T, the first ATG is located at position 488 and the ORF is composed of 289 amino acids, encoding a protein with predicted molecular mass of 30 245 Da. These molecular masses are in good agreement with that of FPP synthases from *E. coli* (Fujisaki *et al.*, 1990) and *Bacillus stearothermophilus* (Koyama *et al.*, 1993) (32 000 Da). All three FPP synthases from the phototrophic organisms have alanine as the major amino acid (21–25% of the

total amino acid content) encoded either by the GCC or GCG codons, as compared with the *E. coli* FPP synthase which has only 14% Ala encoded by the GCC or GCA codons. In *E. coli*, Gly is encoded by the GGT codon, whereas in the three phototrophic bacteria it is encoded by the GGC codon. The three ORFs have a G + C content ranging from 67.7 to 69.9 mol %, which is in the range of the G + C content of *Rhodobacter* and *Rhodovulum* species.

The amino acid sequences deduced from these nucleotide sequences have a remarkable similarity to known GGPP synthases from bacteria and plants (Stanley Fernandez *et al.*, 2000). To differentiate our sequences from that of GGPP synthase, sequences of two GGPP synthase genes from *Rhodobacter capsulatus* (accession no. X52291) and *Rhodobacter sphaeroides* (X82458) were included in the alignment to examine for the presence of conserved functional domains (Koyama *et al.*, 1993). Seven well-conserved regions (regions I to VII; Fig. 2) found in FPP and GGPP synthases (Ogura *et al.*, 1997) were also present in our sequences. One of these regions, the first aspartate-rich motif (FARM), which determines the product specificity of the molecule (Ohnuma *et al.*, 1997), was examined to differentiate the FPP synthases from GGPP synthases. Our sequences have a common pattern of aromatic amino acid substitution (tyrosine) in the fourth position upstream from the FARM instead of alanine as in the GGPP synthase of *Rhodobacter capsulatus* ATCC 11166^T and *Rhodobacter sphaeroides* ATCC 11167^T, as well as proline and cysteine insertions, which are characteristic of bacterial FPP synthases (Ohnuma *et al.*, 1996, 1997; Wang & Ohnuma, 1999).

Table 3 shows the overall percentages of amino acid similarity of the bacterial FPP synthase and other FPP synthase sequences retrieved from the database, revealing a significant overall homology. The amino acid sequence similarity between *Rhodobacter capsulatus* ATCC 11166^T, *Rhodobacter sphaeroides* ATCC 11167^T and *Rhodovulum sulfidophilum* W4^T ranged from 66.0 to 78.3%. The FPP synthase of *Rhodovulum sulfidophilum* W4^T is more similar to the FPP synthase of *Rhodobacter sphaeroides* ATCC 11167^T than to that of *Rhodobacter capsulatus* ATCC 11166^T. The three FPP synthases from the phototrophic strains exhibited 37.7–47.1% similarity to those from *Bradyrhizobium japonicum* USDA 110, *Mesorhizobium loti* MAFF 303099, *E. coli* K-12^T, *Haemophilus influenzae* Rd, *Bacillus subtilis* 168, *Bacillus halodurans* C-125, 'Micrococcus luteus' B-P 26, *Helicobacter pylori* 26695 and *Campylobacter jejuni* NCTC 11168.

The identity of our sequences as the FPP synthase gene is therefore firmly established by the similarity with other bacterial FPP synthase genes, the presence of conserved motifs in the amino acid sequence, and the amino acids before and in the FARM. These findings further showed that the gene is amenable for ampli-

Fig. 2. Alignment of FPP synthase amino acid sequences of *Rhodobacter capsulatus* ATCC 11166^T (designated *fps* Rcapsu), *Rhodobacter sphaeroides* ATCC 11167^T (*fps* RspHae) and *Rhodovulum sulfidophilum* W4^T (*fps* Rsulfi) with FPP synthase of *E. coli* (*fps* Ecoli), and GGPP synthase amino acid sequences of *Rhodobacter capsulatus* (ggpps Rcap) and *Rhodobacter sphaeroides* (ggpps RspHae). ' ', Positions which have a conserved residue; ' ', positions in which a 'strong' group is conserved; ' ', positions in which a 'weak' group is conserved. Amino acids are given as their one-letter code. Roman numerals I–VII above the bold lines indicate conserved regions, with conserved sequences in bold. Numbering starts from the start codon of *E. coli ispA*.

Table 3. FPP synthase amino acid (lower triangle) and 16S rDNA nucleotide sequence (upper triangle) similarities (%) of both phototrophic and non-phototrophic bacterial strains

Strain	Amino acid and nucleotide sequence similarity (%)											
	1	2	3	4	5	6	7	8	9	10	11	12
1. <i>Rhodobacter capsulatus</i>		95.6	93.8	89.1	87.0	79.0	78.7	75.7	77.3	78.9	79.6	79.3
2. <i>Rhodobacter sphaeroides</i>	72.1		93.3	89.8	88.1	80.4	80.0	76.1	77.5	79.9	80.0	80.0
3. <i>Rhodovulum sulfidophilum</i>	66.0	78.3		90.6	87.1	78.7	79.5	77.0	78.1	79.8	79.3	79.1
4. <i>Mesorhizobium loti</i>	44.3	45.9	47.1		90.0	81.7	80.9	77.2	77.4	81.3	81.9	81.5
5. <i>Bradyrhizobium japonicum</i>	43.9	45.5	47.1	92.6		79.9	79.1	77.6	78.5	80.1	80.3	79.7
6. <i>Escherichia coli</i>	45.9	44.7	43.9	46.3	45.5		89.3	78.7	76.7	79.3	79.5	79.3
7. <i>Haemophilus influenzae</i>	44.3	42.2	40.6	42.2	42.2	56.1		76.6	76.6	78.2	78.1	78.0
8. <i>Campylobacter jejuni</i>	39.3	37.7	39.3	36.5	37.3	38.1	40.6		85.7	75.9	75.9	77.2
9. <i>Helicobacter pylori</i>	41.0	42.6	41.8	34.0	33.2	35.2	40.2	51.2		76.9	76.8	76.4
10. <i>Bacillus halodurans</i>	44.3	41.4	41.8	36.9	36.9	43.0	42.6	39.8	39.3		95.4	93.2
11. <i>Bacillus subtilis</i>	44.7	41.8	41.0	36.9	36.5	42.2	40.6	37.3	39.8	59.8		93.9
12. 'Micrococcus luteus' B-P 26	42.2	38.5	37.7	36.1	35.7	42.6	45.9	44.7	39.8	48.8	50.0	

fication and sequencing. This is important in systematic studies since the molecule of choice must be available for a number of species and the gene must be easily cloned and sequenced in cases when the sequence is not available. The absolute conservation of functionally important protein sequences and the presence of the ribosome-binding sites suggest that the characterized genes are transcribed and translated and function as FPP synthase. Southern blot analysis (data

not shown) have shown that the genes encoding FPP synthase in the *Rhodobacter* and *Rhodovulum* genomes exists in one copy number per cell.

Phylogenetic analysis

Fig. 3 (left) shows the dendrogram derived from the FPP synthase amino acid sequences. Species from the α -, γ - and ϵ -*Proteobacteria*, as well as some members of

Fig. 3. Comparison of the FPP synthase amino acid sequence (left) and 16S rRNA (right) gene trees showing the phylogenetic relationships between *Rhodobacter capsulatus* ATCC 11166^T, *Rhodobacter sphaeroides* ATCC 11167^T, *Rhodovulum sulfidophilum* W4^T and other representative members of the α -, ε - and γ -Proteobacteria as well as low-G+C Gram-positive bacteria and those belonging to *Aquificales* and *Thermotogales*. Archaeal GGPP synthases were used as the outgroup. Phylogenetic trees were constructed using the neighbour-joining method; bold lines indicate branches supported by high bootstrap values (> 90%). Branch points with closed circles in the FPP synthase gene tree indicate > 90% bootstrap support using the maximum-likelihood method. The lengths of the horizontal lines are proportional to evolutionary distances; lengths of the vertical lines are meaningless. Strains under study are in bold; boxes indicate similar phylogenetic groupings based on 16S rRNA sequences. The positions of *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum* were interchanged in the FPP synthase tree as indicated by the arrows.

the Gram-positive bacteria, are represented in this tree. Seven archaeal GGPP synthases were included in the alignment and used as the outgroup due to the remarkable similarity of FPP synthase with GGPP synthase; the archaeal GGPP synthases are more ancient than the FPP synthases and other prenyl transferases and therefore, could predate the divergence of bacterial FPP synthases (Chen *et al.*, 1994). Fig. 3 (right) shows the 16S rDNA phylogenetic tree of exactly the same species set included in the FPP synthase tree, allowing for a more direct comparison between the two molecular markers. Such a comparison is useful since differences in branching patterns between trees of different molecules can help identify genetic diversity and unusual evolutionary processes or inaccuracy in one or both trees (Eisen, 1995).

The overall topology of the FPP synthase dendrogram is, in principle, consistent with the topology of the 16S rRNA phylogenetic tree, further suggesting that the FPP synthase gene has a similar evolutionary pattern as the 16S rRNA. For instance, the groupings of the ε -Proteobacteria consisting of *Helicobacter pylori* 26695 and *C. jejuni* NCTC 11168, as well as those of the γ -Proteobacteria consisting of *Haemophilus influenzae* Rd and *E. coli* K-12^T emerged from the FPP synthase tree and were supported by a high bootstrap value using both inference methods. The grouping of the Gram-positive phylum represented by the low-G+C-containing *Bacillus subtilis* 168, *Bacillus halodurans* C-125 and '*Micrococcus luteus*' B-P 26 is also evident in the tree. The positions of *Thermotoga maritima* MSB8 and '*Aquifex aeolicus*' VF-5 in the FPP synthase tree, although tentative due to a low bootstrap value, are outside of the other bacterial groups and are in good

agreement with the result of the 16S rRNA phylogenetic analyses (Woese, 1987; Snel *et al.*, 1999).

The phototrophic strains *Rhodobacter capsulatus* ATCC 11166^T, *Rhodobacter sphaeroides* ATCC 11167^T and *Rhodovulum sulfidophilum* W4^T formed a highly supported cluster (100%) in the FPP synthase gene tree, confirming the phylogenetic affiliations of the genus *Rhodobacter* and *Rhodovulum* (Kawasaki *et al.*, 1993; Hiraishi & Ueda, 1994). The most notable feature of the FPP synthase gene tree is the highly supported grouping (100% bootstrap value) of the non-phototrophic *Bradyrhizobium japonicum* USDA 110 and *Mesorhizobium loti* MAFF 303099. These species are root- and stem-nodulating bacterial symbionts of legumes and are relatives of physiologically diverse phototrophic anaerobes such as *Rhodobacter* and *Rhodopseudomonas* spp. (So *et al.*, 1994). Reconstruction of the 16S rRNA phylogeny showed that the 16S rRNA nucleotide sequence similarity of *Mesorhizobium loti* MAFF 303099 with that of the three phototrophic strains is approximately the same (89.1–90.6%) as the *Mesorhizobium loti* MAFF 303099/*Bradyrhizobium japonicum* USDA 110 sequence similarity (90.0%). Nevertheless, the similarity value of the FPP synthase (Table 3) between *Mesorhizobium loti* MAFF 303099 and *Bradyrhizobium japonicum* USDA 110 is high (92.6%) compared with that between *Mesorhizobium loti* MAFF 303099 and the phototrophic strains (44.3–47.1%), suggesting that the FPP synthase genes of the two non-phototrophic rhizobia are highly homologous. The long branch length in the gene tree suggests a recent acquisition of the FPP synthase gene by these two nodule-forming bacteria. The monophyletic

groupings of members of the α -*Proteobacteria* included in this study seemed to correlate with their phenotypic characteristics (i.e. phototrophic ability and the ability to form nodules in their host plants).

The use of a single gene in studying evolutionary relationships among different groups of organisms provides little or insufficient information. In particular, the intermingling of the phototrophic and non-phototrophic bacteria in the *Proteobacteria* could not be explained by 16S rRNA sequence analysis alone. Other genes of taxonomic importance should be used to infer this intermingled relationship. Genes involved in photosynthesis will not provide any direct evidence since these genes are not present in non-phototrophic bacteria. This study therefore focussed on the genes involved in the early steps of carotenoid biosynthesis, which are present in both phototrophic and non-phototrophic bacteria. In this study, the gene encoding FPP synthase was introduced as a potential molecule for investigating the evolutionary relationship between phototrophic and non-phototrophic bacteria in the *Proteobacteria*; the phylogeny of this molecule, which is similar and thus lends credence to the 16S rRNA phylogeny, reflects both intra- and intergeneric relationships of the strains studied. Establishing the relationship at this level is essential since the phototrophic bacteria are interspersed in many phylogenetic branches based on 16S rRNA phylogeny. The presence of the FPP synthase gene in the highly divergent lineages investigated and the overall phylogenetic congruence of the FPP synthase gene tree with the 16S rRNA tree suggest that the two markers have the same evolutionary pattern, and therefore the gene evolved vertically from a common ancestral gene. The observed incongruence from the 16S rRNA phylogeny (i.e. the monophyletic grouping of *Bradyrhizobium japonicum* USDA 110 and *Mesorhizobium loti* MAFF 303099), however, indicates that, in addition to vertical transmission, some genetic rearrangements (e.g. lateral gene transfer) have been involved in the evolution of FPP synthase. It is hoped that this observed difference between the two phylogenetic trees could provide helpful information on the phylogenetic relationships of the phototrophic and non-phototrophic bacteria in the *Proteobacteria*.

An extension of this study to various bacterial groupings, wherein phototrophic and non-phototrophic bacteria are closely related, is deemed important in investigating the phylogenetic relationships of these organisms to finally come up with a conclusive idea concerning this type of relationship. Nevertheless, it is noteworthy to conclude that the FPP synthase gene is useful for clearly differentiating species within and above the genus level due to its greater sequence variation compared with 16S rRNA. Furthermore, the characteristics of the FPP synthase gene, i.e. it is a housekeeping gene (Rivera *et al.*, 1998) and hence easily transferred laterally or lost during the course of evolution, and its occurrence in one copy number per cell, can be very useful for inferring the

probable cause of unusual evolutionary processes such as gene transfer events. Expansion of the FPP synthase gene database will contribute not only to phylogenetic studies but also to studies of the function and secondary structure of the gene itself.

ACKNOWLEDGEMENTS

This work was supported by a Grant-in-Aid (12760056) from the Japan Society for the Promotion of Science, Ministry of Education, Culture, Sports, Science and Technology, Japan, to H. Kawasaki-Nakagawa. The authors are indebted to Dr D. Honda (Konan University, Japan) for the assistance with the MOLPHY software and to Professor T. Koyama (Tohoku University, Japan) for generously providing the '*Micrococcus luteus*' B-P 26 strain. Special thanks to S. Kosaki for the help extended during the conduct of this study.

REFERENCES

- Adachi, J. & Hasegawa, M. (1996). *MOLPHY version 2.3. Programs for Molecular Phylogenetics Based on Maximum Likelihood*. Computer Science Monographs no. 28. Tokyo: Institute of Statistical Mathematics.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Anderson, M. S., Yarger, J. G., Burck, C. L. & Poulter, C. D. (1989). Farnesyl diphosphate synthetase. Molecular cloning, sequence, and expression of an essential gene from *Saccharomyces cerevisiae*. *J Biol Chem* **264**, 19176–19184.
- Armstrong, G. A., Cook, D. N., Ma, D., Alberti, M., Burke, D. H. & Hearst, J. E. (1993). Regulation of carotenoid and bacteriochlorophyll biosynthesis genes and identification of an evolutionarily conserved gene required for bacteriochlorophyll accumulation. *J Gen Microbiol* **139**, 897–906.
- Attucci, S., Aitken, S. M., Gulick, P. J. & Ibrahim, R. K. (1995). Farnesyl pyrophosphate synthase from white lupin: molecular cloning, expression, and purification of the expressed protein. *Arch Biochem Biophys* **321**, 493–500.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1995). *Current Protocols in Molecular Biology*. New York: Wiley.
- Chen, A., Kroon, P. A. & Poulter, C. D. (1994). Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. *Protein Sci* **3**, 600–607.
- 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.
- Fujisaki, S., Hara, H., Nishimura, Y., Horiuchi, K. & Nishino, T. (1990). Cloning and nucleotide sequence of the *ispA* gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*. *J Biochem* **108**, 995–1000.
- Hiraishi, A. & Ueda, Y. (1994). Intragenetic structure of the genus *Rhodobacter*: transfer of *Rhodobacter sulfidophilus* and related marine species to the genus *Rhodovulum* gen. nov. *Int J Syst Bacteriol* **44**, 15–23.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with CLUSTAL X. *Trends Biochem Sci* **23**, 403–405.
- Kawasaki, H., Hoshino, Y. & Yamasato, K. (1993). Phylogenetic diversity of phototrophic purple non-sulfur bacteria in the *Proteobacteria* α group. *FEMS Microbiol Lett* **112**, 61–66.
- Koyama, T., Obata, S., Osabe, M., Takeshita, A., Yokoyama, K., Uchida, M., Nishino, T. & Ogura, K. (1993). Thermostable farnesyl diphosphate synthase of *Bacillus stearothermophilus*: molecular cloning,

sequence determination, overproduction and purification. *J Biochem (Tokyo)* **113**, 355–363.

Lang, H. P., Cogdell, R. J., Takaichi, S. & Hunter, C. N. (1995). Complete DNA sequence, specific Tn5 insertion map, and gene assignment of the carotenoid biosynthesis pathway of *Rhodobacter sphaeroides*. *J Bacteriol* **177**, 2064–2073.

Nagashima, K. V. P., Hiraishi, A., Shimada, K. & Matsuura, K. (1997). Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J Mol Evol* **45**, 131–136.

Ogura, K., Koyama, T. & Sagami, H. (1997). Polyprenyl diphosphate synthases. In *Subcellular Biochemistry*, vol. 28, *Cholesterols: its Functions and Metabolism in Biology and Medicine*, pp. 57–87. Edited by R. Bittman. New York: Plenum Press.

Ohnuma, S., Suzuki, M. & Nishino, T. (1994). Archaeobacterial ether-linked lipid biosynthetic gene: expression, cloning, sequencing and characterization of geranylgeranyl diphosphate synthase. *J Biol Chem* **269**, 14792–14797.

Ohnuma, S., Narita, K., Nakazawa, T., Ishida, C., Takeuchi, Y., Ohto, C. & Nishino, T. (1996). A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. *J Biol Chem* **271**, 30748–30754.

Ohnuma, S., Hirooka, K., Ohto, C. & Nishino, T. (1997). Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase. Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. *J Biol Chem* **272**, 5192–5198.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Perrière, G. & Gouy, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364–369.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A* **95**, 6239–6244.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–5467.

Snel, B., Bork, P. & Huynen, M. A. (1999). Genome phylogeny based on gene content. *Nat Genet* **21**, 108–110.

So, R. B., Ladha, J. K. & Young, J. P. (1994). Photosynthetic symbionts of *Aeschynomene* spp. form a cluster with bradyrhizobia on the basis of fatty acid and rRNA analyses. *Int J Syst Bacteriol* **44**, 392–403.

Stackebrandt, E., Rainey, F. A. & Ward-Rainey, N. (1996). Anoxygenic phototrophy across the phylogenetic spectrum: current understanding and future perspectives. *Arch Microbiol* **166**, 211–223.

Stanley Fernandez, S. M., Kellogg, B. A. & Poulter, C. D. (2000). Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity. *Biochemistry* **39**, 15316–15321.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Wang, K. & Ohnuma, S. (1999). Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends Biochem Sci* **24**, 445–451.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.

Xiong, J., Inoue, K. & Bauer, C. E. (1998). Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proc Natl Acad Sci U S A* **95**, 14851–14856.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.