*Photosynthesis Research* **76:** 173–183, 2003.



*Minireview*

# Evolutionary relationships among photosynthetic bacteria

Radhey S. Gupta
*Department of Biochemistry, McMaster University, Hamilton, Ontario, L8N 3Z5 Canada*
*(e-mail: gupta@mcmaster.ca; fax: +1-905-522-9033)*

Received November 18, 2002; accepted in revised form December 29, 2002

*Key words:* bacterial phylogeny, R.E. Blankenship, branching order, genome sequences, lateral gene transfer, photosynthesis, protein signatures, R.S. Gupta, reaction centers, C.R. Woese

## Abstract

To understand the evolution of photosynthetic bacteria it is necessary to understand how the main groups within *Bacteria* have evolved from a common ancestor, a critical issue that has not been resolved in the past. Recent analysis of shared conserved inserts or deletions (indels) in protein sequences has provided a powerful means to resolve this long-standing problem in microbiology. Based on a set of 25 indels in highly conserved and widely distributed proteins, all main groups within bacteria can now be defined in clear molecular terms and their relative branching orders logically deduced. For the 82 presently completed bacterial genomes, the presence or absence of these signatures in various proteins was found to be almost exactly as predicted by the indel model, with only 11 exceptions observed in 1842 observations. The branching order of different bacterial groups as deduced using this approach is as follows: low G+C Gram-positive (*Heliobacterium chlorum*) → high G+C Gram-positive → *Clostridium–Fusobacterium–Thermotoga* → *Deinococcus–Thermus* → green nonsulfur bacteria (*Chloroflexus aurantiacus*) → Cyanobacteria → *Spirochetes* → *Chlamydia–Cytophaga–Flavobacteria*–green sulfur bacteria (*Chlorobium tepidum*) → *Aquifex* → *Proteobacteria* ($\delta$ and $\epsilon$) → *Proteobacteria* ($\alpha$) → *Proteobacteria* ($\beta$) and → *Proteobacteria* ($\gamma$). The *Heliobacterium* species, which contain an Fe–S type of reaction center (RC 1) and represent the sole photosynthetic phylum from the Gram-positive or monoderm bacteria (i.e., bounded by only a single membrane), is indicated to be the most ancestral of the photosynthetic lineages. Among the Gram-negative or diderm bacteria (containing both inner and outer cell membranes) the green nonsulfur bacteria, which contain a pheophytin-quinone type of reaction center (RC 2), are indicated to have evolved first. The later emerging photosynthetic groups which contain either one or both of these reaction centers could have acquired such genes from the earlier branching lineages by either direct descent or by means of lateral gene transfer.

*Abbreviations:* a.a. – amino acid(s); CFBG – Chlamydia–Cytophaga–Bacteroides–Green sulfur bacteria; Fe–S – iron–sulfur; $G^+$ or $G^-$ – Gram-Positive or Gram-negative bacteria; Hsp – heat shock protein; indel – insert or deletion PAC-transformylase – 5′-phosphoribosyl 5-aminoimidazole-4-carboxamide formyltransferase; Proteo – proteobacteria; PRPP synthetase – phosphoribosyl pyrophosphate synthetase; PS – Photosystem; RC(s) – reaction center(s)

## Introduction

The advent of photosynthesis by prokaryotic organisms represented a seminal event in the history of life. In order to understand the evolution of photosynthesis, it is necessary to resolve how different groups of photosynthetic organisms are related to each other and how they branched off from a common ancestor. Within prokaryotes, photosynthetic capability is present within five major groups of bacteria, which include: (i) the low G+C Gram-positive bacteria such as heliobacteria (Gest and Favinger 1983);

(ii) green nonsulfur or filamentous bacteria such as *Chloroflexus aurantiacus*; (iii) green sulfur bacteria such as *Chlorobium*; (iv) proteobacteria; and (v) cyanobacteria (Blankenship 1992; Kondratieva et al. 1992; Olson and Pierson 1987; Vermaas 1994; Woese 1987; Yurkov and Beatty 1998). Of these, only cyanobacteria which contain two different reaction centers (RC), RC 1 and RC 2 (or PS I and PS II) linked to each other are capable of carrying out oxygenic photosynthesis. All other photosynthetic bacteria carry out only anoxygenic photosynthesis and contain a single reaction center. The RCs of heliobacteria and green sulfur bacteria contain iron–sulfur clusters (Fe–S) as electron acceptors and they are similar to the PS I of cyanobacteria. In contrast, the RCs of green nonsulfur bacteria and proteobacteria are analogous to the PS II of cyanobacteria and they contain quinones as electron acceptors. The observed similarities in different RCs in their component parts, and the overall charge transfer mechanisms, indicate that they all originated from a common ancestor (Blankenship 1994; Golbeck 1993; Nitschke and Rutherford 1991; Vermaas 1994). To understand which of these RCs evolved first, it is essential to understand how different groups of bacteria have evolved.

The phylogenetic trees based on 16S rRNA provide the present framework for understanding the evolutionary relationships within prokaryotes (Ludwig and Klenk 2001; Olsen et al. 1994; Woese 1987). Based on these trees a number of main groups or phyla within *Bacteria* have been identified. These include *Thermotoga, Aquifex, Deinococcous–Thermus,* low G+C Gram-positive (*Firmicutes*), high G+C Gram-positive (*Actinobacteria*), green nonsulfur bacteria, cyanobacteria, Spirochetes, Chlamydiae, *Cytophaga–Flavobacteria–Bacteroides*–green sulfur bacteria (CFBG group), and proteobacteria. The proteobacteria are further divided into five subdivisions ($\alpha$, $\beta$, $\gamma$, $\delta$ and $\epsilon$). However, despite the extensive use of 16S rRNA trees, these trees are unable to resolve the evolutionary relationships among different groups within *Bacteria*, which is central to understanding bacterial evolution (Ludwig and Klenk 2001; Woese 1987). Recently, we have described a new approach based on shared conserved inserts and deletions found in various proteins (referred to as indels or signature sequences) that provides a powerful means to reliably resolve these relationships (Gupta 1998, 2000a, 2001, 2002; Gupta and Griffiths 2002). In this review, I will first briefly describe this approach and its usefulness in deducing the branching order of the bacterial groups. Subsequently, the application of this approach for determining the evolutionary relationships among photosynthetic bacteria and evolution of photosynthesis will be discussed.

## The use of signature sequences for deducting the branching order of bacterial groups

The usefulness of a conserved indel for understanding evolutionary relationship is based on the rationale that if a conserved indel of defined length and sequence is found at the same position in a given protein (or gene) in all members from one or more groups of bacteria, then its simplest and most parsimonious explanation is that this indel was introduced only once in a common ancestor of these species. Thus, based on the presence or absence of various shared indels, different species can be divided into distinct groups (Gupta 1998, 2000a). Well-defined indels in proteins can also serve as useful milestones for evolutionary purposes, since it is expected that all species emerging from an ancestral cell in which the indel was first introduced will contain the signature, whereas all species that either existed prior to this event or which did not evolve from this ancestor will lack the indel. However, to understand the evolutionary significance of a given indel and to determine whether it is the result of an insertion or deletion event, a reference point is needed. A useful reference point for such purposes is provided by the root of the prokaryotic tree. Based on several lines of molecular sequence evidence discussed in an earlier review (Gupta 1998), the root of the prokaryotic tree has been placed between archaebacteria and Gram-positive bacteria. Cells of both these groups of prokaryotes are bounded by a single unit lipid membrane (thus termed monoderms), which is indicated to be an ancestral characteristic in comparison to the bacteria bounded by both an inner and outer cell membranes (i.e., diderms) (Gupta 1998). Based on this rooting, the signature sequences in various proteins can be logically interpreted to deduce the branching order of different groups.

We have identified a large number of conserved indels in different proteins which provide valuable markers for such studies (Gupta 1998, 2000a, 2001). Of these, 25 indels indicated in Figure 1 have proven most useful for determining the relative branching orders of different groups. Based upon the presence or absence of these indels in the particular proteins, most of the major groups within *Bacteria* can be clearly

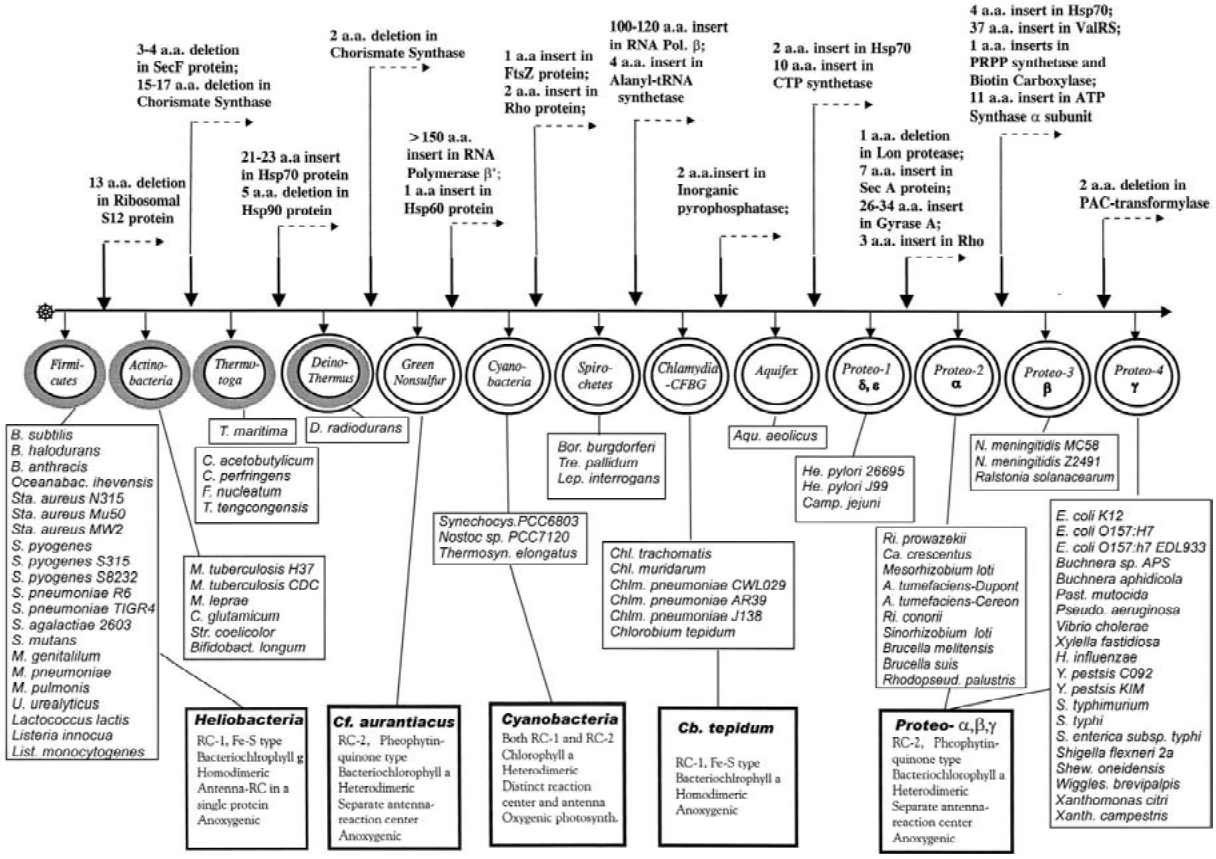


*Figure 1.* The branching order of the main groups within *Bacteria* based on conserved indels present in various proteins. The thick arrows above the line show the evolutionary stages where the identified indels are postulated to have been introduced. The phylogenetic placements of species from completed bacterial genomes based on these signatures are as shown. The model predicts that once a signature has been introduced in the main lineage, all groups to the right of it should contain the signature whereas all groups preceding it should lack it. The model is strongly supported by the sequence data for completed bacterial genomes with only 11 exceptions from the predicted pattern observed in 1842 observations. The order of emergence of photosynthetic groups and some characteristics of their RCs are described in the boxes below. Abbreviations in the protein names are: PRPP synthetase – phosphoribosyl pyrophosphate synthetase; PAC-transformylase – 5-aminoimidazole-4-carboxamide formyltransferase.

distinguished and their relative branching order from a common ancestor can be deducted. The detailed descriptions of most of these signatures can be found in our published work (Gupta 1998, 2000a, 2001; Griffiths and Gupta 2002). However, an example of a new signature in the RNA polymerase $\beta$ subunit is presented in Figure 2. This protein is a core component of the RNA polymerase and it is present in all sequenced bacterial genomes. The signature in this case consists of a large insert of between 90–133 a.a. which is commonly shared by various proteobacteria, *Aquifex* and the CFBG group of species, but not found in any of the other groups of bacteria. The insert is flanked on both sides by conserved sequences, indicating its reliability. This indel was originally described by Klenk et al. (1999), but its evolutionary significance was not clear. However, according to our model, similar to the indel in the alanyl-tRNA synthetase, this insert was introduced in a common ancestor of the proteobacteria, *Aquifex* and CFBG groups after branching of the other bacterial phyla which lack this indel (Figure 1).

The model shown in Figure 1 indicating the evolutionary stages where the identified indels have been introduced makes very specific predictions as to which of these indels should be present or absent in different species (Gupta 2001, 2002). According to the model, once an indel has been introduced in an ancestral lineage, various groups emerging after that point should all contain the indel, whereas all species from dif-

ferent groups that existed prior to the introduction of the indel should lack the signature. We have tested the reliability and predictive power of the model by examining the presence of these indels in all 81 bacterial genomes which have thus far been sequenced. (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html), as well as for *Rhodopseudomonas palustris*, whose sequence is available in GenBank and at http://genome.ornl.gov/microbial/rpal/1/rpa 1_1.html. The presence or absence of these indels in various genomes was determined by means of BLAST analysis and sequence alignments. The results of these analyses are presented in Table 1. The table lists the number of genomes where these proteins are found, and also the number of species that are predicted to either contain or lack each of these indels based on the positions of the indels and the deduced branching order of the groups. For example, for the large 21–23 amino acid insert in Hsp70, based on its position, the model predicts that 50 of the 82 species should contain this indel whereas the other 32 should be lacking it. Similarly, for the signatures in Ala-tRNA synthetase or RNA polymerase $\beta$ subunit (RpoB), 43 of the 82 species which are to the right of the arrow are all expected to contain the indel, whereas all 39 species on the left should not possess it. The actual results obtained for different indels are also summarized in Table 1. Since each of these genes/proteins is not present in all bacterial genome, the total number of observations are not the same for all indels/proteins. The results of these analyses are strikingly clear (Table 1), as the presence or absence of these indels in different genomes followed exactly the pattern as predicted by the model. In a total of 1842 observations concerning the placement of indels in different species, only 11 exceptions to the predicted pattern were observed. Most of these exceptions are found in the protein chorismate synthase. The remainder of the proteins, a large number of which are found in all Bacteria, behaved exactly as predicted by the model. These results provide compelling evidence that the signature approach provides a reliable and internally consistent means for determining the relative branching orders of the bacterial groups. The nearly perfect concordance seen between the predicted and observed results in over 1800 observations also provides strong evidence that the genes containing these indels, most of which are for highly conserved housekeeping functions, have not been affected or corrupted by factors such as lateral gene transfer (LGT) (Doolittle 1999; Jain et al. 1999; Koonin et al. 2001). If these genes were affected by LGTs, or if these indels were introduced independently into these genes in different species, then it would not have been possible to predict the presence of absence of these indels with such a high degree of accuracy ($> 99.0\%$). Of the various genes containing these indels, significant number of anomalies are observed only in Chorismate synthase, which could be the result of either LGTs or independent occurrence of indels in some of the species.

## Phylogenetic placements and branching orders of photosynthetic bacteria

To determine the relative branching order of various photosynthetic bacteria using the indel approach, we have previously cloned and sequenced gene fragments containing some of these signatures from several photosynthetic bacteria (Gupta et al. 1999). Sequence information for other signatures from photosynthetic bacteria was obtained from published and unpublished microbial genome databases by BLAST analyses. Results of these studies for different photosynthetic bacteria phyla are presented in Table 2 and discussed below.

### *Heliobacterium chlorum*

Signature sequences in a number of proteins provide evidence concerning the placement of *H. chlorum* in the low G+C Gram-positive (*Firmicutes*) group. These include the presence in this species of the 13 a.a. insert in ribosomal S12 protein which is distinctive of the *Firmicutes* group (E. Griffiths and R.S. Gupta, unpublished results) as well as the absence of the large insert in the Hsp70 protein, which distinguishes Gram-positive (monoderm) bacteria from Gram-negative (diderm) bacteria (Gupta et al. 1999). As expected, *H. chlorum* also lacked the insert in the Hsp60 protein (Gupta et al. 1999), which is indicated to have been introduced after the branching of *Deinococcus-Thermus and* green nonsulfur groups of bacteria. The status of most other indels in *H. chlorum* has not been examined, but based on its phylogenetic placement it is now possible to predict whether they will be present or absent (Table 2). Phylogenetic analyses based on 16S rRNA also places *H. chlorum* in the low G+C Gram-positive group (Olsen et al. 1994; Vermaas 1994).

*Table 1.* Predicted *versus* observed distribution of indels in proteins from completed bacterial genomes

| Protein | Signature description | No. of genomes with protein | No. of genomes with indels expected/found | No. of genomes lacking the indel expected/found | Exceptions observed |
|---|---|---|---|---|---|
| Ribosomal S12 protein | 13 a.a. low G+C signature | 81 | 20/20 | 61/61 | 0 |
| Chorismate synthase | 15–17 a.a. deletion after *Actinobacteria* | 72 | 23/26 | 49/46 | 3[a] |
| SecF protein | 3–4 a.a. deletion after *Actinobacteria* | 67 | 13/13 | 54/54 | 0 |
| Hsp70/DnaK | 21–23 a.a. G+/G− insert | 82 | 50/50 | 32/32 | 0 |
| Hsp90 | 5 a.a. G+/G− insert | 45 | 10/10 | 35/35 | 0 |
| Chorismate synthase | 2 a.a. deletion after *Deinococcus* | 72 | 28/29 | 44/43 | 5[b] |
| Hsp60/GroEL | 1 a.a. insert after *Deinococcus* | 80 | 49/49 | 31/31 | 0 |
| RNA polymerase $\beta'$-subunit (RpoC) | >150 a.a. insert after cyanobacteria | 82 | 46/46 | 36/36 | 0 |
| FtsZ protein | 1 a.a. insert after cyanobacteria | 74 | 41/41 | 33/33 | 0 |
| Rho protein | 2 a.a. insert before spirochetes | 67 | 46/47 | 21/20 | 1 (tm)[c] |
| Ala-tRNA synthetase | 4 a.a. insert after spirochetes | 82 | 43/43 | 39/39 | 0 |
| RNA polymerase $\beta$-subunit | 100–120 a.a. insert after spirochetes | 82 | 43/43 | 39/39 | 0 |
| Inorganic pyro-phosphatase | 2 a.a. insert common to *Aquifex* and proteo | 56 | 36/36 | 20/20 | 0 |
| Hsp70/DnaK | 2 a.a. Proteo insert | 82 | 36/36 | 46/46 | 0 |
| CTP synthetase | 10 a.a. insert before proteobacteria | 75 | 36/36 | 39/39 | 0 |
| Lon protease | 1 a.a. deletion in $\alpha\beta\gamma$ proteobacteria | 59 | 26/26 | 33/33 | 0 |
| Rho protein | 3 a.a. insert in $\alpha\beta\gamma$ proteobacteria | 67 | 33/33 | 34/34 | 0 |
| DNA gyrase A subunit | 26–34 a.a. insert in $\alpha\beta\gamma$ proteobacteria | 82 | 33/33 | 49/49 | 0 |
| SecA protein | 7 a.a. insert in $\alpha\beta\gamma$ proteobacteria | 82 | 33/33 | 49/49 | 0 |
| HSP70/DnaK | 4 a.a. $\beta\gamma$ insert | 82 | 23/24 | 59/58 | 1[d] |
| Valyl-tRNA synthetase | 37 a.a. $\beta\gamma$ insert | 82 | 23/23 | 59/59 | 0 |
| ATP synthase $\alpha$-subunit | 11 a.a. $\beta\gamma$ insert | 74 | 23/23 | 41/41 | 0 |

*Table 1.* Continued

| Protein | Signature description | No. of genomes with protein | No. of genomes with indels expected/found | No. of genomes lacking the indel expected/found | Exceptions observed |
|---|---|---|---|---|---|
| PRPP synthetase | 1 a.a. $\beta\gamma$ insert | 76 | 23/23 | 53/53 | 0 |
| Biotin carboxylase | 1 a.a. $\beta\gamma$ insert | 72 | 22/23 | 50/49 | 1[e] |
| PAC-transformylase | 2 a.a. $\gamma$-proteo deletion | 67 | 47/47 | 20/20 | 0 |

[a]Smaller inserts, which could be of independent origin, are present in this region *in Deinococcus, Aquifex* and *Chlorobium tepidum*.
[b]Exceptions observed are *Fusobacterium nucleatum, Clostridium perfringens, Clo. acetobutylicum, Aquifex aeolicus* and *Cb. tepidum*.
[c]*T. maritima* contains the indel, which is not expected.
[d]The presence of this insert in *Bifidobacterium longum* constitutes the exception.
[e]*Shewanella oneidensis* lacks the expected indel.
Abbreviations in the protein names are: PAC – 5-aminoimidazole-4 carboxamide formyltransferase; PRPP – phosphoribosyl pyrophosphate.

Chloroflexus aurantiacus *(green nonsulfur bacteria)*

Signature sequences in various proteins provide evidence that *C. aurantiacus* is one of the earliest branching lineage within gram-negative (diderm) bacteria (Table 2). The presence of the large insert in the Hsp70 protein as well as the absence of indels in the Hsp90, SecF and chorismate synthase provide evidence that it belongs to the Gram-negative or diderm group of bacteria (Gupta et al. 1999) (Table 2). The absence of indels in Hsp60 and RNA polymerase $\beta'$ proteins in *C. aurantiacus* indicate that it branched off prior to the cyanobacterial group and that it is one of the earliest branching lineage within Gram-negative bacteria. Although in our earlier work, no signatures were known that could distinguish between the *Deinococcus–Thermus* group and the green nonsulfur bacteria, a new signature in the Chorismate synthase that we have identified indicates that *Chloroflexus* has branched after the *Deinococcus–Thermus* group. The earlier branching of the *Deinococcus–Thermus* group is also consistent with the observation that many species belonging to this group contain a thick cell wall similar to the Gram-positive bacteria (Murray 1986).

*Cyanobacteria*

The branching position of cyanobacteria in the indicated position (Figure 1) is supported by all of the identified signatures (Figure 1 and Table 2). The signature sequences in Hsp60, RNA polmerase $\beta'$, FtsZ and Rho proteins provide evidence that this group branched after the green nonsulfur bacteria but prior to green sulfur group (*Cb. tepidum*).

Chlorobium tepidum *(green sulfur bacteria)*

The genome sequence of *Chlorobium tepidum* has recently been determined (Eisen et al. 2002). The signature sequences in various proteins provide evidence that *Cb. tepidum* branches in a similar position as the *Chlamydia–Flavobacteria–Bacteriodes* (CFBG) group of species. This inference is supported by all of the signatures without any exceptions (Table 2).

*Proteobacteria*

Among proteobacteria the photosynthetic capability has been observed within members of $\alpha$-, $\beta$- as well as $\gamma$-subdivisions (Kondratieva et al. 1992; Woese 1987; Yurkov and Beatty 1998). Although these groups are presently referred to as different subdivisions of proteobacteria, based on the facts that these groups can be clearly distinguished from each other in phylogenetic trees as well as by large number of signature sequences, and their relative branching order can be clearly deduced, we have argued that these groups should be accorded a similar status as the other main divisions within *Bacteria* (Gupta 2000a; Gupta and Griffiths 2002). Based on signature sequences in different proteins (Table 2), these groups are indicated to be the last of the photosynthetic groups that have evolved, with the relative branching orders as shown

*Table 2.* Presence of various identified indels in different photosynthetic phyla

| Signature | *Hel. chlorum* | *Cf. aurantiacus* | Cyano-bacteria | *Cb. tepidum* | *Rhodo. palustris* | $\beta$-Proteo-bacteria | $\gamma$-Proteo-bacteria |
|---|---|---|---|---|---|---|---|
| Ribosome S12 protein indel | + | ND (−)[a] | − | − | − | − | − |
| Hsp70 G+/G− indel | − | + | + | + | + | + | + |
| Hsp90 G+/G− indel | ND | − | − | − | − | − | − |
| Chorismate synthase 15–17 a.a. indel | ND (+) | − | − | +/−[b] | − | − | − |
| Sec F protein Indel | ND (+) | − | − | − | − | − | − |
| Chorismate synthase (2 a.a.) | ND (+) | − | − | − | − | − | − |
| Hsp60 indel | − | − | + | + | + | + | + |
| RNA Pol. $\beta'$-subunit indel | ND (−) | − | + | + | + | + | + |
| FtsZ indel | ND (−) | − | − | + | + | + | + |
| Rho factor Spirochete indel | ND (−) | − | 0[c] | + | + | + | + |
| Ala–tRNA synth. indel | ND (−) | ND (−) | − | + | + | + | + |
| RNA Pol. $\beta$-subunit indel | ND (−) | − | − | + | + | + | + |
| Inorganic pyrophosphatase indel | ND (−) | ND(−) | − | − | + | + | + |
| Hsp70 proteobacteria indel | − | − | − | − | + | + | + |
| CTP Synthase proteo indel | ND (−) | ND (−) | − | − | + | + | + |
| Lon protease $\alpha\beta\gamma$ indel | ND (+) | + | 0 | 0 | − | − | − |
| Rho protein $\alpha\beta\gamma$ indel | ND (−) | − | 0 | − | + | + | + |
| DNA gyrase A $\alpha\beta\gamma$ indel | ND (−) | − | − | − | + | + | + |
| Sec A protein $\alpha\beta\gamma$ indel | ND (−) | − | − | − | + | + | + |
| Hsp70 $\beta\gamma$ indel | − | − | − | − | − | + | + |
| Val-tRNA synthetase, $\beta\gamma$ indel | ND (−) | − | − | − | − | + | + |
| ATP synthase $\beta\gamma$ indel | ND (−) | − | − | − | − | + | + |
| PRPP synthetase $\beta\gamma$ indel | ND (−) | − | − | − | − | + | + |
| Biotin caroboxylase indel | ND (−) | − | − | − | − | + | + |
| PAC-transformylase indel | ND (−) | − | − | − | − | − | + |

[a]ND – not determined. The (−) or (+) sign indicates the prediction of the model for whether the indicated indel will be present or not.
[b]This indel is shorter than that found in other species and could have resulted independently.
[c]0 indicates that the protein was not identified in this species by BLAST analysis.

in Figure 1. The distribution of various indels in the photosynthetic bacterium *Rhodopseudomonas palustris* was found to be the same as in various $\alpha$ proteobacteria, confirming its assignment to this group and its relative branch order (Figure 1 and Table 2).

## Implications of the observed branching order for the evolution of photosynthesis

The problem of understanding the evolutionary relationships among photosynthetic bacteria is an integral part of the broader question as to how the different main groups within *Bacteria* have branched off from a common ancestor. Using the indel approach described here, we have been able to logically and reliably deduce the branching order of different bacterial groups as shown in Figure 1. In recent years, several investigators have used proteins or pigments that are part of the photosynthetic apparatus to infer the branching order of photosynthetic bacteria (Burke et al. 1993; Mulkidjanian and Junge 1997; Xiong et al. 1998, 2000). These studies have led to inferences that are quite different from those obtained here. For example, Xiong et al. (1998, 2000) have reported detailed phylogenetic studies based on a number of genes involved in photosynthesis-related functions. Based on their analysis, they inferred that among photosynthetic bacteria, proteobacterial photosynthesis genes represented the earliest branching lineage. The derived phylogenetic tree(s) also suggested a specific relationship between *Chlorobium* and *Chloroflexus*, and cyanobacteria were indicated to be the last or most recently evolved photosynthetic lineage. These

inferences are quite unusual and contrary to our understanding derived from other sources. In various published phylogenetic trees based on 16S rRNA or various other widely distributed genes/proteins, proteobacteria consistently appear as one of the most recently evolved groups, with green nonsulfur bacteria, cyanobacteria, Gram-positive bacteria and green sulfur bacteria showing much deeper branching (Brown and Doolittle 1997; Eisen 1995; Gupta 1998; Olsen et al. 1994; Woese 1987). The observed tree topology and the inferred relationships thus are of concern. An important concern in these studies is the suitability of the genes or components involved in photosynthesis for inferring the branching order of photosynthetic phyla. Since such genes can confer enormous selective advantage to a recipient organism in suitable environment, they may have been horizontally transferred in the past from one species to others (Gupta 1998; Igarashi et al. 2001; Raymond et al. 2002). By restricting phylogenetic analysis to such genes that are found only among photosynthetic lineages, it is difficult to detect such LGT events, which in turn could lead to misleading inferences. In contrast, most of the genes/proteins on which the indel model is based are found in all different groups of bacteria and they likely have been acquired by vertical descent. The lateral transfer of such genes is thus expected to confer no selective advantage in recipient species and therefore it is less likely to occur, as is strongly indicated by the observed results (Table 1).

The inferences deduced from indel analysis are supported by a recent detailed study on whole genome analysis of photosynthetic bacteria. In this study, Raymond et al. (2002) made use of orthologs of various proteins which are found in all five photosynthetic taxa, and determined which of the 15 possible phylogenetic relationships among these phyla was supported by different proteins. Although each of the possible relationship was supported by at least some proteins, the most common pattern observed showed a grouping of *H. mobilis, C. aurantiacus* and cyanobacteria in one cluster, and *Cb. tepidum* and *R. capsulatus* in a separate one. Once the observed results were corrected for the abnormally long branch lengths, about two-thirds of the total of 188 proteins supported the above relationships. Further, based on the average number of substitutions per site and the shortest between-taxa distances, their analysis pointed to a specific relationship between *Heliobacterium* and *Chloroflexus*, followed by a pairing of *Chloroflexus* and *Synechocystis*, while the green sulfur bacteria (*Chlorobium*) and proteobacteria (*R. capsulatus*) were distantly related. Although this study provided no indication as to which of these groups was ancestral, their results are in complete agreement with those derived based on various indels (Figures 1 and 2). Raymond et al. (2002) also carried out similar analysis on genes/proteins involved in photosynthesis-related functions. Unlike the genes responsible for widely distributed functions, analysis of the 'photosynthesis-specific' genes supported no coherent relationships among different photosynthetic bacteria, indicating that such genes have been subjects of lateral gene transfers.

From the branching order of different groups it can now be inferred that of the different RCs, the one in heliobacteria is the most ancient. The ancestral nature of the RC in heliobacteria is also in accordance with a number of other observations (Blankenship 1992; Gest and Favinger 1983; Olson and Pierson 1987; Vermaas 1994; Nelson and Ben-Shem 2002): (i) Unlike other photosynthetic organisms, except possibly cyanobacteria, where the antenna and reaction center complexes are present on different polypeptides, in heliobacteria both these activities are non-separable and contained within a single protein; (ii) the reaction center complex in heliobacteria (and also green sulfur bacteria) has a simpler homodimeric structure as opposed to being heterodimeric in other photosynthetic bacteria; (iii) the RC in heliobacteria contains a unique photosynthetic pigment Bchl *g*, which based upon its biosynthetic pathway and chemical structure is indicated to be primitive in comparison to the pigments found in other photosynthetic organisms. Lastly, of all the photosynthetic bacteria, only heliobacteria are bounded by a single unit lipid membrane, which is indicated to be an ancestral characteristic in comparison to the cells containing both an inner and outer cell membranes (Gupta et al. 1999). Following heliobacteria, green nonsulfur bacteria are indicated to be the next group of photosynthetic organisms that branched off from the common ancestor. Since the RCs in different bacteria are related, this suggests that the pheophytin-quinone type of reaction center (RC 2) present in green nonsulfur bacteria likely evolved from the Fe–S type of reaction center (RC 1) found in heliobacteria. The branching of both heliobacteria and green nonsulfur bacteria prior to cyanobacteria also provides evidence that both RC 1 and RC 2 (or PS I and PS II) had already evolved prior to the emergence of cyanobacteria.

Although it is now possible to infer with high degree of confidence the relative branching order of

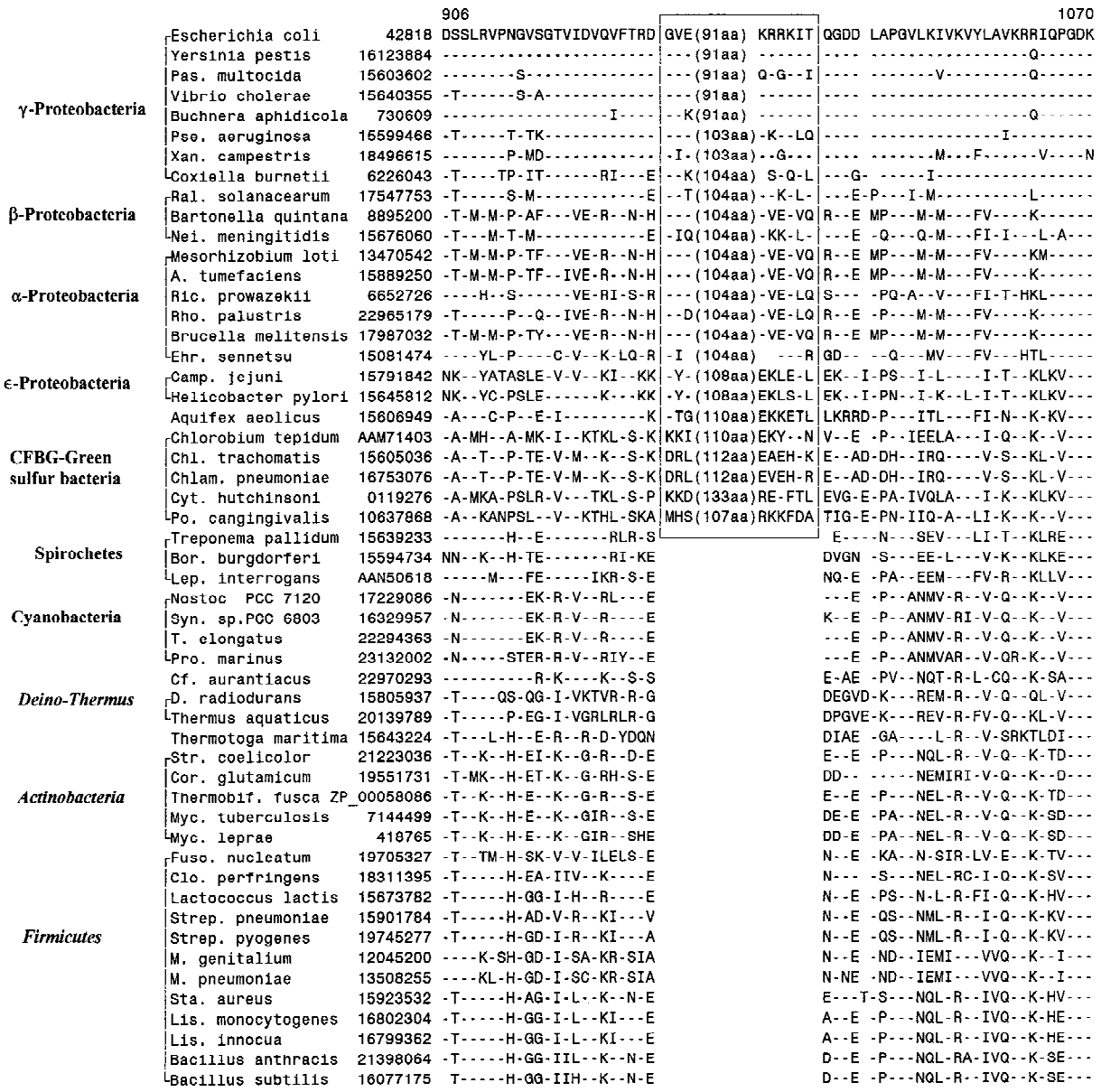


*Figure 2.* A signature sequence (boxed) consisting of a large insert in the RNA polymerase $\beta$-subunit that is common to various proteobacteria, *Aquifex* and the CFBG group of species but not found in any other bacteria. This insert was likely introduced in a common ancestor of these groups of bacteria. Because of the large size of the insert, the entire insert region is not shown but the number of additional a.a. residues present in between are indicated. The dashes show sequence identity with the amino acids present in the top line of the alignment. The numbers on the top line indicate the position of the sequence in the *E. coli* protein. The accession numbers of sequences are shown in the second column. The abbreviations of the species names are: *A.* – *Agrobacterium*; *Bor.* – *Borrelia*; *Camp.* – *Campylobacter*; *Cb.* – *Chlorobium*; *Cf.* – *Chloroflexus*; *Chl.* – *Chlamydia*; *Chlam.* – *Chlamydophila*; *Clo.* – *Clostridium*; *Cor.* – *Cornyebacterium*; *Cyt.* – *Cytophaga*; *D.* – *Deinococcus*; *Ehr.* – *Ehrlichia*; *Ent.* – *Enterococcus; Fuso.* – *Fusobacterium; Lep.* – *Leptospira; Lis.* – *Listeria; M.* – *Mycoplasma; Myc.* – *Mycobacterium; Nei.* – *Neisseria; Pas.* – *Pasteurella; Pse.* – *Pseudomonas; Po.* – *Porphyromonas; Pro.* – *Prochlorococcus; Ral.* – *Ralstonia; Ric. Rickettsia; Rho.* – *Rhodopseudomonas; Sta.* – *Staphylococcus; Str.* – *Streptomyces; Strep.* – *Streptococcus; Syn.* – *Synechocystis; T.* – *Thermosynechococcus; Xan.* – *Xanthomonas*.

photosynthetic organisms (Figure 1), and by inference which RC is most ancient, the question as to how RCs in later emerging groups (namely, cyanobacteria, green sulfur bacteria and various proteobacteria)

have been derived is more complex and it is far from being resolved. The presence of either RC 1, or RC 2, or both RC 1 and RC 2, in the latter emerging photosynthetic lineages could be explained by at least two different types of models (Blankenship 1992; Nitschke and Rutherford 1991; Vermaas 1994). One possibility is that in an ancestral organism that already contained the RC 1, duplication of genes for the photosynthetic components followed by extensive divergence led to the evolution of RC 2. These RCs somehow became linked in cyanobacteria, enabling them to oxidize water to produce molecular oxygen. In this case, the latter emerging photosynthetic organisms could have evolved from an evolutionary intermediate organism possessing both types of RCs (unlinked) by loss of either RC 1 (green nonsulfur bacteria and $\alpha$-, $\beta$- and $\gamma$-proteobacteria) or RC 2 (green sulfur bacteria). This possibility would necessitate loss of both RCs from several intermediate branching phyla (e.g., *Spirochetes, Aquifex*, $\delta$- and $\epsilon$-proteobacteria) which do not have any known photosynthetic member (Kondratieva 1992; Woese 1987). An alternative possibility is that an organism such as *H. chlorum,* which contains the Fe–S type of reaction center (RC 1), gave rise to an organism such as *Cf. aurantiacus*, which contained only the pheophytin-quinone type of RC (RC 2). In this case, the later emerging photosynthetic organisms which contain either only one of these two types of RCs (RC 2 in green sulfur bacteria and RC 1 in proteobacteria) or both (cyanobacteria) could have acquired the genes for the RC(s) components by means of LGT from earlier branching species. Although this latter possibility is attractive because acquisition of photosynthetic capability can confer much selective advantage on a recipient species, a combination of these possibilities is more likely to account for the properties of the RCs found in the later emerging groups. It should be possible to distinguish between these possibilities by examining how the RC components in the later branching groups are related to the earlier branching taxa.

## Acknowledgments

I thank various investigators whose published and unpublished sequence data in the public databases form the foundation of our work. Due to space constraints, references to various completed bacterial genomes are not provided. Such information is available at http://www.ncbi.nlm.nih.gov/PMGifs/Genomes /micr. html. The research work from the author's lab was supported by a research grant from the Canadian Institute of Health Research.

At the suggestion of the editors, I have included here a recent photograph of myself (Figure 3).

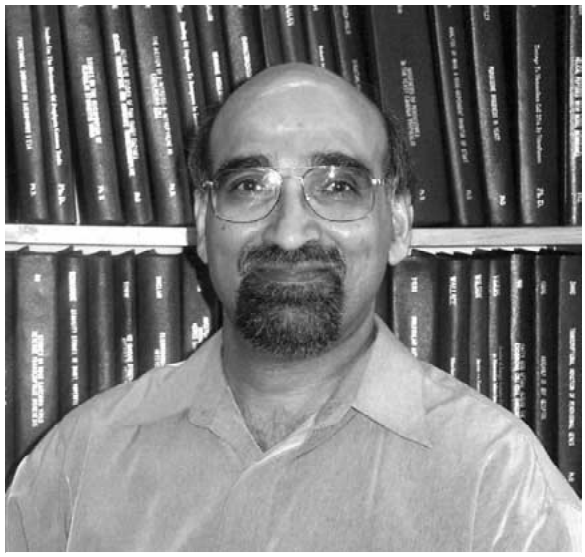

*Figure 3.* A recent photograph of the author (RSG).

## References

Blankenship RE (1992) Origin and early evolution of photosynthesis. Photosynth Res 33: 91–111

Blankenship RE (1994) Protein structure, electron transfer and evolution of prokaryotic photosynthetic reaction centers. Antonie van Leeuwenhoek 65: 311–329

Brown JR and Doolittle WF (1997) *Archaea* and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 61: 456–502

Burke DH, Hearst JE and Sidow A (1993) Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. Proc Natl Acad Sci USA 90: 7134–7138

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284: 2124–2128

Eisen JA (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

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, DeBoy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA and Fraser CM (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci USA 99: 9509–9514

Gest H and Favinger J (1983) *Heliobacterium chlorum*, an anoxygenic brownish-green photosynthetic bacterium containing a 'new' form of bacteriochlorophyll. Arch Microbiol 136: 11–16

Golbeck JH (1993) Shared thematic elements in photochemical reaction centers. Proc Natl Acad Sci USA 90: 1642–1646

Gupta RS (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62: 1435–1491

Gupta RS (2000a) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol Rev 24: 367–402

Gupta RS (2000b) The *natural* evolutionary relationships among prokaryotes. Crit Rev Microbiol 26: 111–131

Gupta RS (2001) The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int Microbiol 4: 187–202

Gupta RS (2002) Phylogeny of Bacteria: are we now close to understanding it? ASM News 68: 284–291

Gupta RS and Griffiths E (2002) Critical issues in bacterial phylogeny. Theor Popul Biol 61: 423–434

Gupta RS, Mukhtar T and Singh B (1999) Evolutionary relationships among photosynthetic prokaryotes (*Heliobacterium chlorum, Chloroflexus aurantiacus,* cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis. Mol Microbiol 32: 893–906

Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K and Nagashima KV (2001) Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. J Mol Evol 52: 333–341

Jain R, Rivera M and Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 96: 3801–3806

Klenk H-P, Meier T-D, Durovic P, Schwass V, Lottspeich F, Dennis PP and Zillig W (1999) RNA Polymerase of *Aquifex pyrophilus*: Implications for the evolution of the bacterial *rpoBC* operon and extremely thermophilic bacteria. J Mol Evol 48: 528–541

Kondratieva EN, Pfennig N and Truper HG (1992) The phototrophic prokaryotes. In: Balows A, Truper HG, Dworkin M, Harder W and Schleifer KH (eds) The Prokaryotes, pp 312–330. Springer-Verlag, New York

Koonin EV, Makarova KS and Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55: 709–742

Ludwig W and Klenk H-P (2001) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systamatics. In: Boone DR and Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology, Vol 1, 2nd ed: The *Archaea* and the Deeply Branching and Phototrophic *Bacteria*, pp 49–65. Springer-Verlag, Berlin

Mulkidjanian AY and Junge W (1997) On the origin of photosynthesis as inferred from sequence analysis. Photosynth Res 51: 27–42

Murray RGE (1986) Family II.*Deinococcaceae* Brooks and Murray 1981, 356VP. In: Sneath PHA, Mair NS, Sharpe ME and Holt JG (eds) Bergey's Manual of Systematic Bacteriology, Vol 2, pp 1035–1043. Williams and Wilkins, Baltimore, Maryland

Nelson N and Ben-Shem A (2002) Photosystem I reaction center: past and future. Photosynth Res 73: 193–206

Nitschke W and Rutherford AW (1991) Photosynthetic reaction centers: variation on a common structural theme? Trends Biochem Sci 16: 241–245

Olsen GJ, Woese CR and Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176: 1–6

Olson JM and Pierson BK (1987) Evolution of reaction centers in photosynthetic prokaryotes. Int Rev Cytol 108: 209–248

Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY and Blankenship RE (2002) Whole-genome analysis of photosynthetic prokaryotes. Science 298: 1616–1620

Vermaas WF (1994) Evolution of heliobacteria: Implications for photosynthetic reaction center complexes. Photosynth Res 41: 285–294

Woese CR (1987) Bacterial evolution. Microbiol Rev 51: 221–271

Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99: 8742–8747

Xiong J, Inoue K and Bauer CE (1998) Tracking molecular evolution of photosynthesis by characterizaton of a major photosynthesis gene cluster from *Heliobacillus mobilis*. Proc Natl Acad Sci USA 95: 14851–14856

Xiong J, Fischer WM, Inoue K, Nakahara M and Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. Science 289: 1724–1730

Yurkov VV and Beatty JT (1998) Aerobic anoxygenic phototropic bacteria. Microbiol Mol Biol Rev 62: 695–724