Chapter 34

Purple Bacterial Genomics

Madhusudan Choudhary[1], Chris Mackenzie[1],
Timothy J. Donohue[2] and Samuel Kaplan[1]*
[1]*Department of Microbiology and Molecular Genetics, The University of Texas Health Science Center, Houston, Texas 77030, U.S.A.;* [2]*Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53076, U.S.A.*

Summary 691
I. Introduction 692
II. Genome Architecture and Characteristics 692
A. Genome Comparisons 692
B. Evolutionary Relationships and the Origin of Mitochondria 694
C. Continual Evolution of Purple Bacteria 694
III. Gene Homologs and Metabolic Versatility 696
A. Flagellum Biosynthesis 696
B. CO_2 Utilization and Photosynthesis 699
C. Energy Production 699
D. Tetrapyrrole Biosynthesis 700
E. Sigma Factors 700
F. Molecular Chaperones 701
IV. Variation in Transcriptional Regulation and Adaptation to Changing Environments 701
V. Transposons and Genomic Rearrangements 702
VI. Circadian Clock and Gas Vesicle Proteins 702
VII. Inorganic Compounds as Reducing Power 703
VIII. Genomic Insights into the Photosynthetic Lifestyle 703
Acknowledgments 704
References 704

Summary

Genomes of several purple bacteria have recently been sequenced and many of these genomes have been fully annotated and are now available on either NCBI or/and other publicly accessible databases. This chapter gives a comparative analysis of three representative genomes, *Rhodobacter* (*Rba.*) *sphaeroides 2.4.1,Rhodospirillum* (*Rsp.*) *rubrum ATCC 11170, and Rhodopseudomonas* (*Rps.*) *palustris CGA009*. The results reveal that these three genomes possess some remarkable similarities, although these species differ in their genome architectures and the numbers and sizes of plasmids. The genomes of these three species encode two or multiple homologs (orthologs or paralogs) of many protein-coding genes representing a wide variety of metabolic pathways, which substantiate the enormous amount of metabolic versatilities shown by organisms belonging to the α-Proteobacteria. Paralogs are genes related by duplication within a genome, and therefore evolve new functions, even if these are related to the original one. In addition to abundant gene paralogs revealed by their genomes, the

*Author for correspondence, email: Samuel.Kaplan@uth.tmc.edu

C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer and J. Thomas Beatty (*eds*)*: The Purple Phototrophic Bacteria, pp. 691–706.*
© 2009 Springer Science + Business Media B.V.

above strains of these species displayed different numbers of sigma factors, transcriptional regulators (activators or repressors) and corresponding DNA binding motifs in their respective genomes. The diversity in these regulatory components indicates the use of different regulatory strategies for the genome-wide transcriptional regulation in these organisms. The different numbers of insertion sequences and transposases found among strains of these three species suggests that transposon-induced genomic variants may play a major role in strain differentiation within species.

I. Introduction

Several studies involving morphology, physiology, and genetics of purple bacteria have been conducted over many decades, and these studies have led to the demonstration that the primary metabolism of these organisms is a photosynthetic process requiring light and a low availability of oxygen. However the biochemical reaction of CO_2 reduction is carried out with reductant derived not from water as in green plants but from a variety of organic substances, as well as H_2 or hydrogen sulfide (van Niel 1944; also see Chapter 1, Madigan and Jung; Chapter 28, Romangnoli and Tabita; Chapter 39, Vignais). Also, members of this group can fix atmospheric nitrogen (Chapter 37, Masepohl and Kranz), and they can grow fermentatively.

Three representative species of purple photosynthetic bacteria, *Rba. sphaeroides 2.4.1, Rsp. rubrum ATCC 11170, and Rps. palustris CGA009*, which belong to the α-subgroup of Proteobacteria (Woese et al., 1984), are chosen for genome comparison as their genomes have been sequenced and the genomic sequence data are publicly available. The sequence of the complete genome of *Rps. palustris* CGA009 is available under GenBank/EMBL/DDBI accession numbers BX571963 (chromosome) and BX571964 (plasmid). The sequence of the *Rba. sphaeroides* 2.4.1 genome is available under NCBI accession numbers NC_007488 to NC_07494. The complete genome sequence of *Rsp. rubrum* ATCC11170 is under NCBI accession numbers NC_007643 (chromosome) and NC_007641 (plasmid).

These representative species along with other members of the α-Proteobacteria are widely distributed in nature and considered the most metabolically versatile group of organisms (Gest 1972). Recent biochemical and genetic studies of these bacteria have shown induced levels of expression of genes encoding the photosynthetic apparatus, and enzymes involved in photosynthesis or anaerobic respiration, under a reduced oxygen tension (Zeiltra-Ryalls et al., 1998; Grammel et al., 2003; Roh et al., 2004; Braatsch et al., 2006; Mackenzie et al., 2007; Chapter 35, Bauer et al.). Although multiple strains of *Rba. sphaeroides* (Choudhary et al., 2007) and *Rps. palustris* (http://img.jgi.doe.gov) have been recently sequenced, unless otherwise stated our genome comparison of these three species is based on a single representative strain selected from each of the three species (http://img.jgi.doe.gov): *Rba. sphaeroides* 2.4.1; *Rsp. rubrum* ATCC 11170; and *Rps. palustris* CGA009.

Abbreviations: ALA – 5-aminolevulinic acid; ALAS – 5 -aminolevulinic acid synthase; CDS – Coding sequence; CI – Large chromosome of *Rba. sphaeroides*; CII – Small chromosome of *Rba. sphaeroides*; IS – Insertion sequence; LH – Light harvesting complex; *P. – Pseudomonas*; PGC – Photosynthesis gene cluster; *Rba. – Rhodobacter*; RC – Reaction center ; *Rps. – Rhodopseudomonas*; *Rsp. – Rhodospirillum*

II. Genome Architecture and Characteristics

The genome characteristics of *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* are described in Table 1, and the original data of these three species are available on the publicly available website (http://img.jgi.doe.gov).

A. Genome Comparisons

The genome sizes of *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* are ~4.60, 4.40, and 5.46 Mb, respectively. The genomes of *Rba. sphaeroides* and *Rsp. rubrum* are of similar sizes, but the genome of *Rps. palustris* is 1.2 Mb larger than the genomes of the other two species. Besides the different genome sizes, the genome architecture of the three species also appears to be different. The genome of *Rba. sphaeroides* is comprised of two circular chromosomes, CI (large chromosome) and CII (small chromosome), and five plasmids (Suwanto and Kaplan, 1989b). The sizes of CI and CII of *Rba. sphaeroides* are 3,188,631 and 943,022 bp, respectively. In contrast the genomes of *Rsp. rubrum* and *Rps. palustris* consist of only one

Table 1. Comparison of Genome Characteristics

Genome characteristics	*Rba. sphaeroides* 2.4.1	*Rsp. rubrum* ATCC 11170	*Rps. palustris* CGA009
Total number of nucleotides	4,603,060	4,406,557	5,467,640
% DNA coding	88.28	88.93	87.90
% GC content	68.79	65.38	65.03
DNA scaffolds	7 (2 chromosomes) CI: 3,188,631 bp and CII: 943,022 bp)	2 (1 chromosome)	2 (1 chromosome)
Number of genes	4367	3917	4895
Protein coding genes	4304 (98.56%)	3850 (98.29%)	4838 (98.84%)
RNA genes	63	67	57
rRNA genes	9	12	8
5S rRNA	3	4	2
16S rRNA	3	4	2
23S rRNA	3	4	2
tRNA	54	55	49
Genes with function prediction	3038 (69.57%)	2861 (73.04%)	3406 (69.58%)
Genes in ortholog clusters	4115 (94.23%)	3627 (92.60%)	4723 (96.49%)
Genes in paralog clusters	2136 (50.33%)	1972 (50.34%)	2712 (55.40%)

Source of data: http://img.jgi.doe.gov, *Rba. sphaeroides* (version 1.6), and *Rps. palustris* and *Rsp. rubrum* (version 2.1).

chromosome and a plasmid. Several strains of *Rsp. rubrum* possess a very similar size ~55 kb plasmid with very little sequence divergence, but these are not identical and the plasmid is required for photosynthesis (Kuhl et al., 1983). Thus, similar plasmids seem to be disseminated throughout strains of *Rsp. rubrum*. Similarly, the genome of *Rps. palustris* contains one chromosome and a single plasmid of ~8.4 kb (Larimer et al., 2004). Besides the size variation of plasmids (8–15 kb) among strains of *Rps. palustris*, the plasmid replicates in closely related species of *Bradyrhizobium,* but not in *Rba. sphaeroides* and *Rhizobium* species. Thus, the origin of this plasmid has been thought to be from one of the strains of *Bradyrhizobium* (Giraud et al., 2000). Despite its large genome size, the genome of *Rps. palustris* contains a single chromosome, which also suggests that genome size is not the sole factor for genome complexity, including the existence of multiple chromosomes in bacteria.

Genomes of these three species exhibit remarkable similarities in gross genome structure. The total number of predicted genes in *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* are 4367, 3917, and 4897, respectively, which reflects their relative genome sizes. Of the total annotated genes, ~98% represent open reading frames, which potentially encode protein functions in all three species. Of these protein coding genes, ~70% can be assigned predicted functions, and >90% of these genes are classified in ortholog clusters in each of the three species. Orthologs are genes that evolve from a common ancestral gene by speciation. Since orthologs conserve the same function through evolution, identification of orthologs is valuable for prediction of gene function in newly sequenced genomes. Also, the coding capability of each of the three genomes is approximately 88%.

The genomes of different bacterial species in varied habitats reveal a narrow variation of their G+C compositions, which indicates that primarily, global environmental factors as opposed to different lifestyles of the microorganisms influence the G+C composition of a microbial community (Foerstner et al., 2005). The variation in G+C composition of different species is either maintained by selective pressure or caused by mutational bias. The observed G+C differences have a direct impact on the amino acid composition of proteins in the organisms living in their respective environments (Bharanidharan et al., 2004). The G+C composition of the genomes of *Rsp. rubrum* and *Rps. palustris* is ~65%, which is ~3% lower than the ~68% G+C content of the genome of

Rba. sphaeroides. The differences in G+C content exist across species, however there is a tendency of large genomes to be G+C-rich and small genomes to be A+T-rich (Glass et al., 2000; Moran, 2002; Rocha and Danchin, 2002). Bacteria with large genomes are found in more complex and diverse environments, and therefore G+C content may possibly be reflective of their niche complexity (Rocha and Danchin, 2002; Moran, 2002). Since the nucleotide composition of the three genomes discussed in this chapter is G+C-rich, it may correlate with a similar distribution of these species in diverse environments. A higher G+C composition of *Rba. sphaeroides* is possibly related to its genome architecture, as the genome of *Rba. sphaeroides* possesses two chromosomes, but this issue remains unsettled.

B. Evolutionary Relationships and the Origin of Mitochondria

Despite the gene level similarities, genome data also suggest significant evolutionary differences among these three bacteria. Based on 16S rDNA sequences, *Rba. sphaeroides* is part of a distinct clade within the α-3 sub-group of Proteobacteria. The group of α-Proteobacteria also contains organisms that associate with eukaryotes, like Rhizobacteria (essential for N_2 fixation in legumes), Agrobacteria (plant pathogens), and Rickettsia (intracellular animal pathogens) (Woese, 1987). The ability of α-Proteobacteria to associate with eukaryotes and their diverse metabolic activities led to the proposal that they are evolutionary ancestors of mitochondria (Woese, 1987). The current thought is that the species belonging to the order Rickettsiales of α-Proteobacteria are the closest relatives of mitochondria (Andersson et al., 1998; Gray et al., 1999; Emelyanov, 2001). More than 150 nucleus-encoded mitochondrial proteins of *Saccharomyces cerevisiae* share significant sequence homology with *Rickettsia prowazekii* proteins, and also several gene clusters in the mitochondrial genome are reminiscent of those found in *Rickettsia*, including a similar repertoire of proteins involved in ATP production, and transport (Andersson et al., 1998). However, the relationship of mitochondria to the *Ricketsiales* is limited to some conserved indels (signature protein sequences), and the closest relationship of mitochondria was seen for *Rsp. rubrum* rather than *Ricketsiales* (Esser et al., 2004). Thus, additional data are required to resolve the above conflicting observations.

C. Continual Evolution of Purple Bacteria

Several hypotheses have been proposed to explain the complex patterns of sequence relationships observed in microbial genomes. The continual 'horizontal transfer' hypothesis suggests that gene acquisitions are ongoing processes in microorganisms (Jain et al., 2002), whereas the early 'massive transfer' hypothesis proposes that massive exchanges occurred early in prokaryotic evolution, long before the diversification of modern bacterial species (Woese, 1998). Indeed, the presence of photosynthetic and non-photosynthetic species among α-Proteobacteria has been used to propose a continual evolution of traits acquiring by horizontal gene transfer in this group (Woese, 1987). Although the Proteobacteria are considered as the earliest lineage among the photosynthetic prokaryotes (Xiong et al., 2000), the evidence suggests that the Cyanobacteria constitute one of the earliest prokaryotic photosynthetic lineages that existed ~2.5 billion years ago. Thus, photosynthetic purple bacteria could have independently evolved or acquired photosynthesis gene clusters from one of the members of Cyanobacteria. Therefore, it is possible that non-photosynthetic purple bacteria may have evolved by losing the photosynthesis gene clusters from their ancestor lineages. Indeed, the notion of continual evolution among α-Proteobacteria is reinforced by genome-wide comparisons of *Rba. sphaeroides*, *Rsp. rubrum* and *Rps. palustris.* Although more than 40 bacterial species, which belong to diverse groups of bacteria, are known to have multiple chromosome-like replicons, most of the species containing multiple chromosomes have been found within the α-subgroup of Proteobacteria (Mackenzie et al., 2004). Among 25 isolates examined, the presence of two chromosomes is found in all these strains of *Rba. sphaeroides* (Nereng and Kaplan, 1999). The presence of two chromosomes in *Rba. sphaeroides* was originally proposed from pulse-field gel electrophoresis, but it has been confirmed by either optical mapping or genome sequence analysis of several isolates of this species (Choudhary et al., 2007; T. Donohue unpublished). In addition, by comparing genome sequence-derived restriction maps or optical maps of *Rba. sphaeroides* (Zhou et al., 2003), *Rsp. rubrum* (Reslewic et al., 2005) and *Rps. palustris,* no large regions of genome similarity except the photosynthesis gene cluster (PGC) are found between these three α-Proteobacteria. The total length of conserved regions among *Rba. sphaeroi-*

des, *Rsp. rubrum,* and *Rps. palustris* is ~330 kb of DNA that span over 107 common collinear blocks (Choudhary and Kaplan, unpublished). The majority of common DNA blocks consist of an average ~3 kb of DNA length. Also, the overall nucleotide similarity over all common collinear blocks was only ~25% among these three species. This is unlike the finding of significant regions of conservation among many *Rba. sphaeroides* isolates (Choudhary et al., 2007), and it is probably not surprising given the various activities and habitats described for these and other photosynthetic α-Proteobacteria (Woese 1987; also see Chapters 1, Madigan and Jung; Chapter 3, Yurkov and Csotonyi).

A comparison of the coding sequences (CDSs) of the *Rsp. rubrum* and *Rps. palustris* genomes when used to query an *Rba. sphaeroides* protein database indicates that each has a similar number of genes that are orthologous to *Rba. sphaeroides* genes (Fig. 1). Each has approximately twice the percentage of polypeptides with high quality matches (low P-values) to *Rba. sphaeroides* as compared to the *E. coli* outlier. The genomes of both *Rsp. rubrum* and *Rps. palustris* also encode relatively few polypeptides with high P-values (poor quality matches) whereas a larger proportion of the *E. coli* CDSs fall into this class. This is perhaps not surprising, since the three phototrophic bacteria are α-proteobacteria and *E. coli* is a γ-Proteobacterium. Interestingly, the *Rsp. rubrum* and *Rps. palustris* genomes have only approximately half of the high quality CDS matches (~2% of the genome having P-values $<10^{-200}$), as does *Rba. sphaeroides* to members of the marine *Roseobacter* clade (Moran et al., 2004) or *Paracoccus denitrificans* (http://genome.ornl.gov) (~5% of the genome having P-values $<10^{-200}$). This suggests that although *Rsp. rubrum*, *Rps. palustris,* and *Rba. sphaeroides* are similar because of their photosynthetic properties, they are in fact quite distantly related when the marine Roseobacters or the non-photosynthetic *Paracoccus*

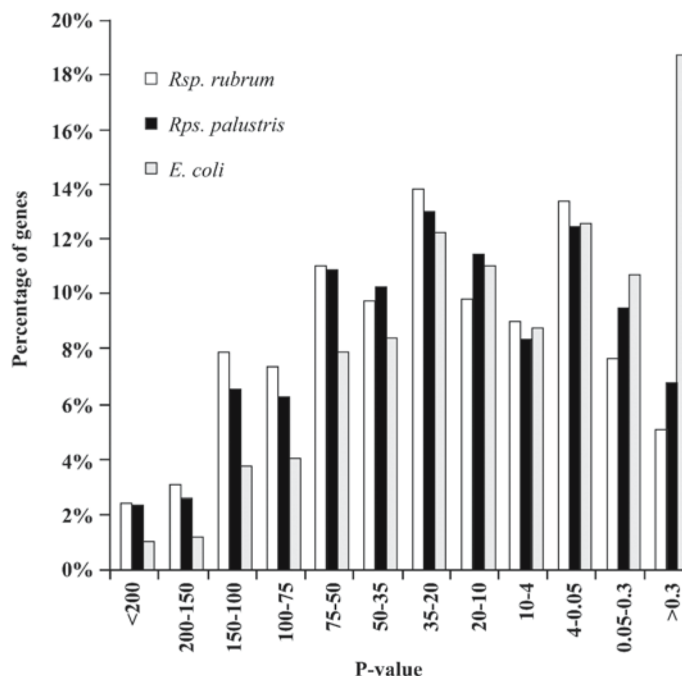

Fig. 1. The CDSs from the genomes of *Rps. palustris* CGA009 and *Rsp. rubrum* ATCC 11170 and E. coli (for comparison) were run in a WU-BLAST 2.0 (W. Gish, personal communication) search using BLASTP mode against an *Rba. sphaeroides* 2.4.1 CDS database. The output of each BLASTP search was extracted to give the top hit found in the *Rba. sphaeroides* genome. The top hits were then combined into a list, then sorted on the basis of their P-value. The percentage of genes in each genome falling within the following P-value ranges; $<10^{-200}$, $10^{-200-150}$, $10^{-150-100}$, $10^{-100-75}$, 10^{-75-50}, $10–5^{0-35}$, 10^{-35-20}, 10^{-20-10}, 10^{10-4}, 10^{-4}–0.05, 0.05–0.03 and >0.3 was then determined (x-axis). To give an indication of the conversion of P-value to amino acid identity/similarity, consider the following two examples; a polypeptide of 275 amino acid (aa) residues having a good match, 73% identity (85% similar) to a polypeptide in the *Rba. sphaeroides* database had a P-value of $5.7e^{110}$, whereas a borderline match was a polypeptide of 219 aa residues with 29% identity (44% similar) having a P-value of $5.4e^{-18}$. An imaginary cutoff for relevant matches lies approximately in the 10^{-20-10} range.

denitrificans are added into the analysis (Mackenzie et al., 2007).

III. Gene Homologs and Metabolic Versatility

A gene related to another gene within or between genomes evolves by descent from a common ancestral DNA sequence and is defined as a homolog. Homologs are classified into two families, orthologs and paralogs (Fitch, 2000). Orthologs represent homologous genes in different species that evolved from a common ancestral gene by the speciation event, which leads to the evolution of two different lineages. Most orthologs retain the same physiological function across species in the course of evolution. In contrast, gene paralogs are related by a gene duplication event followed by sequence divergence of these duplicated genes within a genome, and therefore paralogs evolve new functions.

Gene paralogs are abundantly found in bacterial genomes, and thought to descend from sequence duplications (Gevers et al., 2004). Genome analysis of *Rba. sphaeroides, Rps. palustris* and *Rsp. rubrum* reveals ~50% of the genes in paralog clusters as shown in Table 1. Although the numbers of rRNA and tRNA genes in all three genomes are similar, the genomes of *Rba. sphaeroides* and *Rsp. rubrum* contain a higher number of these genes than *Rps. palustris* as shown in Table 1. The genomes of *Rps. palustris*, *Rba. sphaeroides*, and *Rsp. rubrum* encode 2, 3, and 4 copies of 16S and 23S rRNA genes, respectively. *Rba. sphaeroides* and *Rsp. rubrum* contain 54 and 55 tRNA genes, respectively, which are higher than the 49 tRNA genes identified in the *Rps. palustris* genome.

Genome analyses of *Rba. sphaeroides* revealed numerous homologous genes representing many metabolic pathways, such as flagellum biosynthesis, photosynthesis, carbon metabolism, and chemosensory pathways (Mackenzie et al., 2001; Choudhary et al., 2004). The majority of gene homologs found in the *Rba. sphaeroides* genome displayed less similarity with each other than to orthologs from species closely related to *Rba. sphaeroides.* Thus, the majority of the gene homologs in *Rba. sphaeroides*, for example *rdxA* and *rdxB*, and *hemA* and *hemT*, are thought to be very old and possibly occurred prior to the evolution of the *Rba. sphaeroides* lineage (Choudhary et al., 2004), and therefore these gene homologs are orthologs, which descend from speciation events. Of the sample of genes listed in Table 2, many homologs exist as paralogs in the genomes of all three species, *Rba. sphaeroides*, *Rsp. rubrum,* and *Rps. palustris*.

A. Flagellum Biosynthesis

In *Rba. sphaeroides*, there are two sets of flagellar biosynthesis genes on CI, albeit the second set is incomplete. The first set of complete flagellar genes (RSP0032-RSP0084) is responsible for the synthesis and rotation of the subequatorial flagellum while the functions of the second set of incomplete flagellar genes (RSP1302-RSP1330) are not known. In contrast, the genome of *Rsp. rubrum* encodes a single operon containing genes for flagellum formation on its chromosome. However, the genome of *Rps. palustris* encodes a single copy of the genes required for flagellum formation, except *flgC*, *flgE*, and *flgG*, which are encoded in gene paralogs.

The *Rba. sphaeroides* flagellar genes in the first operon are expressed in all growth conditions examined, such as aerobic, semiaerobic, and photosynthetic (J. Roh, unpublished results). Although microarray expression of the second set of *Rba. sphaeroides* genes was undetected in any of the growth conditions tested (J. Roh, unpublished), and therefore the functions of these genes are unknown, the finding of two copies of most of the structural genes for flagellum formation is very significant. The cluster of the second set of genes in *Rba. sphaeroides* appears to be horizontally acquired, and these genes were shown to coexist with the endogenous (first) set of flagellar genes (Poggio et al., 2007). It was speculated that the second set of genes for flagellum biosynthesis in *Rba. sphaeroides* could be required for surface translocation using lateral flagella during biofilm production or another hypothetical type of alternative life style of this organism, as has been shown in other organisms (Capdevila et al., 2004; Kanbe et al., 2007). Bacterial cell motility in other organisms is very important in colonization since non-motile mutants of *Pseudomonas* (*P.*) *fluorescens* are severely impaired in root colonization (Capdevila et al., 2004). It has been noticed that the predominance of flagella variants with enhanced surface motility reach the distal part of the rhizosphere, which is not easily traveled by the wild type strains of *P. fluorescens* (Sanchez-Contreras et al., 2002). It remains to be seen if the second set of flagellar genes of *Rba. sphaeroides* has strong homology with the flagellar genes of organisms such

Table 2. List of gene homologs and number of genes in functional classes

Metabolic pathways/Genes	Number of genes encoded in genomes of		
	Rba. sphaeroides	*Rsp. rubrum*	*Rps. palustris*
Flagellum biosynthesis			
motA	2	1	1
flgA	2	1	1
flgC	1	1	2
flgE	2	1	2
flgG	1	1	2
flgH	2	1	1
flgI	2	1	1
flgL	2	1	1
fliE	2	1	1
fliF	2	1	1
fliG	2	1	1
fliH	2	1	1
fliM	2	1	1
fliQ	2	1	1
fliR	2	1	1
flhB	2	1	1
Chemotaxis			
cheA	3	3	3
cheB	2	3	2
cheY	10	21	3
cheW	4	4	4
cheR	3	4	4
Photosynthesis			
pucA	2	2	6
pucB	2	2	8
pucC	1	2	1
CO_2 fixation			
cbbA	2	1	1
cbbF	2	1	1
cbbP	2	1	1
cbbL	2	1	2
cbbG	3	1	1
cbbT	2	2	2
dxs	2	2	1
Energy production (Extra ATPase subunit)			
nuoA	2	1	1
nuoB	2	2	2
nuoC	2	2	2
nuoD	2	2	2
nuoE	2	1	2
nuoF	2	1	2
nuoG	2	1	2
nuoH	2	2	2
nuoI	1	1	2
nuoJ	2	1	2
nuoK	1	2	2
nuoL	1	1	2
nuoM	2	2	2
nuoN	1	2	2
atp (b subunit)	2	2	2

Table 2. Continued

Metabolic pathways/Genes	Number of genes encoded in genomes of		
	Rba. sphaeroides	*Rsp. rubrum*	*Rps. palustris*
DNA Replication and partitioning			
parA	4	1	3
repA	3	3	1
parB	7	1	3
repB	2	1	1
dnaE	2	1	2
dnaN	1	3	2
Amino acid biosynthesis			
trpB	2	2	2
glnA	5	2	4
Lipid metabolism			
Ech (Enoyl-CoA hydratase)	6	6	23
Heat shock proteins			
groES	1	2	2
groEL	3	2	2
Tetrapyrrole biosynthesis			
hemA/hemT	2	2	2
hemN	2	1	1
Sigma factors			
rpoN	4	1	1
rpoH	2	1	1
Transcription regulators			
DoeR family	4	6	1
LysR family	23	26	27
LuxR family	13	6	11
MarR family	5	10	20
ArsR family	6	6	9
AsnC family	10	10	5
Crp family	8	5	13
GntR family	27	12	13
MerR family	6	5	3
TetR family	9	18	39
IclR Family	2	2	6
AraC Family	10	18	23
Helix-turn-helix, Fis type	1	16	8
Helix-turn-helix, CopG family	2	10	2
Histidine kinases	29	54	64
Response regulators	51	47	70
Sigma factors	17	17	19
Others			
Transposases	34	26	16
Dehydrogenases	195	111	194

Source of data: (http://img.jgi.doe.gov; Choudhary et al., 2004; Larimer et al., 2004).

as *P. fluorescens,* where the functions of these genes have been implicated for root colonization. Flagellar motility and behavioral responses are discussed in this book elsewhere (Chapter 32, Armitage).

All three of *Rba. sphaeroides*, *Rsp. rubrum* and *Rps. palustris* possess a number of gene paralogs in chemosensory pathways. There are 21 copies of *cheY* (chemotaxis-specific response regulator) in *Rsp. rubrum*, which is twice the number of *cheY* paralogs (10 copies) present in the *Rba. sphaeroides* genome.

However, the genome of *Rps. palustris* encodes only 3 copies of *cheY*. Thus, *Rsp. rubrum* and *Rba. sphaeroides* display a high level of *cheY* diversity compared to *Rps. palustris,* which may allow the former two species to exploit more diverse environments. However, the number of methyl accepting chemotaxis genes, such as *cheA*, *cheB*, and *cheR* moderately vary (2 to 4 copies) in the three species, as shown in Table 2. Thus, methyl accepting chemotaxis genes, being the best indicators of the ability of a bacterium to exploit diverse environments, do not indicate great difference in these three species.

B. CO_2 Utilization and Photosynthesis

The purple photosynthetic bacteria serve as model organisms for the study of autotrophy and its relationship to photosynthesis. This group of organisms can grow in different growth conditions that use the Calvin cycle for CO_2 assimilation. CO_2 fixation and carbon metabolism are described elsewhere in this book (Chapter 28, Romangnoli and Tabita). Genes encoding enzymes and regulatory proteins of the Calvin cycle have been identified from a variety of bacteria, but the organization and the regulation of these genes are known in detail in only some organisms, including *Rba. sphaeroides*. Gene paralogs in the pathway of Calvin cycle are more abundant in *Rba. sphaeroides* than in *Rsp. rubrum* and *Rps. palustris*. For example, there are two homologs of *cbbA*, *cbbF*, *cbbP*, *cbbL*, *cbbT*, and *dxs*; and three homologs of *cbbG* in *Rba. sphaeroides* as shown in Table 1. However, *Rsp. rubrum* possesses two homologs of *cbbT* and *dxs*, while the genome of *Rps. palustris* contains paralogs of *cbbT* and *cbbL*, but on the other hand *Rsp. rubrum* does not have paralog of *cbbL*.

Photosynthesis is the essential characteristic of this group of organisms. Most of the genes involved in photosynthesis are located in a photosynthesis gene cluster (PGC) that contains reaction center (RC), light-harvesting complexes, and bacteriochlorophyll and carotenoid biosynthesis genes. The *pucBA* structural genes encode the apoproteins of the light-harvesting complex, LH2 (B800-850) β- and α-polypeptides. Paralogs of *pucBA* are located about 1.36 Mb apart on CI of *Rba. sphaeroides*. The *puc2B*-encoded polypeptide is almost identical to the *puc1B*-encoded polypeptide. However, the N-terminal 48 amino acid residues of *puc2A* and *puc1A* exhibit a lower level of sequence conservation, and it is suspected that the *puc2A* gene is a pseudogene or/and may be in the process of acquiring a new function (Zeng et al., 2003). The expressions of these two *puc* operons in *Rba. sphaeroides* are both oxygen and light dependent. In contrast, the genome of *Rsp. rubrum* lacks *pucBA* (Bérard et al., 1986; Hessner et al., 1991), while the genome of *Rps. palustris* encodes four complete sets and one incomplete set of LH2 genes (Larimer et al., 2004). Two of the four complete sets of *pucBA* genes are located adjacent to the bacteriophytochrome genes that may regulate LH2 complex gene expression. However, only one *pucBA* operon is located in the PGC of *Rba. sphaeroides*. The high number of LH2 (*pucBA*) genes appear to encode different types of LH2 complexes in *Rps. palustris*, which help to adapt *Rps. palustris* in harvesting different wavelengths of light (Hartigan et al., 2002; Tharia et al., 1999; Chapter 40, Evans et al.). Thus, *Rps. palustris* appears to be more versatile for photosynthetic growth. The *pufBA* genes, which encode LH1 polypeptides, are highly conserved in *Rba. sphaeroides*, *Rps. palustris* and *Rsp. rubrum*. In addition, the genes encoding the RC L, M, and H polypeptides of *Rps. palustris* are highly conserved and are 45 to 60% similar at the amino acid sequence level to the homologs of *Rba. sphaeroides*, but these genes are more closely related to *Bradyrhizobium* ORS278 (Giraud et al., 2000), and therefore these three genes (*pufL*, *pufM*, and *pufH*) may have been acquired by horizontal transfer from *Bradyrhizobium*, or vice versa.

C. Energy Production

The NADH: ubiquinone oxidoreductase (complex I), an enzyme of the photosynthetic as well as aerobic and anaerobic respiratory chains (Hirst, 2005), consists of 13 subunits, NuoA through NuoN. The genomes of *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* encode two *nuo* operons, one of which contains an incomplete set of *nuo* genes. Many of these *nuo* genes were identified as paralogs in these species as shown in Table 2, however their functions are not yet known. Expression of the *nuo* genes in *P. fluorescens* is increased under low oxygen tension, and their levels of expression vary during different growth phases (Camacho Carvajal et al., 2002). Mutation in the *nuo* gene(s) in *P. fluorescens* has been shown to impair the root-colonization ability in tomato rhizosphere, due to the low oxygen availability (Camacho Carvajal et al., 2002). A better understanding of the regulation of the paralogs of *nuo* genes could possibly indicate

whether *Rba. sphaeroides*, *Rsp. rubrum* or *Rps. palustris* is involved in a symbiotic association involving the gene products of the nuo genes.

Enzymes in the F_0F_1 family of proton-translocating ATPases and ATP synthases are essential in all bacterial growth conditions. The general structure of ATP synthase is highly conserved (Borghese et al., 1998a,b), and it was found that there are ten genes contained in two operons, *atpHAGDC* and *atpFXEBI*, in *Rba. capsulatus* and *Rba. sphaeroides*. This type of gene organization seems to be unique to other members of *Rhodospirillaceae*. All three representatives of the purple photosynthetic bacteria examined in this chapter possess a total of these 10 genes, of which eight genes encode essential subunits of the ATP synthase. In addition to these essential subunits, *atpX* and *atpI* encode for b′ and I subunits, respectively. While the function of the *atpI* gene remains unknown, a duplicated and divergent copy of the b subunit gene (b′ subunit) was identified in all three species and its function was linked to proton transport. In *Cyanobacteria*, it has been shown that when an additional b subunit gene is present, a b_2 heterodimer (bb′) is formed (Dunn et al., 2001), that could lead to different rates of proton transport (Turina et al., 2006). Since the duplication of a divergent b subunit gene was found only in the photosynthetic prokaryotes and the plant chloroplast (Borghese et al., 1998b), the role of gene duplication has been implicated in an alternative life style or photosynthesis, which may possibly require different levels of protons per ATP hydrolyzed in order to adapt to a specific growth condition (Cross and Taiz, 1990).

Rba. sphaeroides possesses gene homologs, *rdxB* and *rdxA* (Neidle and Kaplan, 1992), encoding membrane bound ferredoxin-like proteins that are homologs of the *Rhizobium meliloti fixG*. RdxB is involved in photosynthesis gene expression under aerobic growth, carotenoid biosynthesis, and the expression of the cbb_3 cytochrome oxidase. These findings suggest that *rdxB* along with other genes in this cluster (*rdxBHIS*) on CI are part of the same signal transduction pathway, and might control directly or indirectly the redox state of the cell. In contrast, the other copy (*rdxA*) on CII is expressed in a number of growth conditions (Pappas et al., 2004), but no specific function has been determined. A homolog of *rdxBHIS* of *Rba. sphaeroides* is also present in the genome of *Rps. palustris*, but no such homolog is found in *Rsp. rubrum*. Therefore, *rdxBHIS* genes are not essential effectors of photosynthesis, but they may possibly modulate photosynthesis gene expression differentially in the environments where both *Rba. sphaeroides* and *Rps. palustris* are naturally established.

D. Tetrapyrrole Biosynthesis

The enzyme 5-aminolevulinic acid synthase (ALAS) catalyzes the first and rate-limiting step in a branched pathway for tetrapyrrole biosynthesis, which is discussed thoroughly elsewhere in this book (Chapter 39, Zeilstra-Ryalls). There are two ALAS genes, *hemA* and *hemT* in the *Rba. sphaeroides* genome, localized on CI and CII, respectively (Neidle and Kaplan, 1993a; 1993b). Both *Rsp. rubrum* and *Rps. palustris* encode homologs of the gene for ALAS on their chromosomes. It has been demonstrated that both *hemA* and *hemT* contribute to the overall cellular levels of ALA in *Rba. sphaeroides*, however the mRNA from *hemT* was not detected (Zeilstra-Ryalls and Kaplan, 1995). In addition to the difference in gene regulation and possibly the biochemical properties of the two isozymes, their cellular localizations are thought to be different; one is localized in the cytoplasm and other appears to be in the membrane bound fraction (Fanica-Gaignier and Clement-Metral 1973; Bolt et al., 1999). It remains to be examined whether paralog copies of ALAS genes also follow the same pattern of expression in *Rsp. rubrum* and *Rps. palustris*.

E. Sigma Factors

Prokaryotic transcription initiation requires different types of sigma (σ) factors, which are activated in response to different environmental conditions. The genomes of *Rsp. rubrum* and *Rps. palustris* contain a single σ^{54} homolog encoded by the *rpoN* gene. However, the *Rba. sphaeroides* genome contains four σ^{54} homologs encoded by $rpoN_I$, $rpoN_{III}$ and $rpoN_{IV}$, located on CI, and $rpoN_{II}$ located on CII. $rpoN_I$ is located within the *nif* gene cluster containing genes for nitrogen fixation (Suwanto and Kaplan, 1989a). Deletion of $rpoN_I$ resulted in reduced diazotrophic growth of *Rba. sphaeroides* (Meijer and Tabita, 1992). In contrast, $rpoN_{II,}$ is not involved in either diazotrophic growth or in the synthesis of nitrogenase activity (Smith and Kaplan, unpublished). The $rpoN_{III}$ gene lies ~100 bp downstream of *fliC* and might be involved in flagellar biosynthesis and/or cell motility (Mackenzie et al., 2001). The physical proximity of

$rpoN_{III}$ and *fliC* suggests their involvement in flagellum formation as in other bacterial species. The negative control of flagellum formation in *Pseudomonas* species was seen to be mediated by either structural proteins of the flagella or regulatory proteins such as *RpoN*, *RpoF* (FliA), or/and FleQ (Kieboom et al., 1998). In *Pseudomonas aeruginosa*, both copies of the *rpoN* gene, which encode the alternative sigma factor, σ^{54}, as well as *fliC*, encoding the flagellar subunit protein flagellin, were found to be essential for flagellum formation and motility (Garrett et al., 1999). There is no obvious operon near the *Rba. sphaeroides* $rpoN_{II}$ or $rpoN_{IV}$ that could provide a clue as to their possible function(s).

Similarly, two *rpoH* genes encoding σ^{32} homologs are present in the *Rba. sphaeroides* genome, but found only in single copy in both *Rsp. rubrum* and *Rps. palustris*. Each of these two alternative sigma factors ($RpoH_{I}$ and $RpoH_{II}$) may contribute to the heat-shock response since cells lacking only one of these proteins are able to grow at elevated temperatures (Karls et al., 1998; Green and Donohue, 2006). However, $RpoH_{II}$ expression appears to be dependent on the extracytoplasmic function of sigma factor, σ^{E} (Anthony et al., 2005), and it was suggested that the role of $RpoH_{II}$ is tightly linked to the singlet oxygen stress response. In contrast, cells lacking $RpoH_{I}$ are more sensitive than wild type cells to oxyanions (Karls et al., 1998), suggesting a role for this protein in another potential stress pathway.

F. Molecular Chaperones

Molecular chaperones are commonly called heat-shock proteins, which help protein folding. Chaperones exist in many different families and they are expressed under conditions of high stress. The GroEL/GroES complex (HSP60), well-characterized in *E. coli*, is a large (~1 MDa) chaperone in bacteria. The genome of *Rba. sphaeroides* contains two *groESL* operons, one of which encodes *groES* and the other *groEL* genes. Transcription of the *groESL1* genes was observed to be low under photoautotrophic growth conditions. However, no transcript was detected for the *groESL2* operon under growth conditions examined so far (Smith and Kaplan, unpublished). Seemingly, *Rsp. rubrum* and *Rps. palustris* also possess one each of *groES* and *groEL*, but their functions are not yet determined.

IV. Variation in Transcriptional Regulation and Adaptation to Changing Environments

Transcriptional regulation serves the primary mechanisms of adaptation in prokaryotes. Bacterial cells sense environmental signals using sensor kinases, and typically transmit these signals by phosphorylation of response regulators, which may activate or repress transcription of specific genes. *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* possess 80, 101, and 134 signal transduction genes (encoding histidine kinases and response regulators), respectively. The greater number of signal transduction genes in *Rps. palustris* may possibly provide this species the ability to control many different physiological processes involving transcriptional regulators. The networks of transcriptional controls are discussed elsewhere in this book (Chapter 35, Bauer et al., Chapter 36, Klug and Masuda), therefore only a brief overview of diverse families of regulatory proteins found in these species is provided here.

The families of transcription regulatory proteins (Martinez-Bueno et al., 2004), such as AsnC, AraC, Crp, GntR, LysR, LuxR, TetR and the number of paralogs encoded in the *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* genomes are listed in Table 2. Many of these proteins are predicted to contain multiple domain motifs (Pareja et al., 2006). Since LysR regulatory proteins act as either activator or repressor, which are commonly found in bacteria and archaeae, it is not surprising that all three species possess a comparable number of LysR in their genomes. LysR regulatory proteins induce gene expression involved in multiple pathways, including pathogenicity, and biofilm production in other species (Kovacikova et al., 2005). However, their specific functions have not been yet determined in these three species.

The *Rba. sphaeroides* genome encodes 27 GntR-like regulatory repressors, which are also found in diverse groups of bacterial species. In contrast, *Rps. palustris* and *Rsp. rubrum* encode only half the number of GntR proteins as compared *Rba. sphaeroides*. GntR regulatory proteins have been shown to sense cellular signals, and stimulate antibiotic production and carbon utilization (Hillerich and Westpheling, 2006). Thus, the possession of a variety of GntR regulators in these species may possibly play an important regulatory role in secondary metabolite production and carbon metabolism.

In comparison to *Rba. sphaeroides* and *Rsp. rubrum*, the genome of *Rps. palustris* encodes more

copies of MarR regulatory proteins. This class of regulator often negatively regulates the expression of antibiotic resistance genes. In *Rps. palustris*, the BadR protein, a member of the MarR family, positively regulates expression of genes for anaerobic benzoate degradation (Egland and Harwood, 1999). In addition, the genome of *Rps. palustris* possesses a much higher number of TetR and AraC regulatory proteins. TetR regulators are involved in transcriptional control of multidrug efflux pumps, antibiotic production, pathogenicity, and responses to osmotic stress (Ramos et al., 2005). Activators belonging to AraC family are involved in quorum sensing, signaling, and the production of virulence factors (Ramos, 2002). In addition to the high copies of LysR type transcriptional regulators, *Rsp. rubrum* also possesses moderately high copies of TetR, AraC, MarR, and AsnC families of transcriptional regulators.

Although the comparative analysis of gene sequences has been helpful in the prediction of the structure-function of proteins, the regulation of these genes depends on their upstream regions with different regulatory motifs. For example, DNA sequences predicted to bind helix-turn-helix motifs are more abundantly found in *Rsp. rubrum* than the same sequences found in the genomes of the other two species (M. Choudhary, unpublished). Therefore, the number of regulatory proteins found within each of these organisms and the corresponding variation in number of DNA binding motifs located in the respective genome, may possibly play an important role in different strategies for global transcriptional regulation in each of the three species.

V. Transposons and Genomic Rearrangements

Insertion sequences (IS) consist of a large group of bacterial mobile elements, and are known to comprise up to 2 to 5% of genomic DNA in most bacterial genomes (Mahillion et al., 1999; Chandler and Mahillion, 2002). IS elements encode transposases required for transposition that helps the bacterial genome to acquire accessory functions and antibiotic resistance genes. *Rba. sphaeroides* and *Rsp. rubrum* contain 34 and 26 transposase genes, respectively. However, *Rps. palustris* contains only seven transposase genes indicating a minor role of transposon-induced genomic variation in this species. In contrast to strain *Rps. palustris* CGA009, other strains of *Rps. palustris* (http://img.jgi.doe.gov) possess equivalent number of transposases as found in either *Rba. sphaeroides* 2.4.1 or *Rsp. rubrum* ATCC 11170. On the contrary to *Rba. sphaeroides* 2.4.1, another strain of *Rba. sphaeroides* (ATCC 17029) contains a reduced number of transposases in its genome (http://img.jgi.doe.gov). Members of the IS66 family are found in all three of the genomes considered here, and are widely distributed in Gram-negative bacteria, such as in *Agrobacterium*, *Rhizobium*, *Escherichia*, *Pseudomonas*, and *Vibrio* (Han et al., 2001).

Although *Rba. sphaeroides*, *Rsp. rubrum*, and *Rps. palustris* harbor similar families of transposons in their genomes, the numbers vary in different strains within species, and therefore transposon-generated functions possibly play an important role in strain differentiation or intra-species variation. This hypothesis has recently been supported by a comparative genome analysis of the three strains of *Rba. sphaeroides*, 2.4.1, ATCC 17029, and ACC 17025, which revealed a rapid nucleotide divergence between orthologs of CII-specific DNA sequences (Choudhary et al., 2007). The DNA divergence could be the result of rearranged DNA sequences mediated by insertion sequences, gene duplications or/and newly acquired genetic elements on CII. The rapid evolution of CII suggests that accumulation of genetic variants in CII may play a major role in strain differentiation in *Rba. sphaeroides*.

VI. Circadian Clock and Gas Vesicle Proteins

The existence of circadian clock genes, *kaiB* and *kaiC*, has been reported among the purple bacteria (Mackenzie et al., 2001; Larimer et al., 2004). In *Rba. sphaeroides*, both genes are located on a plasmid while in *Rsp. rubrum* and *Rps. palustris*, these two circadian clock genes are located on the chromosome. Although circadian clock genes have previously been identified in many cyanobacterial species (Johnson and Golden, 1999), the function of these genes in the anoxygenic photosynthetic bacteria has yet to be determined. These species generate enough energy by photophosphorylation during daylight hours, but an inadequate amount of sunlight is available at night. Therefore, these organisms may use other organic compounds as an energy source when sunlight is absent or limiting. Thus, circadian regulation of redox reactions such as nitrogen fixa-

tion or other physiological processes could require these time-keeping genes. For example, gas vesicle proteins are encoded on CII of *Rba. sphaeroides*. Surprisingly, gas vesicle proteins are not encoded in the genome of either *Rsp. rubrum* or *Rps. palustris*. We speculate that gas vesicles may possibly play an important role in the flotation of *Rba. sphaeroides* at low light intensities or sinking of the cells under intense light, in conjunction with day-night cycles. It remains to be seen whether the expression of gas vesicle genes and circadian clock genes are linked in *Rba. sphaeroides*. If so, *Rba. sphaeroides* and *Rps. palustris* may use different strategies to adjust to different light intensities.

VII. Inorganic Compounds as Reducing Power

In addition to carbon dioxide and nitrogen assimilation, many inorganic compounds such as thiosulfate, hydrogen gas, as well as organic compounds such as formate, serve as reductants for autotrophic growth. The genomes of *Rps. palustris*, *Rba. sphaeroides*, and *Rsp. rubrum* encode a cluster of 20 genes for the synthesis of a nickel-containing uptake hydrogenase (Larimer et al., 2004), which may provide reducing power for autotrophic growth.

Genes homologous to the *sox* genes encode cytochrome *c* oxidoreductases and are found in many sulfur-oxidizing organisms (Friedrich et al., 2001). Genomes of *Rba. sphaeroides*, *Rps. palustris*, and *Rsp. rubrum* encode *sox* genes, which may allow autotrophic growth by oxidation of reduced sulfur-compounds. In addition, the genomes of *Rba. sphaeroides*, *Rps. palustris*, and *Rsp. rubrum* encode carbon monoxide dehydrogenases and a formate dehydrogenase. This is not a surprising observation since these three species exist in diverse ecological niches where carbon monoxide and formate may be produced by either usual metabolism or/and photooxidation of atmospheric hydrocarbons (Conrad, 1988). Thus, it is possible that all three of these organisms may be able to oxidize CO and reduced sulfur compounds.

VIII. Genomic Insights into the Photosynthetic Lifestyle

These genome sequences have also provided new perspectives on the photosynthetic lifestyle of bacteria. As discussed in section III.B on CO_2 utilization and photosynthesis, *Rba. sphaeroides* and several other purple photosynthetic bacteria contain a ~65 kb region of the genome, which is known as the photosynthesis gene cluster (PGC) that houses many of the genes required for photosynthetic growth. Genes within this PGC encode enzymes for bacteriochlorophyll and carotenoid biosynthesis, the structural genes for pigment-binding proteins of the light-harvesting and RC complexes, electron carriers like cytochrome c_2, and conserved proteins of unknown function (Choudhary and Kaplan, 2000). This is a striking example of a large region with high conservation among the genomes of *Rba. sphaeroides*, *Rsp. rubrum* and *Rps. palustris*. Given what is known about the control of *Rba. sphaeroides* photosynthetic membrane assembly, it is not surprising that expression of the genes in this region respond to signals that regulate the synthesis of the photosynthetic apparatus (Roh et al., 2004). Genome-wide expression studies have also provided insight into new gene products or metabolic pathways, which are required for photosynthetic growth in *Rba. sphaeroides* (Tavano et al., 2005) and possibly other bacteria. Given the differences in gene expression that have been noted among other photosynthetic bacteria, it will be interesting to see if these patterns are conserved across genus and species lines.

Another genome-based approach to identify what is needed for bacterial photosynthesis has been based on obtaining protein blueprints of whole cells or subcellular fractions that contain the photosynthetic apparatus. For example, by determining the protein abundance and composition of either whole cells or purified cell fractions (cytoplasm, cytoplasmic membrane, periplasm, outer membrane), including the purified photosynthetic membrane (chromatophores) ~2000 proteins have been identified (Fejes et al., 2003; Callister et al., 2006a,b; VerBerkmoes et al., 2006). When analyzing the protein composition of *Rba. sphaeroides* chromatophores of known purity, ~100 proteins have been identified that were present or enriched in this fraction, including the light-harvesting, reaction center and bioenergetic enzymes (cytochrome bc_1 complex, F_1F_0 ATPase) that were known to be present in these photosynthetic membranes (Zeng et al., 2007). However, many chromatophore proteins are either of unknown function or are homologs of proteins that were not found in the photosynthetic apparatus. In contrast, many of the PGC proteins that are required for pigment biosynthesis were not found

in the purified chromatophores, suggesting that these pigments are synthesized elsewhere and subsequently inserted into intact pigment-protein complexes before or during insertion into the specialized domain of the inner membrane. Indeed, most of the chromatophore proteins are encoded by genes that lie outside the PGC (Fejes et al., 2003; Zeng et al., 2007). This illustrates the types of new information that can be derived from genome-enabled analysis of solar energy utilization in these and other bacteria.

Acknowledgments

This work was supported by: NIH grant GM15590-39 to SK; NIH grant GM075273 and DOE grant DE-FG02-05ER15653 to TJD; and the Department of Energy (DOE) grant DOE ER63232-1018220-0007203 to SK and TJD.

References

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH and Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–140

Anthony JR, Warczak K and Donohue TJ (2005) A transcriptional response to singlet oxygen, a toxic byproduct of photosynthesis. Proc Natl Acad Sci USA 102: 6502–6507

Bérard J, Bélanger G, Corriveau P and Gingras G (1986) Molecular cloning and sequence of the B880 holochrome gene from *Rhodospirillum rubrum*. J Biol Chem 262: 82–87

Bharanidharan DG, Bhargavi R, Uthanumallian K and Gautam N (2004) Correlation between nucleotide frequencies and amino acid composition in 115 bacterial species. Biochem Biophys Res Commun 315: 1097–1103

Bolt EL, Kryszak L, Zeilstra-Ryalls J, Shoolingin-Jordon PM and Warren MJ (1999) Characterization of the *Rhodobacter sphaeroides* 5-aminolaevulinic acid synthase isozynes, HemA and HemT, isolated from *Escherichia coli*. Eur J Biochem 265: 290–299

Borghese R, Crimi M, Fava L and Melandri BA (1998a) The ATP synthase *atpHAGDC* (F_1) operon from *Rhodobacter capsulatus*. J Bacteriol 180: 416–421

Borghese R, Turina P, Lambertini L and Melandri BR (1998b) The *atpIBEXF* operon coding for the F_0 sector of the ATP synthase from the purple nonsulfur photosynthetic bacterium *Rhodobacter capsulatus*. Arch Microbiol 170: 385–388

Braatsch S, Bernstein JR, Lessner F, Morgan J, Liao JC, Harwood CS and Beatty JT (2006) *Rhodopseudomonas palustris* CGA009 has two functional *ppsR* genes, each of which encodes a repressor of photosynthesis gene expression. Biochemistry 45:14441–14451

Callister SJ, Dominguez MA, Nicora CD, Zeng X, Tavano CL, Kaplan S, Donohue TJ, Smith RD and Lipton MS (2006a) Application of the accurate mass and time tag approach to the proteome analysis of sub-cellular fractions obtained from *Rhodobacter sphaeroides* 2.4.1. Aerobic and photosynthetic cell cultures. J Proteome Res 5: 1940–1947

Callister SJ, Nicora CD, Zeng X, Roh JH, Dominguez MA, Tavano CL, Monroe ME, Kaplan S, Donohue TJ, Smith RD and Lipton MS (2006b) Comparison of aerobic and photosynthetic *Rhodobacter sphaeroides* 2.4.1 proteomes. J Microbiol Methods 67: 424–436

Camacho Carvajal MM, Wijfjes AH, Mulders IH, Lugtenberg BJ and Bloemberg GV (2002) Characterization of NADH dehydrogenases of *Pseudomonas fluorescens* WCS365 and their role in competitive root colonization. Mol Plant Microbe Interact 15: 662–671

Capdevila S, Martinez-Granero FM, Sanchez-Contreras M, Rivilla R and Martin M (2004) Analysis of *Pseudomonas fluorescens* F113 genes implicated in flagellar filament synthesis and their role in competitive root colonization. Microbiology 150: 3889–3897

Chandler M and Mahillon J (2002) Insertion sequences revisited. In: Craig NL, Craigie R, Gellert M and Lambowitz AM (ed) Mobile DNA II, pp 305–366. ASM Press, Washington, DC

Choudhary M and Kaplan S (2000) DNA sequence analysis of the photosynthesis region of *Rhodobacter sphaeroides* 2.4.1. Nucl Acids Res 28: 862–867

Choudhary M, Fu YX, Mackenzie C and Kaplan S (2004) DNA sequence duplication in *R. sphaeroides* 2.4.1: Evidence of an ancient partnership between chromosome I and II. J Bacteriol 186: 2019–2027

Choudhary M, Zanhua X, Fu YX and Kaplan S (2007) Genome analysis of three strains of *Rhodobacter sphaeroides*: Evidence of rapid evolution of chromosome II. J Bacteriol 189: 1914–1921

Conrad R (1988) Biogeochemistry and ecophysiology of atmospheric CO and H_2. Adv Microb Ecol 10: 231–283

Cross R L and Taiz L (1990) Gene duplication as a means for altering H^+/ATP ratio during the evolution of F_0F_1 ATPases and synthases. FEBS Lett 259: 227–229.

Dunn SD, Kellner E and Lill H (2001) Specific heterodimer formation by the cytoplasmic domains of the b and b´ subunits of cyanobacterial ATP synthase. Biochemistry 40: 187–192

Egland PG and Harwood CS (1999) BadR, a new marR family member, regulates anaerobic benzoate degradation by *Rhodopseudomonas palustris* in concert with AadR and Fnr family members. J Bacteriol 181: 2102–2109

Emelyanov VV (2001) Evolutionary relationship of *Rickettsiae* and mitochondria. FEBS Lett 501:11–18

Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K, Kretschmann E, Richly E, Leister D, Bryant D, Steel MA, Lockhart PJ, Penny D and Martin W (2004) A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol Biol Evol 21:1643–1660

Fanica-Gaignier M and Clement-Metral J (1973) Cellular compartmentation of two species of 5-aminolevulinic acid synthetase in a facultative photohetrotrophic bacterium (*Rps. sphaeroides Y*). Biochem Biophys Res Comm 55: 610–615

Fejes AP, Yi EC, Goodlett DR and Beatty JT (2003) Shotgun proteomic analysis of a chromatophore-enriched preparation from the purple phototrophic bacterium *Rhodopseudomonas*

palustris. Photosynth Res 78: 195–203

Fitch WM (2000) Homology: A personal view on some of the problems. Trends Genet 16: 227–231

Friedrich CG, Rother D, Bardischewsky F, Quentmeier A and Fischer J (2001) Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? Appl Environ Microbiol 67: 2873–2882

Foerstner, KU, von Mering C, Hooper, SD and Bork P (2005) Environments shape the nucleotide composition of genomes. EMBO Reports 6: 1208–1213

Garrett ES, Perlegas D and Wozniak DJ (1999) Negative control of flagellum synthesis in *Pseudomonas aeruginosa* is modulated by the alternative sigma factor AlgT (AlgU). J Bacteriol 181: 7401–7404

Gest H (1972) Energy conservation and generation of reducing power in bacterial photosynthesis. Adv Microb Physiol 7: 243–282

Gevers D, Vandepoele K, Simillion C and Deeper YV (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. Trends Microbiol 12: 148–154

Giraud E, Hannibal L, Fardoux J, Verméglio A and Dreyfus B (2000) Effects of *Braydirhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitive*. Proc Natl Acad Sci USA 97: 14795–14800

Glass JI, Lefkowitz, EJ, Glass JS, Heiner CR, Chen EY and Cassell GH (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature 407; 757–762

Grammel H, Gilles ET and Ghosh R (2003) Microaerophilic cooperation of reductive and oxidative pathways allows maximal photosynthetic membrane biosynthesis in *Rhodospirillum rubrum*. Appl Envir Microbiol 69: 6577–6586

Gray MW, Burger G and Lang BF (1999) Mitochondrial evolution. Science 283:1476–1481

Green HA and Donohue TJ (2006) Activity of *Rhodobacter sphaeroides* $RpoH_{II}$, a second member of the heat shock sigma factor family. J Bacteriol 188: 5712–5721

Han C, Shiga Y, Tobe T, Sasakawa C and Ohtsubo E (2001) Structural and functional characterization of IS679 and IS66-family elements. J Bacteriol 183: 4296–4304

Hartigan N, Tharia HA, Sweeney F, Lawless AM and Papiz MZ (2002) The 7.5-Å electron density and spectroscopic properties of a novel low-light B800 LH2 from *Rhodopseudomonas palustris*. Biophys J 82: 963–977

Hessner MJ, Wejksnora PJ and Collins ML (1991) Construction, characterization, and complementation of *Rhodospirillum rubrum puf* region mutants. J Bacteriol 173:5712–5722

Hillerich B and Westpheling J (2006) A new GntR family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. J Bacteriol 188: 7477–7487

Hirst J (2005) Energy transduction by respiratory complex I-an evaluation of current knowledge. Biochem Soc Trans 33: 525–529

Jain R, Rivera MC, Moore JE and Lake JA (2002) Horizontal gene transfer in microbial genome evolution. Theor Popul Biol 61:489–495

Johnson CH and Golden SS (1999) Circadian programs in cyanobacteria: Adaptiveness and mechanism. Annu Rev Microbiol 53: 389–409

Kanbe M, Yagasaki J, Zehner S, Gottfert M and Aizawa S (2007) Characterization of two sets of subpolar flagella in *Bradyrhizobium japonicum*. J Bacteriol 189: 1083–1089

Karls RK, Brooks J, Rossmeissl P, Luedke J and Donohue TJ (1998) Metabolic roles of a *Rhodobacter sphaeroides* member of the σ^{32} family. J Bacteriol 180: 10–19

Kieboom J, Dennis JJ, de Bont AMJ and Zylstra GJ (1998) Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. J Biol Chem 273: 85–91

Kovacikova G, Lin W and Skorupski K (2005) Dual regulation of genes involved in acetoin biosynthesis and motility/biofilm formation by the virulence activator AphA and the acetate-responsive LysR-type regulator AlsR in *Vibrio cholerae*. Mol Microbiol 57: 420–433

Kuhl SA, David WN and Yoch DC (1983) Characterization of *Rhodospirillum rubrum* plasmid: Loss of photosynthetic growth in plasmidless strains. J Bacteriol 156: 737–742

Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA, Beatty JT, Lang AS, Tabita FR, Gibson JL, Hanson TE, Bobst C, Torresy Torres JL, Peres C, Harrison FH, Gibson J and Harwood CS (2004) Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palutris.* Nature Biotechnol 22: 55–61

Mackenzie C, Choudhary M, Larimer FW, Predki PF, Stilwagen S, Armitage JP, Barber RD, Donohue TJ, Hosler JP, Newman JE, Shapleigh JP, Sockett RE, Zeilstra-Ryalls J and Kaplan S (2001) The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. Photosynth Res 70: 19–41

Mackenzie C, Kaplan S and Choudhary, M (2004) Multiple chromosomes. In: Miller RV and Day MJ (ed) Microbial Evolution, pp 82–101. ASM Press, Washington, DC

Mackenzie C, Eraso JM, Choudhary M, Roh JH, Zeng X, Brucella P, Puskás A and Kaplan S (2007) Post-genomic adventures with *Rhodobacter sphaeroides*. Annu Rev Microbiol 61: 283–307

Mahillion J, Leonard C and Chandler M (1999) IS elements as constituents of bacterial genomes. Res Microbiol Mol Biol Rev 150: 675–687

Martinez-Bueno M, Molina-Henares AJ, Praeja E, Ramos JL and Tobes R (2004) BactTregulators: A database of transcriptional regulators in bacteria and archaea. Bioinformatics 20: 2787–2791

Meijer WG and Tabita R (1992) Isolation and characterization of the *nifUSVW-rpoN* gene cluster from *Rhodobacter sphaeroides*. J Bacteriol 174: 3855–3866

Moran NA (2002) Microbial minimalism: Genome reduction in bacterial pathogens. Cell 108: 583–586

Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J and Ward N (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature 432: 910–913

Neidle E and Kaplan S (1992) *Rhodobacter sphaeroides rdxA*, a homolog of *Rhizobium meliloti fixG*, encodes a membrane protein which may bind cytoplasmic (4Fe-4S) clusters. J Bacteriol 74: 6444–6454

Neidle E and Kaplan S (1993a) Expression of *Rhodobacter*

sphaeroides hemA and *hemT* genes encoding two aminolevulinic acid synthase isozymes. J Bacteriol 175: 2292–2303

Neidle E and Kaplan S (1993b) 5-aminolevulinic acid availability and control of spectral complex formation in HemA and HemT mutants of *Rhodobacter sphaeroides*. J Bacteriol 175: 2304–2313

Nereng K and Kaplan S (1999) Genomic complexity among strains of the facultative phototrophic bacterium, *Rhodobacter sphaeroides*. J Bacteriol 181: 1684–1688

Pappas CT, Sram J, Moskvin OV, Ivanov PS, Mackenzie RC, Choudhary M, Land ML, Larimer FW, Kaplan S and Gomelsky M (2004) Construction and validation of the *Rhodobacter sphaeroides* 2.4.1 DNA microarray: Transcriptome flexibility at diverse growth modes. J Bacteriol 186: 4748–4758.

Pareja E, Pareja-Tobes P, Manrique M, Pareja-Tobes E, Bonal J and Tobes R (2006) Extra Train: A database of extragenic regions and transcriptional information in prokaryotic organisms. BMC Microbiol 6: 29–38

Poggio S, Abreu-Goodger C, Fabela S, Osorio A, Dreyfus G, Vinuesa P and Camarena L. (2007) A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in *Rhodobacter sphaeroides*. J Bacteriol 189: 3208–3216

Ramos JL (2002) AraC-XylS database: a family of positive transcriptional regulators in bacteria. Nucl Acid Res 30: 318–321

Ramos JL, Martinez-Bueno M, Molina-Henares AJ, Teran W, Watanabe K, Zhang X, Gallegoes MT, Brennan R and Tobes R (2005) The TetR family of transcriptional repressors. Microbial Mol Biol Rev 69: 326–356

Reslewic S, Zhou S, Place M, Zhang Y, Briska A, Goldstein S, Churas C, Runnheim R, Forrest D, Lim A, Lapidus A, Han CS, Roberts GP and Schwartz DC (2005) Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*. Appl and Environ Microbiol 71: 5511–5522

Rocha EP and Danchin A (2002) Base composition bias might result from competition for metabolic resources. Trend Genet 18: 291–294

Roh JH, Smith WE and Kaplan S (2004) Effects of oxygen and light intensity on transcriptome expression in *Rhodobacter sphaeroides* 2.4.1. J Biol Chem 279: 9146–9155

Sanchez-Contreras M, Martin M, Villacieros M, O'Gara F, Bonilla I and Rivilla R (2002) Phenotypic selection and phase variation occur during alfalfa root colonization by *Pseudomonas fluorescens* F113. J Bacteriol 184: 1587–1596

Suwanto A and Kaplan S (1989a) Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: Genome size, fragment identification, and gene localization. J Bacteriol 171: 5840–5849

Suwanto A and Kaplan S (1989b) Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: Presence of two unique circular chromosomes. J Bacteriol 171: 5850–5859

Tavano CL, Podevels A and Donohue TJ (2005) Gene products required to recycle reducing power produced under photosynthetic conditions. J Bacteriol 187: 5249–5258

Tharia HA, Nightingale TD, Papiz MZ and AM Lawless AM (1999) Characterisation of hydrophobic peptides by RP-HPLC from different spectral forms of LH2 isolated from *Rhodopseudomonas palustris*. Photosynth Res 61: 157–167

Turina P, Rebecchi A, D'Alessandro M, Anefors S and Melandri BA (2006) Modulation of proton pumping efficiency in bacterial ATP synthases. Biochim Biophys Acta 1757: 320–325

Van Niel CB (1944) The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. Bacteriol Rev 8: 1–118

VerBerkmoes NC, Shah MB, Lankford PK, Pelletier DA, Strader MB, Tabb DL, McDonald WH, Barton JW, Hurst GB, Hauser L, Davison BH, Beatty JT, Harwood CS, Tabita FR, Hettich RL and Larimer FW (2006) Determination and comparison of the baseline proteomes of the versatile microbe *Rhodopseudomonas palustris* under its major metabolic states. J Proteome Res 5: 287–298

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

Woese C (1998) The universal ancestor. Proc Natl Acad Sci U S A.95: 6854–6859

Woese CR, Stachebrandt E, Weisburg WG, Paster BJ, Madigan MT, Fowler CRM, Hahn CM, Blanz P, Gupta R, Nealson KH and Fox GE (1984) The phylogeny of the purple bacteria: the α-subdivision. Syst Appl Microbiol 5: 315–326

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

Zeilstra-Ryalls J and Kaplan S (1995) Regulation of 5-aminolevulinic acid synthase in *Rhodobacter sphaeroides* 2.4.1: The genetic basis of mutant H-5 auxotrophy. J Bacteriol 177: 2760–2768

Zeilstra-Ryalls J, Gomelsky M, Eraso JM, Yeliseev A, O'Gara J and Kaplan S (1998) Control of photosystem formation in *Rhodobacter sphaeroides*. J Bacteriol 180: 2801–1809

Zeng X, Choudhary M and Kaplan S (2003) A second and unusual *pucBA* operon of *Rhodobacter sphaeroides* 2.4.1: Genetics and function of the encoded polypeptides. J Bacteriol 185: 6171–6184

Zeng X, Roh JH, Callister SJ, Tavano CL, Donohue TJ, Lipton ML and Kaplan S (2007) Proteomic characterization of the *Rhodobacter sphaeroides* photosynthetic membrane. J Bacteriol 189: 7464–7474

Zhou SE, Kile AK, Severin J, Forrest D, Runnheim R, Churas C, Anantharaman TS, Hickman JW, Mackenzie C, Choudhary M, Donohue TJ, Kaplan S and Schwartz DC (2003) Whole-genome shotgun optical mapping of *Rhodobacter sphaeroides* 2.4.1 and its use for whole-genome shotgun sequence assembly. Genome Res 13: 2142–2151