

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



The Core- and Pan-Genomes of Photosynthetic Prokaryotes

Jeffrey W. Touchman and Yih-Kuang Lu
Arizona State University
USA

1. Introduction

Genome sequencing projects are revealing new information about the distribution and evolution of photosynthesis and phototrophy, particularly in prokaryotes. Although coverage of the five phyla containing photosynthetic prokaryotes (Chlorobi, Chloroflexi, Cyanobacteria, Proteobacteria and Firmicutes) is limited and uneven, full genome sequences are now available for 82 strains from these phyla. In this chapter, we present data and comparisons that reflect recent advances in phototroph biology as a result of insights from genome sequencing. By performing a comprehensive analysis of the core-genome (the pool of genes shared by all phototrophic prokaryotes) and pan-genome (the global gene repertoire of all phototrophic prokaryotes: core genome + dispensable genome) along with available biological data for each organism, we address the following key questions: 1) what are the principal drivers behind the evolution and distribution of phototrophy and 2) how do environmental parameters correlate with genomic content to define niche partitioning and ecotype distributions in photic environments?

Over a decade has passed since the first phototrophic prokaryote, the cyanobacterium *Synechocystis* sp. PCC 6803, was completely sequenced (Kaneko et al., 1996). Since then, availability of an increasing diversity of newly sequenced species is accumulating in public databases at a sustained pace and there is little indication that this trend will level off in the near future (Raymond & Swingley, 2008). A deepening archive of complete genomes has enabled comparative genomic analyses, which has heavily influenced our views of genome evolution and uncovered the extent of gene sharing between organisms (Pallen & Wren, 2007). The analysis of pan- and core-genomes in particular allows us to link genome content to the relationship of organisms to one another and to their physical surroundings. For example, a low pan-genome diversity due to extensive overlap of metabolic function among groups of bacteria could reflect shared environmental habitats and resource utilization, while distinctive species that adapt to disparate environments would be expected to have a high pan-genome diversity. This approach was first developed by Tettelin et al. (2005) and Hogg et al. (2007) for tracking the number of unique genes among multiple strains of *Streptococcus agalactiae* and *Haemophilus influenzae*, respectively. Such analysis resulted in the determination of core-genes that encode functions related to the basic metabolism and phenotype of the species, and a pan-genome that consists of dispensable or unique genes that impart specific functionalities to individual strains.

Within prokaryotes, photosynthetic capability is present within five major groups, which include heliobacteria, green filamentous bacteria (*Chloroflexus sp.*), green sulfur bacteria (*Chlorobium sp.*), *Proteobacteria*, and *Cyanobacteria* (Blankenship, 1992; Gest & Favinger, 1983; Olson & Pierson, 1987; Vermaas, 1994). While only *Cyanobacteria*, which contain two distinct reaction centers linked to each other, are capable of oxygenic photosynthesis, other photosynthetic bacteria primarily carry out anoxygenic photosynthesis with a single reaction center. Traditionally, the phylogenetic relationship of these five distinct photosynthetic groups has been constructed by comparing sequences of the small subunit 16S rRNA gene (Ludwig & Klenk, 2001). But the use of the 16S rRNA gene is unable to resolve the relationships among these phototrophs with confidence, which is central to understanding their evolution. For example, phylogenetic trees based on a comparison of different combinations of 527 shared genes amongst all five photosynthetic prokaryote groups shows that no less than 15 different tree topologies can be constructed depending on the subset of genes used in the analysis, only one of which matches the traditional 16S rDNA tree (Raymond et al., 2002). In fact, comparing just those genes involved in photosynthesis supports no coherent relationship among the different photosynthetic bacteria either, indicating that such genes may have been subjects of lateral gene transfers (ibid).

Recent genome sequencing efforts have made whole genome data available for many more representatives of each of the five phyla of bacteria with photosynthetic members. To resolve the complicated relationship between bacterial phototrophy and evolutionary history, we describe an analysis of the 82 fully-sequenced photosynthetic prokaryotes to construct the pan- and core-genomes across all available strains. We present results showing various gene-based indicators of the relationship between genome and phenotype among these organisms. Not surprisingly, our findings describe new relationships between gene content and environmental habitat. These results add to a complete gene-based functional annotation of the phototrophic prokaryotes, and set the groundwork for continuing studies on genetic and evolutionary dynamics of this important photosynthetic community.

2. Whole-genome analysis of phototrophic prokaryotes

The list and summary details of 82 fully-sequenced photosynthetic species used in this study are shown in Table 1 (Liolios et al., 2006). Every species exhibits common characteristics with other relatives in the same phylum. For example, the *Chlorobia* and heliobacteria (*Firmicutes*) are strictly anaerobic while the *Chloroflexia* and *Proteobacteria* are facultatively anaerobic. The *Chloroflexia* are alkali-trophic thermophiles whereas other phylal members are neutral pH mesophiles. Genome size is generally uniform among the *Chlorobia* and *Chloroflexia*, but varies widely among the *Cyanobacteria* and *Proteobacteria*. Furthermore, both *Chloroflexia* and *Proteobacteria* possess a pheophytin-quinone reaction center, while *Heliobacteria* and *Chlorobia* use an iron-sulfur reaction center. *Cyanobacteria* exclusively possesses two types of reaction centers. Both *Chlorobia* and *Cyanobacteria* are two phyla comprised entirely of photosynthetic representatives. Although most of the photosynthetic species are free-living organisms, *Nostoc sp.* PCC 7120, *Nostoc punctiforme* PCC 73102, and *Acaryochloris marina* MBIC11017 in the *Cyanobacteria* and *Bradyrhizobium* BTAi1, ORS278 and some *Methylobacterium* strains in the *Proteobacteria* form a mutual relationship with terrestrial plants and coral. The *Heliobacteria* (e.g., *Heliobacterium modesticaldum*) are the only photosynthetic members of the *Firmicutes*. The genome of *Heliobacillus mobilis*, the strain most studied biochemically, still remains proprietary and was not included in our analysis.

2.1 Clustering of ortholog groups of photosynthetic prokaryotes

All of the 312,254 protein sequences from 82 photosynthetic prokaryote genomes were collected and clustered with the Markov clustering algorithm Ortho-MCL (Chen et al., 2006). Ortho-MCL is a graph-clustering algorithm designed to identify homologous proteins based on sequence similarity and distinguish true orthologs from paralogous relationships without computationally intensive phylogenetic analysis. Upon clustering, 41,824 proteins (13.3%) were removed due to the absence of detectable sequence similarities (BLASTP; $E=10^{-5}$) and 272,686 (86.7%) were assigned to clusters. To assess clustering performance we modified a method described by Frech and Chen (2010) whereby both false-positive (the number of proteins that are found in two or more separate clusters) and false-negative (number of proteins that are found in wrong clusters) results were calculated using both the KEGG (Kanehisa & Goto, 2000) and COG (Natale et al., 2000) databases as a reference. An inflation index is then calculated that controls cluster granularity and gene family size while limiting error (Huerta-Cepas et al., 2008). The inflation parameter impacts the calculation of the number of shared orthologs in each phylum.

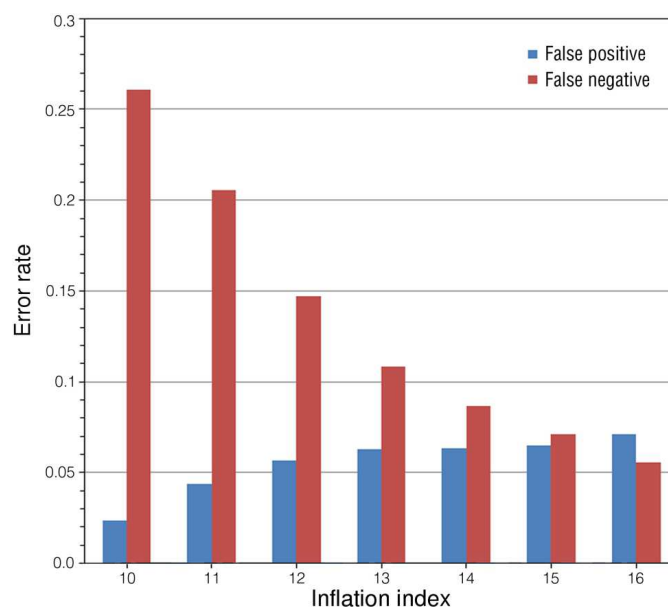


Fig. 1. Estimate of false positive and false negative error rate during ortholog clustering

As Figure 1 shows, an increasing false-positive rate is anti-parallel to decreasing false-negative rate in the inflation parameter. In order to obtain an adequate clustering result, we adjusted the Ortho-MCL program parameters such that reference ortholog clusters compared to both KEGG and COG are classified correctly to minimize erroneous clustering of orthologous groups (inflation index of 15). In our analysis, each predicted orthologous group was evaluated and corrected based on information from both KEGG and COG databases.

2.2 The assembly of core- and pan-genomes

The pan-genome of all 82 species contains 312,254 genes that form 23,362 ortholog clusters. Based on the clustering results, we observed that every photosynthetic prokaryote shares large portions of its genes with others. 204,074 genes that represent 74.8% of the entire data were found to co-exist in at least two organisms from any phyla (“multi-shares”; Figure 2).

The number of gene clusters specific to a particular phylum is much smaller. Both *Cyanobacteria* and *Proteobacteria* possess 32,316 (11.8%) and 30,717 (11.3%) phylum-specific gene clusters, respectively, whereas both *Chlorobi* and *Chloroflexi* have 2,290 (0.8%) and 3,123 (1.1%) gene clusters, respectively. Additionally, 16,665 genes of all species (6.1%) are in common (that is, are contained in the phototrophic prokaryote core-genome). On the surface, this result suggests a remarkable degree of overlap in the gene composition across all five major phyla of photosynthetic prokaryotes.

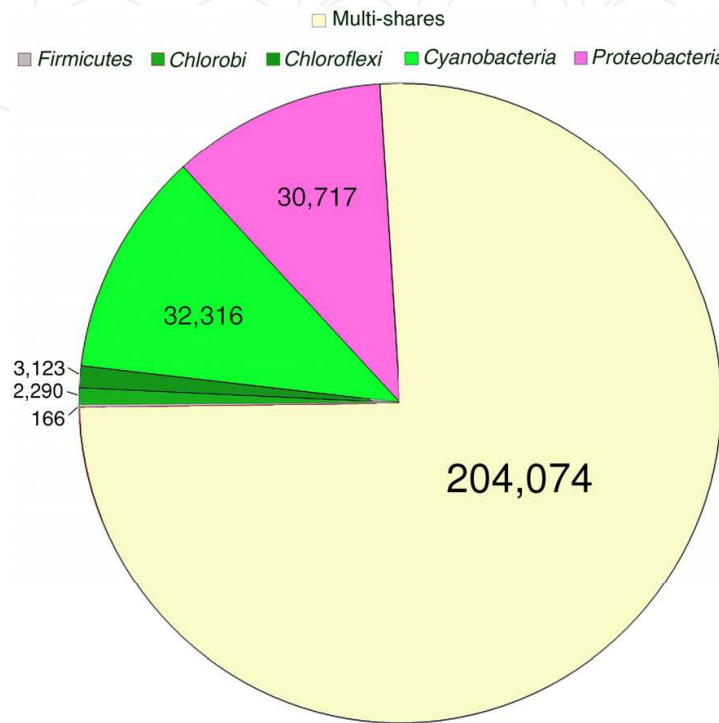


Fig. 2. Distribution of clustered genes within the pan-genome of the five photosynthetic bacterial phyla. Numbers indicate the number of genes specific to each group.

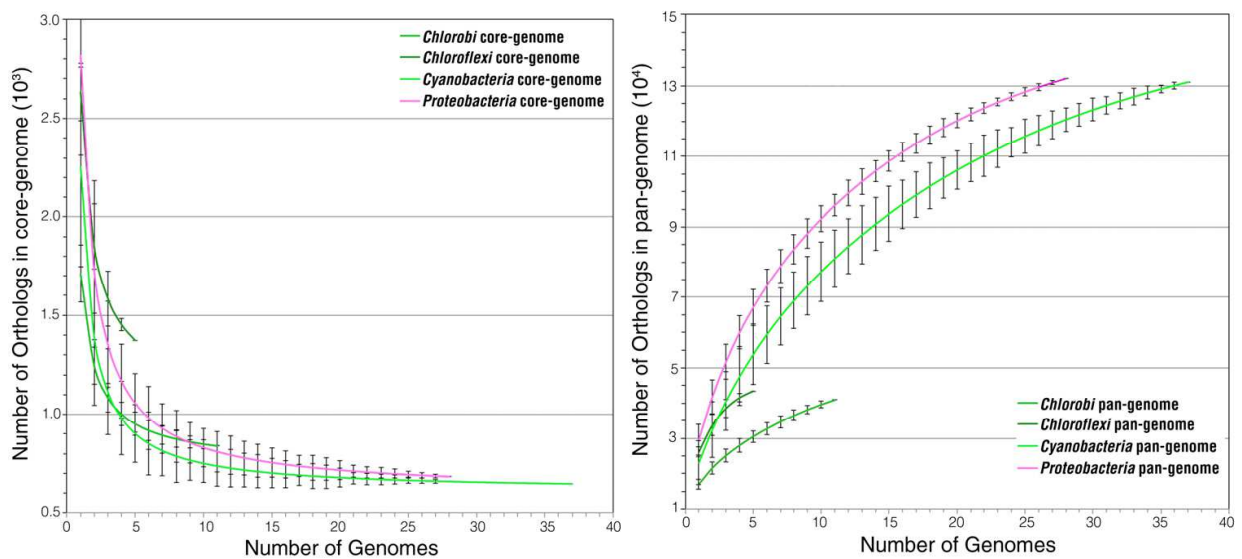


Fig. 3. Plot of the contraction of the core-genome (left) or expansion of the pan-genome (right) as the number of photosynthetic prokaryote genomes analyzed is increased.

To estimate the change of the core-genome size within a particular phyla upon sequential addition of each new genome sequence, a plot was extrapolated by fitting a power law function to the data (Figure 3). As more genomes are compared, there is an asymptotic decline in the number of core orthologs in every phyla, similar to observations for *Streptococcus* (Lefebure & Stanhope, 2007) and *Prochlorococcus* genomes (Kettler et al., 2007). The pan-genome, in contrast, was determined by the plot of the numbers of new orthologs, which fit a decaying exponential curve (Figure 3). The gene accumulation curve for each phyla is clearly far from saturated.

2.3 The core- and pan-genomes of phototrophic prokaryotes

The results of the clustering analysis to determine the core- and pan-genome sizes for each phylum and all phyla together is shown in Figure 4. The size of the phylum core-genomes are: 819 genes in the *Chlorobi*, 1,392 in the *Chloroflexi*, 619 in the *Cyanobacteria*, and 644 in the *Proteobacteria*. The core-genome of all 82 phototrophs considered together consists of 268 genes shared by all organisms. This overall core-genome encompasses a large number of housekeeping genes involved in genetic processes and metabolism and a small number of genes involved in cellular and environmental processes. The housekeeping genes involved in genetic processes include DNA polymerase, ligase, and helicase for DNA replication; RNA polymerase, ribosomal proteins, and tRNA synthetases for translation; and chaperones and signal peptidase for post-translational processes. The housekeeping genes involved in metabolism are mainly involved in the biosynthesis of amino acids, nucleotides, and coenzymes, and a few key enzymes such as transketolase, phosphoglycerate mutase, phosphoglycerate kinase of the glycolysis, acetyl-CoA carboxylase of the tricarboxylic acid (TCA) cycle, H⁺-transporting ATPase, acyl carrier protein, and UDP-N-acetylmuramate-L-alanine ligase for the biosynthesis of bacterial cell wall are preserved. Moreover, we identified the chlorophyll-synthesizing enzymes that include porphobilinogen synthase, oxygen-independent coproporphyrinogen III oxidase, magnesium chelatase, chlorophyll synthase, magnesium-protoporphyrin O-methyltransferase, and light-independent protochlorophyllide reductase. For both cellular and environmental processes, glycosyltransferase for cell membrane biogenesis, phosphate transport system proteins, signal recognition SRP54, and sec-independent protein TatC for membrane transport were identified, suggesting that transferring phosphate and translocating membrane proteins are universal in photosynthetic organisms. The large proportion of housekeeping genes responsible for nearly all major genetic functions and the biosynthesis for both amino acids and nucleotides is understandable since these genes are essential for basic life functions. We observed a paucity of genes involved in both cellular and environmental processes in the overall core-genome. This observation supports the view in which essential life functions are unchanging while nonessential or environment-specific functions are found in a flexible genome (Kettler et al., 2007).

Core-genomes were also calculated in a pairwise fashion between photosynthetic phyla to gauge the number of shared orthologs in a given pair of phyla pan-genomes (Figure 5). Each circle in the figure is proportional to the size (number in the circle) of the shared orthologs. Although these results are heavily influenced by the size of the dataset for an individual phyla, it provides a provisional measure of shared genes between phyla.

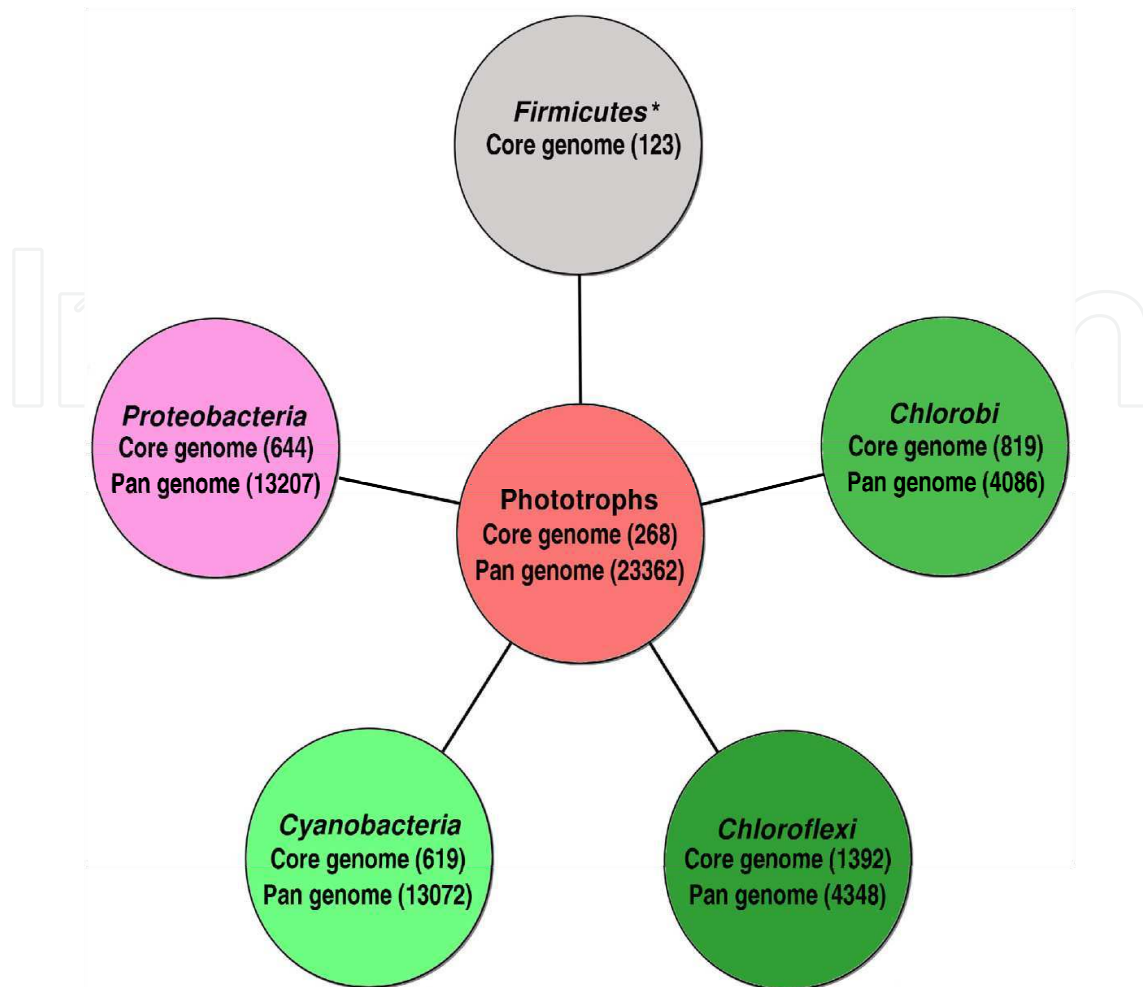


Fig. 4. Both core- and pan-genomes present in all five photosynthetic phyla. Each colored circle represents a phylum: *Firmicutes* (grey), *Chlorobi* (green), *Chloroflexi* (dark-green), *Proteobacteria* (purple), *Cyanobacteria* (light-green), and all phyla together (red). Numbers represent the ortholog clusters contained within the core genome or pan genome of each phylum. * The *Firmicutes*, with only a single sequenced genome, lack a pan genome.

2.3.1 *Heliobacteria* (*Firmicutes*)

Given that there was only a single fully-sequenced genome from the phototrophic *Heliobacteria* (*Firmicutes*) available for study, the core genome of *Heliobacteria* was provisionally constructed by excluding those genes that are homologous to any known genes from the other sequence-available *Firmicutes*. It is worth noting that although there are four *Heliobacteria* genera containing a total of ten species that have been formally described: *Heliobacterium*, *Heliobacillus*, *Heliophilum*, and *Heliorestis*, the phototrophic *Heliobacterium modesticaldum* is the only sequenced bacteria representing them (Sattley et al., 2008). When *H. modesticaldum* was compared with the available bacterial genomes of the *Firmicutes*, we identified 123 ortholog clusters tentatively assigned to the core-genome of this organism. Genes encoding proteins involved in major genetic, cellular, and environmental processes and metabolism are very limited. This may be partly due to their mutualistic relationship with plants. Other major ortholog groups in this core-genome are involved in sporulation. The previous examination of several other *Heliobacteria* species for sporulating genes has

indicated that sporulation gene presence may be universal within the heliobacteria (Kimble-Long & Madigan, 2001). It should be noted that a set of genes involved in bacteriochlorophyll (Bchl) *g* biosynthesis, not found in other phototrophs, were frequently reported in other heliobacteria species (reviewed in Asao & Madigan, 2010). These enzymes were not clustered into the core-genome due to their absence in the non-phototrophic *Firmicutes*. Additional genome sequences from the heliobacteria group will aid in our understanding of their specific core-genome.

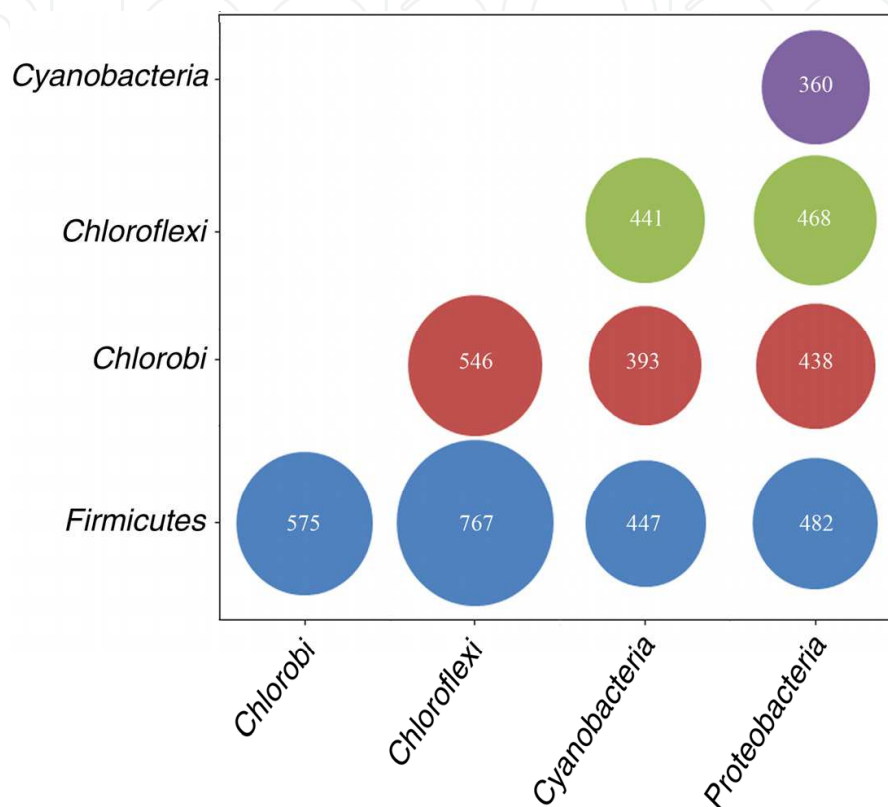


Fig. 5. Pairwise comparison of shared ortholog groups between phyla

2.3.2 *Chlorobi*

The *Chlorobia* core-genome contains 819 genes representing 30-40% of the total genes in a given *Chlorobia* genome. As a phyla, they are very similar with respect to gene content compared to the other phototrophic prokaryotes. In addition to the components of the core-genome for all species, the *Chlorobia* core orthologs are composed of major metabolism genes such as the electron transport chain that supports photosynthesis and sulfur oxidation, the reductive TCA cycle supporting carbon fixation and transport, and others for the biosynthesis of amino acids, lipids, and coenzymes. The core-genome also contains the type I reaction center unique to the *Chlorobi*. Our findings are similar to other recent reports (Davenport et al., 2010). In addition to those identified orthologs for central metabolism, we also identified genes involved in the biosynthesis of Bchl, carotenoids, and the photosynthetic “chlorosome” apparatus. Most pigment-synthesizing enzymes operate downstream of the metabolic pathways for final products like Bchl *c*, *d*, chlorobactene, and γ -carotene, which are located on the chlorosome to harvest light. A few metal and inorganic compound transporters for iron, nickel, and molybdate as well as the major facilitator

superfamily (MFS) transporter, were identified. Since the *Chlorobi* species are capable of fixing nitrogen, preserving these transport systems is necessary to support this process. A total of 1,774 ortholog clusters are assigned to the pan-genome of *Chlorobia*. Most of these are associated with phylogenetically close species and have functions such as secretion, extracellular constituents, and cell wall biogenesis. These are conspicuous features of the genera *Chlorobium* and *Prosthecochloris*. Although the *Chlorobia* have been well-characterized biochemically and microbiologically (Frigaard & Bryant, 2004), our finding that the *Chlorobia* possess a relatively uniform core-genome complemented by a relatively limited set of accessory genes enhancing cellular activities provides further insight into their anoxygenic phototrophic lifestyle. The core-and pan-genome pattern suggests a largely vertical inheritance that has preserved the core-genome needed for major cellular activities--a result of living in environmentally stable niches.

2.3.3 *Chloroflexi*

The *Chloroflexi* core-genome contains 1,392 ortholog clusters, the largest size among the five phototrophic groups. It reflects roughly 35% of the genes of a *Chloroflexus* species' genome. The functional composition of this core-genome is somewhat similar to that of the *Chlorobi* core-genome, since many core genes involved in both genetic and cellular processes were cross-identified. However, the core set also contains type II reaction centers, NADH dehydrogenase, and cytochrome *c* oxidase, similar to the *Proteobacteria* but different from the *Chlorobi*. Moreover, many transporters for metal ions, inorganic and organic compounds, as well as two-component histidine kinases for signal transduction were identified in the *Chloroflexi* but not seen in the *Chlorobi*. The conservation of functionally diverse transporters with signal-transduction histidine kinases may be related to a more dynamic life-style. Generally, *Chloroflexus* is a photoheterotroph and usually found in the lower layers of microbial mats with cyanobacteria growing above it that provide organic byproducts. The *Chloroflexi* core-genome possesses numerous heat shock proteins, chaperones, and signal peptidases involved in protein folding and translocating processes that likely serve to reinforce protein structures in the thermophilic *Chloroflexus* species. For genes involved in major metabolic pathways, the core-genes appear to be largely conserved across all photosynthetic phyla. Although the *Chloroflexi* are found to be distinctive from the *Chlorobi*, they do have some common characteristics such as the absence of intra-cytoplasmic membrane structures and chlorosomes on their plasma membranes. They also use the same Bchl *a* and *c* biosynthesis pathways.

The *Chloroflexi* pan-genome contains 4,348 genes, and in contrast to the *Chlorobi* pan-genome it is comprised of more putative genes for extracellular constituents, inter-cellular communication, and other physiological and biochemical activities. For example, genes involved in the 3-hydroxypropionate pathway for carbon fixation were found. But generally, the *Chloroflexi* core-genome equips most of the major functional genes for a wide range of metabolisms such as synthesis of organic compounds, energy production, transport, genetic processing, etc. Such coverage throughout most cellular activities makes the core-genome of the *Chloroflexi* similar in character to that of *Chlorobia*.

2.3.4 *Cyanobacteria*

Representing the largest sampled phylogenetic clade of the phototrophic prokaryotes, the *Cyanobacteria* have 37 completely sequenced genomes available for analysis, resulting in the

smallest core- (619 genes) and largest pan-genome (13,072 genes) of all five phyla. The proportion of genes designated “core” with respect to any cyanobacterial genome varies from less than 10% (in the the non-*Prochlorococcus*/*Synechococcus* genomes) to nearly 38% for *Prochlorococcus* and *Synechococcus* strains. The core orthologs are responsible for several major reactions such as the Calvin cycle, glycolysis, the incomplete TCA cycle, and pathways to synthesize amino acids and cofactors. Two types of photosystems (PS I and PS II) and the participating electron transport chain for oxygenic photosynthesis are also included.

The large pan-genome of cyanobacteria appears to support diverse abilities and processes. There are many genes found in the pan-genome that carry out metabolic activities unrelated to photosynthesis. For example, *M. aeruginosa* NIES-843 produces a diverse range of toxins with the non-ribosomal peptide synthetases (Kaneko et al., 2007). These enzymes produce neurotoxins and hepatotoxins that cause a variety of human illnesses, and are responsible for deaths in native and domestic animals. *T. erythraeum* IMS101 can perform nitrogen fixation in the presence of oxygen (Sandh et al., 2011). *Nostocaceae* species generally have an unbranched filamentous cell type, develop heterocysts, and possess multiple plasmids. In contrast, *Prochlorococcus* and *Synechococcus* species have a small round shape with no plasmids. Finally, a large number of genes identified in members of the *Nostocaceae* and *A. marina* MBIC11017 have unknown functions. Judging by the life style of these cyanobacterial species, which have a mutualistic relationship with terrestrial plants (Baker et al., 2003) and coral (Marquardt et al., 1997), it is possible that these genes are involved in supporting inter-communication and mutualism with their host.

2.3.5 Proteobacteria

The *Proteobacteria* contain 644 core gene clusters and 13,207 non-redundant genes in the pan genome. The percentage of core genes in any of the *Proteobacteria* genomes varies from 10% to 25%, similar to the results obtained for the *Cyanobacteria*. This is because the *Proteobacteria* is the second major photosynthetic group with 28 completely sequenced genomes from phylogenetically distinct clades. The *Proteobacteria* core-genome preserves most of the key enzymes essential to major cellular activities, similar to other core-genomes. The type II reaction center and light-harvesting proteins are in the core genome, the former of which was also identified in the *Chloroflexi*. Nevertheless, the additional orthologs coding for the bacterial flagella, chemotaxis, and respiratory electron transfer chain proteins unique to the *Proteobacteria* were also identified. Both flagella and chemotaxis help cells move either toward nutrients or away from unfavourable living conditions and both anaerobic and aerobic respiration supports chemo-heterotrophic growth when phototrophic growth is not possible. Thus, integrating both cell mobility and respiration to the *Proteobacteria* core genome suggests an ecological advantage of adaptation to a broader range of living environments than other phototrophic phyla.

In contrast to the core-genome, the characteristics of the pan-genome are widely diverse--from variant types of nitrogen assimilation, carbon assimilation, and hydrogen metabolism to inter-cellular communications and nodulation. Such vast variety in the functional repertoire associated with the pan-genome can give the *proteobacteria*, such as *Rhodobacter*, *Rhodopseudomonas*, and *Rhodospillium* genera, a broad range of growth conditions for anaerobic phototrophy and aerobic chemoheterotrophy in the absence of light (Larimer et al., 2004; Lu et al., 2010; Mackenzie et al., 2007). Over 50 genes associated with nodulation

were identified in the diazotrophic *Bradyrhizobiaceae* and *Rhizobiaceae*. Most photosynthetic bacteria in these two orders are capable of forming mutualistic symbiosis with terrestrial plants by fixing nitrogen inside special structures called legumes. Genes for hydrogen production or metabolizing C₁-compounds such as methane were identified in *Rhodopseudomonas*, some *Rhodobacteraceae*, and *Methylobacterium*. These traits have garnered much-warranted attention for their potential ability to reduce CH₄ (greenhouse gas) emission (Eller & Frenzel, 2001; Lidstrom & Chistoserdova, 2002). Based on the construction of both core- and pan-genomes, photosynthetic members in the *Proteobacteria* exhibit the greatest gene diversity amongst all phyla studied. This diversity reflects their ability to grow chemoheterotrophically as well as phototrophically, which makes them better at living in a broader range of environments than the *Cyanobacteria*.

Taken together, we have identified core-genes responsible for phylum-specific reactions. We have also observed a wide variety of accessory functions supporting smaller groups of bacteria. The core-genome assembled from a group of closely related bacteria represents a backbone of essential components regulating the adaptability to specific niches. Our results indicate that the gene content of each phylum-specific core is distinctive and can exemplify the very different evolutionary histories of the major photosynthetic groups, where the accessory components comprising the pan-genomes provide fitness advantages in distinct habitats.

2.4 Phylogeny of photosynthetic prokaryotes using the pan-genome

Construction of both core- and pan-genomes of all photosynthetic bacteria provides a novel opportunity to determine the phylogenetic relationship among these prokaryotes. Several methods have been used to evaluate the phylogenies of different bacterial groups such as single-gene phylogenies (e.g., 16S rDNA), concatenated sequences of photosynthesis-related proteins (Rokas et al., 2003), and signature sequences of house-keeping proteins (Gupta, 2003). The sequences compared in these methods are necessarily present in all analyzed species. Here, we present a phylogeny that is formulated using the clustered pan-genome that does not rely on a universally shared collection of genes. Hierarchical clustering with resampling 100 times was performed based on a relative Manhattan distance calculated on the presence/absence of an ortholog between a given pair of genomes (Snipen & Ussery, 2010). It in essence generates a tree based on shared gene content. Figure 6 shows the resulting tree.

There is broad agreement with this tree and traditional single-gene phylogenies. But surprisingly, the topology of the tree shows that both *A. vinosum* DSM 180 and *H. halophila* SL1, both belonging to the γ -*Proteobacteria* class, are situated outside of the *Proteobacteria* clade and positioned between the *Chlorobi* and *Firmicutes*. We investigated in detail the gene content of these two organisms and found that *A. vinosum* DSM 180 has lost most of the *Proteobacteria*-specific orthologs, while *H. halophila* SL1 contains more shared orthologs with *A. vinosum* DSM 180 than between the other purple bacteria species. Another unusual topology is found in the *Proteobacteria* clade where both *Rhodobacter* and *Rhodospirillum* families are closer to the *Rhodopseudomonas* genus, which belongs phylogenetically to the *Rhizobia* within the *Brydorrhizobia* and *Methylobacter* families.

We further utilized the pan-genome to reveal a three-dimensional relationship between individual species and the major photosynthetic lineages. By performing a multidimensional scaling analysis of the ortholog distribution across all 82 species, we found that related species were clustered in groups reflecting their phyla (Figure 7). While the *Heliobacteria*, *Chlorobi*, and *Chloroflexi* species occupied a central space, the *Proteobacteria* and *Cyanobacteria* were greatly

separated and located on opposing poles. The relative position of the *Proteobacteria* and *Cyanobacteria* groups apart from each other indicate that their ortholog profiles have diverged substantially. Additionally, the distribution of both *Cyanobacteria* and *Proteobacteria* species is also consistent with their phylogenetic positions in Figure 6. However, the γ -*Proteobacterial* species, *A. vinosum* DSM 180 and *H. halophila* SL1 were exceptionally close to both *Chlorobi* and *Chloroflexi*, a result similar to the two-dimensional pan-genome-based phylogenetic tree. Clustering organisms by determining the occurrence of the specific patterns of orthologs shared by a group of species reveals an overall pattern consistent with both 16s rDNA- and pan-genome-based phylogenies. Yet, the observation of shared orthologs in one or a group of species can highlight functional divergence or convergence in groups that can be quantified by gene analysis but missed by single-gene-based phylogenies.

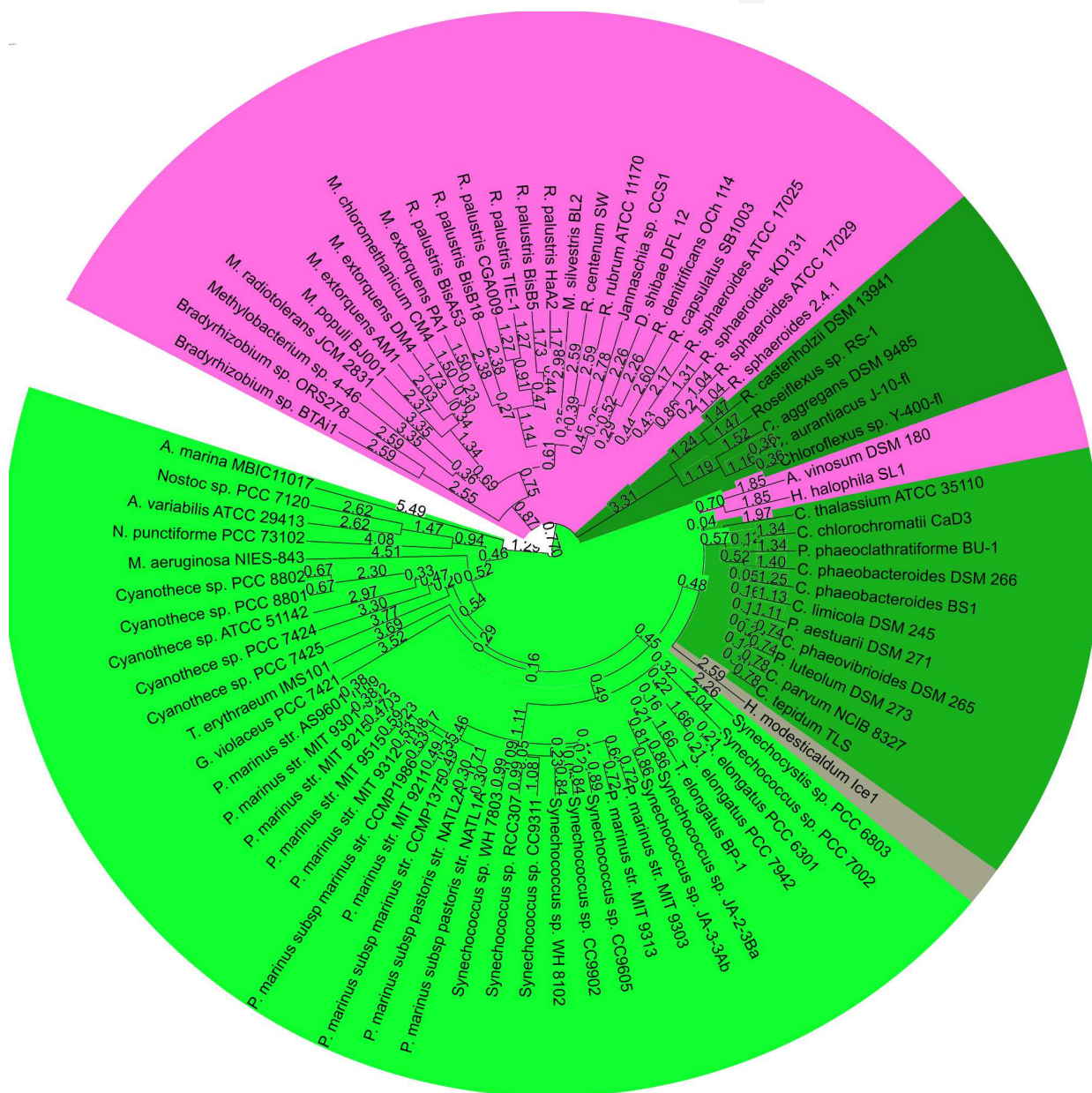


Fig. 6. Phylogenetic relationship of 82 photosynthetic prokaryotes reconstructed with the pan-genome

3. Acknowledgements

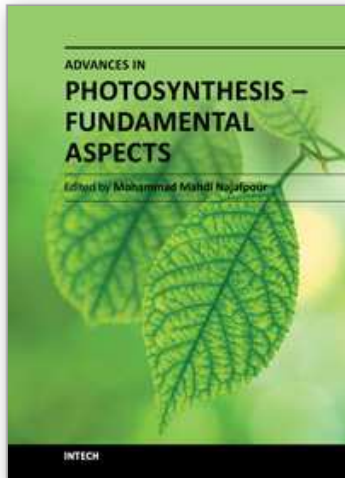
This work was supported by the U.S. National Science Foundation Phototrophic Prokaryotes Sequencing Project, grant number 0950550.

4. References

- Asao, M., & Madigan, M.T. (2010). Taxonomy, phylogeny, and ecology of the heliobacteria. *Photosynthesis Research*, Vol.104, No.2-3, (2010), pp. 103-111, ISSN 1573-5079; 0166-8595
- Baker, J.A., Entsch, B., & McKay, D.B. (2003). The cyanobiont in an *Azolla* fern is neither *Anabaena* nor *Nostoc*. *FEMS microbiology letters*, Vol.229, No.1, (2003), pp. 43-47, ISSN 0378-1097; 0378-1097
- Blankenship, R.E. (1992). Origin and early evolution of photosynthesis. *Photosynthesis Research*, Vol.33, No.2, (1992), pp. 91-111, ISSN 0166-8595
- Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., & Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res*, Vol.34, No.Database issue, (Jan 1 2006), pp. D363-368, ISSN 1362-4962; 0305-1048
- Davenport, C., Ussery, D.W., & Tummler, B. (2010). Comparative genomics of green sulfur bacteria. *Photosynthesis Research*, Vol.104, No.2-3, (2010), pp. 137-152, ISSN 1573-5079; 0166-8595
- Eller, G., & Frenzel, P. (2001). Changes in activity and community structure of methane-oxidizing bacteria over the growth period of rice. *Applied and Environmental Microbiology*, Vol.67, No.6, (2001), pp. 2395-2403, ISSN 0099-2240
- Frech, C., & Chen, N. (2010). Genome-wide comparative gene family classification. *PLoS One*, Vol.5, No.10, (2010), pp. e13409, ISSN 1932-6203
- Frigaard, N.U., & Bryant, D.A. (2004). Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. *Archives of Microbiology*, Vol.182, No.4, (2004), pp. 265-276, ISSN 0302-8933; 0302-8933
- Gest, H., & Favinger, J.L. (1983). *Heliobacterium chlorum*, an anoxygenic brownish-green photosynthetic bacterium containing a "new" form of bacteriochlorophyll. *Archives of Microbiology*, Vol.136, No.1, (1983), pp. 11-16, ISSN 0302-8933
- Gupta, R.S. (2003). Evolutionary relationships among photosynthetic bacteria. *Photosynth Res*, Vol.76, No.1-3, (2003), pp. 173-183, ISSN 0166-8595
- Hogg, J.S., Hu, F.Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J.C., & Ehrlich, G.D. (2007). Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol*, Vol.8, No.6, (2007), pp. R103, ISSN 1465-6914; 1465-6906
- Huerta-Cepas, J., Bueno, A., Dopazo, J., & Gabaldon, T. (2008). PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic acids research*, Vol.36, No.Database issue, (2008), pp. D491-496, ISSN 1362-4962; 0305-1048
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, Vol.28, No.1, (2000), pp. 27-30, ISSN 0305-1048; 0305-1048

- Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., Kawashima, K., et al. (2007). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA research*, Vol.14, No.6, (2007), pp. 247-256, ISSN 1756-1663; 1340-2838
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiura, M., Sasamoto, S., et al. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research*, Vol.3, No.3, (Jun 30 1996), pp. 109-136, ISSN 1340-2838
- Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., et al. (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS genetics*, Vol.3, No.12, (2007), pp. e231, ISSN 1553-7404; 1553-7390
- Kimble-Long, L.K., & Madigan, M.T. (2001). Molecular evidence that the capacity for endospore formation is universal among phototrophic heliobacteria. *FEMS microbiology letters*, Vol.199, No.2, (2001), pp. 191-195, ISSN 0378-1097; 0378-1097
- Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., et al. (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospirillum rubrum*. *Nature biotechnology*, Vol.22, No.1, (2004), pp. 55-61, ISSN 1087-0156; 1087-0156
- Lefebvre, T., & Stanhope, M.J. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome biology*, Vol.8, No.5, (2007), pp. R71, ISSN 1465-6914; 1465-6906
- Lidstrom, M.E., & Chistoserdova, L. (2002). Plants in the pink: cytokinin production by methylobacterium. *Journal of Bacteriology*, Vol.184, No.7, (2002), pp. 1818, ISSN 0021-9193; 0021-9193
- Liolios, K., Tavernarakis, N., Hugenholtz, P., & Kyrpides, N.C. (2006). The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res*, Vol.34, No.Database issue, (Jan 1 2006), pp. D332-334, ISSN 1362-4962; 0305-1048
- Lu, Y.K., Marden, J., Han, M., Swingle, W.D., Mastrian, S.D., Chowdhury, S.R., Hao, J., Helmy, T., Kim, S., Kurdoglu, A.A., et al. (2010). Metabolic flexibility revealed in the genome of the cyst-forming alpha-1 proteobacterium *Rhodospirillum rubrum*. *BMC Genomics*, Vol.11, No.1, (May 25 2010), pp. 325, ISSN 1471-2164; 1471-2164
- Ludwig, W., & Klenk, H.P. (2001). Overview: A Phylogenetic Backbone and Taxonomic Framework for Prokaryotic Systematics. In: *The Archaea and the Deeply Branching and Phototrophic Bacteria*, D.R. Booke, & R.W. Castholz, eds., Springer-Verlag, pp. 49-65. ISBN 0387987711, 9780387987712, Berlin
- Mackenzie, C., Eraso, J.M., Choudhary, M., Roh, J.H., Zeng, X., Bruscella, P., Puskas, A., & Kaplan, S. (2007). Postgenomic adventures with *Rhodobacter sphaeroides*. *Annual Review of Microbiology*, Vol.61, (2007), pp. 283-307, ISSN 0066-4227; 0066-4227

- Marquardt, J., Senger, H., Miyashita, H., Miyachi, S., & Morschel, E. (1997). Isolation and characterization of biliprotein aggregates from *Acaryochloris marina*, a Prochloron-like prokaryote containing mainly chlorophyll d. *FEBS letters*, Vol.410, No.2-3, (1997), pp. 428-432, ISSN 0014-5793; 0014-5793
- Natale, D.A., Galperin, M.Y., Tatusov, R.L., & Koonin, E.V. (2000). Using the COG database to improve gene recognition in complete genomes. *Genetica*, Vol.108, No.1, (2000), pp. 9-17, ISSN 0016-6707; 0016-6707
- Olson, J.M., & Pierson, B.K. (1987). Evolution of Reaction Centers in Photosynthetic Prokaryotes. In *International Review of Cytology*, K.W.J. G.H. Bourne, & M. Friedlander, eds. (Academic Press), pp. 209-248.
- Pallen, M.J., & Wren, B.W. (2007). Bacterial pathogenomics. *Nature*, Vol.449, No.7164, (Oct 18 2007), pp. 835-842, ISSN 1476-4687; 0028-0836
- Raymond, J., & Swingley, W.D. (2008). Phototroph genomics ten years on. *Photosynth Res*, Vol.97, No.1, (Jul 2008), pp. 5-19, ISSN 0166-8595
- Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y., & Blankenship, R.E. (2002). Whole-genome analysis of photosynthetic prokaryotes. *Science*, Vol.298, No.5598, (Nov 22 2002), pp. 1616-1620, ISSN 1095-9203
- Rokas, A., Williams, B.L., King, N., & Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, Vol.425, No.6960, (2003), pp. 798-804, ISSN 1476-4687; 0028-0836
- Sandh, G., Ran, L., Xu, L., Sundqvist, G., Bulone, V., & Bergman, B. (2011). Comparative proteomic profiles of the marine cyanobacterium *Trichodesmium erythraeum* IMS101 under different nitrogen regimes. *Proteomics*, Vol.11, No.3, (2011), pp. 406-419, ISSN 1615-9861; 1615-9853
- Sattley, W.M., Madigan, M.T., Swingley, W.D., Cheung, P.C., Clocksin, K.M., Conrad, A.L., Dejesa, L.C., Honchak, B.M., Jung, D.O., Karbach, L.E., et al. (2008). The genome of *Heliobacterium modesticaldum*, a phototrophic representative of the Firmicutes containing the simplest photosynthetic apparatus. *J Bacteriol*, Vol.190, No.13, (Jul 2008), pp. 4687-4696, ISSN 1098-5530
- Snipen, L., & Ussery, D.W. (2010). Standard operating procedure for computing pangenome trees. *Standards in genomic sciences*, Vol.2, No.1, (2010), pp. 135-141, ISSN 1944-3277; 1944-3277
- Tettelin, H., Maignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America*, Vol.102, No.39, (2005), pp. 13950-13955, ISSN 0027-8424; 0027-8424
- Vermaas, W. (1994). Evolution of heliobacteria: Implications for photosynthetic reaction center complexes. *Photosynthesis Research*, Vol.41, No.1, (1994), pp. 285-294, ISSN 0166-8595



Advances in Photosynthesis - Fundamental Aspects

Edited by Dr Mohammad Najafpour

ISBN 978-953-307-928-8

Hard cover, 588 pages

Publisher InTech

Published online 15, February, 2012

Published in print edition February, 2012

Photosynthesis is one of the most important reactions on Earth. It is a scientific field that is the topic of many research groups. This book is aimed at providing the fundamental aspects of photosynthesis, and the results collected from different research groups. There are three sections in this book: light and photosynthesis, the path of carbon in photosynthesis, and special topics in photosynthesis. In each section important topics in the subject are discussed and (or) reviewed by experts in each book chapter.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jeffrey W. Touchman and Yih-Kuang Lu (2012). The Core- and Pan-Genomes of Photosynthetic Prokaryotes, *Advances in Photosynthesis - Fundamental Aspects*, Dr Mohammad Najafpour (Ed.), ISBN: 978-953-307-928-8, InTech, Available from: <http://www.intechopen.com/books/advances-in-photosynthesis-fundamental-aspects/the-core-and-pan-genomes-of-photosynthetic-prokaryotes>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen