S9-001

Further evidence and new molecular insight on the origin and early evolution of photosynthesis

Jin Xiong and Carl E. Bauer

Department of Biology, Indiana University, Bloomington, IN 47405, USA.
jxiong@bio.indiana.edu; cbauer@bio.indiana.edu

Keywords: evolution, phylogeny, bacteriochlorophylls, photosynthetic reaction center

Introduction

Photosynthesis is an ancient biological process. The advent of photosynthesis, especially oxygenic photosynthesis, fundamentally changed the redox balance on Earth and permitted the development of more advanced life forms. However, due to the involvement of multiple biochemical components in the process of photosynthesis and the lack of molecular sequence information for these components, the nature of the evolutionary path of photosynthesis has eluded researchers for decades. Previous evolutionary analysis of photosynthetic reaction centers which are the best indicator of evolution of photosynthesis, has established that both type I and type II reaction center apoproteins may share a common evolutionary origin (Schubert et al. 1998). However, no unified phylogenetic tree can be constructed for the two types of reaction centers because little sequence homology exists among the reaction center core apoproteins (Blankenship 2001).

As one of the first comprehensive phylogenetic comparisons, we recently analyzed Mg-tetrapyrrole biosynthesis genes and enzymes across the entire photosynthetic domain and used them as molecular markers to indicate the early evolution of photosynthesis (Xiong et al. 2000). Here we address potential problems regarding phylogenetic artifacts and provide additional phylogenetic evidence on early evolution of photosynthesis which indicates that bacteriochlorophyll biosynthesis genes may indeed be ancestral to chlorophyll biosynthesis genes.

Methods

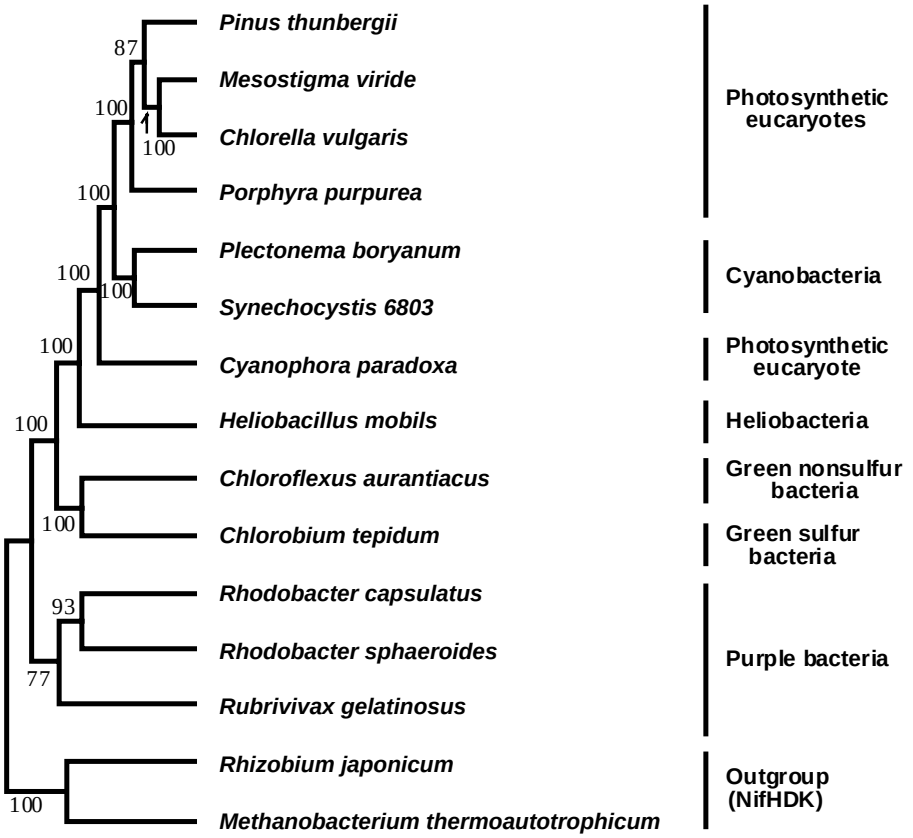
Amino acid sequence homologs of BchL, BchN and BchB proteins were extracted from the GenBank database. The sequence alignment was performed as described in Xiong et al. (2000). Individually sequence alignments were concatenaed into a large data set that contained 1337 positions. A phylogenetic tree was derived using the grouped corrected parsimony (GCP) analysis (Willson 1999) which was developed to minimize effects of long branch attraction errors. The tree building process utilized an error-rectification algorithm to strengthen the phylogenetic signal to resolve a correct tree. A consensus tree was built from 100 replicates. Ancestral amino acid sequences of reaction center proteins at internal nodes of phylogenetic trees were inferred with a likelihood-based Bayesian method using the CODEML program in the PAML package (Yang 2000), which performed inference using the Jones, Taylor and Thornton amino acid substitution model (JTT model, Jones et al. 1992).

Results and discussion

Based on analyses of 9 different bacteriochlorophyll biosynthesis genes and one concatenated set of genes using the distance neigbor-joining, maximum likelihood (ML) and maximum

parsimony (MP) methods, Xiong et al. (2000) showed that purple bacteria were most likely the basal lineage. Other conclusions include that (1) chlorophyll biosynthesis in oxygenic phototrophs evolved after the bacteriochlorophyll biosynthesis pathway which is found in anoxygenic phototrophs, which contradicts the long-favored Granick hypothesis (Granick 1965); (2) photopigment biosynthesis genes from heliobacteria have a close relationship with chlorophyll biosynthesis genes from cyanobacteria and (3) green sulfur bacteria and green non-sulfur bacteria have a closely related set of bacteriochlorophyll biosynthesis genes. This latter point is striking given that these different green bacterial lineages contain vastly different reaction center apoproteins (type I in green sulfur bacteria and type II in green non-sulfur bacteria).

However, as in all phylogenetic analysis, there is always a concern that the analysis result may be biased by the so-called "long branch attraction" effect, which means that a highly divergent taxon that forms a long branch in a phylogenetic tree may potentially disrupt the true phylogeny by clustering with another very long branch regardless of its true position. In order to avoid potential phylogenetic artifacts, we took multiple precautions to minimize long branch attraction errors. For example, we used the largest possible taxon sampling size for individual genes/proteins, which allowed the phylogenetic data to escape from the "Felsenstein zone" thereby reducing the formation of long branches. In our rooted phylogenetic analysis, we used only conserved and reliably aligned sequence regions from the outgroup sequences in order to minimize the artifactual effects derived from the use of distant outgroups. To assess the stability of the ingroup tree topology, which could be influenced by the addition of outgroup lineages due to long branch attraction, we also analyzed the phylogenetic trees with and without chosen outgroups. This additional analysis found no alteration of the ingroup topology. We also constructed a large concatenated data set, which can further minimize potential long branch attraction by simultaneously analyzing genes with different functions. Since it is highly unlikely for a single taxa to have very high evolutionary rates for all the genes involved, the construction of a large concatenated set of genes will dampen the effect of a single gene that may have a higher rate of evolution. The trees from the concatenated data set were very similar to those observed with individual gene trees. When performing the ML analysis, we also employed extensive rate categorizations which account for multiple substitutions at individual sites. This method downweights the changes at fast evolving sites which are often the cause of long branch attraction. In addition, we have carried out a grouped corrected parsimony (GCP) analysis which is a higher order parsimony method developed to further minimize the effect of long branch attraction (Willson 1999). Our GCP analysis essentially corroborates the results from our previous analysis. A GCP tree of concatenated BchL, BchN and BchB proteins with concatenated NifH, NifD and NifK as outgroup is shown in Fig. 1. It is in congruence with our previous analyses that purple bacteria are the basal group. Thus, we believe that the Xiong et al. (2000) paper provided carefully constructed and correctly rooted phylogenetic trees.



BchLNB-NifHDK

Fig. 1. Phylogenetic tree of concatenated BchL, BchN and BchB protein sequences using concatenated NifH, NifD and NifK sequences as outgroup (1337 positions for 15 taxa). Bootstrap frequencies are indicated at each node. Major photosynthetic groups to which the ingroup species belong are indicated on the right of the taxon names.

The above conclusions are supported by several additional lines of evidence. For example, the conclusion that anoxygenic photosynthesis evolved prior to oxygenic photosynthesis is supported by the ferrous iron oxidation experiment (Widdel et al. 1993). The conclusion that heliobacteria are closest relatives to cyanobacteria is supported by a recent phylogenetic analysis of ribosomal protein genes showing cyanobacteria and gram-positive were sister groups (Hansmann and Martin 2000). The close phylogenetic relationship between green sulfur and green non-sulfur bacteria is supported by the common antenna pigment, bacteriochlorophyll *c*, and the bacteriochlorophyll *c*-containing chlorosomes for light harvesting present in these two lineages.

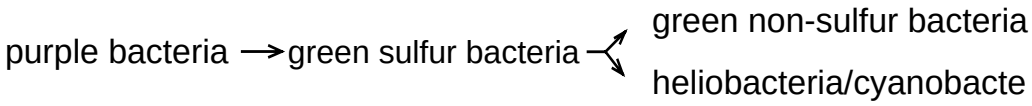
In addition, there is another published theory regarding the origin of photosynthesis that seems to be in line with our result that purple photosynthetic bacteria represent the earliest photosynthetic bacterial lineage. Nisbet et al. (1995) proposed that photosynthesis arose from bacteriochlorophyll-containing organisms which initially used the pigments for phototaxis under infrared or near-infrared radiation from oceanic hydrothermal systems. Based on the spectral analysis showing a close match of the absorption spectra of bacteriochlorophylls *a* and *b* from purple bacteria with the emission spectra of the geothermal light, Nisbet et al. suggested that an early bacterium capable of movement may have first developed a bacteriochlorophyll-like pigment absorbing infrared light to detect the geothermal light so that the organism could preferentially occupy an optimum habitat in water. Rudimentary photosynthesis may have subsequently evolved as a supplement to chemotrophy, giving the

organism added selective advantage. Through further adaptation of the primitive photosystem, the organism would start making use of the near-infrared part of sunlight when moving to shallow water. Eventually chlorophylls would be developed to make use of higher-energy (visible) light to split water. The hypothesis appears to be supported by the observation of infrared phototaxis of purple bacteria (Ragatz et al. 1994). Furthermore, the amount of energy emitted from the geothermal light has been shown to be sufficient to drive photochemical reactions (White et al. 2000).

The phylogeny based on bacteriochlorophyll biosynthesis genes is clearly incongruent with 16S rRNA based phylogeny (Pace 1997), which suggests that lateral gene transfer process may have been involved in the evolution of photosynthesis. As more and more prokaryotic genomes have been sequenced, the chimeric or mosaic nature of these genomes is being more and more recognized. Thus these results supports the notion that a gene phylogeny in prokaryotes may maximally reflect the evolution of a metabolic process instead of the entire genome. Lateral gene transfers are in fact being recognized as the major driving force of prokaryotic evolution (Koonin et al. 2000). Furthermore, Doolittle (1999) has suggested that the evolutionary history of life should be best represented by a "mangrove-like network" rather than by the well-known "universal tree".

A major ramification of the above new evolutionary thinking is that cyanobacteria or green bacteria etc. may have existed in the past for a long period of time as non-photosynthetic organisms before acquiring photosynthesis genes. Thus, this brings to an interesting discussion of whether the well known 3.5 billion year old microfossils that look like current-day cyanobacteria may represent evidence for the earliest emergence of photosynthesis. Despite recent analysis suggesting that the cells represented by fossils could well be green non-sulfur bacteria such as *Chloroflexus* (Pierson 1994) and the analysis by Buick (1991) questioning the biogenic nature of the objects, we believe that the objects may well be genuine cyanobacterial microfossils. However, linking microfossil morphology with metabolic evolution can be problematic. In light of the pervasive lateral gene transfers in the ancient bacterial community, the genomes of the so-called cyanobacteria may not have encoded any photosynthesis genes at that particular time period. Additionally, even if the 3.5 billion year old "cyanobacteria" were photosynthetic (Schopf 1993) containing a similar genome as current cyanobacteria, there still remains a high possibility that anoxygenic photosynthesis by purple bacteria evolved earlier than oxygenic photosynthesis performed by cyanobacteria because the earliest record of photosynthesis was dated back to 3.8 billion years ago (Schidlowski 1988) leaving a long enough time frame (300 million years) for the development of photosynthesis in the anoxygenic form. Thus, in light of the tremendous controversies in deciphering the origin of photosynthesis, we believe that more definitive geochemical and molecular phylogenetic evidence may be needed for the resolution of the problem.

Regarding the issue of the direction of lateral gene transfers, we provided further evidence by infering ancestral sequences of *bch* genes at internal nodes of phylogenetic trees. This was conducted by using a likelihood-based Bayesian method (Yang et al. 1995; Koshi et al. 1996) with a JTT amino acid substitution model (Jones et al. 1992). Using such techniques, we have reconstructed ancestral amino acid sequences at internal nodes of several "purple basal" phylogenetic trees. For these trees, our Bayesian analysis shows that the ancestral sequence for all ingroup sequences belongs to purple bacteria. The ancestral sequence at the node before the divergence of green bacteria and heliobacteria is green sulfur, and the nodal sequence between green sulfur and green non-sulfur is also green sulfur. Thus, for these purple-basal phylogenetic trees, a more accurate route for the bacteriochlorophyll biosynthesis gene transfer appears to be:



This resolution of direction of evolution by the Baysian analysis indicates that the lateral gene transfer for (bacterio)chlorophyll biosynthesis genes may not be multidirectional but rather in the direction consistent with the phylogenetic interpretation by Xiong et al. (2000).

Overall, we believe that this additional analysis has provided new molecular insight on the origin and early evolution of photosynthesis and also pointed out some issues that may remain candidates for debates in the years to come. Through these debates, we believe a wider range of research will be stimulated and a clearer picture of the evolutionary process of photosynthesis will eventually emerge.

Acknowledgements

We thank H. Gest and R. Blankenship for helpful comments. The work is supported by an NIH grant (GM53940) and an Arizona State University NASA astrobiology grant.

References

Blankenship, RE (2001) *Trends in Plant Sciences* **6**, 4-6.
Buick, R (1991) *Palaios* **5**, 441-459.
Doolittle, WF (1999) *Science* **284**, 2124-2128.
Granick, S (1965) *In* Bryson, V, Vogel, HJ [Eds] *Evolving genes and proteins.* Academic Press, New York, pp. 67-68.
Hansmann, S, Martin, W (2000) *International Journal of Systematic and Evolutionary Microbiology* **50**, 1655-1663.
Jones DT, Taylor WR, Thornton, JM (1992) *Computational Applications in Biosciences* **8**, 275-282.
Koonin, EV, Aravind, L, Kondrashov, AS (2000) *Cell* **101**, 573-576.
Koshi, JM, Goldstein, RA (1996) *Journal of Molecular Biology* **42**, 313-320.
Nisbet, EG, Cann, JR, van Dover, CL (1995) *Nature* **373**, 479-480.
Pace, NP (1997) Science **276**, 734-740.
Pierson, BK (1994) *In* Bengston, S [Ed]. *Early Life on Earth.* Nobel Symposium, No. 84, Columbia University Press, New York, pp. 161-180.
Ragatz, L, Jiang, Z-Y, Bauer, C, Gest, H (1994) *Nature* **370**, 104.
Schidlowski, M (1988) *Nature* **333**, 313-318.
Schopf, JW (1993) *Science* **260**, 640-646.
Schubert W-D, Klukas, O, Saenger, W, Witt, HT, Fromme, P, Krauss, N (1998) *Journal of Molecular Biology* **280**, 297-314.
White, SN, Chave, AD, Reynolds, GT, Gaidos, EJ, Tyson, JA, van Dover, CL (2000) *Geophysical Research Letters* **27**, 1151-1154.
Widdel, F, Schnell, S, Heising, S, Ehrenreich, A, Assmus, B, Schink, B (1993) *Nature* **362**, 834-836.
Willson, SJ (1999) *Mol. Biol Evol.* **16**, 694-705.
Xiong, J, Fisher, WM, Inoue, K, Nakahara, M, Bauer, CE (2000) *Science* **289**, 1724-30.
Yang, Z (2000) *Phylogenetic Analysis by Maximum Likelihood (PAML).* (University College London, London, England), version 3.0.
Yang, Z, Kumar, S, Nei, M (1995) *Genetics* **141**, 1641-1650.