

- at 3 km. The $\delta^{18}\text{O}$ record for *G. ruber* is very similar to the previously published record for *G. sacculifer* from 806B, as well as other piston cores from the OJP, but our sampling resolution is twice that of the previous study of Hole 806B. We slightly modified the existing age model using our *G. ruber* $\delta^{18}\text{O}$ record, generated at twice the resolution, correlated to the standard SPEC-MAP chronology. We also corrected the core-top age to the typical radiocarbon age of core-tops from this region and depth (34). The average sedimentation rate is about 2 cm/ky.
34. W. S. Broecker, E. Clark, D. C. McCorkle, I. Hajdas, G. Bonani, *Paleoceanography* **14**, 13 (1999).
 35. N. J. Shackleton, J. Le, A. Mix, M. A. Hall, *Quat. Sci. Rev.* **11**, 387 (1992).
 36. Spectral analysis of Hole 806B data indicates that the overall cross correlation between Mg/Ca and $\delta^{18}\text{O}$ is $r = -0.73$, with Mg/Ca leading by 3 ky. The 100-, 41-, and 23-ky orbital periods in the $\delta^{18}\text{O}$ and Mg/Ca records are all coherent at the 95% CI.
 37. We determined $\delta^{18}\text{O}$ and Mg/Ca in *G. sacculifer* shells from the top 150 cm of Hole 806B. As is observed for TR163-19, the Mg/Ca values for *G. sacculifer* are systematically lower than for *G. ruber* and the $\delta^{18}\text{O}$ values are systematically more positive. The *G. sacculifer* Mg/Ca change over termination I is 2.8 to 3.6 mmol/mol, equivalent to a 2.8°C increase in SST and essentially identical to the SST change calculated from the *G. ruber* data.
 38. M. T. McCulloch et al., *Science* **283**, 202 (1999).
 39. H. M. Stoll, D. P. Schrag, S. C. Clemens, *Geochim. Cosmochim. Acta* **63**, 3535 (1999).
 40. P. A. Martin, D. W. Lea, T. A. Mashiotto, T. Papenfuss, M. Sarntin, *Geochim. Geophys. Geosyst.* **1**, Paper Number 1999GC000006 (1999).
 41. J. F. McManus, D. W. Oppo, J. L. Cullen, *Science* **283**, 971 (1999).
 42. L. W. Kroenke et al., *Proc. Ocean Drill. Program*, E. M. Barbu, Ed. (Ocean Drilling Program, College Station, TX, 1991), vol. 130.
 43. S. Levitus and T. P. Boyer, *World Ocean Atlas 1994, Volume 4: Temperature*, NOAA Atlas NESDIS (U.S. Department of Commerce, Washington, DC, 1994). Accessed at <http://ingrid.ldeo.columbia.edu/SOURCES/LEVITUS94/>
 44. D. J. Andreasen and A. C. Ravelo, *Paleoceanography* **12**, 395 (1997).
 45. We used a low-light paleotemperature equation derived for *Orbulina universa* (46): $\delta^{18}\text{O}_w = (T - 16.5 + 4.8 \cdot \delta^{18}\text{O}_{\text{calite}}) / 4.8 + 0.27$. Because the slopes of $\delta^{18}\text{O}$ change versus temperature are similar (0.20 to 0.23‰ per °C), the choice of paleotemperature equations does not have a large impact on the calculations. The estimated uncertainty of the $\delta^{18}\text{O}_w$ values is $\pm 0.18\text{‰}$, calculated from the uncertainty in the paleotemperature equation, the standard error of the Mg/Ca-SST calibration, and the reproducibility of $\delta^{18}\text{O}$ and Mg/Ca in a typical interval.
 46. B. E. Bemis, H. J. Spero, J. Bijma, D. W. Lea, *Paleoceanography* **13**, 150 (1998).
 47. S. Levitus, R. Burgett, T. P. Boyer, *World Ocean Atlas 1994, Volume 3: Salinity*, NOAA Atlas NESDIS (U.S. Department of Commerce, Washington, D.C., 1994). Accessed at <http://ingrid.ldeo.columbia.edu/SOURCES/LEVITUS94/>
 48. R. G. Fairbanks et al., *Coral Reefs* **16**, S93 (1997).
 49. R. G. Fairbanks, M. Sverdløve, R. Free, P. H. Wiebe, A. W. H. Bé, *Nature* **298**, 841 (1982).
 50. D. P. Schrag, G. Hampt, D. W. Murray, *Science* **272**, 1930 (1996).
 51. W. S. Broecker, *Paleoceanography* **4**, 207 (1989).
 52. I. J. Winograd et al., *Science* **258**, 255 (1992).
 53. J. Imbrie et al., in *Milankovitch and Climate, Part 1*, A. L. Berger, J. Imbrie, J. Hays, G. Kukla, B. Saltzman, Eds. (Reidel, Dordrecht, Netherlands, 1984), pp. 269–305.
 54. J. R. Petit et al., *Nature* **399**, 429 (1999).
 55. We thank J. Kennett and D. McCorkle for samples; T. Crowley, E. Bard, H. Elderfield, N. Pisias, and P. Martin for comments on earlier drafts; M. Kashgarian and T. Guilderson for radiocarbon dating; P. Howell for time-series software and advice on its use; I. Winograd for the suggestion to compare to the Devils Hole record; Q. Xie, G. Paradi, and H. Berg for mass spectrometer operation and maintenance; D. Gates, A. Schilla, P. Dekens, A. Davé, P. von Langen, L. Juranek, and M. Thomas for sample preparation; and E. Christian for clerical assistance. This research was funded by the NSF.

29 March 2000; accepted 13 July 2000

Molecular Evidence for the Early Evolution of Photosynthesis

Jin Xiong,¹ William M. Fischer,¹ Kazuhito Inoue,² Masaaki Nakahara,² Carl E. Bauer^{1*}

The origin and evolution of photosynthesis have long remained enigmatic due to a lack of sequence information of photosynthesis genes across the entire photosynthetic domain. To probe early evolutionary history of photosynthesis, we obtained new sequence information of a number of photosynthesis genes from the green sulfur bacterium *Chlorobium tepidum* and the green nonsulfur bacterium *Chloroflexus aurantiacus*. A total of 31 open reading frames that encode enzymes involved in bacteriochlorophyll/porphyrin biosynthesis, carotenoid biosynthesis, and photosynthetic electron transfer were identified in about 100 kilobase pairs of genomic sequence. Phylogenetic analyses of multiple magnesium-tetrapyrrole biosynthesis genes using a combination of distance, maximum parsimony, and maximum likelihood methods indicate that heliobacteria are closest to the last common ancestor of all oxygenic photosynthetic lineages and that green sulfur bacteria and green nonsulfur bacteria are each other's closest relatives. Parsimony and distance analyses further identify purple bacteria as the earliest emerging photosynthetic lineage. These results challenge previous conclusions based on 16S ribosomal RNA and Hsp60/Hsp70 analyses that green nonsulfur bacteria or heliobacteria are the earliest phototrophs. The overall consensus of our phylogenetic analysis, that bacteriochlorophyll biosynthesis evolved before chlorophyll biosynthesis, also argues against the long-held Granick hypothesis.

The advent of photosynthesis is one of the central events in the early development of life on Earth. The origin and evolution of photo-

synthesis, however, have long remained unresolved. Studies have demonstrated that photosynthetic eukaryotes acquired photosynthetic properties from endosymbiosis with cyanobacteria (1). This observation, coupled with the fact that no Mg-tetrapyrrole-based photosynthesis has been found in Archaea, supports the notion that photosynthesis is a bacterially derived process (2). To obtain insight into the early evolution of photosynthe-

sis, it is essential to conduct detailed phylogenetic analysis of many photosynthesis genes from each of the five known photosynthetic bacterial lineages. However, a paucity of photosynthesis gene sequences across the entire spectrum of photosynthetic bacteria has required that previous analyses rely on the use of nonphotosynthesis genes, which have given conflicting results for the evolution of photosynthesis and of photosynthetic organisms. For example, phylogenetic analysis of small-subunit rRNA suggests that green nonsulfur bacteria are the earliest evolving photosynthetic lineage (3). In contrast, using portions of the Hsp60 and Hsp70 heat shock proteins as markers, Gupta et al. (4) concluded that heliobacteria are the earliest evolving photosynthetic lineage and that this lineage subsequently diverged to green nonsulfur bacteria, cyanobacteria, green sulfur bacteria, and purple bacteria, in that order. The conflicting trees derived from such studies indicate that extrapolating the evolution of photosynthesis from nonphotosynthesis gene trees may be invalid.

Another problem arises when only a single set of photosynthesis genes is used for phylogeny. Previous attempts to analyze the evolution of photosynthesis using photosynthetic reaction center apoproteins failed to construct a phylogeny that includes all five photosynthetic bacterial lineages, because anoxygenic photosynthetic bacteria contain only one type of photosynthetic reaction center (type I or type II), whereas cyanobacteria contain both types of reaction center. Though the two types of reaction centers share significant structural similarities (5), their sequences have diverged to such an extent that it is virtually impossible to perform a statistically

¹Department of Biology, Indiana University, Bloomington, IN 47405, USA. ²Department of Biological Sciences, Kanagawa University, Hiratsuka, Kanagawa, 259-1293, Japan.

*To whom correspondence should be addressed. E-mail: cbauer@bio.indiana.edu

meaningful phylogenetic analysis using reaction center apoproteins as markers. Thus, tracking the evolution of photosynthesis must instead rely on analysis of an alternative set of photosynthesis genes present in all known photosynthetic lineages. Here, we have attempted to resolve the issue of the early evolutionary path of photosynthesis by obtaining a large number of photosynthesis gene sequences from the green sulfur bacterium *Chlorobium tepidum* and the green non-sulfur bacterium *Chloroflexus aurantiacus*, representatives of the two main photosynthetic lineages for which only a few photosynthesis genes have previously been sequenced so far. The additional sequence information on Mg-tetrapyrrole biosynthesis has allowed the first detailed phylogenetic reconstruction of photosynthesis genes for all photosynthetic lineages.

New photosynthesis gene sequences. Using a combination of functional complementation, degenerate polymerase chain reaction (PCR), inverse PCR, and automated DNA-sequencing techniques, we obtained eight DNA segments totaling 57,534 base pairs (bp) of genomic DNA containing 50 open reading frames (ORFs) from *C. tepidum*, and six DNA segments with a total of 40,766 bp of DNA containing 38 ORFs from *C. aurantiacus* (Fig. 1). Functions of 53 ORFs from both species were assigned with a relatively high degree of confidence based on sequence homology analysis (Table 1). Among these, 22 ORFs from *C. tepidum* and nine ORFs from *C. aurantiacus* were identified as encoding proteins involved in various steps of photosynthesis, with a total of 23 genes dedicated to bacteriochlorophyll biosynthesis (Table 1). The remaining photosynthesis genes are divided between those encoding proteins involved in carotenoid biosynthesis (one gene), protoporphyrin IX biosynthesis (two genes), carbon fixation (one gene), light-harvesting (one gene), and photosynthetic electron transfer (three genes). Another 22 ORFs were functionally assigned to products not directly related to photosynthesis (Table 1). There are also 10 ORFs from *C. tepidum* and four ORFs from *C. aurantiacus* that match conserved hypothetical proteins in the database that have no known functions. A total of three ORFs from *C. tepidum* and 10 ORFs from *C. aurantiacus* were found to be unique to the two green bacteria, having no database matches.

Unlike purple bacteria and heliobacteria, which both have large, tightly linked photosynthesis gene clusters (6), there are no major photosynthesis gene clusters in *C. tepidum* or *C. aurantiacus*. However, there are small clusters of two to three photosynthesis genes that are fully conserved in linkage among the four anoxygenic photosynthetic bacterial lineages (Fig. 2). A notable feature is that these

gene groups tend to encode products that show physical interactions with each other and make up subunits of a single enzyme or protein complex. The interactions of some of the protein pairs shown in Fig. 2 have already been experimentally confirmed. These include Bchl/ChlI and BchD/ChlD (7); BchD/ChlD and BchH/ChlH (8), which are subunits of Mg-chelatase; BchN/ChlN and BchB/ChlB (9), subunits of light-independent protochlorophyllide reductase; and PetB and PetC, subunits of the cytochrome bc_1 complex (10). The coupling of gene order with functional and physical interactions at the encoded protein level has been previously

observed in the genomes of a number of divergent nonphotosynthetic bacterial and archaeal species (11). In addition, we have observed a partial conservation of gene linkage between *bchM* and *bchE*, *bchE* and *bchJ*, and *bchJ* and *bchG*. These gene pairs tend to encode enzymes catalyzing two consecutive steps in the biosynthetic pathway.

Phylogenetic analysis. The newly obtained green bacterial photosynthesis genes allowed us to perform the first detailed evolutionary analysis of (bacterio)chlorophyll biosynthesis genes from all photosynthetic lineages. For this analysis, phylogenetic trees were constructed for both protein and DNA

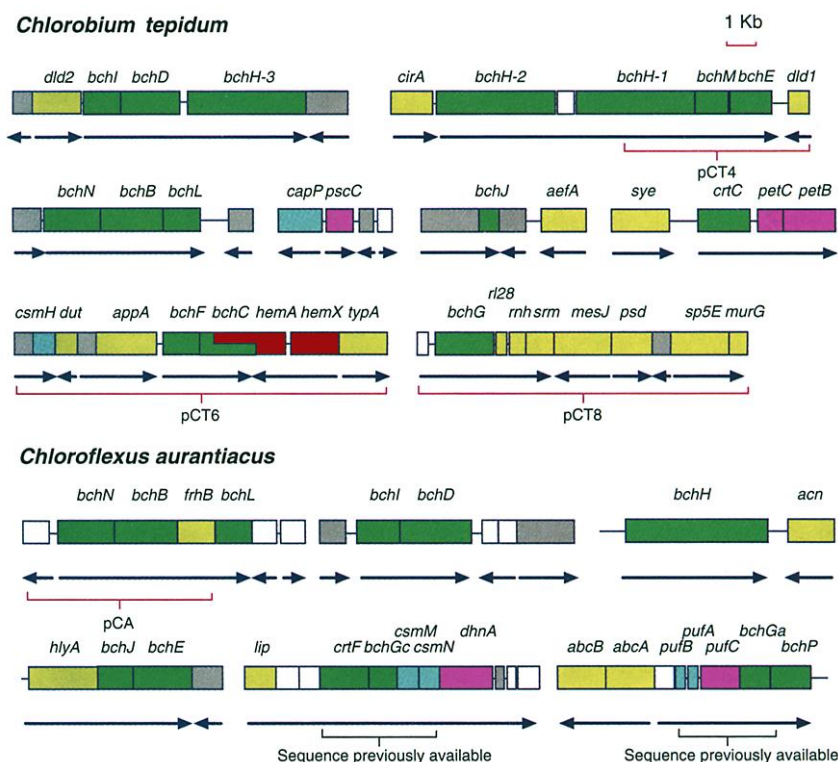
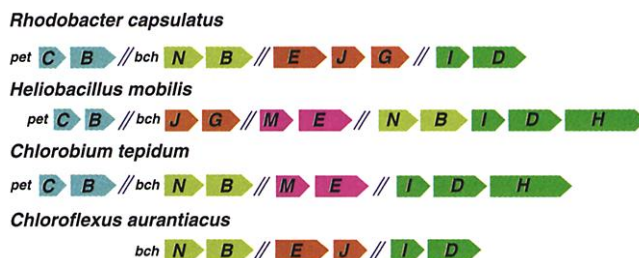


Fig. 1. Linear representation of newly identified photosynthesis genes and their flanking sequences from *C. tepidum* and *C. aurantiacus*. The predicted protein coding regions are colored by biological role: green boxes for bacteriochlorophyll and carotenoid biosynthesis genes, red for porphyrin biosynthesis genes, magenta for genes encoding electron transfer proteins, cyan for genes involved in carbon fixation, yellow for nonphotosynthesis genes, and gray for genes of conserved hypothetical proteins with no known functions; white boxes indicate unknown ORFs. Arrows represent the orientations of transcription. Cosmid insert sequences identified from functional complementation of bacteriochlorophyll-deficient strains of *Rhodobacter capsulatus* (25) are delineated by brackets.

Fig. 2. Comparison of linkage of photosynthesis genes among anoxygenic photosynthetic bacteria. Conserved small clusters of two to three genes shown here tend to encode proteins that physically interact with each other. Directions of transcription for each gene are indicated by arrowed boxes. Intergenic regions of unspecified lengths are indicated by "/".



sequences using neighbor-joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) methods (12). In all cases, the largest number of taxa available was sampled to avoid systematic errors such as long-branch attraction. Our ML analysis of gene sequences employed a method that takes into account site-to-site variability in evolutionary rates. Failure to address this phenomenon can result in long-branch attraction artifacts (13). In addition, in our rooted phylogenetic analysis, we used only conserved and reliably aligned sequence regions from the outgroup sequences in order to minimize potential phylogenetic reconstruction artifacts derived from the use of distant outgroups. In order 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 analyzed the phylogenetic trees with and without chosen outgroups: No alteration of the ingroup topology was found. Our nucleotide sequence analyses included only the first and second codon positions, in order to avoid potential substitutional saturation and compositional bias at the third position.

A phylogenetic tree of a gene encoding one of the three subunits of light-independent protochlorophyllide reductase, *bchB/chlB*, was rooted with its close homologs *nifK* and *nifD*, which encode corresponding subunits of nitrogenase, and *nifE* and *nifN*, which are responsible for the nitrogenase FeMo cofactor biosynthesis (14). Strong structural and functional similarities between *bchB/chlB* and the *nifD/nifK* gene products have been previously demonstrated (9). All three treeing methods, NJ, MP, and ML, yielded the same tree topology with *bchB/chlB* and the *nif* genes forming two distinct monophyletic groups, suggesting their evolution from an ancient gene duplication event (Fig. 3A). Thus, the *nif* genes can indeed be effectively used as outgroups to root the photosynthetic *bchB/chlB* subtree. In the *bchB/chlB* subtree, the purple bacterial taxa form a monophyletic group that was placed as the most basal lineage among the photosynthetic ingroup with strong bootstrap support (89%, 91%, and 98% for NJ, MP, and ML methods, respectively). The next-diverging lineage consists of a green bacterial clade with both *C. aurantiacus* and *C. tepidum* as sister taxa. The next-diverging lineage, the heliobacterium *Heliobacillus mobilis*, branches before the divergence of the cyanobacteria/plant lineages. The *bchB/chlB* DNA tree is completely consistent with the BchB/ChlB protein tree (15).

A similar phylogenetic tree topology was observed when the MP method was used for the *bchH* gene encoding the H subunit of Mg-chelatase (Fig. 3B, left). The tree is rooted with a close homolog, *cobN*, encoding a subunit of Co-chelatase. Another closely re-

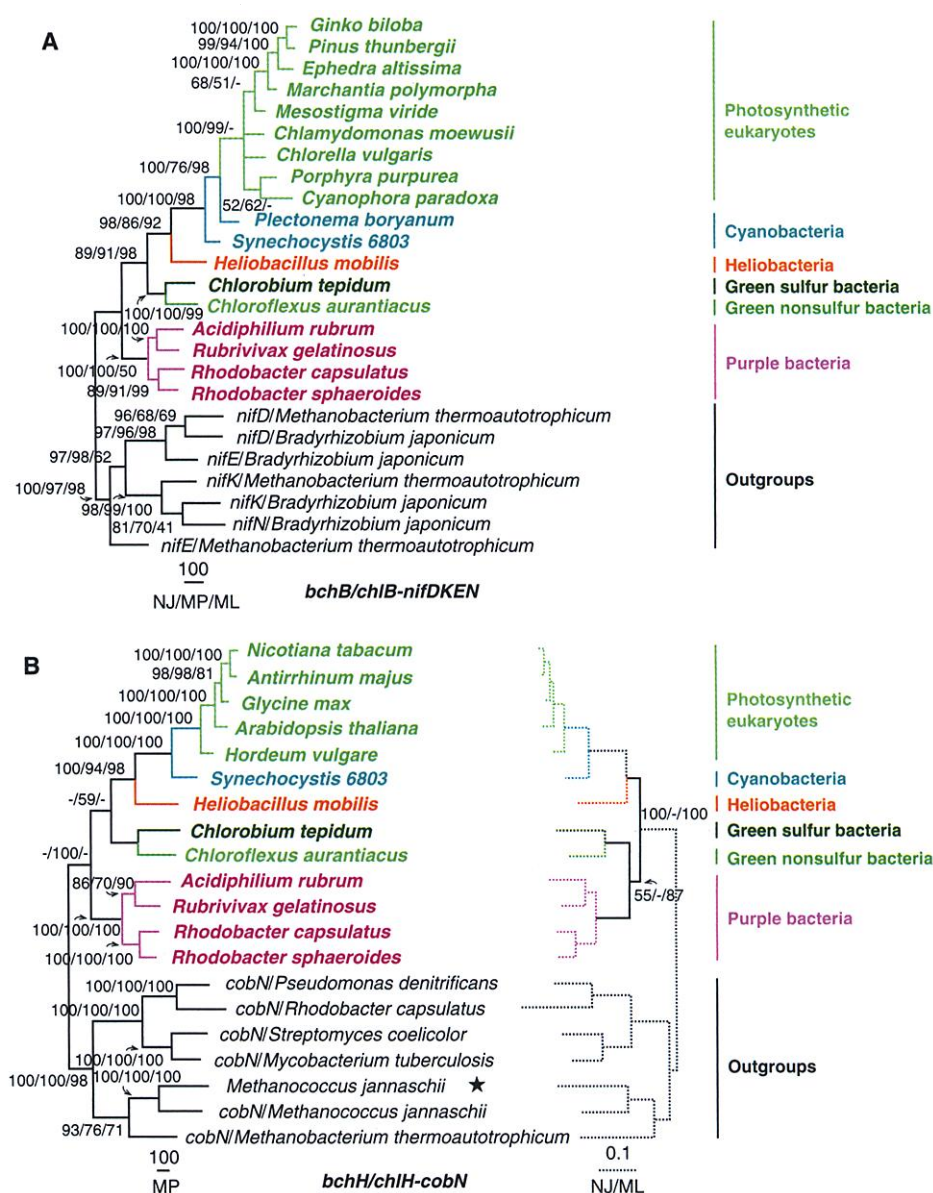


Fig. 3. Photosynthesis phylogeny based on multiple photosynthesis gene markers. (A) Phylogenetic tree for the *bchB/chlB* gene (with first and second codon positions) using close homologs *nifD*, *nifK*, *nifE*, and *nifN* from Bacteria and Archaea as outgroups. The phylogeny was derived using NJ, MP, and ML analyses. All three methods resulted in identical tree topology. Only the branch lengths for the MP tree are shown (length, 6601 steps; CI, 0.42; RI, 0.47). Bootstrap values >40% for NJ (first), MP (second), and ML (third) are indicated near the base of each branch. Major photosynthetic groups to which the ingroup species belong are indicated on the right of the taxon names. (B) Phylogenetic trees of *bchH* (with first and second codon positions) using *cobN* from Bacteria and Archaea and a putative Ni-chelatase in Archaea (indicated by a star) as outgroups. The three phylogenetic methods resolved slightly differently at the basal node for the photosynthetic ingroup. The MP tree (length, 10674 steps; CI, 0.52; RI, 0.54) is shown on the left, and the alternative topology from the NJ and ML analyses is shown on the right of the taxon names. (C) (Left panel) Phylogenetic tree for concatenated *bchI/chlI*, *bchD/chlD*, *bchH/chlH*, *bchL/chlL*, *bchN/chlN*, *bchB/chlB*, and *bchG/chlG* sequences common to all photosynthetic lineages (8504 characters for the first and second codon positions). The taxon named "Plants" is the concatenated sequence of *chlI*, *chlD*, and *chlH* from *Nicotiana tabacum*; *chlL*, *chlN*, and *chlB* from *Pinus thunbergii*; and *chlG* from *Arabidopsis thaliana*. A concatenated homologous sequence for *bchI*, *bchD*, *cobN*, *nifH*, *nifD*, *nifK*, and *bchG* from *M. thermoautotrophicum* is used as outgroup to all the photosynthetic lineages. The MP and NJ trees (left) differed slightly from the ML tree (right) at the basal node. (Right panel) The parsimony/distance gene phylogeny is thought to correlate with the structural divergence of (bacterio)chlorophyll pigments, for which key positions of difference are highlighted.

lated gene encoding a putative Ni-chelatase (indicated by a star) in methanobacteria was also included in the outgroup. For the photo-

synthetic ingroup, purple bacteria again form a moderately supported (59% bootstrap value) basal group. However, the ML and NJ

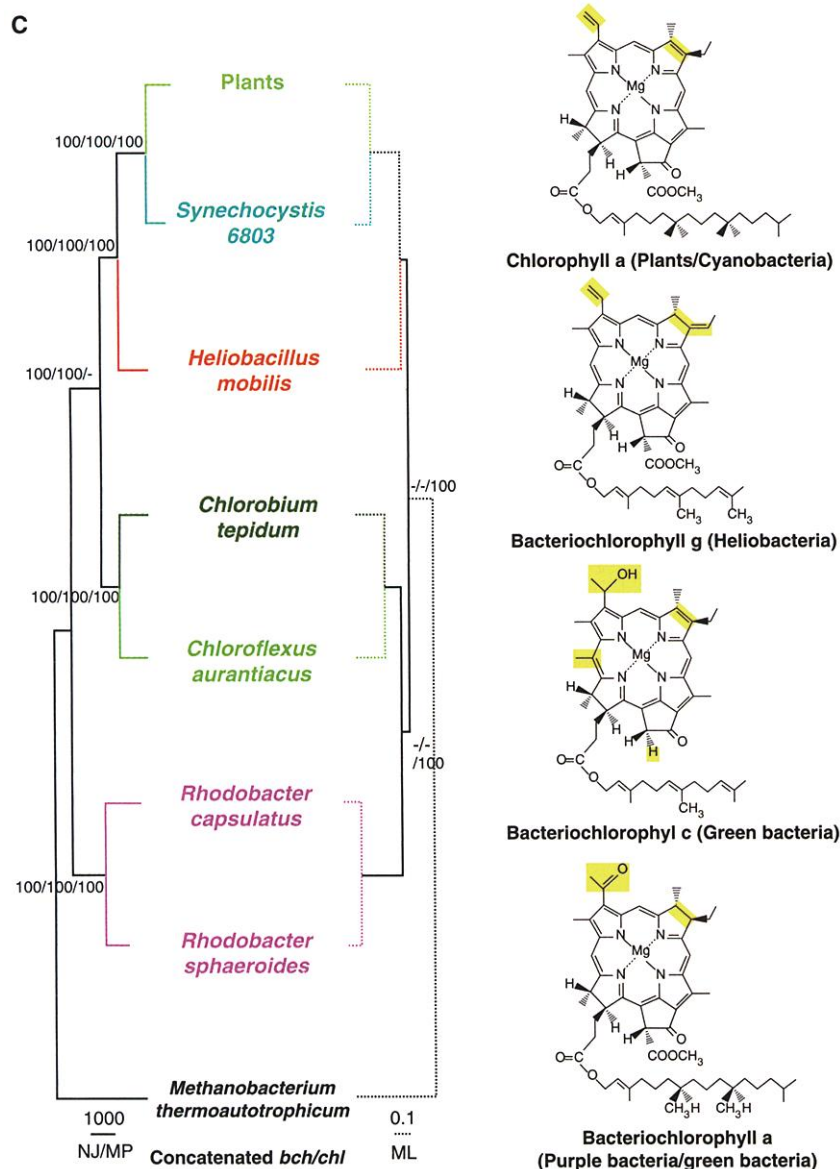


Fig. 3. (continued)

analysis shows a different topology at this basal node for the photosynthetic bacterial ingroup: purple bacteria and green bacteria cluster together to become a sister group to the heliobacterial/oxygenic clade (Fig. 3B, right). The protein BchH/ChlH tree topology derived from all three methods was similar to that of the DNA tree inferred by the ML and NJ methods (15). In all cases, green sulfur and green nonsulfur bacteria consistently are sister groups.

For seven other photosynthesis gene trees (*bchL/chlL*, *bchN/chlN*, *bchI/chlI*, *bchD/chlD*, *bchG/chlG*, *bchM/chlM*, and *bchJ/chlJ*) and protein trees (BchL/ChlL, BchN/ChlN, BchI/ChlI, BchD/ChlD, BchG/ChlG, BchM/ChlM, and BchJ/ChlJ), they typically either resolve into the same topology as observed for the *bchB/chlB* tree, which places purple bacteria as the most basal group, or into the

alternative topology as the *bchH/chlH* tree, which places purple and green bacteria as sister groups (Table 2).

In light of observed incongruence among some of the individual gene/protein trees, we carried out an additional analysis combining multiple photosynthesis genes or proteins into a single large data set. This approach is thought to improve the resolution of phylogenetic reconstruction by avoiding the biases that can result from analysis of individual genes or proteins (16). For this purpose, homologs of all (bacterio)chlorophyll biosynthesis genes (gene products) shared by all photosynthetic bacterial and plant lineages were individually aligned and concatenated into a single large data set that was used to infer a phylogenetic tree (17). The combined DNA data set (*bchI/chlI*, *bchD/chlD*, *bchH/chlH*, *bchL/chlL*, *bchN/chlN*, *bchB/chlB*, and

bchG/chlG) contains 8504 characters, and the combined protein data set (BchI/ChlI, BchD/ChlD, BchH/ChlH, BchL/ChlL, BchN/ChlN, BchB/ChlB, and BchG/ChlG) contains 3967 characters. The corresponding homologs of the above genes or gene products from *Methanobacterium thermoautotrophicum*, which are thought to be involved in cofactor F430 biosynthesis, Co chelation, and nitrogen reduction, were concatenated and used as an outgroup. Though the number of taxa in the combined data set is limited, rendering it more sensitive to the long-branch attraction, the simultaneous use of various genes with different functions makes it unlikely that a single taxon would have very high evolutionary rates for all the genes used. As shown in Fig. 3C (left side of left panel), the purple bacteria basal topology seen in most of the individual data sets is strongly supported by the MP and NJ analysis (100% bootstrap value). The tree also has 100% bootstrap support for all other nodes. The result obtained from the combined DNA data set is also consistent with that for the protein data set in the MP and NJ analyses. This confirms the conclusions that (i) green sulfur and green nonsulfur bacteria are each other's closest relatives that are rooted intermediately between heliobacteria and purple bacteria, and (ii) that heliobacteria are closest to the last common ancestor of all oxygenic lineages. Though generally congruent with the MP and NJ trees, the phylogeny derived from the ML analysis (Fig. 3C, right side of left panel) has a nearest-neighbor interchange in the node between the purple bacterial lineage and the green bacterial lineage, resulting in a topology similar to that of the ML tree for *bchH/chlH* (Fig. 3B, right). In view of the overall consensus of the individual and combined gene/protein analyses, purple bacteria are shown to receive the most support as the basal lineage for the photosynthetic bacteria. However, unambiguous phylogenetic determination of this branching order may have to await larger sampling of photosynthetic bacterial taxa and larger collection of photosynthesis gene sequences.

The consensus phylogenetic relationship based on the MP and NJ analysis is also supported by biochemical evidence such as structural divergence of photosynthetic pigments. For example, bacteriochlorophyll g synthesized by heliobacteria closely resembles chlorophyll a of cyanobacteria. This is consistent with the close relationship between heliobacteria and cyanobacteria determined by the phylogenetic analysis for Mg-tetrapyrrole biosynthesis genes (Fig. 3C, right panel). Indeed, the only difference between chlorophyll a and bacteriochlorophyll g ring structure is an isomerization of a double bond at ring II that occurs spontaneously in bacteriochlorophyll g upon exposure to air (19). Bacteriochlorophyll c, a light-harvesting pigment

RESEARCH ARTICLES

Table 1. The predicted functions and database similarity matches of the ORFs of *Chlorobium tepidum* and *Chloroflexus aurantiacus* identified in this study (26). Abbreviations of organisms of best database matches are shown in brackets (the first being the best match for *C. tepidum*, the second being best match for *C. aurantiacus*): Ar, *Acidiphilium rubrum*; Bs, *Bacillus subtilis*; Cf, *Chloroflexus aurantiacus*; Cl, *Chlorobium limicola*; Crv, *Chlorella vulgaris*; Ct, *Chlorobium tepidum*; Cv, *Chlorobium vibrioforme*; Ec, *Escherichia coli*; Ef,

Enterococcus faecalis; Hi, *Haemophilus influenzae*; Hm, *Heliobacillus mobilis*; Hs, *Homo sapiens*; Mt, *Methanobacterium thermoautotrophicum*; Pb, *Plectonema boryanum*; Ps, *Pseudomonas* sp. KWI-56; Rg, *Rubrivivax gelatinosus*; Ro, *Rhodothermus obamensis*; Rs, *Rhodobacter sphaeroides*; Sc, *Saccharomyces cerevisiae*; Sy, *Synechocystis* sp. PCC 6803; Th, *Thermus thermophilus*; Tm, *Thermotoga maritima*; Tt, *Thermoanaerobacterium thermosulfurigenes*; and Xc, *Xanthomonas campestris*.

Gene		Most significant database match, [organism], functional role	Percent similarity (Ct/Cf)
<i>Chlorobium tepidum</i>	<i>Chloroflexus aurantiacus</i>		
Photosynthesis-related genes			
<i>bcbB</i>	<i>bcbB</i>	B subunit of light-independent protochlorophyllide reductase, [Hm/Pb], bacteriochlorophyll biosynthesis	58/58
<i>bcbC</i>		2-Desacetyl-2-hydroxyethyl bacteriochlorophyllide a dehydrogenase, [Rs], bacteriochlorophyll biosynthesis	43
<i>bcbD</i>	<i>bcbD</i>	D subunit of Mg-chelatase, [Cv/Cv], bacteriochlorophyll biosynthesis	97/69
<i>bcbE</i>	<i>bcbE</i>	Mg-protoporphyrin IX monomethyl ester oxidative cyclase, [Hm/Hm] bacteriochlorophyll biosynthesis	84/83
<i>bcbF</i>		2-Vinyl bacteriochlorophyll hydratase, [Ar], bacteriochlorophyll biosynthesis	67
<i>bcbG</i>		Geranylgeranyl bacteriochlorophyll synthase, [Rs], bacteriochlorophyll biosynthesis	61
<i>bcbH-1</i>	<i>bcbH</i>	H subunit of Mg-chelatase, [Hm/Hm], bacteriochlorophyll biosynthesis	62/61
<i>bcbH-2</i>		H subunit of Mg-chelatase, [Cv], bacteriochlorophyll biosynthesis	79
<i>bcbH-3</i>		H subunit of Mg-chelatase, [Cv], bacteriochlorophyll biosynthesis	98
<i>bchl</i>	<i>bchl</i>	I subunit of Mg-chelatase, [Cvl/Hm], bacteriochlorophyll biosynthesis	95/78
<i>bchJ</i>	<i>bchJ</i>	4-Vinyl reductase, [Hm/Hm], bacteriochlorophyll biosynthesis	68/64
<i>bchL</i>	<i>bchL</i>	L subunit of light-independent protochlorophyllide reductase, [Sy/Sy], bacteriochlorophyll biosynthesis	73/72
<i>bchM</i>		Mg-protoporphyrin IX methyl transferase, [Sy], bacteriochlorophyll biosynthesis	52
<i>bchN</i>	<i>bchN</i>	N subunit of light-independent protochlorophyllide reductase, [Hm/Crv], bacteriochlorophyll biosynthesis	64/62
	<i>bchP</i>	Geranylgeranyl hydrogenase, [Sy], bacteriochlorophyll biosynthesis	59
<i>capP</i> (partial)		Phosphoenolpyruvate carboxylase, [Ro], carbon fixation	50
<i>crtC</i>		Hydroxyneurosporene synthase, [Rg], carotenoid biosynthesis	45
<i>csnH</i>		Chlorosome envelope protein, [Ct], light harvesting	100
<i>hemA</i>		Glutamyl-tRNA reductase, [Cv], porphyrin biosynthesis	91
<i>hemX</i>		HemX protein, [Bs], porphyrin biosynthesis (down-regulation of <i>hemA</i>)	45
<i>petB</i>		Cytochrome b of the cytochrome bc complex, [Cl], electron transfer	97
<i>petC</i>		Rieske iron sulfur protein of cytochrome bc complex, [Cl], electron transfer	87
<i>pscC</i>		Cytochrome c, [Cr], electron transfer	100
Non-photosynthesis genes			
<i>aefA</i>		AEF membrane protein [Ec], virulence factor	
	<i>abcA</i>	ABC transporter, [Tt], ATP-binding component of a transport system	59
	<i>abcB</i>	ABC transporter, [Sy], ATP-binding component of a transport system	61
	(partial)		
	<i>acn</i>	Aconitase, [Xc], tricarboxylic acid cycle and regulation of translation by binding to iron metabolism mRNA	76
	(partial)		
<i>appA</i>		Oligopeptide binding protein, [Bs], a component of oligopeptide permease for binding and transport of oligopeptides	45
<i>cirA</i> (partial)		Similar to colicin I receptor, [Ec], iron transport.	46
	<i>dhnA</i>	NADH dehydrogenase [Bs], respiratory electron transfer	62
<i>dldI</i> (partial)		Dehydrolipoamide dehydrogenase, [Bs], catalysis of NAD ⁺ -dependent oxidation of dihydrolipoyl groups of lipoate acyltransferase	58
<i>dld2</i>		Dehydrolipoamide dehydrogenase, [Cv], catalysis of NAD ⁺ -dependent oxidation of dihydrolipoyl groups of lipoate acyltransferase	89
<i>dut</i>		Deoxyuridine 5'-triphosphate nucleotidohydrolase, [Ct], nucleotide metabolism	100
	<i>frhB</i>	β subunit of coenzyme F420 reducing hydrogenase, [Mt], coenzyme F420 reduction	44
	<i>hlyA</i>	Hemolysin, [Tm], bacterial exotoxin	65
	<i>lip</i>	Lipase, [Ps], lipid degradation	48
	(partial)		
<i>mesJ</i>		Cell cycle protein, [Hm], cell division	51
<i>murG</i> (partial)		Undecaprenyl-PP-N-acetylmuramic acid-pentapeptide N-acetylglucosamine transferase, [Ef], cell wall synthesis	69
<i>rl28</i>		50S ribosomal protein L28, [Hi], protein biosynthesis	74
<i>rnH</i>		Ribonuclease H, [Th], RNA degradation	69
<i>psd</i>		Phosphatidylserine decarboxylase, [Mt], lipid biosynthesis	65
<i>srm</i>		RNA binding protein [Hs], involved in RNA splicing; or cell surface protein [Sc]	
<i>sp5E</i>		Stage V sporulation protein E, [Bs], sporulation	57
<i>syE</i> (partial)		Glutamyl-tRNA synthetase, [Bs], protein biosynthesis	57
<i>typA</i> (partial)		GTP-binding protein (elongation factor EF-G), [Sy], protein biosynthesis	59

unique to green sulfur and green nonsulfur bacteria, is structurally more distantly related to chlorophyll a, because there are alterations in several functionally important side groups (Fig. 3C, right panel). In our analyses, green sulfur and green nonsulfur bacteria, which are grouped together, are also phylogenetically more distant from cyanobacteria than are heliobacteria. The structure of bacteriochlorophyll a, synthesized by both purple and green bacteria but more exclusively by purple bacteria, is most divergent relative to chlorophyll a, with alterations in both the macrocyclic ring and several side groups (Fig. 3C). These differences in bacteriochlorophyll structures are consistent with our phylogenetic placement of purple bacteria as the most divergent lineage relative to cyanobacteria on the basis of MP and NJ analysis.

Implications for early evolution of photosynthesis. Our conclusion that chlorophyll a biosynthesis evolved from a more complex bacteriochlorophyll biosynthesis pathway argues against the oft-cited Granick hypothesis for the evolution of chlorophylls (20). Granick proposed that chlorophyll a, which requires fewer biosynthetic steps in its production, evolved before bacteriochlorophylls. The molecular phylogeny for the *bch/chl* genes in this, and in other studies (21), strongly argues that cyanobacteria were late-evolving. If so, cyanobacteria-based chlorophyll a biosynthesis is a recent development in the course of evolution of photosynthetic pigments, which may have occurred by a shortening of the bacteriochlorophyll biosynthetic pathway. Because chlorophyll a absorbs more energetic wavelengths of light than do bacteriochlorophylls, it may be that there was a selection to use shorter wavelengths, which would provide the energy needed to drive oxidation of water.

It is of particular interest that our phylo-

genetic analysis of the Mg-tetrapyrrole biosynthesis genes and enzymes consistently placed green nonsulfur bacteria and green sulfur bacteria as closest relatives. The close relationship between these two different phyla is also supported by the fact that they both synthesize bacteriochlorophyll c-containing chlorosomes for light harvesting. This is in striking contrast to the different photosystem types contained in each lineage, with green nonsulfur bacteria having the type II photosynthetic reaction center and green sulfur bacteria having the type I reaction center. Phylogenetic analysis of the reaction center core polypeptides (22) shows that the reaction center apoprotein phylogenetic trees are incongruent with the pigment biosynthesis protein trees. As mentioned above, it is practically unfeasible to build a tree to compare both type I and type II apoproteins. Our further analysis (23) suggests that the evolution of pigment biosynthesis is a limiting or better determining factor for studying the evolution of photosynthesis, for which a linear phylogenetic representation of all photosynthetic lineages can be established.

It should be emphasized that there is a conceptual difference between the evolution of photosynthesis and the evolution of photosynthetic organisms. The former involves only a limited number of genes for this bioenergetic process, whereas the latter involves the whole genome, the evolution of which is often represented by strictly vertically inherited genes such as the small-subunit rRNA gene. The phylogeny of the small-subunit rRNA gene does not necessarily reflect the phylogeny of the genes for specific metabolic pathways. Indeed, detailed comparison of the phylogenetic trees of the 16S rRNA gene versus Mg-tetrapyrrole biosynthesis genes shows incongruence at the deep branches (3), suggesting that horizontal gene transfers of

the photosynthesis genes may have taken place during the evolution of Bacteria. Because the Mg-tetrapyrrole biosynthesis genes are shown to have largely co-evolved, these genes may have been transferred as an entity during early radiation from ancient purple bacteria, which are known to have all of their photosynthesis genes present in a tightly linked "photosynthesis gene cluster" (6). Since the prevalence of lateral gene transfer among prokaryotic microorganisms has been overwhelmingly demonstrated [see reviews in (24)], the gene phylogenies in prokaryotes may indicate only the evolution of specific metabolic processes rather than the evolution of the whole genome.

References and Notes

1. M. W. Gray, *Int. Rev. Cytol.* **141**, 233 (1992); C. W. Morden, C. F. Delwiche, M. Kuhsel, J. D. Palmer, *Biosystems* **28**, 75 (1992).
2. J. M. Olson and B. K. Pierson, *Int. Rev. Cytol.* **108**, 209 (1987); R. E. Blankenship, *Antonie Van Leeuwenhoek* **65**, 311 (1994); W. Nitschke, U. Mühlenhoff, U. Liebl, in *Photosynthesis: A Comprehensive Treatise*, A. Raghavendra, Ed. (Cambridge Univ. Press, Cambridge, 1997), pp. 285–304.
3. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987); G. J. Olsen, C. R. Woese, R. Overbeek, *J. Bacteriol.* **176**, 1 (1994); J. M. Neefs, Y. Van de Peer, P. De Rijk, S. Chapelle, R. De Wachter, *Nucleic Acids Res.* **21**, 3025 (1993).
4. R. S. Gupta, T. Mukhtar, B. Singh, *Mol. Microbiol.* **32**, 893 (1999).
5. W.-D. Schubert et al., *J. Mol. Biol.* **280**, 297 (1998); K. H. Rhee, E. P. Morris, J. Barber, W. Kühlbrandt, *Nature* **396**, 283 (1998).
6. M. Alberti, D. H. Burke, J. E. Hearst, in *Anoxygenic Photosynthetic Bacteria*, R. E. Blankenship, M. T. Madigan, C. E. Bauer, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1995), pp. 1083–1106; J. Xiong, K. Inoue, C. E. Bauer, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14881 (1998).
7. J. Papenbrock, S. Gräfe, E. Kruse, F. Hänel, B. Grimm, *Plant J.* **12**, 981 (1997).
8. S. Gräfe, H.-P. Saluz, B. Grimm, F. Hänel, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1941 (1999).
9. Y. Fujita and C. E. Bauer, *J. Biol. Chem.* **275**, 23583 (2000).
10. B. L. Trumpower, *Microbiol. Rev.* **54**, 101 (1990).
11. T. Dandekar, B. Snel, M. Huynen, P. Bork, *Trends Biochem. Sci.* **23**, 324 (1998).
12. Phylogenetic trees were generated by distance (neighbor-joining), maximum parsimony, and maximum likelihood methods for both DNA and protein sequences. NJ trees were generated by CLUSTALX (version 1.63b) using the method of N. Saitou and M. Nei [*Mol. Biol. Evol.* **4**, 406 (1987)]. The alignments were bootstrapped by 1000 replications with correction for multiple substitutions. MP trees were built using the PAUP* program (version 4.0b3a) with 100 full heuristic replicates while maintaining groupings with frequency >50%. Each search run was done with random taxon-addition-order replicates and with tree-bisection-reconnection (TBR) branch-swapping. Whenever applicable, a branch-and-bound search approach was also used to verify the heuristic search results. ML analyses for DNA sequences used the F84 model as incorporated in FASTDNAML [G. J. Olsen et al., *Comput. Appl. Biosci.* **10**, 41 (1994)]. To aid these analyses, we used PUZZLE [K. Strimmer and A. von Haessler, *Mol. Biol. Evol.* **13**, 964 (1996); version 4.02] to estimate a single-parameter transition/transversion (Ti/Tv) ratio under the Tamura-Nei model of evolution with the parameter estimation as "exact" and model of rate heterogeneity as "uniform." Five FASTDNAML taxon-addition-order replicates were performed using the "jumble" option and global swapping. The tree with the best log-likelihood score was used as an input tree topology for estimation of site-to-site rate

Table 2. Results of phylogenetic analysis of Mg-tetrapyrrole biosynthesis genes and enzymes using three analysis methods. NJ, distance-based neighbor joining; MP, maximum parsimony; ML, maximum likelihood. PB, purple bacterial most-basal phylogeny as in *bchB/chlB*; PG, purple bacteria and green bacteria as sister group or bifurcated basal phylogeny as in *bchH/chlH*; OT, other phylogenies without any common pattern.

Gene/protein (<i>bch/chl</i>)/ (<i>Bch/Chl</i>)	Number of taxa in ingroup/ outgroup	Phylogenetic methods					
		NJ		MP		ML	
		DNA	Protein	DNA	Protein	DNA	Protein
L	18/7	OT	PB	OT	PB	OT	PB
N	16/9	OT	PB	PB	PB	PB	OT
B	18/7	PB	PB	PB	PB	PB	PB
I	23/2	OT	PG	OT	PG	OT	OT
D	11/2	PG	PB	PB	PB	PG	PG
H	13/7	PG	PG	PB	PG	PG	PG
G	9/3	OT	OT	PG	OT	OT	OT
M	9/1	PB	PB	PB	PB	PB	PB
J	5/1	PB	PB	OT	PB	OT	PB
Concatenated data set	7/1	PB	PB	PB	PB	PG	PG

- variation categories using the DNARates program [G. J. Olsen, S. Pracht, R. Overbeek, <http://geta.life.uiuc.edu/~gary/programs/DNARates.html> (1998)]. The Ti/Tv ratio for each of these second-round FASTDNAML analyses was estimated with PUZZLE using 16 gamma-distributed rate categories. The process of DNA rate categorization and tree inference was repeated until a stable tree topology was achieved. Data sets for ML bootstrapping were generated using the SEQBOOT program in PHYLIP [J. Felsenstein, *Cladistics* **5**, 164 (1989)] and analyzed with FASTDNAML using the no-rates Ti/Tv ratio with global swapping across all nodes for each run. One hundred individual ML trees were obtained and used as input for DNARates and subsequent rounds of FASTDNAML, with the 16 gamma-rate Ti/Tv ratio. The resulting phylogenetic trees were compiled and analyzed using the CONSENSE program from PHYLIP. For protein analysis, we used an approximation to ML analysis: a heuristic search in PAUP* (with 10 random taxon-addition-order replicates and TBR branch-swapping) was based on a ML distance matrix calculated with PUZZLE. The distance matrix was generated using JTT substitution frequencies, amino acid usage estimated from the data, and site-to-site variation modeled on a gamma distribution with eight categories plus invariant sites and the shape parameter estimated from the data. Several different alignments and different masking of the conserved sequence regions were tested: No significant differences were found among the trees.
13. S.-M. Chaw *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4086 (2000); Y. Van de Peer and R. De Wachter, *J. Mol. Evol.* **45**, 619 (1997).
 14. P. J. Goodwin *et al.*, *Biochemistry* **37**, 10420 (1998).
 15. J. Xiong, W. Fischer, K. Inoue, M. Nakahara, C. E. Bauer, data not shown.
 16. Y.-L. Qiu *et al.*, *Nature* **402**, 404 (1999); J. P. Huelsenbeck, J. J. Bull, C. W. Cunningham, *Trends Ecol. Evol.* **11**, 152 (1996).
 17. To generate a composite phylogenetic tree of all (bacterio)chlorophyll biosynthesis genes or gene products, all *bch/chl* (Bch/Chl) homologs shared among the bacterial and plant lineages (*bchl/chlI*, *bchD/chlD*, *bchH/chlH*, *bchl/chlL*, *bchN/chlN*, *bcbB/chlB*, and *bchG/chlG* for DNA sequences and Bch/ChlI, BchD/ChlD, BchH/ChlH, Bchl/ChlL, BchN/ChlN, BcbB/ChlB, and BchG/

- ChlG for protein sequences) were also individually aligned and concatenated into a large data set containing 8504 positions in each nucleotide sequence (with the exclusion of the third codon position) and 3976 positions in each amino acid sequence. The combined data sets were used to infer phylogenetic trees. As shown in Fig. 1, there are three members in the *bchH* gene family in *C. tepidum*. Use of any one of the members in the calculation for the composite tree resulted in no difference in tree topology, though individual tree analysis showed different branching patterns for *bchH-1* (BchH-1) (Fig. 3B) relative to *bchH-2* (BchH-2) and *bchH-3* (BchH-3).
18. D. Moreira, H. L. Guyader, H. Phillippe, *Nature* **405**, 69 (2000).
 19. P. Beer-Romero, J. L. Favinger, H. Gest, *FEMS Microbiol. Lett.* **49**, 451 (1988).
 20. S. Granick, in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 67–68.
 21. J. Xiong, K. Inoue, C. E. Bauer, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14881 (1998); D. H. Burke, J. E. Heast, A. Sidow, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7134 (1993).
 22. R. E. Blankenship, *Photosynth. Res.* **33**, 91 (1992).
 23. J. Xiong and C. E. Bauer, in preparation.
 24. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000); W. F. Doolittle, *Science* **284**, 2124 (1999).
 25. Genomic DNA from *C. tepidum* (strain TLS, ATCC 49652) and *C. aurantiacus* (strain J10-fl) was isolated as previously described [A. Pospiech and B. Neumann, *Trends Genet.* **11**, 217 (1995)]. Photosynthesis genes from the two species were initially obtained through functional complementation of bacteriochlorophyll biosynthesis genes using pigment biosynthesis-deficient mutants of *Rhodobacter capsulatus*, as described [D. A. Young, C. E. Bauer, J. C. Williams, B. L. Marrs, *Mol. Gen. Genet.* **218**, 1 (1989)]. Three cosmids from *C. tepidum* (pCT4 for complementing *bchM⁻* strain, pCT6 for *bchF⁻*, and pCT8 for *bchG⁻*) and one cosmid from *C. aurantiacus* (pCA for *bchN⁻*) containing bacteriochlorophyll biosynthesis genes were obtained (Fig. 1). The cosmid inserts were subject to sequence analysis with the primer walking technique. DNA sequencing was performed with a fluorescent dye-labeled dideoxynucleotide sequencing reaction kit (Amersham) with an Ap-

- plied Biosystems DNA sequencer (model 377, Perkin-Elmer). Oligonucleotide primers were synthesized by Operon Technologies. All of the DNA sequences we reported were determined completely on both strands. The flanking sequence of some of the cosmids was further extended by using inverse PCR according to the established protocol [H. Ochman, A. S. Gerber, D. L. Hartl, *Genetics* **120**, 621 (1988)]. The rest of the new photosynthesis gene sequences were initially obtained by PCR using degenerate primers based on alignments of the most highly conserved regions in homologous protein sequences. Inverse PCR was subsequently employed to obtain further flanking sequences. DNA sequences reported here have been annotated and deposited in GenBank with the following accession numbers: AF286047, AF287480, AF287481, AF287482, AY005135, AY005136, AY005137, AY005138, AF288458, AF288459, AF288460, AF288461, AF288462, and AF288602.
26. The obtained nucleotide sequence was analyzed with the Sequencer software (version 3.0, Gene Codes). Sequence compilation and ORF prediction and translation were performed with the aid of Sequencer and the GCG sequence analysis package (version 9.0). Homologous sequences were searched and retrieved from GenBank using BLAST [S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997)] against a nonredundant protein database in GenBank. Multiple amino acid sequence alignments were generated with the aid of the computer program CLUSTALX (version 1.63b) and were further improved manually with the aid of SEQAPP (version 1.9) and SEQPUP (version 0.7) programs. The NH₂- or COOH-terminal overhangs and loops shared by <20% of the taxa as well as ambiguously aligned regions were excluded from further phylogenetic analysis. DNA sequences were aligned in frame according to the aligned corresponding amino acid sequences.
 27. We thank Y. Fujita, D. Kehoe, H. Gest, S. Turner, and J. Palmer for critical discussions and comments during the preparation of the manuscript and M. Madigan for providing the *Chloroflexus aurantiacus* culture. Supported by NIH grant GM53940 (C.E.B.), an Arizona State University NASA astrobiology grant (C.E.B.), and a U.S.–Japan Cooperative Program on Photoconversion and Photosynthesis Research (K.I.).

29 December 1999; accepted 25 July 2000

REPORTS

Quantized Phonon Spectrum of Single-Wall Carbon Nanotubes

J. Hone,^{1*} B. Batlogg,² Z. Benes,³ A. T. Johnson,^{1†} J. E. Fischer³

The electronic spectra of carbon nanotubes and other nanoscale systems are quantized because of their small radii. Similar quantization in the phonon spectra has been difficult to observe because of the far smaller energy scale. We probed this regime by measuring the temperature-dependent specific heat of purified single-wall nanotubes. The data show direct evidence of one-dimensional quantized phonon subbands. Above 4 kelvin, they are in excellent agreement with model calculations of individual nanotubes and differ markedly from the specific heat of two-dimensional graphene or three-dimensional graphite. Detailed modeling yields an energy of 4.3 millielectron volts for the lowest quantized phonon subband and a tube-tube (or "lattice") Debye energy of 1.1 millielectron volts, implying a small intertube coupling in bundles.

The electronic structure of single-wall carbon nanotubes (SWNTs) has been extensively studied and is known to reflect confinement of electron waves to the one-dimensional (1D) molecular cylinder. How-

ever, the low-energy phonon structure of SWNTs is largely unexplored experimentally despite considerable theoretical work. The low-energy phonons are related to the mechanical properties and define the ther-

mal conductivity (1, 2), which will determine whether applications such as thermal management in molecular electronics are feasible. In addition, detailed knowledge of the phonon structure is important for understanding electron-phonon scattering in nanotubes (3, 4).

The phonon spectrum in SWNTs should display quantum size effects, whereby the two-dimensional (2D) phonon bands of graphene fold into a set of quantized 1D subbands, as is seen in the electronic band structure (5, 6). In reduced dimensions, the fundamental physics of phonon scattering

¹Department of Physics and Astronomy and Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, PA 19104–6272, USA. ²Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA. ³Department of Materials Science and Engineering and Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, PA 19104–6272, USA.

*Present address: Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA.

†To whom correspondence should be addressed.