The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives

Akiko Tomitani[†‡], Andrew H. Knoll[§¶], Colleen M. Cavanaugh[§], and Terufumi Ohno[†]

[†]The Kyoto University Museum, Kyoto University, Yoshida-honmachi, Sakyo, Kyoto 606-8501, Japan; and [§]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138

Cyanobacteria have played a significant role in Earth history as primary producers and the ultimate source of atmospheric oxygen. To date, however, how and when the group diversified has remained unclear. Here, we combine molecular phylogenetic and paleontological studies to elucidate the pattern and timing of early cyanobacterial diversification. 16S rRNA, *rbcL*, and *hetR* genes were sequenced from 20 cyanobacterial strains distributed among 16 genera, with particular care taken to represent the known diversity of filamentous taxa. Unlike most other bacteria, some filamentous cyanobacteria evolved a degree of cell differentiation, producing both specialized cells for nitrogen fixation (heterocysts) and resting cells able to endure environmental stress (akinetes). Phylogenetic analyses support the hypothesis that cyanobacteria capable of cell differentiation are monophyletic, and the geological record provides both upper and lower bounds on the origin of this clade. Fossil akinetes have been identified in 1,650- to 1,400-mega-annum (Ma) cherts from Siberia, China, and Australia, and what may be the earliest known akinetes are preserved in ≈2,100-Ma chert from West Africa. Geochemical evidence suggests that oxygen first reached levels that would compromise nitrogen fixation (and hence select for heterocyst differentiation) 2,450–2,320 Ma. Integrating phylogenetic analyses and geological data, we suggest that the clade of cyanobacteria marked by cell differentiation diverged once between 2,450 and 2,100 Ma, providing an internal bacterial calibration point for studies of molecular evolution in early organisms.

molecular evolution | phylogeny | Proterozoic

Cyanobacteria are oxygenic photosynthetic bacteria that are widely distributed in aquatic and terrestrial environments, including such extreme habitats as hot springs, deserts, and polar regions (1). They are globally important primary producers today and have been through much of our planet's history (2, 3). Some diazotrophic cyanobacteria are reported to be important agents in the global nitrogen budget (3, 4); therefore, the group plays a significant role in the nitrogen cycle as well as in the cycles of oxygen and carbon. Moreover, it is generally accepted that the chloroplasts of plants and algae are derived from a cyanobacterial ancestor (5), implicating the blue–green bacteria in eukaryotic evolution. Without doubt, then, cyanobacteria are key to any understanding of Earth's early biological and environmental history.

Cyanobacteria are monophyletic but morphologically diverse. In traditional classifications, morphological distinctions have been used to divide the group into five subsections (6, 7). Cyanobacteria of subsection I (formerly Chroococcales) and II (Pleurocapsales) are unicellular coccoids. Cells of subsection I divide by binary fission, whereas those of subsection II can also undergo multiple fission to produce small, easily dispersed cells called baeocytes. Cyanobacteria of subsections III–V form filaments of varying morphological complexity. Filamentous cyanobacteria of subsection III (Oscillatoriales) have only vegetative cells, but in subsections IV (Nostocales) and V

(Stigonematales), vegetative cells can differentiate into morphologically and ultrastructurally distinct heterocysts, cells that specialize in nitrogen fixation under aerobic conditions (8), or akinetes, resting cells that survive environmental stresses such as cold and desiccation (9), depending on growth conditions. In addition, filaments of subsection V have complicated branching patterns. Their developmental variety and complexity are thus among the most highly developed in prokaryotes.

The geologic record offers some clues to the reconstruction of early cyanobacterial history. 2-methylhopanoids, molecular fossils known to be sourced by cyanobacteria, have been identified in 2,700-mega-annum (Ma) shales from Australia (10), although incomplete phylogenetic sampling leaves open the possibility that other bacteria might also produce this biomarker (11). Consistent with the cyanobacterial interpretation of early 2-methylhopanoids, 2,700-Ma lacustrine stromatolites, also from Australia, display features interpreted as products of carbonate accretion by cyanobacterial mats in a setting where electron donors other than water were limited (12). Microfossils with detailed and diagnostic morphological counterparts among modern cyanobacteria in subsections I–III occur in the ≈1,900-Ma Belcher Supergroup of Canada (13). However, given the generally poor preservation of microfossils in older rocks, these fossils must be regarded as a minimum constraint on cyanobacterial diversification.

Molecular phylogeny has become a powerful tool in elucidating evolutionary pattern, and analyses based on 16S rRNA sequences (14–16) indicate that cyanobacteria producing baeocytes (subsection II), heterocysts (subsections IV and V), and true-branching filaments (subsection V) are each phylogenetically coherent. In contrast, phylogenies reconstructed by using *nifH* (17) and *nifD* (18), structural genes for nitrogenase, the enzyme that catalyses biological nitrogen fixation, do not support monophyly of subsection V; the *nifH* phylogeny also indicates paraphyly of subsection II. Not all cyanobacteria fix nitrogen and, therefore, genes such as *nifH* or *nifD* cannot be used to analyze non-nitrogen-fixing species. Because even well resolved molecular phylogenies show only the relative timing of diversification events, direct evidence from the geologic record is required to constrain actual divergence times.
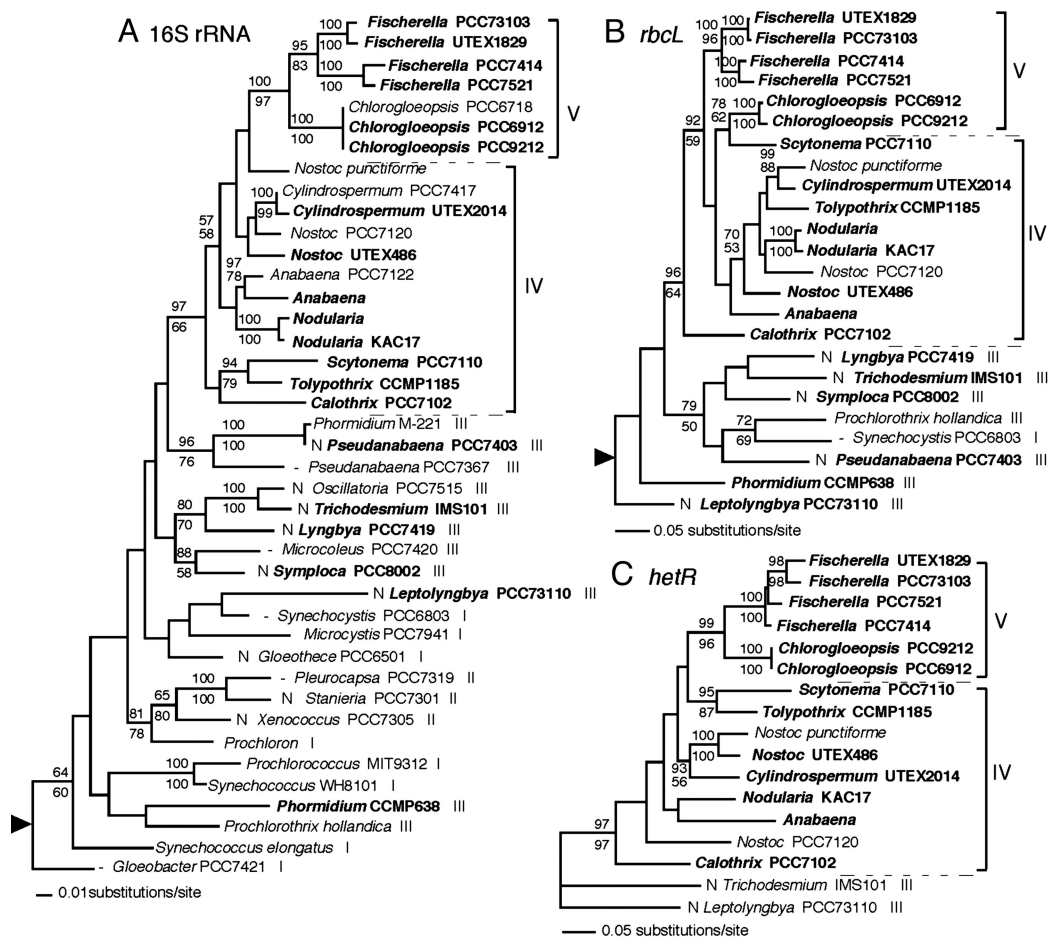
Fig. 1. The phylogenetic relationships of cyanobacteria inferred from 16S rRNA (*A*), *rbcL* (*B*), and *hetR* (*C*) nucleotide sequences; trees were constructed by the ML method (19). Roman numerals denote cyanobacterial subsections I–V; sequences obtained in this study are indicated in boldface. Numbers at each branch point are the bootstrap values (22) for percentages of 1,000 replicate trees calculated by NJ (20) (*Upper*) and MP (21) (*Lower*) methods; only values >50% are shown. Arrowheads mark the deepest roots determined by using *Agrobacterium* and *Prochlorococcus* as outgroups for 16S rRNA and *rbcL*, respectively. N, nonheterocystous cyanobacteria known to fix nitrogen; -, cyanobacteria for which nitrogen fixation has been tested but not detected.

To reveal how and when cyanobacteria evolved, we carried out molecular phylogenetic and paleontological studies in tandem. First, to elucidate the phylogenetic pattern of cyanobacterial diversification, we conducted a molecular phylogenetic study, choosing 20 cyanobacterial strains from 15 genera to represent the known diversity of filamentous taxa. We focused on filamentous cyanobacteria because of their developmental and paleoenvironmental interest. Because differentiated cell types have played an important role in the classification of both living and fossil cyanobacteria, we asked whether these characters fall within a single clade whose members can be recognized from well preserved fossils. In addition, cyanobacterial development is controlled by growth conditions, so the evolution of cell differentiation may reflect key events in Earth's environmental history. Three genes were analyzed here: 16S rRNA; *rbcL*, which codes for the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase, the key enzyme for CO_2 fixation by means of the Calvin cycle; and *hetR*, a gene essential for heterocyst differentiation. Trees obtained in these analyses were calibrated by using both original paleontological observations and published geochemical data that place upper and lower constraints on the timing of cyanobacterial diversification.

Results

Molecular Phylogenetic Analyses of Cyanobacterial 16S rRNA, *rbcL*, and *hetR* Sequences. Phylogenetic analyses of the 20 16S rRNA sequences show that filamentous cyanobacteria without cell differentiation (subsection III) mix with unicellular species (I and II), indicating that filamentous morphology has polyphyletic origins within cyanobacteria. Some cyanobacteria of subsections I–IV form clusters with conserved internal topology in all trees constructed by maximum likelihood (ML) (19), neighbor joining (NJ) (20), and maximum parsimony (MP) (21) methods (e.g., *Prochloron* and cyanobacteria of subsection II). However, relationships among these clusters varied depending on the analytic method used. In contrast, heterocyst- and akinete-bearing taxa (IV and V) form a monophyletic clade (Fig. 1*A*). Monophyly is supported by the ML method, by NJ with a bootstrap value (22) of 97%, and to a lesser extent by MP algorithms (bootstrap value of 66%). The 16S rRNA tree also indicated that members of subsection V are a monophyletic group nested within the broader clade of heterocyst- and akinete-bearing taxa.

To examine further the monophyly of cyanobacteria that produce differentiated cells, part of the *rbcL* gene was amplified from the same set of the 20 cyanobacteria and sequenced. The *rbcL* gene and its homologs have been sequenced previously from a number of organisms and fall into several subgroups, IA–ID and II–IV (23). All of the cyanobacterial *rbcL* genes obtained in this study fall in the subgroup IB (data not shown), a subgroup that contains the majority of cyanobacteria *rbcL* sequences and those of "green" chloroplasts (23). Based on the *rbcL* sequences, phylogenetic trees constructed by NJ, MP, and

ML methods all support the monophyly of heterocystous cyanobacteria (Fig. 1*B*). Monophyly also was supported when the third nucleotide in each codon was excluded from the analyses (data not shown). Several topological differences occur among the trees constructed by the three methods: (*i*) the positions of *Tolypothrix* and *Nostoc* PCC 7120 (formerly described as *Anabaena*) in the heterocystous clade in the NJ tree differ from MP and ML, (*ii*) *Phormidium* and *Leptolyngbya* form a cluster in the NJ and MP trees but are distributed separately in ML, and (*iii*) positions of *Symploca* and *Pseudanabaena* vary in a cluster composed of *Lyngbya*, *Trichodesmium*, *Symploca*, *Prochlorothrix*, *Synechocystis* PCC 6803, and *Pseudanabaena*, depending on the analytical method.

Partial *hetR* genes were amplified and sequenced from the 13 heterocystous cyanobacteria among our 20 experimental strains. Interestingly, genes homologous to *hetR* have been detected from some nonheterocystous nitrogen-fixing filamentous cyanobacteria (subsection III), although the function of those genes is not yet certain (24). The obtained sequences were aligned with *hetR* sequences of two nostocalean cyanobacteria whose genomes are completely sequenced and *hetR*-like genes of the oscillatorialeans *Leptolyngbya* (formerly described as *Plectonema*) PCC 73110 and *Trichodesmium* IMS 101. No gene known so far has significant similarity to *hetR*, and so an outgroup was not included in the calculation. Overall topology is similar among trees constructed by the NJ, MP, and ML methods. The members of subsection V form a monophyletic clade in the *hetR* tree constructed by ML (Fig. 1*C*). Monophyly of subsection V is also supported by the NJ and MP methods, with bootstrap values of 99% and 96%, respectively. Two cyanobacteria of subsection III fall outside of a cluster of heterocyst-producing species. The differences in topology among the trees constructed by the three methods are (*i*) relationships within *Fischerella* strains and (*ii*) the position of a cluster composed of *Nodularia* KAC17 and *Anabaena*. The outer location of subsection III and the monophyly of subsection V were also supported when the third nucleotide in each codon was excluded from *hetR* analyses, although relationships within subsection IV varied depending on the analytical method used (data not shown).

Fossil Akinetes. The genus *Archaeoellipsoides* consists of large ellipsoidal or cylindrical microfossils, which mostly occur as solitary individuals in rocks. They are preserved abundantly in ≈1,500-Ma cherts from the Billyakh Group of Siberia (Fig. 2*B*). Based on morphometric comparison with akinetes of the extant nostocalean genus *Anabaena* (e.g., Fig. 2*A*), Golubic *et al.* (25) interpreted *Archaeoellipsoides* as fossilized akinetes. The Billyakh fossils show no evidence of cell division, expected in vegetative cells, but do display features similar to those formed during akinete germination. They also occur in close association with short trichomes interpreted as the products of akinete germination (25).

Silicified carbonates of the ≈1,650-Ma (26) Amelia Dolomite of northern Australia also contain well preserved *Archaeoellipsoides* (Fig. 2*C*), as do 1,631 ± 5-Ma cherts from the Kheinjua Formation in India (27), preserved with other fossils representing a broad cross section of cyanobacterial diversity. The oldest fossils attributed to *Archaeoellipsoides* come from the ≈2,100-Ma Franceville Group of Gabon (28) (Fig. 2*D*). Although relatively poorly preserved, these fossils exhibit the same morphological features as those found in the more securely interpreted mid-Proterozoic populations.

Discussion

Cyanobacterial Phylogeny and Monophyly of Heterocystous Taxa. Our 16S rRNA analyses support the monophyly of heterocyst- and akinete-bearing cyanobacteria (subsections IV and V; Fig.
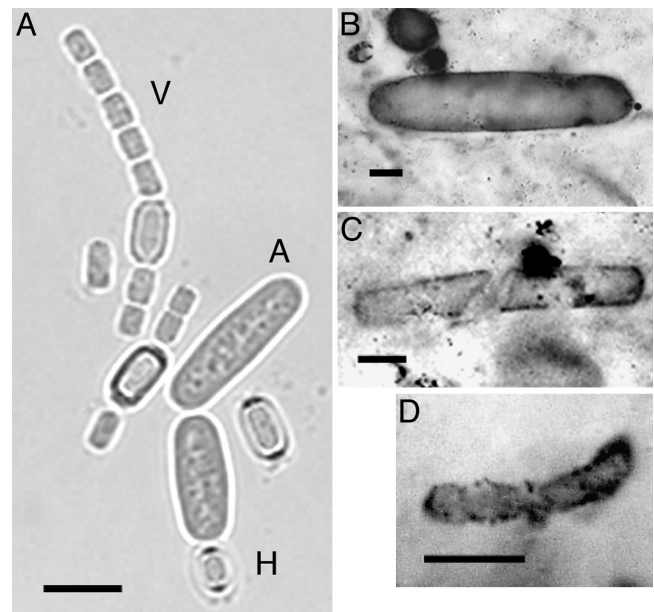

Fig. 2. Modern cyanobacterial akinetes and *Archaeoellipsoides* fossils. (*A*) Three-month-old culture of living *A. cylindrica* grown in a medium without combined nitrogen. A, akinete; H, heterocyst; V, vegetative cells. (*B*–*D*) Shown are *Archaeoellipsoides* fossils from 1,500-Ma Billyakh Group, northern Siberia (*B*); 1,650-Ma McArthur Group, northern Australia (*C*); and 2,100-Ma Franceville Group, Gabon (*D*). (Scale bars, 10 μm.)

1*A*). This finding is consistent with previous 16S rRNA phylogenies (14–16) as well as analyses of *nifH* (17) and *nifD* (18) and that of 36 genes collected from 14 cyanobacterial genomes (29). However, in the *nifH* tree, two nostocalean sequences did not cluster with other heterocystous–cyanobacterial *nifH* genes and thus were considered to be derived from gene duplication and/or gene transfer (17). In the genome-based tree (29), available sequences are still limited, and subsection V was not included. The monophyly of heterocystous taxa is also supported by the analyses of *rbcL* sequences (Fig. 1*B*). This clade is supported by the analyses using the ML, NJ, and MP methods for both 16S rRNA and *rbcL* (with and without the third nucleotide positions) sequences, although the MP method showed lower bootstrap support (66% and 64% for 16S rRNA and *rbcL*, respectively). In the *hetR* tree, the two oscillatorialean cyanobacteria that contain *hetR*-like genes lie outside of the heterocystous cluster (Fig. 1*C*), consistent with the 16S rRNA and *rbcL* results.

Turner *et al.* (15) grouped cyanobacteria into 10 monophyletic groups, including a group of plastid sequences, based on 16S rRNA sequence analysis. Although the composition of sampled organisms was not completely the same, some clusters in our 16S rRNA phylogeny were almost identical to Turner's groupings, including the *Nostoc* (NOST), *Pseudanabaena*, *Oscillatoria* (OSC), and *Synechococcus* sequence groups. However, statistical support for the clade comparable to OSC was not high (bootstrap value <50%) in our 16S rRNA tree. The *Synechocystis*/ *Pleurocapsa*/*Microcystis* sequence group identified by Turner *et al.* did not appear as a stable grouping in our 16S rRNA analysis, although part of its membership, *Prochloron* and the pleurocapsalean cyanobacteria, formed a cluster. In the *rbcL* phylogeny, the OSC and NOST sequence groups were supported. Congruence with other sequence groups is unknown because available *rbcL* sequences from subsections I–III are limited.

In our 16S rRNA phylogeny, heterocystous cyanobacteria with branching filaments (subsection V) are also monophyletic, nested within the broader grouping of heterocystous taxa. In the *rbcL* tree, however, subsection V (*Chlorogloeopsis* and *Fischerella*
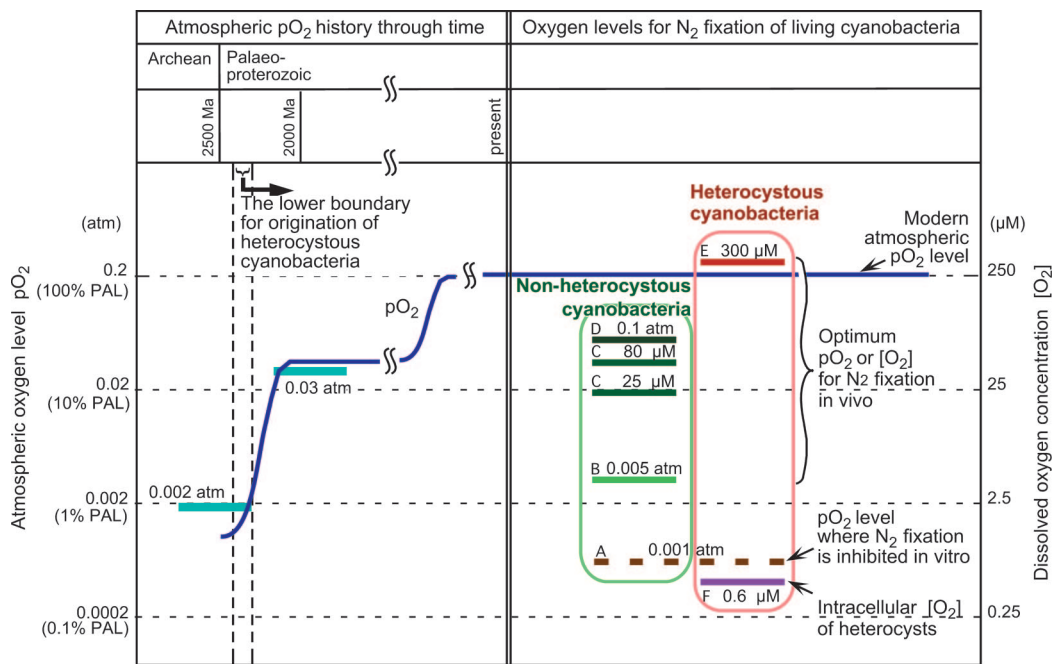
Fig. 3. The geochemical record of atmospheric oxygen as it relates to nitrogen fixation and the evolution of heterocystous cyanobacteria. The curve presents an estimate of atmospheric P_{O2} through time, as constrained by paleosol (42) and S isotopic (43, 44, 46) data. Numbers in *Right* indicate the P_{O2} level at which *in vitro* nitrogenase activity is inhibited in nonheterocystous *Plectonema* and heterocystous *Anabaena* (34) (A), and P_{O2} or $[O_2]$ optima for *in vivo* nitrogenase activity of *Plectonema* (36) (B), nonheterocystous *Gloeothece* (37) (C), *Trichodesmium* (38) (D), and *Anabaena* (39) (E). F indicates the intracellular $[O_2]$ of *Anabaena* heterocysts. The $[O_2]$ (the left vertical axis) corresponds to P_{O2} (right axis) when $[O_2]$ is 250 μM in air-saturated water. PAL, present atmospheric level.

in this study) did not form a cluster; instead, the two *Chlorogloeopsis* strains clustered with the nostocalean *Scytonema*. *hetR*, which plays a key role in the early stage of heterocyst differentiation (8) and is unique to filamentous cyanobacteria, should provide better resolution of the relationship between subsections IV and V. Sampled members of subsection V formed a monophyletic clade in our *hetR* trees, consistent with the 16S rRNA phylogeny (Fig. 1*A*), but not with analyses of *rbcL* (Fig. 1*B*), *nifH* (17), and *nifD* (18) phylogenies. 16S rRNA analysis of taxonomically more diverse stigonematalean cyanobacteria indicated polyphyletic origins among filamentous taxa (30); further investigation is necessary to settle their phylogenetic placement.

Species able to fix nitrogen (all taxa in subsections IV and V and some among I–III, marked with an "N" in Fig. 1) do not form a monophyletic group, indicating either that the ability to fix nitrogen existed in the common ancestor of cyanobacteria and was subsequently lost independently in multiple descendants or that this capability spread through the group by lateral gene transfer. Phylogenetic analyses of *nifH* and *nifD* from various nitrogen-fixing bacteria and archaea suggest that cyanobacterial genes form a cluster (17, 18). Although all cyanobacterial nitrogen fixation may have a common origin, it remains unclear whether the present distribution of this trait reflects vertical descent and secondary losses or horizontal gene transfer within the cyanobacteria.

The Upper Time Limit of Divergence of Cyanobacteria with Differentiated Cells. With varying degrees of confidence, many Precambrian (>543 Ma) microfossils have been linked to modern cyanobacteria of subsections I–III (31). Putative subsection V microfossils with complex branching filaments and possible heterocysts are preserved in the Lower Devonian (≈400 Ma) Rhynie Chert in Scotland (32), but fossil heterocysts have not been reliably identified in Precambrian rocks (25).

In contrast to heterocysts, akinetes survive postmortem decay better than vegetative cells (9, 25), and, as noted above, fossil cells interpreted as akinetes can be traced back well into the Paleoproterozoic Era. Akinetes of extant cyanobacteria occur only in the heterocystous species (7). Although phylogeny provides no way of determining unambiguously the relative timing of akinete and heterocyst origins, geochemical data discussed below suggest that the paleoenvironmental drivers of heterocyst evolution were in place before the earliest known records of akinetes.

The Lower Time Limit of Divergence of Heterocystous Cyanobacteria. Nitrogenase is widely distributed among prokaryotes and in all cases shows an extreme sensitivity to O_2 *in vitro* (33). For example, partial pressures of oxygen (P_{O2}) as low as 0.001 atmosphere (atm) inhibit *in vitro* nitrogenase activity of both the oscillatorialean *Plectonema boryanum* and the nostocacean *Anabaena cylindrica* (34) (Fig. 3*A*).

Most cyanobacteria capable of aerobic N_2 fixation produce heterocysts (33), whereas a variety of unicellular and simple filamentous forms can fix nitrogen only under reducing conditions (35). *P. boryanum*, for example, showed optimum nitrogen fixation at a P_{O2} of <0.005 atm (36) (Fig. 3*B*). A few nonheterocystous cyanobacteria fix nitrogen aerobically, but most of them show optimum nitrogen-fixing activity under low oxygen tension [e.g., at dissolved oxygen concentrations ($[O_2]$) of 25 and 80 μM (the exponential and stationary growth phases, respectively) for the chroococcalean *Gloeothece* (ref. 37 and Fig. 3*C*)]. An exception is the oscillatorialean *Trichodesmium*, which fixes nitrogen during the day in air-saturated water. *Trichodesmium* strain NIBB 1067 maximized nitrogen-fixing activity at a P_{O2} of 0.1 atm (Fig. 3*D*), although the culture used for this measurement was not axenic (38). In contrast, intact cells of *Anabaena variabilis* showed optimum nitrogen fixation at an $[O_2]$ of 300 μM (ref. 39 and Fig. 3*E*) at 35°C, which is above modern $[O_2]$ values in air-saturated water (220 μM at 35°C and 280 μM at 20°C). The intracellular $[O_2]$ of *Nostoc* sp. PCC 7120 heterocysts has been estimated to remain as low as 600 nM (J. Elhai, personal

communication; Fig. 3*F*), suggesting that heterocyst formers have a more efficient system to protect nitrogenase from oxygen than do nonheterocystous taxa. Considering that heterocyst development is arrested under anaerobic conditions (40) and that the rate of heterocyst differentiation is enhanced by increases in P_{O2} (41), the evolution of heterocysts may have accompanied the initial appearance of oxygen-rich environments in Earth's history.

Chemical analyses of iron retention in ancient weathering horizons suggest that P_{O2} rose from <0.002 atm to as much as 0.03 atm between ≈2,400 and 2,200 Ma (42). Sulfur-isotope analyses independently indicate a rapid increase of atmospheric oxygen between ≈2,450 and 2,320 Ma (43, 44). One might argue that heterocysts could have evolved earlier, in oxic habitat islands within thick mats, but the ecology of extant heterocyst-formers militates against this occurrence. Most heterocyst-forming cyanobacteria live in soil or freshwater environments where they are in direct contact with air and oxic waters; they are uncommon in marine microbial mats where oxic microhabitats might develop (45). Accordingly, cyanobacterial heterocysts likely appeared after 2,450–2,320 Ma, when P_{O2} first reached levels that inhibit nitrogenase activity (Fig. 3).

Time Calibration for the Tree of Life. Integrating molecular phylogenetic, physiological, paleontological, and geochemical data, we propose that the clade of cyanobacteria marked by heterocyst and akinete differentiation evolved once between 2,450 and 2,100 Ma. This constraint on the timing of a relatively late branch within the cyanobacteria provides a calibration point that may illuminate molecular evolutionary studies of deep nodes in the Tree of Life. As complete genome sequences of several cyanobacterial species have been determined, including the heterocystous *Nostoc* sp. PCC 7120 (47) and *Nostoc punctiforme* (48), the calibration can also be applied to expanding research in comparative genomics. Continuing study of Precambrian microfossils and taxon-specific biomarkers promises to place increasingly precise constraints on the timing of early cyanobacterial evolution. Conversely, continuing molecular biological research on living populations will add depth to the portrait of cyanobacterial evolution gleaned from an incomplete geological record.

Materials and Methods

Gene Isolation and Sequencing. Twenty filamentous cyanobacteria were selected for sequencing: six strains from subsection III, eight from IV, and six from V (shown in boldface in Fig. 1). Axenic cultures of *Fischerella* strain UTEX 1829, *Nostoc* UTEX 486, and *Cylindrospermum* UTEX 2014 were obtained from the Culture Collection of Algae at the University of Texas, Austin. *Chlorogloeopsis* PCC 6912 and 9212; *Fischerella* PCC 73103, 7414, and 7521; and *Pseudanabaena* PCC 7403 came from the Pasteur Culture Collection of Cyanobacteria (Paris). Isolates of *Anabaena* and *Nodularia*, together with *Calothrix* PCC 7102, *Lyngbya* PCC 7419, and *Scytonema* PCC 7110, were provided by J. B. Waterbury (Woods Hole Oceanographic Institution, Woods Hole, MA). DNA from cyanobacterial cells was extracted by standard methods designed for bacteria (49). Genomic DNA from *Tolypothrix* CCMP 1185 and *Phormidium* CCMP 638 was obtained from the Provasoli–Guillard National Center for Culture of Marine Phytoplankton (Boothbay Harbor, ME). Genomic DNA from *Nodularia* strain KAC17, *Symploca* PCC 8002, *Leptolyngbya* PCC 73110, and *Trichodesmium* IMS 101 was provided by S. Janson (University of Kalmar, Kalmar, Sweden). 16S rRNA was amplified by using a pair of primers designed specially for cyanobacteria (PLG1-1 and PLG2-1) (50).

Two pairs of primers for *rbcL* (GF-AB:GARTCTTCIA-CYGGTACYTGGAC and GR-D:TGCCAIACGTGGATAC-CACC; GF-C:CTTYACYCAAGACTGGGCTTC and GR-E:AACTCRAACTTGATTTCYTTCC) were designed based on the conserved regions in *rbcL* sequences of *Nostoc* sp. PCC 7120 (GenBank accession no. L02522), *Synechocystis* PCC 6803 (GenBank accession no. D64000), and *Prochlorothrix* (GenBank accession no. X57359). A pair of primers for *hetR* (HETR-F1:TATCTGGCTTTTAGYGCCATG and HETR-R1:CTTGGTGATATTTATCWGCCC) was designed based on conserved regions between *hetR* sequences of *Nostoc* PCC 7120 (GenBank accession no. M37779) and *N. punctiforme* (GenBank accession no. AF318069). PCRs were performed in a T-gradient thermal cycler (Biometra, Tampa, FL) by using the AmpliTaq Gold kit (Applied Biosystems). Amplified PCR products were cloned by using either the TOPO TA cloning kit (Invitrogen) or the pSTBlue-1 AccepTor kit (Novagen). Both strands of a minimum of three clones from each PCR product were sequenced with the Big Dye Terminator Cycle Sequencing kit version 2 (Applied Biosystems) by using a DNA sequencer (3100 Genetic Analyser, Applied Biosystems).

Phylogenetic Analyses. Twenty-one 16S rRNA sequences of cyanobacteria (GenBank accession nos. AB003165, AB039009, AB39006, AF001480, AF027655, AF053398, AF091108, AF091150, AF132777, AF132783, AF132789, AF132790, AF132792, AF132933, D64000, D83715, U40340, X59559, X63141, X70770, and X78680) and that of *Agrobacterium* (GenBank accession no. D14500) were obtained from GenBank, as were partial *rbcL* sequences of *Nostoc* PCC 7120 (GenBank accession no. L02522), *Synechocystis* PCC 6803 (GenBank accession no. D64000), *Prochlorothrix* (GenBank accession no. X57359), and *Prochlorococcus* MED 4 (accession no. BX572091), two *hetR* genes of *Nostoc* PCC 7120 (GenBank accession no. M37779) and *N. punctiforme* (GenBank accession no. AF318069), and two *hetR*-like genes of *Leptolyngbya* PCC 73110 (GenBank accession no. AF410433) and *Trichodesmium* IMS 101 (accession no. AF410432). The *rbcL* sequence of *N. punctiforme* was obtained from the web site of the Department of Energy Joint Genome Institute (Walnut Creek, CA).

Nucleotide sequences were aligned by using CLUSTAL X (51) with manual refinement. Sites where gaps existed or sequences were ambiguous were excluded. Secondary structures (52) were taken into consideration when the 16S rRNA sequences were aligned with *Agrobacterium* as an outgroup. The 20 *rbcL* sequences were aligned with the form IB *rbcL* sequences from three cyanobacteria (*Synechocystis* PCC 6803, *Nostoc* sp. PCC 7120, and *N. punctiforme*) whose genomes are completely sequenced and that of *Prochlorothrix*, with *Prochlorococcus* MED 4 (a form IA *rbcL*) as an outgroup (data not shown). The *hetR* sequences were aligned with sequences of two nostocalean cyanobacteria whose genomes are completely sequenced and *hetR*-like genes of the oscillatorialeans *Leptolyngbya* PCC 73110 and *Trichodesmium* IMS 101 (data not shown). Sequence lengths obtained here were as follows (base pairs of full sequence/base pairs used for analyses): 16S rRNA (1,145–1,160/1,117), *rbcL* (1,183/1,181), and *hetR* (742/737).

Phylogenetic trees were generated by using PAUP* 4.0b10 (53). The Hasegawa–Kishino–Yano 1985 model was used to model nucleotide substitution in the ML (19) and NJ (20) analyses. Base frequencies and transition transversion ratios were estimated from each data set. Each MP (21) calculation was performed by heuristic search of 100 replicates with the TBR (tree bisection and reconnection) option. For 16S rRNA, ML analyses were first done by heuristic search with the NNI (nearest-neighbor interchange) option, using the NJ and MP results as starting trees. To further investigate the 16S rRNA tree topology, heuristic search by ML with the TBR option was performed, using a topology constraint file that fixed groupings that appeared in all 16S rRNA trees obtained by the NJ, MP, and ML (NNI option) methods. The result obtained by ML with the TBR option was the same as that by ML with NNI. For *rbcL* and *hetR*, ML

Tomitani *et al.*

analyses were carried out by heuristic search with the TBR option, using results obtained from the NJ and MP analyses as starting trees.

Nitrogen fixation ability of PCC strains in subsections I–III ("N" in Fig. 1) is based on information provided by the Pasteur Culture Collection.

Paleontological Analyses. Fossils were identified by optical microscopy on thin sections of silicified carbonates deposited in coastal marine environments in a series of Paleoproterozoic and Mesoproterozoic basins. Materials from the ≈1,500-Ma Billyakh Group of northern Siberia and the 1,600- to 1,500-Ma McArthur Group of northern Australia collected by A.H.K. are reposited in the Paleobotanical Collections of the Harvard University Herbaria, as are fossiliferous cherts of the 2,100-Ma Franceville Group of West Africa. Additional fossiliferous samples of the Amelia Dolomite (McArthur Group) are reposited in the Commonwealth Palaeontological Collection in Canberra, Australia (collection CPC 15609 for the specimen shown in Fig. 2). A living cyanobacterial culture of *A. cylindrica* was provided by D. G. Adams (University of Leeds, Leeds, U.K.).

1. Whitton, B. A. & Potts, M. (2000) in *The Ecology of Cyanobacteria*, eds. Whitton, B. A. & Potts, M. (Kluwer, Dordrecht, The Netherlands), pp. 1–11.
2. Liu, H., Nolla, H. A. & Campbell, L. (1997) *Aquat. Microb. Ecol.* **12,** 39–47.
3. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. (1997) *Science* **276,** 1221–1229.
4. Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D. & Stal, L. (2002) *Biogeochemistry* **57/58,** 47–98.
5. Delwiche, C. F. & Palmer, J. D. (1997) in *Origins of the Algae and Their Plastids*, ed. Bhattachartya, D. (Springer, Berlin), pp. 53–96.
6. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. (1979) *J. Gen. Microbiol.* **111,** 1–61.
7. Castenholz, R. W. (2001) in *Bergey's Manual of Systematic Bacteriology*, eds. Boone, D. R. & Castenholz, R. W. (Springer, New York), 2nd Ed., Vol. 1, pp. 474–487.
8. Wolk, C. P., Ernst, A. & Elhai, J. (1994) in *The Molecular Biology of Cyanobacteria*, ed. Bryant, D. A. (Kluwer, Dordrecht, The Netherlands), pp. 769–823.
9. Herdman, M. (1987) in *The Cyanobacteria*, eds. Fay, P. & Van Baalen, C. (Elsevier, Amsterdam), pp. 227–250.
10. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. (1999) *Science* **285,** 1033–1036.
11. Knoll, A. H. (2003) *Geobiology* **1,** 3–14.
12. Buick, R. (1992) *Science* **255,** 74–77.
13. Hofmann, H. J. (1976) *J. Paleontol.* **50,** 1040–1073.
14. Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. & Pace, N. R. (1988) *J. Bacteriol.* **170,** 3584–3592.
15. Turner, S., Prayer, K. M., Miao, V. P. & Palmer, J. D. (1999) *J. Eukaryot. Microbiol.* **46,** 327–338.
16. Wilmotte, A. & Herdman, M. (2001) in *Bergey's Manual of Systematic Bacteriology*, eds. Boone, D. R. & Castenholz, R. W. (Springer, New York), 2nd Ed., Vol. 1, pp. 487–493.
17. Zehr, J. P., Mellon, M. T. & Hiorns, W. D. (1997) *Microbiology* **143,** 1443–1450.
18. Henson, B. J., Watson, L. E. & Barnum, S. R. (2004) *J. Mol. Evol.* **58,** 390–399.
19. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376.
20. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
21. Fitch, W. M. (1977) *Am. Nat.* **111,** 223–257.
22. Felsenstein, J. (1985) *Evolution (Lawrence, Kans.)* **39,** 783–791.
23. Tabita, F. R. (1999) *Photosynth. Res.* **60,** 1–28.
24. Janson, S., Matveyev, A. & Bergman, B. (1998) *FEMS Microbiol. Lett.* **168,** 173–179.
25. Golubic, S., Sergeev, V. N. & Knoll, A H. (1995) *Lethaia* **28,** 285–298.
26. Page, R. W., Jackson, M. J. & Krassay, A. A. (2000) *Aust. J. Earth Sci.* **47,** 431–459.
27. Srivastava, P. (2005) *Orig. Life Evol. Biosph.* **35,** 175–185.
28. Amard, B. & Bertrand-Sarfati, J. (1997) *Precambrian Res.* **81,** 197–221.
29. Sánchez-Baracaldo, P., Hayes, P. K. & Blank, C. E. (2005) *Geobiology* **3,** 145–165.
30. Gugger, M. F. & Hoffmann, L. (2004) *Int. J. Syst. Evol. Microbiol.* **54,** 349–357.
31. Knoll, A. H. & Golubic, S. (1992) in *Early Organic Evolution: Implications for Mineral and Energy Resources*, eds. Schidlowski, M., Golubic, S. & Kimberley, M. M. (Springer, Heidelberg), pp. 450–462.
32. Croft, W. N. & George, E. A. (1959) *Bull. Brit. Mus. Nat. Hist. Geol.* **3,** 339–353.
33. Gallon, J. R. (1992) *New Phytol.* **122,** 571–609.
34. Haystead, A., Robinson, R. & Stewart, W. D. P. (1970) *Arch. Mikrobiol.* **74,** 235–243.
35. Rippka, R. & Waterbury, J. B. (1977) *FEMS Microbiol. Lett.* **2,** 83–86.
36. Weare, N. M. & Benemann, J. R. (1974) *J. Bacteriol.* **119,** 258–265.
37. Maryan, P. S., Eady, R. R., Charplin, A. E. & Gallon, J. R. (1986) *J. Gen. Microbiol.* **132,** 789–796.
38. Ohki, K. & Fujita, Y. (1988) *Marine Biol.* **98,** 111–114.
39. Jensen, B. B. & Cox, R. P. (1983) *Arch. Microbiol.* **135,** 287–292.
40. Rippka, R. & Stanier, R. Y. (1978) *J. Gen. Microbiol.* **105,** 83–94.
41. Fay, P. (1992) *Microbiol. Rev.* **56,** 340–373.
42. Rye, R. & Holland, H. D. (1998) *Am. J. Sci.* **298,** 621–672.
43. Farquhar, J., Bao, H. & Thiemens, M. (2000) *Science* **289,** 756–759.
44. Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., III, Stein, H. J., Hannah, J. L., Coetzee, L. L. & Beukes, N. J. (2004) *Nature* **427,** 117–120.
45. Stal, L. J. (2000) in *The Ecology of Cyanobacteria: Their Diversity in Space and Time*, ed. Whitton, B. A. (Luver, Dordrecht, The Netherlands), pp. 61–121.
46. Canfield, D. E. & Teske, A. (1996) *Nature* **382,** 127–132.
47. Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., *et al.* (2001) *DNA Res.* **8,** 205–213, 227–253.
48. Meeks, J. C., Elhai, J., Thiel, T., Potts, M., Larimer, F., Lamerdin, J., Predki, P. & Atlas, R. (2001) *Photosynth. Res.* **70,** 85–106.
49. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidmann, J. G., Smith, J. A. & Sturuhl, K., eds. (1987) in *Current Protocols in Molecular Biology* (Wiley, New York), Sect. 2.4.
50. Urbach, E., Robertson, D. L. & Chisholm, S. W. (1992) *Nature* **355,** 267–270.
51. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998) *Trends Biochem. Sci.* **23,** 403–405.
52. Wilmotte, A., Van der Auwera, G. & De Wachter, R. (1993) *FEBS Lett.* **317,** 96–100.
53. Swofford, D. L. (2002) PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA), Version 4.

EVOLUTION