BIOINFORMATICS DISCOVERY NOTE Vol. 28 no. 5 2012, pages 603–606
doi:10.1093/bioinformatics/bts008

Phylogenetics

Advance Access publication January 11, 2012

The evolution of nitrogen fixation in cyanobacteria

Natasha Latysheva[1], Vivien L. Junker[1], William J. Palmer[1,2], Geoffrey A. Codd[3] and Daniel Barker[1,*]

[1]Centre for Evolution, Genes and Genomics, School of Biology, University of St Andrews, St Andrews, Fife, KY16 9TH, UK, [2]Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK and [3]Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee, DD1 4HN, UK

Associate Editor: David Posada

ABSTRACT

Motivation: Fixed nitrogen is an essential requirement for the biosynthesis of cellular nitrogenous compounds. Some cyanobacteria can fix nitrogen, contributing significantly to the nitrogen cycle, agriculture and biogeochemical history of Earth. The rate and position on the species phylogeny of gains and losses of this ability, as well as of the underlying *nif* genes, are controversial.

Results: We use probabilistic models of trait evolution to investigate the presence and absence of cyanobacterial nitrogen-fixing ability. We estimate rates of change on the species phylogeny, pinpoint probable changes and reconstruct the state and *nif* gene complement of the ancestor. Our results are consistent with a nitrogen-fixing cyanobacterial ancestor, repeated loss of nitrogen fixation and vertical descent, with little horizontal transfer of the genes involved.

Contact: db60@st-andrews.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 28, 2011; revised on December 30, 2011; accepted on January 4, 2012

1 INTRODUCTION

The cyanobacteria are a biochemically and morphologically diverse clade of Gram-negative bacteria with major environmental and economic roles and effects, including global primary productivity (Iturriaga and Mitchell, 1986; Paerl, 2000), potential uses in biofuel production (Hu *et al*., 2008) and in presenting water-borne health hazards via toxin production (Codd *et al*., 2005). Cyanobacteria are key biocatalysts in the N_2 cycle (Vitousek *et al*., 2002). The rice paddies of Asia, which feed over half of the world's population, depend upon cyanobacterial N_2 fixation (Irisarri *et al*., 2001). The cyanobacterial genus *Trichodesmium* is responsible for ~42% of the total global nitrogen fixation of 240 Tg N_2 y^{-1} (Berman-Frank *et al*., 2003).

Cyanobacteria may have existed for 2.7–2.9 billion years or more (Blank and Sánchez-Baracaldo, 2010; Noffke *et al*., 2008; Olson, 2006). A question of primary importance to our view of the biological environment of the early Earth, and to cyanobacterial evolution, is whether the last cyanobacterial common ancestor (LCCA) could fix N_2 (reduce N_2 to NH_3). Sánchez-Baracaldo *et al*. (2005) concluded that LCCA was not N_2-fixing. A study of 13 cyanobacterial genomes agreed with this conclusion (Shi and Falkowski, 2008). A study of a greater number of genomes, including basal N_2-fixers *Synechococcus* sp. JA-3-3Ab (Cyanobacteria bacterium Yellowstone A-Prime) and *Synechococcus* sp. JA-2-3B'a(2-13) (Cyanobacteria bacterium Yellowstone B-Prime), reconstructed the status of LCCA as ambiguous (Larsson *et al*., 2011). But due to use of parsimonious ancestral state reconstruction, the relative support for presence versus absence of N_2 fixation could not be quantified.

Where present in cyanobacteria, N_2 fixation is performed by proteins encoded by *nif* genes. The central enzyme is nitrogenase 1 (N_2ase), a molybdenum-dependent ATP-hydrolyzing complex of two metalloproteins: a dinitrogenase $\alpha_2\beta_2$ heterotetramer that contains the active site for the reduction of N_2 (its α and β subunits are coded by the *nifD* and *nifK* genes, respectively) and a dinitrogenase reductase γ_2 homodimer (coded by *nifH*) that transfers high-energy electrons to dinitrogenase. N_2ase is inactivated upon oxygen binding, suggesting an origin prior to the Great Oxygenation Event (Broda and Peschek, 1983) at the start of the Proterozoic eon between 2400 and 2200 million years ago. There are about 16 *nif* genes in cyanobacteria, of which 8 are regarded as core to the N_2 fixation pathway, or 9 including the crucial non-*nif* transcriptional regulator *patB* (Stucken *et al*., 2010). Larsson *et al*. (2011) included the *nif* genes *nifEHDKUB* and *patB* in the LCCA genome that they predicted using parsimony reconstructions, though they state that this reconstruction is ambiguous. *nif* genes have been highly evolutionarily conserved despite being present in a wide range of bacterial and archaeal taxa (Ruvkun and Ausubel, 1980; Young, 1992). A vanadium-dependent system, nitrogenase 2, is found in addition to nitrogenase 1 in the cyanobacterium *Anabaena variabilis*. Iron-dependent nitrogenase 3 is known in other groups (Lyons and Thiel, 1995).

On a phylogeny of 49 cyanobacterial genomes, we perform empirical Bayes ancestral state reconstructions for the ability to fix N_2 and for the presence/absence of 27 *nif* orthologous groups. Empirical Bayes ancestral state reconstruction quantifies ambiguity by giving a posterior probability for each state at each ancestral node (Pagel, 1999; Yang, 2006). We address the question of whether there has been any gain of N_2 fixation within the cyanobacterial

*To whom correspondence should be addressed.

clade, using a likelihood ratio test (LRT). To reduce the influence of phylogenetic uncertainty, we perform these analyses of trait evolution on a bootstrap sample of phylogenies (Felsenstein, 1988; c.f. Lutzoni *et al*., 2001). We also use stochastic mapping to estimate the number of gains and losses of the ability to fix N_2 within cyanobacteria and the position of these evolutionary events on the phylogeny (Heulsenbeck *et al*., 2003; Neilsen, 2002).

2 METHODS

Protein sets for all 49 cyanobacterial genomes in release 108 of integr8 (http://www.ebi.ac.uk/integr8) and a proteobacterial outgroup (cf. Swingley *et al*., 2008) were downloaded (Fig. 1; Supplementary Table S1). Orthologous groups predicted using OrthoMCL 2.0 (Li *et al*., 2003) with MCL 09-308 (van Dongen, 2000) on BLAST results with 'm S' masking and an *E*-value cutoff of 10^{-5} (Altschul *et al*., 1997). From trial inflation parameters 1.2, 1.6, 2, 2.8, 3.6, 4.4, 5.2, 6, 8, 11, 14, 17 and 20, the final value, 1.6, was chosen to maximize the number of orthologous groups in a single copy in each of the 65 species (cf. Swingley *et al*., 2008). A maximum likelihood (ML) species phylogeny was estimated, and 50 bootstrap replicates obtained, as described in Supplementary Material S1. For main analyses, trees were rooted using the outgroup, which was removed to leave a rooted tree of the 49 cyanobacterial species (Barker *et al*., 2007). For trait model selection and ancestral reconstructions, additional analyses were performed with the outgroup retained and the tree rooted at the mid-point of the branch between the ingroup and the outgroup (Barker and Pagel, 2005).

Data indicating presence or absence of the ability to fix N_2 in the species or strains were compiled from the literature (Supplementary Table S1). Parameters of continuous time Markov processes representing evolution of this binary trait were estimated for the phylogeny of species and these data using ML in the StochChar package (version 1.1) of Mesquite (version 2.74; W.P. Maddison and D.R. Maddison, http://mesquiteproject.org). Model selection (Asymm.2 or Mk1) was performed with an LRT. Empirical Bayes marginal ancestral state reconstructions were obtained in Mesquite using the 'global' approach of Pagel (1999) but with equilibrium state frequencies (Yang, 2006: 124). To test for N_2 fixation gain within the cyanobacterial clade, the Asymm.2 model with free parameters was compared with Asymm.2 with the rate of gain fixed close to zero, i.e. 1×10^{-10} (LRT). To estimate the number of gains and losses of N_2 fixation and their positions on the phylogeny, Mesquite's stochastic character mapping was performed 100 times on the ML tree. The distribution of each *nif* gene across extant genomes was obtained by text searches of sequence headers in the integr8 database and the assumption that all sequences in an orthologous group containing a *nif* gene were also products of that *nif* gene. Where a gene was represented by two or more orthologous groups, it was regarded as present where at least one of these groups was present, and absent where all the orthologous groups were absent. Exceptions were *nifS* and *nifU* for which, to avoid masking major differences in orthologous group distribution, ancestral state reconstructions were performed for the orthologous groups separately. Also, *nifD* was assumed present in *Nostoc punctiforme* (Larsson *et al*., 2011: 12). Ancestral state reconstructions of presence/absence of each *nif* gene on the ML tree were carried out as above.

3 RESULTS

Supplementary Table S2 shows the orthologous groups predicted by OrthoMCL (with species/strain abbreviations shown in Supplementary Table S1). The ML phylogeny based on a concatenation of the 147 universal, single-copy orthologous groups is given in Figure 1 and (with outgroup) in Supplementary Figure S1, and in Nexus format in Supplementary Material. *nif* genes and their

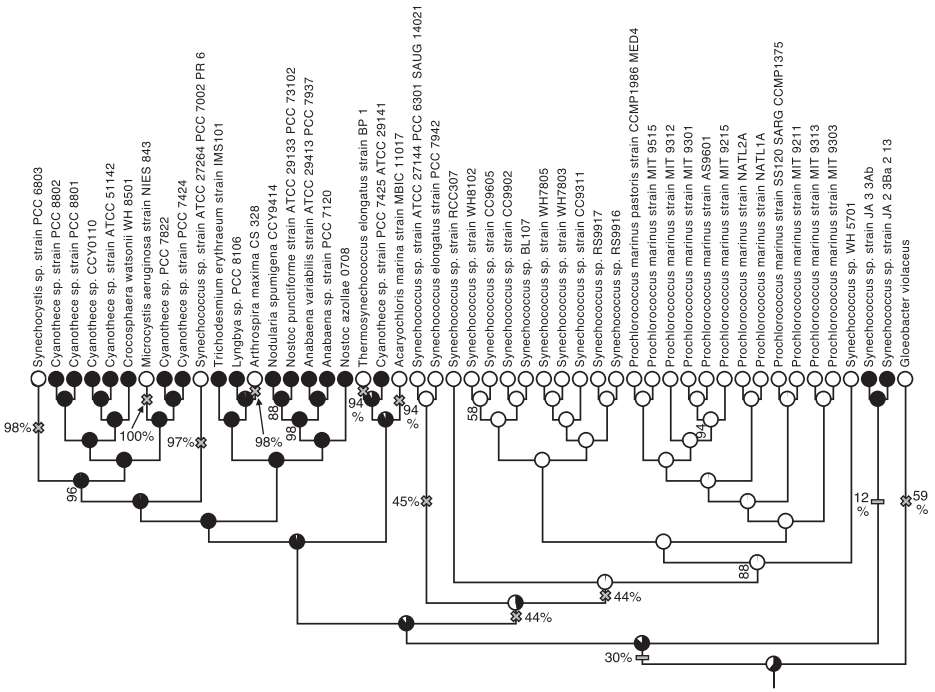

Fig. 1. Rooted phylogeny of cyanobacterial species. Bootstrap support is 100% for all nodes except where shown (vertical percentages). Pie charts represent the posterior probability of the presence (black) and absence (white) of nitrogen fixation according to the Asymm.2 model of trait evolution. Grey crosses on branches indicate the most probable locations of trait loss according to stochastic mapping, and grey bars indicate the most probable sites of gain through horizontal transfer (frequency in mappings $\geq$12%, excluding change and reversal on the same branch; frequencies shown as horizontal percentages).

orthologous groups and cross-species distribution patterns are given in Supplementary Table S3. Within cyanobacteria, *nifHDKTEXWZ* have a pattern of presence and absence identical to that of the N_2 fixation trait.

For the N_2 fixation trait, the Asymm.2 model was selected in preference to Mk1 whether the outgroup was excluded (ML tree, likelihood ratio statistic $2\Delta\ell = 2.70$, 1 d.f., $P = 0.0202$; bootstrap sample, $n = 48$, maximum $2\Delta\ell = 2.755$, $P = 0.0189$, minimum $2\Delta\ell = 2.63$, $P = 0.0218$) or included (ML tree, $2\Delta\ell = 6.47$, $P = 3.22 \times 10^{-4}$). On the ML tree, the accepted Asymm.2 model gave a posterior probability for N_2 fixation in LCCA of 0.61 (Fig. 1) and the rejected Mk1 model gave a posterior probability of 0.42. Including the outgroup changed the Asymm.2 posterior probability only slightly, to 0.64. Across the outgroup-excluding bootstrap sample, the mean posterior probability of N_2 fixation at LCCA was 0.60 ($n = 48$, SD = 0.050, range 0.363–0.636; Supplementary Fig. S2).

The ratio of rate of gain of N_2 fixation to rate of loss was 0.15 on the ML tree, and across the bootstrap sample had a mean of 0.15 ($n = 48$, SD = 0.00108, range 0.149–0.153). The rate of trait gain is significantly different from 1×10^{-10} (ML tree, $2\Delta\ell = 11.38$, 1 d.f., $P = 1.84 \times 10^{-6}$; bootstrap sample, maximum $2\Delta\ell = 20.59$, $P = 1.38 \times 10^{-10}$, minimum $2\Delta\ell = 20.52$, $P = 1.49 \times 10^{-10}$). The mean number of gain events across stochastic mappings on the ML tree was 0.64 ($n = 100$, SD = 0.75, range 0–3), of losses was 7.97 ($n = 100$, SD = 0.95, range 5–11) and the total number of evolutionary events (gains + losses) was 8.61 ($n = 100$, SD = 0.69, range 8–11), excluding changes occurring in both directions on the same branch (Supplementary Fig. S3). The locations of the trait gains most often seen in stochastic mappings suggest most probable positions for the horizontal gene transfer (HGT) of *nif* genes (Fig. 1; Supplementary Table S3 for the *nif* orthologous group distributions). The posterior probability of *nif* genes at LCCA, with and without the outgroup included and at the cyanobacterial–proteobacterial ancestor, is shown in Supplementary Figure S4. Stochastic mapping of *nif* genes largely suggests the same pattern of gain and loss as N_2 fixation, with a few deviations (see Section 4).

4 CONCLUSIONS

Our phylogeny of species is broadly congruent with others (e.g. Gupta, 2009; Larsson *et al*., 2011; Swingley *et al*., 2008; Tomitani *et al*., 2006). The reconstruction for presence versus absence of N_2 fixation in LCCA favours N_2 fixation, but only weakly (probability = 0.61). This contrasts with Sánchez-Baracaldo *et al*. (2005), Shi and Falkowski (2008), Boyd *et al*. (2011) and David and Alm (2011), but is in accord with Leigh (2000), Mehta and Baross (2006), Tomitani *et al.* (2006), Swingley *et al*. (2008) and Criscuolo and Gribaldo (2011). We expect this reconstruction to be refined as further taxa are included (Supplementary Table S4; Buschbom and Barker, 2006), but do not expect N_2 fixation in LCCA to be ruled out.

An N_2 fixing LCCA would imply that N_2 fixation and *nif* genes had arisen by ∼3 billion years ago. It is thought that fixed N was a limiting resource in the early Earth environment (Kasting and Siefert, 2001; Raven and Yin, 1998). The early Earth appears to have featured a mildly reducing atmosphere, in which fixed nitrogenous compounds would have been stable (Catling *et al*., 2001; Kasting and Siefert, 2001, 2002). A dramatic decrease in atmospheric CO_2 in the early Archaean (∼3.5 billion years ago) may have led to a fixed N crisis by limiting the amount of NO_x formation from N_2 and CO_2 (Navarro-González *et al*., 2001), creating evolutionary pressures for biological N_2 fixation at an early stage of prokaryotic evolution (Towe, 2002), despite the high energetic cost of this process (at least 16 ATP molecules per N_2 fixed; Scherer *et al*., 1988). Alternatively, the origin of N_2ases may pre-date the origin of biological N_2 fixation. The first N_2ases may have been N_2-using respiratory enzymes or cyanide detoxification centres (Fani *et al*., 2000; Postgate and Eady, 1988).

The means of the spread of N_2 fixation after its appearance has been another outstanding question. It has been unclear whether N_2 fixation has spread by HGT, or if vertical descent has had a larger impact. The rate of gain of N_2 fixation within the cyanobacterial clade was found to be significant, but much lower than the rate of loss. Consistent with this, stochastic character mapping most frequently suggested zero trait gains (51% of mappings; 35% showed one gain) and always suggested a higher number of trait losses than gains. In the 35% of instances in which the N_2 fixing trait was not ancestral and was gained only once, the most common position (86% frequency) of trait gain occurred very close to the root, just after the split between *Gloeobacter violaceus* and all remaining cyanobacteria (Fig. 1). This may be a site of HGT. There is very slight support for additional potential HGT events: 6% of mappings showed gain of N_2 fixation in the branch leading to *Cyanothece* sp. PCC 7425; 12% showed a gain in the ancestor of *Synechococcus* sp. JA-3-3ab and *Synechococcus* sp. JA-2-3B'a(2-13). Additionally, 9% of mappings included at least one instance of a gain of N_2 fixation, followed by loss on the same branch, manifested as absence in the extant descendant(s), which usually occurred amidst recent *Prochlorococcus* ancestors; 4% showed a single branch lose and subsequently gain N_2-fixing ability. Zehr *et al*. (1997) suggested that any HGT in cyanobacterial N_2 fixation would have happened early, and our results indeed suggest the most probable HGTs have occurred near the root of the phylogeny (Fig. 1), aside from the case of the potential HGT event in the terminal branch culminating in *Cyanothece* sp. PCC 7425.

Stochastic character mapping suggests that one loss of N_2 fixation occurred either just before (in 44% of mappings) or after (44%) the split between the *Synechococcus elongatus* PCC 7942/*Synechococcus sp.* ATCC 27144 clade and its sister group. Most other losses occurred recently in independent strains (Fig. 1). Our phylogeny supports both the monophyly of heterocyst-forming cyanobacteria (Henson *et al*., 2004) and their evolution within an N_2-fixing lineage, dated at 2450–2100 million years ago (Tomitani *et al*., 2006).

Ancestral state reconstructions on *nif* genes are broadly similar to the N_2 fixation results (Supplementary Fig. S4; Supplementary Material S2). Our reconstructions for LCCA (posterior probability > 0.5) agree with Larsson *et al*. (2011) in the presence of *nifEHDKU* and potentially *nifB*. Our results also predict *nifT*, *nifN*, *nifX*, *nifS*, *nifV*, *nifW*, *nifZ*, and have finer resolution of *nifS* and *nifU* orthologous groups. Raymond *et al*. (2004) suggest that the N_2ase gene family has evolved as a unit, with a highly conserved operon structure.

Our analysis has shown that loss of N_2 fixation has been more prevalent than HGT in cyanobacteria, despite evidence for HGT of N_2ase in other systems (Raymond *et al.*, 2004). Our results are consistent with a pre-cyanobacterial origin of N_2 fixation, perhaps

even in the Last Universal Common Ancestor, followed by repeated independent loss (Fani *et al.,* 2000; Normand *et al*., 1992; Young 1992). We conclude that trait loss has been the predominant agent in sculpting the modern distribution of N_2 fixation in cyanobacteria.

ACKNOWLEDGEMENTS

We thank Dannie Durand for helpful discussion, and three anonymous referees for their helpful comments.

Funding: Callan Memorial Scholarship (to N.L.).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.

Barker,D. and Pagel,M. (2005) Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput. Biol*., **1**, e3.

Barker,D. *et al.* (2007) Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. *Bioinformatics*, **23**, 14–20.

Berman-Frank,I. *et al*. (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol.,* **154**, 157–164.

Blank,C.E. and Sánchez-Baracaldo,P. (2010) Timing of morphological and ecological innovations in the cyanobacteria – a key to understanding the rise in atmospheric oxygen. *Geobiology*, **8**, 1–23.

Boyd,E.S. *et al.* (2011) A late methanogen origin for molybdenum-dependent nitrogenases. *Geobiology*, **9**, 221–232.

Broda,E and Peschek,G.A. (1983) Nitrogen fixation as evidence for the reducing nature of the early biosphere, *Biosystems*, **16**, 1–8.

Buschbom,J. and Barker,D. (2006) Evolutionary history of vegetative reproduction in *Porpidia s.l.* (lichen-forming Ascomycota). *Syst. Biol.*, **55**, 471–484.

Catling,D.C. *et al.* (2001) Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth. *Science,* **293**, 839–843.

Codd,G.A. *et al.* (2005) Cyanobacterial toxins: risk management for health protection. *Toxicol. Appl. Pharm.*, **203**, 264–272.

Criscuolo,A and Gribaldo,S. (2011) Large-scale phylogenomic analyses indicate a deep origin of primary plastids within Cyanobacteria. *Mol. Biol. Evol*., **28**, 3019–3032.

David,L.A. and Alm,E.J. (2011) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature*, **469**, 93–96.

Fani,R. *et al.* (2000) Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J. Mol. Evol*., **51**, 1–11.

Felsenstein,J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet*., **22**, 521–565.

Gupta,R.S. (2009) Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *Int. J. Syst. Evol. Microbiol.*, **59**, 2510–2526.

Henson,B.J. *et al.* (2004) Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on *nifD*. *Int. J. Syst. Evol. Micr.,* **54**, 493–497.

Hu,Q. *et al*. (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J*., **54**, 621–639.

Huelsenbeck,J.P. *et al*. (2003) Stochastic mapping of morphological characters. *Syst. Biol*., **52**, 131–158.

Irisarri,P. *et al.* (2001) Cyanobacteria in Uruguayan rice fields: diversity, nitrogen fixing ability and tolerance to herbicides and combined nitrogen. *J. Biotechnol.,* **91**, 95–103.

Iturriaga,R. and Mitchell,B.G. (1986) Chroococcoid cyanobacteria: a significant component in the food web dynamics of the open ocean. *Mar. Ecol. Prog. Ser*., **28**, 291–297.

Kasting,J.F. and Siefert,J.L. (2001) Biogeochemistry: the nitrogen fix. *Nature*, **412**, 26–27.

Kasting,J.F. and Siefert,J.L (2002) Life and the evolution of Earth's atmosphere. *Science*, **296**, 1066–1068.

Larsson,J. *et al.* (2011) Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol. Biol.*, **11**, 187.

Leigh,J.A. (2000) Nitrogen fixation in Methanogens: an Archaeal perspective. *Curr. Issues Mol. Biol*., **2**, 125–131.

Li,L. *et al.* (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.*, **13**, 2178–2189.

Lutzoni,F. *et al.* (2001) Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, **411**, 937–940.

Lyons,E.M. and Thiel,T. (1995). Characterization of *nifB*, *nifS*, and *nifU* genes in the cyanobacterium *Anabaena variabilis*: nifB is required for the vanadium-dependent nitrogenase. *J. Bacteriol*., **177**, 1570–1575.

Mehta,M.P. and Baross,J.A. (2006) Nitrogen fixation at 92°C by a hydrothermal vent archaeon. *Science*, **314**, 1783–1786.

Navarro-González,R. *et al.* (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature*, **412**, 61–64.

Nielsen,R. (2002) Mapping mutations on phylogenies. *Syst. Biol*., **51**, 729–739.

Noffke,N. *et al*. (2008) An actualistic perspective into Archean worlds – (cyano-) bacterially induced sedimentary structures in the siliciclastic Nhlazatse Section, 2.9 Ga Pongola Supergroup, South Africa. *Geobiology*, **6**, 5–20.

Normand,P. *et al.* (1992) Nucleotide sequence of *nifD* from *Frankia alni* strain Ar13: phylogenetic inferences. *Mol. Biol. Evol.,* **9**, 495–506.

Olson,J.M. (2006) Photosynthesis in the Archean era. *Photosyn. Res.*, **88**, 109–117.

Paerl,H.W. (2000) Marine plankton. In Whitton,B.A. and Potts,M. (eds) *The Ecology of Cyanobacteria*. Kluwer Academic Publishers, Netherlands, pp. 121–148.

Pagel,M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.*, **48**, 612–622.

Postgate,J.R. and Eady,R.R. (1988) The evolution of biological nitrogen fixation. In Bother,H. *et al*. (eds) *Nitrogen Fixation: Hundred Years After*. *Proceedings of the 7th International Congress on Nitrogen Fixation*. Gustav Fischer, Stuttgart, pp. 31–40.

Raven,J.A. and Yin,Z.H. (1998) The past, present and future of nitrogenous compounds in the atmosphere, and their interactions with plants. *New Phytol*., **139**, 205–219.

Raymond,J. *et al.* (2004) The natural history of nitrogen fixation. *Mol. Biol. Evol*., **21**, 541–554.

Ruvkun,G.B. and Ausubel,F.M. (1980) Interspecies homology of nitrogenase genes. *Proc. Natl Acad. Sci. USA* **77**, 191–195.

Sánchez-Baracaldo,P. *et al.* (2005) Morphological and habitat evolution in the cyanobacteria using a compartmentalization approach. *Geobiology*, **3**, 145–165.

Scherer,S. *et al.* (1988) Interaction of photosynthesis, respiration and nitrogen fixation in cyanobacteria. *Photosynth. Res*., **15**, 95–114.

Shi,T. and Falkowski,P.G. (2008) Genome evolution in cyanobacteria: The stable core and the variable shell. *Proc. Natl Acad. Sci. USA*, **105**, 2510–2515.

Stucken,K. *et al.* (2010) The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. *PLoS ONE*, **5**, e9235.

Swingley,W.D. *et al.* (2008) Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol. Biol. Evol*., **25**, 643–654.

Tomitani,A. *et al.* (2006) The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proc. Natl Acad. Sci. USA*, **103**, 5442–5447.

Towe,K.M. (2002) Evolution of nitrogen fixation. *Science*, **295**, 798–799.

van Dongen,S. (2000) A Cluster Algorithm for Graphs. *Technical Report INS-R0010*, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam.

Vitousek,P.M. *et al.* (2002) Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry,* **57–58**, 1–45.

Yang,Z. (2006) *Computational Molecular Evolution*. Oxford University Press, Oxford.

Young,J.P.W. (1992). Phylogenetic classification of nitrogen-fixing organisms. In Stacey,G. *et al.* (eds) *Biological Nitrogen Fixation*. Chapman & Hall, New York, pp. 43–86.

Zehr,J.P. *et al.* (1997) Phylogeny of cyanobacterial *nifH* genes: evolutionary implications and potential applications to natural assemblages. *Microbiology*, **143**, 1443–1450.