

Expanding phylogenetic trees and constraining the divergence of cyanobacteria

KELSEY R. MOORE^{1*}, CARA MAGNABOSCO², LILY MOMPÉR¹, GREGORY FOURNIER¹, TANJA BOSAK¹

¹ Massachusetts Institute of Technology, Cambridge, MA 02139, USA (*correspondence: krmoore@mit.edu)

² Simons Foundation for Center for Data Analysis, New York, NY 10010, USA

The pathway for oxygenic photosynthesis evolved only once in Earth history—within the ancestor lineage of cyanobacteria. These organisms are therefore key to understanding photosynthesis and the rise of oxygen on our planet. However, the timeline of the evolution and diversification of cyanobacteria remains poorly constrained. We use phylogenetic trees and molecular clocks to expand and better constrain this important timeline through the addition of novel genomes and the testing of fossil calibrations. Using large and small ribosomal subunit protein sequences from multiple cyanobacterial strains, plastid lineages and bacteria outgroups, we build phylogenetic trees to determine the genetic relationships between different groups. We then test four well-supported, morphologically distinct and characteristic cyanobacterial fossil calibration points to time-calibrate our phylogenetic trees and construct molecular clocks.

In addition to a core set of cyanobacterial genomes representative of the five major cyanobacterial groups and plastid lineages whose genomes are available on the *Genbank* database, we add newly sequenced cyanobacteria with unique morphologies, habitats and environmental tolerances. Some of these cyanobacteria have morphologies and lifestyles identical to those of the oldest known cyanobacterial fossils. The new sequences, obtained by the whole genome sequencing of environmental samples and pure cultures, expand the diversity of deeply branching cyanobacteria, filamentous forms found in pustular and tufted mats, *Chroococcidiopsis* and other endolithic species, and a novel organism isolated from the anoxic zone of Green Lake (New York State) that is tolerant to high concentrations of sulfide and inhibitors of oxygenic photosynthesis. We use both RAxML and Phylobayes to build species trees and molecular clocks for cyanobacteria. These analyses provide insight into the evolution and diversification of cyanobacteria, use more rigorous molecular clock dating techniques, increase the number of available fossil calibrations and test the influence of particular calibrations on the inferences about cyanobacterial evolution.