A p p e n d i x 5

CROWN GROUP OXYPHOTOBACTERIA POSTDATE THE RISE OF OXYGEN

Patrick M. Shih, James Hemp, Lewis M. Ward, Nicholas J. Matzke, and Woodward W. Fischer. "Crown group Oxyphotobacteria postdate the rise of oxygen." Geobiology 15.1 (2017): 19-29. DOI: 10.1111/gbi.12200

Abstract:

The rise of oxygen *ca.* 2.3 billion years ago (Ga) is the most distinct environmental transition in Earth history. This event was enabled by the evolution of oxygenic photosynthesis in the ancestors of Cyanobacteria. However, longstanding questions concern the evolutionary timing of this metabolism, with conflicting answers spanning more than one billion years. Recently, knowledge of the Cyanobacteria phylum has expanded with the discovery of non-photosynthetic members, including a closely related sister group termed Melainabacteria, with the known oxygenic phototrophs restricted to a clade recently designated Oxyphotobacteria. By integrating genomic data from the Melainabacteria, cross-calibrated Bayesian relaxed molecular clock analyses show that crown group Oxyphotobacteria evolved *ca.* 2.0 billion years ago (Ga), well after the rise of atmospheric dioxygen. We further estimate the divergence between Oxyphotobacteria and Melainabacteria *ca.* 2.5-2.6 Ga, which—if oxygenic photosynthesis is an evolutionary synapomorphy of the Oxyphotobacteria—marks an upper limit for the origin of oxygenic

photosynthesis. Together these results are consistent with the hypothesis that oxygenic photosynthesis evolved relatively close in time to the rise of oxygen.

Introduction:

Oxygenic photosynthesis was responsible for the most profound environmental shift in Earth history: the rise of oxygen. It was long recognized that this metabolism evolved in the Cyanobacteria phylum, and that this unique ability was a necessary precondition for the rise of oxygen at *ca*. 2.35 Ga [1]. However the evolutionary origins of Cyanobacteria remains uncertain, due to many conflicting lines of evidence ranging from microfossils [2], geochemical data [3-6], biomarkers [7], and geochemical models [8, 9]. These different proxies for Cyanobacteria provide divergence estimates that span more than one billion years of Earth history. This quandary leaves a major gap in knowledge regarding the O_2 cycle; consequently it remains unclear whether the rise of oxygen was directly related to the evolution of oxygenic photosynthesis [4, 10], or alternatively driven by a change in Earth's geophysical processes [8, 9].

The origins of Cyanobacteria has been intensely debated over the last half-century. Historically, hypotheses have been grounded in observations from the fossil and sedimentary rock records. However, the validity of some of the landmark studies, which have become core to many fundamental assumptions of the field, has been called into question, specifically concerning the biogenicity and taxonomic affinity of microfossils [11-13] and the recent reevaluation of Archean molecular fossils as sample contaminants

[14]. Because of the paucity of unequivocal evidence from early Precambrian sedimentary rocks, phylogenetics and comparative genomics provide an independent approach to constrain the evolutionary timing of Cyanobacteria, offering a useful point of comparison with the geological and fossil records [15].

A natural way to reconstruct and evaluate the evolution of oxygenic photosynthesis by Cyanobacteria from comparative biology is to identify their closest living relatives, which until recently remained a mystery. The notion that all Cyanobacteria evolved from a common photosynthetic ancestor hinged upon limited sampling of their extant diversity. Adding to the uncertainty, previous phylogenetic efforts to distinguish bacterial phylum-level relationships and identify the closest relative to Cyanobacteria have yielded conflicting results [16-19]. However, rapidly growing genomic and metagenomic datasets have substantially added to our understanding of the Cyanobacteria phylum. 16S rDNA surveys first described a substantial diversity of microbes related to oxygenic Cyanobacteria from aphotic environments [20]. Recently, genomes from a wide range of aphotic environments, such as gut microbiomes and groundwater samples, were analyzed that revealed the presence of a close sister group to oxygenic Cyanobacteria, named Melainabacteria, filling in an important gap in the diversity and evolution of the Cyanobacteria phylum [20-24]. The close evolutionary relationships of Melainabacteria and previously known Cyanobacteria fit with widely applied guidelines using 16S rDNA sequence identity for inclusion in microbial phyla [22]; this is a similar degree of sequence identity, for example, as that observed between *Escherichia coli* and *Pseudomonas*

aeruginosa—two common and well studied members of the class γ-Proteobacteria in the phylum Proteobacteria [25]. The placement of Melainabacteria as a close sister group to all known phototrophic Cyanobacteria is objective and reproducible on the basis a myriad of different gene and protein comparisons [21-23, 26, 27]. Additionally, the known oxygenic Cyanobacteria are much more closely related to the Melainabacteria than they are to any of the currently known members of other phyla capable of anoxygenic phototrophy, which are placed at far greater evolutionary distances, with numerous non-phototrophic groups interspersed between them [26]. These data illustrated that the Cyanobacteria phylum hosts a greater degree of physiological diversity than previously recognized – a condition similar to all the other known phototrophic phyla, which contain both phototrophic and non-phototrophic members [26].

With the discovery of a substantial diversity of close-living relatives to the oxygenic Cyanobacteria, it has been proposed to update the systematics of the Cyanobacteria phylum by adding a Melainabacteria class and relegating the oxygenic Cyanobacteria to the Oxyphotobacteria class [22]. For ease of discussion, we refer to these two sister clades using the following nomenclature: Oxyphotobacteria (class *Oxyphotobacteria* containing all known oxygenic Cyanobacteria) and Melainabacteria, wherein the Melainabacteria class to date includes four orders *Gastranaerophilales* [20, 21], *Obscuribacterales*, *Vampirovibrionales*, *Caenarcaniphilales* [22] (Figure 1). Interestingly, out of all the Melainabacteria genomes sequenced, none contain any genes associated with photosynthesis, and many lack genes necessary for aerobic or anaerobic respiration [21,

22]. Despite substantial environmental and genomic diversity, due to the current lack of observed photosynthetic basal lineages it is important to consider the hypothesis that oxygenic photosynthesis is a derived, and perhaps relatively recent, feature of the Cyanobacteria phylum (Figure 1). Thus, the newly identified Melainabacteria can add useful phylogenetic information in the sparsely covered regions closer to the base of the Cyanobacterial phylum.

To estimate the origin of Oxyphotobacteria, molecular clock studies have typically used either molecular fossils [28, 29] or cyanobacterial-like microfossils [30-32] as calibration constraints. It was once common to interpret microfossils as specific living lineages on the basis of morphological traits; however, phylogenetic analyses have revealed multiple independent acquisitions and widespread convergences of many of the classical cyanobacterial morphotypes, such as baeocystous and filamentous cells [33, 34], highlighting classic challenges in assigning microfossils to extant clades [35]. Moreover, 2-methylhopane molecular fossils, once thought to have been specific to the oxygenic Cyanobacteria, appear to have evolved in other phyla [36] and may not offer unique calibration constraints, concerns about syngeneity aside [14]. A number of studies have used the rise of oxygen as a calibration point for the minimal age of Oxyphotobacteria [28, 30, 31, 37]—this placement implicitly assumes that crown group Oxyphotobacteria were responsible for the rise of oxygen. However, it is equally plausible that extinct lineages existing before the most recent common ancestor of Oxyphotobacteria (*i.e.*, stem lineages) sourced the O_2 fluxes connected with the rise of oxygen. In order to test these assumptions,

it is useful to relax these constraints. Instead, valuable evolutionary insights into the origins of Oxyphotobacteria comes from endosymbiosis, which lends constraints from the fossil records from plants and algae – characteristics that most other bacterial phyla do not share. Additionally, by incorporating new molecular data from Melainabacteria taxa molecular clock analyses can now obtain a tighter estimate of the divergence time of Oxyphotobacteria.

Here, we revisit this evolutionary problem by performing cross-calibrated Bayesian relaxed molecular clock analyses with increased taxonomic sampling encompassing Melainabacteria, Oxyphotobacteria, plastids, and mitochondria. Distinct from previous cross-calibration efforts, we expand upon the technique by using a concatenated dataset composed of slowly evolving proteins and genes found in both plastids and mitochondria, which permit calibrations to be used multiple times across a phylogenetic tree [38].

Methods:

Generation of Concatenated Dataset. Sequences from subunits of ATP synthase, the ribosomal large subunit, the ribosomal small subunit, elongation factor Tu, and 16S rDNA were gathered. The protein sequences gathered from ATP synthase machinery consisted of AtpA, AtpB, AtpE, AtpF, AtpH, and AtpI. The ribosomal protein subunits collected for this study were Rpl2, Rpl16, Rps3, and Rps12. All sequences and their corresponding accessions are listed in Table S1. Sequences were collected to maximize the coverage of the Oxyphotobacteria and plastid-bearing eukaryotes. Plant mitochondrial genomes were

used along with α-Proteobacteria to serve as both an appropriate outgroup and enable cross-calibration between the corresponding plant mitochondrial and plastid lineages. Alignments for each protein or nucleotide family were performed using the –maxiterate strategy in the alignment program MAFFT [39], and then concatenated to generate the final dataset. The dataset was partitioned into two parts: concatenated protein and 16S nucleotide sequences, respectively.

Age Calibrations. Dating calibration priors were primarily chosen to avoid biasing analyses with the introduction of controversial microfossil occurrences, instead relying on well-accepted divergence times of plant fossils estimated. The posterior of divergence times from the comprehensive molecular clock analysis of land plants by Smith et al. were used as priors in this study [40]. A summary of all constraints used is described in Table 1. Normal distributions of 217 ± 40, 327 ± 30, 432 ± 30, and 477 ± 70 Ma were used as divergence time calibration points for Angiospermae, Gymnospermatophyta, Tracheophyta, and land plants, respectively. Importantly, the use of land plant divergence events enabled the cross-calibration between divergence events happening simultaneously in lineages that contain both plastids and mitochondria. In BEAST runs incorporating the fossil *Bangiomorpha pubescens* were constrained using a normal distribution between 1174-1222 Ma, as done in Yoon et al [41]. Strict geochronologic constraints for the strata from which these fossils were found – the Hunting Formation on Somerset Island, Nunavut, Canada – are between *ca*. 1200 and *ca*. 900 Ma. We examined a constraint associated with the oldest estimate placed between 1174 and 1222 Ma. This was chosen in

part to provide a conservative (*i.e.*, oldest possible) estimate on the divergence of crown group Oxyphotobacteria [42, 43]. Because the timing of the evolution of oxygenic photosynthesis remains controversial based on geological observations, we used a uniform prior spanning between 2400 and 3000 Ma and tested this prior on the two different nodes that represent 1) the divergence between Melainabacteria and Oxyphotobacteria and 2) the radiation of crown group Oxyphotobacteria. The "Rise of Oxygen" constraint provides a minimum age of 2400 Ma [44], marking the oldest ages hypothesized for the rise of oxygen, and an upper age of 3000 Ma as suggested by various geological studies [5, 45]. A uniform distribution enables the Markov chain Monte Carlo (MCMC) search to agnostically explore with no initial bias before ultimately converging onto a date within the provided hard upper and lower bounds of 2400-3000 Ga. Finally, a uniform prior between 3800 and 2400 Ma was used as a calibration for the last common ancestor of all taxa used in this study. The large range was used to permit an unbiased and largely unrestricted constraint, which assumes that oxygenic photosynthesis must predate the rise of oxygen and that these taxa had not diverged prior to the Late Heavy Bombardment *ca*. 3800 Ma [46].

Molecular Clock Analysis. Dated phylogenies were estimated using the program BEAST [47] using the CIPRES Science Gateway v. 3.3 server [48]. Cross-calibrated analyses were coded into BEAST XML files as previously described [38]. For each macrofossil that provides an age calibration, the same date is applied to every node in the molecular phylogeny that corresponds to the speciation event. Thus, if paralogs of a gene are found in

the nucleus, chloroplast, and mitochondrion, one calibration fossil supplies calibration distributions for several nodes in the gene phylogeny. The CpREV model was chosen as the best-fitting amino acid substitution model based on ProtTest analysis [49]. The CpREV model was used for the concatenated amino acid sequences and the 16S rDNA sequences used the GTR+G model in accordance to the 16S rDNA molecular clock study using BEAST by Schirrmeister et al [30]. A variety of BEAST runs were calculated using different combinations of date calibration priors, described below. For all runs, we ran five MCMC chains for the maximum limit allowed on the CIPRES Science Gateway, sampling every 10,000th generation. On average, more than 20 million generations were collected from each MCMC chain, and the initial 5 million generations were discarded as burnin (for exact numbers of post-burnin generations reported in Table 2). Maximum clade credibility (MCC) trees were generated using TreeAnnotator v1.7.5. All BEAST runs are summarized in Table 2. In general, the resulting chronograms of runs T64, T65, T68, and T69 were consistent with the overall phylogenomic analyses revealing the sister relationship between Oxyphotobacteria and Melainabacteria described by both Soo et al [22] and Di Rienzi et al [21]. In contrast, the topology of T72 and T73 placed Melainabacteria sister to α-Proteobacteria.

Regression Analysis of Node Age Uncertainty. The 95% Highest Posteriori Density (HPD) widths of node dates from the MCC trees generated by each analysis were extracted and plotted in R. For each pair of analyses, the list of HPD widths was reduced to the set of nodes common between the trees (the trees had mostly common nodes), and

the HPD widths were regressed on each other and plotted. A 1:1 line was also plotted for visual comparison. To statistically test for deviations from a 1:1 line (indicating that one analysis had a higher or lower amount of uncertainty in node dates than the other), the expected 1:1 relationship was subtracted and the regression was repeated; the resulting p-value on slope (printed on each plot) thus tests for significant deviation from the expected 1:1 relationship.

Expanded Phylogeny of Melainabacteria and Oxyphotobacteria. 16S rDNA sequences from members of the Cyanobacteria phylum (including the Oxyphotobacteria, Melainabacteria, and ML635J-21 classes) greater than 1250 base pairs in length were acquired from the Silva database [50]. Alignments were performed using the Infernal aligner [51] implemented by the Ribosomal Database Project [52]. Multiple phyla (Chloroflexi, OP11, Armatimonodetes, BD1-5, OD1, SHA-109, SR1, TM7, WWE3, and WS6) were used as out groups for the 16s rDNA phylogeny. Phylogenetic analyses were performed at the CIPRES Science Gateway using RAxML under the GTR model and default parameters for nucleotides. Trees were imaged using Interactive Tree of Life software [53].

Phylogenetic Analysis of O_2 reductases. Sequences for the A and C-family heme-copper O_2 reductases were acquired from public genomic and metagenomic databases and aligned using MAFFT. The resulting alignments were manually curated using structural and biochemical data. Phylogenetic analyses were performed at the CIPRES Science Gateway

using RAxML [54] under the LG model. The A and C-family O_2 reductases were each mid-point rooted due to uncertainty in the true position of the roots for these related protein families. Trees were imaged using Interactive Tree of Life software [53].

Results and Discussion:

Crown group Oxyphotobacteria postdate the rise of oxygen.

We have previously shown that cross-calibrated of Bayesian relaxed molecular clock analyses improve estimates of timing of evolutionary events, and can be useful to date Precambrian divergences [38]. Here, we built upon this technique by constructing and analyzing a dataset of slowly evolving proteins and genes found in both plastids and mitochondria, which permit fossil calibrations to be used multiple times across a phylogenetic tree. This included proteins from ATP synthase, the ribosomal large subunit, the ribosomal small subunit, and elongation factor Tu, as well as 16S rDNA sequences from Oxyphotobacteria, Melainabacteria, plastids, α-Proteobacteria, and plant mitochondria (Table S1-S3).

By expanding the cross-calibrated analysis to include Melainabacteria, we estimate reveal the divergence of crown group Oxyphotobacteria at *ca*. 2.0 Ga, postdating the rise of oxygen by at least 300 million years (Table 3, Table S2). A useful way to test the robustness of this estimate is by examining the results under different combinations of age calibration constraints. Permutations of the two deepest Proterozoic calibrations—the "Rise of Oxygen" constraint and the *Bangiomorpha* fossil constraint—all show a

Paleoproterozoic radiation of crown group Oxyphotobacteria, ranging in time between 2.071-1.741 Ga (Table 3, Figure 2, Figure S1-S4). This result contrasts with previous molecular clock studies that explicitly constrained the most recent common ancestor of all Oxyphotobacteria to the rise of oxygen, under the assumption that the radiation of the crown group was responsible for the rise of oxygen [28, 30, 31, 37]. However, this finding is in agreement with the fossil record, where the first widely accepted fossil Cyanobacteria (*Eoentophysalis sp*.) occur in shallow marine carbonate strata at *ca*. 1.9 Ga [55]. In contrast, when the "Rise of Oxygen" constraint was placed on the radiation of crown Oxyphotobacteria, the Melainabacteria clade instead became sister to the α-Proteobacteria/mitochondria clade – a result in conflict with all phylogenetic studies to date [20-22], which suggested that this calibration set poorly fit the data (Figure S5 & S6). Moreover, comparisons of node age uncertainty between BEAST analyses supported placement of the "Rise of Oxygen" constraint on the divergence of Melainabacteria and Oxyphotobacteria over the radiation of crown Oxyphototbacteria (Figure 3).

Dating the divergence between Oxyphotobacteria and Melainabacteria.

Next, we examined the hypothesis that Oxyphotobacteria and Melainabacteria diverged prior to the rise of oxygen. With a sister group relationship from current genomic data there is a degree of ambiguity regarding metabolic reconstruction of the last common ancestor of these clades. Two reasonable evolutionary scenarios exist (Figure 4). 1) The common ancestor to both Oxyphotobacteria and Melainabacteria was aphototrophic and oxygenic photosynthesis evolved after their divergence in the Oxyphotobacteria lineage. If correct,

the divergence of these clades would mark an upper age limit on the evolution of oxygenic photosynthesis. 2) Oxygenic photosynthesis evolved prior to the divergence of Oxyphotobacteria and Melainabacteria and was subsequently lost from the Melainabacteria clade. The BEAST results can, in principle, help select between these possibilities. For example, if we observe that they diverged after the rise of oxygen, that result would support a loss of photosynthesis from the Melainabacteria (scenario 2).

Our cross-calibrated analyses estimate the divergence of Oxyphotobacteria and Melainabacteria to have occurred *ca*. 2.5-2.6 Ga (Table 3). This result is consistent with the hypothesis that stem group Oxyphotobacteria evolved oxygenic photosynthesis after their divergence from the Melainabacteria, relatively close in time to the rise of oxygen [10, 56, 57]. This is supported by the observation that the Oxyphotobacteria are nested within a diverse clade of organisms from aphotic environments (Figure 1). Furthermore, no characterized Melainabacteria genomes encode genes involved in photosynthesis, whereas oxygenic photosynthesis is conserved within all extant Oxyphotobacteria except for a few clear cases of losses in obligate symbionts [58, 59]. Additional support for scenario 1 is provided by the contrasting evolutionary histories of other aspects of high potential metabolism (e.g. aerobic respiration) between Oxyphotobacteria and Melainabacteria. As mentioned above, no Melainabacteria contain genes for photosynthesis, but all extant Oxyphotobacteria and some Melainabacteria are able to perform oxidative phosphorylation [22], leading to the question of when this ability was acquired within the Cyanobacteria phylum. The F_0F_1-ATP synthase is the only phylogenetically congruent protein shared

between the Oxyphotobacteria and Melainabacteria used for oxidative phosphorylation. Phylogenetic analyses of the other respiratory proteins described below illustrate that they were acquired after the divergence of the two clades, suggesting that aerobic respiration and high potential metabolism was acquired independently in the Oxyphotobacteria and Melainabacteria.

For aerobic respiration, Oxyphotobacteria use evolutionarily conserved A-family O_2 reductases; these proteins display a characteristically low affinity for O_2, and imply that Oxyphotobacteria acquired their A-family O_2 reductases after the evolution of oxygenic photosynthesis and oxygenation of their environment. In contrast, some Melainabacteria in the *Obscuribacterales*, *Vampirovibrionales*, and *Caenarcaniphilales* orders are capable of aerobic respiration using C-family O_2 reductases. Notably, the C-family O_2 reductases are only very distantly related to the A-family O_2 reductases from Oxyphotobacteria [60], and occur within an operon with genes for Complex III. This suggests that the Melainabacteria capable of aerobic respiration acquired this ability by lateral gene transfer after the evolution of oxygenic phototrophy (Figure S7 & S8). Thus aerobic respiration was acquired independently at least twice within this phylum (once each in the Oxyphotobacteria and Melainabacteria). Oxyphotobacteria and Melainabacteria also utilize two substantially different Complex IIIs for aerobic respiration. The Oxyphotobacteria use a b_6f complex that functions for both phototrophy and aerobic respiration. The b_6f complex has a split cytochrome *b*, an additional cytochrome c_i near the Q_i plastoquinol binding site, and a novel cytochrome *f*. The aerobic Melainabacteria instead have a Complex III with a

full-length cytochrome *b* and no cytochrome *c*-containing subunit [22]. Thus, current genomic data supports a scenario in which aerobic respiration evolved after the origin of oxygenic photosynthesis in this phylum. Again this is more consistent with scenario 1: Oxygenic photosynthesis evolved after the divergence of Oxyphotobacteria from Melainabacteria, with the independent acquisition of aerobic respiration within these groups after the rise of oxygen. Although several lines of evidence favor scenario 1 over scenario 2, a formal test of the hypothesis described in scenario 2 will require genomic data from yet more basal members of this bacterial phylum (e.g. members of ML635J-21).

Compared with geological data, a Neoarchean 2.5-2.6 Ga divergence result for Oxyphotobacteria from their closest living relatives is also consistent with observations of a transitional photosystem using Mn prior to water-splitting at 2.415 Ga [61], as well as a number of studies suggesting small photosynthetic fluxes of O_2 in Neoarchean environments prior to the rise of oxygen [62, 63]. On the other hand, this result is discordant with hypotheses that stretch Oxyphotobacteria deep into Mesoarchean and older intervals [3, 5, 45], and raises the possibility that if those geochemical data are interpreted correctly, they may reflect non-cyanobacterial sources of O_2 or other oxidants [26, 64, 65].

Evolution of photosynthesis in Oxyphotobacteria.

Oxygenic photosynthesis would eventually become the core engine of the carbon cycle, but this need not have occurred synchronously with the evolution of water-splitting. All analyses show several hundred million years between the divergence of Melainabacteria

and Oxyphotobacteria, and the radiation of crown group Oxyphotobacteria (Figure 2). This length of time is important because a large number of evolutionary characters (including a number of multi-subunit complexes) are associated with extant oxygenic photosynthetic organisms (and missing from Melainabacteria), including photosystems I and II, the b_6f complex, RuBisCO and the Calvin cycle, and aerobic respiration – all are synapomorphies of Oxyphotobacteria [66]. It is unlikely that these traits all evolved at the same time, and comparisons of genomic data from Oxyphotobacteria, Melainabacteria, and additional basal clades can help ordinate the relative timing of the appearance of these characters.

The new dates for the divergence of Oxyphotobacteria from Melainabacteria, and the crown group radiation of Oxyphotobacteria, also provide a test of certain models for the evolution of phototrophy. As mentioned above, although all phototrophy is built around the same common molecular machinery (reaction centers, chlorins, and complex III or alternative complex III), it is also clear that the six phyla that are known to contain phototrophic members today are not closely related to one another; and consequently, lateral gene transfer was an important evolutionary vector for the sparse and scattered distribution of phototrophy observed today [26, 67]. In one class of hypotheses for origin of phototrophy—termed the fusion hypothesis—the Type I and Type II reaction centers evolved in different lineages (none of which must remain extant), and that ancestral Oxyphotobacteria then ultimately acquired both types of reaction centers by lateral gene transfer [26, 68]. If it is correct that oxygenic photosynthesis is derived within the Cyanobacteria phylum, then our estimates of 2.5-2.6 Ga for the divergence between

Oxyphotobacteria and Melainabacteria support the fusion hypothesis, because geological observations highlight that anoxygenic phototrophy was present at *ca*. 3.4 Ga [69].

Conclusions:

We presented the first molecular clock estimates for key divergences in the Cyanobacteria phylum that include data from close-living non-phototrophic relatives, but do not employ Archean lipid biomarkers, putative Cyanobacteria microfossils, or the rise of oxygen as calibration constraints. The results from cross-calibrated molecular analyses that include genomic data from the newly discovered Melainabacteria suggest that all known oxygenic Cyanobacteria did not appear until relatively late in Earth history [10, 26]. Estimates of the timing of divergence between Oxyphotobacteria and Melainabacteria are ~ 2.5-2.6 Ga. If it can be uniquely determined that oxygenic photosynthesis is a synapomorphy of the Oxyphotobacteria – and not simply loss or losses of both phototrophy and respiration from the Melainabacteria – this constraint would correspond to a maximum age of oxygenic photosynthesis (Figure 4, scenario 1). Current data best support this interpretation, but a formal test of this hypothesis awaits more genomic data from basal members of this phylum.

The results also consistently highlight that crown group oxygenic Cyanobacteria substantially postdate the rise of oxygen—which might instead be attributed to O_2 fluxes provided by organisms belonging to stem lineages. The results also consistently highlight that crown group Oxyphotobacteria (and thus all currently known Cyanobacteria capable of

oxygenic photosynthesis) substantially postdate the rise of oxygen—which might instead be attributed to O_2 fluxes provided by organisms belonging to stem lineages. Many previous studies have placed this crown group divergence on or prior to the rise of oxygen, but based on our results that show this assumption is not well supported; we suggest that it may be useful for future molecular clock studies to examine analyses with and without it. Further comparative genomic analyses reveal that aerobic respiration was acquired at least twice in the Cyanobacteria phylum after the evolution of oxygenic phototrophy.

Ultimately, our understanding of the major metabolic innovations that underpin Earth's biogeochemical cycles is limited to interpretations of the geochemical and fossil data in the context of frameworks built by knowledge of extant biology. This study highlights the opportunity to leverage the exponentially growing amount of genomic and metagenomic data in efforts to update and improve the quality of these frameworks.

References:

1. Luo, G., Ono, S., Beukes, N.J., Wang, D.T., Xie, S., and Summons, R.E. (2016). Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. Science Advances *2*.
2. Schopf, J.W., and Packer, B.M. (1987). Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. Science *237*, 70-73.

3. Rosing, M.T., and Frei, R. (2004). U-rich Archaean seafloor sediments from Greenland—indications of >3700 Ma oxygenic photosynthesis. Earth Planet. Sci. Lett. *217*, 237-244.

4. Kopp, R.E., Kirschvink, J.L., Hilburn, I.A., and Nash, C.Z. (2005). The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. Proc Natl Acad Sci *102*, 11131-11136.

5. Crowe, S.A., Dossing, L.N., Beukes, N.J., Bau, M., Kruger, S.J., Frei, R., and Canfield, D.E. (2013). Atmospheric oxygenation three billion years ago. Nature *501*, 535-538.

6. Johnson, J.E., Gerpheide, A., Lamb, M.P., and Fischer, W.W. (2014). O2 constraints from Paleoproterozoic detrital pyrite and uraninite. Geol Soc Am Bull.

7. Brocks, J.J., Logan, G.A., Buick, R., and Summons, R.E. (1999). Archean Molecular Fossils and the Early Rise of Eukaryotes. Science *285*, 1033-1036.

8. Holland, H.D. (2009). Why the atmosphere became oxygenated: A proposal. Geochim Cosmochim Acta *73*, 5241-5255.

9. Kump, L.R., and Barley, M.E. (2007). Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. Nature *448*, 1033-1036.

10. Ward, L., Kirschvink, J., and Fischer, W. (2015). Timescales of Oxygenation Following the Evolution of Oxygenic Photosynthesis. Orig Life Evol Biosph, 1-15.

11. Brasier, M.D., Green, O.R., Jephcoat, A.P., Kleppe, A.K., Van Kranendonk, M.J., Lindsay, J.F., Steele, A., and Grassineau, N.V. (2002). Questioning the evidence for Earth's oldest fossils. Nature *416*, 76-81.

12. Butterfield, N.J. (2015). Proterozoic photosynthesis – a critical review. Palaeontology *58*, 953-972.

13. Knoll, A.H. (2003). The geological consequences of evolution. Geobiology *1*, 3-14.

14. French, K.L., Hallmann, C., Hope, J.M., Schoon, P.L., Zumberge, J.A., Hoshino, Y., Peters, C.A., George, S.C., Love, G.D., Brocks, J.J., et al. (2015). Reappraisal of hydrocarbon biomarkers in Archean rocks. Proc Natl Acad Sci *112*, 5915-5920.

15. Shih, P.M. (2015). Photosynthesis and early Earth. Curr Biol *25*, R855-R859.

16. Zhaxybayeva, O., and Gogarten, J.P. (2002). Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses. BMC Genomics *3*, 4.

17. Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature *462*, 1056-1060.

18. Wolf, Y., Rogozin, I., Grishin, N., Tatusov, R., and Koonin, E. (2001). Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol Biol *1*, 8.

19. Daubin, V., Gouy, M., and Perrière, G. (2002). A Phylogenomic Approach to Bacterial Phylogeny: Evidence of a Core of Genes Sharing a Common History. Genome Res *12*, 1080-1090.

20. Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc Natl Acad Sci *102*, 11070-11075.

21. Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., et al. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife *2:e01102*.

22. Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., Steen, J.A., Parks, D.H., Tyson, G.W., and Hugenholtz, P. (2014). An Expanded Genomic Representation of the Phylum Cyanobacteria. Genome Biol Evol *6*, 1031-1045.

23. Johnson, J.E., Webb, S.M., Thomas, K., Ono, S., Kirschvink, J.L., and Fischer, W.W. (2013). Reply to Jones and Crowe: Correcting mistaken views of sedimentary geology, Mn-oxidation rates, and molecular clocks. Proc Natl Acad Sci *110*, E4119-E4120.

24. van der Lelie, D., Taghavi, S., McCorkle, S.M., Li, L.-L., Malfatti, S.A., Monteleone, D., Donohoe, B.S., Ding, S.-Y., Adney, W.S., Himmel, M.E., et al. (2012). The Metagenome of an Anaerobic Microbial Community Decomposing Poplar Wood Chips. PLoS ONE *7*, e36740.

25. Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998). Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. J Bacteriol *180*, 4765-4774.

26. Fischer WW, H.J., Johnson JE (2016). Evolution of oxygenic photosynthesis. Annu Rev Earth Planet Sci *44*.

27. Fischer WW, H.J., Valentine JS (2016). How did life survive Earth's Great Oxidation? Curr Opin Chem Biol *in review*.

28. Battistuzzi, F., Feijao, A., and Hedges, S.B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol Biol *4*, 44.

29. Battistuzzi, F.U., and Hedges, S.B. (2009). A Major Clade of Prokaryotes with Ancient Adaptations to Life on Land. Mol Biol Evol *26*, 335-343.

30. Schirrmeister, B.E., de Vos, J.M., Antonelli, A., and Bagheri, H.C. (2013). Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. Proc Natl Acad Sci.

31. Sánchez-Baracaldo, P., Ridgwell, A., and Raven, John A. (2014). A Neoproterozoic Transition in the Marine Nitrogen Cycle. Curr Biol *24*, 652-657.

32. Tomitani, A., Knoll, A.H., Cavanaugh, C.M., and Ohno, T. (2006). The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives. Proc Natl Acad Sci *103*, 5442-5447.

33. Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage

of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci *110*, 1053-1058.

34. Turner, S., Pryer, K.M., Miao, V.P.W., and Palmer, J.D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol *46*, 327-338.

35. Knoll, A.H., and Golubic, S. (1992). Proterozoic and Living Cyanobacteria. In Early Organic Evolution, M. Schidlowski, S. Golubic, M. Kimberley, D. McKirdy, Sr. and P. Trudinger, eds. (Springer Berlin Heidelberg), pp. 450-462.

36. Ricci, J.N., Michel, A.J., and Newman, D.K. (2015). Phylogenetic analysis of HpnP reveals the origin of 2-methylhopanoid production in Alphaproteobacteria. Geobiology, DOI: 10.1111/gbi.12129.

37. Falcon, L.I., Magallon, S., and Castillo, A. (2010). Dating the cyanobacterial ancestor of the chloroplast. ISME J *4*, 777-783.

38. Shih, P.M., and Matzke, N.J. (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. Proc Natl Acad Sci *110*, 12355-12360.

39. Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res *33*, 511-518.

40. Smith, S.A., Beaulieu, J.M., and Donoghue, M.J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc Natl Acad Sci.

41. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol *21*, 809-818.

42. Butterfield, N.J., Knoll, A.H., and Swett, K. (1988). Exceptional preservation of fossils in an Upper Proterozoic shale. Nature *334*, 424-427.

43. Butterfield, N.J., Knoll, A.H., and Swett, K. (1990). A bangiophyte red alga from the Proterozoic of arctic Canada. Science *250*, 104-107.

44. Hoffman, P.F. (2013). The Great Oxidation and a Siderian snowball Earth: MIF-S based correlation of Paleoproterozoic glacial epochs. Chem Geol *362*, 143-156.

45. Planavsky, N.J., Asael, D., Hofmann, A., Reinhard, C.T., Lalonde, S.V., Knudsen, A., Wang, X., Ossa Ossa, F., Pecoits, E., Smith, A.J.B., et al. (2014). Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. Nature Geosci *7*, 283-286.

46. Cohen, B.A., Swindle, T.D., and Kring, D.A. (2000). Support for the Lunar Cataclysm Hypothesis from Lunar Meteorite Impact Melt Ages. Science *290*, 1754-1756.

47. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol *7*, 214.

48. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Gateway Computing Environments Workshop (GCE), 2010. pp. 1-8.

49. Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics *21*, 2104-2105.

50. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res *41*, D590-D596.

51. Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335-1337.

52. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res *42*, D633-D642.

53. Letunic, I., and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res *39*, W475-W478.

54. Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics.

55. Hofmann, H.J. (1976). Precambrian Microflora, Belcher Islands, Canada: Significance and Systematics. J Paleontol *50*, 1040-1073.

56. Fischer, W.W. (2008). Biogeochemistry: Life before the rise of oxygen. Nature *455*, 1051-1052.

57. Shih, Patrick M. (2015). Cyanobacterial Evolution: Fresh Insight into Ancient Questions. Current Biology *25*, R192-R193.

58. Nakayama, T., Kamikawa, R., Tanifuji, G., Kashiyama, Y., Ohkouchi, N., Archibald, J.M., and Inagaki, Y. (2014). Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. Proc Natl Acad Sci *111*, 11407-11412.

59. Tripp, H.J., Bench, S.R., Turk, K.A., Foster, R.A., Desany, B.A., Niazi, F., Affourtit, J.P., and Zehr, J.P. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature *464*, 90-94.

60. Pereira, M.M., Santana, M., and Teixeira, M. (2001). A novel scenario for the evolution of haem–copper oxygen reductases. Biochim Biophys Acta *1505*, 185-208.

61. Johnson, J.E., Webb, S.M., Thomas, K., Ono, S., Kirschvink, J.L., and Fischer, W.W. (2013). Manganese-oxidizing photosynthesis before the rise of cyanobacteria. Proc Natl Acad Sci.

62. Anbar, A.D., Duan, Y., Lyons, T.W., Arnold, G.L., Kendall, B., Creaser, R.A., Kaufman, A.J., Gordon, G.W., Scott, C., Garvin, J., et al. (2007). A Whiff of Oxygen Before the Great Oxidation Event? Science *317*, 1903-1906.

63. Godfrey, L.V., and Falkowski, P.G. (2009). The cycling and redox state of nitrogen in the Archaean ocean. Nature Geosci *2*, 725-729.

64. Liang, M.-C., Hartman, H., Kopp, R.E., Kirschvink, J.L., and Yung, Y.L. (2006). Production of hydrogen peroxide in the atmosphere of a Snowball Earth and the origin of oxygenic photosynthesis. Proc Natl Acad Sci *103*, 18896-18899.

65. Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., et al. (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature *464*, 543-548.

66. Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., et al. (2006). The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci *103*, 13126-13131.

67. Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y., and Blankenship, R.E. (2002). Whole-Genome Analysis of Photosynthetic Prokaryotes. Science *298*, 1616-1620.

68. Hohmann-Marriott, M.F., and Blankenship, R.E. (2011). Evolution of Photosynthesis. Annu Rev Plant Biol *62*, 515-548.

69. Tice, M.M., and Lowe, D.R. (2004). Photosynthetic microbial mats in the 3,416-Myr-old ocean. Nature *431*, 549-552.

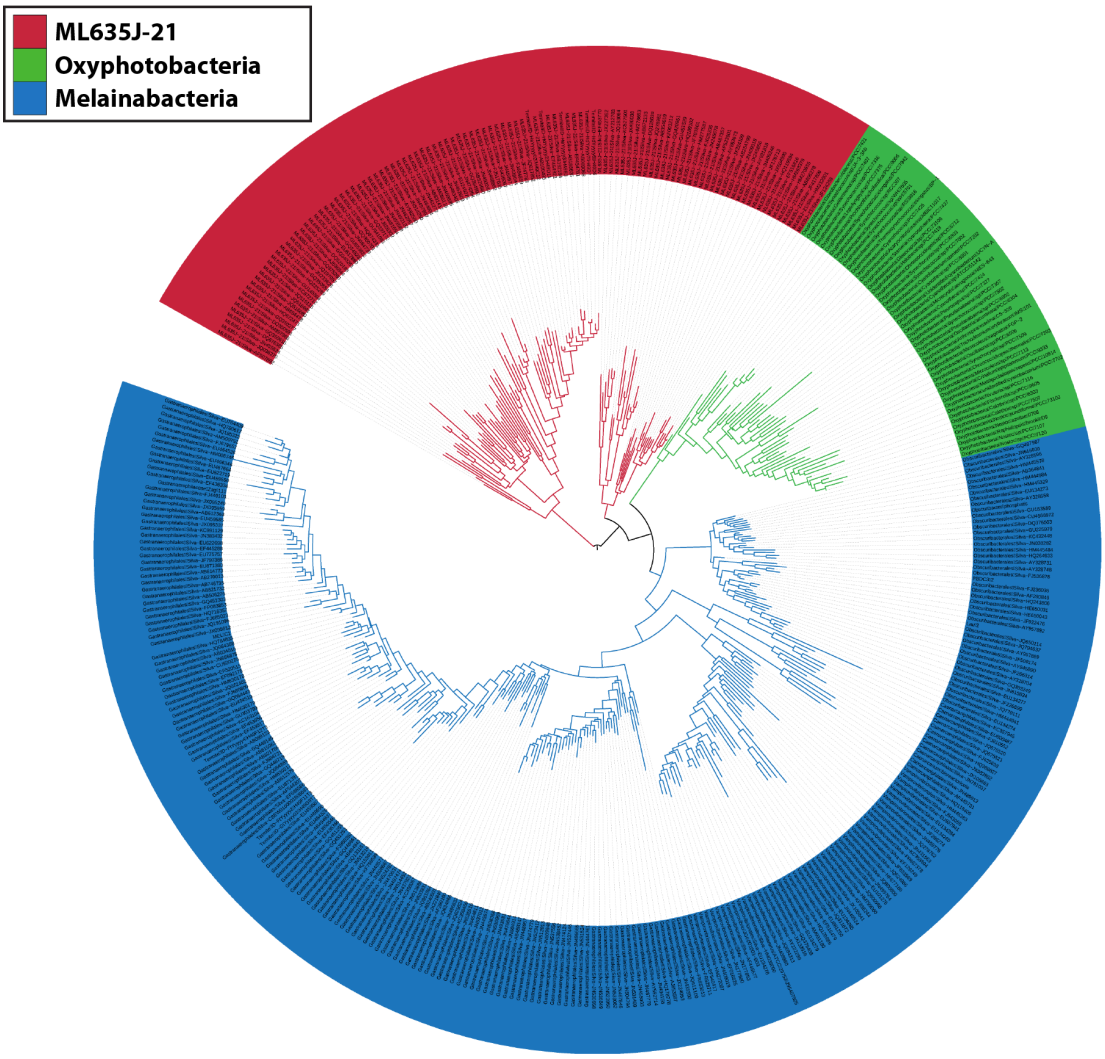

Figure 1. Complete labeled 16S rDNA tree of the Cyanobacterial phylum. Class Oxyphotobacteria in green, class Melainabacteria (orders Gastranaerophilales, Obscuribacterales, Vampirovibrionales, Caenarcaniphilales) in blue, and ML635J-21 clades shown in red. The majority of these Melainabacteria and ML635J-21 sequences were collected from aphotic and anaerobic environments, consistent with the physiologies reconstructed from existing Melainabacteria genomes. This topology warrants consideration of the hypothesis that oxygenic photosynthesis was derived from within the Cyanobacterial phylum.

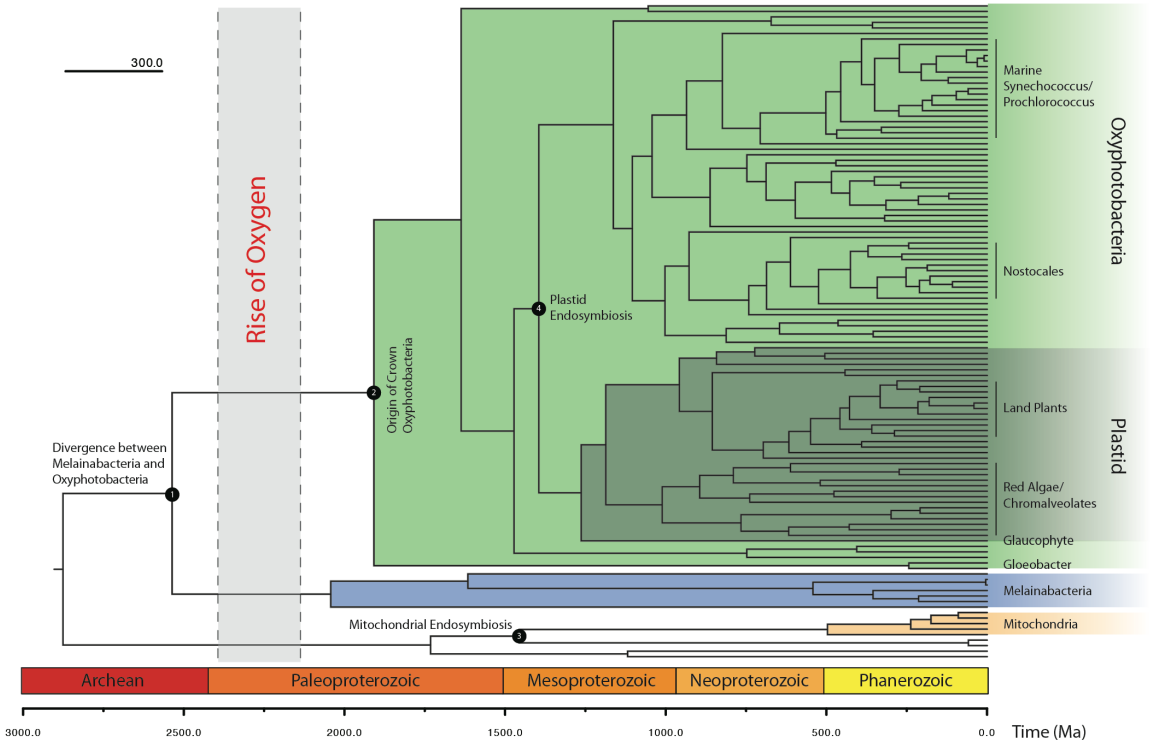

Figure 2. Divergence time estimates for the Cyanobacterial phylum. Dated phylogeny generated from cross-calibrated Bayesian analysis of a concatenated dataset (Run T65). The alignment is composed of conserved proteins found in plastids, mitochondria, and bacteria as well as their 16S rDNA sequences. All analyses illustrate that crown group Oxyphotobacteria postdate the rise of oxygen, which in turn reflects O_2 sourced from stem group lineages. Error bars on ages of nodes are shown and summarized in Figure S2.

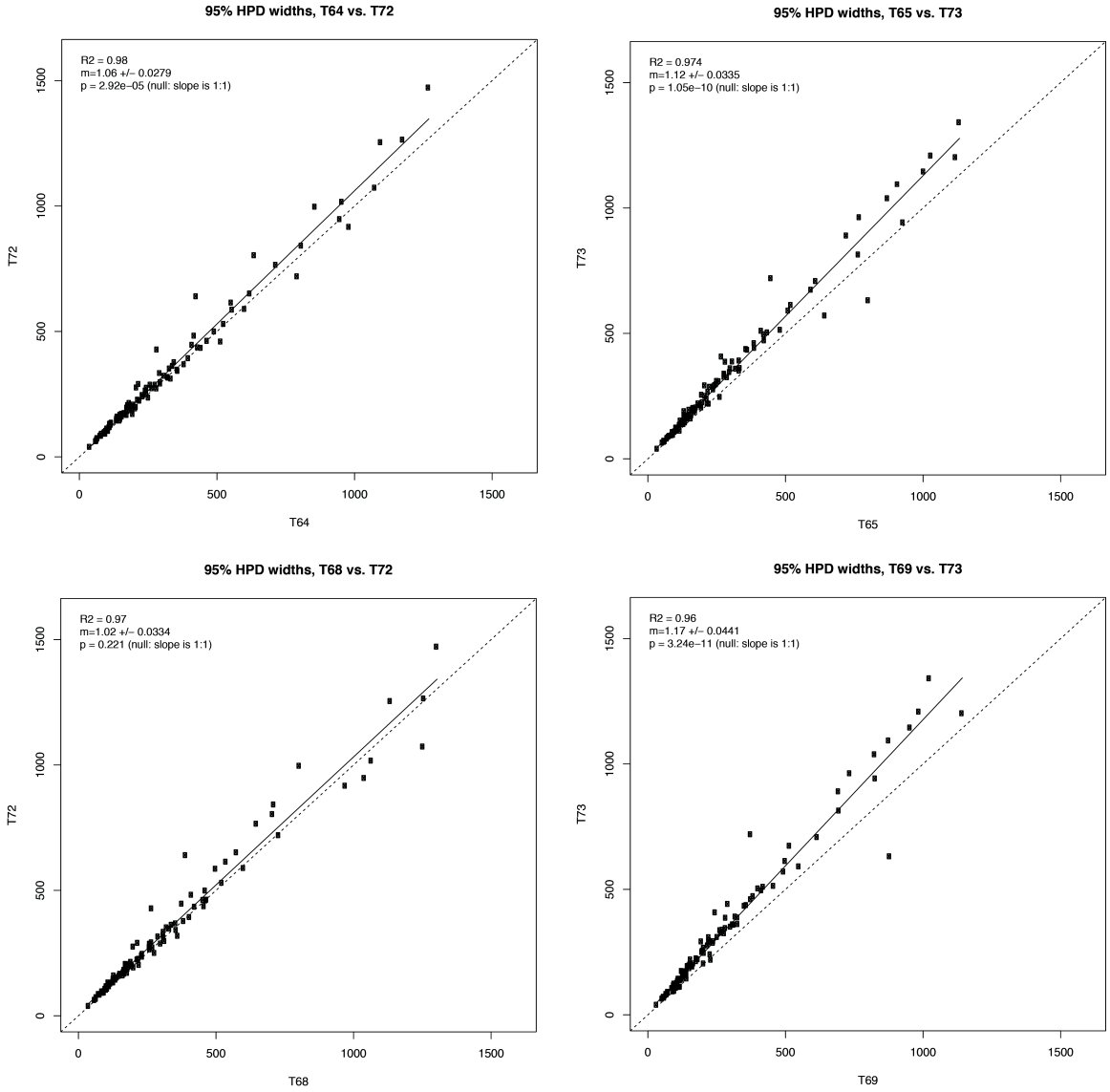

Figure 3. Higher observed age uncertainty when the "Rise of Oxygen" constraint is placed on the radiation of crown group Oxyphotobacteria. Each plot is a comparison of two BEAST analyses, where the width of the 95% HPD represents the amount of dating uncertainty. Each dot represents a corresponding node-date estimate in both trees. All plots are in comparison to one of the two analyses that place the "Rise of Oxygen" constraint on the radiation of crown group Oxyphotobacteria (T72 and T73). In all plots, the observed trend shows that there are higher levels of uncertainty for the T72 and T73 analyses,

suggesting that the placement of this constraint is in poorer agreement with the dataset than the combinations of constraints placed with analyses T64, T65, T68, and T69. *p*-values were calculated using a 1:1 slope as the null hypothesis, assuming that there is no difference in age uncertainty between both analyses.

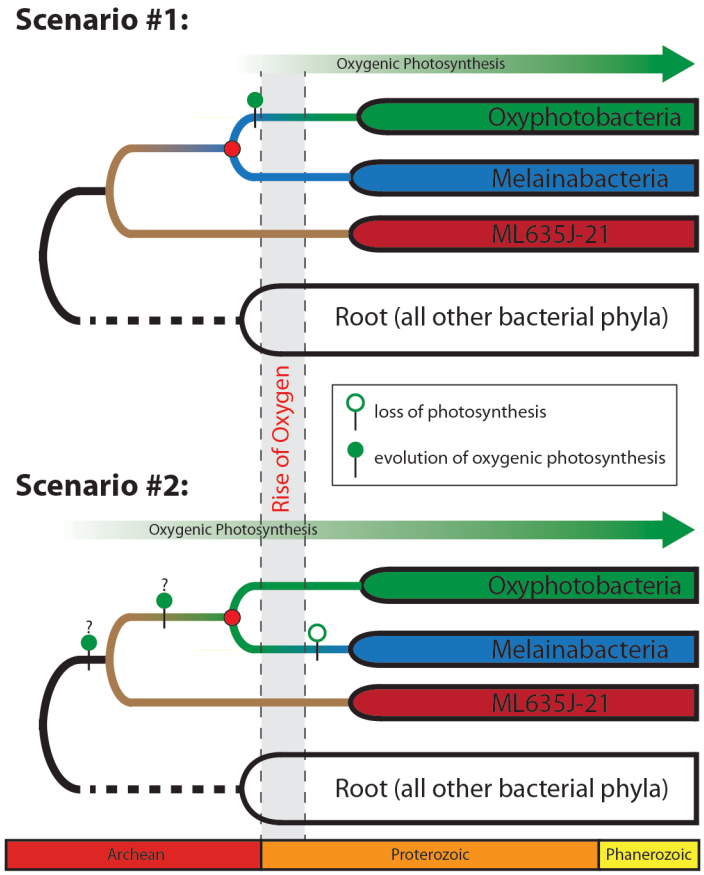

Figure 4. Schematic of two possible scenarios for the timing of oxygenic photosynthesis in relation to the divergence of Oxyphotobacteria and Melainabacteria. Scenario 1. Stem group Oxyphotobacteria diverge from their common ancestor with Melainabacteria (red circle) and evolve oxygenic photosynthesis later. This is the most parsimonious scenario given current data. In this scenario, the divergence of Oxyphotobacteria and Melainabacteria forms a maximum age constraint for oxygenic photosynthesis *ca*. 2.5-2.6 Ga., and is consistent with a wide range of geological and geochemical data for the rise of oxygen at *ca*. 2.35 Ga. **Scenario 2.** Oxygenic photosynthesis evolved prior to the divergence of the Oxyphotobacteria and

Melainabacteria, with subsequent loss(es) in the Melainabacteria. Genomic data from members of ML635J-21 clades will help resolve this ambiguity.

Table 1. Summary of calibration constraints used in this study.

Divergence event	Type of Distribution	Constraint in Mya (±std dev)
Monocotyledoneae	Normal	156 (±14)
Angiospermae	Normal	217 (±40)
Gymnospermatophyta	Normal	327 (±30)
Tracheophyta	Normal	432 (±30)
Land Plants	Normal	477 (±70)
Bangiomorpha	Uniform	1174-1222
"Rise of Oxygen"	Uniform	2400-3000
Last Common Ancestor	Uniform	2400-3800

Table 2. Summary of cross-calibrated BEAST runs generated in this study.

Name	Description	Generations
T64	"Rise of Oxygen" prior set on Mel/Oxyph divergence. Bangiomorpha prior used.	78.04 million post-burnin generations across 5 runs (each run had a burnin of 5 million), 103.04 million total
T65	"Rise of Oxygen" prior set on Mel/Oxyph divergence. Bangiomorpha prior omitted.	79.08 million post-burnin generations across 5 runs (each run had a burnin of 5 million), 104.08 million total
T68	"Rise of Oxygen" prior omitted. Bangiomorpha prior used.	99.12 million post-burnin generations across 5 runs (each run had a burnin of 5 million), 124.12 million total
T69	"Rise of Oxygen" prior omitted. Bangiomorpha prior omitted.	133.14 million post-burnin generations across 5 runs (each run had a burnin of 5 million), 158.14 million total
T72	"Rise of Oxygen" prior set on crown cyano divergence. Bangiomorpha prior used.	81.14 million post-burnin generations using a burnin of 5 million, 106.14 million total
T73	"Rise of Oxygen" prior set on crown cyano divergence. Bangiomorpha prior omitted.	81.78 million post-burnin generations using a burnin of 5 million, 106.78 million total

Table 3. Age estimates for key evolutionary divergences within the oxygenic Cyanobacteria using cross-calibrated methods. Despite substantially different combinations of deep time constraints ("Rise of Oxygen" and "*Bangiomorpha*" constraints), all BEAST runs robustly estimate a Neoarchean divergence between oxygenic Cyanobacteria and Oxyphotobacteria, as well as a Paleoproterozoic radiation of crown group Oxyphotobacteria. Each column identifies the age constraints that were used in different model runs in the study. Dates are in units of millions of years, and parentheses denote the limits of the 95% highest posterior density.

Run	**"Rise of Oxygen" constraint**	**Bangiomorpha constraint**	**Melainabacteria/ Oxyphotobacteria split**	**Crown Oxyphotobacteria**
T64	yes	yes	2630 (2400-2930)	2071 (1805-2393)
T65	yes	no	2536 (2400-2853)	1909 (1556-2254)
T68	no	yes	2541 (2090-3066)	2024 (1723-2374)
T69	no	no	2238 (1750-2790)	1741 (1361-2161)