Timing the Evolutionary Advent of Cyanobacteria and the Later Great Oxidation Event Using Gene Phylogenies of a Sunscreen

Ferran Garcia-Pichel,[a,b] Jonathan Lombard,[c] Tanya Soule,[d] Sean Dunaj,[d] Steven H. Wu,[e*] Martin F. Wojciechowski[a]

[a]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[b]Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, Arizona, USA
[c]Department of Cellular and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[d]Department of Biology, Purdue University, Fort Wayne, Indiana, USA
[e]Center for Personalized Diagnostic, Biodesign Institute, Arizona State University, Tempe, Arizona, USA

ABSTRACT The biosynthesis of the unique cyanobacterial (oxyphotobacterial) indole-phenolic UVA sunscreen, scytonemin, is coded for in a conserved operon that contains both core metabolic genes and accessory, aromatic amino acid biosynthesis genes dedicated to supplying scytonemin's precursors. Comparative genomics shows conservation of this operon in many, but not all, cyanobacterial lineages. Phylogenetic analyses of the operon's aromatic amino acid genes indicate that five of them were recruited into the operon after duplication events of their respective housekeeping cyanobacterial cognates. We combined the fossil record of cyanobacteria and relaxed molecular clock models to obtain multiple estimates of these duplication events, setting a minimum age for the evolutionary advent of scytonemin at 2.1 ± 0.3 billion years. The same analyses were used to estimate the advent of cyanobacteria as a group (and thus the appearance of oxygenic photosynthesis), at 3.6 ± 0.2 billion years before present. *Post hoc* interpretation of 16S rRNA-based Bayesian analyses was consistent with these estimates. Because of physiological constraints on the use of UVA sunscreens in general, and the biochemical constraints of scytonemin in particular, scytonemin's age must postdate the time when Earth's atmosphere turned oxic, known as the Great Oxidation Event (GOE). Indeed, our biological estimate is in agreement with independent geochemical estimates for the GOE. The difference between the estimated ages of oxygenic photosynthesis and the GOE indicates the long span (on the order of a billion years) of the era of "oxygen oases," when oxygen was available locally but not globally.

IMPORTANCE The advent of cyanobacteria, with their invention of oxygenic photosynthesis, and the Great Oxidation Event are arguably among the most important events in the evolutionary history of life on Earth. Oxygen is a significant toxicant to all life, but its accumulation in the atmosphere also enabled the successful development and proliferation of many aerobic organisms, especially metazoans. The currently favored dating of the Great Oxidation Event is based on the geochemical rock record. Similarly, the advent of cyanobacteria is also often drawn from the same estimates because in older rocks paleontological evidence is scarce or has been discredited. Efforts to obtain molecular evolutionary alternatives have offered widely divergent estimates. Our analyses provide a novel means to circumvent these limitations and allow us to estimate the large time gap between the two events.

KEYWORDS UV photobiology, cyanobacteria, deep evolution, secondary metabolism

Address correspondence to Ferran Garcia-Pichel, ferran@asu.edu.

* Present address: Steven H. Wu, BioConsortia Inc., Davis, California, USA.

Most paleoenvironmental models argue that the composition of the Earth's atmosphere was mildly reducing and anoxic (1–4) before the emergence of oxygenic photosynthesis in the cyanobacteria (5–7). For possibly hundreds of millions of years, the newly produced oxygen reacted in the environment with reduced compounds, so that molecular oxygen was only temporarily and locally available in "oxygen oases" (1, 8, 9) where cyanobacteria were present with sufficient biomass. Once the pool of reduced minerals had been oxidized, though, molecular oxygen started to accumulate in the atmosphere in what is known as the Great Oxidation Event (GOE)—an accumulation that eventually would result in the oxygen-rich atmosphere that we know today (3). Most evidence from the geochemical rock record places the GOE sometime between 2.45 and 2.22 billion years ago (Ga) (2, 4, 10, 11), even though recent estimates show evidence of oxygenated environments as far back as 3 Ga (8, 9).

And yet, the oxygenation of the atmosphere and oceans brought major changes to the ecological nature of the planet, given that oxygen was, and remains, a reactive toxicant to life (12). The GOE also brought about the formation of the stratospheric ozone layer (13), lessening the penetration of highly damaging solar UV radiation in the UVC and UVB ranges (i.e., below 320 nm in wavelength). In contrast, UVA is unaffected by stratospheric ozone absorption, and its harmful photosensitized reactions are mediated by O_2, so the UVA portion of the solar spectrum (320 to 400 nm) became biologically injurious only after O_2 started to accumulate in the environment (14, 15). In other words, the accumulation of oxygen in the atmosphere must have turned the UVA into a new ecological stress factor that required novel physiological adaptations. Among such protective adaptations against UVA photodamage, one must count the synthesis of the indole-phenol alkaloid scytonemin, which many cyanobacteria synthesize, excrete, and accumulate in their extracellular sheaths, where it serves as a sunscreen to protect themselves from UVA-mediated photodamage (16). Using sunscreens, however, is a costly photoprotective strategy for microbes, since it requires very large accumulations of pigment to make up for the very short optical path-lengths inherent to microbial cells (17). Sunscreens, by definition, are passive photoprotective mechanisms that require no metabolic activity beyond their initial synthesis. Therefore, microbial sunscreens like scytonemin become of particular fitness value when active photodamage repair mechanisms are impaired or suppressed due to slow metabolic activity or quiescence (18, 19). The distribution of scytonemin producers in nature is consistent with this notion, as they are restricted to habitats where only pulsed growth episodes are possible (terrestrial surfaces, the upper intertidal area, or alpine areas) and no scytonemin producers are found in oceanic or freshwater plankton (14). Scytonemin also undergoes facile reduction even under mildly reducing conditions, losing much of its UV-absorbing capacity and interfering with photosynthetically active radiation (16), no longer providing the photoprotection of the oxidized form. In summary, the evolution of scytonemin makes sense under conditions in which quiescent cyanobacteria were exposed to UVA and oxygen concurrently. In order to maintain oxygenation and photoprotection during periods of organismal quiescence, the oxygenation of the environment must be preserved even when the organisms are not active, so it cannot rely solely on oxygen released by photosynthesis. Yet, since oxygenic photosynthesis is halted while the organisms are inactive, photoprotection provided by scytonemin would only have had a fitness value after the GOE itself. Scytonemin's advent must thus postdate the GOE.

In this work, we applied divergence-time estimation using relaxed molecular clock models to analyze scytonemin's biosynthesis genes in an attempt to provide a geochemically independent, minimal age estimate of the GOE. We analyzed the presence, conservation, and synteny of the genes present in the scytonemin operon; studied their phylogenetic relationships; and determined their origin within the broader bacterial evolutionary history. Informative gene phylogenies were then dated using relaxed model analyses that had been calibrated using data from the fossil record. As a result, we obtained multiple, independent evaluations for the timing of relevant events in
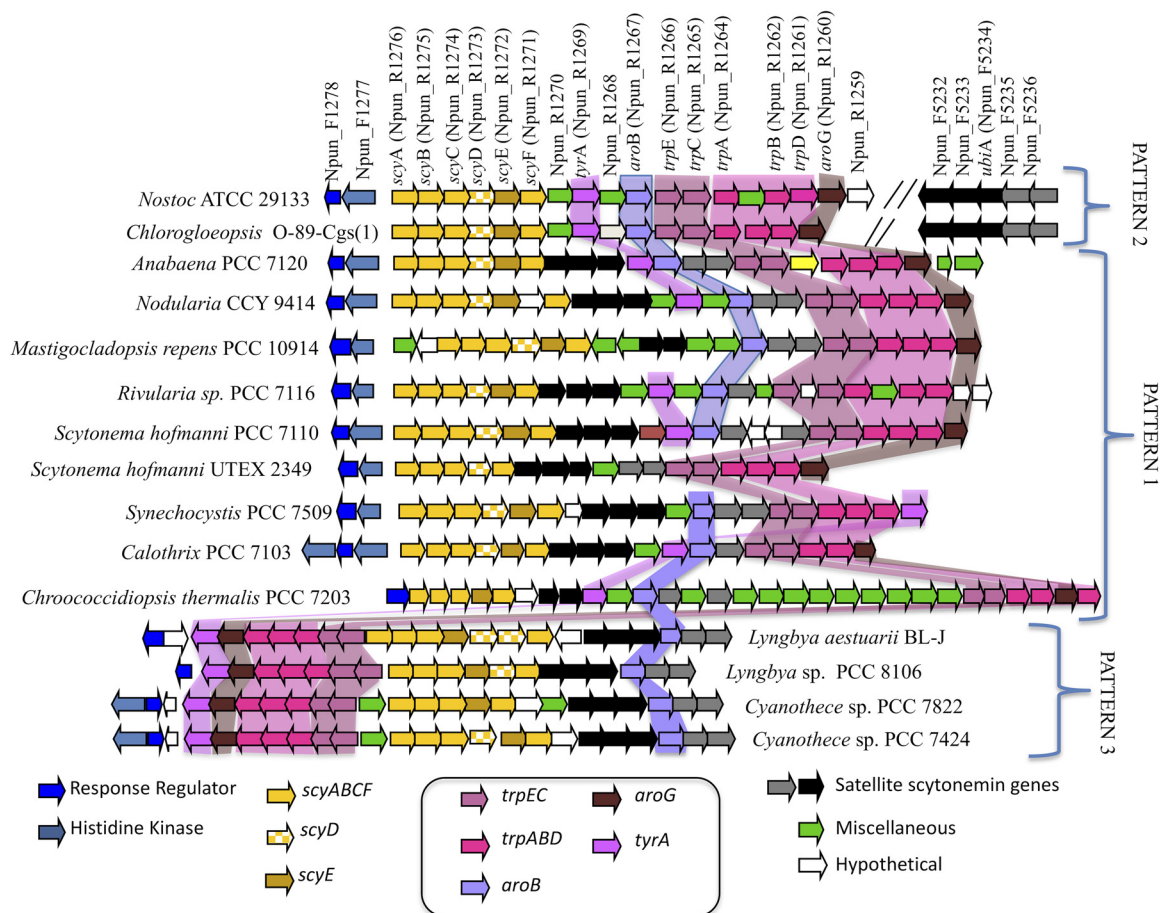
FIG 1 Genomic organization of the cyanobacterial scytonemin operon. Gene nomenclature is based on the annotation of *Nostoc punctiforme* ATCC 29133. Core biosynthetic genes are shown in yellow. Regulatory genes are shown in blue. Genes shown in black or gray involve a region dedicated to monomer export, and genes shown in various shades of purple/pink code for dedicated aromatic amino acid biosynthetic (AAAB) genes that supply biosynthetic precursors.

cyanobacterial evolution. Our results refine previous molecular estimates, are congruent with the accepted geochemical dating, and, importantly, provide a duration estimate of the era of oxygen oases.

RESULTS

Scytonemin operon architecture and gene synteny across cyanobacteria. The basic comparative structure of the scytonemin operon has been presented elsewhere (20, 21). Here (Fig. 1), we have updated our previous analyses using genomic information currently available in public databases. The core of the operon is made up of the biosynthetic genes *scyABCE*, which are required for biosynthesis of scytonemin from intermediaries of the aromatic amino acid pathway, and the biosynthetic adjuvant *scyF* (22). *scyD*, which is not required for scytonemin production and possesses the characteristics of a pseudogene (22), is missing in at least two strains used in our analyses. This core is otherwise conserved in sequence and shares synteny throughout the phylum. A distinct second region codes for enzymes supplying building blocks to the scytonemin biosynthetic pathway (denoted by various shades of purple in Fig. 1). We refer to those as aromatic amino acid biosynthesis (AAAB) genes. We can divide the AAAB genes into three functional classes: those homologous to tryptophan biosynthesis genes (*trpCE*, *trpABD*); one homolog of a phenylalanine biosynthesis gene, *tyrA*; and *aroBG*, which are homologous to key enzymes in the shikimic acid pathway that feeds precursors to the aromatic amino acid pathways. These AAAB genes represent additional copies of their respective housekeeping versions in all strains that have a

scytonemin operon. They are dedicated to scytonemin biosynthesis and coexpressed with the core genes when induced by UVA exposure (23). Scytonemin-operon AAAB genes are highly conserved and frequently share their synteny, particularly for the sequence *trpECABD-aroG*. The placement of the *tyrA* and *aroB* genes within the operon seems to be less constrained throughout the phylum. A third group of five highly conserved genes (black/gray in Fig. 1) occurs just downstream of the biosynthetic core in most strains but in a different locus in the genomes of *Nostoc punctiforme* and *Chlorogloeopsis* sp. (hence the name "satellite genes"). Aside from their distant placement in those two strains, conservation and internal shared synteny of these genes are quite conspicuous. They seem to be involved in the secretion of scytonemin monomers to the periplasm for final condensation (24). A recent contribution (25) has identified the group of satellite genes as the cyanobacterial version of a much more widespread and ancient operon, the "*ebo* operon" conserved among several bacterial phyla and eustigmatophyte plastid genomes. Finally, a set of conserved, homologous regulatory elements, typically forming a two-component regulator, occurs in close proximity to the scytonemin operon, and their presence is also required to allow the synthesis of scytonemin in response to UVA radiation (26).

Three organizational patterns exist among the cyanobacterial genomes with complete scytonemin operons (Fig. 1). The most frequent (pattern 1) consists of the region sequence regulatory/core/satellite/AAAB, with all but the regulatory elements transcribed in the same direction. Pattern 2 (present in *Nostoc punctiforme* and *Chlorogloeopsis* sp.) is characterized by a transposition of the "satellite" genes onto a distant locus (regulator/core/AAAB/—/satellite). Pattern 3 is found only in nonheterocystous filamentous forms and unicellular strains (*Lyngbya* sp., *Cyanothece* sp.) with the sequence regulators/AAAB/core/satellite. Here, transcription proceeds in opposite directions from the AAAB/core boundary. Since pattern 3 occurs in two lineages that, according to standard cyanobacterial phylogenies (27, 28), diverged earlier from the other cyanobacterial groups that contain these operons, the most parsimonious hypothesis is to assume that this order was the first one to emerge. Then, a rearrangement would have led to pattern 1 before the origin of heterocystous cyanobacteria, and a second, rather recent, rearrangement would have yielded pattern 2 within heterocystous *Nostoc* and close allies. The alternative sequence of pattern 1 first giving rise to pattern 3 and later to pattern 2, however, is also plausible.

In addition to these cyanobacterial genomes with full operons, a few strains also contain evidence of genomic remnants of the scytonemin operon. For example, there are recognizable stretches of the core genes in *Microchaete* strain PCC 7126 and stretches of the AAAB gene region (*trp/aro* clusters, often sharing synteny) in *Chamaesiphon minutus* PCC 6605, *Fischerella muscicola* PCC 73103, and *Chlorogloeopsis fritschii* PCC 6912. Also, the genomes of *Scytonema hoffmanii* PCC 7110 and *Rivularia* PCC 7116 contain a duplicated version of the AAAB gene region, besides the AAAB locus characteristic of the scytonemin operon.

Phylogeny of core biosynthetic genes. The single-gene phylogenies derived from analyses of the core biosynthetic genes *scyABCEF* show clear monophyletic relationships in all strains containing scytonemin operons (see Fig. S1 in the supplemental material), where *scyD* seems to be a nonfunctional paralogue of *scyE* (22). These genes have only low similarity to functionally unrelated genes from other cyanobacteria, bacteria, or fungi (depending on the gene; data not shown). Despite the small number of cyanobacterial genomes that contain the scytonemin operon, *scyABC* phylogenies are congruent with the generally accepted cyanobacterial phylogeny (27, 29) in that the heterocystous groups evolved secondarily, and they lend support to the ancient nature of genomic pattern 3 (see above). For *scyE* (20) and *scyF*, such a pattern is not resolved. In any event, in our data set no cyanobacterial version of any of these genes exists that is not integrated in the *scy* operon, which precludes the placing of their origins in a larger context, so we did not pursue those analyses further.
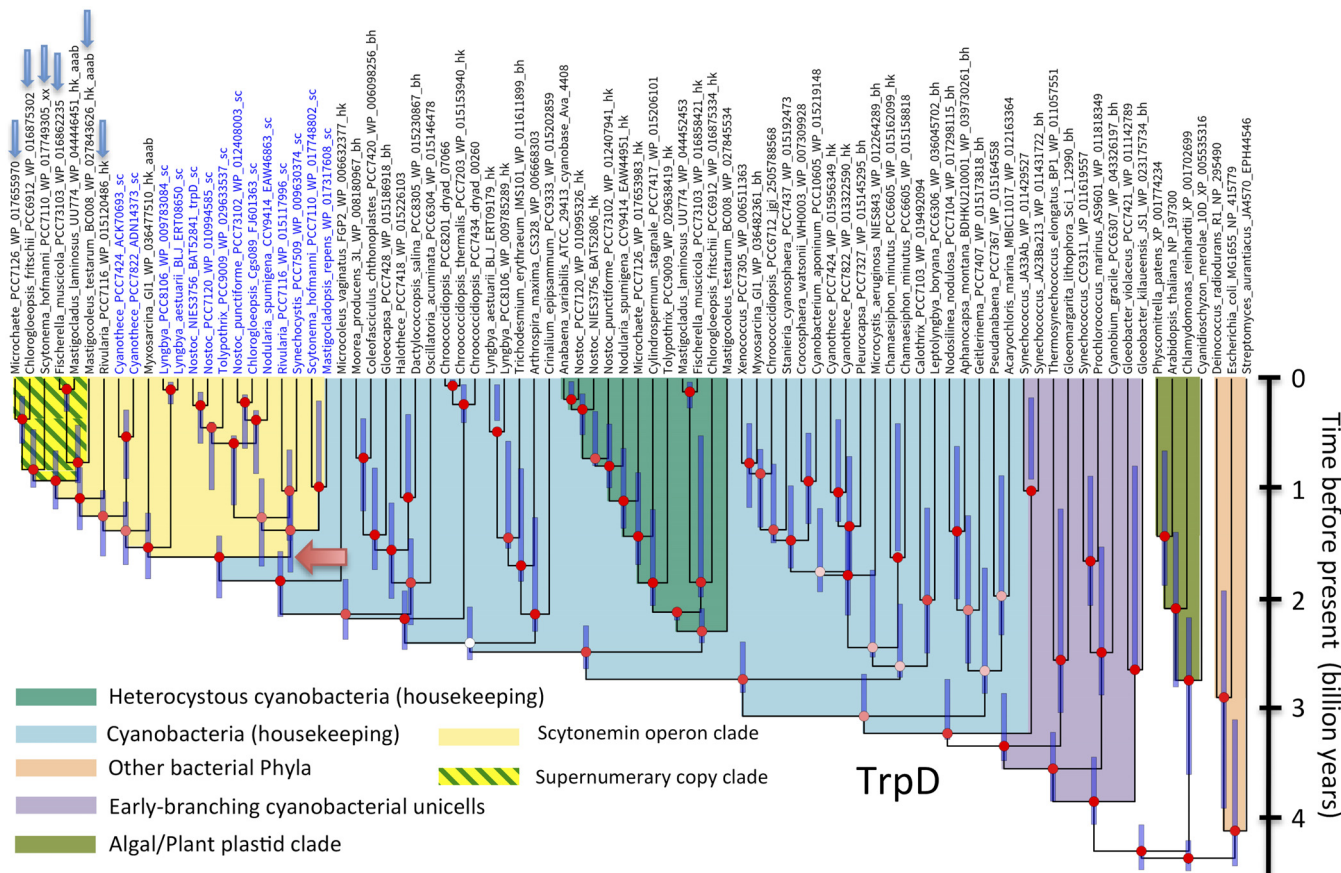
FIG 2 Phylogeny derived from BEAST analyses of TrpD amino acid sequences supports an origin of a scytonemin-associated clade in a single internal duplication of housekeeping genes, shown as exemplary of the two phylogenies (TrpB and TrpC) supporting such an origin. The corresponding phylogeny for TrpC is in Fig. S8. Entries in blue type correspond to homologs found within full scytonemin operons. Those marked by a blue arrow were found in remnant scytonemin operons or correspond to supernumerary homologs. Bayesian posterior probabilities (BPP) at the nodes are color coded as follows: red for BPP ≥ 0.8, pink for 0.8 ≥ BPP ≥ 0.5, and white for BPP ≤ 0.5. The red arrow marks the node corresponding to the most recent common ancestor of the oldest clade containing scytonemin-associated homologs.

AAAB gene product phylogenies. The fact that the scytonemin-dedicated AAAB genes typically represent additional versions of their respective housekeeping cognates (20) suggests that they may have evolved by duplication of these housekeeping genes in ancestral cyanobacteria. This was of much interest because, if confirmed, the timing of these duplications would represent a conservative estimate of the evolutionary origin of scytonemin. To test this hypothesis, we constructed phylogenetic trees from analyses of the amino acid sequences of each AAAB gene, including all bona fide orthologs present in *scy* operons available from cyanobacteria, some of those found in scytonemin operon remnants, and a phylogenetically representative set of housekeeping orthologs from the cyanobacteria and plant plastids, and included some from other bacterial phyla, to serve as outgroups. If this hypothesis were correct, one would expect to find the *scy* operon orthologs sharing an ancestor within their respective cyanobacterial lineages. Using translated amino acid sequences, we constructed preliminary neighbor-joining trees to explore general relationships (Fig. S1 to S4). These trees clearly showed that the different scytonemin-dedicated AAAB proteins tended to cluster together, although not always branching within the main housekeeping group of AAAB enzymes from cyanobacteria, but the trees did not show high support at relevant nodes. The same data sets were then used to obtain full Bayesian trees derived from BEAST analyses (Fig. 2 to 4; also Fig. S5 to S9). Our neighbor-joining and Bayesian results were largely congruent, and they have been summarized in Table 1. For five of the eight loci, the hypothesis of an internal duplication of the AAAB genes within their

TABLE 1 Summary of phylogenetic results for the evolution of aromatic amino acid biosynthesis loci contained in the scytonemin operon of cyanobacteria as determined from model-based Bayesian analyses of relevant locus-specific data sets

Locus	Whether orthologs form monophyletic group	Whether scytonemin clade(s)[a] originate(s) within cyanobacterial phylogeny	Housekeeping phylogeny consistent with cyanobacterial phylogeny[b]
trpB	Unclear	Yes	Yes
trpC	Yes	Yes	Yes
trpD	Yes	Yes	Yes
trpA	No, two clades	Yes	Yes
aroG	No, multiple clades	Yes	Yes
trpE	Yes	No, lateral transfer	Yes
aroB	Yes	No, lateral transfer	Yes[c]
tyrA	Yes	No, lateral transfer	Yes

[a]Including a clade of supernumerary and scytonemin operon remnant copies.
[b]See text for criteria.
[c]However, it contains two clades of heterocystous cyanobacteria.

cyanobacterial housekeeping counterparts was well supported (Bayesian posterior probability [BPP] = 1 in AroG, TrpA, TrpC, and TrpD, and BPP > 0.8 in TrpB). For the remaining three (TrpE, AroB, and TyrA) (Fig. 3 and Fig. S5 and S9), the Bayesian phylogenies strongly support an ancient recruitment of these genes into the operon
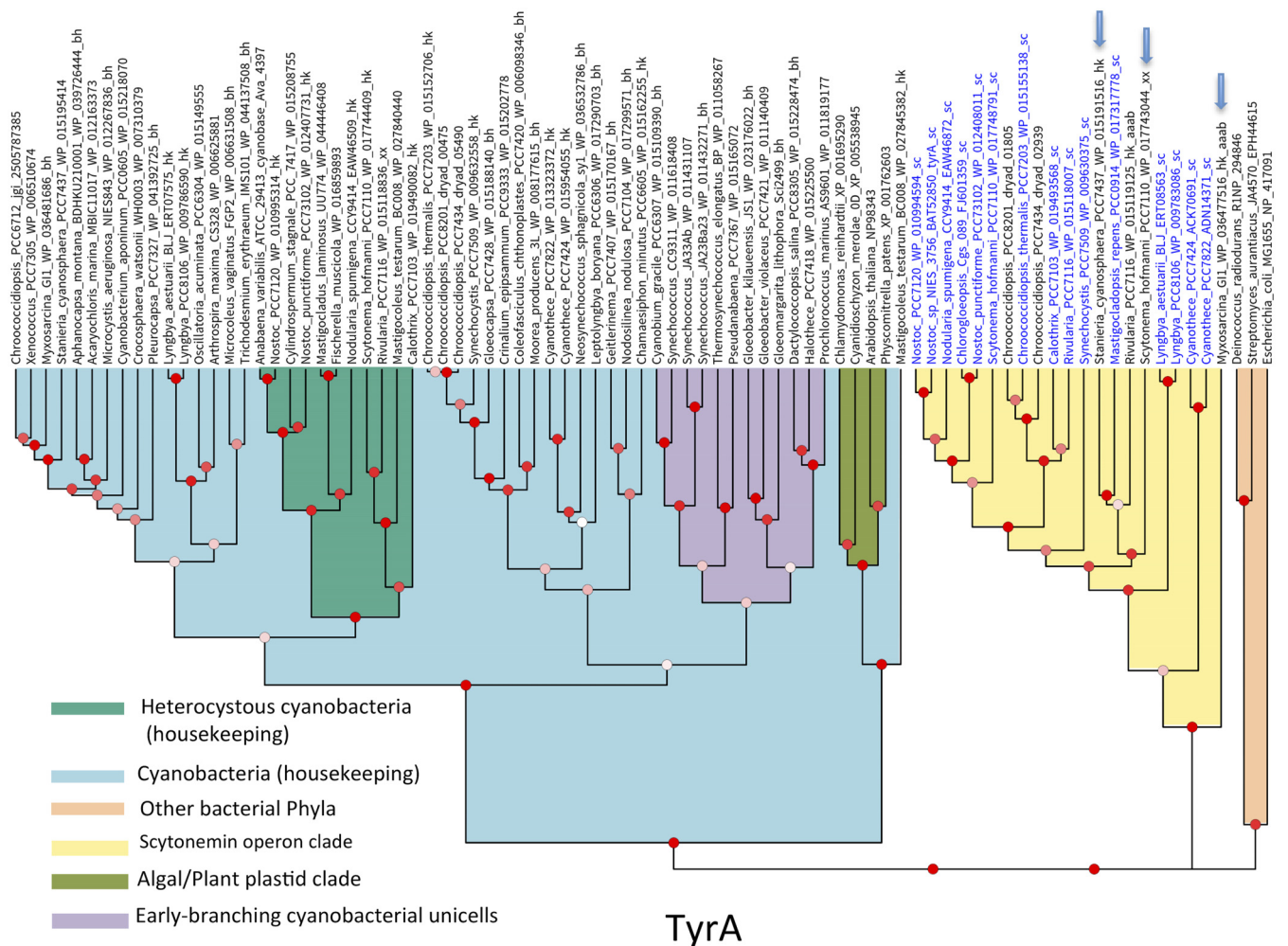

FIG 3 Phylogeny derived from BEAST analyses of TyrA amino acid sequences supports an origin by lateral transfer from bacterial phyla other than cyanobacteria of the homologs in the scytonemin operon. It is shown as exemplary of the three phylogenies (AAAB loci *trpE*, *aroB*, and *tyrA*) supporting such an origin. Phylogenies for the other genes are found in Fig. S5 and S9. Entries in blue type correspond to homologs found within full scytonemin operons. Those marked by a blue arrow were found in remnant scytonemin operons or correspond to supernumerary homologs. Bayesian posterior probabilities (BPP) at the nodes are color coded as follows: red for BPP ≥ 0.8, pink for 0.8 ≥ BPP ≥ 0.5, and white for BPP ≤ 0.5.

through lateral gene transfer from other bacterial lineages (BPP = 1 in the three cases), while their respective cyanobacterial housekeeping clades also show strong support (BPP ≥ 0.99).

A recurrent pattern in most phylogenies (except for AroB) was that AAAB homologs assignable to supernumerary gene copies (additional to the housekeeping and scytonemin-operon versions, many of them belonging to remnant scytonemin operons) were nested within the clades containing "true" scytonemin-operon copies, indicating that they indeed represented old scytonemin-associated copies or had originated as secondary duplications of scytonemin homologs. Thus, in the following narrative, we include such traceable homologs in our definition of the scytonemin clade for any given AAAB phylogeny. In AroB (Fig. S5), a well-supported clade of supernumerary copies that was internal to the scytonemin clade has been related to the biosynthesis of secondary metabolites other than scytonemin (25).

Among five loci that lent support to an origin within the cyanobacteria, phylogeny reconstructions placed the scytonemin-operon AAAB orthologs in well-supported monophyletic groups in two cases (TrpD, BPP = 0.97; TrpC, BPP = 0.98) (Fig. 2 and Fig. S8). The node marking the most recent common ancestor of the scytonemin operon in these cases is obvious. In the other three genes there was always a main clade of scytonemin sequences, but a few of them branched outside this main group. For instance, in TrpA (Fig. S7), *Lyngbya* homologs were clearly paraphyletic to the main scytonemin-associated clade, indicating the presence of two independent recruitment duplications from the housekeeping clade. Incidentally, *Cyanothece* and *Lyngbya* scytonemin operons are those presenting the ancient organizational "pattern 3" and show the deepest-branching patterns for ScyABC (Fig. S1) in our trees. A similar case is observed in AroG (Fig. S6), where the scytonemin homologs of *Cyanothece* are paraphyletic to the main scytonemin-associated clade. Finally, for TrpB sequences, two scytonemin-operon clades could be defined (0.8 > BPP > 0.5), a main clade containing most homologs and a second clade containing the *Lyngbya* homologs and a group of supernumerary copies (Fig. 4). The poor resolution of that part of the tree, however, implies that one cannot exclude the possibility that the separation in two clades may not represent the true evolutionary history of these genes. In the latter three genes, we take the common ancestors of the most deeply branching clade as oldest evidence for the existence of the scytonemin operon.

In order to determine if the AAAB gene phylogenies are congruent with conventional knowledge of cyanobacterial evolution, we tested them for compliance with a set of criteria that can be derived from current molecular phylogenetic and paleontological evidence (5, 30–32). The criteria included the following: (i) the deepest/earliest branches should correspond to unicells (minimally *Gloeobacter*, but often including *Gloeomargarita*, and thermophilic *Synechococcus* homologs), (ii) followed in distance from the root by simple filamentous forms, with (iii) clades of heterocystous forms appearing later (i.e., more derived) within the cyanobacterial radiation. Additionally, (iv) the algal/plant plastid lineages had to be monophyletic and branch off close to the base of the cyanobacterial lineage. All eight AAAB housekeeping protein trees met these criteria if representatives of the groups involved were present (Table 1; Fig. 2 to 4 and Fig. S5 to S9). Two potential, if likely inconsequential, exceptions are to be noted. First, one could have doubts about the compliance of TrpD with additional criterion iv. Second, the marine unicellular *Prochlorococcus* clade in all trees branched deeper/earlier than in standard 16S rRNA gene trees, potentially reflecting long branch artifacts due to the higher divergence in the well-described process of genome simplification that these taxa have experienced (33).

Timing of evolutionary events. To calibrate the timing of the evolutionary appearance of scytonemin and to avoid the use of geochemical markers that rely on redox changes, we used exclusively fossil evidence and refrained from using geochemically derived constraints other than the age of the Earth. We also avoided using dates derived from molecular fossils (hopanoids), as these have recently been discredited
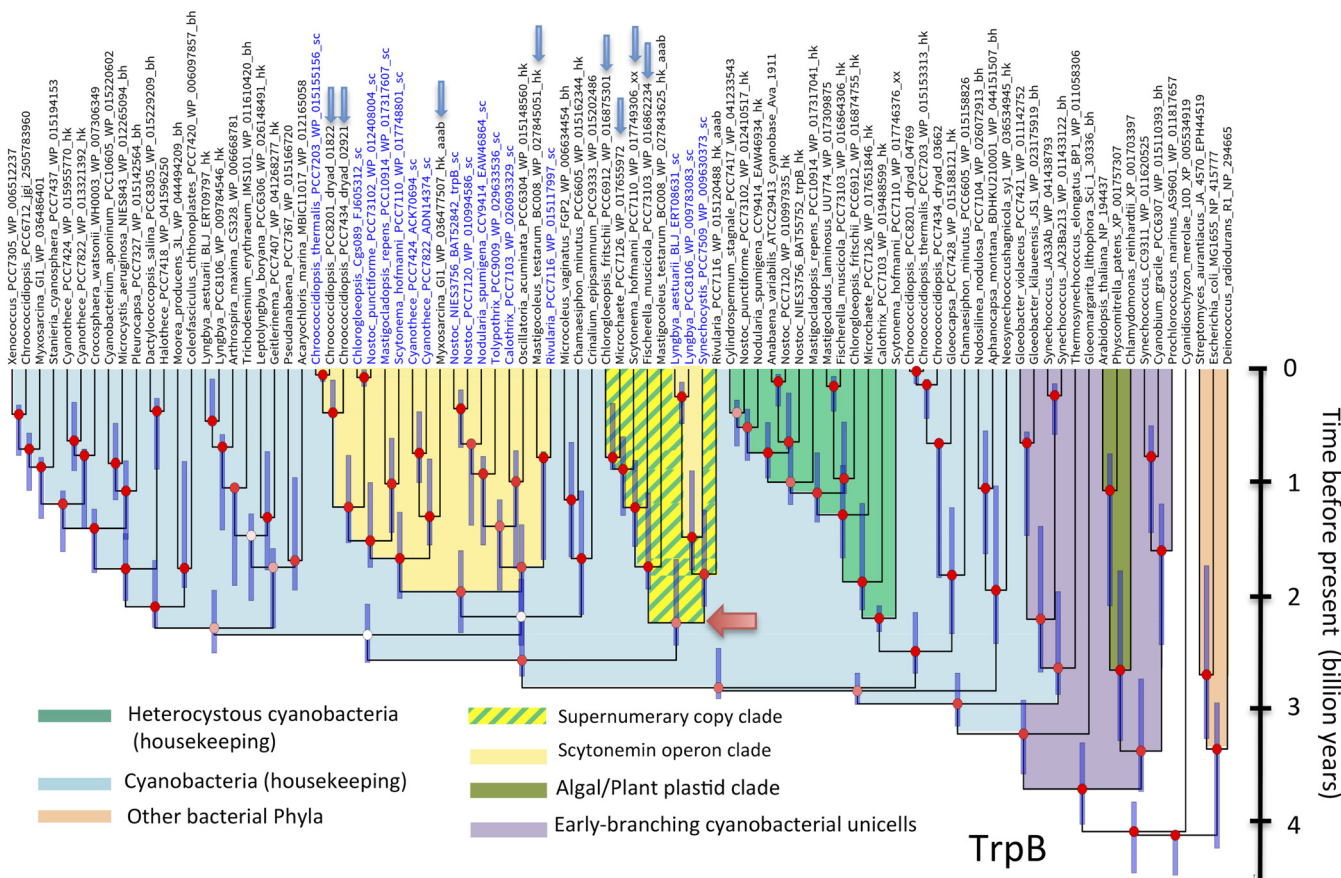
FIG 4 Phylogeny derived from BEAST analyses of TrpB amino acid sequences supports an origin of scytonemin-associated homologs by more than one internal duplication of housekeeping genes, shown as exemplary of the three phylogenies (*trpB*, *trpA*, and *aroG*) supporting such an origin. Phylogenies for the other genes are found in Fig. S6 and S7). Entries in blue type correspond to homologs found within full scytonemin operons. Those marked by a blue arrow were found in remnant scytonemin operons or correspond to supernumerary homologs. Bayesian posterior probabilities (BPP) at the nodes are color coded as follows: red for BPP ≥ 0.8, pink for 0.8 ≥ BPP ≥ 0.5, and white for BPP ≤ 0.5. The red arrow marks the node corresponding to the most recent common ancestor of the oldest clade containing scytonemin-associated homologs.

(34). We also refrained from using the 3.3-Ga age of the oldest cyanobacterium-like fossils of Schopf and Packer (35), which have been used by others (29, 36), because their biogenicity has been effectively challenged (37). We used an absolute 2-point calibration. For the first point, we chose the appearance of filamentous cyanobacteria at 2.7 Ga based on dating of the Tumbiana microfossils (38), which remains uncontroversial and is backed up by the contemporaneous evidence for structures built exclusively by filamentous cyanobacteria (38) and the presence of oxygen bubbles in fossil stromatolites (39). Given that the deepest branches in our trees are populated by unicellular strains, we assigned this first calibration point to the node containing the oldest filamentous strains. The second point was the age of the clade comprising the orders *Nostocales* and *Stigonematales* (heterocyst forming). An age of 2.1 Ga is widely used for this point (29, 31, 36, 40), but a recent publication has challenged the interpretation of the original fossils (41), claiming that the oldest fossils for this group are from the base of the Tonian (some 0.9 Ga old), at the beginning of the Neoproterozoic. For our interpretation, the new fossils are also problematic in that they are too large to be cyanobacteria and do not show heterocysts. Given this situation, we have pragmatically constrained the origin of the heterocystous clade to each of the two possible dates (2.1 and 0.9 Ga) and have run the model-based analyses independently. The results proved rather insensitive to this source of uncertainty (Fig. 5). The models were only otherwise constrained with the condition that the deepest node in each tree could not be older than 4.5 Ga, the age of the Earth. All AAAB genes but those showing an origin in lateral
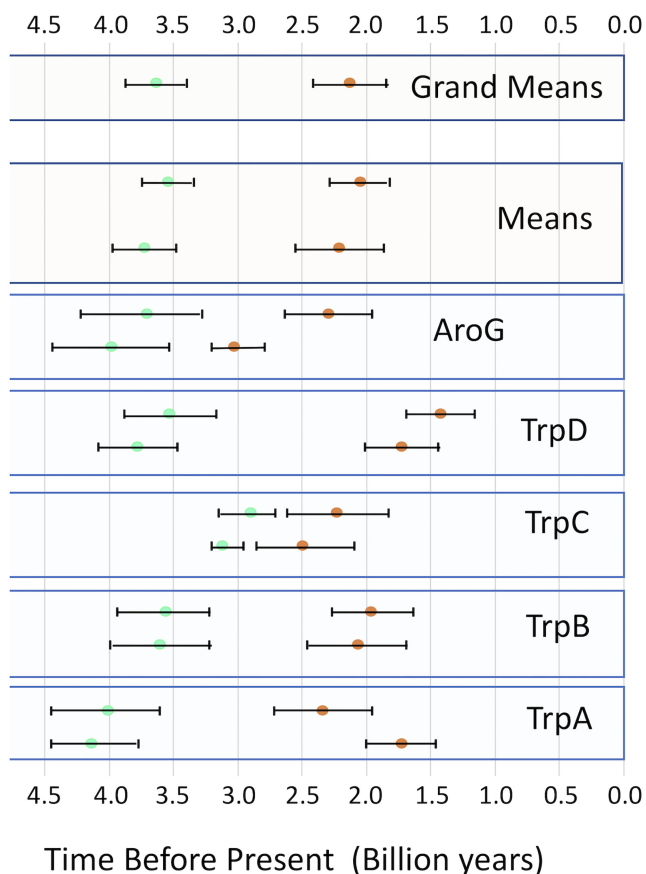
FIG 5 Time estimates for the origin of cyanobacteria (blue-green) and scytonemin (orange) derived from relaxed molecular clock models applied to the phylogeny of gene products associated with the scytonemin operon. Each of the lower 5 boxes corresponds to a gene product as marked. In each of these boxes, the upper estimates used an age of 0.9 billion years for the common ancestor of all heterocystous cyanobacteria, whereas the lower estimates used an age of 2.1 billion years. Error bars are 95% CI for each estimate. The two uppermost boxes include averages of single-gene estimates for each of the two age scenarios (Means) or for all estimates (Grand Means).

gene transfer from other phyla were used for the divergence time estimation, and the results are summarized in Fig. 5. The estimates for the minimal age of the scytonemin operon were determined as the age of the common ancestor of all bona fide scytonemin-operon homolog proteins in the trees that had a monophyletic distribution. For the genes that produced paraphyletic scytonemin clades, the estimated age of the earliest-branching clade was used instead. The estimates were obtained as the geometric mean of the probability distribution for the relevant node and ranged from 3.0 to 1.7 Ga, with a grand mean ($n = 10$) of 2.12 (± 0.28, 95% CI) Ga. Similarly, we estimated the age of cyanobacteria ($=$oxyphotobacteria) as the age of the respective nodes encompassing all housekeeping homologs; they ranged from 2.7 to 4.1 Ga, with a mean estimate of 3.63 $\pm$ 0.24 Ga.

Once the evolutionary history of the scytonemin operon has been traced through the appropriate AAAB genes, it is possible to interpret phylogenies of standard marker genes under that light. Using the same set of strains as in the AAAB phylogenies, and the same time calibrations, 16S rRNA trees (Fig. 6) yielded topologies congruent with the conventional criteria stated above. The estimates for the origin of scytonemin-operon-containing cyanobacteria (2.1 to 2.6 Ga) and the origin of cyanobacteria (3.7 to 4.1 Ga) here are remarkably consistent with the estimates based on the AAAB genes. Alternatively, a concatenated tree of conserved genes could have been used to decrease the variation in age estimates, and any potential bias, when one attempts to obtain an evolutionary history of a species or group thereof, as long as all genes shared
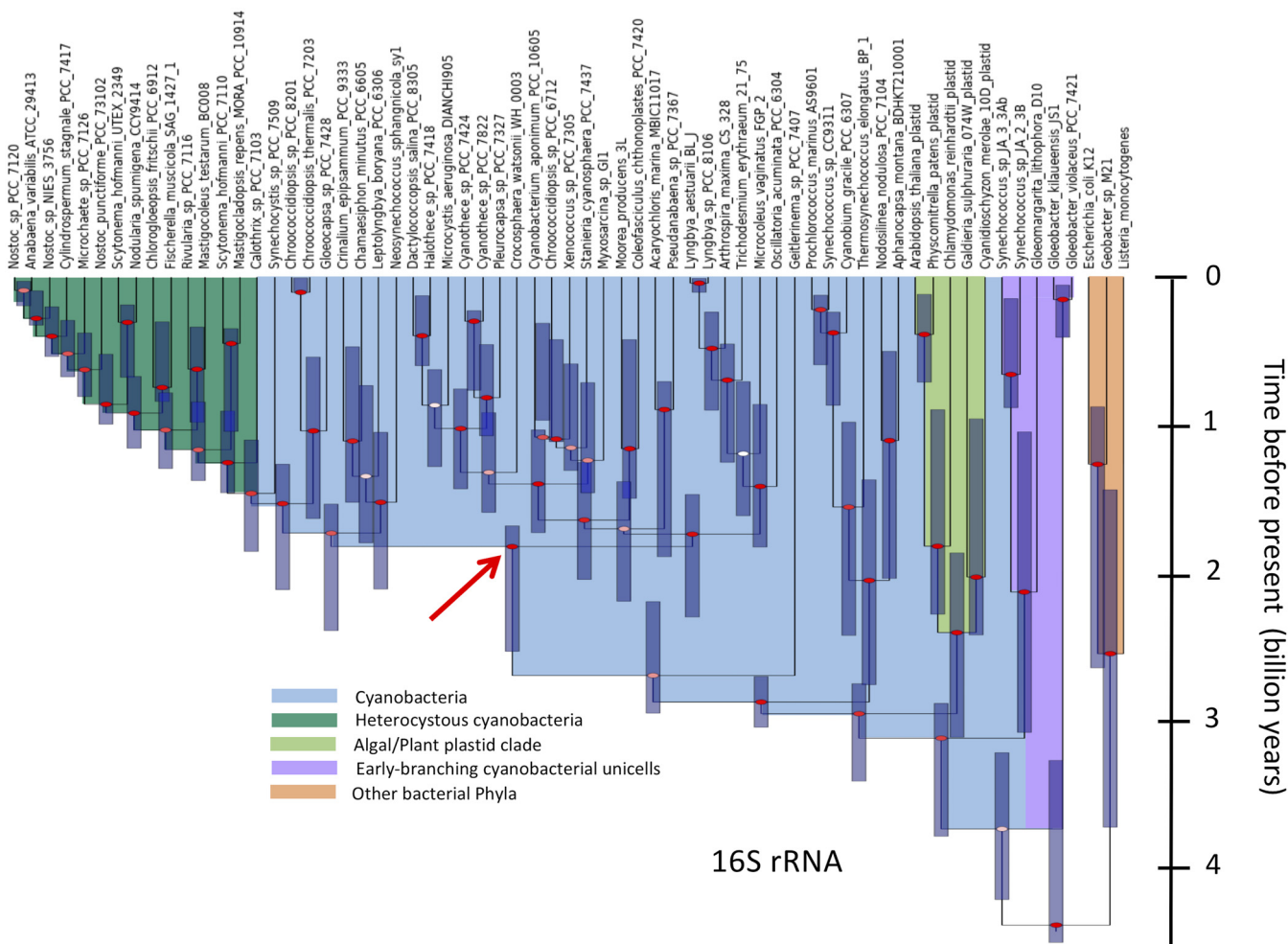
FIG 6 Phylogeny derived from BEAST analyses of full 16S rRNA sequences of cyanobacteria. Bayesian posterior probabilities (BPP) at the nodes are color coded as follows: red for BPP ≥ 0.8, pink for 0.8 ≥ BPP ≥ 0.5, and white for BPP ≤ 0.5. The red arrow marks the node corresponding to the most recent common ancestor of the clades encompassing scytonemin-operon containing genomes.

a congruent evolutionary history. It should prove interesting to see if our findings using multiple genes within the operon would find confirmation using concatenated core genes, as they did when using the widely used 16S rRNA gene for this purpose.

DISCUSSION

The high degree of conservation in sequence, synteny, and regulatory elements of the scytonemin operon speaks for a genetic locus of secondary metabolism that has presumably conferred a fitness advantage to cyanobacteria through much of their history. Although many cyanobacteria possess this locus, they are by no means a majority of those strains with fully sequenced genomes. This likely reflects the well-known preferential distribution of scytonemin among cyanobacteria from environments that allow only periodic growth (16). The operon appears to have originated once but has probably been lost multiple times during diversification, as attested by the presence of operon remnants in some strains. The synteny of these genes has been also quite conserved, although a major reorganization apparently took place early on, from an ancient arrangement now typical of filamentous and secondarily unicellular/colonial cyanobacteria, to yield a more modern one characteristic of the heterocystous groups. The operon itself is conspicuously absent from many of the sequenced isolates in the earliest branches of cyanobacterial diversification and from the early-branching lineage that gave rise to plastids (27), implying that scytonemin appeared sometime

early during the diversification of cyanobacteria but not at its root. This is consistent with the notion of a time gap between the appearance of cyanobacteria and the GOE, during which a microbial UVA sunscreen would have provided little fitness value (see the introduction). The presumably most ancient core part of the operon benefited from the recruitment of genes dedicated to supplying the aromatic amino acid precursors. Our analyses clearly suggest that three AAAB genes were acquired through lateral transfer and the rest by duplication of housekeeping genes. The latter allowed us to obtain multiple independent estimates of the sequence of events and the timing of their duplication.

Our estimates of the age of the scytonemin operon should be considered conservative in that an ancient but functional operon must have existed that contained only the core genes before the AAAB genes were recruited into it in order to optimize precursor supply. Estimates for the minimum age of the scytonemin operon, with an average of 2.12 ± 0.28 Ga, are consistent with a variety of previous estimates for the GOE (2, 4, 10, 11). It is a possibility that including the "remnant genes" in the scytonemin clade may have artificially pushed back our estimate of its origin, if the remnants were nonfunctional and thus not subject to purifying selection; unfortunately, we can provide no evidence for their functional status one way or the other. In any event, our estimates have added value in that they do not rely on any geochemical evidence; they are based exclusively on molecular evolution models and fossil evidence and thus constitute a truly independent avenue to the same end. In view of the final estimates, this alternative avenue converges with geochemical estimates.

Our data sets also allow us to obtain multiple estimates of the origin of the cyanobacteria (=oxyphotobacteria), perhaps the single most important evolutionary event in the history of the Earth. While minimum ages for the presence of cyanobacteria can principally be derived from the (micro)fossil record, rocks of this age are exceedingly rare and fossils in them are few and poorly preserved. There is little doubt among paleontologists that the actual origin of cyanobacteria very likely and significantly predates the best-preserved, noncontroversial fossils (7). Using the molecular evolutionary models anchored with fossil calibrations may be a robust way of circumventing this limitation. And yet, the use of fossil markers, particularly the oldest ones, remains a source of uncertainty in this and other studies (29, 31, 36, 40), even when we strived to select a deep-time marker that was congruent among various lines of evidence. Importantly, we also contend that the convergence of the estimates of the age of scytonemin with the already-available geochemical estimates of the GOE provides an internal cross-disciplinary check on the goodness of our methodology, lending additional credence to our concurrent estimates of the age of the most recent common ancestor of cyanobacteria at 3.63 ± 0.24 Ga. Results from similar analyses carried out on 16S rRNA genes (but using the same time markers) from the same group of strains are consistent with this timing. Rigorously, however, the inference, based on the redox chemistry of scytonemin, that its advent must postdate the GOE (see the introduction), may be regarded as too indirect. If this stance is taken, then the current results can be interpreted as meaning that scytonemin evolution indeed postdated the geochemically established GOE, rather than adding an independent estimate to it.

The key innovation that the evolution of oxygenic photosynthesis represents meant that for the first time a phototroph could use a ubiquitous source of electrons, water, for the reduction of CO_2. This liberated photosynthesis from the ecological servitude of more localized and limited sources of reductant, possibly enabling the colonization of the continents (42, 43), where the occurrence of scytonemin is prevalent (44). From our study, this event can be traced back deep into the Archaean Eon. These estimates are remarkably close to the indirect, synthetic view offered in Schopf's 2011 review (45). In contrast, however, recent molecular clock estimates using concatenated central metabolism genes (46) provide much more recent dates for the same event, almost coincident with that of the GOE. That study, however, used calibrations based on the plant fossil record applied to the branch of the cyanobacterial phylogeny, even when

it can hardly be expected that conserved genes in the free-living cyanobacteria and their symbiotic relatives within plants be subject to similar evolutionary rates.

The absence of concurrent time estimates for the origin of cyanobacteria and the GOE based on the same data sets has prevented rigorously gauging the length of the lag period between the appearance of oxygenic photosynthesis and the GOE, a time of "oxygen oases," geochemical "whiffs of oxygen" (1), exceedingly extreme photochemical harshness (15), and likely great evolutionary significance. Our estimates, whose means differ by 1.5 billion years, suggests that it may have taken cyanobacteria this long to change the Earth's Archaean biosphere from neutral/reducing to oxidizing in nature, a period 3-fold as long as the age of the oldest green plants or metazoans. It is conceivable that fully oxygenated niches, some perhaps quite large, existed during this time in closed basins, providing opportunities for the evolution of biological oxidative metabolism and oxidative damage defense mechanisms in multiple, independent lineages (47) and, in time, for the eventual tipping point in the evolutionary trajectory of life that the GOE represents.

MATERIALS AND METHODS

Genomic data and reconstructions. DNA sequences of the genes of the scytonemin operon of *Nostoc punctiforme* ATCC 29133 (PCC 73102) that had homologs in the aromatic amino acid biosynthesis (AAAB) pathway, as well as the corresponding AAAB amino acid sequences, were used as queries for similarity searches (see Text S1 in the supplemental material). Initial BLASTp searches (48) were carried out against GenBank on a broad selection of genomes from cyanobacteria and a few representatives from bacteria and plastid-bearing eukaryotes (Text S1). Owing to the high number of homologs detected, however, genomes were recovered from GenBank (www.ncbi.nlm.nih.gov/genome/), CyanoBase (http://genome.microbedb.jp/cyanobase/), JGI (https://genome.jgi.doe.gov/), and Dryad (https://datadryad.org/) to build a local cyanobacterial database that was both taxonomically representative of this phylum and enriched in organisms with a scytonemin operon or particular relevance for subsequent calibration purposes (Text S1). The final BLASTp searches on cyanobacteria were carried out with default parameters and permissive E values (<1e−05) against that local database.

For each relevant protein-coding gene detected in a cyanobacterial genome, the five protein-coding sequences up- and downstream of that locus were recorded to determine if they were likely part of a larger scytonemin operon ("SC copies") or isolated and therefore likely involved in housekeeping ("HK copies") or other functions. The proteins tentatively annotated as SC copies were further evaluated by examining up- and downstream sequences until there were three proteins in a row that were dissimilar to proteins in the *N. punctiforme* scytonemin operon in each direction. This approach was used to delimit the scytonemin protein cluster in all of the chosen taxa and consequently provide greater confidence in annotating them as SC or HK copies. In some cases, we found supernumerary copies (beyond SC or those initially labeled HK). These were also included, and we determined *post hoc* if they were related to the main HK or the SC lineages. Preliminary alignments of each gene were carried out using MUSCLE v3.8.31 (49), the sequences were trimmed using the program NET of the MUST package (50), and preliminary phylogenies were reconstructed using FastTree v.2.1.7 (51). The examination of these preliminary trees revealed the presence in some of our data sets of highly divergent sequences, which generally clustered with their respective outgroup in the preliminary phylogenies and in some cases could be attributed to other functions based on searches carried out against the Pfam 28.0 (http://pfam.xfam.org/), PDB (http://www.rcsb.org/pdb/home/home.do), or KEGG (http://www.genome.jp/kegg/) database. These preliminary analyses were particularly important to determine the relationship of SC-TrpEG, SC-AroG, and SC-TrpD with their respective HK copies or with more distant relatives. We removed the sequences that were too divergent from our data sets, as well as those too short, too long, identical to others, or incomplete, as they could introduce unnecessary noise to our analyses. New MUSCLE v3.8.31 alignments were carried out on the curated data sets, and the statistical support of each alignment was checked using the Guidance server (http://guidance.tau.ac.il/ver2/) (52). The final alignments were trimmed using the program BMGE (53) and analyzed for similarity by neighbor-joining methods (total character distance) with nonparametric bootstrap analyses using PAUP* (54).

Divergence time estimation methods. Times to the most recent common ancestor (tmrca) were estimated using BEAST v1.8.1 (55). Four independent chains were run for each data set, and each was run for 50 million iterations and sampled every 5,000 iterations, initially. Some data sets were run for 100 million iterations to ensure convergence, and if the chain failed to converge, up to 8 additional runs were used to search for convergence. The posterior probability distributions were inspected manually using Tracer v1.6 (56). At least the first 10% of the samples were discarded as "burn-in," and the 95% highest-posterior-density region (HPD) was calculated using Tracer. Phylogenies were summarized by TreeAnnotator v1.8.1 from the BEAST package. The tmrca were estimated using the following parameter settings: each data set was fitted with the LG amino acid substitution model (57) $+ \Gamma +$ invariant sites, and a Yule speciation model with uncorrelated lognormal-distributed relaxed clock model was employed. Random starting phylogenies were used under conditions in which the initial maximum age for a monophyletic cyanobacterium was set to 3,000 million (Ma) years ago. For time calibration, we used the appearance of filamentous cyanobacteria, and the appearance of heterocystous cyanobacteria as

described in Results. The prior distribution for filamentous cyanobacteria was a truncated normal distribution with the mean 2,700 Ma and standard deviation 295, truncated between 2,700 and 3,200 Ma. The prior distribution for heterocystous cyanobacteria was set as an exponential distribution with mean of 210 Ma and offset (lower bound) of 2,100 Ma. The prior distribution for the age of the root node was a uniform distribution between 2,100 and 4,500 Ma. This means that the deepest node was constrained at the age of the Earth. All other priors used the default BEAST priors.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .00561-19.

TEXT S1, PDF file, 0.1 MB.
FIG S1, PDF file, 0.3 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 0.2 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.6 MB.
FIG S6, PDF file, 0.5 MB.
FIG S7, PDF file, 0.5 MB.
FIG S8, PDF file, 0.6 MB.
FIG S9, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J, Buick R. 2007. A whiff of oxygen before the great oxidation event? Science 317:1903. https://doi.org/10.1126/science.1140325.

2. Bekker A, Holland HD, Wang PL, Rumble D, III, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ. 2004. Dating the rise of atmospheric oxygen. Nature 427:117–120. https://doi.org/10.1038/nature02260.

3. Catling DC, Claire MW. 2005. How Earth's atmosphere evolved to an oxic state: a status report. Earth Planet Sci Lett 237:1–20. https://doi.org/10.1016/j.epsl.2005.06.013.

4. Kump LR. 2008. The rise of atmospheric oxygen. Nature 451:277–278. https://doi.org/10.1038/nature06587.

5. Knoll AH. 2008. Cyanobacteria and Earth history, p 1–19. In Herrero A, Flores E (ed), The cyanobacteria: molecular biology, genomics, and evolution. Caister Academic Press, Poole, United Kingdom.

6. Blankenship RE, Hartman H. 1998. The origin and evolution of oxygenic photosynthesis. Trends Biochem Sci 23:94–97. https://doi.org/10.1016/S0968-0004(98)01186-4.

7. Buick R. 2008. When did oxygenic photosynthesis evolve? Philos Trans R Soc B Biol Sci 363:2731. https://doi.org/10.1098/rstb.2008.0041.

8. Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE. 2013. Atmospheric oxygenation three billion years ago. Nature 501: 535–538. https://doi.org/10.1038/nature12426.

9. Lalonde SV, Konhauser KO. 2015. Benthic perspective on Earth's oldest evidence for oxygenic photosynthesis. Proc Natl Acad Sci U S A 112: 995–1000. https://doi.org/10.1073/pnas.1415718112.

10. Canfield DE. 2005. The early history of atmospheric oxygen: homage to Robert M. Garrels. Annu Rev Earth Planet Sci 33:1–36. https://doi.org/10.1146/annurev.earth.33.092203.122711.

11. Scott C, Lyons TW, Bekker A, Shen Y, Poulton SW, Chu X, Anbar AD. 2008. Tracing the stepwise oxygenation of the proterozoic ocean. Nature 452:456–459. https://doi.org/10.1038/nature06811.

12. Imlay JA. 2002. How oxygen damages microbes: oxygen tolerance and obligate anaerobiosis. Adv Microb Physiol 46:111–153. https://doi.org/10.1016/S0065-2911(02)46003-1.

13. Kasting JF, Donahue T. 1980. The evolution of atmospheric ozone. J Geophys Res 85:3255–3263. https://doi.org/10.1029/JC085iC06p03255.

14. Castenholz RW, Garcia-Pichel F. 2000. Cyanobacterial responses to UV-radiation, p 591–611. In Whitton BA (ed), The ecology of cyanobacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.

15. Garcia-Pichel F. 1998. Solar ultraviolet and the evolutionary history of cyanobacteria. Orig Life Evol Biosph 28:321–347. https://doi.org/10.1023/A:1006545303412.

16. Garcia-Pichel F, Castenholz RW. 1991. Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment. J Phycol 27:395–409. https://doi.org/10.1111/j.0022-3646.1991.00395.x.

17. Garcia-Pichel F. 1994. A model for internal self-shading in planktonic organisms and its implications for the usefulness of ultraviolet sunscreens. Limnol Oceanogr 39:1704–1717. https://doi.org/10.4319/lo.1994.39.7.1704.

18. Garcia-Pichel F, Sherry ND, Castenholz RW. 1992. Evidence for an ultraviolet sunscreen role of the extracellular pigment scytonemin in the terrestrial cyanobacterium Chlorogloeopsis sp. Photochem Photobiol 56: 17–23. https://doi.org/10.1111/j.1751-1097.1992.tb09596.x.

19. Soule T, Stout V, Swingley WD, Meeks JC, Garcia-Pichel F. 2007. Molecular genetics and genomic analysis of scytonemin biosynthesis in Nostoc punctiforme ATCC 29133. J Bacteriol 189:4465–4472. https://doi.org/10.1128/JB.01816-06.

20. Soule T, Palmer K, Gao Q, Potrafka RM, Stout V, Garcia-Pichel F. 2009. A comparative genomics approach to understanding the biosynthesis of the sunscreen scytonemin in cyanobacteria. BMC Genomics 10:336. https://doi.org/10.1186/1471-2164-10-336.

21. Gao Q, Garcia-Pichel F. 2011. Microbial ultraviolet sunscreens. Nat Rev Microbiol 9:791–802. https://doi.org/10.1038/nrmicro2649.

22. Ferreira D, Garcia-Pichel F. 2016. Mutational studies of putative biosynthetic genes for the cyanobacterial sunscreen scytonemin in Nostoc punctiforme ATCC 29133. Front Microbiol 7:735. https://doi.org/10.3389/fmicb.2016.00735.

23. Soule T, Garcia-Pichel F, Stout V. 2009. Gene expression patterns associated with the biosynthesis of the sunscreen scytonemin in Nostoc punctiforme ATCC 29133 in response to UVA radiation. J Bacteriol 191:4639–4646. https://doi.org/10.1128/JB.00134-09.

24. Klicki K, Ferreira D, Hamill D, Dirks B, Mitchell N, Garcia-Pichel F. 2018. The widely conserved ebo cluster is involved in precursor transport to the periplasm during scytonemin synthesis in Nostoc punctiforme. mBio 9:e02266-18. https://doi.org/10.1128/mBio.02266-18.

25. Yurchenko T, Ševčíková T, Strnad H, Butenko A, Eliáš M. 2016. The plastid

genome of some eustigmatophyte algae harbours a bacteria-derived six-gene cluster for biosynthesis of a novel secondary metabolite. Open Biol 6:160249. https://doi.org/10.1098/rsob.160249.

26. Naurin S, Bennett J, Videau P, Philmus B, Soule T. 2016. The response regulator Npun_F1278 is essential for scytonemin biosynthesis in the cyanobacterium Nostoc punctiforme ATCC 29133. J Phycol 52:564–571. https://doi.org/10.1111/jpy.12414.

27. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U S A 110:1053–1058. https://doi.org/10.1073/pnas.1217107110.

28. Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, Tandeau de Marsac N, Gugger M, Lockhart PJ, Allen JF, Brune I, Maus I, Pühler A, Martin WF. 2013. Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol Evol 5:31–44. https://doi.org/10.1093/gbe/evs117.

29. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. Proc Natl Acad Sci U S A 103:5442–5447. https://doi.org/10.1073/pnas.0600999103.

30. Garcia-Pichel F, Nübel U, Muyzer G. 1998. The phylogeny of unicellular, extremely halotolerant cyanobacteria. Arch Microbiol 169:469–482. https://doi.org/10.1007/s002030050599.

31. Schirrmeister BE, de Vos JM, Antonelli A, Bagheri HC. 2013. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. Proc Natl Acad Sci U S A 110:1791–1796. https://doi.org/10.1073/pnas.1209927110.

32. Castenholz RW. 2001. Cyanobacteria, p 474–487. In Boone DR, Castenholz RW (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol 1. Springer, New York, NY.

33. Dufresne A, Garczarek L, Partensky F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol 6:R14. https://doi.org/10.1186/gb-2005-6-2-r14.

34. French KL, Hallmann C, Hope JM, Schoon PL, Zumberge JA, Hoshino Y, Peters CA, George SC, Love GD, Brocks JJ, Buick R, Summons RE. 2015. Reappraisal of hydrocarbon biomarkers in Archean rocks. Proc Natl Acad Sci U S A 112:5915–5920. https://doi.org/10.1073/pnas.1419563112.

35. Schopf J, Packer B. 1987. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. Science 237:70–73. https://doi.org/10.1126/science.11539686.

36. Bahl J, Lau MCY, Smith GJD, Vijaykrishna D, Cary SC, Lacap DC, Lee CK, Papke RT, Warren-Rhodes KA, Wong FKY, McKay CP, Pointing SB. 2011. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. Nat Commun 2:163. https://doi.org/10.1038/ncomms1167.

37. Brasier MD, Green OR, Jephcoat AP, Kleppe AK, Van Kranendonk MJ, Lindsay JF, Steele A, Grassineau NV. 2002. Questioning the evidence for Earth's oldest fossils. Nature 416:76–81. https://doi.org/10.1038/416076a.

38. Walter M. 1972. Stromatolites and the biostratigraphy of the Australian Precambrian and Cambrian. Spec Pap Palaeontol 11:1–190.

39. Bosak T, Liang B, Sim MS, Petroff AP. 2009. Morphological record of oxygenic photosynthesis in conical stromatolites. Proc Natl Acad Sci U S A 106:10939–10943. https://doi.org/10.1073/pnas.0900885106.

40. Blank CE, 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. https://doi.org/10.1111/j.1472-4669.2009.00220.x.

41. Pang K, Tang Q, Chen L, Wan B, Niu C, Yuan X, Xiao S. 2018. Nitrogen-fixing heterocystous cyanobacteria in the Tonian period. Curr Biol 28:616–622.e1. https://doi.org/10.1016/j.cub.2018.01.008.

42. Thomazo C, Couradeau E, Garcia-Pichel F. 2018. Possible nitrogen fertilization of the early Earth Ocean by microbial continental ecosystems. Nat Commun 9:2530. https://doi.org/10.1038/s41467-018-04995-y.

43. Beraldi-Campesi H, Farmer JD, Garcia-Pichel F. 2014. Modern terrestrial sedimentary biostructures and their fossil analogs in Mesoproterozoic subaerial deposits. Palaios 29:45–54. https://doi.org/10.2110/palo.2013.084.

44. Couradeau E, Karaoz U, Lim HC, Nunes da Rocha U, Northen T, Brodie E, Garcia-Pichel F. 2016. Bacteria increase arid-land soil surface temperature through the production of sunscreens. Nat Commun 7:10373. https://doi.org/10.1038/ncomms10373.

45. William Schopf J. 2011. The paleobiological record of photosynthesis. Photosynth Res 107:87–101. https://doi.org/10.1007/s11120-010-9577-1.

46. Shih PM, Hemp J, Ward LM, Matzke NJ, Fischer WW. 2017. Crown group Oxyphotobacteria postdate the rise of oxygen. Geobiology 15:19–29. https://doi.org/10.1111/gbi.12200.

47. Soo RM, Hemp J, Parks DH, Fischer WW, Hugenholtz P. 2017. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science 355:1436–1440. https://doi.org/10.1126/science.aal3794.

48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

49. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

50. Philippe H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res 21:5264–5272. https://doi.org/10.1093/nar/21.22.5264.

51. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

52. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res 38:W23–W28. https://doi.org/10.1093/nar/gkq443.

53. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210.

54. Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.

55. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075.

56. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

57. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307–1320. https://doi.org/10.1093/molbev/msn067.