

Exploring thylakoid emergence: evolution of membrane biogenesis and photosystem II assembly in early-diverging cyanobacteria

Louise Hambücker , Denis Baurain  and Luc Cornet 

InBioS-PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, Liège, 4000, Belgium

Author for correspondence:
Luc Cornet
Email: luc.cornet@uliege.be

Received: 11 February 2026
Accepted: 27 April 2026

New Phytologist (2026)
doi: 10.1111/nph.71284

Key words: cyanobacteria evolutionary biology, membrane dynamics, photosynthesis, phylogenomics, PSII assembly, thylakoids.

Summary

- Thylakoid membranes (TM) in cyanobacteria and chloroplasts host the light-dependent reactions of oxygenic photosynthesis. *Gloeobacterales*, the earliest-diverging cyanobacterial lineage, lack TM and perform photosynthesis in the cytoplasmic membrane (CM), representing an ancestral state relative to other cyanobacteria (Phycobacteria). This study investigates the evolutionary origin of TM.
- Phylogenomic analyses were performed across a phylogenetically diverse set of cyanobacteria, including extensive representation of basal lineages (*Gloeobacterales*, *Thermotichales*, *Gloeomargaritales*, and *Pseudanabaenales*), as well as micro- and macrocyanobacteria, using orthologous proteins involved in membrane dynamics and photosystem II (PSII) assembly, together with structural modeling using AlphaFold3.
- We identified two candidate proteins associated with membrane trafficking that may contribute to TM biogenesis, including the SPFH (Stomatin, Prohibitin, Flotillin, en HflK/C) family member Slr1106, proposed to have been acquired by lateral gene transfer. Analysis of 36 PSII assembly factors revealed modifications in late-stage assembly, notably in manganese homeostasis. Structural changes in the YidC translocase may have facilitated the relocation of linear electron transfer components from the CM to TM.
- Altogether, these phylogenetic and functional prediction analyses provide new insight into the molecular innovations that led to TM emergence, including membrane trafficking systems, PSII assembly changes, and protein targeting adaptations.

Introduction

Thylakoids are specialized membranes present in both cyanobacteria and chloroplasts that carry out the light-dependent reactions of oxygenic photosynthesis. They generate NADPH and ATP for subsequent CO₂ fixation in the Calvin–Benson cycle through a linear electron transport (LET) chain composed of multi-subunit complexes, notably photosystem II (PSII) and photosystem I (PSI), together with associated electron carriers. The emergence of this intracellular compartment provided cyanobacteria with an expanded membrane surface for light energy conversion, which is hypothesized to have contributed to the Great Oxidation Event (GOE) (Guéguen & Maréchal, 2022) *c.* 2.4 billion years ago, profoundly transforming Earth's biosphere (Bekker *et al.*, 2004).

Among extant cyanobacteria, members of the order *Gloeobacterales*, the earliest-diverging lineage, lack thylakoid membranes (TM) (Rippka *et al.*, 1974; Rahmatpour *et al.*, 2021). Instead, they carry out oxygenic photosynthesis within specialized regions of their cytoplasmic membrane (CM) (Rexroth *et al.*, 2011). The absence of TM in this basal group suggests that these structures arose after the divergence of *Gloeobacterales* but before the

emergence of *Thermotichales*, the second earliest-diverging cyanobacterial lineage, which already possesses TM. Consequently, *Gloeobacterales* retain the ancestral evolutionary state preceding TM formation and provide a unique model for exploring the transition from a LET located in the CM to a LET integrated within TM (Cornet, 2025). The proteins mediating the assembly of photosynthetic complexes within the TM may retain evolutionary signatures of this ancestral relocation, offering valuable insights into how cyanobacteria with TM, known as Phycobacteria, have emerged (Cornet, 2025).

The mechanisms underlying TM biogenesis remain poorly understood. Four main hypotheses have been proposed: fusion between CM and TM, vesicular transport, direct lipid transfer via soluble lipid carriers, or self-assembly driven by lipid phase transitions (Rast *et al.*, 2015; Guéguen & Maréchal, 2022) (Fig. 1). The Vesicle-Inducing Protein in Plastids (VIPP1), an homolog of the eukaryotic Endosomal Sorting Complex Required for Transport III (ESCRT-III) system implicated in membrane deformation and scission, was suggested to play an essential role in the biogenesis of TM, thereby fitting the first scenario (Westphal *et al.*, 2001; Heidrich *et al.*, 2017; Tan, 2024).

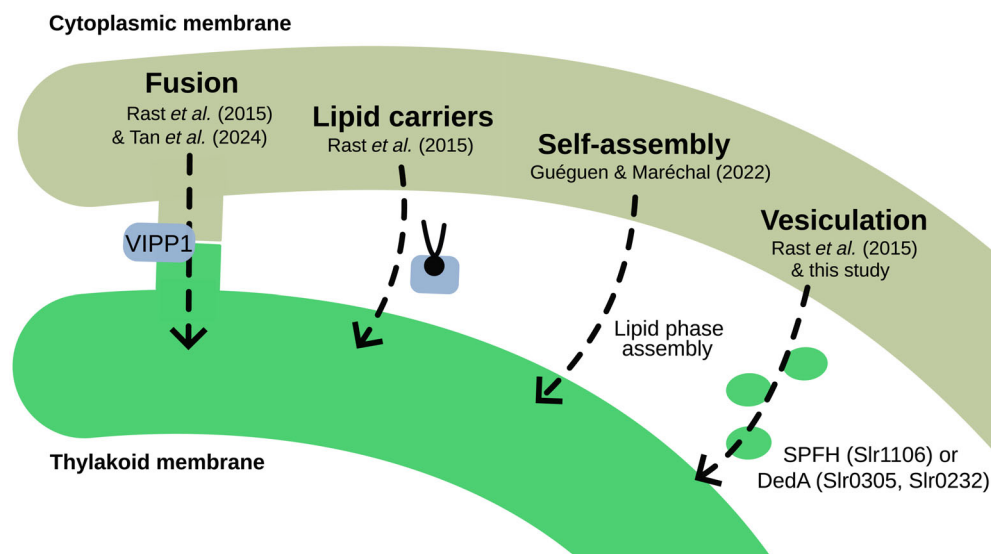


Fig. 1 Proposed mechanisms of thylakoid membrane (TM) biogenesis. Schematic overview of the four main hypotheses proposed to explain TM biogenesis: (1) membrane fusion between the cytoplasmic membrane (CM) and the TM (Rast et al., 2015); (2) direct lipid transfer mediated by soluble lipid carriers (Rast et al., 2015); (3) self-assembly driven by lipid phase transitions (Guéguen & Maréchal, 2022); photosys and (4) vesicular transport from the CM (Rast et al., 2015).

This conclusion was supported by the fact that VIPP1 was only found in oxygenic photosynthetic organisms and that its presence is essential for thylakoid formation (Westphal et al., 2001). VIPP1 evolved from a duplication of the stress response phage shock protein A (PspA) at the root of Phycobacteria and differs from PspA by an additional C-terminal extension (Tan, 2024). However, a recent study showed that the *Gloeobacteriales* *pspA* gene could complement a *vipp1*-null mutant in *Arabidopsis* (Ma et al., 2025). Furthermore, proteomic data from TM biogenesis indicate that VIPP1 accumulates at late stages of TM formation (Huang et al., 2023). Overall, these recent findings question the previously assumed central role of VIPP1 in the process.

Although the mechanisms underlying the assembly of most of the protein complexes forming the LET remain poorly studied, *de novo* assembly of PSII was shown to be a highly synchronized spatial and temporal process in *Synechocystis*, mediated by at least 36 proteins known as assembly factors (Van De Meene et al., 2006; Heinz et al., 2016; Rast et al., 2019). During the early steps of PSII biogenesis, the precursor of D1 (pD1) is translocated across the CM by the Sec translocon and the YidC/Alb3 insertase, followed by Chl incorporation and D1 maturation. The D1–D2 heterodimer then forms within a specialized region of the CM enriched in assembly factors, known as the PrtA-defined membrane (PDM) (Stengel et al., 2012). During the later steps of PSII biogenesis, the inner antenna subunits CP47 and CP43 are sequentially incorporated, and the Mn_4CaO_5 cluster is assembled and activated (Rengstl et al., 2011). These later steps likely occur within the developing TM.

Here, we present a phylogenomic analysis of 13 proteins involved in membrane trafficking that identifies one member of the SPFH family and one member of the DedA family as potential contributors to TM biogenesis, providing an alternative to the VIPP1-based model. We further report the first phylogenetic analysis of the 36 PSII assembly factors known so far, highlighting patterns of gene presence/absence and structural variations that suggest important modifications in the late stages of PSII

assembly, notably in manganese homeostasis, after the divergence of *Gloeobacteriales*. In parallel, we pinpoint structural changes in the translocase YidC and the Sec machinery that may have facilitated the relocation of the LET components from CM to TM.

Materials and Methods

Genome selection and phylogenomic species tree

A total of 45 316 representative species genomes of Terrabacteria were retrieved from the Genome Taxonomy Database (GTDB) release r220 (Parks et al., 2022) (Figshare: doi: 10.6084/m9.figshare.31135354). To limit computational burden during orthologous gene inference, a subset of 950 genomes were selected based on genome quality and phylogenetic position. Briefly, only the 43 827 genomes that passed the Genome UNClutterer (GUNC) (Orakov et al., 2021) contamination filter were retained. Within each GTDB order, genomes were then ranked by the difference between their CheckM2 (Chklovski et al., 2023) completeness and contamination scores. For orders represented by a single genus, the two highest-quality genomes were kept; for all other orders, the best genome of each genus was selected, yielding 1452 genomes in total, excluding the Cyanobacteriota phylum (Supporting Information Table S1). Sampling within cyanobacteria was more exhaustive, resulting in 432 genomes, that is one or two representative genomes were chosen per genus, prioritizing those with > 80% completeness and < 10% contamination, as estimated by CheckM2 (Chklovski et al., 2023) (Table S2). To ensure coverage of early-diverging cyanobacterial lineages, all available genomes from the orders *Gloeobacteriales*, *Thermotrichales*, and *Gloeomargaritales* were incorporated, whereas the more abundant *Pseudanabaenales* were dereplicated using ToRQUEMADA v.0.2.1, with default parameters, resulting in 343 genomes (Léonard et al., 2020).

Protein-coding sequences were obtained directly from the National Center for Biotechnology Information (NCBI)

proteomes when available or predicted using PRODIGAL v.2.6.3 (Hyatt *et al.*, 2010) (Table S3). A phylogenomic tree based on ribosomal proteins was reconstructed using the ORPER workflow with default parameters from the GEN-ERA toolbox v.1 (Cornet *et al.*, 2023), with the exception that MAFFT v.7.471 was used for the alignment of the ribosomal proteins (Katoh, 2002) and that the ribosomal tree was inferred using IQ-TREE v.2.4.0 (Minh *et al.*, 2020) with the best-fit substitution model (Q.YEAST + I + R10) and 1000 ultrafast bootstrap replicates (Fig. S1). The resulting phylogeny was pruned using PARNAS v.0.1.6 (Markin *et al.*, 2023) to retain a final set of 950 representative species (Table S4; Figs S2, S3), encompassing *c.* 90% of the total phylogenetic diversity, as estimated by diversity score metrics (Fig. S4), and comprising 30 cyanobacterial species.

Orthologous group inference

The orthologous groups were inferred with ORTHOFINDER v.2.5.5 (Emms & Kelly, 2019) from the 950 genomes selected using default parameters. An approach based on *Hierarchical Orthologous Groups* was first explored to account for multigene families stemming from duplications. However, this strategy proved to be less accurate than the one based on standard *Orthologous Groups* (OGs), which is detailed below.

Identification of the orthologous groups of interest

The OGs corresponding to known assembly factors and membrane dynamics proteins were identified by performing BLASTP searches (v.2.9.0+) (Camacho *et al.*, 2009) of every protein from the 950 proteomes against a database containing reference sequences of known PSII assembly factors and proteins involved in membrane dynamics (Tables S5, S6). The *e*-value threshold was fixed at 10^{-3} , and only the first hit was kept. The hits were then plotted based on their bitscore distributions for each OG (Figs S5–S50). For each assembly factor and membrane dynamics protein, the OG with the highest median bitscore for cyanobacterial hits was selected (Table S7). To check for optimal OG delineation, a reciprocal BLASTP of every reference sequence against a database containing the 950 proteomes was also performed. The best hits were then filtered based on their *e*-value, sequence length, number of copies per organism (first best hit per organism for the best score), and the fraction of the sequence covered by the alignment with the reference protein using OmpaPa (<https://metacpan.org/dist/Bio-MUST-Apps-OmpaPa>; D. Baurain). We then cross-validated the OGs derived from the filtered sequences against those obtained from the bitscore distribution plots.

Orthologous gene tree inference

The gene trees of each OG of interest were generated using IQ-TREE v.3.0.0 (Hoang *et al.*, 2018; Wong *et al.*, 2025) with the best-fit substitution model and 1000 ultrafast bootstrap replicates (Table S7; Figs S51–S95). The sequences corresponding to the assembly factors and membrane dynamics proteins were then pinpointed in the gene trees based on the bitscores. When

subsequent analyses were needed, for either taxonomic distribution determination or tree topology resolution, a Hidden Markov Model profile (pHMM) was built with the cyanobacterial sequences and searched against a database comprising the inferred proteomes of 343 Cyanobacteriota species (the dereplicated cyanobacterial dataset of 432 genomes, hereafter referred to as the cyanobacterial dataset) using HMMER v.3.4 (Finn *et al.*, 2011). The best hits were then filtered based on their *e*-value, sequence length, number of copies per organism (the three best hits per organism were kept), and the fraction of the sequence covered by the pHMM alignment (minimum 0.7) using OmpaPa. As the selected parameters varied depending on the protein under study, the filtering thresholds are detailed in figshare doi: [10.6084/m9.figshare.31135354](https://doi.org/10.6084/m9.figshare.31135354). Minimal lengths ranged from 8 to 556 amino acids, with a median of 154, and minimal $-\log_{10}$ *e*-values ranged from 0 to 281, with a median of 33.

The selected sequences were then aligned using MAFFT v.7.505 (Katoh, 2002). Alignment sites were conserved if they had at most 70% gaps and the sequences with a minimum of 30% length (with respect to the longest unaligned sequence) were exported using ali2phylipl.pl (<https://metacpan.org/release/Bio-MUST-Core>; D. Baurain). Finally, a gene tree was inferred from the filtered alignment using IQ-TREE v.3.0.0 (Hoang *et al.*, 2018; Wong *et al.*, 2025) with the best-fit substitution model and 1000 ultrafast bootstrap replicates (Figs S96–S136). The presence or absence of the proteins was mapped onto a species tree of the cyanobacterial dataset (Finn *et al.*, 2011) using iTOL (Letunic & Bork, 2024).

Functional protein and 3D protein structure prediction

Functional protein domains were annotated with INTERPRO v.5.48–83.0 (Jones *et al.*, 2014). When the experimental structures were not available, the 3D protein structures were modeled using the ALPHAFOLD3 server (Abramson *et al.*, 2024). The resulting structures were visualized and aligned using CHIMERA-X v.1.11 (Meng *et al.*, 2023). Electrostatic potential calculations were performed using the APBS plugin via the Poisson–Boltzmann server (Jurrus *et al.*, 2018) (<https://server.poissonboltzmann.org/>) with the AMBER force field and default parameters.

All figures were created using Inkscape.

Results

Membrane trafficking proteins

SPFH proteins (implicated in vesicular lipid transport), DedA family proteins (associated with lipid rearrangement), and dynamin-like proteins (involved in membrane fission and fusion) have been proposed by Siebenaller & Schneider (2023) as potential contributors to lipid trafficking in cyanobacteria (Siebenaller & Schneider, 2023). To further explore their evolutionary history among Terrabacteria, we conducted a separate phylogenetic analysis for 13 of these proteins (Fig. 2). Their taxonomic distribution revealed that two are absent from *Gloeobacterales* but present in most Phycobacteria included in this study: Slr1106 and Slr0305.

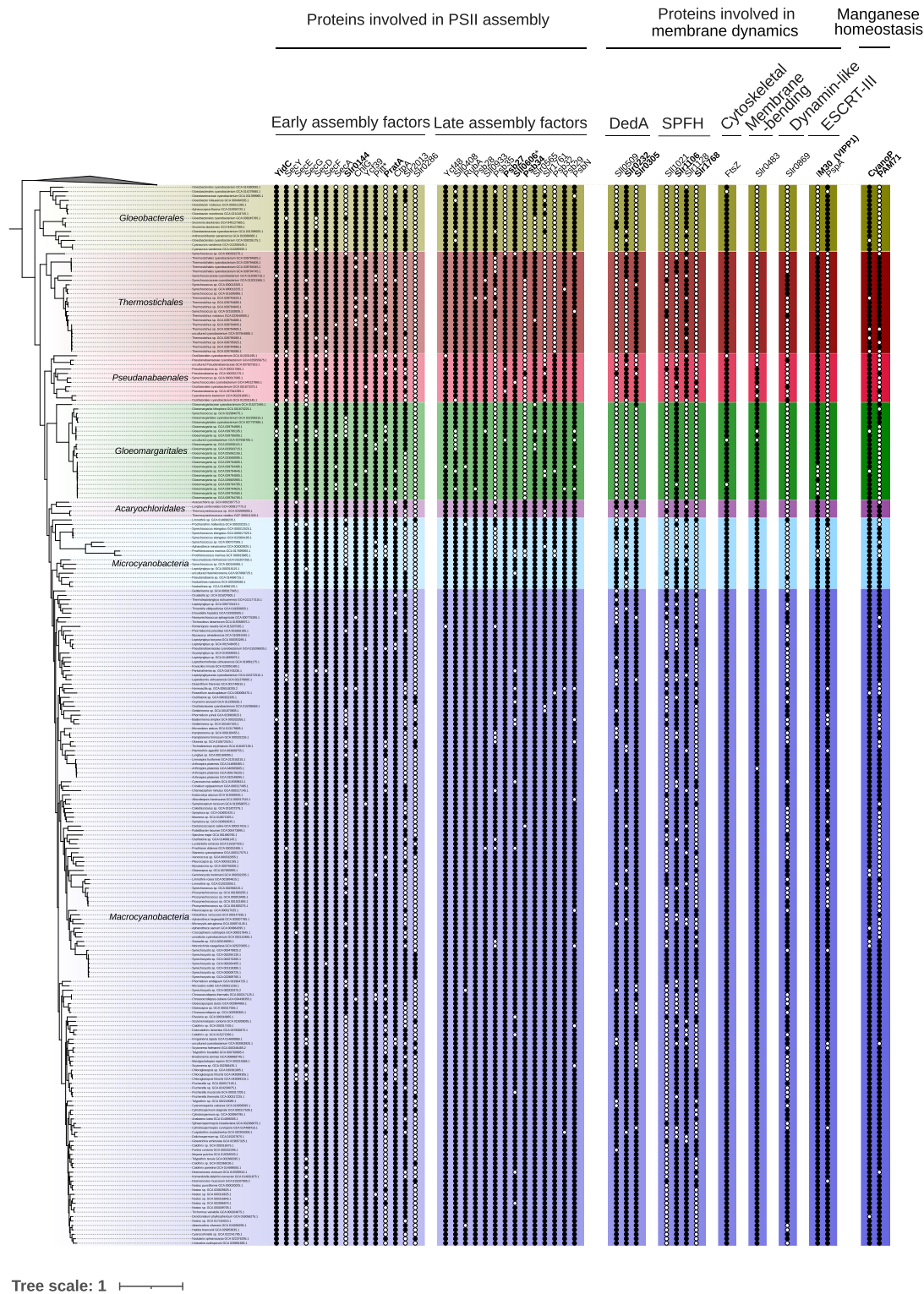


Fig. 2 Taxonomic distribution of the photosystem II assembly factors and proteins involved in membrane dynamics in *Synechocystis*. A black dot indicates the presence of a protein, whereas a white dot indicates its absence. Highlighted proteins are discussed in the manuscript. Proteins marked with an asterisk indicate cases in which the *Gloeobacterales* homolog may not share the same evolutionary origin as the corresponding phycobacterial protein.

The prohibitin Slr1106 is a member of the SPFH family. Its phylogeny within Terrabacteria (Fig. 3) suggests for both proteins an origin through lateral gene transfer (LGT), possibly from

Patescibacteria. Slr1106 appears to have been acquired at the base of the Phycobacteria, whereas Slr1768 may have been acquired later, before the emergence of the *Pseudanabaenales* but after that

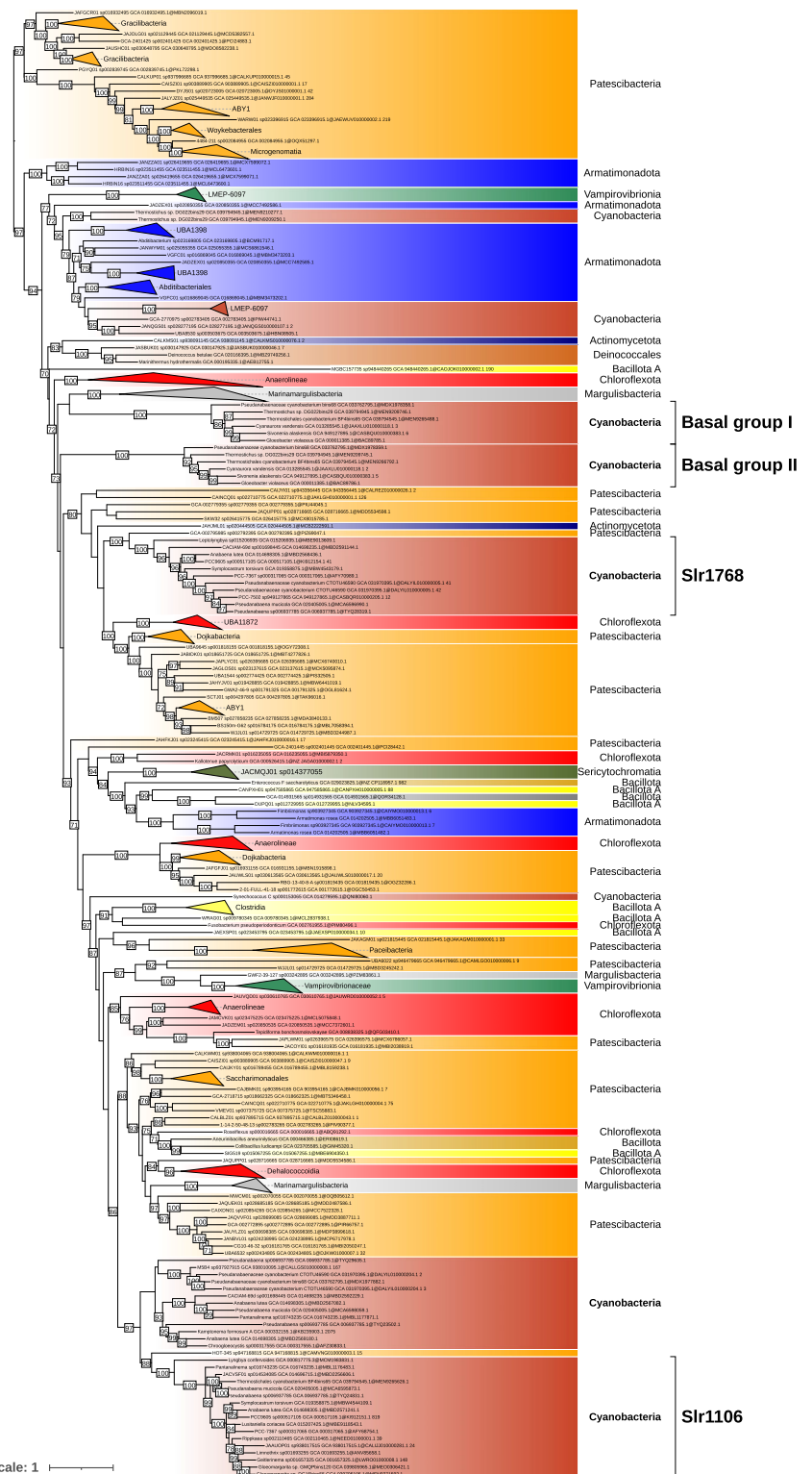


Fig. 3 Phylogenetic tree of the cyanobacterial orthologous sequences of Slr1106 and Slr1768 (SPFH family) in the Terrabacteria dataset. The phylogenetic tree was constructed from 273 sequences across 510 unambiguously aligned positions using IQ-TREE (Q.PLANT + F + I + R6 model) with 1000 ultrafast bootstrap replicates. Bootstrap values $\geq 70\%$ are shown. The tree annotation corresponds to the region showing the highest similarity to the reference *Synechocystis* sequences (NCBI Protein Accessions: BDH96643.1 (Slr1106) and BAA17070.1 (Slr1768)), as determined by BLAST.

of the *Thermotrichales* (Fig. 2). Our analyses showed that the tree of Slr1106 and Slr1768 also contains two groups of basal Cyanobacteria, including *Gloeobacterales*, which lack orthologues in other cyanobacterial orders (Fig. 3). Moreover, structural predictions revealed an additional β -sheet positioned outside of the ring structure formed during oligomerization in the proteins of the

basal group I (Fig. S137). The DedA protein Slr0305 was not detected in *Gloeobacterales* but was present from the root of Phycobacteria (Fig. 4). Another DedA protein, Slr0232, is for its part conserved across all cyanobacteria. However, its phylogeny exhibits a long branch at the base of Phycobacteria (Fig. 5), albeit without discernible structural or sequence divergence.

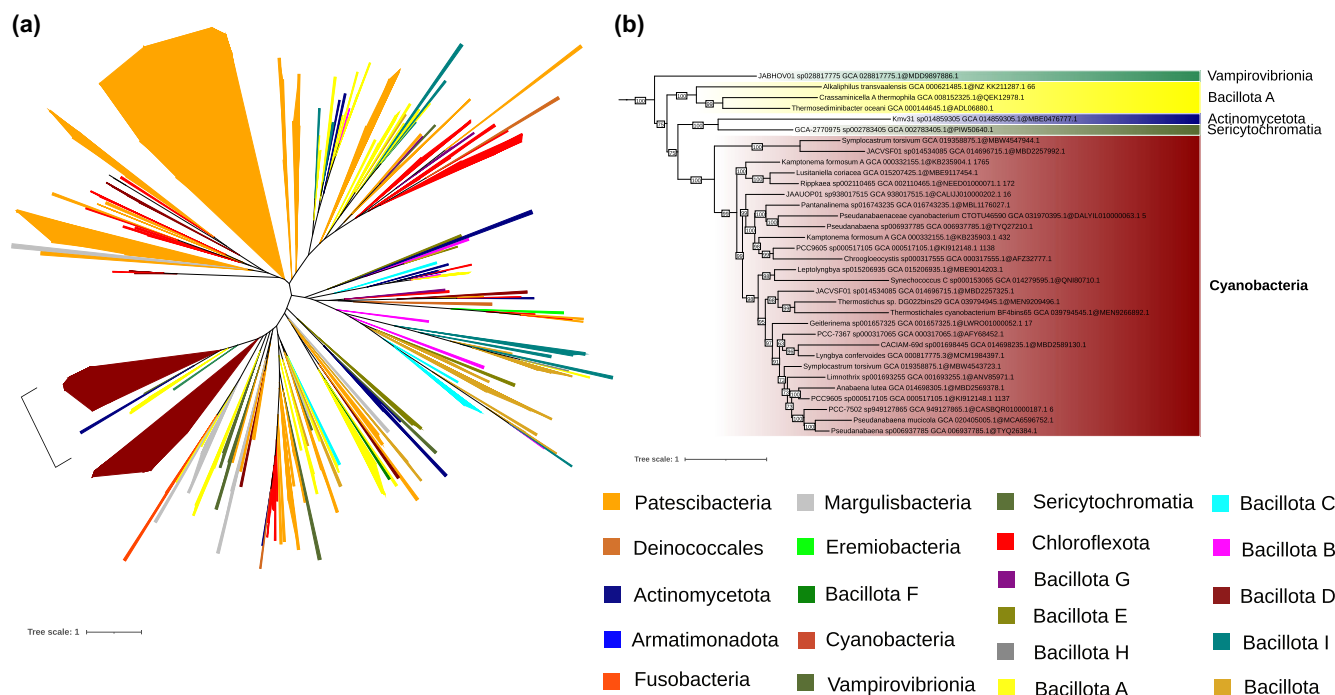


Fig. 4 Phylogenetic tree of the cyanobacterial orthologous sequences of Slr0305 (DedA family) in the Terrabacteria dataset. (a) The phylogenetic tree was constructed from 347 sequences across 504 unambiguously aligned positions using IQ-TREE (LG + F + R7 model) with 1000 ultrafast bootstrap replicates. The region emphasized by brackets corresponds to the region showing the highest similarity to the reference *Synechocystis* sequence (NCBI Protein Accession: BAA10672.1), as identified by BLAST. (b) The region labeled as Slr0305 in (a) was zoomed. Bootstrap values $\geq 70\%$ are shown.

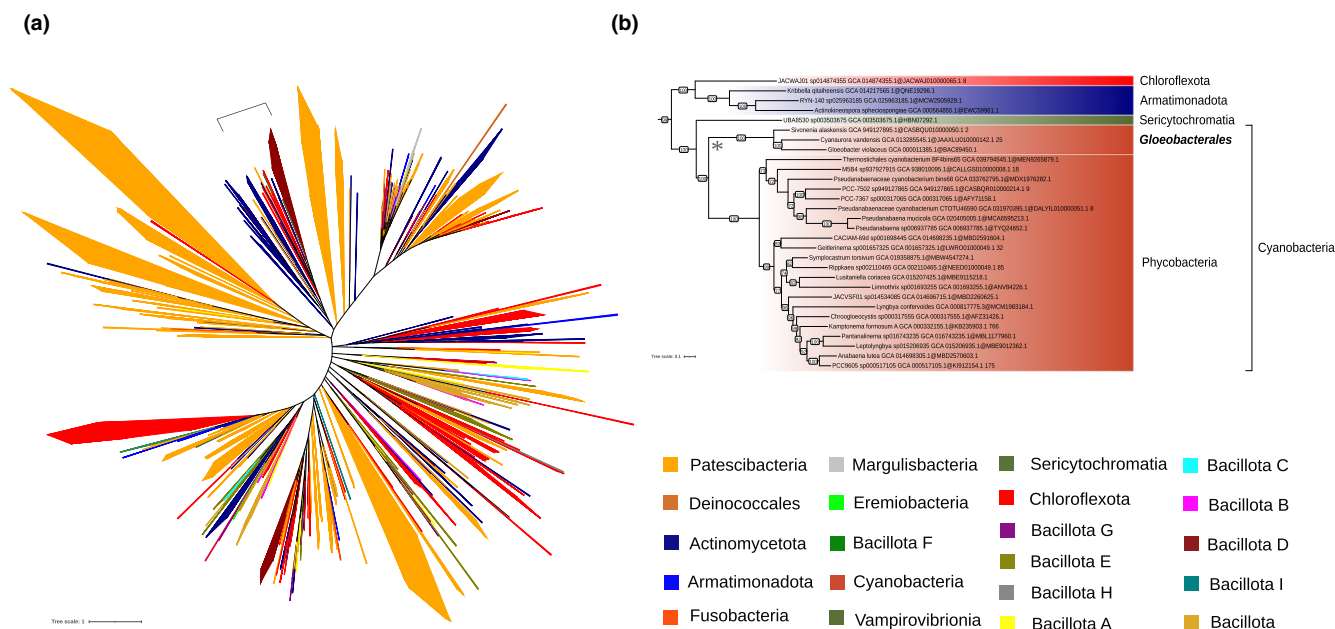
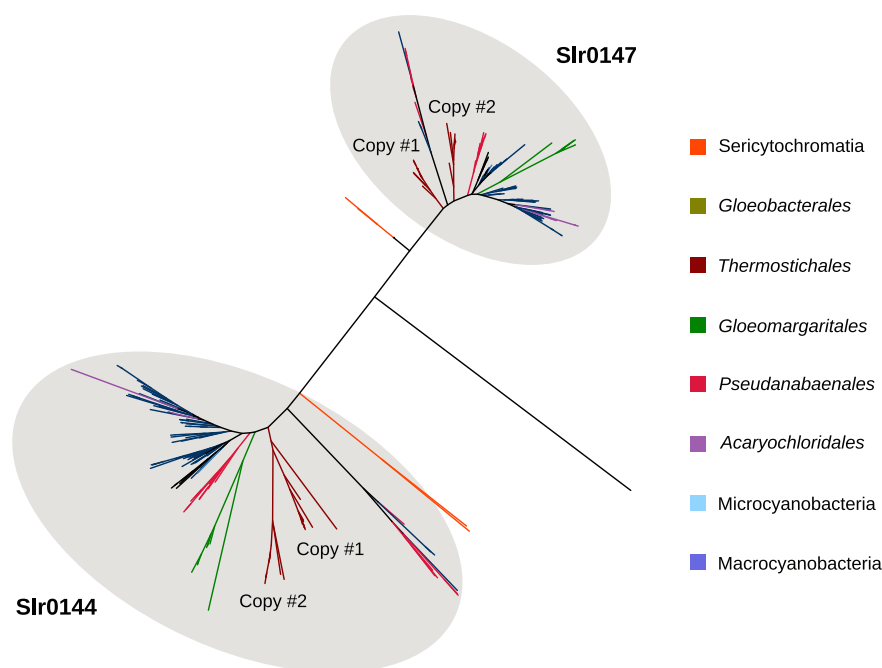


Fig. 5 Phylogenetic tree of the cyanobacterial orthologous sequences of Slr0232 (DedA family) in the Terrabacteria dataset. (a) The phylogenetic tree was constructed from 987 sequences across 522 unambiguously aligned positions using Iq-TREE (LG + F + R10 model) with 1000 ultrafast bootstrap replicates. The region emphasized by brackets in the figure corresponds to the region showing the highest similarity to the reference *Synechocystis* sequence (NCBI Protein Accession: BAA10237.1), as identified by BLAST. (b) The region labeled as Slr0232 in (a) was zoomed. The long branch at the basis of Phycobacteria is indicated by an asterisk. Bootstrap values $\geq 70\%$ are shown.

Fig. 6 Phylogenetic tree of the cyanobacterial orthologues sequences of Slr0144 and Slr0147 in the cyanobacterial dataset. The phylogenetic tree was constructed from 311 sequences across 234 unambiguously aligned positions using IQ-TREE (Q.PFAM+R6) with 1000 ultrafast bootstrap replicates. The tree annotation corresponds to the region showing the highest similarity to the reference *Synechocystis* sequences (NCBI Protein Accessions: BAA18544.1 (Slr0144) and BAA18547.1 (Slr0147)), as determined by BLAST.



Assembly factors of PSII

Late-stage assembly factors The largest differences between *Gloeobacterales* and Phycobacteria were observed at the level of late-stage assembly factors (Fig. 2). Hence, Psb27 and Psb34 were not identified in *Gloeobacterales* in our analyses (Fig. 2). Psb34 is thought to bind CP47 and compete with HliA/B to enable the final stage of PSII assembly (Rahimzadeh-Karvansara *et al.*, 2022), while Psb27 binds both unassembled and assembled CP43 in the PSII complex (Komenda *et al.*, 2012). Sll0606, a protein of unknown function but leading to a massive reduction in D1 and CP43 in PSII assembly upon deletion (Zhang *et al.*, 2010), was not detected either in *Gloeobacterales* (Fig. 2). Domain analysis of Sll0606 suggests a potential role in the transport of hydrophobic components, and distant homologs in *Gloeobacterales* have been proposed (Zhang *et al.*, 2010). In this work, we also identified such a homolog in *Gloeobacterales*: BAC90294.1. Its sequence is not orthologous to the *Synechocystis* PCC6803 reference sequence, but InterPro domain analysis annotates three domains capable of binding hydrophobic components at the N-terminal region (Fig. S138). Hlips family phylogenies are not well-resolved in our analyses, as evidenced by the lack of statistically supported clades separating HliA, HliB, HliC, and HliD. Instead, these proteins were scattered over several OGs (Fig. S9), and a combined phylogenetic tree did not reveal clear distinct monophyletic groups for each gene (Fig. S55).

Except for the accession GCA_038245785.1 (*Gloeobacterales* cyanobacterium), *Gloeobacterales* were also shown to lack Slr0144, a protein containing a 4-vinyl reductase (V4R) domain that binds small hydrophobic molecules and is inferred to interact with Chl precursors in *Rhodobacter capsulatus* (Anantharaman *et al.*, 2001; Singh *et al.*, 2004). In *Synechocystis* sp. PCC6803, Slr0144 is encoded within the *slr0144–slr0152* operon

(Anantharaman *et al.*, 2001), which is involved in PSII stabilization (Yang *et al.*, 2014). Slr0144 and Slr0147, the only two proteins of *Synechocystis* sp. PCC6803 containing a V4R domain, are the result of a gene duplication (Fig. 6). Slr0144 and Slr0147 differ notably by Slr0144 having an N-terminal extension, along with changes in its C-terminal region (Fig. S139).

Early-stage assembly factors The early stages of PSII biogenesis do not show any differences in terms of presence or absence of assembly factors between *Gloeobacterales* and Phycobacteria. Indeed, whether considering the processing and incorporation of the D1 precursor (pD1), D2 folding, or the formation of the reaction center, all corresponding assembly factors are present in *Gloeobacterales* (Fig. 2). However, a structural difference exists in the periplasmic domain of the translocase YidC and SecD between *Gloeobacterales* and Phycobacteria, with the latter harboring a specific features in its periplasmic domain.

YidC is part of the Oxa1/Alb3/YidC family of proteins, which are involved in the insertion and assembly of membrane proteins and have homologs across all three domains of life (Hennon *et al.*, 2015). Notably, bacterial YidC shares homology with mitochondrial inner membrane Oxa1 and chloroplast thylakoid Alb3 proteins. YidC has been shown to participate in PSII biogenesis alongside the SecYEG machinery by mediating the membrane integration of the D1 precursor (pD1) in *Synechocystis* PCC6803, a process homologous to the one described in chloroplasts (Gathmann *et al.*, 2008). In both cyanobacteria and chloroplasts, this process involves the SecYEG translocon, YidC, SecDF, and SecA. SecYEG forms the core membrane channel for protein translocation; YidC acts as a membrane insertase facilitating the integration of membrane proteins; SecDF is an accessory complex that enhances translocation using the proton motive force; and SecA is an ATPase that drives preprotein translocation

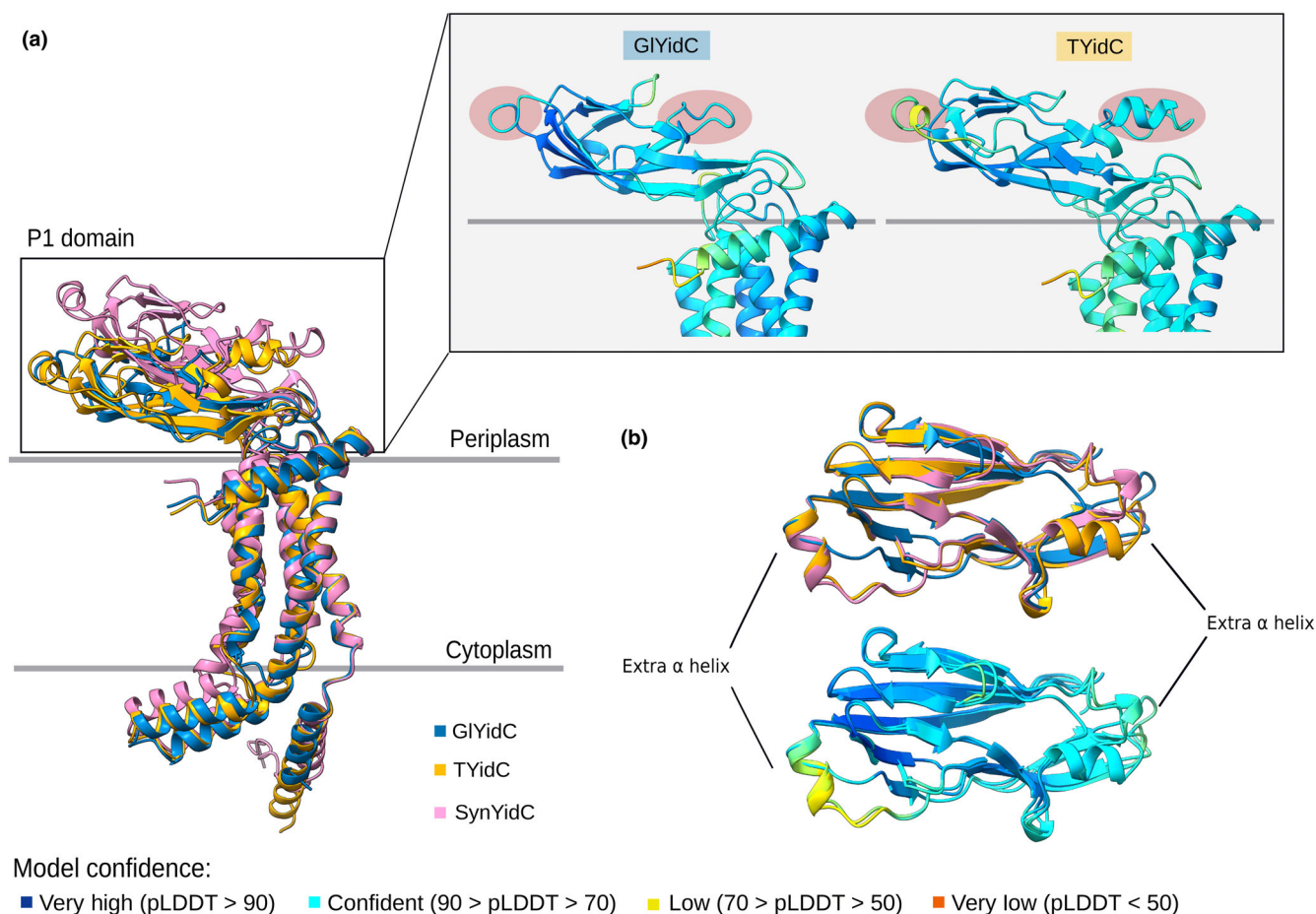


Fig. 7 AlphaFold Prediction of the 3D structures of *Gloeobacter violaceus* PCC7421 (BAC90172.1–GIYidC), *Thermotrichales cyanobacterium* GCA_039794945.1 (MEN9209976.1–TYidC), and *Synechocystis* sp. PCC 6803 substr. Kazusa (P74155.1–SynYidC). (a) Structure alignment of GIYidC (blue), TYidC (yellow), and SynYidC (pink). Positions of the additional α helices are highlighted in red. (b) Top view of the structure alignment of GIYidC, TYidC, and SynYidC periplasmic domains. The Predicted Local Distance Difference Test (pLDDT) scores mapped onto the GIYidC and TYidC indicate confidence levels of the predicted models.

through SecYEG (Frain *et al.*, 2016). Notably, both YidC and SecDF contain periplasmic domains. The periplasmic domain of YidC is only present in diderm bacteria, such as cyanobacteria, as its size is markedly reduced in other bacterial and chloroplastic homologs (Hennon *et al.*, 2015). Interestingly, our results indicate that the periplasmic domain of cyanobacterial YidC is largely conserved across the cyanobacterial phylum, except in *Gloeobacteriales*. Sequence alignment reveals well-conserved N-terminal and C-terminal regions; however, the periplasmic domain of *Gloeobacteriales* YidC differs, with missing amino acids in the sequence (Fig. S140). Besides, 3D structure prediction of the periplasmic domain highlights these differences, showing that Phycobacteria YidC possesses a P1 domain with structural changes and two additional helix loops compared to the *Gloeobacteriales* homolog, one predicted with confidence (pLDDT ≥ 70) and the other with low support (70 > pLDDT ≥ 50) (Fig. 7). In order to investigate how these changes affect YidC interaction with the Sec machinery, we performed the same analyses for SecD and SecF, the only two components of the Sec

machinery that harbor periplasmic domains (Nouwen & Driesen, 2002). Our data show that SecD is predicted to contain an additional helix–loop insertion in *Thermotrichales*, supported with relatively low confidence (70 > pLDDT ≥ 50), with this loop being even larger in *Synechocystis* and modeled with high confidence (overall pLDDT ≥ 70). (Fig. S141). By contrast, no structural differences were observed in the periplasmic domain of *Thermotrichales* SecF, whereas such differences were identified in *Synechocystis* (Fig. S142).

PratA, a periplasmic tetratricopeptide repeat protein strongly associated with the PDM region (Rengstl *et al.*, 2011), has been shown to facilitate efficient Mn delivery to pD1 (Klinkert *et al.*, 2004). Our taxonomic distribution analyses showed the absence of PratA in most *Gloeobacteriales*, at the exception of two accessions: GCA_018389385.1 (*Anthocerotibacter panamensis*) and GCA_038245785.1 (*Gloeobacteriales cyanobacterium*) (Fig. 2). CyanoP, the cyanobacterial homolog of the PsbP protein known in land plants, is a Mn-binding protein that plays a role either in the binding association of CP47 and CP43 or in

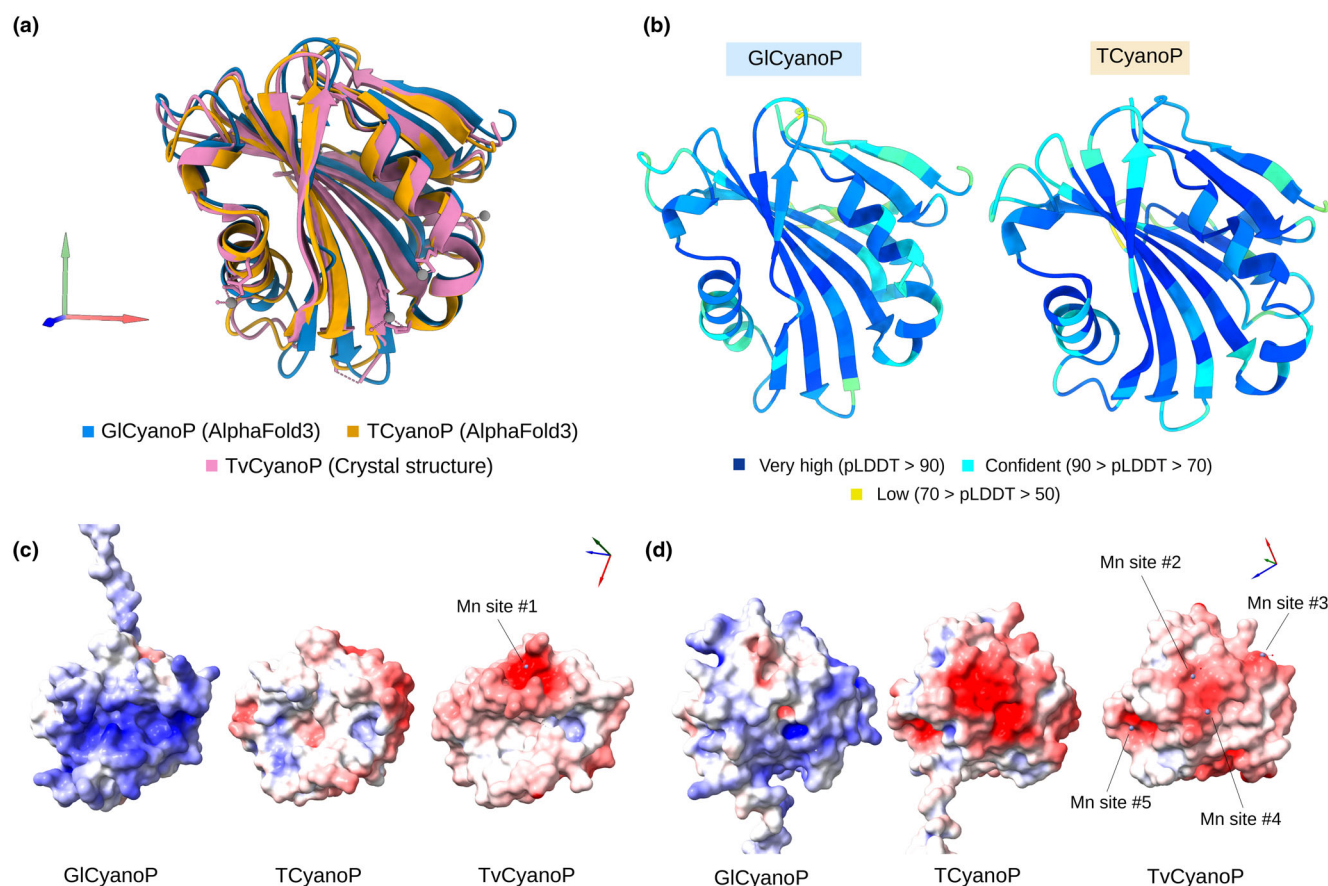


Fig. 8 AlphaFold Prediction of the 3D structures of *Gloeobacter violaceus* PCC7421 (BAC89381.1–GICyanoP), *Thermotrichales cyanobacterium* GCA_039794945.1 (MEN9209059.1–TCyanoP), and PDB 2XB3 CyanoP structure from *Thermosynechococcus vestitus* BP-1 (TvCyanoP) with APBS electrostatics estimation. (a) Structure alignment of the predicted GICyanoP (blue) and TCyanoP (yellow) AlphaFold structures and the TvCyanoP crystal structure (PDB: 2XB3) (pink). (b) Predicted Local Distance Difference Test (pLDDT) scores mapped onto CyanoP structures GICyanoP and TCyanoP, indicating confidence levels of the predicted models. (c, d) APBS electrostatics estimation of GICyanoP, TCyanoP, and TvCyanoP structures with two different views displaying the putative manganese-binding sites. The GICyanoP and TCyanoP predicted structures were aligned to TvCyanoP in order to find the position of the ion-binding sites. The APBS electrostatics was estimated based on the APBS plugin via the Poisson–Boltzmann server (<https://server.poissonboltzmann.org/>) with the AMBER force field and default parameters.

the photoactivation of the oxygen-evolving complex (Cormann *et al.*, 2014). While our analyses show the presence of CyanoP in *Gloeobacterales*, its sequence has diverged substantially compared to Phycobacteria (Fig. S143). Adaptive Poisson–Boltzmann Solver (APBS) electrostatics calculations suggest that *Gloeobacterales* CyanoP either binds divalent ions differently compared to its Phycobacteria counterpart or does not bind ions at all (Fig. 8). Photosynthesis Affected Mutant 71 (PAM71) is a Mn transporter involved in manganese homeostasis, sequestering excess cytoplasmic manganese into the periplasmic and luminal compartments (Gandini *et al.*, 2017). Its role in Mn uptake is considered to support PSII oxygen-evolving complex performance, suggesting a PrtA-independent Mn delivery pathway (Eisenhut, 2019). Our results indicate its absence in *Gloeobacterales* (Fig. 2). MncA plays an important role in manganese binding within the periplasmic space (Totter *et al.*, 2008). However, its OG could not be determined, as we could not extract an OG that contained most cyanobacterial sequences with high sequence similarity (Fig. S9).

Discussion

To facilitate interpretation, the key results of this study are summarized in Fig. 9, which outlines a possible evolutionary scenario. The emergence of TM is necessarily linked to a biological innovation involving lipid environments (Rast *et al.*, 2015; Guéguen & Maréchal, 2022). VIPP1 has long been considered a key factor in TM formation, a hypothesis supported by Tan (2024), who demonstrated that VIPP1 originated from a duplication and subsequent C-terminal extension of the stress response protein PspA at the root of Phycobacteria (Tan, 2024). However, the functional role of VIPP1 remains ambiguous. In the two model cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002, VIPP1 does not appear to participate in TM biogenesis but rather in the assembly of photosynthetic complexes (Gao & Xu, 2009; Zhang *et al.*, 2014). This observation is consistent with the proteomic analysis of TM by Huang *et al.* (2023), which detected VIPP1 only during late stages of TM development

Structural differences in *Gloeobacterales*



homology to Tvp38, a putative lipid scramblase in yeast. Furthermore, Slr0305 is closely related to the *Arabidopsis thaliana* plastid protein AT5G19070.1, which is currently annotated as a SNARE-associated Golgi protein (Keller & Schneider, 2013). The phylogenetic distribution of these membranes trafficking proteins (emerging with *Thermotichales*), together with their reported function and membrane localization (known for Slr1106), is consistent with a role in TM emergence. This hypothesis contrasts with the CM-TM fusion model proposed for the ESCRT-III homolog VIPP1, whose central role in TM formation has been challenged by recent studies (Huang *et al.*, 2023; Ma *et al.*, 2025). However, our conclusions are based on comparative genomics analyses, and direct functional evidence is currently lacking for these candidate genes. Future experimental validation will be essential to clarify their precise roles and to determine whether they operate independently of, or inversely in coordination with, VIPP1.

The differences observed in late-stage PSII assembly factors are consistent with the respective membrane architecture of *Gloeobacterales* and Phycobacteria. Yet, evolutionary changes in assembly factors during cyanobacterial diversification, notably in photoprotection or PSII repair, could also have contributed to the patterns observed here, without necessarily being directly linked to TM emergence. On the one hand, the absence (or distant *Gloeobacterales* homologs) of assembly factors related to the late stages of PSII biogenesis, namely Psb27, Psb34, and Slr0606, may indicate an increased reliance on late-stage assembly steps during PSII repair in TM (Komenda *et al.*, 2024). On the other hand, the absence of TM facilitates the biogenesis of the LET components, as it occurs within a single cellular compartment, the CM. This is particularly clear in the case of Mn utilization. Mn is stored in the periplasmic space (Keren *et al.*, 2002). In *Gloeobacterales*, the Mn cluster of PSII is located at the periplasmic side, and therefore does not require Mn transport across membranes for PSII assembly, unlike in Phycobacteria. The absence of PrtA and PAM71 can thus be explained by the lack of TM. Interestingly, the SynPAM71 deletion mutant in *Synechocystis* sp. PCC6803 leads to a massive upregulation of the Mn-binding CyanoP, which suggests a role of CyanoP both in PSII assembly and in Mn homeostasis (Gandini *et al.*, 2017). Our results show that *Gloeobacterales* CyanoP exhibits a less negatively charged surface compared with its orthologs in Phycobacteria (Fig. 8), which would suggest a change in Mn homeostasis from *Gloeobacterales* to Phycobacteria. Nevertheless, the Mn-binding protein Mn cupin A (MncA), which is associated with Mn storage in the periplasmic space (Totter *et al.*, 2008), might also be expected in *Gloeobacterales*, although we were unable to confirm its occurrence. Little is known about the function of the V4R domain of Slr0144 and Slr0147 in *Synechocystis*. Their absence in the majority of *Gloeobacterales*, along with duplication events at the base of Phycobacteria and within *Thermotichales* (Fig. 6), underscores unresolved questions regarding their roles in PSII assembly as part of the photosynthetically related *slr0144–slr0152* operon (Wegener *et al.*, 2008).

No close homologs of PSII assembly factor proteins were found for most photosynthesis-specific proteins in the

nonphotosynthetic sister lineages of Cyanobacteria (Table S8), while all assembly factors involved in early stages of PSII biogenesis are already present in *Gloeobacterales*. Therefore, gene gain or loss cannot account for the relocation of PSII at the root of Phycobacteria. By contrast, the presence of additional helix loops in the predicted periplasmic domains of the phycobacterial YidC and SecD from *Thermotichales* might explain the differential localization of PSII between *Gloeobacterales* and Phycobacteria (Fig. 7). In *Escherichia coli*, YidC directly interacts with Sec components, particularly SecDF, which both contain periplasmic domains (Nouwen & Driessen, 2002; Xie *et al.*, 2006). Interestingly, our data indicate that SecD of Phycobacteria harbors two additional helix–loop insertions in its periplasmic domain (Fig. S141). By contrast, phycobacterial SecF differed only by its cytoplasmic domain in *Thermotichales*, whereas an additional helix–loop region was observed in *Synechocystis* sp. (Fig. S142). However, these structural interpretations are not empirical evidence of functional impact and remain hypothetical. At this stage, the interaction with YidC cannot be confirmed, as it has primarily been reported in *E. coli*, whose periplasmic domains differ from those found in cyanobacteria, thereby limiting direct structural comparison (Fig. S141C). Moreover, SecD localization within TM has not yet been established. Deletion of 90% of the periplasmic domain of *E. coli* YidC does not impair inner membrane protein biogenesis or cell viability (Xie *et al.*, 2006), and functional complementation experiments indicate that this domain is not strictly required for protein translocation, as Oxa1, despite lacking this domain, can complement YidC function in *E. coli* (Van Bloois *et al.*, 2005). Altogether, these observations suggest that the periplasmic domain of YidC has a yet-undefined role in the cell, despite its conservation across diderm bacteria (Oliver & Paetzel, 2008). Cyanobacteria possess only a single set of Sec genes that operates in both the CM and the TM, raising questions about how translocated proteins are selectively targeted to these two compartments (Nakai *et al.*, 1993; Srivastava *et al.*, 2005). By contrast, the YidC translocase is specifically localized to the PDM and TM. Since the luminal space of Phycobacteria is homologous to the periplasmic space of *Gloeobacterales* (Herrmann *et al.*, 2009), the additional helix loops and structural modifications observed in the periplasmic domain of phycobacterial YidC could potentially underlie changes in its functional properties and contribute to YidC's affinity for TM.

In the evolutionary scenario for the origin of TM, proposed by Cornet (2025), a series of intermediate stages allow transitioning from an electron transport chain located in the CM, that is the ancestral state still observed in *Gloeobacterales*, to the fully functional TM found in Phycobacteria (Cornet, 2025). This hypothesis is notably based on the high level of coordination among assembly factors, which reveals significant differences in the biogenesis of photosynthetic complexes, particularly between the two photosystems, suggesting distinct relocation processes (Cornet, 2025). In this framework, the first evolutionary stage corresponds to the formation of a membrane compartment capable of hosting relocated photosynthetic complexes. At the other end, the final stage corresponds to the establishment of the fully developed phycobacterial TM system as it is observed today, whereas

intermediate stages are hypothesized to involve alternative electron flux pathways. Our results on membrane trafficking proteins suggest an alternative to the VIPP1-based model for the first evolutionary stage. Indeed, they indicate that the emergence of compartmentalization may have involved proteins other than VIPP1. Alternatively, this first step may have required the prior establishment of a broader membrane context, including membrane-associated proteins and specific lipids, leading to the formation of early TM-like structures through vesiculation processes. Those structures could have supported simplified photosynthetic processes such as alternative electron flows, hence providing an evolutionary edge to these intermediates (Cornet, 2025). The observation that most PSII assembly factors are present in both *Gloeobacterales* and *Thermotrichales* suggests that they were also encoded in the last common ancestor of extant Cyanobacteria, consistent with a long evolutionary history of oxygenic photosynthesis predating the diversification of modern Cyanobacteria (Cardona, 2019; Oliver *et al.*, 2021). This implies that the establishment of fully functional TM, the final stage, was driven primarily by structural innovations, such as those observed in YidC and SecD, rather than by gene gain. Once established, TM could have led to increased oxygen release into the atmosphere and potentially contributing to the GOE, as already suggested by Guéguen & Maréchal (2022).

Our results on PSII, the photosynthetic complex with the best-characterized biogenesis, highlight the value of evolutionary analyses of assembly factors for understanding the emergence of TM and the relocation of photosynthetic complexes from the ancestral CM-localized state to modern TM. Extending this approach to other complexes (PSI, the cytochrome *b₆f*, ATP synthase, NDH-1) would provide a framework for elucidating intermediate evolutionary stages in TM emergence. Nevertheless, the biogenesis of these other complexes remains poorly characterized, with only a limited number of assembly factors identified to date (Rast *et al.*, 2015). This underscores the need to search for additional genes involved in the biogenesis of photosynthetic complexes if one wants to fully reconstruct the evolutionary emergence and diversification of the cyanobacterial photosynthetic apparatus.

Conclusion

Our comparative analyses of membrane trafficking proteins and PSII assembly factors highlight key molecular innovations associated with the emergence of TM in Phycobacteria. The acquisition of prohibitins and DedA-family proteins, such as Slr1106, possibly via LGT, and DedA-family proteins such as Slr0305 probably facilitated the organization of lipid microdomains and the development of membrane compartments, providing a structural basis for thylakoids. By contrast, *Gloeobacterales* retain a simplified system, lacking V4R-domain proteins, late-stage PSII assembly factors (Psb27, Psb34, Sll0606), and manganese transporters, consistent with PSII biogenesis occurring entirely within the CM. Early-stage assembly factors are surprisingly conserved, but structural modifications in periplasmic domains of YidC and SecD in Phycobacteria may contribute to the selective targeting

of photosynthetic complexes to TM. Overall, thylakoid evolution may have required gene acquisitions for compartment formation and structural innovations enabling PSII migration.

Acknowledgements

We thank Edi Sudianto for kindly providing an additional dataset of Cyanobacteriota genomes and the corresponding species tree. This work was supported by a research grant (PDR T.0018.24 OR-OX-PHOT-IN-CYN) from the Belgian National Fund for Scientific Research (F.R.S.-FNRS) to DB. LC is supported by a mandate from the Belgian National Fund for Scientific Research (F.R.S.-FNRS). LH is a FRIA grantee of the Fonds de la Recherche Scientifique – FNRS. Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the F.R.S.-FNRS under Grant No. 2.5020.11 and by the Walloon Region. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. ChatGPT (OpenAI) was used as a support tool for grammatical correction and text refinement.

Competing interests

None declared.

Author contributions

LH performed all analyses and drew figures. LH, DB and LC wrote the manuscript. All authors read and approved the final version of the manuscript.

ORCID

Denis Baurain  <https://orcid.org/0000-0003-2388-6185>

Luc Cornet  <https://orcid.org/0000-0002-3420-4488>

Louise Hambücker  <https://orcid.org/0009-0000-1861-4738>

Data availability

Multiple sequence alignments, phylogenetic trees, 3D structure predictions, and analysis scripts used for OrthoFinder analyses, orthologous group boxplots, phylogenetic trees, and ChimeraX are available on Figshare (doi: [10.6084/m9.figshare.31135354](https://doi.org/10.6084/m9.figshare.31135354)). Detailed datasets of selected genome accessions are provided in Table S4 (XLS format) and phylogenetic trees, including genome and protein accession numbers of orthologous proteins within Terrabacteria, are provided in Figs S51–S95, together with homologous sequences within Cyanobacteriota in phylogenetic trees Figs S96–S136. The protein accession numbers supporting the findings of this study are, for CyanoP: BAC89381.1 (*Gloeobacter violaceus* PCC7421 GCA 000011385.1),

MEN9209059.1 (*Thermotichales cyanobacterium* GCA_039794945.1), and the 3D PDB structure 2XB3 (*Thermosynechococcus vestitus* BP-1); and for YidC: BAC90172.1, MEN9209976.1 (*Thermotichales cyanobacterium* GCA_039794945.1), P74155.1 (*Synechocystis* sp. PCC 6803 substr. Kazusa). All other supporting data, including the Supporting Information figures file (containing phylogenetic trees, orthologous group selection and the ribosomal species tree) and the Supporting Information tables file (containing formatted tables and descriptions), are available in the Supporting Information.

References

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J *et al.* 2024. Accurate structure prediction of biomolecular interactions with ALPHAFOLD 3. *Nature* **630**: 493–500.
- Anantharaman V, Koonin EV, Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *Journal of Molecular Biology* **307**: 1271–1292.
- 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.
- Browman DT, Hoegg MB, Robbins SM. 2007. The SPFH domain-containing proteins: more than lipid raft markers. *Trends in Cell Biology* **17**: 394–402.
- Bryan SJ, Burroughs NJ, Evered C, Sacharz J, Nenninger A, Mullineaux CW, Spence EM *et al.* 2011. Loss of the SPFH homologue Slr1768 leads to a catastrophic failure in the maintenance of thylakoid membranes in *Synechocystis* sp. PCC 6803. *PLoS ONE* **6**: e19625.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Cardona T. 2019. Thinking twice about the evolution of photosynthesis. *Open Biology* **9**: 180246.
- Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CHECKM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods* **20**: 1203–1212.
- Cormann KU, Bartsch M, Rägner M, Nowaczyk MM. 2014. Localization of the CyanoP binding site on photosystem II by surface plasmon resonance spectroscopy. *Frontiers in Plant Science* **5**: 1459.
- Cornet L. 2025. On the non-oxygenic origins of thylakoids. *Communications Biology* **8**: 1697.
- Cornet L, Durieu B, Baert F, D'hooge E, Colignon D, Meunier L, Lupo V, Cleenwerck I, Daniel HM, Rigouts L *et al.* 2023. The GEN-ERA toolbox: unified and reproducible workflows for research in microbial genomics. *GigaScience* **12**: 587.
- Eisenhut M. 2019. Manganese homeostasis in cyanobacteria. *Plants* **9**: 18.
- Emms DM, Kelly S. 2019. ORTHOFINDER: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: W29–W37.
- Frain KM, Gangl D, Jones A, Zedler JAZ, Robinson C. 2016. Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics* **1857**: 266–273.
- Gandini C, Schmidt SB, Husted S, Schneider A, Leister D. 2017. The transporter Syn PAM 71 is located in the plasma membrane and thylakoids, and mediates manganese tolerance in *Synechocystis* sp. PCC 6803. *New Phytologist* **215**: 256–268.
- Gao H, Xu X. 2009. Depletion of Vipp1 in *Synechocystis* sp. PCC 6803 affects photosynthetic activity before the loss of thylakoid membranes. *FEMS Microbiology Letters* **292**: 63–70.
- Gathmann S, Rupprecht E, Kahmann U, Schneider D. 2008. A conserved structure and function of the YidC homologous protein Slr1471 from *Synechocystis* sp. PCC 6803. *Journal of Microbiology and Biotechnology* **18**: 1090–1094.
- Guéguen N, Maréchal E. 2022. Origin of cyanobacterial thylakoids via a non-vesicular glycolipid phase transition and their impact on the great oxygenation event. *Journal of Experimental Botany* **73**: 2721–2734.
- Heidrich J, Thurotte A, Schneider D. 2017. Specific interaction of IM30/Vipp1 with cyanobacterial and chloroplast membranes results in membrane remodeling and eventually in membrane fusion. *Biochimica Et Biophysica Acta (BBA) - Biomembranes* **1859**: 537–549.
- Heinz S, Liauw P, Nickelsen J, Nowaczyk M. 2016. Analysis of photosystem II biogenesis in cyanobacteria. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics* **1857**: 274–287.
- Hennon SW, Soman R, Zhu L, Dalbey RE. 2015. YidC/Alb3/Oxa1 family of insertases. *Journal of Biological Chemistry* **290**: 14866–14874.
- Herrmann JM, Kauff F, Neuhaus HE. 2009. Thiol oxidation in bacteria, mitochondria and chloroplasts: common principles but three unrelated machineries? *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research* **1793**: 71–77.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.
- Huang F, Grauslys A, Huokko T, Caamaño Gutiérrez E, Jones AR, Liu LN. 2023. Dynamic changes in the thylakoid proteome of cyanobacteria during light-regulated thylakoid membrane development. *Plants* **12**: 3967.
- Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. INTERPROSCAN 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
- Juruss E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K *et al.* 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27**: 112–128.
- Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.
- Keller R, Schneider D. 2013. Homologs of the yeast Tvp38 vesicle-associated protein are conserved in chloroplasts and cyanobacteria. *Frontiers in Plant Science* **4**: 2587.
- Keren N, Kidd MJ, Penner-Hahn JE, Pakrasi HB. 2002. A light-dependent mechanism for massive accumulation of manganese in the photosynthetic bacterium *Synechocystis* sp. PCC 6803. *Biochemistry* **41**: 15085–15092.
- Klinkert B, Ossenbühl F, Sikorski M, Berry S, Eichacker L, Nickelsen J. 2004. Prata, a periplasmic tetratricopeptide repeat protein involved in biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. *Journal of Biological Chemistry* **279**: 44639–44644.
- Komenda J, Knoppová J, Kopečná J, Sobotka R, Halada P, Yu J, Nickelsen J, Boehm M, Nixon PJ. 2012. The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiology* **158**: 476–486.
- Komenda J, Sobotka R, Nixon PJ. 2024. The biogenesis and maintenance of PSII: recent advances and current challenges. *Plant Cell* **36**: 3997–4013.
- Léonard RR, Leleu M, Van Vlierberghe M, Kerff F, Baurain D. 2020. ToRQuEMaDA: tool for retrieving queried eubacteria. *Metadata and Dereplicating Assemblies* **25**: 3259.
- Letunic I, Bork P. 2024. Interactive tree of life (iTOL) v.6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**: W78–W82.
- Ma L, Dong B, Sun M, Hao R, Wang X, Yu H, Han C, Muhire A, Gachie SW, Li D *et al.* 2025. VESICLE-INDUCING PROTEIN IN PLASTIDS 1 from thylakoid-lacking *Gloeobacter* promotes thylakoid formation in Arabidopsis. *Plant Physiology* **199**: kiaz359.
- Markin A, Wagle S, Grover S, Vincent Baker AL, Eulenstein O, Anderson TK. 2023. PARNAS: objectively selecting the most representative taxa on a phylogeny. *Systematic Biology* **72**: 1052–1063.
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF CHIMERAx: tools for structure building and analysis. *Protein Science* **32**: e4792.

- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Nakai M, Sugita D, Omata T, Endo T. 1993. Sec-Y protein is localized in both the cytoplasmic and thylakoid membranes in the cyanobacterium *Synechococcus* PCC9742. *Biochemical and Biophysical Research Communications* 193: 228–234.
- Nouwen N, Driessen AJM. 2002. SecDFyajC forms a heterotetrameric complex with YidC. *Molecular Microbiology* 44: 1397–1405.
- Oliver DC, Paetzel M. 2008. Crystal structure of the major periplasmic domain of the bacterial membrane protein assembly facilitator YidC. *Journal of Biological Chemistry* 283: 5208–5216.
- Oliver T, Sánchez-Baracaldo P, Larkum AW, Rutherford AW, Cardona T. 2021. Time-resolved comparative molecular evolution of oxygenic photosynthesis. *Biochimica Et Biophysica Acta (BBA) – Bioenergetics* 1862: 148400.
- Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, Schmidt TSB, Bork P. 2021. GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biology* 22: 178.
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* 50: D785–D794.
- Rahimzadeh-Karvansara P, Pascual-Aznar G, Bečková M, Komenda J. 2022. Psb34 protein modulates binding of high-light-inducible proteins to CP47-containing photosystem II assembly intermediates in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynthesis Research* 152: 333–346.
- Rahmatpour N, Hauser DA, Nelson JM, Chen PY, Villarreal JC, Ho MY, Li FW. 2021. A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. *Current Biology* 31: 2857–2867.
- Rast A, Heinz S, Nickelsen J. 2015. Biogenesis of thylakoid membranes. *Biochimica Et Biophysica Acta (BBA) – Bioenergetics* 1847: 821–830.
- Rast A, Schaffer M, Albert S, Wan W, Pfeffer S, Beck F, Plitzko JM, Nickelsen J, Engel BD. 2019. Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nature Plants* 5: 436–446.
- Rengstl B, Oster U, Stengel A, Nickelsen J. 2011. An intermediate membrane subfraction in cyanobacteria is involved in an assembly network for photosystem II biogenesis. *Journal of Biological Chemistry* 286: 21944–21951.
- Rexroth S, Mullineaux CW, Ellinger D, Sendtko E, Rögner M, Koenig F. 2011. The plasma membrane of the cyanobacterium *Gloeobacter violaceus* contains segregated bioenergetic domains. *Plant Cell* 23: 2379–2390.
- Rippka R, Waterbury J, Cohen-Bazire G. 1974. A cyanobacterium which lacks thylakoids. *Archives of Microbiology* 100: 419–436.
- Siebenaller C, Schneider D. 2023. Cyanobacterial membrane dynamics in the light of eukaryotic principles. *Bioscience Reports* 43: BSR20221269.
- Singh AK, Li H, Sherman LA. 2004. Microarray analysis and redox control of gene expression in the cyanobacterium *Synechocystis* sp. PCC 6803. *Physiologia Plantarum* 120: 27–35.
- Srivastava R, Pisareva T, Norling B. 2005. Proteomic studies of the thylakoid membrane of *Synechocystis* sp. PCC 6803. *Proteomics* 5: 4905–4916.
- Stengel A, Gügel IL, Hilger D, Rengstl B, Jung H, Nickelsen J. 2012. Initial steps of photosystem II *de novo* assembly and preloading with manganese take place in biogenesis centers in *Synechocystis* sp. *Plant Cell* 24: 660–675.
- Takekawa N, Isumi M, Terashima H, Zhu S, Nishino Y, Sakuma M, Kojima S, Homma M, Imada K. 2019. Structure of *vibrio* FlIL, a new stomatin-like protein that assists the bacterial flagellar motor function. *MBio* 10: e00292.
- Tan S. 2024. Exploring the origins and evolution of oxygenic and anoxygenic photosynthesis in deeply branched Cyanobacteriota. *Molecular Biology and Evolution* 41: 1584.
- Totter S, Waldron KJ, Firbank SJ, Reale B, Bessant C, Sato K, Cheek TR, Gray J, Banfield MJ, Dennison C *et al.* 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* 455: 1138–1142.
- Van Bloois E, Nagamori S, Koningstein G, Ullers RS, Preuss M, Oudega B, Harms N, Kaback HR, Herrmann JM, Lührink J. 2005. The sec-independent function of *Escherichia coli* YidC is evolutionary-conserved and essential. *Journal of Biological Chemistry* 280: 12996–13003.
- Van De Meene AML, Hohmann-Marriott MF, Vermaas WJF, Roberson RW. 2006. The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803. *Archives of Microbiology* 184: 259–270.
- Wegener KM, Welsh EA, Thornton LE, Keren N, Jacobs JM, Hixson KK, Monroe ME, Camp DG II, Smith RD, Pakrasi HB. 2008. High sensitivity proteomics assisted discovery of a novel operon involved in the assembly of photosystem II, a membrane protein complex. *Journal of Biological Chemistry* 283: 27829–27837.
- Westphal S, Heins L, Soll J, Vothknecht UC. 2001. *Vipp1* deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? *Proceedings of the National Academy of Sciences, USA* 98: 4243–4248.
- Wong TKF, Ly-Trong N, Ren H, Demotte P, Baños H, Roger AJ, Susko E, Bielow C, De Maio N, Goldman N *et al.* 2025. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Molecular Biology and Evolution*. doi: <https://doi.org/10.1093/molbev/msg117>.
- Xie K, Kiefer D, Nagler G, Dalbey RE, Kuhn A. 2006. Different regions of the nonconserved large periplasmic domain of *Escherichia coli* YidC are involved in the SecF interaction and membrane insertase activity. *Biochemistry* 45: 13401–13408.
- Yang H, Liao L, Bo T, Zhao L, Sun X, Lu X, Norling B, Huang F. 2014. Slr0151 in *Synechocystis* sp. PCC 6803 is required for efficient repair of photosystem II under high-light condition. *JIPB* 56: 1136–1150.
- Zhang S, Frankel LK, Bricker TM. 2010. The Sll0606 protein is required for photosystem II assembly/stability in the cyanobacterium *Synechocystis* sp. PCC 6803. *Journal of Biological Chemistry* 285: 32047–32054.
- Zhang S, Shen G, Li Z, Golbeck JH, Bryant DA. 2014. *Vipp1* is essential for the biogenesis of photosystem I but not thylakoid membranes in *Synechococcus* sp. PCC 7002. *Journal of Biological Chemistry* 289: 15904–15914.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Phylogenomic tree based on ribosomal proteins.

Fig. S2 Genome completeness by Terrabacteria phylum using CheckM2.

Fig. S3 Genome contamination by Terrabacteria phylum using CheckM2.

Fig. S4 Diversity curve generated using PARNAS v.0.1.6 for the pruning of the ribosomal tree.

Figs. S5–S50 Boxplots based on the bitscore distributions of BLASTP results.

Figs. S51–S95 Gene trees of each OG of interest.

Figs. S96–S136 Gene trees based on the best hits of a HMM search on the Cyanobacterial dataset (343 Cyanobacteriota species).

Fig. S137 Additional β -sheet positioned outside of the ring structure formed during oligomerization in the proteins of the basal group I.

Fig. S138 Interpro predictions of functional domains within the closest homologs of Sll0606 in Gloeobacterales.

Fig. S139 Alignment of the cyanobacterial Slr0144 and Slr0147 sequences.

Fig. S140 Alignment of the cyanobacterial YidC sequences.

Fig. S141 AlphaFold3 predicted 3D structures of SecD from *Gloeobacter violaceus*, *Thermotrichales cyanobacterium* GCA_039794545.1, *Synechocystis* sp. PCC 6803.

Fig. S142 AlphaFold3 predicted 3D structures of SecF from *Gloeobacter violaceus*, *Thermotrichales cyanobacterium* GCA_039794545.1, and *Synechocystis* sp. PCC 6803.

Fig. S143 Alignment of the cyanobacterial CyanoP sequences.

Fig. S144 Visualization of the selection of the sequences similar to Slr1106 with Ompa-Pa.

Fig. S145 Visualization of the selection of the sequences similar to Slr1768 with Ompa-Pa.

Fig. S146 Preliminary multigenic family tree of Slr1106.

Fig. S147 Preliminary multigenic family tree of Slr1768.

Table S1 Manual selection of the Terrabacteria representative species recovered from GTDB r220 Parks *et al.* (2022).

Table S2 Selection of the 343 genomes of Cyanobacteriota (Cyanobacterial dataset).

Table S3 Sources of proteomes for the manually-selected accessions.

Table S4 Selection of the 950 genomes based on a manual filtering and ribosomal tree pruning using PARNAS.

Table S5 Description of selected assembly factors for this study.

Table S6 Description of selected proteins involved in cyanobacterial membrane dynamics.

Table S7 Summary of the BLASTP search and the phylogenetic trees.

Table S8 Identification of homologs in closely related taxa.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.