# From Cytoplasmic Membrane to Thylakoids: Evolution of Membrane Biogenesis and Photosystem II assembly in early-diverging Cyanobacteria

Louise Hambücken[1], Denis Baurain[1], Luc Cornet [1*]

1 InBioS–PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, 4000 Liège, Belgium

* Author to whom correspondence should be addressed.
luc.cornet@uliege.be

# Abstract

Thylakoid membranes (TM) in cyanobacteria and chloroplasts host the light-dependent reactions of oxygenic photosynthesis which involve a linear electron transfer (LET) chain composed of multi-subunit complexes, including Photosystem II (PSII). Gloeobacterales, the earliest-diverging cyanobacterial lineage, lack TM and perform photosynthesis within specialized regions of the cytoplasmic membrane (CM), thereby representing an ancestral state with respect to other cyanobacteria, all equipped with TM and termed Phycobacteria. The emergence of TM, which increased the membrane surface available for oxygenic photosynthesis, was a key innovation that likely contributed to the Great Oxidation Event. This evolutionary transition involved the formation of a distinct membrane compartment, followed by the relocation of LET components from the CM to TM. Here, we present a phylogenomic analysis identifying three candidate proteins associated with membrane trafficking that may contribute to TM biogenesis, including the SPFH family member Slr1106, which we show was acquired via lateral gene transfer from Patescibacteria. Moreover, evolutionary analysis of 37 PSII assembly factors reveals key modifications in late-stage PSII assembly, particularly in manganese homeostasis, and highlights structural changes in the YidC translocase , which may have facilitated the relocation of LET components from the CM to TM. Altogether, these results bring new insights into the molecular innovations that led to the emergence of TM.

# Main

## 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 $CO_2$ fixation in the Calvin–Benson cycle, through a linear electron transport chain (LET) composed of multi-subunit complexes, among which Photosystem I (PSI), Photosystem II (PSII), and 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 triggered the Great Oxidation Event (GOE)[1] approximately 2.4 billion years ago, profoundly transforming Earth's biosphere[2].

Among extant cyanobacteria, members of the order Gloeobacterales, the earliest-diverging lineage, lack thylakoid membranes (TM)[3,4]. Instead, they carry out oxygenic photosynthesis within specialized regions of their cytoplasmic membrane (CM)[5]**.** The absence of TM in the basal group suggests that these structures arose after the divergence of *Gloeobacterales* but before the emergence of *Thermostichales*, the second earliest-diverging cyanobacterial lineage, which already possesses TM.

Consequently, *Gloeobacterales* represent the ancestral evolutionary state preceding TM formation and provide a unique model for exploring the transition from a LET located in the CM to one integrated within TM [6]. The evolution of TM likely occurred through a two-step process[6]on the. First, the formation of a distinct intracellular lipidic compartment, and second, the relocation of LET components from the CM to this newly formed structure. The proteins mediating this latter transition, namely the assembly factors, may retain evolutionary signatures of this ancestral relocation, offering valuable insights into how cyanobacteria with TM, termed Phycobacteria, emerged[6].

The mechanisms by which the lipid composition of TM is established 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[1,7].VIPP1, a protein from the ESCRT-III family, was suggested to play an essential role in the biosynthesis of TM, thus supporting the first scenario[8–10]. However, a recent study showed that the *Gloeobacterales* VIPP1 could complement a *vipp1* mutant in *Arabidopsis*[11]. Furthermore, proteomic data from TM biogenesis indicate that VIPP1 accumulates at later stages of TM formation[12]. 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 37 proteins known as assembly factors[13–15]. 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 chlorophyll incorporation and D1 maturation. The D1–D2 heterodimer then forms within a specialized region of the CM enriched in assembly factors, known as the PratA-defined membrane (PDM)[16]. 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[17].These later steps likely occur within developing TM.

Here, we present a phylogenomic analysis of ten proteins involved in membrane trafficking that identify two members 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 present the first evolutionary analysis of the 37 PSII assembly factors known so far, revealing important modifications in the later stages of PSII assembly, notably in manganese homeostasis, after the divergence of *Gloeobacterales*. In parallel, we pinpoint structural changes in the YidC machinery that may have facilitated the relocation of LET composoant from CM to TM.

# Materials and methods

## Genome selection and phylogenomic species tree

Initially, the 45,316 representative species genomes of Terrabacteria were retrieved from the GTDB release r220 [18]. To limit computational burden during Orthologous gene inference, a subset of 950 genomes was eventually selected based on genome quality and phylogenetic position. Briefly, only the 43827 genomes that passed the GUNC contamination filter were retained. Within each GTDB order, genomes were then ranked by the difference between their CheckM2[19] 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 was selected, yielding 1,452 genomes in total without the Cyanobacteria phylum. Sampling within the Cyanobacteria phylum was more exhaustive, resulting in 432 genomes, i.e., one or two representative genomes were chosen per genus, prioritizing those with >80% completeness and <10% contamination, as estimated by CheckM2[20]. To ensure coverage of early-diverging cyanobacterial lineages, all available genomes from the orders *Gloeobacterales*, *Thermostichales* and *Gloeomargaritales* were incorporated, whereas the more abundant *Pseudanabaenales* were dereplicated using ToRQuEMaDA v0.2.1 with default parameters[21].

Protein-coding sequences were obtained directly from NCBI proteomes when available or predicted using Prodigal v2.6.3[22]. A phylogenomic tree based on ribosomal proteins was reconstructed using the GEN-ERA toolbox v3[23] and IQ-TREE v2.4.0[24] with the best-fit substitution model and 1,000 ultrafast bootstrap replicates. The resulting phylogeny was pruned using PARNAS v0.1.6[25] to retain a final set of 950 representative species, encompassing approximately 90% of total phylogenetic diversity based on diversity score metrics.

## Orthologous group inference

The orthologous groups were inferred with OrthoFinder v2.5.5[26] from the 950 genomes selected. An approach based on Hierarchical Orthologous Groups (HOGs) was first explored to account for multigene families stemming from duplications. However, this strategy, described in Supplementary File 1, proved to be less accurate than the one based on standard Orthologous Groups (OGs), detailed below.

## Identification of the orthologous groups of interest

The OGs corresponding to known assembly factors were identified by performing BLASTp searches (v2.9.0+)[27] of every protein from the 950 proteomes against a database containing reference sequences of known PSII assembly factors and proteins involved in membrane dynamics (**Table S1**). 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. For each assembly factor, the OG that contained the highest mean bitscore for cyanobacterial hits wa then selected. To check for optimal OG delineation, a reciprocal BLASTp of every reference sequence of assembly factors 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 alignment coverage 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 v3.0.0[28] with the best-fit substitution model and 1,000 ultrafast bootstrap replicates. The sequences genuinely corresponding to the assembly factors were then pinpointed in the gene trees based on the bitscores. When subsequent analyses were needed, either for taxonomic distribution determination or tree topology resolution, a HMM profile (pHMM) was built with the cyanobacterial sequences and searched against a database comprising the inferred proteomes of 343 Cyanobacteriota species (hereafter named the cyanobacterial dataset) using HMMER v3.4[29]. The best hits were then filtered based on their e-value, sequence length, number of copies per organism (the three best hits per organisms were kept) and alignment coverage (minimum 0.7) using OmpaPa. The selected sequences were then aligned using MAFFT v7.505[30]. Alignment sites were conserved if they had at most 30% gaps and the sequences with a minimum of 30% length (with respect to the longest unaligned sequence) were exported using ali2phylip.pl (https://metacpan.org/release/Bio-MUST-Core;D. Baurain). Finally, a gene tree was inferred from the filtered alignment using IQ-TREE v3.0.0[28] with the best-fit substitution model and 1,000 ultrafast bootstrap replicates.

## Functional protein and 3D protein structure prediction

Functional protein domains were annotated with InterPro v5.48-83.0[31]. When the experimental structures were not available, the 3D protein structures were modelled using the AlphaFold 3 server [32]. The structures were then visualized and aligned using PyMOL v2.5.5[33].

# Results

## Membrane trafficking proteins

SPFH proteins (implicated in vesicular lipid transport), DedA family proteins (putative lipid scramblases), and dynamin-like proteins (involved in membrane fission and fusion) have been proposed by Siebenaller and Schneider (2023) as potential contributors to lipid trafficking in Cyanobacteria. To further explore their evolutionary history among Terrabacteria, we conducted a separate phylogenetic analysis for ten of these proteins (**Table 1**). Their taxonomic distribution revealed that two are absent from the *Gloeobacterales* but present in the Phycobacteria: Slr1106 (an SPFH protein) and Slr0305 (DedA family). The prohibitin Slr1106 is a member of the PAM71 family, whose phylogeny within Terrabacteria (Figure 1), which also includes another SPFH protein, Slr1768, suggests an origin through lateral gene transfer from Patescibacteria to Cyanobacteria. Slr1106 appears to have been transferred at the base of the Phycobacteria, whereas Slr1768 was acquired later, prior to the emergence of the *Pseudanabaenales* but after that of the *Thermostichales* (**Table 1**). Our analyses showed that the tree of Slr1106 and Slr1768 also contains two groups of basal Cyanobacteria, including *Gloeobacterales*, which lack obvious orthologues in other cyanobacterial orders (**Figure 1**). Moreover, structural predictions revealed an additional β-sheet positioned outside of the ring structure formed during oligomerization in the proteins of these basal group I (**Figure S1**). Regarding the DedA protein Slr0305, it was not detected in *Gloeobacterales* but is present from the root of Phycobacteria (**Figure 2**), while no apparent difference in sequence or structure could be detected. Another DedA protein, Slr0232, is for its part conserved across all cyanobacteria. However, its phylogeny exhibits an conspicuously long branch at the base of Phycobacteria (**Figure 3**), albeit again without discernible structural or sequence divergence.

## Assembly factors of PSII

### Late-stage assembly factors

The largest differences between *Gloeobacterales* and Phycobacteria were observed at the level of late-stages assembly factors. The assembly factors Psb27 and Psb34 were not identified in *Gloeobacterales* in our analyses (**Table 1**). Psb34 is thought to bind CP47 and compete with HliA/B to enable the final stage of PSII assembly[34], while Psb27 binds both unassembled and assembled CP43 in the PSII complex[35]. Sll0606, a protein of unknown function but leading to a massive reduction of D1 and CP43 in PSII assembly upon deletion[36], was also not detected in *Gloeobacterales* (**Table 1**). Domain analysis of Sll0606 suggests a potential role in the transport of hydrophobic components, and distant homologues in *Gloeobacterales* have been proposed[36]. In this work, we also identified such a homolog in *Gloeobacterales*: its sequence is not orthologous to the *Synechocystis* PCC6803 reference sequence, but InterPro domain analysis indeed annotates three domains capable of binding hydrophobic components at the N-terminal region **(Figure S2).**

*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 chlorophyll precursors in *Rhodobacter capsulatus*[37,38]. In *Synechocystis* sp. PCC6803, Slr0144 is encoded by the slr0144 gene within the slr0144–slr0152 operon[38], which is involved in PSII stabilization[39]. Slr0144 and Slr1471, the only two proteins in *Synechocystis* sp. PCC6803 containing a V4R domain, are the result of a gene duplication event (**Figure 4**). Sequence-wise, Slr0144 and Slr1471 differ notably by Slr0144 featuring an N-terminal extension, along with changes in its C-terminal region (**Figure S3**).

Manganese is bound to D1 and incorporated during the late stages of PSII assembly, once PSII is already in TM. PratA, a periplasmic tetratricopeptide repeat (TPR) protein strongly associated with the PDM region [16], has been shown to facilitate efficient manganese delivery to pD1[40]. Our taxonomic distribution analyses showed the absence of PratA in *Gloeobacterales* (**Table 1**). CyanoP, the

cyanobacterial homolog of the PsbP protein known in land plants, is a manganese-binding protein that plays a role either in the binding association of CP47 and CP43 or in the photoactivation of the oxygen-evolving complex[41]. While our analyses show the presence of CyanoP in *Gloeobacterales*, its sequence has diverged substantially compared to Phycobacteria (**Figure S4**). APBS electrostatics calculations suggest that *Gloeobacterales* CyanoP either binds divalent ions differently compared to Phycobacteria or does not bind ions at all (**Figure 5**). PAM71 is a manganese transporter involved in its homeostasis, storing excess manganese both in the periplasmic and lumenal spaces[42]. Its role in manganese uptake is considered to support PSII oxygen-evolving complex performance, suggesting a PratA-independent manganese delivery pathway [43]. Our result indicates its absence in *Gloeobacterales* (**Table 1**). MncA plays an important role in manganese binding within the periplasmic space[44]. However, its OG could not be determined, as we could not extract an OG that contained most cyanobacterial sequences with high sequence similarity.

**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* (**Table 1**). However, a structural difference exists in the translocase YidC between *Gloeobacterales* and Phycobacteria, with the latter harboring a specific P1 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[45] 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 analogous to the one described in chloroplasts[46]. The periplasmic domain of YidC plays is only present in diderm bacteria, such as cyanobacteria, as its size is markedly reduced in other bacterial homologs[45].

Interestingly, our results indicate that the periplasmic domain of cyanobacterial YidC is largely conserved across the cyanobacterial phylum, except in *Gloeobacterales.* Sequence alignment reveals well-conserved N-terminal and C-terminal regions; however, the periplasmic domain of *Gloeobacterales* YidC differs, with missing amino acids in the sequence (**Figure S5**). 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 *Gloeobacterales* homolog (**Figure 6**).

# Discussion

The emergence of TM is necessarily linked to a biological innovation involving lipid environments[1,7]. VIPP1 has long been considered a key factor in TM formation, a hypothesis supported by Tan *et al.* (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[9]. 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[47,48]. 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[12]. Moreover, a recent study identified a VIPP1 homolog in *Gloeobacterales* capable of complementing a *vipp1* mutant in *Arabidopsis*[11]. Although its function is still undefined, these results collectively indicate that VIPP1 is unlikely to be directly involved in TM formation.

In contrast, the prohibitin Slr1106, acquired through lateral gene transfer (LGT) from Patescibacteria to the ancestor of Phycobacteria, may provide a plausible explanation for the emergence of TM. Patescibacteria being characterized by reduced genomes and limited metabolic capabilities, they are generally regarded as parasitic or symbiotic bacteria[49,50]. Slr1106 belongs to the prohitin family, within the SPFH protein superfamily, typically associated with the inner mitochondrial membrane or the nucleus in eukaryotic cells, where they contribute to structural stabilization, apoptosis regulation, and cell signaling[51]. Members of this family are characterized by four α-helices and a twisted β-sheet structure, and they assemble into large oligomeric complexes[52]. Slr1106 is localized in both CM and TM[53], consistent with a potential role in membrane organization and the formation of lipid microdomains. Cornet (2025) hypothesized that the complex stepwise relocation of the LET components from CM to TM involved intermediate evolutionary stages between *Gloeobacterales* and Phycobacteria. One possible scenario is that an ancestral stage was associated with a Patescibacterium, which subsequently led to the lateral gene transfer of Slr1106[6]. The OG corresponding to the prohibitin Slr1106 includes proteins that are restricted to basal cyanobacteria, including *Gloeobacterales*, suggesting that these ancestral variants may have participated in the formation of primitive photosynthetic regions in CM, as observed in *Gloeobacterales*[54]. In addition, our analyses identified two other proteins involved in membrane trafficking that could have played a role in TM emergence: Slr0305, acquired at the root of the Phycobacteria, and Slr0232, which exhibits a unusually long branch in its phylogeny between *Gloeobacterales* and Phycobacteria.

The differences observed in the late stages of PSII biogenesis are consistent with the membrane architecture of *Gloeobacterales*. Indeed, 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 manganese utilization. Manganese is stored in the periplasmic space[55] and therefore *Gloeobacterales* does not require transport across membranes, unlike Phycobacteria. The absence of PratA and PAM71 can thus be explained by the lack of TM. Nevertheless, one would expect Mn-binding protein Mn cupin A (MncA), which is associated with manganese storage in the periplasmic space[44], to be found in *Gloeobacterales* as well, although we were unable to confirm its occurrence.

The presence of all assembly factors involved in the early stages of PSII biogenesis is somewhat unexpected, as one might have anticipated the emergence of new assembly factors at the root of Phycobacteria. However, the presence of two additional helix loops in the periplasmic space of the phycobacterial YidC might explain the differential localization of PSII between *Gloeobacterales* and Phycobacteria. In *Escherichia coli*, YidC directly interacts with Sec components, particularly SecDF, which contain periplasmic domains[56,57]. Nevertheless, deletion of 90% of the periplasmic domain of *E. coli* YidC does not impair inner membrane protein biogenesis or cell viability[57], and functional complementation experiments indicate that this domain is not strictly required for protein translocation, as Oxa1 can complement YidC function in *E. coli*[58]. Altogether, these observations suggest that the periplasmic domain of YidC has a yet-undefined role in the cell, despite its conservation across diderm bacteria[59]. Cyanobacteria possess only a single set of Sec genes, operating in both CM and TM, thus raising questions about how proteins are selectively targeted to these two compartments[60,61]. In contrast, the YidC translocase is specifically localized to the PDM and TM. Since the lumenal space of Phycobacteria is homologous to the periplasmic space of *Gloeobacterales* [62]. the additional helix loops and structural modifications observed in the periplasmic domain of Phycobacterial YidC could potentially underlie its membrane modulation properties**.** This represents the only structural difference identified between *Gloeobacterales* and Phycobacteria early stage assembly factors, and given that the biogenesis of PSII is well characterized, it is plausible that this modification could contribute to YidC's affinity for TM.

# Conclusion

Our evolutionary analysis of proteins involved in membrane trafficking revealed three candidate proteins potentially contributing to the emergence of TM, including Slr1106, which is localized to both the CM and TM and was acquired via a lateral gene transfer from Patescibacteria to the root of

Phycobacteria. Moreover, our study of PSII assembly factors highlighted the absence of genes associated with manganese homeostasis during PSII biogenesis. Finally, we report a structural modification—the only notable change among the early-stage PSII assembly factors—consisting of two additional helix loops in the periplasmic space of the phycobacterial version of the YidC translocase, potentially enabling the differential targeting of PSII between *Gloeobacterales* and Phycobacteria.

# 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.

# Acknowledgments

We thank Edi Seduanto for kindly providing an additional dataset of Cyanobacteriota genomes. 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).

# Conflict of Interest

The authors declare no competing interest.

# References


1. Guéguen, N. & Maréchal, E. 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 (2022).
2. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
3. Rahmatpour, N. *et al.* A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. *Current Biology* **31**, 2857-2867.e4 (2021).
4. Rippka, R., Waterbury, J. & Cohen-Bazire, G. A cyanobacterium which lacks thylakoids. *Arch. Microbiol.* **100**, 419–436 (1974).
5. Rexroth, S. *et al.* Functional Characterization of the Small Regulatory Subunit PetP from the Cytochrome $b_6f$ Complex in *Thermosynechococcus elongatus*. *Plant Cell* **26**, 3435–3448 (2014).
6. Luc Cornet. On the Non-Oxygenic Origins of Thylakoids. *Communications Biologu (in press)* (2025).
7. Rast, A., Heinz, S. & Nickelsen, J. Biogenesis of thylakoid membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1847**, 821–830 (2015).
8. Heidrich, J., Thurotte, A. & Schneider, D. 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 (2017).
9. Tan, S. Exploring the Origins and Evolution of Oxygenic and Anoxygenic Photosynthesis in Deeply Branched Cyanobacteriota. *Molecular Biology and Evolution* **41**, (2024).
10. Westphal, S., Heins, L., Soll, J. & Vothknecht, U. C. *Vipp1* deletion mutant of *Synechocystis* : A connection between bacterial phage shock and thylakoid biogenesis? *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4243–4248 (2001).
11. Ma, L. *et al.* VESICLE-INDUCING PROTEIN IN PLASTIDS 1 from thylakoid-lacking Gloeobacter promotes thylakoid formation in Arabidopsis. *Plant Physiology* **199**, kiaf359 (2025).
12. Huang, F. *et al.* Dynamic Changes in the Thylakoid Proteome of Cyanobacteria during Light-Regulated Thylakoid Membrane Development. *Plants* **12**, 3967 (2023).
13. Heinz, S., Liauw, P., Nickelsen, J. & Nowaczyk, M. Analysis of photosystem II biogenesis in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1857**, 274–287 (2016).
14. Rast, A. *et al.* Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nat. Plants* **5**, 436–446 (2019).
15. Van De Meene, A. M. L., Hohmann-Marriott, M. F., Vermaas, W. F. J. & Roberson, R. W. The three-dimensional structure of the cyanobacterium Synechocystis sp. PCC 6803. *Arch Microbiol* **184**, 259–270 (2006).

16. Stengel, A. *et al.* Initial Steps of Photosystem II de Novo Assembly and Preloading with Manganese Take Place in Biogenesis Centers in *Synechocystis*. *Plant Cell* **24**, 660–675 (2012).
17. Rengstl, B., Oster, U., Stengel, A. & Nickelsen, J. An Intermediate Membrane Subfraction in Cyanobacteria Is Involved in an Assembly Network for Photosystem II Biogenesis. *Journal of Biological Chemistry* **286**, 21944–21951 (2011).
18. Parks, D. H. *et al.* 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 (2022).
19. Orakov, A. *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* **22**, 178 (2021).
20. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods* **20**, 1203–1212 (2023).
21. Léonard, R. R., Leleu, M., Van Vlierberghe, M., Kerff, F. & Baurain, D. ToRQuEMaDA: Tool for Retrieving Queried Eubacteria, Metadata and Dereplicating Assemblies. Preprint at https://doi.org/10.1101/2020.11.15.363259 (2020).
22. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
23. Cornet, L. *et al.* The GEN-ERA toolbox: unified and reproducible workflows for research in microbial genomics. *GigaScience* **12**, (2023).
24. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
25. Markin, A. *et al.* PARNAS: Objectively Selecting the Most Representative Taxa on a Phylogeny. *Systematic Biology* **72**, 1052–1063 (2023).
26. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
27. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
28. Wong, T. K. F. *et al.* IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. *EcoEvoRxiv* https://doi.org/10.32942/X2P62N (2025) doi:https://doi.org/10.32942/X2P62N.
29. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**, W29–W37 (2011).
30. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).

31. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

32. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

33. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. (2015).

34. Rahimzadeh-Karvansara, P., Pascual-Aznar, G., Bečková, M. & Komenda, J. Psb34 protein modulates binding of high-light-inducible proteins to CP47-containing photosystem II assembly intermediates in the cyanobacterium Synechocystis sp. PCC 6803. *Photosynth Res* **152**, 333–346 (2022).

35. Komenda, J. *et al.* The Psb27 Assembly Factor Binds to the CP43 Complex of Photosystem II in the Cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiology* **158**, 476–486 (2012).

36. Zhang, S., Frankel, L. K. & Bricker, T. M. The Sll0606 Protein Is Required for Photosystem II Assembly/Stability in the Cyanobacterium Synechocystis sp. PCC 6803. *Journal of Biological Chemistry* **285**, 32047–32054 (2010).

37. Anantharaman, V., Koonin, E. V. & Aravind, L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains11Edited by F. Cohen. *Journal of Molecular Biology* **307**, 1271–1292 (2001).

38. Singh, A. K., Li, H. & Sherman, L. A. Microarray analysis and redox control of gene expression in the cyanobacterium *Synechocystis* sp. PCC 6803. *Physiologia Plantarum* **120**, 27–35 (2004).

39. Yang, H. *et al.* Slr0151 in *Synechocystis* sp. PCC 6803 is required for efficient repair of photosystem II under high-light condition. *JIPB* **56**, 1136–1150 (2014).

40. Klinkert, B. *et al.* PratA, a Periplasmic Tetratricopeptide Repeat Protein Involved in Biogenesis of Photosystem II in Synechocystis sp. PCC 6803. *Journal of Biological Chemistry* **279**, 44639–44644 (2004).

41. Cormann, K. U., Bartsch, M., RÃ¶gner, M. & Nowaczyk, M. M. Localization of the CyanoP binding site on photosystem II by surface plasmon resonance spectroscopy. *Front. Plant Sci.* **5**, (2014).

42. Gandini, C., Schmidt, S. B., Husted, S., Schneider, A. & Leister, D. The transporter Syn PAM 71 is located in the plasma membrane and thylakoids, and mediates manganese tolerance in *Synechocystis* PCC 6803. *New Phytologist* **215**, 256–268 (2017).

43. Eisenhut, M. Manganese Homeostasis in Cyanobacteria. *Plants* **9**, 18 (2019).

44. Tottey, S. *et al.* Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **455**, 1138–1142 (2008).

45. Hennon, S. W., Soman, R., Zhu, L. & Dalbey, R. E. YidC/Alb3/Oxa1 Family of Insertases. *Journal of Biological Chemistry* **290**, 14866–14874 (2015).

46. Gathmann, Sven, Eva Rupprecht, Uwe Kahmann, & Dirk Schneider. A Conserved Structure and Function of the YidC Homologous Protein Slr1471 from Synechocystis sp. PCC 6803. *Journal of microbiology and biotechnology* **18**, 1090–1094 (2008).

47. Gao, H. & Xu, X. Depletion of Vipp1 in *Synechocystis* sp. PCC 6803 affects photosynthetic activity before the loss of thylakoid membranes. *FEMS Microbiology Letters* **292**, 63–70 (2009).

48. Zhang, S., Shen, G., Li, Z., Golbeck, J. H. & Bryant, D. A. 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 (2014).

49. Ji, Y. *et al.* Widespread but Poorly Understood Bacteria: Candidate Phyla Radiation. *Microorganisms* **10**, 2232 (2022).

50. Kuroda, K. *et al.* Symbiosis between *Candidatus* Patescibacteria and Archaea Discovered in Wastewater-Treating Bioreactors. *mBio* **13**, e01711-22 (2022).

51. Browman, D. T., Hoegg, M. B. & Robbins, S. M. The SPFH domain-containing proteins: more than lipid raft markers. *Trends in Cell Biology* **17**, 394–402 (2007).

52. Takekawa, N. *et al.* Structure of *Vibrio* FliL, a New Stomatin-like Protein That Assists the Bacterial Flagellar Motor Function. *mBio* **10**, e00292-19 (2019).

53. Siebenaller, C. & Schneider, D. Cyanobacterial membrane dynamics in the light of eukaryotic principles. *Bioscience Reports* **43**, BSR20221269 (2023).

54. Rexroth, S. *et al.* The Plasma Membrane of the Cyanobacterium *Gloeobacter violaceus* Contains Segregated Bioenergetic Domains. *The Plant Cell* **23**, 2379–2390 (2011).

55. Keren, N., Kidd, M. J., Penner-Hahn, J. E. & Pakrasi, H. B. A Light-Dependent Mechanism for Massive Accumulation of Manganese in the Photosynthetic Bacterium *Synechocystis* sp. PCC 6803. *Biochemistry* **41**, 15085–15092 (2002).

56. Nouwen, N. & Driessen, A. J. M. SecDFyajC forms a heterotetrameric complex with YidC. *Molecular Microbiology* **44**, 1397–1405 (2002).

57. Xie, K., Kiefer, D., Nagler, G., Dalbey, R. E. & Kuhn, A. 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 (2006).

58. Van Bloois, E. *et al.* The Sec-independent Function of Escherichia coli YidC Is Evolutionary-conserved and Essential. *Journal of Biological Chemistry* **280**, 12996–13003 (2005).

59. Oliver, D. C. & Paetzel, M. Crystal Structure of the Major Periplasmic Domain of the Bacterial Membrane Protein Assembly Facilitator YidC. *Journal of Biological Chemistry* **283**, 5208–5216 (2008).

60. Masato Nakai, Daigo Sugita, Tatsuo Omata, & Toshiya Endo. Sec-Y protein is localized in both the cytoplasmic and thylakoid membranes in the cyanobacterium Synchococcus PCC9742. *Biochemical and Biophysical Research Communications* **193**, 228–234 (1993).
61. Srivastava, R., Pisareva, T. & Norling, B. Proteomic studies of the thylakoid membrane of ***Synechocystis*** sp. PCC 6803. *Proteomics* **5**, 4905–4916 (2005).
62. Herrmann, J. M., Kauff, F. & Neuhaus, H. E. Thiol oxidation in bacteria, mitochondria and chloroplasts: Common principles but three unrelated machineries? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1793**, 71–77 (2009).

# Table and Figures

| | | Gloeobacterales | Thermostichales | Gloeomargaritales | Pseudanabaenales | Microcyanobacteria | Macrocyanobacteria |
|---|---|---|---|---|---|---|---|
| precursor of the D1 protein (pD1) insertion | YidC | | | | | | |
| | SecY | | | | | | |
| | SecE | | | | | | |
| | SecG | | | | | | |
| | SecD | | | | | | |
| | SecF | | | | | | |
| | SecA | | | | | | |
| Chlorophyll-binding molecule | Slr0144 | | | | | | |
| Chlorophyll incorporation into pD1 | ChlG | | | | | | |
| | Ycf39* | | | | | | |
| | HliC | | | | | | |
| | HliD | | | | | | |
| | Pitt | | | | | | |
| Manganese-preloading interaction with pD1 | PratA | | | | | | |
| Cleavage of the C-terminal extension of pD1 | CtpA | | | | | | |
| D2 folding | Slr0286 | | | | | | |
| | Slr2013 | | | | | | |
| RC complex formation | Ycf48 | | | | | | |
| | Sll0408 | | | | | | |
| | RubA | | | | | | |
| RC47 intermediate formation | Psb28 | | | | | | |
| | Sll0933 | | | | | | |
| | HliA | | | | | | |
| | HliB | | | | | | |
| | Psb35 | | | | | | |
| PSII lacking manganese cluster formation | Psb27 | | | | | | |
| | Sll0606* | | | | | | |
| | Psb34 | | | | | | |
| | TerC | | | | | | |
| Photoactivation of the oxygen-evolving manganese cluster | CyanoP | | | | | | |
| | Slr0565 | | | | | | |
| Protection from oxidative stress and photodamage | Psb32 | | | | | | |
| Supercomplex assembly | Slr1761 | | | | | | |
| PSII repair | Psb29 | | | | | | |
| | Slr0151 | | | | | | |
| | PsbN | | | | | | |
| | | | | | | | |
| DedA proteins | Slr0232 | | | | | | |
| | Slr0305 | | | | | | |
| | Sll0509 | | | | | | |
| SPFH proteins | Slr1106 | | | | | | |
| | Slr1768 | | | | | | |
| | Sll1021 | | | | | | |
| | Slr1128 | | | | | | |
| | Sll0815 | | | | | | |
| Cytoskeletal proteins | Sll1633 (FtsZ) | | | | | | |
| Membrane-bending proteins | Slr0483 | | | | | | |
| Dynamin-like proteins | Slr0869 | | | | | | |
| ESCRT-III proteins | Sll0617 (IM30) | | | | | | |
| | Slr1188 (PspA) | | | | | | |

**Table 1. Taxonomic distribution of the PSII assembly factors and proteins involved in membrane dynamics in *Synechocystis*.** Green represents the presence, while white represents the absence of the assembly factors. Orange represents the presence of homolog proteins but that were not found to be orthologous to the reference sequence. Grey represents undetermined orthogroups.

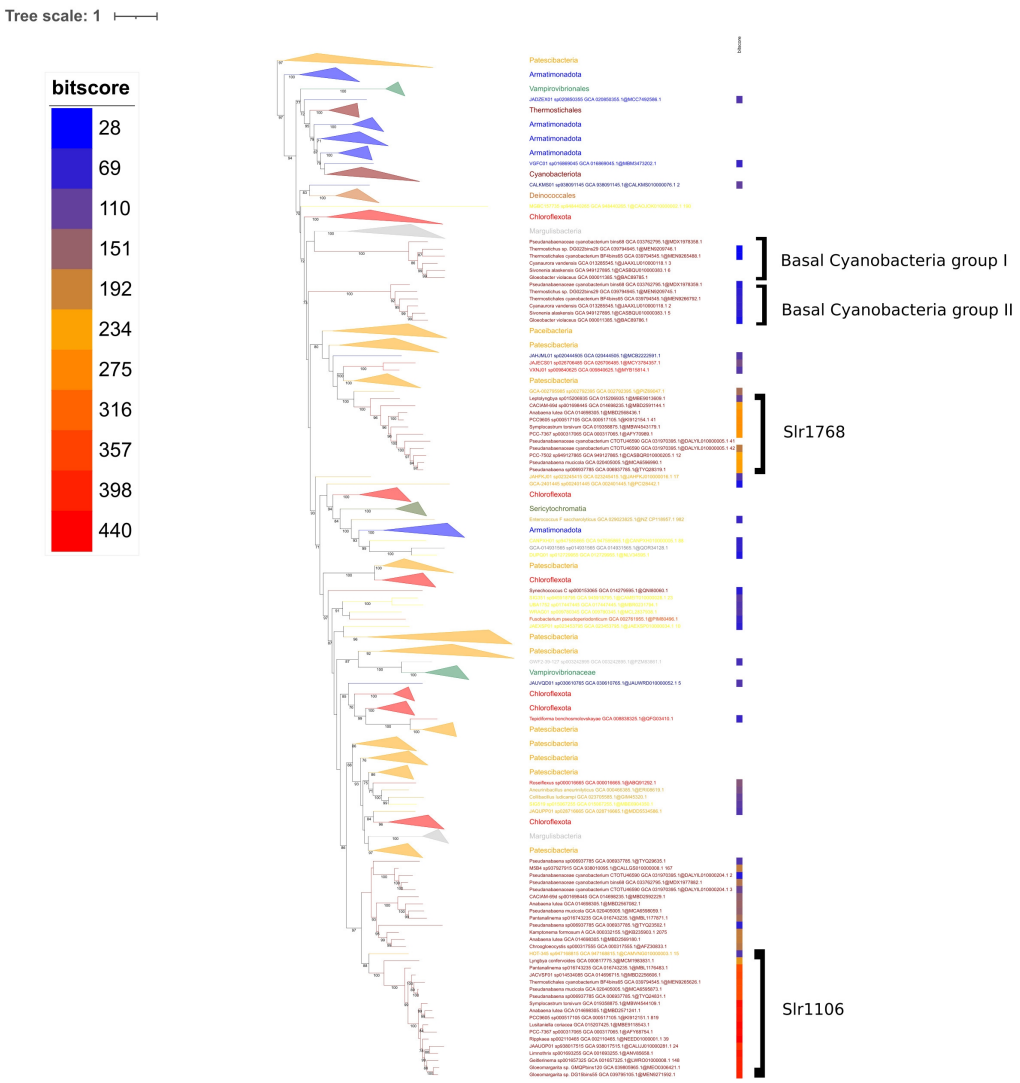


**Figure 1. Phylogenetic tree of the cyanobacterial orthologous sequences of Slr1106 and Slr1768 in the Terrabacteria dataset.** The phylogenetic tree was constructed from 273 sequences across 510 unambiguously aligned positions using IQ-TREE (Q.PLANT+F+I+R6 model) with 1000 ultrafast bootstrap replicates. Bootstrap values are shown above or equal to 70.

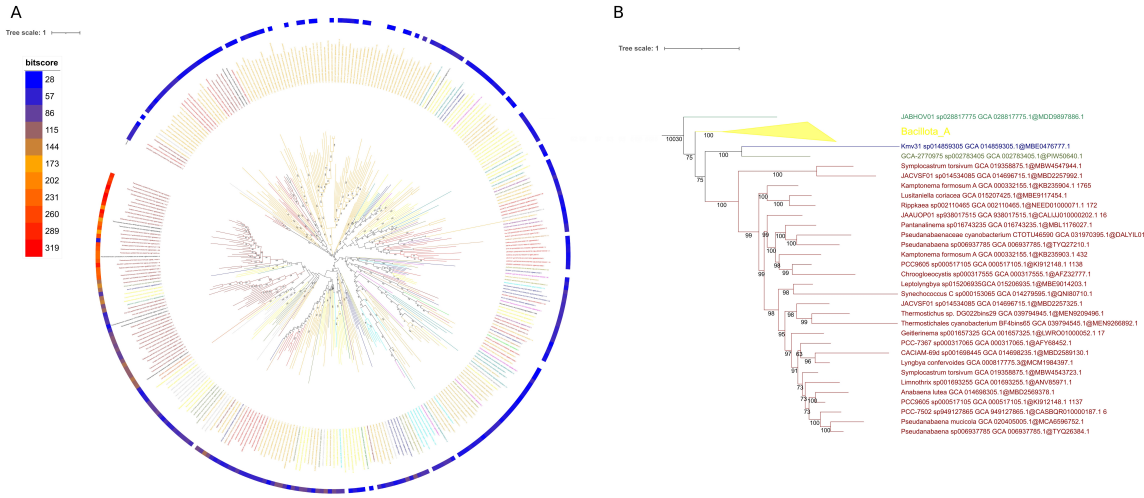


**Figure 2. Phylogenetic tree of the cyanobacterial orthologous sequences of Slr0305 (DedA family) in the Terrabacteria dataset.** (A) The phylogenetic tree was constructe from 347 sequences across 504 unambiguously aligned positions using IQ-TREE (LG+F+R7 model) with 1000 ultrafast bootstrap replicates. Bootstrap values are shown above or equal to. (B) The region matching the most with the reference sequence Slr0305 was zoomed.

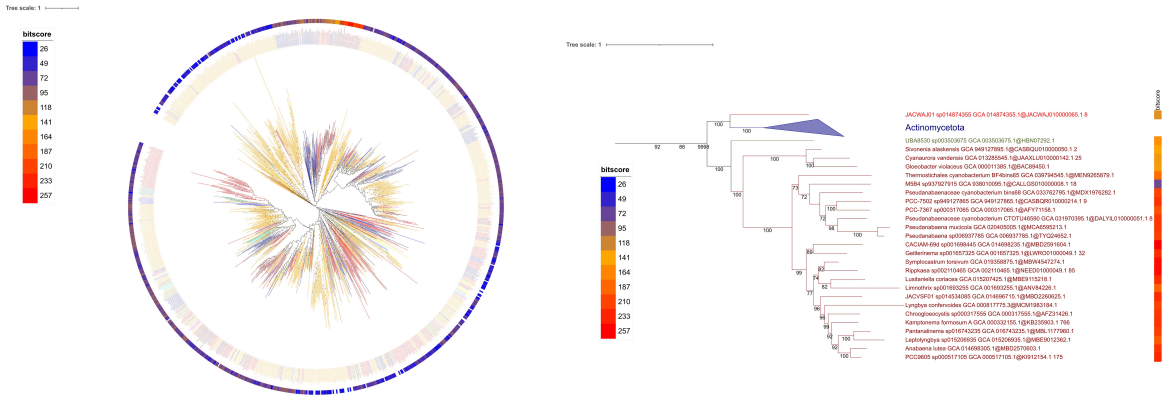


**Figure 3. Phylogenetic tree of the cyanobacterial orthologous sequences of Slr0232 (DedA family) in the Terrabacteria dataset.** (A) The phylogenetic tree was constructed from 987 sequences across 522 unambiguously aligned positions using IQ-TREE (LG+F+R10 model) with 1000 ultrafast bootstrap replicates. Bootstrap values are shown above or equal to. (B) The region matching the most with the reference sequence Slr0305 was pruned. The long branch at the basis of Phycobacteria is indicated by an arrow.

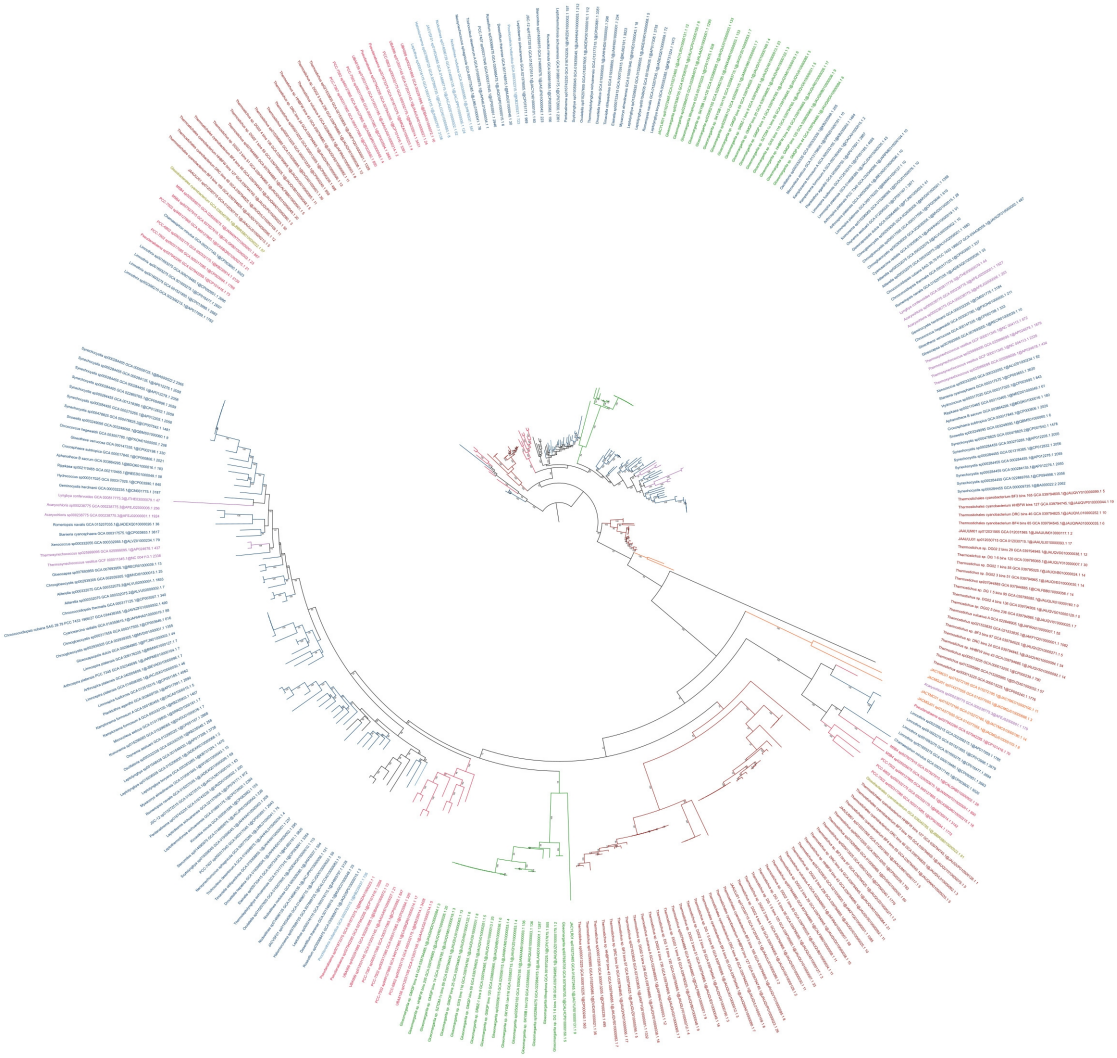

**Figure 4. Phylogenetic tree of the cyanobacterial orthologues sequences of Slr0144 and Slr0147 in the cyanobacterial dataset.** The phylogenetic tree was constructed from 311 sequences across 234 unambiguously aligned positions using IQ-TREE (Q.PFAM+R6) with 1000 ultrafast bootstrap replicates. Bootstrap values are shown above or equal to. The two reference sequences are indicated by arrows.

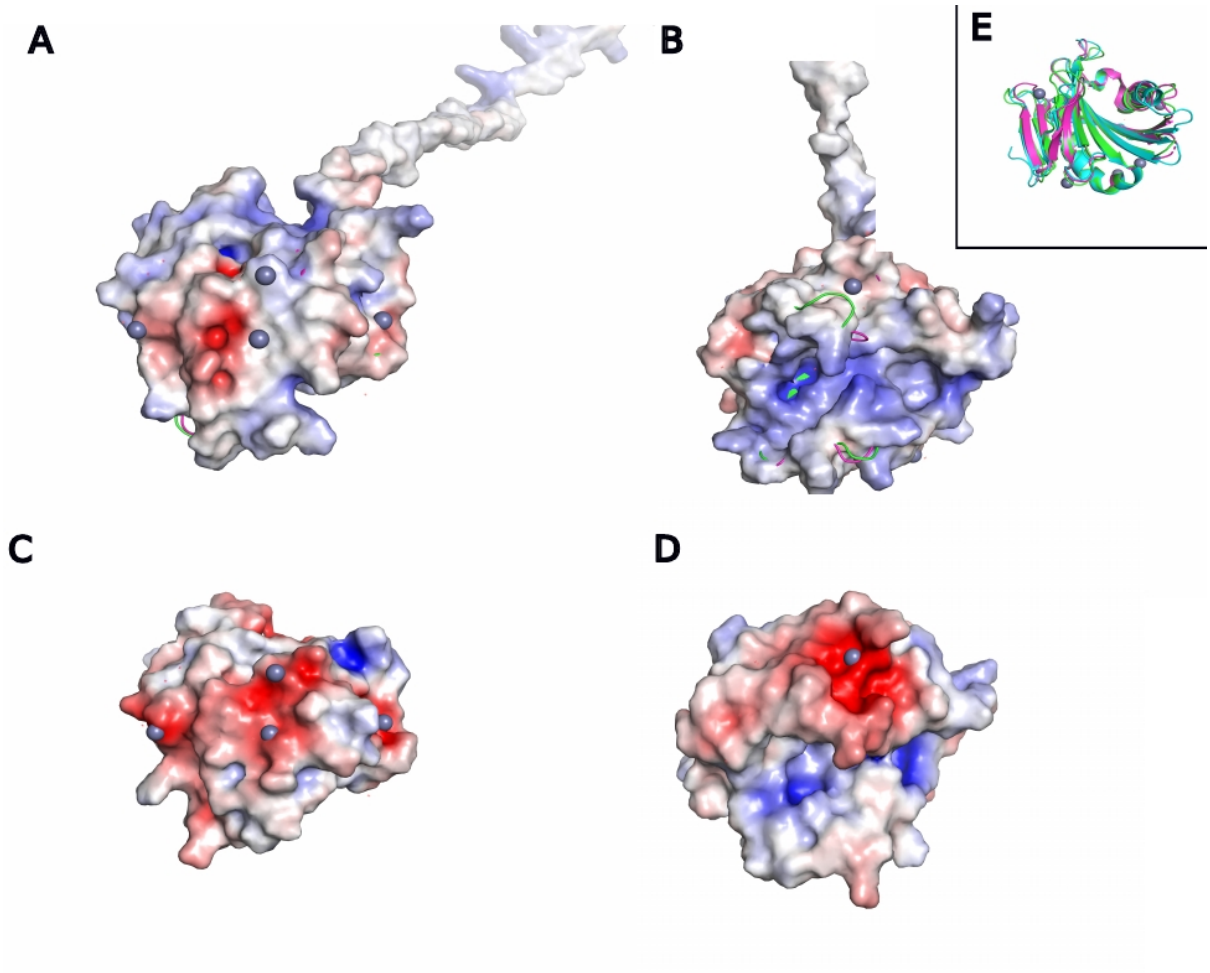


**Figure 5. Prediction of the 3D structure with APBS electrostatics estimation of *Gloeobacterales* and 2LNJ CyanoP structure from *Synechocystis* sp. PCC 6803.** The APBS electrostatics was esteemed based on the PyMOL plugin. The *Gloeobacterales* sequence was aligned to the PDB structure 2LNJ in order to find the ion-binding sites. (A, B) APBS electrostatics estimation of *Gloeobacterales* structure with two different views. (C, D) APBS electrostatics estimation of 2LNJ structure with two different views. (E) Alignment of *Gloeobacterales* CyanoP and 2LNJ structures.

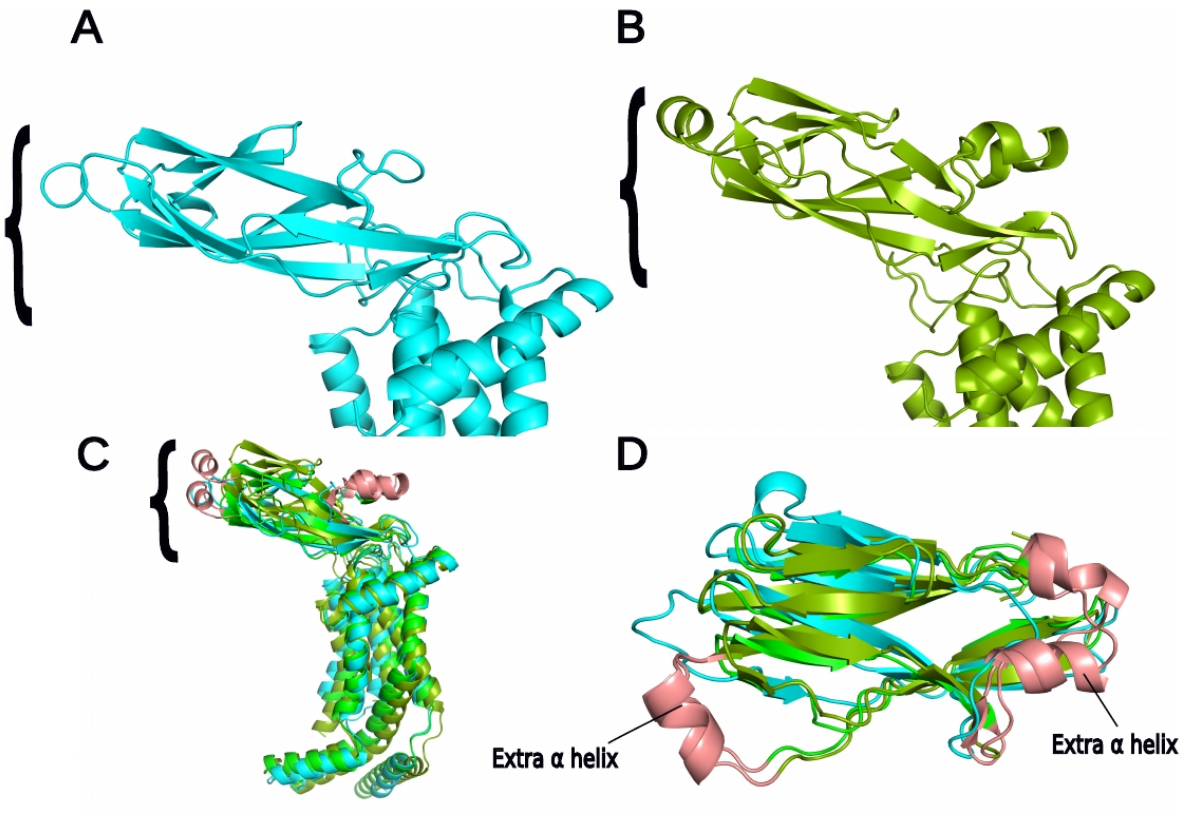


**Figure 6. Prediction of the 3D structure of Cyanobacterial YidC.** (A,B) The P1 domain (braces) is found in *Gloeobacterales* (cyan) and *Thermostichales* (olive green). (C) Structure alignment of *Gloeobacterales* (cyan), *Thermostichales* (olive green) and *Synechocystis* (green) structures with the P1 domain in braces. (D) Extra helices found in the P1 domain in Thermostichales (olive green) and *Synechocystis* (green) are highlighted in salmon.