

Mass spectrometry reveals the evolutionary conservation of phycobiliprotein complexes

Received: 28 November 2024

Accepted: 4 February 2026

Published online: 16 February 2026

 Check for updates

Jaspreet K. Sound ¹, Giorgio Bianchini ², Thrupthi A. Ashok¹,
Cecilia Rad-Menéndez ^{3,4}, David H. Green ³, Patricia Sánchez-Baracaldo ² &
Aneika C. Leney ¹ ✉

Cyanobacteria are a highly taxonomically and ecologically diverse group of oxygenic phototrophs that have colonized many different environments on our planet. Despite their differences, almost all cyanobacteria rely on highly efficient light-harvesting protein complexes, termed phycobilisomes, for effective photosynthesis. Phycobilisomes, along with the phycobiliproteins that make them up, have maintained their function throughout evolutionary history while also diversifying to optimize energy capture and transfer in different conditions. Here, we use a combination of evolutionary proteomics, phylogenomics, and structural bioinformatics to probe how phycobiliproteins have maintained their function while adapting to different habitats. Using high-resolution native mass spectrometry, we show that the two most abundant phycobiliprotein complexes, phycocyanin and allophycocyanin, are highly dynamic. Moreover, upon mixing phycobiliproteins from cyanobacterial strains representing diverse environments and evolutionary lineages, heterologous phycobiliprotein complexes rapidly form, comprising building blocks from different cyanobacterial strains. Bioinformatics and structural prediction methods allow us to identify critical residues involved in these interactions. We thus demonstrate that key structural features within the phycobiliprotein components have remained conserved over three billion years of cyanobacterial evolution, ensuring effective photosynthesis across a wide variety of natural environments.

Cyanobacteria are a group of photosynthetic prokaryotes that originated over three billion years ago^{1–3}, making them one of the most ancient and diverse organisms on Earth. As the only prokaryotes capable of performing oxygenic photosynthesis, cyanobacteria were pivotal in the development of the early Earth's atmosphere⁴. Even to this day they play key roles in important global processes such as the carbon cycle⁵ and the nitrogen cycle⁶. The cyanobacterial photosynthetic machinery is extremely efficient⁷ and involves multiple components, such as the photosystems, the light-harvesting complexes, and the oxygen-evolving complex. Investigating the structural

characteristics and behavior of these components is a key step towards understanding the evolutionary processes that led to their origin.

Cyanobacteria use large antenna complexes for energy transfer during photosynthesis. These complexes, termed phycobilisomes, comprise hundreds of linear tetrapyrrole (bilin)-containing proteins, known as phycobiliproteins, joined together by linker proteins⁸. The careful arrangement of bilin chromophores throughout the phycobilisome enables light energy between the 450–670 nm region to be absorbed and transferred towards chlorophyll within photosystems I and II with high quantum efficiency. While most cyanobacteria possess

¹School of Biosciences, University of Birmingham, Birmingham, UK. ²School of Geographical Sciences, University of Bristol, Bristol, UK. ³Scottish Association for Marine Science, Argyll, UK. ⁴Culture Collection of Algae and Protozoa (CCAP), Scottish Marine Institute, Oban, UK. ✉e-mail: a.leney@bham.ac.uk

phycobilisomes, the overall structure and size of these complexes can vary widely across different strains. For instance, rod bundle/paddle-shaped phycobilisomes are found in thylakoid-less cyanobacteria such as *Gloeobacter violaceus*^{9,10}, while hemi-discoidal phycobilisomes are the most common^{11,12}, with some hemi-ellipsoidal phycobilisomes arrangements also observed in red algae⁸. In general, phycobilisomes consist of a central core made from the phycobiliprotein allophycocyanin, from which rod-like assemblies protrude, comprising multiple copies of the structurally similar phycobiliprotein phycocyanin. In some cyanobacterial species, additional phycobiliproteins, such as phycoerythrin and phycoerythrocyanin, are also present within the rod-like structures, albeit at lower abundance¹³. Within hemi-discoidal phycobilisomes, there is further sub-division into those with di, tri, or penta-cylindrical cores¹⁴.

While there are many variations to the overall phycobilisome architecture¹⁵ across cyanobacteria, the structures of the individual phycobiliproteins are highly conserved due to their specialized function of absorbing and transferring specific wavelengths of light. Each phycobiliprotein is composed of dimeric building blocks consisting of two polypeptide chains, α and β , which assemble into donut-shaped ($\alpha\beta$)₃ hexameric complexes^{16,17} (Supplementary Figs. S1, 2). Within the phycobilisome, these ($\alpha\beta$)₃ complexes stack upon one another; their arrangement assisted by linker proteins, which act to hold the phycobiliproteins together and aid light energy transmission^{18–20}. In cyanobacteria, the phycobilisome is dynamic and rapidly assembles and disassembles in response to fluctuating light exposure and nutrient levels²¹. Moreover, the allophycocyanin core can even exchange its α and β subunits for alternative variants, α -B and β -18, respectively, to further alter energy transmission from the phycobilisome to photosystem I/II in response to environmental stimuli²².

Due to the similarities amongst the phycobiliproteins, it has long been suggested that α and β subunits evolved from a single ancestral protein^{23–25}. The individual phycobiliprotein complexes then evolved to ensure highly efficient unidirectional energy flow within the phycobilisome²⁶, whilst increasing their capacity to absorb longer wavelengths of light. However, questions still remain on how the structural framework of the phycobiliproteins has remained conserved in cyanobacteria over billions of years of diversification.

Native mass spectrometry (MS), the analysis of proteins and protein complexes whilst maintaining their native structure^{27–29}, has proven a fruitful tool within evolutionary proteomics to monitor hemoglobin³⁰, heat shock proteins³¹, and transcription factors³², but also to determine the evolutionary pressures that have enabled protein complexes to assemble in an ordered manner^{33,34}. By focusing specifically on proteins that are expressed, native MS enables the ability to draw conclusions from functionally acting proteins, rather than those that are encoded in the genome but are not actively translated. Within the context of phycobilisomes, native MS has been applied to study concentration-based self-assembly of phycocyanin³⁵, metal-based fluorescence quenching³⁶, linker protein modifications³⁷, and to

monitor structural changes within phycobiliprotein complexes^{38,39} and their associated proteins^{40,41}. In addition, through the use of recent advances in high-resolution MS, phycobiliproteins from a mixture of cyanobacterial strains can be rapidly separated and uniquely identified⁴².

Here, we use an evolutionary proteomics approach involving native MS to analyze phycobiliproteins from cyanobacterial strains from a range of different environments and evolutionary lineages. We show that across all the strains, phycobiliproteins are dynamic and can rapidly disassemble and reform novel heterologous complexes, composed of subunits coming from entirely different species. By monitoring the formation of heterologous phycobiliprotein complexes in real time, and combining these findings with structural predictions using AlphaFold2⁴³, we reveal how photosynthesis within cyanobacteria has remained a highly dynamic functional process.

Results

Phycobiliprotein dynamics vary between phycocyanin rods and allophycocyanin core

We first investigated the intrinsic dynamics within the phycocyanin rods and the allophycocyanin core of the phycobilisome from cyanobacterial strains found in a range of environments. Eight different cyanobacterial strains were first selected and their phycobiliprotein complexes extracted and analyzed directly by native MS (Table 1, Supplementary Table S1, Supplementary Figs. S3–9). Two mass spectra from the salt-tolerant strains *Limnospira maxima* CCAP 1475/9 (*L. maxima*) and *Gloeocapsopsis crepidinum* CCAP 1425/1 (*G. crepidinum*) are shown in Fig. 1. Native MS analysis of the phycobiliprotein extract from *L. maxima* shows two dominant charge state distributions between 4500 and 5500 *m/z* corresponding to the functional hexameric ($\alpha\beta$)₃ phycocyanin and hexameric ($\alpha\beta$)₃ allophycocyanin complexes (Fig. 1a). Despite peaks associated with phycocyanin being overall of higher relative intensity, the hexameric ($\alpha\beta$)₃ complex is in dynamic equilibrium with its dimeric $\alpha\beta$ building block (Fig. 1a, b). This is consistent with previous studies that have shown that ionic strength^{44–46}, protein concentration³⁵ and metal binding³⁶ can all influence phycocyanin's oligomeric state. On the other hand, allophycocyanin from both *L. maxima* and *G. crepidinum* remains hexameric ($\alpha\beta$)₃, with less than 5% $\alpha\beta$ dimer observed relative to the ($\alpha\beta$)₃ hexamer (Fig. 1b, d). Moreover, regardless of the strain chosen for analysis, when both phycocyanin and allophycocyanin are observed, phycocyanin is found to be consistently more dimeric in comparison with allophycocyanin (Supplementary Fig. S3–9).

Phycobiliproteins from different environments form heterologous complexes

Considering the high level of variation in oligomeric states and the structural similarity of phycobiliproteins across cyanobacterial strains, we next sought to determine whether phycobiliproteins from *L. maxima* could co-form complexes with phycobiliproteins from *G.*

Table 1 | Taxonomic classification and isolation environment for experimentally analyzed cyanobacterial strains

Strain	Species	Order	Environment
CCAP 1475/9	<i>Limnospira maxima</i>	Oscillatoriales	Hypersaline
CCAP 1425/1	<i>Gloeocapsopsis crepidinum</i>	Chroococciopsidales	Marine, thermotolerant
CCAP 1453/12	<i>Nostoc muscorum</i>	Nostocales	Freshwater
SAMS O1UC	<i>Kamptenema</i> sp.	Oscillatoriales	Freshwater
ANT.L61.2	<i>Phormidesmis priestleyi</i>	Leptolyngbyales	Freshwater, psychrophiles
CCAP 1475/3	<i>Spirulina major</i>	Spirulinales	Brackish
CCAP 1403/21	<i>Dolichospermum circinale</i>	Nostocales	Freshwater
CCAP 1437/1	<i>Gloeomargarita lithophora</i>	Gloeomargaritales	Freshwater
CCAP 1479/10	<i>Synechococcus</i> sp.	Synechococcales	Freshwater

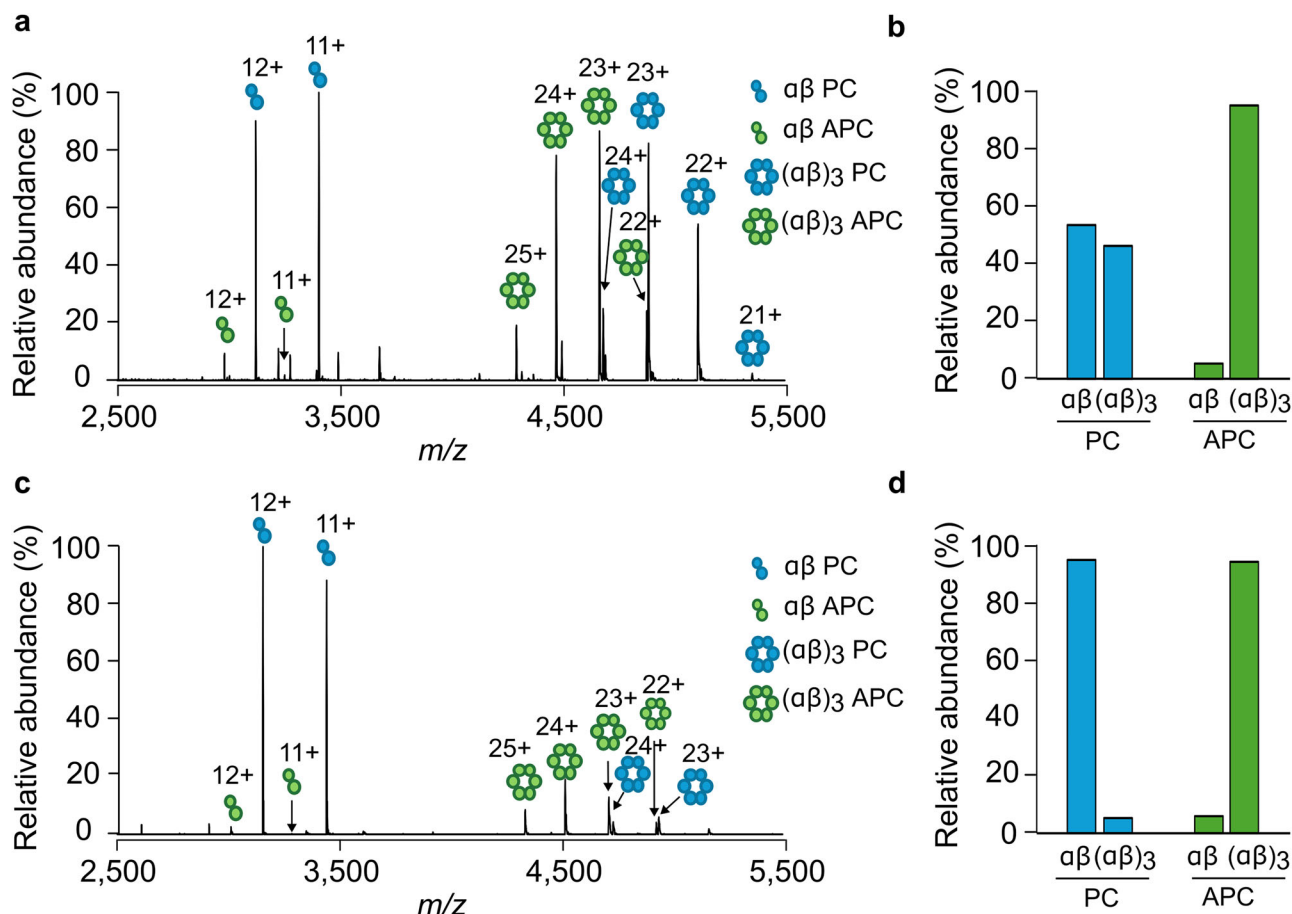


Fig. 1 | Phycocyanin rods are intrinsically more dynamic than the allophycocyanin core. Native mass spectrum of *L. maxima* (a, b) and *G. crepidinum* (c, d) phycobiliprotein extracts showing the dynamic equilibrium between the

dimeric ($\alpha\beta$) and hexameric ($\alpha\beta)_3$ phycocyanin (PC, blue) and allophycocyanin (APC, green) phycobiliprotein complexes.

crepidinum. To address this, phycobiliprotein extracts from *L. maxima* and *G. crepidinum* were mixed at an equal total protein ratio and the resulting complexes analyzed by native MS (Fig. 2a, Supplementary Tables S2, 3). Due to the high resolving power of the Orbitrap mass analyzer and the subtle differences between the primary sequences of phycobiliproteins from different cyanobacterial strains⁴², pure phycocyanin and allophycocyanin from *L. maxima* ($(\alpha\beta)_3$ _{L.maxima}) and *G. crepidinum* ($(\alpha\beta)_3$ _{G.crepidinum}) could be clearly resolved and identified. In addition, peaks corresponding to mixed complexes of both phycocyanin and allophycocyanin were observed in the high m/z region corresponding to $(\alpha\beta)_3$ _{2xL.maxima,1xG.crepidinum} and $(\alpha\beta)_3$ _{1xL.maxima,2xG.crepidinum}.

Next, we chose to determine the extent at which cyanobacterial strains from different environments could form heterologous phycobiliprotein complexes with one another. We mixed phycobiliproteins extracted from the hypersaline strain *L. maxima* and the freshwater strain *Nostoc muscorum* CCAP 1453/12 (*N. muscorum*) (Fig. 2b, Table 1), two fresh water strains with distinct morphologies *Kamptomena* sp. SAMS 01UC (*Kamptomena* sp.) and *N. muscorum* (Fig. 2c, Table 1), and two strains *G. crepidinum* and *Phormidesmis priestleyi* ANT.L61.2 (*P. priestleyi*) that can survive in hot and cold climates, respectively (Fig. 2d, Table 1) and analyzed the resulting complexes by native MS (Supplementary Tables S2, 3). In all cases, when the individual phycobiliproteins were present in the cell lysates from both strains in the mixture, heterologous hexameric ($\alpha\beta)_3$ complexes containing $\alpha\beta$ dimers from both strains were observed. Thus, this data indicates that if two distinct copies of the phycobiliprotein genes were present and

expressed at the same time within a single strain, mixed phycobiliprotein complexes would form. Interestingly, despite their high structural homology, hexameric ($\alpha\beta)_3$ complexes consisting of $\alpha\beta$ dimers from both allophycocyanin and phycocyanin were not observed regardless of the strains mixed (Fig. 2, Supplementary Fig. S10), highlighting the specificity in heterologous complex formation. Moreover, the heterogeneity is restrained to the hexameric ($\alpha\beta)_3$ complexes, with no mixed complexes evident in the dimer region of the spectrum (Supplementary Fig. S11). We hypothesized that low-abundant linker proteins within our protein extracts could be influencing heterologous complex assembly. Thus, next we took two protein extracts from *L. maxima* and *S. major*, and purified their phycocyanin and allophycocyanin complexes using a combination of ammonium sulfate precipitation and anion exchange chromatography (Supplementary Figs. S12, S13). The highly purified phycocyanin from *L. maxima* formed mixed complexes with phycocyanin from *S. major* (Supplementary Fig. S14a). Likewise, high-purity allophycocyanin from *L. maxima* formed mixed complexes with high-purity allophycocyanin from *S. major* (Supplementary Fig. S14b). Thus, although we cannot rule out that linker proteins may assist in mixed complex self-assembly, mixed complexes can form effectively without their presence.

Heterologous complex formation occurs in evolutionary distinct strains

To determine the evolutionary relationships between the strains used in Figs. 1 and 2, we performed a phylogenomic analysis (Fig. 3), including the strains within Table 1, as well as a representative sample

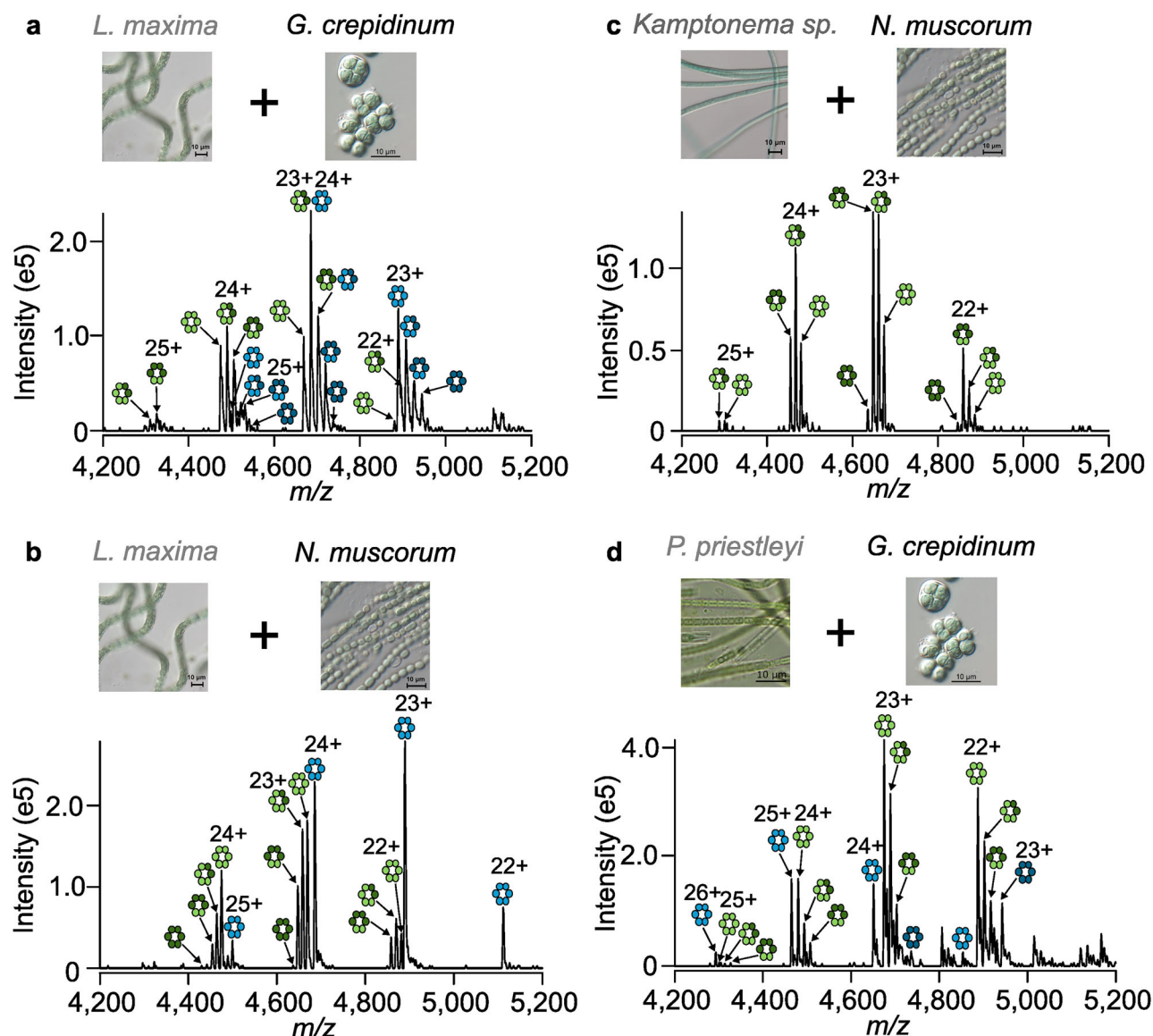


Fig. 2 | Heterologous phycobiliprotein complexes form containing proteins from different cyanobacterial strains. Native mass spectra of $(\alpha\beta)_3$ allophycocyanin (green) and phycocyanin (blue) complexes formed upon mixing phycobiliproteins from strains **a** *L. maxima* (light) and *G. crepidinum* (dark); **b** *L. maxima*

(light) and *N. muscorum* (dark); **c** *Kamptonema* sp. (light) and *N. muscorum* (dark); **d** *P. priestleyi* (light) and *G. crepidinum* (dark). The $\alpha\beta$ dimeric building blocks are colored (light or dark) according to the strain from which they originate.

of most orders of Cyanobacteria⁴⁷. *L. maxima* and *Kamptonema* sp. belong to the order Oscillatoriales, *N. muscorum* to the Nostocales, *G. crepidinum* to the Chroococciopsidales, and *P. priestleyi* to the Lep- tolyngbyales. Most of our selected strains so far are filamentous (with the exception of *G. crepidinum*, which does, however, form aggregates⁴⁸), and all are members of the Macrocyanobacteria, a monophyletic group that contains strains that generally have a cell diameter $>3\ \mu\text{m}$ ⁴⁹.

To determine whether the phycobiliproteins from evolutionarily distinct strains exchange differently, we performed three additional mixing experiments. Phycobiliproteins from the Nostocales strain *Dolichospermum circinale* CCAP 1403/21 (*D. circinale*) were mixed with the Oscillatoriales strain *L. maxima*, with the Microcyanobacteria strain *Synechococcus* sp. CCAP 1479/10 (*Synechococcus* sp.), and with the early-branching strain *Gloeomargarita lithophora* CCAP 1437/1 (*G. lithophora*), from the order Gloeomargaritales, which represents the group of cyanobacteria most closely related to the chloroplast^{50,51}. As

predicted, allophycocyanin $\alpha\beta$ dimers from *D. circinale* readily exchanged with those from *L. maxima*, resulting in mixed allophycocyanin $(\alpha\beta)_3$ complexes (Fig. 4a–c). Furthermore, despite being from distant evolutionary backgrounds, we also observed heterologous complexes formed from the exchange of $\alpha\beta$ dimers between *D. circinale* with *G. lithophora* (Fig. 4d–f) and *Synechococcus* sp. (Fig. S12) for both allophycocyanin and phycocyanin. Across these phycobiliprotein mixing experiments, the degree of heterologous complex formation reflects the difference in allophycocyanin:phycocyanin expression between strains.

Phycobiliprotein sequence conservation maintains essential function across distinct cyanobacterial lineages

To investigate structurally how these mixed complexes might form, we used AlphaFold2^{43,52} to predict the structure of mixed $(\alpha\beta)_3$ phycobiliprotein complexes containing $\alpha\beta$ dimers from each of the two distantly related strains *D. circinale* (Nostocales) and *G. lithophora*

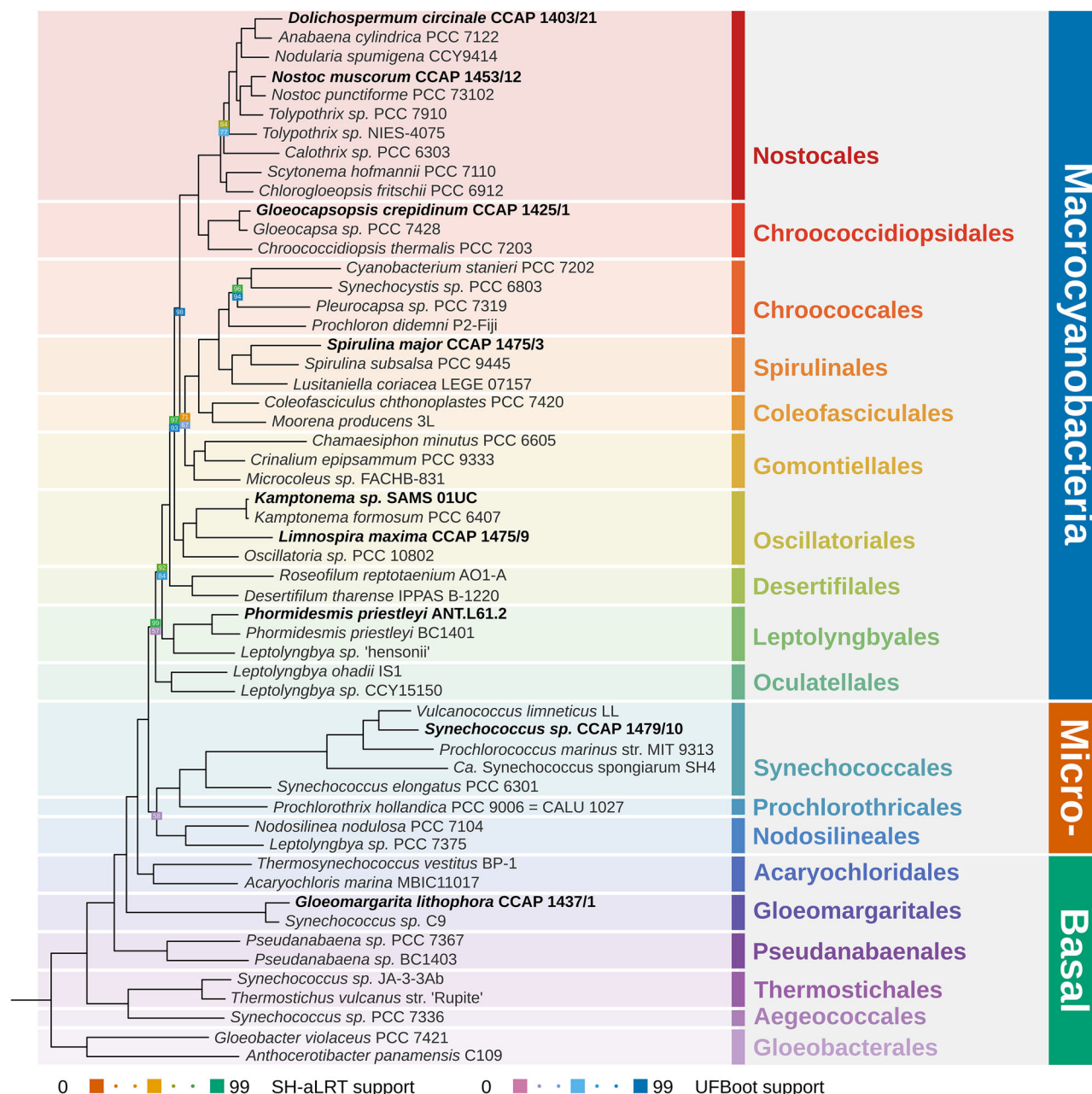


Fig. 3 | Maximum-likelihood phylogenomic tree of 55 strains of Cyanobacteria. This is estimated based on 136 protein-coding genes and 16S and 23S ribosomal RNA genes. Strains analyzed in our experiments are highlighted in bold. The SH-

aLRT and UFBoot support values are indicated on nodes where they are <100%. The tree was built in IQ-TREE v2.2.5⁶⁸ using a 136-gene phylogenomic dataset⁴, and visualized using TreeViewer v2.2.0⁶⁹.

(Gloeomargaritales), as well as complexes containing only dimers from an individual strain. The structure of allophycocyanin containing two ($\alpha\beta$)_{D.circinale} dimers and one ($\alpha\beta$)_{G.lithophora} dimer, a complex observed in our native MS data, is shown in Fig. 5a. To assess the confidence of AlphaFold2 in predicting interactions, we compared the number of contacts observed at the predicted dimer interface within a 3 Å distance for the mixed allophycocyanin structure with the allophycocyanin structure composed entirely from *D. circinale* (Fig. 5a). We also analyzed this using the predicted aligned error (PAE), a measure of the uncertainty in the relative position of two residues, and the predicted local distance difference test (pLDDT), a confidence metric estimating how well the prediction agrees with an experimental structure^{43,53}. These measures have been shown to successfully discriminate between interacting and non-interacting proteins⁵⁴, and here we used

them to qualitatively compare the binding strength between phycobiliprotein dimers. Consistent with the observation of the mixed allophycocyanin hexamer in vitro, the predicted mixed complex had a low PAE and a high interface pLDDT score, both of which were comparable to those of the *D. circinale* structure alone (Fig. 5b, Supplementary Fig. S16).

Due to the high structural similarity between allophycocyanin and phycocyanin, we next sought to predict why ($\alpha\beta$)₃ complexes containing both phycocyanin and allophycocyanin $\alpha\beta$ dimers were not observed by native MS. Thus, AlphaFold2 was used to predict a structure containing one phycocyanin $\alpha\beta$ dimer from *D. circinale* and two allophycocyanin $\alpha\beta$ dimers from the same strain. The predicted contact interface has a higher PAE, a lower pLDDT, and a smaller number of contacts compared to the ($\alpha\beta$)₃ complexes observed in

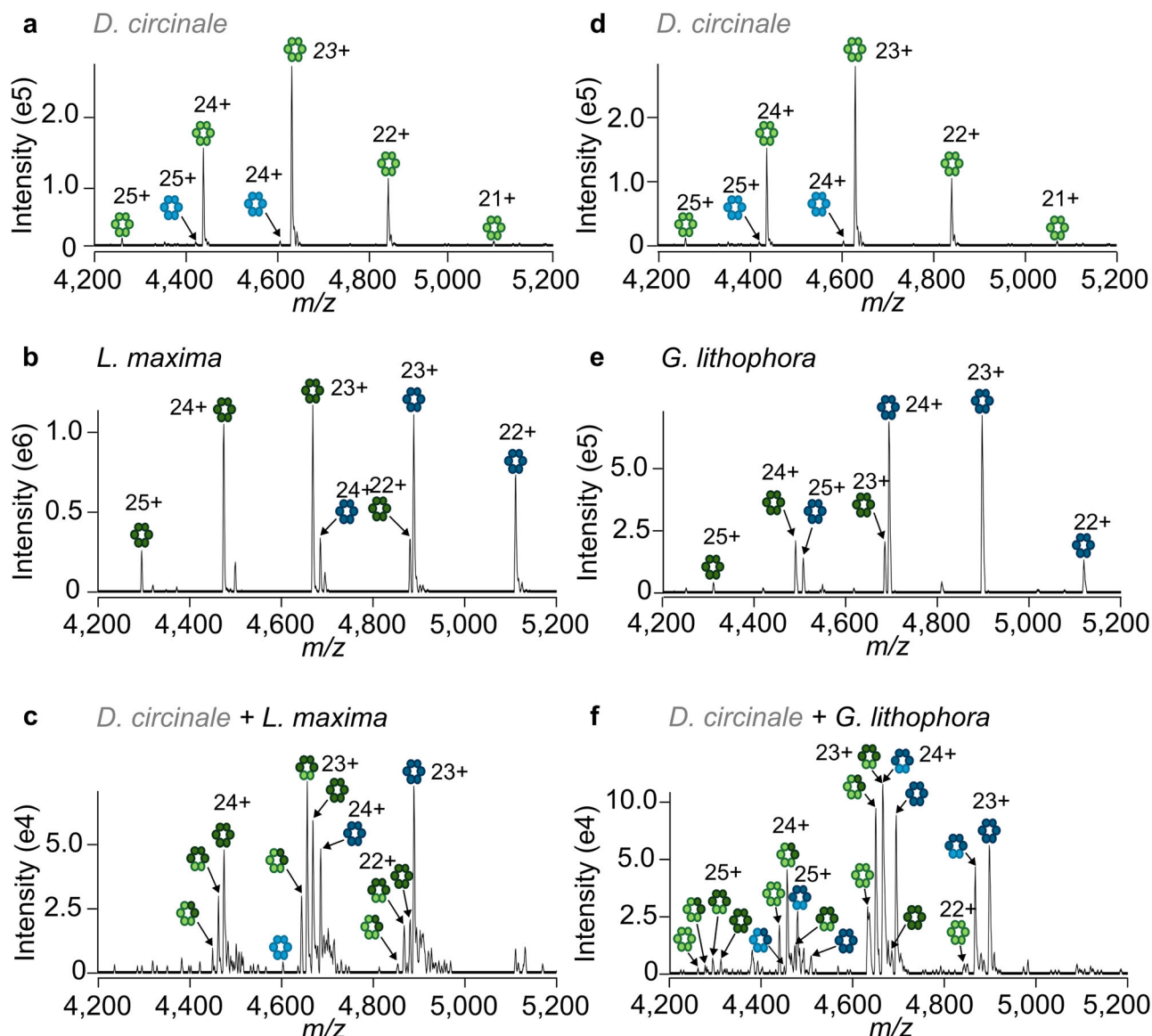


Fig. 4 | Heterologous phycobiliprotein complexes form containing proteins from distantly related strains. Native mass spectra of $(\alpha\beta)_3$ allophycocyanin (green) and phycocyanin (blue) complexes from *D. circinale* (a, d), *L. maxima* (b)

and *G. lithophora* (e) alone, and from *D. circinale* mixed with *L. maxima* (c) or *G. lithophora* (f). The $\alpha\beta$ dimeric building blocks are colored (light or dark) according to the strain from which they originate.

vitro (Fig. 5c, Supplementary Fig. S16). To identify key differences in the residues involved in the contacts between allophycocyanin and phycocyanin dimers, we aligned the sequences of the α and β subunits of allophycocyanin and phycocyanin from 51 strains of cyanobacteria (Supplementary Fig. S23) and inspected conserved residues in the contact interface of the predicted *D. circinale* structures. Helices E and F of the α subunit in both proteins interact with the β subunit's B and E helices and B-E loop. We noted a high degree of conservation of the residues at these critical contact interfaces between $\alpha\beta$ dimers across the different cyanobacterial strains, with an average of 97% (99%) and 88% (91%) identity (similarity) in allophycocyanin and phycocyanin, respectively. This supports the fact that heterologous phycobiliprotein complexes comprising subunits from different cyanobacterial strains can be observed by native MS. In addition, we noted that whilst in allophycocyanin a conserved leucine (residue 123 in the alignment, position 119 in the unaligned sequence for *D. circinale*) in helix α -F contacts a conserved tyrosine (residue 83 in the alignment, position 78 in the unaligned sequence for *D. circinale*) in helix β -E (Supplementary

Fig. S24a), phycocyanin has a conserved phenylalanine residue at this position. This different side chain causes the phycocyanin residue to contact with a conserved phenylalanine 62 in helix β -B, rather than with helix β -E as is the case with allophycocyanin (Supplementary Fig. S24b). Furthermore, the allophycocyanin helix α -E has conserved methionine and threonine (rarely serine) residues at positions 81 and 84, which interact with a conserved tyrosine at 63 in the B-E loop of the β subunit. Instead, PC α -E has a lysine at position 84; this amino acid is both has a longer side chain than methionine or threonine and is cationic, and hence the interaction with the B-E loop of the β subunit happens through the isoleucine at position 70. These structural differences highlight why phycocyanin and allophycocyanin do not co-form complexes in vitro. To corroborate this, we predicted the structure of a hypothetical hexamer consisting of two native *D. circinale* allophycocyanin dimers, together with a single *D. circinale* phycocyanin dimer where the regions of the α and β subunits predicted to interact with the other dimers were altered to match those of allophycocyanin. This resulted in a lower interface PAE and higher pLDDT

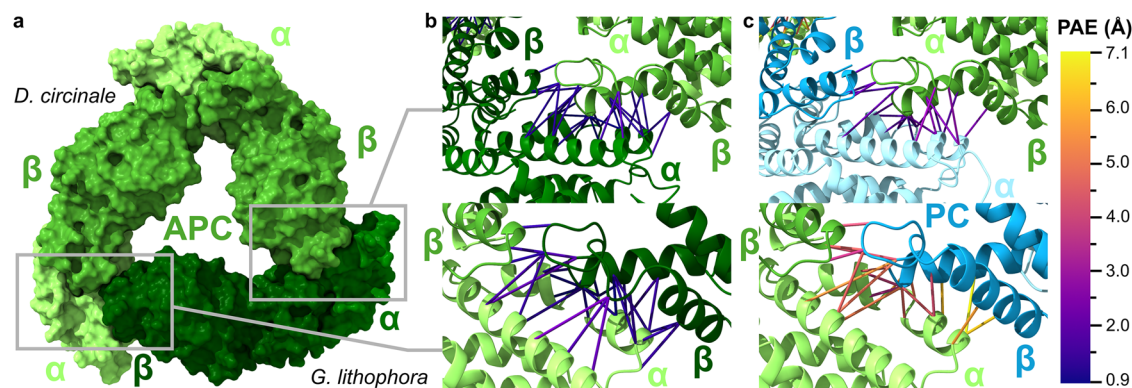


Fig. 5 | AlphaFold models provide rationale behind heterologous complex formation. a AlphaFold2 model of a heterologous allophycocyanin complex, containing two $\alpha\beta$ *D. circinalis* dimers (light green) and one $\alpha\beta$ *G. lithophora* dimer (dark green). **b** Predicted interfaces between the *G. lithophora* $\alpha\beta$ dimer and the

two *D. circinalis* $\alpha\beta$ dimers. **c** Predicted interfaces between a *D. circinalis* phycocyanin $\alpha\beta$ dimer and two allophycocyanin $\alpha\beta$ dimers from the same strain. The predicted alignment error (PAE) for pairs of residues within the dimer interfaces is highlighted.

than the mix of native *D. circinalis* allophycocyanin and phycocyanin, as well as the same number of inter-dimer contacts as native allophycocyanin (Supplementary Fig. S16), thus suggesting that this hypothetical complex would be more stable than a mixed hexamer containing native allophycocyanin and phycocyanin. Finally, we sought to investigate the functional advantage of heterologous complex formation. UV-visible absorption spectroscopy data showed different profiles for their phycobiliprotein extracts largely reflecting their abundances by native MS. (Supplementary Fig. S26). Upon mixing phycobiliproteins from two species, the absorbance profile matched the average of the absorbance profile for each individual species. Thus, if a poorly absorbing complex was mixed with a more efficient light energy absorbing complex, its absorbance profile would overall increase, making this beneficial for the organism that originally had an inefficient complex.

Discussion

Throughout Earth's history, cyanobacteria have evolved and colonized a wide range of habitats, including soil, streams, lakes, oceans, glaciers, deserts, endolithic communities, and hot springs (up to 72 °C)⁵⁵. They also engage in various symbiotic relationships with algae, fungi, plants, and animals. While phycobilisomes play a critical role in the light-harvesting and energy transfer processes during photosynthesis, their evolutionary history remains an open question. Here, we have shown that heterologous allophycocyanin and phycocyanin ($\alpha\beta$)₃ complexes can form consisting of $\alpha\beta$ dimers from different cyanobacterial strains. Moreover, this process occurs within minutes and is independent of cyanobacteria strain environment and relatedness. Observations of this nature were only possible due to the high resolving power of the mass spectrometer and its ability to decipher the complex compositions based on the subtle differences in phycobiliprotein masses derived from their amino acid sequences. In fact, suggestions that hybrid complexes can form between unicellular and filamentous cyanobacteria have been previously reported^{56,57}, however, knowledge on the detailed composition of these heterologous complexes and their quaternary structure was not feasible by the techniques used at that time.

Due to the universal presence of phycocyanin and allophycocyanin within cyanobacteria, it is understood that the two genes encoding the α and β subunits were duplicated before the last common ancestor of crown group cyanobacteria²⁵, which has been dated to over 3.1 Ga^{1–3,58}. Consistent with this hypothesis, our native MS data showed that allophycocyanin $\alpha\beta$ dimers do not form heterologous complexes with phycocyanin $\alpha\beta$ dimers regardless of the strain they were

extracted from (Supplementary Fig. S10). In addition, the predicted AlphaFold2 structures of heterologous hexamers containing both phycocyanin and allophycocyanin have high PAE and low pLDDT scores compared to their respective homogenous phycobiliprotein complexes (Supplementary Fig. S13). This suggests that ancestral phycobiliproteins were already optimized for function, and capturing energy from multiple wavelengths has always been highly beneficial. We speculate that, at the early stages of evolution, both the phycocyanin and allophycocyanin sets of genes were subject to strong evolutionary pressures. This is because, just after their initial divergence, heterologous complexes could form containing both allophycocyanin and phycocyanin subunits. Thus, structural mutations that initially differentiated phycocyanin from allophycocyanin likely happened on the contact interfaces between different dimers, resulting in the prevention of mixed phycocyanin/allophycocyanin hexamers and thus allowing their divergent evolution.

Overall, our data implies that any phycobiliprotein $\alpha\beta$ dimer should be functional in nearly any cyanobacterium in which it is integrated. This is intriguing since phycobiliprotein single gene trees and pan-genome trees are highly concordant, suggesting that phycobiliprotein genes have not undergone significant horizontal gene transfer (HGT). However, excluding these cyanobacterial genes from HGT is unlikely given the significant evidence for HGT in cyanobacteria^{59,60} that include core photosystem genes⁶¹. Moreover, it has recently been shown that the Far-Red Light Photoacclimation (FaRLiP) cluster which contains phycobiliprotein variants can be transferred laterally⁶². Indeed, cyanobacterial cells usually contain multiple copies of their genome per cell⁶³, meaning that HGT integration events may affect one of many genomes, and the multiplicity of phycobiliprotein genes on the non-integrated genomes dominate the phycobilisomes, maintaining photosystem homeostasis. In support of this, cyanobacterial strains often contain multiple duplicate genes for allophycocyanin⁶⁴ and phycocyanin⁶⁵ subunits. In addition, there are also allophycocyanin and phycocyanin genes encoding variants involved in phycobilisome assembly and energy transfer (e.g., the α -B subunit encoded by the *apcD* gene, or the β -18 subunit encoded by *apcF*) and adaptations to different light intensities or wavelengths (e.g., FaRLiP and LOLIP phenotypes⁶⁶). Overall, understanding whether HGT occurs with phycobiliprotein genes and how stringency is enforced remains a subject of further study. Indeed, research into whether heterologous complexes can form in vivo is a critical next step. Regulatory mechanisms may be involved in vivo that either accelerate or prevent heterologous phycobilisome complex assembly. Moreover, by monitoring the assembly/disassembly kinetics of heterologous complexes with linker

proteins present, further insight can be gained into the stability and functional relevance of these complexes in vivo. Nevertheless, if heterologous complexes of this nature can be induced by HGT and are able to form in vivo, this opens the opportunity to biochemically engineer the phycobilisome to change its light absorption characteristics.

Methods

Cyanobacterial growth

The strains *L. maxima*, *G. crepidinum* and *S. major* were grown in 50:50 artificial seawater/blue-green (ASW:BG11) culture medium. The strain *N. muscorum* was grown in BG11₀ culture medium. The strains *Kamptonema* sp., *G. lithophora* and *Synechococcus* sp. were grown in BG11. The strain *D. circinale* was grown in Jaworski's Medium culture medium. All CCAP and SAMS strains were kept on a rocker at 20 °C under a 12:12 h light/dark regime. The strains were examined under a Zeiss Axioimager A2 light microscope (Carl Zeiss, Germany). Micrographs were taken with an AxioCam 506 color camera (Carl Zeiss, Germany) using the Zen free software, v2.6.

P. priestleyi was supplied by the BCCM/ULC culture collection and grown in BG11 at 18 °C under constant illumination at $-10 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$. Microphotographs were captured on a Nikon Eclipse Ts2-FL inverted microscope with a 100x objective and a Nikon DS-Fi2 camera.

Cyanobacterial lysis and phycobiliprotein extraction

Cyanobacterial growth conditions are described in Supplementary methods. Fresh cyanobacterial cell pellets were taken and lysed in an equivalent volume of ultra-pure water using a combination of freeze-thaw (-70 to $+25$ °C) cycles and sonication. The lysate was then centrifuged at $16,000 \times g$ for 5 min to remove insoluble material. The use of ultra-pure water during cell lysis not only enabled rapid cell lysis, but also ensured a low ionic strength of the solution so that the phycobilisome could dissociate into its phycobiliprotein sub-complexes^{44–46}. A visible blue colored supernatant and a strong absorbance reading in 620–650 nm region indicated successful cell lysis. The lysate was then immediately buffer exchanged into 100 mM ammonium acetate pH 6.8 using an Amicon Ultra 0.5 mL centrifugal concentrator with a 30 kDa MWCO (Merck Millipore). This ensured all low molecular weight contaminants were removed. Phycobiliprotein purity was confirmed by native mass spectrometry (Supplementary Figs. S3–9). In all species, the phycobiliproteins peaks in the mass spectrum contribute to $\geq 65\%$ total ion signal. To further purify phycocyanin and allophycocyanin complexes from *L. maxima* and *S. major* to $>95\%$ purity, a combination of ammonium sulfate precipitation and anion exchange chromatography was used. For ammonium sulfate precipitation, 25% ammonium sulfate in 5 mM potassium phosphate pH 7.4 was initially added and any unwanted precipitated proteins removed by centrifugation. The phycobiliproteins were then precipitated in 60% ammonium sulfate. Phycocyanin and allophycocyanin were then separated using anion exchange chromatography. The ammonium sulfate precipitated phycobiliproteins were diluted >10 -fold into 5 mM potassium phosphate pH 7.4 to remove any residual ammonium sulfate and the proteins loaded onto a self-packed CHT Ceramic Hydroxyapatite column (Type 1, $40 \mu\text{m}$) (Bio-Rad) pre-equilibrated in 5 mM potassium phosphate pH 7.4. The absorbance at 280 nm, 620 nm and 650 nm was monitored using ÄKTA pure 25 chromatography system (Cytiva). Phycocyanin interacted weakly with the column and eluted immediately. 50 mM potassium phosphate pH 7.4 was used to elute allophycocyanin. High purity of phycocyanin and allophycocyanin complexes were confirmed by a combination of SDS-PAGE (Supplementary Fig. S12) and native mass spectrometry analysis (Supplementary Fig. S13). The concentration of phycobiliproteins was determined from the absorbance at 280 nm (Jenway 7315 Spectrophotometer) assuming an extinction co-efficient of $1 (\text{mg/ml})^{-1} \text{ cm}^{-1}$. All phycobiliproteins extracts were stored at 4 °C in the dark.

Native mass spectrometry

Phycobiliprotein extracts and purified phycobiliprotein complexes from individual strains were analyzed at a final concentration of 0.2 mg/mL in 100 mM ammonium acetate pH 6.8. To form mixed complexes, phycobiliprotein extracts from different strains were mixed at an equal final concentration of 0.2 mg/mL, incubated for 1 h at 4 °C in the dark, and analyzed immediately by UV-visible absorbance spectroscopy and native MS.

All native MS experiments were performed on an Orbitrap Eclipse Tribrid mass spectrometer equipped with a nano-electrospray ionization source (Thermo Fisher Scientific). For nano-electrospray ionization, borosilicate emitter tips (1.2 mm o.d., 0.68 mm i.d.) were pulled in-house (P-1000 micropipette puller, Sutter Instruments) and gold coated (sputter coater, Agar Scientific). The instrument was operated in positive ion mode and calibrated with FlexMix (Thermo Fisher Scientific). The capillary voltage was set to 1.2 kV, transfer tube at 250 °C, in-source dissociation 0 V and S-lens RF 120 %. Intact protein and high-pressure mode were used throughout. The Orbitrap was used for all mass spectra acquisition, typically with a mass range of 1000–8000 m/z and resolution of 7500 (at 400 m/z). The normalized automatic gain control was set to 100 %, maximum injection time 50–200 ms. All spectra were averaged over 1 min and the data processed using Xcalibur v4.1 (Thermo Fisher Scientific).

Predicted (theoretical) masses of the allophycocyanin and phycocyanin dimers and hexamers for each strain were calculated from the amino acid sequences of the allophycocyanin α , α -B, and β chains and phycocyanin α and β chains. The calculated masses were adjusted to include all predicted post-translational modifications (Supplementary Table S1). For allophycocyanin, the mass of the α monomer was modified to include the addition of 1 phycocyanobilin chromophore (+586.7 Da) and the loss of the initiator methionine (-131.2 Da) while the mass of the β monomer was modified to include the addition of 1 phycocyanobilin chromophore and the methylation of Asn71 to N4-methylAsn71 (+14 Da). For phycocyanin, the mass of the α was modified to include the addition of 1 phycocyanobilin chromophore while the mass of the β monomer was modified to include the addition of 2 phycocyanobilins, the methylation of Asn72 to N4-methylAsn72 and between 0 and 1 losses of the initiator methionine. The measured mass error on the experimentally determined molecular weight, and the percentage error between the theoretical and experimentally determined molecular weights were used to verify the presence of the allophycocyanin and phycocyanin complexes (Supplementary Tables S1, S2), whereby percentage errors of $<0.1\%$ indicated the presence of the corresponding protein complex.

Bioinformatics

To perform phylogenomic analysis, a set of 55 annotated cyanobacterial genomes (Supplementary Table S5) was selected, including 9 strains experimentally analyzed in this study and representing 19 cyanobacterial orders⁴⁷. For each genome, we used blastp v2.11.0+ to identify ortholog sequences for a phylogenomic dataset consisting of 145 protein-coding genes; genes for which a single copy ortholog was found in fewer than 30 strains were excluded from the analysis, leaving a total 136 genes. All sequence alignments were performed using MAFFT v7.511⁶⁷ with the $--\text{localpair}$ $--\text{maxiterate}$ 1000 options. The sequences for each gene were aligned and the resulting alignments were manually inspected and trimmed with the online Alignment-Viewer utility, removing positions with more than 85% of gaps and probably mis-aligned regions. The final dataset consisted of 50426 aligned amino acid positions. The best evolutionary model for each gene was selected using IQ-TREE v2.2.5^{68,69} with the $-m$ MF $--\text{mset}$ LG,Poisson,cpREV,Dayhoff,JTT,WAG,VT,DCMut,PMBJTTDCMut,-Blosum62,Q.LG,Q.pfam,Q.pfam_gb options, and a maximum-likelihood partitioned analysis was performed in IQ-TREE with the $-p$ option (which allows each partition to have its own evolutionary rate).

Support values for the tree were estimated using ultrafast bootstrap (1000 replicates, with the `--bnni` option to refine the sampled tree) and the SH-aLRT test (1000 replicates). The analysis was repeated three times and the tree with the maximum likelihood was selected, after verifying that there were no major discrepancies between the trees recovered by each analysis. The tree was visualized using TreeViewer v2.2.0⁶⁹.

To investigate phycobiliprotein diversity across the phylum, sequences for the *apcA*, *apcB*, *cpcA*, and *cpcB* genes (representing, respectively, the α and β subunits of APC and the α and β subunits of PC) were identified using blastp v2.11.0+ with the query sequences listed in Supplementary Table S6. The protein sequences for the five best hits from each genome were retrieved and aligned using MAFFT v7.511 with the `--localpair --maxiterate 1000` options, together with the query sequences and with sequences from other known homologs of each gene (Supplementary Table S7). For each gene, a maximum-likelihood phylogenetic tree was built using IQ-TREE v2.2.5 with the `--mset Q.pfam_gb` option and visualized using TreeViewer v2.2.0, in order to identify orthologs of the gene of interest. The ortholog sequences were re-aligned, and the resulting alignments were manually inspected. Minor improvements were made manually (i.e., aligning the starting methionine residues and correcting the position of loop insertions). A C# script was used to plot the alignment and assess sequence identity and similarity.

Protein structure predictions were carried out using a local install of AlphaFold v2.3.2⁴³, downloaded with the UniRef30 2023-02 release (Supplementary Figs. S17–22). Structure predictions were run in multimer mode, with template date cutoff 12th December 2023. The protein sequences used for structure predictions are listed in Supplementary Table S8. The predicted structures were visualized and analyzed using UCSF ChimeraX v1.7.1; the amino acid residues on the contact interface between individual monomers were identified using the “alphafold contacts” command. PAE and pLDDT values were extracted from the AlphaFold output files and plotted using a C# script.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw mass spectrometry and UV-vis absorbance spectroscopy data files are freely available via UoB edata archive [<https://doi.org/10.25500/edata.bham.00001196>]. The *P. priestleyi* ANT.L61.2 genome was deposited in GenBank under the accession JHBLFI0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_053471545.1/]. The BioProject number for CCAP/SAMS strains is PRJNA1127564. The CCAP1403/21 genome was deposited under genome accession code JBHYCU0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_051861405.1/], CCAP1425/1 under JBHYDD0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_051861225.1/], CCAP1437/1 under JBHYDV0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_051860525.1/], CCAP1453/12 under JBHYEF0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_051860325.1/], CCAP1475/3 under JBIMLG0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_051859885.1/], CCAP1475/9 under JBIMLI0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_051859755.1/] and SAMS01UC under the genome accession code JBHYGB0000000000 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_051859225.1/]. Accession numbers for the genomes used in the phylogenomic analysis are provided in Supplementary Table S5; raw, aligned, and trimmed sequence files are accessible from Zenodo (<https://doi.org/10.5281/zenodo.17991306>). Sequences used for structure prediction are listed in Supplementary Table S8; sequence files, predicted structures, and confidence metrics

are accessible from Zenodo (<https://doi.org/10.5281/zenodo.17991306>).

Code availability

C# scripts used in the methods and complete sequence alignments are available in GitHub (<https://github.com/arklumpus/SoundEtAl>) and archived on Zenodo (<https://doi.org/10.5281/zenodo.17478265>).

References

- Bianchini, G., Hagemann, M. & Sánchez-Baracaldo, P. Stochastic character mapping, Bayesian model selection, and biosynthetic pathways shed new light on the evolution of habitat preference in cyanobacteria. *Syst. Biol.* **73**, 644–665 (2024).
- Boden, J. S., Konhauser, K. O., Robbins, L. J. & Sánchez-Baracaldo, P. Timing the evolution of antioxidant enzymes in cyanobacteria. *Nat. Commun.* **12**, 4742 (2021).
- Fournier, G. P. et al. The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proc. Biol. Sci.* **288**, 20210675 (2021).
- Lyons, T. W. et al. Co-evolution of early Earth environments and microbial life. *Nat. Rev. Microbiol.* **22**, 572–586 (2024).
- Crockford, P. W., Bar On, Y. M., Ward, L. M., Milo, R. & Halevy, I. The geologic history of primary productivity. *Curr. Biol. CB* **33**, 4741–4750.e5 (2023).
- Sánchez-Baracaldo, P., Ridgwell, A. & Raven, J. A. A neoproterozoic transition in the marine nitrogen cycle. *Curr. Biol.* **24**, 652–657 (2014).
- Mulkidjanian, A. Y. et al. The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci.* **103**, 13126–13131 (2006).
- Bryant, D. A. & Gisriel, C. J. The structural basis for light harvesting in organisms producing phycobiliproteins. *Plant Cell* **36**, 4036–4064 (2024).
- Guglielmi, G., Cohen-Bazire, G. & Bryant, D. A. The structure of Gloeobacter violaceus and its phycobilisomes. *Arch. Microbiol.* **129**, 181–189 (1981).
- Jiang, H. W. et al. A structure of the relict phycobilisome from a thylakoid-free cyanobacterium. *Nat. Commun.* **14**, 8009 (2023).
- Bryant, D. A. & Canniffe, D. P. How nature designs light-harvesting antenna systems: design principles and functional realization in chlorophototrophic prokaryotes. *J. Phys. B At. Mol. Opt. Phys.* **51**, 033001 (2018).
- Sidler, W. A. “Phycobilisome and Phycobiliprotein Structures” in *The Molecular Biology of Cyanobacteria*, D. A. Bryant, Ed. (Springer Netherlands, pp. 139–216 (1994).
- Basheva, D. et al. Content of phycoerythrin, phycocyanin, allophycocyanin and phycoerythrocyanin in some cyanobacterial strains: applications. *Eng. Life Sci.* **18**, 861–866 (2018).
- Singh, N. K., Sonani, R. R., Rastogi, R. P. & Madamwar, D. The phycobilisomes: an early requisite for efficient photosynthesis in cyanobacteria. *EXCLI J.* **14**, 268–289 (2015).
- Lea-Smith, D. J. et al. Phycobilisome-deficient strains of *Synechocystis* sp. PCC 6803 have reduced size and require carbon-limiting conditions to exhibit enhanced productivity. *Plant Physiol.* **165**, 705–714 (2014).
- Brejč, K., Ficner, R., Huber, R. & Steinbacher, S. Isolation, crystallization, crystal structure analysis and refinement of allophycocyanin from the cyanobacterium *Spirulina Platensis* at 2.3 Å resolution. *J. Mol. Biol.* **249**, 424–440 (1995).
- Schirmer, T. et al. Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*. *J. Mol. Biol.* **188**, 651–676 (1986).
- Zheng, L. et al. Cryo-EM and femtosecond spectroscopic studies provide mechanistic insight into the energy transfer in CpcL-phycobilisomes. *Nat. Commun.* **14**, 3961 (2023).

19. Domínguez-Martin, M. A. et al. Structures of a phycobilisome in light-harvesting and photoprotected states. *Nature* **609**, 835–845 (2022).
20. Kawakami, K. et al. Core and rod structures of a thermophilic cyanobacterial light-harvesting phycobilisome. *Nat. Commun.* **13**, 3389 (2022).
21. Grossman, A. R., Schaefer, M. R., Chiang, G. G. & Collier, J. L. The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol. Rev.* **57**, 725–749 (1993).
22. Calzadilla, P. I., Muzzopappa, F., Sétif, P. & Kirilovsky, D. Different roles for ApcD and ApcF in *Synechococcus elongatus* and *Synechocystis* sp. PCC 6803 phycobilisomes. *Biochim. Biophys. Acta BBA Bioenerg.* **1860**, 488–498 (2019).
23. Glazer, A. N. et al. Biliproteins of cyanobacteria and Rhodophyta: Homologous family of photosynthetic accessory pigments. *Proc. Natl. Acad. Sci.* **73**, 428–431 (1976).
24. Glazer, A. N. Structure and evolution of photosynthetic accessory pigment systems with special reference to phycobiliproteins. *UCLA Forum Med. Sci.* **21**, 221–244 (1979).
25. Apt, K. E., Collier, J. L. & Grossman, A. R. Evolution of the phycobiliproteins. *J. Mol. Biol.* **248**, 79–96 (1995).
26. Huber, R. A structural basis of light energy and electron transfer in biology (Nobel Lecture). *Angew. Chem.* **28**, 848–869 (1989).
27. Karch, K. R., Snyder, D. T., Harvey, S. R. & Wysocki, V. H. Native mass spectrometry: recent progress and remaining challenges. *Annu. Rev. Biophys.* **51**, 157–179 (2022).
28. Leney, A. C. & Heck, A. J. R. Native mass spectrometry: what is in the name?. *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
29. Boeri Erba, E., Signor, L. & Petosa, C. Exploring the structure and dynamics of macromolecular complexes by native mass spectrometry. *J. Proteomics* **222**, 103799 (2020).
30. Pillai, A. S. et al. Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).
31. Hochberg, G. K. A. et al. Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions. *Science* **359**, 930–935 (2018).
32. Joerger, A. C. et al. Structural evolution of p53, p63, and p73: Implication for heterotetramer formation. *Proc. Natl. Acad. Sci.* **106**, 17705–17710 (2009).
33. Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265 (2008).
34. Marsh, J. A. et al. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461–470 (2013).
35. Eisenberg, I. et al. Concentration-based self-assembly of phycocyanin. *Photosynth. Res.* **134**, 39–49 (2017).
36. Bellamy-Carter, J., Sound, J. K. & Leney, A. C. Probing heavy metal binding to phycobiliproteins. *FEBS J.* **289**, 4646–4656 (2022).
37. Tamara, S., Hoek, M., Scheltema, R. A., Leney, A. C. & Heck, A. J. R. A colorful pallet of B-phycoerythrin proteoforms exposed by a multimodal mass spectrometry approach. *Chem* **5**, 1302–1317 (2019).
38. Kaldmäe, M., Sahin, C., Saluri, M., Marklund, E. G. & Landreh, M. A strategy for the identification of protein architectures directly from ion mobility mass spectrometry data reveals stabilizing subunit interactions in light harvesting complexes. *Protein Sci. Publ. Protein Soc.* **28**, 1024–1030 (2019).
39. Leney, A. C., Tschanz, A. & Heck, A. J. R. Connecting color with assembly in the fluorescent B-phycoerythrin protein complex. *FEBS J.* **285**, 178–187 (2018).
40. Zhang, H. et al. Molecular mechanism of photoactivation and structural location of the cyanobacterial orange carotenoid protein. *Biochemistry* **53**, 13–19 (2014).
41. Lu, Y. et al. Native mass spectrometry analysis of oligomerization states of fluorescence recovery protein and orange carotenoid protein: two proteins involved in the cyanobacterial photoprotection cycle. *Biochemistry* **56**, 160–166 (2017).
42. Sound, J. K. et al. Rapid cyanobacteria species identification with high sensitivity using native mass spectrometry. *Anal. Chem.* **93**, 14293–14299 (2021).
43. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
44. Anderson, L. K., Rayner, M. C., Sweet, R. M. & Eiserling, F. A. Regulation of *Nostoc* sp. phycobilisome structure by light and temperature. *J. Bacteriol.* **155**, 1407–1416 (1983).
45. Zilinskas, B. A. & Glick, R. E. Noncovalent intermolecular forces in phycobilisomes of *Porphyridium cruentum*. *Plant Physiol.* **68**, 447–452 (1981).
46. Gantt, E., Lipschultz, C. A. & Zilinskas, B. Further evidence for a phycobilisome model from selective dissociation, fluorescence emission, immunoprecipitation, and electron microscopy. *Biochim. Biophys. Acta BBA Bioenerg.* **430**, 375–388 (1976).
47. Strunecký, O., Ivanova, A. P. & Mareš, J. An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis. *J. Phycol.* **59**, 12–51 (2023).
48. Jung, P. et al. Emendation of the coccoid cyanobacterial genus *Gloeocapsopsis* and description of the new species *Gloeocapsopsis diffuens* sp. nov. and *Gloeocapsopsis dulcis* sp. nov. isolated from the coastal range of the Atacama Desert (Chile). *Front. Microbiol.* **12**, 671742 (2021).
49. Sánchez-Baracaldo, P. Origin of marine planktonic cyanobacteria. *Sci. Rep.* **5**, 17418 (2015).
50. Ponce-Toledo, R. I. et al. An early-branching freshwater cyanobacterium at the origin of plastids. *Curr. Biol. CB* **27**, 386–391 (2017).
51. Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci.* **114**, E7737–E7745 (2017).
52. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at <https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2> (2022).
53. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
54. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
55. Castenholz, R. W. et al. “Phylum BX. Cyanobacteria” in *Bergey’s Manual® of Systematic Bacteriology: Volume One: The Archaea and the Deeply Branching and Phototrophic Bacteria*, 473–599 (Springer, 2001).
56. Glazer, A. N. & Fang, S. Formation of hybrid proteins from the α and β subunits of phycocyanins of unicellular and filamentous blue-green algae. *J. Biol. Chem.* **248**, 663–671 (1973).
57. Cohen-Bazire, G., Béguin, S., Rimon, S., Glazer, A. N. & Brown, D. M. Physico-chemical and immunological properties of allophycocyanins. *Arch. Microbiol.* **111**, 225–238 (1977).
58. Sánchez-Baracaldo, P., Bianchini, G., Wilson, J. D. & Knoll, A. H. Cyanobacteria and biogeochemical cycles through Earth history. *Trends Microbiol.* **30**, 143–157 (2022).
59. Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**, 1099–1108 (2006).
60. Tooming-Klunderud, A. et al. From green to red: horizontal gene transfer of the phycoerythrin gene cluster between *Planktothrix* strains. *Appl. Environ. Microbiol.* **79**, 6803–6812 (2013).
61. Yerrapragada, S., Siefert, J. L. & Fox, G. E. Horizontal gene transfer in cyanobacterial signature genes. *Methods Mol. Biol. Clifton NJ* **532**, 339–366 (2009).

62. Gisriel, C. J., Bryant, D. A., Brudvig, G. W. & Cardona, T. Molecular diversity and evolution of far-red light-acclimated photosystem I. *Front. Plant Sci.* **14**, 1289199 (2023).
63. Watanabe, S. Cyanobacterial multi-copy chromosomes and their replication. *Biosci. Biotechnol. Biochem.* **84**, 1309–1321 (2020).
64. Jiang, H.-W. et al. A structure of the relic phycobilisome from a thylakoid-free cyanobacterium. *Nat. Commun.* **14**, 8009 (2023).
65. Tang, J., Zhou, H., Yao, D., Du, L. & Daroch, M. Characterization of molecular diversity and organization of phycobilisomes in thermophilic cyanobacteria. *Int. J. Mol. Sci.* **24**, 5632 (2023).
66. Gan, F., Shen, G. & Bryant, D. A. Occurrence of far-red light photoacclimation (FaRLiP) in diverse cyanobacteria. *Life Basel Switz* **5**, 4–24 (2014).
67. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
68. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
69. Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol. Evol.* **14**, e10873 (2024).

Acknowledgements

The mass spectrometry research was supported by the Biotechnology and Biological Sciences Research Council (BBSRC, BB/T015640/1) (A.C.L.). J.K.S. was funded through University of Birmingham funded Midlands Integrative Biosciences Training Partnership (BB/M01116X/1). The Orbitrap Eclipse Tribrid mass spectrometer was funded by the BBSRC (BB/S019456/1) (A.C.L.). We would like to thank Sarah Mansfield-Ford, Jeddiah Bellamy-Carter, and Pasquale Miglionico for helpful preliminary discussions. Hannah E. Wedgwood assisted with protein purification. We thank the Advanced Mass Spectrometry facility at the University of Birmingham for set-up and maintenance of the mass spectrometers used in this work. Genome sequencing was performed at the Centre for Genomic Research of the University of Liverpool. Genome assembly and phylogenomic analyses were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bristol.ac.uk/acrc/>. This work was supported by a Royal Society University Research Fellowship to P.S.-B. G.B. was funded by a Royal Society postdoctoral grant given to PSB.

Author contributions

A.C.L. and J.K.S. designed the mass spectrometry analysis. P.S.B. and G.B. designed the phylogenetic analysis. C.R.M. provided the CCAP

cultures and associated images. Cyanobacterial strains were maintained by T.A.A. and J.K.S. J.K.S. collected the mass spectrometry data. J.K.S. and A.C.L. analyzed the mass spectrometry data. G.B. performed the phylogenetic analysis and AlphaFold structural predictions. J.K.S., A.C.L., G.B., P.S.B. and D.H.G. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-69558-y>.

Correspondence and requests for materials should be addressed to Aneika C. Leney.

Peer review information *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026