

Genomes of N₂-fixing endosymbionts and host-independence

Abstract

Background

The projected 2.7-fold increase in population in sub-Saharan Africa by the end of the century demands consideration as to how agricultural output can keep pace. Augmenting nitrogen inputs is a practical necessity, but this must be accomplished in such a way that avoids the environmental costs of past advances and also places the resource in the hands of those who will be the most affected. Biological nitrogen fixation might play an important role. The realization that certain algae are able to provide for their own nitrogen needs by fixing atmospheric N₂ raises the possibility that an endosymbiont responsible for the nitrogen might be transferred to crop plants. For this to take place, it is necessary that the endosymbionts be (or be made to be) sufficiently independent of their hosts so that they may establish themselves in crop plants appropriate to African agriculture.

Results

Genomes from six endosymbionts from diatoms within the family Rhopalodiaceae were analyzed. They were compared to genomes from free-living cyanobacteria and to those of the nitroplast UCYN-A and chromatophore from *Paulinella*, to which they are related. Unlike the latter two endosymbionts, the six from *Rhopalodia* encode all the enzymes considered that underlie metabolic processes and provide the energy to power N-fixation. Some of the endosymbionts also appear able to synthesize cofactors essential for central metabolism. The analysis points to possible carbon sources the endosymbionts might take up from their hosts, including glycerol and chitobiose. Possible routes of nitrogen export to the host were also examined.

Conclusions

Within the limits of genome analysis, some of the Rhopalodian endosymbionts appear to be metabolically independent of their hosts, except for requiring a carbon source. However, the choice of carbon source and the likely means of nitrogen export are not compatible with crop plants. Genetic modification would surely be necessary for any prospect of propagation of an endosymbiont in a plant of agricultural importance, and significant questions must first be answered in the laboratory. To this end, the endosymbiont of *Epithemia clementina* may be best suited for such investigations, eventually after transfer to the model diatom *Phaeodactylum tricornutum*.

Keywords

Nitrogen fixation; symbiosis; endosymbiosis; metabolism; carbon transport; nitrogen transport; evolution

Background

For the past 60 years, the increase in global agricultural production has depended on a concomitant increase in the application of nitrogenous fertilizer.^{1,2} This has come at a cost, both economic³ and

environmental,^{4,5} and climate change promises to make these costs even greater.⁶ Historically, the price of fertilizer has been too great for the majority of farmers in sub-Saharan Africa,^{7,8} where food needs are most intense and which will contribute most of the population rise in the coming decades.⁹ To keep pace, it would seem that the world must find a way to increase the production of nitrogenous fertilizer, especially in sub-Saharan Africa, but that would also increase the substantial harm that accompanies its use. While this may seem like a global problem, it is best viewed as local, addressed by technologies that can adapt to the great heterogeneity confronting African farmers and that can take advantage of input provided by the farmers themselves.⁸

In principle, biological nitrogen fixation offers a way out. N₂-fixing rhizobia within specialized root nodules of legumes meet the nitrogen needs of their plant hosts in a way that's low cost and environmentally benign. However, the interactions between rhizobia and their hosts are famously complex,¹⁰ and we remain a long way from extending the benefits of root nodule N₂-fixation to crop plants beyond the legumes.¹¹ Other approaches have been explored – inoculation with associative N₂-fixing bacteria, engineering plants to express nitrogenase and related genes, changes in agronomic practices -- each with its own problems.^{12,13,14}

An alternate strategy has not been given much attention, inspired by the example of the chloroplast. Chloroplasts reduce atmospheric CO₂ to sugar for use by the plant. Why not an organelle that reduces abundant atmospheric N₂ to ammonia?^{15,16} This organelle – call it a nitroplast – might provide for the nitrogen needs of diverse crop plants, just as chloroplasts provide for their carbon needs. Admittedly, this sounds like the stuff of science fiction, but such organelles already exist!¹⁷ The haptophyte alga *Braarudosphaera bigelowii* contains an endosymbiont (called UCYN-A), phylogenetically related to unicellular cyanobacteria, that provides its algal host fixed nitrogen from N₂. Its division is integrated with the cell division cycle of its host. It is everything one could ask for in a nitroplast...

...except for portability. Even chloroplasts within the same species may exhibit different levels of compatibility with different genomes,¹⁸ although successful transfer of chloroplasts to members of different closely related species has been demonstrated.¹⁹ A narrow range of compatibility is expected because the encoding of chloroplast proteins is split between the nucleus and the chloroplast, and the proteins encoded by the nucleus of one species may not entirely match the proteins needed by a chloroplast of another. In addition, multiprotein complexes may not work when mixing protein components or regulatory mechanisms of different species.^{18,20} The nitroplast UCYN-A could not possibly function in crop plants. Like chloroplasts, it imports hundreds of proteins from its host, many required for basic functions.¹⁷ It would be incredible if a land plant had the ability to replace these proteins, even if the protein import systems²¹ of the old and new hosts were compatible.

Now consider the nitroplast-like endosymbionts (also called spheroid bodies) found within diatoms of the family Rhopalodiaceae.^{22,23} They too are related to cyanobacteria, provide nitrogen to their hosts,²⁴ and appear to be stably inherited.^{25,26} However, they are much younger than the nitroplast of *B. bigelowii* (UCYN-A), initially acquired by a diatom an estimated 12 to 34 million years ago,^{27,28} compared to the 100 million years nitroplasts in *B. bigelowii* have been around²⁹ (and the more than 1.6 billion years since the acquisition of the chloroplast³⁰). One would expect

that diatom nitroplasts have not progressed as much in the integration of organelle and host. If the diatom nitroplasts are sufficiently host-independent, it is conceivable that they could be transferred to a new host, perhaps even to a crop plant.

Consider also our seemingly favored moment in the course of evolutionary history. Chloroplasts evidently arose from a single acquisition of an ancient cyanobacterium. A second acquisition led to the chromophore of *Paulinella* a 100 million years ago. UCYN-A represents a third acquisition -- three cyanobacterial endosymbioses over the course of 1.6 billion years. Yet we are witness to multiple acquisitions in relatively recent times -- at least one cyanobacterium by a Rhopalodian diatom, another by a diatom within the genus *Climacodium*,³¹ a third of a heterocyst-forming cyanobacterium (the ancestor of *Richelia euintercellularis*) by a diatom within the genus *Hemiaulus*,³² and another of a marine *Synechococcus* by a dinoflagellate.³³ Perhaps it is just luck that we can observe four rare events in evolution that happen to have occurred not so distant from today.

Alternatively, it is possible that there is no favored time, that cyanobacterial acquisition is common,³³ just as is endosymbiosis by heterotrophic bacteria,³⁴ but they generally don't persist. Except for a few rare cases, they come and go. In that case, the recent acquisitions give us an opportunity to investigate why the ancestors of these endosymbionts were successful, at least for the moment. What cyanobacterial characteristics are favorable to endosymbiosis that persist at least as far as they have?

Genomes have now been sequenced from at least six endosymbionts from Rhopalodian diatoms.^{24,35,36,37,38} Each genome is much reduced from the genomes of related free-living cyanobacteria, 2.5Mb to 3.1Mb for the endosymbionts compared to about 5Mb for their closest cyanobacterial relatives.^{39,40} The availability of these genomes makes it possible to assess what are the metabolic capabilities of the endosymbionts, to what degree they are dependent on their diatom hosts, and what may be required for a successful transfer to a crop plant, without need to otherwise modify the plant. Furthermore, if each endosymbiont has to some extent lost genes independently from the others, then a comparison of their genomes offers the prospect of assessing what common genes may be under selection within the endosymbiotic environment.

Results

Genome characteristics and phylogeny of the nitrogen-fixing endosymbionts

Before comparing the genomes of the endosymbionts to gain insight into the endosymbiotic state, it is important to see how the genomes are related to each other and to those of free-living cyanobacteria (see Fig. 1 and Supplemental Table S1 for the genomes considered in this study and their abbreviations). One shouldn't place too much stock in the names of the host diatoms. The diatom genera *Epithemia* and *Rhopalodia* are both paraphyletic, with species of one intermixed with the other in phylogenetic trees.^{37,41} For this reason, I've lumped them together as 'Rhopalodian'.

Cyanobacterium	Nickname	Size(Mb)	GC%	Contigs	Genes	Density	Origin
<i>Crocospaera</i> ATCC 51142	Cro51142	4.93	37.9%	1	4817	87%	Marine, intertidal
<i>Crocospaera</i> plasmids		0.52	38.6%	5	542	86%	
<i>Crocospaera</i> CCY 0110	Cro0110	~5.88	36.7%	163	6520	86%	Marine sediment
<i>Crocospaera watsonii</i> WH 8501	Cro8501	~6.22	37.1%	315	6793	81%	Marine
<i>Gloeotheca</i> PCC 7424	Glo7424	5.94	38.6%	1	5284	82%	Soil of rice field
<i>Gloeotheca</i> plasmids		0.61	37.5%	6	483	76%	
<i>Gloeotheca</i> PCC 7822	Glo7822	6.09	40.2%	1	5482	84%	Soil of rice field
<i>Gloeotheca</i> plasmids		1.75	38.8%	6	1220	78%	
<i>Rippkaea</i> PCC 8801	Rip8801	~4.58	39.7%	29	4409	86%	Soil of rice field
<i>Rippkaea</i> PCC 8802	Rip8802	4.67	39.8%	1	4372	85%	Soil of rice field
<i>Rippkaea</i> plasmids		0.13	39.7%	4	124	76%	
<i>Synechocystis</i> PCC 6803 Kazusa	Syn6803	3.57	47.7%	1	3314	88%	Freshwater lake
<i>Synechocystis</i> plasmids		0.38	44.0%	7	408	83%	

Endosymbiont host	Nickname	Size(Mb)	GC% (Δ GC%)	Genes	Density	Origin
<i>Epithema pelagica</i>	EpSB	2.48	33.7% (-6.1%)	1730	66%	Marine
<i>Epithemia clementina</i>	EcSB	3.07	34.4% (-5.4%)	2050	60%	Freshwater
<i>Rhopalodia gibberula</i> *	RgSB	3.02	33.9% (-5.9%)	1957	61%	Pond
<i>Rhopalodia gibberula</i> 17Bon1*	RaSB	3.06	33.9% (-5.9%)	1973	57%	Mat, river
<i>Epithemia adnata</i> Bon19*	EaSB	2.78	33.7% (-6.1%)	1841	58%	Mat, river
<i>Epithemia turgida</i> *	EtSB	2.79	33.4% (-6.4%)	1993	58%	Lake
<i>Braarudosphaera bigelowii</i>	UCYN-A	1.44	31.1% (-6.8%)	1241	81%	Marine
<i>Paulinella chromatophora</i>	Paul-C	1.02	38.0% (-28.0%)	915	80%	Freshwater

Fig. 1 Genomes used in this study. Genome sources are given for the six Rhopalodian endosymbionts central to this study, along with those of eight closely related cyanobacteria. The genomes of two other cyanobacteria-derived endosymbionts are also listed. When the organism has plasmids in addition to the chromosome, information related to the chromosome is given in black and that related to the plasmids (taken as a whole) is given in red. In the three cases where the genome sequence is incomplete, no distinction can be made between chromosome and plasmid, and information is given in purple. Four endosymbionts (marked by asterisks) have small plasmids which have not been considered in this study. The gene density was calculated for a given entity by dividing as the sum of nucleotides participating in coding and non-coding genes by the number of nucleotides in that entity, where 'entity' is the chromosome or plasmids if the genome is complete or the entire genome, if incomplete. See Supplemental Table S1 for sources.

Fig. 2 shows a phylogeny based on 29 proteins with orthologs encoded in the genomes of all endosymbionts and closely related free-living cyanobacteria. It is concordant with trees based on hundreds of proteins^{38,42} but discordant with trees based on just the NifH protein plus 16s rRNA.^{24,37} All the Rhopalodian endosymbionts cluster together, with *Rippkaea* PCC 8801 (Rip8801) and PCC 8802 (Rip8802), the most closely related free-living cyanobacteria. They are separated

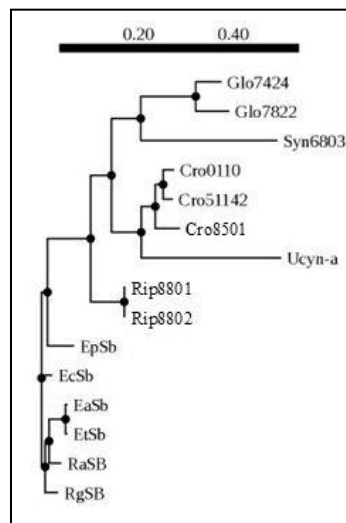


Fig. 2 Phylogenetic relationships amongst Rhopalodian endosymbionts and close relatives. The tree was estimated as described in Materials and Methods. Each bipartite node has a bootstrap value of 100%. The scale bar represents number of inferred substitutions per site.

from the well-studied nitroplast from *B. bigelowii*, UCYN-A, which is more closely related to members of the cyanobacterial genus *Crocospaera*. The tree implies that there were at least two primary acquisitions of cyanobacteria to form N₂-fixing endosymbionts – one leading to UCYN-A and its close relatives⁴³ and at least one leading to the Rhopalodian endosymbionts.

Four of the six Rhopalodian endosymbiont genomes are closely related to each other (Fig. 2). The genome most distant within this group, that of EpSB, comes from the lone marine diatom in the set of six, raising the possibility that the association might have arisen from a separate acquisition of a cyanobacterium similar to the progenitor of the other Rhopalodian endosymbionts. This idea is explored in the Discussion section.

The sizes of the genomes of the Rhopalodian endosymbionts range from 2.5 to 3.1 Mb, compared to the 4.6 Mb size of the genome of its closest free-living relative and the 1.4 Mb size of the genome of the non-Rhopalodian endosymbiont, UCYN-A (Fig. 1A). All of their genomes exhibit an decrease in GC% relative to their closest free-living relative. Both size reduction and reduced GC fraction are typical of obligate endosymbionts.^{34,44} In four cases (EaSB, EtSB, RaSB, RgSB), a small (~6 kb) plasmid was reported in addition to the chromosome. All four plasmids have a high degree of identity and share the same handful of genes.³⁸ It isn't known whether the remaining endosymbiont genomes (EcSB, EpSB) have plasmids.

However, the difference in genome sizes amongst the Rhopalodian endosymbionts is deceptive. Supplemental Table S2A lists the proteins and orthologous relationships of all the endosymbionts and Rip8802 and *Crocospaera* ATCC 51142 (Cro51142). Proteins that are conserved amongst all of the close relatives of the endosymbionts (Fig. 1), called here the **core** proteins (1779 proteins, indicated in Supplemental Table S2A, column **Core?**), appear in the endosymbiont genomes in approximately the same numbers, somewhat less in EtSB and EaSB (Fig. 3, Supplemental Table S3A). The apparent difference in protein number amongst the endosymbionts is explained by differences in the numbers of non-conserved proteins, particularly proteins unique to a genome, which are overwhelmingly biased towards small proteins (<100 amino acids). This characteristic suggests that much of the differences in the number of proteins amongst the endosymbionts may be artifactual, attributable to differences in gene-calling programs and their variable tolerance to small sizes. The differences in the genome sizes are probably better interpreted not as differences in coding capability but rather the degree to which broken genes have remained in the genome. This idea is considered later.

The availability of a set of core proteins makes it possible to gain insight into what conserved proteins that were almost surely present in the free-living ancestor(s) of the Rhopalodian endosymbionts proved to be unnecessary in the symbiotic state and disappeared. To address this issue, a second set, **No-RE**, was defined: those proteins found in free-living relative Rip8802 with no ortholog in any Rhopalodian endosymbiont (Supplemental Table S2A, column **status**). Its intersection with core proteins is given in Supplemental Table S3B, with highlights in Table 1. Some on the list are to be expected. The endosymbionts have lost the ability to fix carbon, so it is no surprise to see the absence of proteins related to Photosystems I and II, antenna, carboxysomes, and carbon fixation. Many of the other entries suggest loss of redundancies, e.g. loss of a conserved

Table 1 Proteins in Core set but not in Rhopalodian endosymbionts^a

Description	Number of proteins
carbon fixation / photorespiration	8
carboxysome proteins	12
chlorophyll biosynthesis	2
cofactor biosynthesis - biotin	1
cofactor biosynthesis - dihydrofolate	1
competence-related proteins	4
cyanophycin metabolism	2
external polysaccharide biosynthesis	14
glyceraldehyde 3-phosphate dehydrogenase (NAD(P)+), phosphorylating)	1
hydrogenase (reversible)	6
nitrate utilization	6
photosystem I	11
photosystem II	24
phycobiliprotein	16
phylloquinone biosynthesis	7
pili	7
superoxide dismutase (Fe-dependent)	1
tocopherol biosynthesis	3
tricarboxylic acid cycle	3
transport proteins	30
others	348

^a See Supplemental Table S3B for list of specific proteins

copy of dihydroorotate dehydrogenase (pyrimidine metabolism). Some of the other entries on the list are more interesting and will be discussed later in context.

Another useful set of proteins are those that are found in all Rhopalodian endosymbionts (*All-RE*) (Supplemental Table S2A, column **status**). One would expect that most proteins required for the symbiotic state would be contained in this group of 1436 proteins. The great majority of proteins encoded by the Rhopalodian endosymbionts are in this set (Supplemental Table S3A). Most of these are also in the Core proteins (Fig. 3), but those that are not Core proteins are of special interest, because they may point to

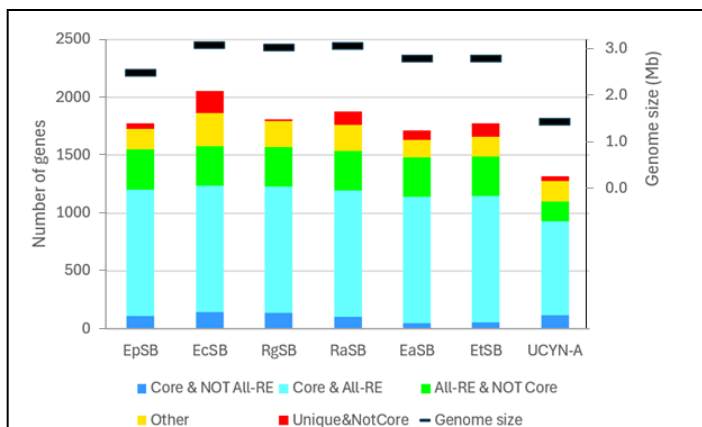


Fig. 3 Fraction of protein-encoding genes in different sets. For each endosymbiont genome, the numbers of protein-encoding genes in the indicated classes are stacked on top of each other. For comparison, the size of each genome is also shown, on a different scale. See Fig. 1 for abbreviations of the endosymbionts and genome sizes. See Supplemental Table S3A for counts of genes in each category.

special characteristics of the ancestor(s) of the endosymbionts that set it apart and enabled it to be successful in the association with its diatom host. Table 2 shows some of the proteins that are in All-RE but are not in Rip8801 and Rip8802, the closest relatives of the endosymbionts (the complete list of 39 such proteins is provided in Supplemental Table S3C). Most are found in one or more of the more distantly related *Crocospaera*. Many of these proteins will be discussed individually in the sections that follow.

Table 2 Proteins in all Rhopalodian endosymbionts but not in *Rippkaea* PCC 8802^a

Prototype ^b	Description	In <i>Crocospaera</i> ^c	Saline habitat ^d
RGRSB-1038	alpha-glucosidase	2	94%
RGRSB-0116	alpha-glucosidase (probable)	2	94%
RGRSB-1815	cytochrome c oxidase (ARTO) subunit I	3	29%
RGRSB-1816	cytochrome c oxidase (ARTO) subunit II	3	31%
RGRSB-1814	cytochrome c oxidase (ARTO) subunit III	3	31%
RGRSB-0743	glucosylglycerol-phosphate phosphatase	2	80%
RGRSB-0176	glucosylglycerol-phosphate synthase	2	73%
RGRSB-0175	glycerol-3-phosphate dehydrogenase (FAD)	2	67%
RGRSB-0112	glyceraldehyde 3-phosphate dehydrogenase (NADP, nonphosphorylating), GapN	1	80%
RGRSB-0442	Superoxide dismutase (Ni-dependent)	1	90%
RGRSB-0441	Superoxide dismutase (Ni-dependent) maturation protein	1	76%
RGRSB-1023	Transport: calcium/sodium antiporter	3	62%
RGRSB-0595	Transport: putative Na ⁺ /K ⁺ /2Cl ⁻ cotransporter	3	77%
RGRSB-0697	Transport: sodium-coupled permease	0	86%
RGRSB-0104	Transport: SulP family inorganic anion transporter	3	70%

^a See Supplemental Table S3C for full list of Rhopalodian orthologs and more granular habitat information

^b Protein from the endosymbiont of *Rhopalodia gibberula* is orthologous to proteins from the other endosymbionts

^c Count of orthologs in *Crocospaera* ATCC 51142, CCY 0110, and WH 8501

^d Fraction of 127 cyanobacteria with orthologs to the given prototype protein, with habitats listed as {"marine", "saline lake", "coastal", "brackish", "intertidal", or "estuary"} and not {"freshwater", "bog", "hot springs", "halo-intolerant", "terrestrial", "rock", "paddy", "sand", "soil", or "root"}

Most proteins within All-RE have orthologs in cyanobacteria heavily biased towards those isolated from saline environments (Table 2 and Supplemental Table S3C). Except for alternative cytochrome oxidase, between 62% to 94% of the orthologs of the proteins are from halophilic cyanobacteria. This is remarkable, because the 127 cyanobacteria considered are overwhelmingly terrestrial and freshwater. Only 34% come from saline environments, only 20% if one focuses on the half of the genomes most closely related to the endosymbionts (thereby excluding the large number of genomes from the marine picocyanobacteria).

Metabolic analysis – Informational proteins

Any endosymbiont needs to maintain its ability to propagate itself. Fig. 4 shows graphically that this minimal bar is likely met by the Rhopalodian endosymbionts without need for additional proteins from their hosts (details given in Supplemental Table S2B). The Rhopalodian endosymbionts encode all the required ribosomal proteins, machinery for transcription, and

Function	Rip8802	EpSB	EcSB	RgSB	RaSB	EaSB	EtSB	UCYN-A	Paul-A
Informational									
rRNA & tRNA	+	+	+	+	+	+	+	+	+
Ribosomal proteins	+	+	+	+	+	+	+	+	+
Aminoacyl-tRNA-synthases	+	+	+	+	+	+	+	95%	+
Transcription	+	+	+	+	+	+	+	+	+
DNA replication	+	+	+	+	+	+	+	+	75%
DNA repair	+	89%	89%	89%	89%	89%	89%	74%	37%
Energy production									
Glycolysis	+	90%+	90%+	90%+	90%+	90%+	90%+	90%	90%
Entner-Doudoroff	+	+	+	+	+	+	+	—	+
Pentose Phosphate	+	+	+	+	+	+	+	80%	80%
TCA (OAA,Pyr → 2-OG)	+	+	+	+	+	+	+	20%	20%
TCA (2-OG → OAA)	+	60%	60%	60%	20%	20%	20%	—	—
NAD(P)H oxidoreductase	+	+	+	+	+	+	+	+	+
Cytochrome b ₆ /f	+	+	+	+	+	+	+	+	75%
Cytochrome c oxidase	+	++	++	++	++	++	++	+	+
Heme biosynthesis	+	+	+	+	+	+	+	+	90%
ATP synthase	+	+	+	+	+	+	+	+	+
Photosystem I	+	—	—	—	—	—	—	+	64%
Photosystem II	+	—	5%	5%	5%	—	—	5%	+
Chlorophyll biosynthesis	+	80%	80%	80%	30%	—	—	90%	+
Antenna proteins	+	—	—	—	—	—	—	—	90%
Carotenoid biosynthesis	+	+	+	+	+	+	+	57%	86%
Nitrogen fixation									
Nitrogenase and cofactor	+	+	+	+	+	+	+	+	10%—
Ammonia assimilation	+	+	+	+	+	+	+	43%	29%
Metabolism									
Glycogen metabolism	+	+	+	+	+	+	+	20%	20%
Phospholipid synthesis	+	+	+	+	+	+	+	83%	83%
Fatty acid synthesis	+	+	+	+	+	+	+	+	+
Purine synthesis	+	+	+	+	+	+	+	35%	35%
Pyrimidine synthesis	+	+	+	+	+	+	+	88%	62%
Amino acid synthesis	Ala	+	+	+	+	+	+	+	—
	Arg	+	+	+	+	+	+	25%	38%
	Asn	+	+	+	+	+	+	+	+
	Asp	+	+	+	+	+	+	50%	—
	Cys	+	+	+	+	+	+	+	67%
	Gln	+	+	+	+	+	+	+	+
	Glu	+	+	+	+	+	+	+	—
	Gly	+	+	+	+	+	+	+	+
	His	+	+	+	+	+	+	—	50%
	Ile	+	+	+	+	+	+	—	88%
	Leu	+	+	+	+	+	+	—	83%
	Lys	+	+	+	+	+	+	+	+
	Met	+	+	+	+	+	+	50%	50%
	Phe	+	+	+	+	+	+	70%	80%
	Pro	+	+	+	+	+	+	33%	67%
	Ser	+	+	+	+	+	+	67%	+
	Thr	+	+	+	+	+	+	80%	80%
	Trp	+	+	+	+	+	+	50%	42%
	Tyr	+	+	+	+	+	+	60%	70%
	Val	+	+	+	+	+	+	—	+
Cofactor synthesis									
Biotin	80%	60%	60%	60%	60%	40%	40%	20%	20%
Folate	89%	56%	67%	67%	67%	67%	56%	56%	56%
Lipoic acid	+	+	+	+	+	+	+	—	+
NAD/NADP	+	+	+	+	+	+	+	50%	—
Pantothenate/CoA	+	78%	89%	89%	89%	89%	89%	44%	44%
Pseudocobalamin	86%	86%	86%	86%	86%	14%	14%	86%	5%
Pyridoxal 5-phosphate	71%	71%	71%	71%	71%	71%	71%	71%	57%
Riboflavin	83%	83%	83%	83%	83%	83%	83%	67%	—
Thiamine	75%	75%	75%	75%	75%	75%	75%	67%	33%

Fig. 4 Summary of metabolic capabilities of endosymbionts and related cyanobacterium. Green represents presence of pathway, red indicates absence, and intermediate colors indicate presence of some but not all enzymes and initial substrates. Percentages indicate the fraction of a pathway recognizable in the genome. Lists of the specific enzymes used in the calculation of percentages are given in Supplemental Table S2. If the percentage is followed by a “+” (or a cell contains an extra “+”), then the indicated pathway is supplemented with a capability that goes beyond that possessed by R8802. A cell surrounded by a red outline indicates that the hole in the pathway was filled by importing a protein encoded by the nucleus of *Braarudosphaera bigelowii* [Coale 2024] or *Paulinella chromatophore* [Singer]. If the outline is dashed, then only one of multiple holes was filled. Organism/endosymbiont abbreviations are explained in Fig. 1.

components for DNA replication. They have retained all but three of 27 proteins related to DNA repair,⁴⁵ even though such proteins are frequently lost in endosymbiosis.¹⁶ In two of those three cases, the function of the missing protein may be assumed by a different protein. In one of them, the missing photolyase PhrB, has the same activity as another photolyase, PhrA, that is found in the endosymbionts. The former has been shown to be of little functional importance in *Synechocystis*.⁴⁶ In the second case, RNase HII (absent in all endosymbionts) overlaps in function with RNase HI^{47,48} (present in all endosymbionts).

The situation is somewhat different with UCYN-A, which lacks a protein, methionyl-tRNA-formyltransferase, that is nearly universal protein amongst prokaryotes.^{49,50} The endosymbiont survives because the protein is one of many imported from its host. Similarly, the chromatophore from *Paulinella* imports from its host the essential DNA replication protein DNA ligase. UCYN-A also lacks seven DNA repair enzymes including the three also absent in the Rhopalodian endosymbionts.

Metabolic analysis – Energy production

Context and lessons learned from free-living cyanobacteria

The *Richelia*/diatom symbiosis devotes 22% of photosynthetic production by the host to support the endosymbiont,³² and it is likely that the same is true for the Rhopalodia. Apart from the general costs of maintaining existence – e.g. ATP to aminoacylate tRNAs for translation and NADPH for the synthesis of membrane lipids – the Rhopalodian endosymbionts bear the added burden of meeting the considerable energy needs of N₂-fixation. The process of fixing a single molecule of N₂ to two molecules of NH₃ requires 16 ATPs and 8 high-energy electrons.⁵¹ Their production almost surely takes place within the endosymbiont, and metabolic processes must be present to allow that to happen.

In addition, there is the cost of protecting the enzymes of nitrogen fixation from irreversible inactivation by oxygen.⁵² Cyanobacteria that form specialized N₂-fixing cells (called heterocysts) protect nitrogenase from inactivation by ambient oxygen and oxygen produced by photosynthesis in adjacent cells through a variety of measures.⁵³ Heterocysts have limited permeability to gases,⁵⁴ and the sites where gases can enter – the poles of the cells – are protected by a high concentration of cytochrome oxidase.^{53,55} Oxygen is also consumed in a light-dependent fashion⁵⁶ by a flavoprotein called Flv3,⁵⁷ and by an uptake hydrogenase that redirects the electrons of the H₂ produced by nitrogen fixation into the electron transport chain.^{58,59} This enables heterocysts to fix

nitrogen during the day using photosynthate from vegetative cells and during the night using photosynthate stored as glycogen.⁶⁰

Single-cell cyanobacteria closely related to Rhopalodian endosymbionts address the oxygen problem differently, fixing nitrogen only at night,⁶¹ at the expense of glycogen accumulated during the day from photosynthesis.^{61,62} Respiration may play a role in reducing the level of ambient oxygen to a level low enough to permit nitrogen fixation.⁶³

A consideration of how heterocyst-forming and N₂-fixing unicellular cyanobacteria address these problems may help in guiding us through the genomes of endosymbionts towards an understanding of their strategies. Heterocyst-forming cyanobacteria when fixing N₂ during the day can use Photosystem I (PSI)-dependent cyclic photophosphorylation to drive ATP production.^{64,65} They are still able to fix N₂ in the dark, albeit at a significantly reduced rate, so long as oxygen is present,⁶⁶ presumably required for oxidative phosphorylation. That process appears to be essential for ATP production in nitrogen-fixing unicellular cyanobacteria as well.^{62,67}

NADPH is necessary to reduce (directly or indirectly) the ferredoxin used by nitrogenase as the electron donor. In heterocyst-forming cyanobacteria, NADPH is produced primarily through the pentose phosphate pathway.⁶⁴ In Cro51142, the genes encoding enzymes in the pentose phosphate pathway are highly expressed at the onset of darkness, preceding peak nitrogenase expression,⁶⁸ consistent with a central role in supporting nitrogen fixation.

Energy production and oxygen protection in endosymbionts

An overview of the metabolic capabilities of the endosymbionts is presented in Fig. 4, and Fig. 5 shows a representation of part of central metabolism. While UCYN-A fixes nitrogen solely during daylight hours⁶⁹ and lacks enzymes for glycogen metabolism, all Rhopalodian endosymbionts have the coding capacity for enzymes of glycogen metabolism (Supplemental Table S2C and Fig. S01), and at least some fix in daylight and to some extent at night. EpSB fixes mostly but not exclusively during the day, and EcSB fixes continuously without regard to light.^{22,24,37}

The time at which nitrogen fixation takes place is important, because it affects the possibilities of how the ATP required for nitrogen-fixation is produced (Fig. 5 and Supplemental Table S2D). UCYN-A has retained PSI proteins and therefore, like heterocysts, can use cyclic photophosphorylation to produce ATP. The Rhopalodian endosymbionts have lost PSI proteins and must rely primarily on oxidative phosphorylation. Glycolysis is not available even for a minor contribution to the level of ATP because all endosymbionts lack the ability to encode the central enzyme phosphofructokinase (Supplemental Fig. S02). The same is true in many free-living cyanobacteria.^{70,98} However, the Entner-Doudoroff pathway is an alternate route to take glucose-6-phosphate to pyruvate, one that is probably of greater quantitative importance in cyanobacteria than glycolysis.^{70,71} This pathway is present in all Rhopalodian endosymbionts (but not in UCYN-A) (Supplemental Table S2D and Fig. S03).

The tricarboxylic acid cycle is present only partially in the Rhopalodian endosymbionts – from oxaloacetate to 2-oxoglutarate -- and from there to succinate in RgSB, EcSB, and EpSB (Fig. 5 and Supplemental Table S2D and Fig. S04). It is entirely absent in UCYN-A. The primary purpose

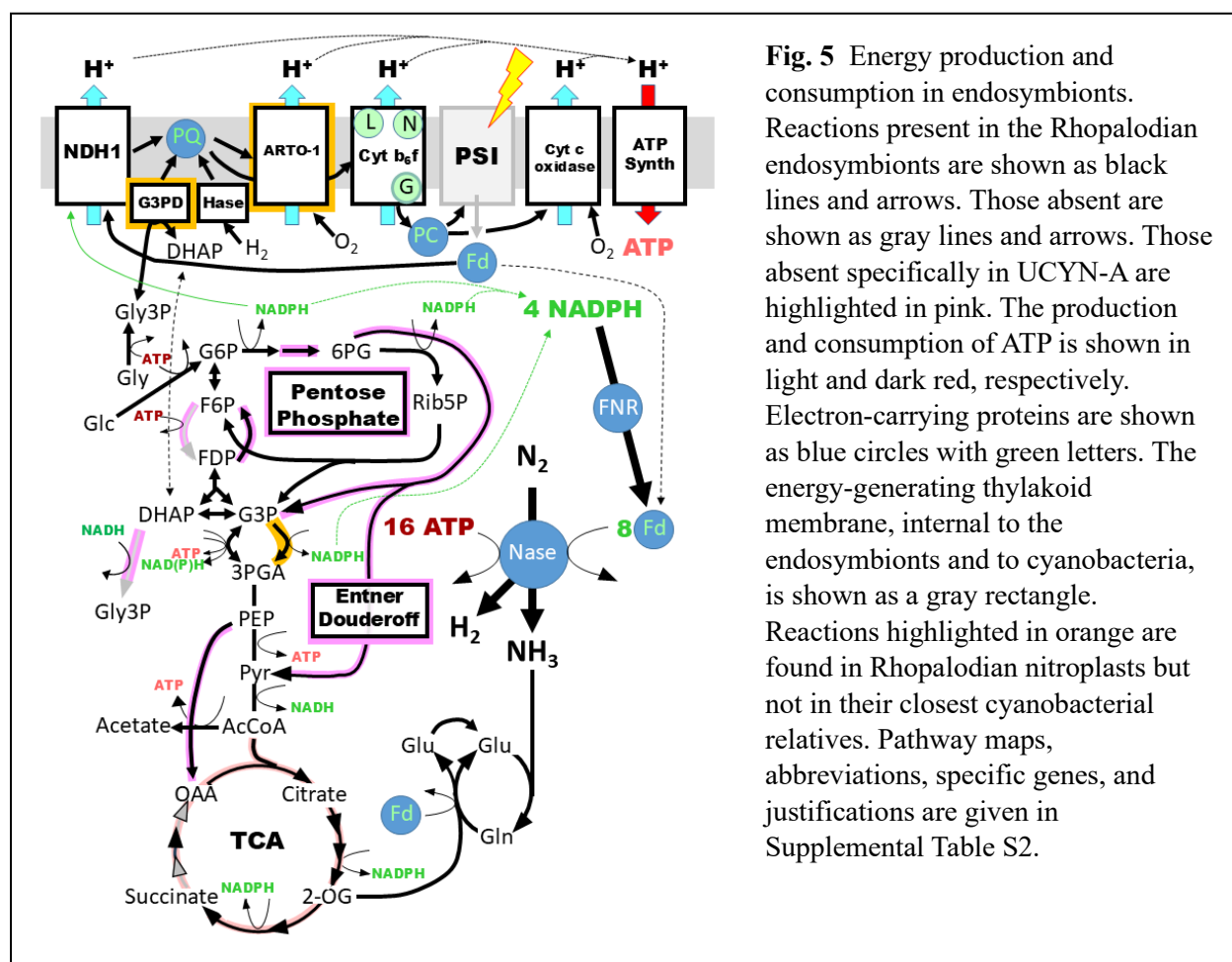


Fig. 5 Energy production and consumption in endosymbionts. Reactions present in the Rhopalodian endosymbionts are shown as black lines and arrows. Those absent are shown as gray lines and arrows. Those absent specifically in UCYN-A are highlighted in pink. The production and consumption of ATP is shown in light and dark red, respectively. Electron-carrying proteins are shown as blue circles with green letters. The energy-generating thylakoid membrane, internal to the endosymbionts and to cyanobacteria, is shown as a gray rectangle. Reactions highlighted in orange are found in Rhopalodian nitroplast but not in their closest cyanobacterial relatives. Pathway maps, abbreviations, specific genes, and justifications are given in Supplemental Table S2.

of the pathway is probably to produce the 2-oxoglutarate required by glutamine synthetase/GOGAT, the key enzymes of ammonia assimilation.⁷² However, the reductant produced by the partial tricarboxylic acid cycle may also contribute to ATP production by oxidative phosphorylation.

The pentose phosphate pathway is intact in all the endosymbionts (Fig. 5 and Supplemental Table S2D and Fig. S03). This includes UCYN-A, which does not encode its own 6-phosphogluconolactonase but imports it from its host.¹⁷ The expression of one central enzyme (glucose-6-phosphate dehydrogenase) has been measured in one endosymbiont and found to be extraordinarily high.²⁴ While this pathway is surely the primary source of NADPH in the endosymbionts, a significant quantity of NADPH may be produced also by an enzyme, GapN,⁶⁵ found in all Rhopalodian endosymbionts but rarely in free-living cyanobacteria. The enzyme is similar in critical sequence characteristics to a version of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) that is unidirectional and generates NADPH but no ATP in the reaction.^{73,74} UCYN-A does not encode this enzyme. In wheat, a similar enzyme in conjunction with the conventional GAPDH has been associated with the coordination of NADPH and ATP production.⁷⁵

By fixing nitrogen at the same time their hosts are producing O₂ through photosynthesis, the Rhopalodian endosymbionts face challenges greater than those faced by their closest free-living relatives. Perhaps for this reason, they possess an alternative respiratory terminal oxidase (ARTO), type 1, in addition to the conventional cytochrome oxidase⁷⁶ (Supplemental Table S2D). In the heterocyst-forming cyanobacterium, *Anabaena* PCC 7120, ARTO is expressed only in heterocysts and is required for nitrogenase activity.⁷⁷ The endosymbionts may also use ARTO to protect nitrogenase from O₂, supplementing the effect of the unidirectional, O₂-consuming uptake hydrogenase,⁷⁸ encoded by the genomes of most cyanobacterial N₂-fixers⁷⁹ and by all of the endosymbionts. All the endosymbionts lack flavodiiron proteins, e.g. Flv3, and so oxygen protection is unlikely through a Mehler-like reaction (reduction of O₂ to H₂O).^{57,80} A true Mehler reactions (the reduction of O₂ to H₂O₂) probably does not occur in any cyanobacterium⁸⁰ or endosymbiont.

Even though all the Rhopalodian endosymbionts lack PSI and PSII proteins, three of them (RgSB, EcSB, and EpSB) nonetheless encode all but three enzymes involved in chlorophyll biosynthesis^{81,82} (Supplemental Table S2D; see also Ref. 36). Why has the partial pathway persisted? The first missing enzyme in the pathway is Lpor, light-dependent protochlorophyllide reductase, conserved in the set of core proteins. However that enzyme can be bypassed by the presence of ChlLNB, light-independent protochlorophyllide reductase,⁸³ present in RgSB, EcSB, and EpSB. The endosymbionts also lack BciB, a reductase dispensible in low light.⁸⁴ Therefore, the absence of these two enzymes is probably not consequential, so long as a high rate of biosynthesis is not necessary.

However, RgSB, EcSB, and EpSB almost surely do not make chlorophyll. For one thing, EcSB at least does not have large amounts of chlorophyll as judged by fluorescence.²⁴ More importantly, the three endosymbionts lack the final enzyme in the pathway, geranylgeranyl diphosphate reductase (ChlP), containing pseudogenes instead. This enzyme converts the geranylgeranyl moiety within chlorophyllide *a* to the phytol tail of chlorophyll *a*. Without this enzyme, *Synechocystis* PCC 6803 cannot grow photoautotrophically under any light regime,⁸⁵ indicating that chlorophyllide *a* cannot substitute for chlorophyll *a* in photosynthesis.

Why maintain the seemingly defective pathway? The answer may lie in a proposal by Vavilin and Vermaas that free chlorophyll and precursors not bound to photosynthetic protein complexes may regulate one or more steps in the biosynthesis of tetrapyrroles.⁸¹ Tetrapyrroles is a class that includes, heme, chlorophyll, and pseudocobalamin.⁸² *Synechocystis* has a set of Small Chlorophyll-*a*-binding Proteins (Scp) with a conserved chlorophyll-binding region at their C-termini. Ferrochetalase (HemH), at the crossroads between heme and chlorophyll biosynthetic pathways, has a similar C-terminus.⁸⁶ The binding of chlorophyll or a precursor to the binding-region was postulated to regulate the transcription or translation of upstream enzymes,⁸¹ and perhaps this is true also in RgSB, EcSB, and SpSB.

The presence or absence in endosymbionts of Scps or extended HemH correspond roughly to their chlorophyll biosynthetic capacities (Supplemental Fig. S09). Of the three Rhopalodian endosymbionts with a nearly complete chlorophyll biosynthesis pathway, one (EpSB) encodes HemH with a conserved C-terminal extension, One (EcSB) has an ortholog of ScpB from

Synechocystis, with its conserved C-terminus. One (RgSB) encodes HemH with a C-terminal extension, although that terminus is dissimilar to C-termini from *Synechocystis*. Of the three endosymbionts with little or no chlorophyll-biosynthetic pathway, two (EaSB and EtSB) lack any Scp, and their HemH proteins have no C-terminal extension. The third, RaSB also lacks an extension on its HemH protein, but it has an ortholog of ScpE whose C-terminus shows six of the eight conserved residues. RaSB also has orthologs of 3 of 12 chlorophyll biosynthetic enzymes, so its pathway may be in an intermediate state of degradation. In summary, it is plausible that chlorophyllide *a* serves a regulatory role in some endosymbionts, not as a precursor to chlorophylls.

Metabolic analysis – Central metabolites and cofactors

The greatest distinction between the capabilities of the Rhopalodian endosymbionts and UCYN-A as implied by their genomes is in the biosynthesis of central metabolites and cofactors. The genomes of the Rhopalodian endosymbionts imply complete biosynthetic pathways for purines, pyrimidines, fatty acids, phospholipids, and all 20 canonical amino acids (Fig. 4 and Supplemental Tables S2D, S2E, and S2F; see also Ref. 74).

The situation is quite different with UCYN-A, which lacks in its genome genes encoding all the enzymes of de novo biosynthesis of purines and most amino acids, to name only a few deficiencies. In some cases, metabolic holes are known to have been plugged by importing key enzymes from the algal host. Examples include threonine synthase and phosphoserine phosphatase, to complete the synthesis of threonine and serine, respectively, and orotate phosphoribosyltransferase to complete nucleotide biosynthesis.¹⁷ In other cases, it must rely on the import of metabolites, e.g. through the purine salvage pathways, which remain intact. Since the UCYN-A genome implies no capacity to make the tricarboxylic acid cycle enzymes leading to 2-oxoglutarate (required for glutamate production), and since there's no evidence of import of the enzymes,¹⁷ the endosymbiont must take in exogenous 2-oxoglutarate or glutamate. It must also procure a great many other metabolites,⁸⁷ well beyond the range of the transporters encoded by its genome.

There is a similar distinction between the Rhopalodian endosymbionts and UCYN-A with respect to cofactor biosynthesis (Fig. 4 and Supplemental Table S2G and Figs S5-S8; see also Ref 74). However, analysis is complicated by the imperfect knowledge of cyanobacterial cofactor biosynthetic pathways in general. At times, the most that can be said is that an endosymbiont's deduced pathway is no more incomplete than that of free-living cyanobacteria, who surely make the cofactor. In this light, the Rhopalodian endosymbionts have as complete pathways as can be expected for the biosynthesis of six of the nine cofactors considered. In contrast, UCYN-A has only two cofactor pathways reasonably complete.

Of the three incomplete pathways, one, pantothenate/Coenzyme A biosynthesis,^{88,89} may not deserve to be on the list (Supplemental Table S2G and Figs S7). Only one enzyme is missing in the Rhopalodian endosymbionts – aspartate 1-decarboxylase (PanD), leading to the production of β -alanine. That enzyme is also missing from all pico-cyanobacteria and 24% of other cyanobacteria. It's a safe bet that all of them make Coenzyme A, so they probably have adopted a different route to produce β -alanine⁸⁹ or may take up the small amount of β -alanine needed for

pantothenate biosynthesis from their environments. Whatever solution they found may also be available to the Rhopalodian endosymbionts, perhaps importing it through the N-II transporter⁹⁰ found in all endosymbionts (Supplemental Table S4B). EpSB, however, is defective in a second enzyme in the pathway, PanB, a core enzyme encoded by the other endosymbionts.

A second incomplete pathway, biotin biosynthesis,⁹¹ is more troublesome (Supplemental Table S2G and Fig. S05). The pathway may be divided into two stages: (a) the production of pimelate esterified to either acyl-carrier protein or coenzyme A, and (b) the formation of the two rings of biotin. The first stage is catalyzed by diverse sets of enzymes. Neither the pathway used by *Escherichia coli* nor the different pathways used by *Bacillus subtilis* use enzymes with orthologs in any cyanobacterium, and it is unknown how any cyanobacterium accomplishes the first stage of biotin biosynthesis. The second stage, in contrast, is highly conserved amongst bacteria⁹¹ and cyanobacteria. Orthologs of BioF, either BioU or BioA, BioD, BioB, and BirA are found in all or nearly all cyanobacteria, completing the pathway to protein-bound biotin. The Rhopalodian endosymbionts have these proteins as well, except (1) BioF is absent in all the endosymbionts, and (2) two related endosymbionts (EaSB and EtSB) lack BioU and BioB. While the biotin biosynthesis pathway is likely to be non-functional in these two endosymbionts, it is possible that the pathway in the other four retain function, if they are able to import the product of BioF (7-amino-8-oxononanoate) from their hosts. The endosymbionts may also import host-synthesized biotin through a protein found in all of them (as well as in UCYN-A) that is orthologous to a proven biotin transporter from *Rhodobacter capsulatus*⁹² (also orthologous to *Synechocystis* protein Slr1365).

The third incomplete cofactor pathway, folate biosynthesis,⁹³ poses problems at one entry point and at its final step (Table S12, Fig. S06). The synthesis of folate requires the production of p-aminobenzoate (PABA). This is conventionally achieved through three reactions starting from chorismite, catalyzed by PabAabc (Also called PabA/PabB and PabC). All the Rhopalodian endosymbionts lack orthologs for the first protein, PabAa, even though orthologs are found in the core set. If the endosymbionts are able somehow to acquire PABA, then the remainder of the pathway is fine until the step leading to the production of tetrahydrofolate, the cofactor required for several enzymatic reactions. The enzyme used by *E. coli* for this final step, dihydrofolate reductase (FolA; b0048), and by most bacteria,⁹³ is similar to proteins found in only 15% of cyanobacteria, none closely related to the endosymbionts. It is not clear how most cyanobacteria catalyze the final reduction to tetrahydrofolate, but many alternatives are possible.⁹³ As a fallback position, the endosymbionts may be able to take up folate or 5-formyl-tetrahydrofolate from their hosts, using transporters orthologous to a *Synechocystis* protein (Slr0642) with proven ability to transport both.⁹⁴ The protein is found in all Rhopalodian endosymbionts but not in UCYN-A.

The role and synthesis of pseudocobalamin warrants special attention, because there is reason to believe that some of the endosymbionts provide their hosts this cofactor in addition to fixed nitrogen. Cobamides are required cofactors in several reactions and are synthesized only by bacteria.⁸² In most cases, the biologically relevant cobamide is cobalamin (Vitamin B₁₂). However, almost all cyanobacteria encode enzymes that lead not to cobalamin but to pseudocobalamin

414 instead,⁹⁵ which differs from cobalamin in the presence of adenine in place of
415 2,6-dimethylbenzamidazol (DMB) as a ligand.

416 While land plants generally rely on cobalamin-independent enzymes, animals and some algae must
417 take up exogenous cobalamin for survival.⁹⁶ This is true for 62% of diatoms and 64% of
418 haptophytes.⁹⁶ Some can also grow when provided with exogenous pseudocobalamin but only
419 when supplemented with DMB.⁹⁵ Coale et al (2024)¹⁷ showed that *B. bigelowii*, the host of
420 UCYN-A, expresses two nuclear-encoded transcripts related to cobalamin biosynthesis
421 (Supplemental Fig. S08), even though the alga encodes no other enzymes of cobamide
422 biosynthesis:

- 423 • KC1-P2-N_k31_Locus_18619_Transcript_1_1: Encodes BluB (5,6-dimethylbenzimi-
424 dazole synthase, which makes DMB
- 425 • KC1-P2_N3_k31_Locus_9846_Transcript_1_1: Encodes conventional CobT (DMB
426 phosphoribosyltransferase), which prepares DMB for attachment to the cobamide.

427 This raises the possibility that the host is able to remodel the pseudocobalamin (or an intermediate)
428 produced by UCYN-A to meet its need for cobalamin. Conceivably, the Rhopalodian hosts have
429 discovered the same trick.

430 Like UCYN-A, four Rhopalodian endosymbionts (RgSB, RaSB, EcSB, and EpSB) encode
431 complete biosynthetic pathways to produce pseudocobalamin, to the extent that the steps are
432 currently understood (Supplemental Table S2G and Fig. S08). Genes encoding the enzymes for
433 two steps (SirC and CobC) have not yet been found in any cyanobacterial or endosymbiont
434 genome. CobC encodes a phosphatase that catalyzes the final step in cobalamin and
435 pseudocobalamin biosynthesis. There is confusion in the literature concerning the other missing
436 enzyme, SirC, encoding precorrin-2 dehydrogenase. In *E. coli*, its function is handled by CysG, an
437 enzyme with three distinct regions, each catalyzing a different reaction.⁹⁷ These regions correspond
438 to three separate enzymes in *Bacillus subtilis* (SirA, SirC, and SirB), catalyzing reactions leading
439 to seroheme.⁹⁸ The first two are also essential for the synthesis of cobalamin (SirA is synonymous
440 with CobA).⁸² While CysG bears a high degree of similarity to cyanobacterial enzymes, the
441 similarity lies only in two of the three regions, not the region corresponding to SirC. Since many
442 cyanobacteria with these uncertainties demonstrably produce pseudocobalamin,⁹⁵ it is likely that
443 this is true as well of RgSB, RaSB, EcSB, EpSB, and UCYN-A.

444 It remains a mystery how EaSB and EtSB manage to survive, unable to synthesize
445 pseudocobalamin but still evidently relying on two pseudocobalamin-dependent enzymes:
446 methionine synthase (MetH) and epoxyqueuosine reductase (QueG; used to modify tRNA).⁹⁹ Both
447 are present in all Rhopalodian endosymbionts. However, unlike many bacteria⁹⁹ and most
448 cyanobacteria, the endosymbionts and their close relatives use the cobamide-independent form of
449 ribonucleotide reductase. Perhaps EaSB and EtSB can scrounge enough pseudocobalamin from
450 the aqueous environment,^{95,100} as does the pseudocobalamin auxotroph *Synechococcus*
451 PCC 7002.¹⁰¹ It is less likely that the two endosymbionts make use of cobalamin in place of
452 pseudocobalamin. Cyanobacterial MetH binds cobalamin very poorly,¹⁰² and mutation would need

to have taken place to enable the two endosymbionts to share the cobalamin used by its diatom host.

Carbon transport

One might take away from the previous sections that the Rhopalodian endosymbionts are independent of their hosts with respect to major metabolic and informational categories. However, they are assuredly dependent in one important area. By having lost photosystems I and II and various proteins important in carbon fixation, they necessarily rely on their hosts for a source of carbon, both to synthesize intermediary metabolites and to power metabolism. It is important to determine to the extent possible what that carbon source is. The discussion below is summarized

in Fig. 6 and Supplemental Table S4A, and a more detailed discussion of this question is available elsewhere regarding RgSB and EtSB.⁷⁴

Glycerol and other simple sugars

Perhaps the first candidate for a carbon source is glycerol. Free-living Cro51142 can be trained to grow well in the dark on 50 mM glycerol. Glucose does not support growth, and its ability to enhance growth in the light is small or non-existent.^{103,104} All the Rhopalodian endosymbionts (but not UCYN-A) have the two enzymes necessary to take glycerol to the glycolytic intermediate dihydroxy acetone phosphate : glycerol kinase and FAD-dependent glycerol-3-phosphate dehydrogenase (Fig. 6 and Supplemental Table S4A). In fact, the endosymbionts are better placed than most free-living strains, such as S8802, which have only NAD(P)-dependent glycerol-3-phosphate dehydrogenase. That form energetically favors the **production** of glycerol-3-phosphate rather than its **consumption**.

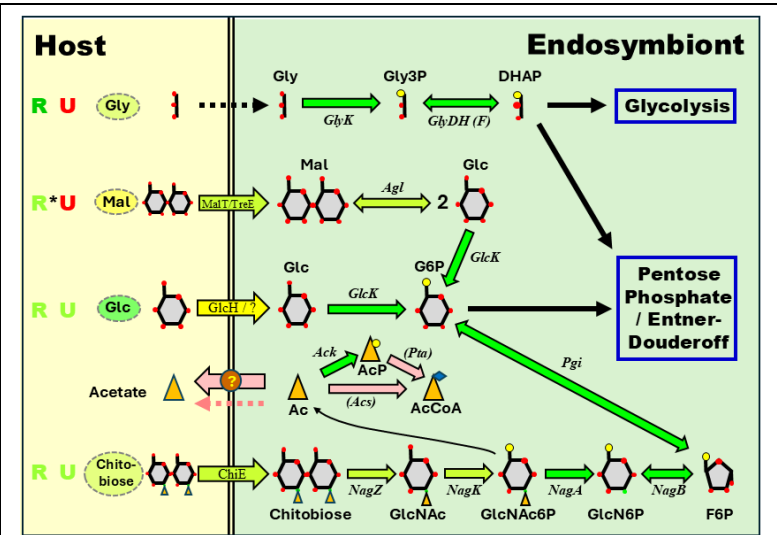


Fig. 6 Candidate pathways for uptake of carbon from host to endosymbiont. See Supplemental Table S4A for list of enzymes, transporters, their abbreviations, and associated genes in the different endosymbionts. Arrows across the double line indicate transporters at a membrane interface (double line) between an endosymbiont and the host. A dashed arrow indicates diffusion of the compound. The colors of the arrows indicate the perceived likelihood that the indicated reaction takes place in Rhopalodian endosymbionts: Green (very likely), yellow-green (possibly), yellow (hypothetical), pink (no evidence, where evidence would be expected). The colors of the ovals in the host compartment indicate in the same way the perceived likelihood that the compound is present in diatom hosts. The colors of “R” and “U” indicate the overall perceived likelihood that the route is present in Rhopalodian endosymbionts or UCYN-A, respectively. An asterisk indicates that the route is surely not present in certain endosymbionts (see Supplemental Table S4A). Abbreviations: Gly, glycerol; Gly3P, glycerol-3-phosphate; DHAP, dihydroxyacetone phosphate; Mal, maltose; Glc, glucose; G6P, glucose-6-phosphate; Ac, acetate; AcP, acetate phosphate; AcCoA, acetyl Coenzyme A; GlcNAc, N-acetylglucosamine; GlcNAc6P, N-acetylglucosamine-6-phosphate; GlcN6P, glucosamine-6-phosphate; F6P, fructose-6-phosphate.

What remains to be established is whether the diatom hosts produce glycerol at a level high enough to enable sufficient uptake by the endosymbionts. Glycerol can enter cells by diffusion, facilitated transport, or active transport.¹⁰⁵ In *E. coli*, a glycerol concentration above 1mM is sufficient to promote growth, unaffected by the presence or absence of a glycerol facilitator protein,¹⁰⁶ and so if the host concentration of glycerol exceeds this amount, then a transporter may not be necessary. Is this likely? Glycerol is used as an osmolyte in some halophilic algae¹⁰⁷ and certain halotolerant diatoms.^{108,109} One of the latter increased its internal concentration of glycerol from 5mM during growth in 150mM salt to 70mM during growth in 700mM salt (a bit higher than the level in seawater).¹⁰⁸ However, one might not expect freshwater diatoms (e.g., the hosts of five of the six endosymbionts considered here) to maintain an appreciable level of glycerol.

Perhaps glycerol is the answer to the carbon-source puzzle, but if not, then sugars come to mind as obvious candidates. This is because glucose, fructose, and sucrose are preferred carbon sources by cyanobacteria that are good heterotrophs.^{110,111} One of them, *Nostoc punctiforme*, requires the glucose transporter GlcP for the establishment of symbiosis with the hornwort *Anthoceros punctatus*.¹¹² Unfortunately, good matches to the Nostoc protein is found in none of the closest relatives of the endosymbionts, and no ortholog is present in any Rhopalodian endosymbiont nor in UCYN-A. The endosymbionts do possess orthologs of GlcH, a protein from the marine picocyanobacterium *Prochlorococcus* ss120 that supports glucose uptake.¹¹³ However, it appears not to have that function in cyanobacteria more closely related to the endosymbionts.⁷⁴ Other routes by which glucose, fructose, or sucrose might serve as a carbon source are unlikely to be pertinent in the endosymbionts.⁷⁴

Maltose

Three other sugars warrant consideration as possible carbon sources. Maltose ($\alpha(1\rightarrow4)$ glucose-glucose disaccharide) is an intriguing possibility (Fig. 6 and Supplemental Table S4A), because four endosymbionts carry proteins orthologous to a proven maltose transporters (MalT) from three diverse heterotrophic bacteria.^{114,115,116} These three members of the major facilitator superfamily (MFS) share 126 positions of identity, of which 71%-72% are shared as well with the endosymbiont orthologs (Supplemental Fig. S10). In contrast, other members of the MFS family that transport lactose or galactose match only 18% of these conserved positions, so the endosymbiont proteins are similar only to MFS proteins that transport maltose.

If we accept that some endosymbionts can take up maltose (not EaSB or EtSB), then they should have a means of metabolizing the sugar. Two routes are known: maltose phosphorylase, which produces glucose-1-phosphate plus unphosphorylated glucose; and α -glucosidase, which (if it acts on maltose) produces two molecules of glucose. No endosymbiont or close relative contains anything similar to the proven maltose phosphorylase from *Bacillus subtilis*,¹¹⁷ but all Rhopalodian endosymbionts encode two proteins (named here aGal1 and aGal2) that may substitute, generally annotated as α -glucosidase (Supplemental Table S4A). It remains to be determined experimentally whether either one acts on $\alpha(1\rightarrow4)$ linkages (such as those in maltose or instead on $\alpha(1\rightarrow6)$ linkages (such as those found in branched polysaccharides). However there is good reason to believe this is the case, the sequences of enzymes with known $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ activities^{118,119,120,121} were collected, aligned, and used as the basis of a phylogenetic tree

(Supplemental Fig. S11). It is apparent that the aGal1 α -glucosidases cluster with proven $\alpha(1\rightarrow4)$ glucosidases, some of which are known to act on maltose, while the aGal2 α -glucosidases cluster with a mixed collection of enzymes. However, Okuyama et al (2016) found that in their set of enzymes, the linkage specificity of the enzymes could be accurately predicted by the amino acid next to a universally conserved catalytic aspartate residue, alanine or threonine predicting $\alpha(1\rightarrow4)$ linkage specificity and valine predicting $\alpha(1\rightarrow6)$ linkage specificity.¹²⁰ Both aGal1 and aGal2 have alanine in the critical position, so possibly both enzymes act on $\alpha(1\rightarrow4)$ linkages.

The endosymbionts can take up and metabolize maltose only if it is present outside their membranes. Here there is a problem. Maltose is common in land plants, which use starch as a storage form. But diatoms instead store glucose as chrysolaminarin,¹²² a mixed β -1,3- and β -1-6-glucan. Breakdown of the polymer should produce not maltose but glucose and β -linked oligosaccharides.¹²² Maltose would therefore not be expected to be available to the endosymbionts.

Amino sugars chitobiose and N-acetylneuraminic acid

The two remaining candidates to discuss are both amino sugars: the disaccharide chitobiose, consisting of two molecules of N-acetylglucosamine (GlcNAc) in a β -1 \rightarrow 4 linkage, and the simple sugar N-acetylneuraminic acid (NeuN5Ac) (Fig. 6 and Supplemental Table S4A). In both cases, the same questions arise: (1) Can the compound be transported into the endosymbiont? (2) Can the endosymbiont metabolize it? and (3) Is the host likely to provide the compound?

In the case of chitobiose, the answers to all three questions would seem to be the same: “probably”. Some diatoms (including one of the same order as the Rhopalodia) have been found capable of synthesizing chitin, a β -(1 \rightarrow 4)-linked polymer of GlcNAc,¹²³ and one might expect to find its breakdown products, including chitobiose, in such diatoms. If the endosymbionts hosts are amongst this group, then it is very likely that the endosymbionts will be able to take up the compound. This is because GlcNAc along with N-acetylmuramic acid form the backbone of peptidoglycan in the cell walls of bacteria, including cyanobacteria.^{124,125} Peptidoglycan has been observed in UCYN-A¹²⁶ and surrounds probably all the endosymbionts. The recycling of peptidoglycan components is common amongst all bacteria¹²⁷ and involves the uptake of short GlcNAc-containing disaccharides and other compounds from outside the cell.¹²⁸

All of the endosymbionts (including UCYN-A) and almost all cyanobacteria appear to possess the enzymes required to convert chitobiose into glycolytic intermediates. The process begins with a β -N-acetylglucosaminidase (NagZ), which can release GlcNAc from a peptidoglycan fragment.^{129,130} An enzyme usually annotated ambiguously as “ β -glucosidase”, exhibits in all endosymbionts and almost all cyanobacteria an amino acid motif diagnostic of β -N-acetylglucosaminidases.¹³¹ In other organisms, some, but not all of these enzymes are able to act on chitobiose (GlcNAc-GlcNAc).^{129,132,133} In the endosymbionts, NagZ lacks a detectable transit peptide, so it would be predicted to act within the cell, like NagZ from Gram-negative bacteria and unlike those from Gram-positive bacteria.¹³⁴ The remaining steps are catalyzed by GlcNAc kinase (NagK), GlcNAc-6-phosphate deacetylase (NagA), and GlcNAc-6-phosphate deamidase (NagB). The latter two appear to be present in all Rhopalodian endosymbionts, while NagK is probably accounted for by an ROK-family enzyme.⁷⁴

The endosymbionts are certainly able to metabolize exogenous peptidoglycan-derived GlcNAc-disaccharides, and should therefore be able to take them up -- perhaps chitobiose as well. This may be achieved through a transporter found in all endosymbionts and most cyanobacteria (Supplemental Table S4A). The most similar transporter in the Transporter Classification Database (TCDB)¹³⁵ is ChiE from *Thermotoga maritima*.¹³⁶ This was judged to be a chitobiose transporter, because: (a) *T. maritima* is able to grow on chitobiose, (b) it is regulated by a protein that binds chitobiose, and (c) It is encoded by a gene within an operon that also contains genes encoding NagZ, NagA, and NagB, suggesting a functional relationship with the enzymes that metabolize chitobiose.

There is also circumstantial evidence connecting ChiE to chitobiose in the endosymbionts. ChiE is the substrate-binding component of an ABC transporter.^{137,138} It lies adjacent in the genome of *T. maritima* to two genes encoding permeases, NagF and NagG, presumably participating in the Chi ABC transporter. Orthologs to the two permeases are found in all endosymbionts. In UCYN-A, one lies nearby the genes encoding NagK and NagZ. All in all, NagF and NagG probably combine with ChiE to form an intact transporter. While chitobiose is not a breakdown product of peptidoglycan, the transporter that brings in GlcNAc-disaccharides may have been repurposed to transport chitobiose as a high volume carbon source for the endosymbionts.

The appeal of the second amino sugar candidate, NeuNAc, lies primarily in the fact that all Rhopalodian endosymbionts have orthologs of NanT, the putative transporter of NeuNAc, even though orthologs are absent in the endosymbionts' closest cyanobacterial relatives and are rare in cyanobacteria in general. However, the presence of NeuNAc has not been described in diatoms, and the metabolism of NeuNAc within the endosymbionts pose certain problems, discussed previously.⁷⁴

Nitrogen transport

N₂-fixing cyanobacteria release very little fixed nitrogen into their environments,¹³⁹ but substantial release is exactly what diatom hosts are asking of their endosymbionts. We have two models of the export of fixed nitrogen by free-living cyanobacteria. One is the export of primarily glutamine and aspartate-arginine from heterocysts to adjacent vegetative cells.^{140,141} The currently prevailing view is that this export takes place through physical junctions between adjacent cells,¹⁴² requiring several proteins, e.g. SepJ, FraC, and FraD.¹⁴⁰ Orthologs to these proteins are not found in unicellular cyanobacteria nor in the endosymbionts.

A more useful model may be the export of nitrogen from symbiotic heterocyst-forming cyanobacteria to their hosts, within extracellular associations (with cycads, bryophytes, ferns, and diatoms)^{139,143} and intracellular associations (with the angiosperm *Gunnera*¹⁴⁴ and the diatom *Hemiaulus hauckii*¹⁴³). In these cases (with the possible exception of the Nostoc-cycad symbiosis¹⁴⁵), the host gains nitrogen in the form of NH₄⁺, made possible by the low activity of glutamine synthetase.¹⁴⁶ Glutamine synthetase catalyzes the first reaction in cyanobacteria to incorporate fixed ammonia into an organic molecule,¹⁴⁷ and low activity, either by direct regulatory control or restriction on the synthesis of the enzyme, should leave an endosymbiont awash in NH₄⁺ during a period of nitrogen fixation.

A high concentration of NH_4^+ within an endosymbiont does not by itself do the host any good. A large amount of the fixed nitrogen must somehow find its way into the host's cytoplasm. A process by which this may happen has been proposed to explain the transfer of fixed nitrogen from rhizobial bacteroids to their legume hosts (Fig. 7, top):^{148,149}

- The NH_3 fixed within bacteroids diffuses across the bacteroid membrane.¹⁵⁰ At physiological pH, it is in equilibrium with NH_4^+ .
- The ammonia is trapped as NH_4^+ outside the bacteroids in a symbiotic cavity, because the host has pumped acid into the cavity surrounding the bacteroid to lower the pH.
- The host is protected from the low pH by a symbiotic membrane, separating the symbiotic cavity from the cytoplasm.
- The host takes up NH_4^+ using a host-encoded active transporter on the symbiotic membrane.
- Import of NH_4^+ by the bacteroid is repressed, to avoid reuptake.

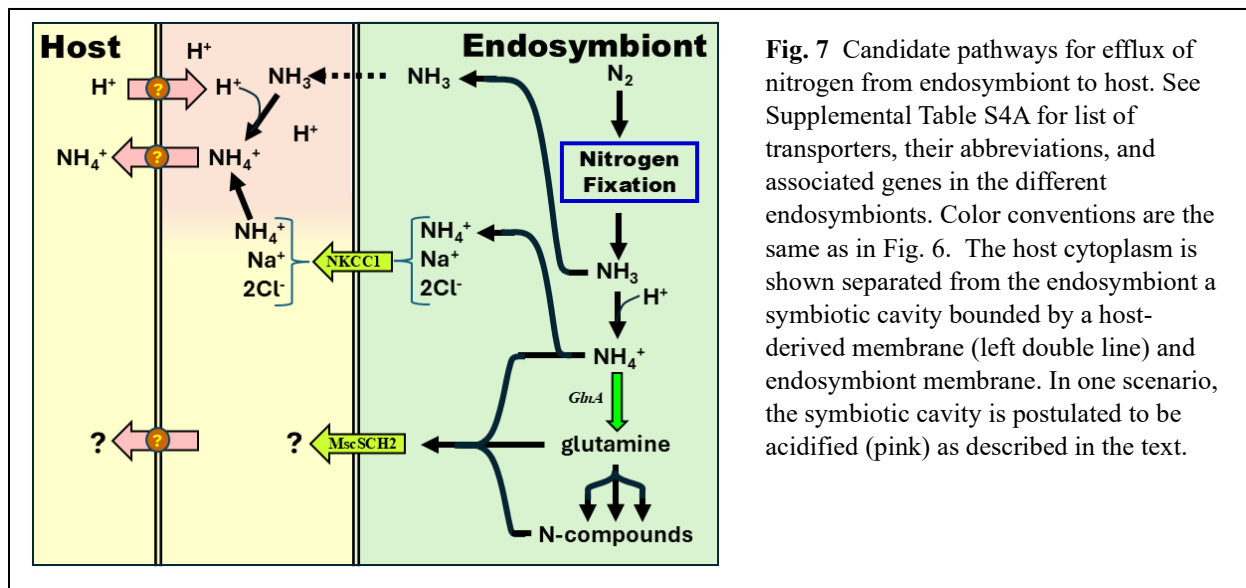


Fig. 7 Candidate pathways for efflux of nitrogen from endosymbiont to host. See Supplemental Table S4A for list of transporters, their abbreviations, and associated genes in the different endosymbionts. Color conventions are the same as in Fig. 6. The host cytoplasm is shown separated from the endosymbiont a symbiotic cavity bounded by a host-derived membrane (left double line) and endosymbiont membrane. In one scenario, the symbiotic cavity is postulated to be acidified (pink) as described in the text.

A critical requirement for this scheme is the presence of a symbiotic space. Such spaces provided by the host (either extracellularly or intracellularly) are known to enclose all characterized cyanobacterial symbioses,¹⁵¹ with the exception of *Richelia*/diatom associations.¹⁵² A Rhopalodian endosymbiont whose genome has not been sequenced also has been shown to be enclosed in a host-derived membrane,¹⁵³ as has UCYN-A.^{126,154} If the diatom host directs an H^+ -transporting ATPase and an NH_4^+ transporter to the symbiotic membrane, there's little left for the endosymbiont left to do but shut down its own Amt1 NH_4^+ transporter (Supplemental Table S4B) and fix N_2 . UCYN-A lacks any NH_4^+ transporter, so that step is unnecessary.

There are alternative routes the endosymbiont might employ to feed NH_4^+ to its host. In a careful study of the transporters encoded by heterocyst-forming cyanobacterial strains in association with diatoms, Nieves-Morión et al identified a remarkable transporter that might fill the bill.¹⁴³ The human protein NKCC1¹⁵⁵ in the Cation-Chloride-Cotransporter (CCC) family is highly similar to a protein in each of the *Richelia* strains they considered¹⁴³ and is also present in all Rhopalodian

endosymbionts as well as in UCYN-A (Fig. 7, Supplemental Table S4B). The typical role of the protein in humans is to cotransport Na^+ , K^+ , and Cl^- in either direction (regardless of the orientation of the transporter), governed by the $[\text{Cl}^-]$ gradient,¹⁵⁶ but its most notable feature for our purposes is that it is also able to transport NH_4^+ (replacing K^+).¹⁵⁷

NKCC1 consists of three regions:¹⁵⁸ an N-terminal soluble region associated with the site of regulation by phosphorylation,¹⁵⁵ a central region with twelve transmembrane domains, and a C-terminal soluble region that may be related to dimer stabilization and that interacts with transmembrane domains.¹⁵⁸ All eubacterial matches to NKCC1 are limited to the central and C-terminal regions, constituting 76% of its length and missing the 24% apparently devoted to regulation. Proteins very similar to NKCC1 are found in many cyanobacteria but rarely in heterotrophic bacteria.

A third possibility is that NH_4^+ or another nitrogenous compound might exit endosymbionts through relatively non-specific transporters of the mechanosensitive channel family (MscS), which function to release solutes in response to hypotonic conditions.¹⁵⁹ Perhaps the best studied is MscSCG and MscSCG2 from *Corynebacterium glutamicum*,^{160,161} which has been exploited for the industrial production of glutamate.¹⁶² These transporters are also able to facilitate the export of other amino acids.^{159,163} MscS transporters are found in all cyanobacteria, almost always in multiple versions. All the endosymbionts have orthologs of MscSCG2 (Supplemental Table S4B). Mutations in *E. coli* that render an MscS-family protein constitutively expressed are lethal,¹⁶⁴ but *C. glutamicum* is viable with constitutive expression of either MscSCG or MscSCG2,^{161,165,166} so it is conceivable that the natural orthologs of the protein in endosymbionts might have been repurposed to mediate the routine export of a nitrogenous compound.

Other less likely candidates in the Dicarboxylate / Amino Acid : Cation Symporter (DAACS) and Drug / Metabolite Transport (DMT) families have been discussed elsewhere.^{74,143}

Discussion

What is required from an existing nitrogen-fixing endosymbiont that would allow it to interact productively with a new host, specifically, with a crop plant? In this work, I have examined three parts of that question: (1) Do the genomes of the Rhopalodian endosymbionts encode the proteins required to meet its metabolic needs? (2) What carbon source from the host is required by the endosymbionts? and (3) By what process is fixed nitrogen exported to the hosts of the endosymbionts?

Metabolic independence of endosymbionts

There is prior reason to expect that the Rhopalodian endosymbionts might live primarily within the limits of their own resources. A proteomic study of the EcSB / *Epithemia clementina* association identified only six host-encoded proteins in the isolated endosymbiont, none of any obvious importance.¹⁶⁷ This is in stark contrast to the hundreds of proteins imported by UCYN-A¹⁷ and by the *Paulinella chromatophore*.¹⁶⁸ However, even one imported protein would be of clear significance, if it were shown that it had indeed traversed the membranes separating the host

cytoplasm from the endosymbiont, since that would signify that a mechanism to import proteins had been established. This remains to be demonstrated.

Given the apparent absence of host contributions to the proteome of Rhopalodian endosymbionts, perhaps it is no surprise that the endosymbionts appear to encode all the proteins it needs for the synthesis of macromolecules and their components, for ATP and NADPH production, and the cofactors needed for metabolic processes (with some reservations — see below). This conclusion must be tempered by humility: there surely are proteins important for survival that I did not consider, proteins whose functions may conceivably be provided by the host. They presumably would be contained within the list of 512 core proteins absent in endosymbiont genomes (Table 1, Supplemental Table S3B). These lists contain many proteins that one would expect to be absent in the endosymbionts (e.g. PSII proteins), and nothing in them stands out to me as a protein that would be needed for endosymbiont function.

Cofactors require special consideration regarding the ability of a Rhopalodian endosymbiont to function in a new host. At one extreme, there is pseudocobalamin. Since land plants do not make or utilize cobamides,¹⁶⁹ it is essential that the endosymbiont be able to synthesize its own (or, alternatively, be engineered to use cobamide-independent enzymes to replace the few that require the cofactor). In the case of some of the endosymbionts (RgSB, RaSB, EcSB, and EpSB), this condition appears to be met.

Their biotin biosynthesis pathways of the endosymbionts are incomplete, however. Land plants may be able to accommodate their needs, since the biotin biosynthetic pathway is split (at least in Arabidopsis), with the initial steps taking place in the cytoplasm and the final steps in the mitochondria.^{169,170} The same four endosymbionts are deficient only in the first enzyme of the pathway, so they may be able to join mitochondria in taking up the intermediate substrate, 7,8-diaminonanoate. All endosymbionts also encode a transporter of biotin, which must exist in the plant cytosol.

Folate biosynthesis is deficient in the endosymbionts because they appear to lack the ability to make PABA, a starting point of the pathway, but this may not be a problem. In Arabidopsis, PABA is synthesized in chloroplasts, exported to the cytosol, and imported by mitochondria, where synthesis of tetrahydrofolate is completed.¹⁷¹ Admittedly, it's not clear how the endosymbionts could take up the cytosolic PABA. If, however, a mechanism exists, then that would explain why the endosymbionts have preserved the downstream enzymes of the pathway (except possibly the last). If it doesn't, then it may be enough that all Rhopalodian endosymbionts have orthologs of a transporter capable of taking up 5-formyl-tetrahydrofolate from the plant host.⁹⁴

In short, the metabolism implied by the genomes of the endosymbionts may suffice to enable them to meet their metabolic needs within a new plant host.

Import of a carbon source and export of fixed nitrogen

All that metabolic capability counts for nothing unless the endosymbionts can gain from their hosts a source of carbon, as starting points for metabolic reactions and as an energy rich compound to power the production of NADPH and ATP. Glycerol would be an obvious choice, as it is the only

carbon source shown to support growth of Rip8802. Furthermore, the endosymbionts have all the enzymes (including an unusual glycerol-3-phosphate dehydrogenase) required to channel glycerol into central metabolism. However, while significant levels of glycerol are not uncommon in marine diatoms, they wouldn't be expected in freshwater diatoms, such as those that host five of the six endosymbionts. Glycerol would be a stronger candidate if it were established that all the symbioses derived from a single acquisition of a cyanobacterium by a marine diatom (see below), one that retained production of glycerol even after its progeny colonized freshwater habitats. Of course, the candidacy of glycerol could be assessed by the simple act of measuring glycerol in the host diatom strains.

It may be that glycerol – agreeable to both the marine diatom and its newly acquired cyanobacterium – was just a temporary solution, replaced over the time required to adapt to freshwater by a different carbon source. I've suggested that chitobiose is a reasonable candidate, as it is found in some diatoms, and endosymbionts are likely to take it up and metabolize it to glycolytic intermediates. This would require ramping up a transporter designed to act on low levels of peptidoglycan breakdown products so that it can provide the high levels of carbon required by nitrogen-fixation. Again, this hypothesis could be tested by measuring chitobiose in the different strains.

Neither glycerol nor chitobiose is likely to be found in the cytosol of crop plants. It will probably be necessary to engineer an endosymbiont to use a glucose transporter. Or perhaps the mysterious endosymbiont protein similar to maltose transporters will serve.

The mechanism of nitrogen export from the endosymbiont to its host is also unclear. One possibility is that NH_3 resulting from N_2 -fixation diffuses into an acidified space between the cyanobacteria-derived and host-derived membranes, analogous to the mechanism by which N-transfer is thought to take place in root nodules between bacteroids and legumes.^{148,149} However, this mechanism relies on the host to provide a membrane to surround the nitroplast and stock the membrane with enzymes to acidify the internal space and to take up NH_4^+ . Crop plants are unlikely to be able to accommodate these needs, so a different transport mechanism would need to be engineered into the nitroplast.

Two possible mechanisms of nitrogen export suggested by the endosymbiont genomes would be more practical in crop plants. Transporters similar to NKCC1 from humans in all Rhopalodian endosymbionts could facilitate the export of fixed nitrogen in the form of NH_4^+ from the endosymbiont to the cytosol of the new host (presuming that there is no host-provided membrane surrounding the endosymbiont). Alternatively, endosymbiont orthologs to MscSCG2 from *C. glutamicum* might act as a conduit for NH_4^+ ; they could also serve to export an organic form of fixed nitrogen. The physiological response of plant cells to a large influx of fixed nitrogen (particularly NH_4^+ ¹⁷²) is a concern. A first step is determining the form of nitrogen that passes between endosymbionts and their natural hosts. The activity of glutamine synthetase in endosymbionts would be particularly interesting to know.

Evolution of the Rhopalodian endosymbionts

The Rhopalodian diatoms are believed to have arisen from a common marine ancestor,⁴¹ from which the freshwater strains were derived. This suggests at least the following three scenarios to explain the presence of the endosymbionts:

- (1) **Multiple acquisitions (marine and freshwater):** A marine Rhopalodian diatom acquired a cyanobacterium, leading to the endosymbiont EpSB, while one or more freshwater Rhopalodian diatoms separately acquired a related cyanobacterium, leading to the other five endosymbionts.
- (2) **Single Acquisition (marine):** An ancestral marine Rhopalodian diatom acquired a cyanobacterium, and all endosymbionts are derived from this ancestor, some of whose progeny learned to live in freshwater.
- (3) **Single Acquisition (freshwater):** A more recent freshwater Rhopalodian diatom acquired a cyanobacterium, and all endosymbionts are derived from this ancestor. An ancestor specifically of EpSB transitioned to a marine environment. While freshwater / marine transitions are not uncommon in the distantly related diatom order Thalassiosirales,¹⁷³ it is not clear whether they take place in the Rhopalodiales.⁴¹

The multiple acquisition hypothesis is made more plausible by the apparent age of EpSB, the endosymbiont of the lone marine diatom considered in this study. EpSB has a smaller genome than the other endosymbionts (Fig. 1), and preliminary evidence suggests that it has far fewer broken genes (see Supplemental Table S2A; filter each column for “(“, the symbol for broken genes), consistent with a gene density that is intermediate between those of other endosymbionts and the older UCYN-A and *Paulinella* chromatophore (Fig. 1). Perhaps these characteristics arise from greater evolutionary time to dispose of the genetic debris.¹⁷⁴ Alternatively, EpSB might simply be more efficient in garbage collection.

There are problems with the idea of multiple acquisition events, however. It implies that the hypothetical freshwater cyanobacterial ancestor of the five freshwater endosymbionts is closer phylogenetically to the hypothetical marine cyanobacterial ancestor of EpSB than to the existing freshwater *Rippkaea* PCC 8802 (Rip8802), the closest known free-living relative. This is possible, but supporting the idea of a single marine ancestor (Hypothesis #2) is the finding that proteins with orthologs in all the endosymbionts but not in Rip8802 have orthologs primarily in marine cyanobacteria (Table 2, Supplemental Table S3C). For example, while most cyanobacteria possess Fe-dependent superoxide dismutase (FeSOD), a minority has a Ni-dependent form of the enzyme (NiSOD), almost all marine strains (Table 2 and Ref 175). All the endosymbionts have NiSOD. The strong association of NiSOD and marine habitats extend to heterotrophic bacteria as well, and it has been suggested that the preference is in response to the deficiency of iron in oceans.¹⁷⁶ The presence of NiSOD in the endosymbionts would make sense if there were a single progenitor, a free-living marine cyanobacterium.

Conclusions

The goal of this work was to assess whether any host-dependency of endosymbionts on their Rhopalodian hosts would preclude the transfer of the endosymbiont to a different host. A close

examination of the genomes of six endosymbionts of Rhopalodian diatoms shows no obvious metabolic dependency on their hosts, except for possibly some cofactors needed in very low amounts. This contrasts with the many dependencies evident from the genomes of two other cyanobacteria-derived endosymbionts: UCYN-A and the chromatophore from *Paulinella chromatophora*. Furthermore, the endosymbionts possess some unusual enzymatic capabilities that may enable them to manage the high need for reductant in the absence of PSII (GapN), withstand the production of O₂ by host photosynthesis (ARTO), and make more efficient use of glycerol as a carbon source (FAD-dependent glycerol-3-phosphate dehydrogenase).

It would be a mistake, however, to jump from this finding to the conclusion that Rhopalodian endosymbionts are ready to thrive within crop plants. This is highly improbable. First, crop plants are unlikely to provide a carbon source the endosymbionts seem equipped to accept and the mechanism of nitrogen export by the endosymbionts may be incompatible with crop plant physiology. Experiments on a model Rhopalodian endosymbiont are necessary to clarify these issues. Second, there was nothing presented here regarding the integration of cell division and division of the endosymbionts, as has been demonstrated with UCYN-A.¹⁷ The Rhopalodian endosymbionts evidently employ some way of ensuring defined partitioning into daughter cells^{26,177} (despite a report of loss of endosymbionts after extended growth in the laboratory³⁷), but until the mechanism is known, it is impossible to say whether it could function in alternative hosts.

Clearly, a good deal of laboratory experimentation is required to form a basis for the transfer of an endosymbiont to a foreign host. Of the six endosymbionts considered here, EcSB may be the most appropriate as a model system. Unlike some of the others (EaSB, EtSB, and RaSB), it appears able to synthesize pseudocobalamin. Unlike EpSB, it has what may be a complete pathway to synthesize pantothenate and more complete pathway for folate. Unlike RgSB, it has been grown in the laboratory, at least as a unialgal culture.²⁴ But there is no need to create all the tools required to tame a new model organism. A good path forward might be to couple basic experiments in EcSB with a study of the EcSB endosymbiont transferred to a related diatom, *Phaeodactylum tricornutum*, that offers a wealth of laboratory capabilities.¹⁷⁸ This prospect has been discussed elsewhere.⁷⁴

A sound experimental foundation may enable the transfer of a suitably modified endosymbiont into diverse crop plants, offering a low cost, environmentally benign source of nitrogen in such a way that might harness the creative input of local farmers and the society the technology is intended to serve.

Materials and methods

The eight endosymbionts and eight closely related free-living cyanobacteria whose genomes were considered in this study are shown in Fig. 1 with their sources shown in Supplemental Table S1. Broader consideration of cyanobacteria was done through searches of the 127 cyanobacterial genomes within BioBIKE¹⁷⁹ and the 277 semi-curated or highly curated cyanobacterial genomes of CyanoCyc,¹⁸⁰ searching by both Blast¹⁸¹ and built-in orthology functions. In the cases of *Crocospaera watsonii* WH 8501 and *Rippkaea orientalis* PCC 8801, the genome versions in BioBIKE differ from the most recent versions. The poorly annotated small plasmids of the

endosymbionts of *Epithemia adnata* Bon19 and *Rhopalodia gibba* 17Bon1³⁸ were not considered in this study, nor were the similar plasmids of endosymbionts of *Epithemia turgida*³⁵ and *Rhopalodia gibberula*.³⁶ It is not known whether the other two endosymbionts have plasmids.

Orthologs of proteins were provisionally defined as bidirectional best Blast hits.¹⁸² A bidirectional best hit is one in which Protein A in Organism X best matches Protein B in Organism Y, and Protein B best matches Protein A. Orthologs were generally found through the ORTHOLOGS-OF function of BioBIKE or a similar function in CyanoBIKE. BioBIKE's function adds an additional restriction that the match must have a better E-value than 10^{-10} . The restriction was relaxed to 10^{-3} for proteins whose lengths are less than 100 amino acids. Sometimes orthology was determined manually, particularly when the ORTHOLOGS-OF function was confused by duplicate genes. To obtain the set of core proteins common to free-living cyanobacteria related to the endosymbionts BioBIKE's COMMON-ORTHOLOGS-OF function was used. For technical reasons, proteins from *Crocospaera watsonii* WH 8501 were not included in determining common orthologs.

Whenever an ortholog existed to a protein in the carefully annotated *Synechocystis* PCC 6803, the annotation from CyanoCyc was used, noting in Supplemental Table S2A whether the annotation is based on experimental evidence. Otherwise the annotation was from RefSeq.¹⁸³ Sometimes, functional annotation was taken from orthologous proteins in heterotrophic bacteria with experimental justification, often found in EcoCyc.¹⁸⁴

The phylogenetic tree of endosymbionts and closely related cyanobacteria was based on concatenated alignments of 29 conserved proteins (Supplemental Table S5), using Clustal W¹⁸⁵ for the alignment, GBlocks¹⁸⁶ to extract informative columns, and PhyML¹⁸⁷ for the final tree. with a model selected by Smart Model Selection,¹⁸⁸ LG as the substitution model, and 100 bootstraps. Quick individual protein trees were made within BioBIKE using its TREE-OF function and neighbor joining.

The metabolic pathways discussed in the Results section, summarized in Fig. 4, listed in Supplemental Table S2, and made visible in Supplemental Figs. S01-S08 were derived from KEGG metabolic maps.¹⁸⁹ Needless to say, they are ultimately arbitrary simplifications of the complex interplay amongst biochemical reactions. Information about transporters was taken largely from the Transporter Classification Database (TCDB).¹³⁵

Supplementary information

Supplemental-figures- S01-through-S11.pdf

Figures S01 through S08 contain maps of pathways with presence or absence of proteins of endosymbionts superimposed

- Figure S01: Glycogen metabolism
- Figure S02: Glycolysis
- Figure S03: Pentose phosphate / Entner-Doudoroff pathways
- Figure S04: Tricarboxylic acid cycle
- Figure S05: Biotin biosynthesis
- Figure S06: Folate and derivatives biosynthesis

875 • Figure S07: Pantothenate /Coenzyme A biosynthesis
876 • Figure S08: Pseudocobalamin biosynthesis

877 Supplemental-Table-S1-genomes.xlsx
878 Genomes used in this study

879 Supplemental-Table-S2-orthologs.xlsx
880 Proteins and RNA related to selected metabolic pathways and biochemical capabilities

881 Supplemental-Table-S3-sets.xlsx
882 • Table S3A: Counts of sets of orthologs and their intersections
883 • Table S3B: Proteins present in Core set with orthologs in no Rhopalodian endosymbiont
884 • Table S3C: Proteins with orthologs in all Rhopalodian endosymbionts but not in Rip8802

885 Supplemental-Table-S4-transport.xlsx
886 • Table S4A: Proteins related to carbon transport and subsequent metabolism
887 • Table S4B: Proteins related to nitrogen transport

888 Supplemental-Table-S5-proteins-for-tree.xlsx
889 Orthologs used to construct the genomic phylogenetic tree
890

891 **Data availability**
892 Not applicable.

893 **Declarations**

894 **Ethics approval and consent to participate**
895 Not applicable.

896 **Consent for publication**
897 Not applicable.

898 **Availability of data and materials**
899 Not applicable.

900 **Competing interests**
901 I declare that I have no competing interests.

902 **Funding**
903 Not applicable.

904 **Authors' contributions**
905 Not applicable.

906 **Acknowledgements**
907 I would like to thank many for stimulating discussions, including: Tyler Coale, Enrique Flores,
908 Rachel Foster, Amanda Shelton, Terry Thiel, Wim Vermaas, and Jon Zehr. I must also thank
909 Carlisle Childress and Mike Davis for their efforts to keep BioBIKE alive, necessary for the
910 completion of this project.

References

1. Sutton MA and 22 others. Our Nutrient World: The challenge to produce more food and energy with less pollution. Global Overview of Nutrient Management. Centre for Ecology and Hydrology, Edinburgh on behalf of the Global Partnership on Nutrient Management and the International Nitrogen Initiative. 2013. <https://doi.org/10.5511/plantbiotechnology.19.145>.
2. Lu C, Tian. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth Syst Sci Data*. 2011;9:181-92. <https://doi.org/10.5194/essd-9-181-2017>.
3. Royal Society. Ammonia: zero-carbon fertiliser, fuel and energy store. Policy briefing. Royal Society, London. 2020. <https://doi.org/10.5194/essd-9-181-2017>.
4. Menegat S, Ledo A, Tirado R. Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture. *Sci Rep*. 2022;12:14490. <https://doi.org/10.1038/s41598-022-18773-w>.
5. Tang X, Griffith G, Malcolm B. A critical review of the external costs of nitrogen fertilizer use. *Australasian Agribusiness Review*. 2023;31:Paper 5. <https://smartfertiliserhub.org.au/publications/articles/a-critical-review-of-the-external-costs-of-nitrogen-fertiliser-use/>.
6. Yang Y and 18 others. Climate change exacerbates the environmental impacts of agriculture. *Science*. 2024;385:eadn3747. <https://doi.org/10.1126/science.adn3747>.
7. Dethier J-J, Effenberger A. Agriculture and development: A brief review of the literature. *Econ Sys* 36:175-205. *Econ Sys*. 2012;36:175-205. <http://doi.org/10.1016/j.ecosys.2011.09.003>.
8. Suri T, Udry C. Agricultural technology in Africa. *J Econ Perspectives*. 2022;36(1). <https://doi.org/10.1257/JEP.36.1.33>.
9. United Nations, Department of Economics and Social Affairs. World Population Prospects 2024. UN DESA/POP/2024/TR/NO. 9. 2012. https://population.un.org/wpp/assets/Files/WPP2024_Summary-of-Results.pdf.
10. Huisman R, Geurts R. A roadmap toward engineered nitrogen-fixing nodule symbiosis. *Plant Comm*. 2020;1:100019. <https://doi.org/10.1016/j.xplc.2019.100019>.
11. Jhu M-Y, Oldroyd GED. Dancing to a different tune, can we switch from chemical to biological nitrogen fixation for sustainable food security? *PLoS Biol*. 2023;21:e3001982. <https://doi.org/10.1371/journal.pbio.3001982>.
12. Bloch SE, Ryu M-H, Ozaydin B, Broglie R. Harnessing atmospheric nitrogen for cereal crop production. *Curr Opin Biotechnol*. 2020;62:181-8. <https://doi.org/10.1016/j.copbio.2019.09.024>.
13. Soumare A, Diedhiou AG, Thuita M, Hafidi M, Ouhdouch Y, Gopalakrishnan S, Kouisni L. Exploiting biological nitrogen fixation: a route towards a sustainable agriculture. *Plants*. 2020;9:1011. <https://doi.org/10.3390/plants9081011>.
14. Imran A, Hakim S, Tariq M, Nawaz MS, Laraib I, Qulzar U, Hanif MK, Siddique MJ, Hayat M, Fraz A, Ahmad M. Diazotrophs for lowering nitrogen pollution crises: Looking deep into the roots. *Front Microbiol*. 2021;12:637815. <https://doi.org/10.3389/fmicb.2021.637815>.
15. Postgate JR, Dixon R, Hill S, Kent H. *Nif* genes in alien backgrounds. *Phil Trans R Soc Lond B*. 1987;317:227-43. <https://doi.org/10.1098/rstb.1987.0059>.
16. Nowack ECM, Weber APM. Genomics-informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. *Annu Rev Plant Biol*. 2018;69:51-84. <https://doi.org/10.1146/annurev-arplant-042817-040209>.
17. Coale TH, Loconte V, Turk-Kubo KA, Vanslebrouck B, Mak WKE, Cheung S, Ekman A, Chen J-H, Haino K, Takano Y, Nishimura T, Adachi M, Le Gros M, Larabell C, Zehr JP. Nitrogen-fixing organelle in a marine alga. *Science*. 2024;384:217-22. <https://doi.org/10.1126/science.adk1075>.
18. Bogdanova VS. Genetic and molecular genetic basis of nuclear-plastid Incompatibilities. *Plants*. 2020;9:23. <https://doi.org/10.3390/plants9010023>.
19. Stegemann S, Keuthe M, Greiner S, Bock R. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci USA*. 2012;109:2434-38. <https://doi.org/10.1073/pnas.1114076109>.

-
20. Greiner S, Bock R. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays*. 2013;35:354-65. <https://doi.org/10.1002/bies.201200137>,
 21. Shi L-X, Theg SM. The chloroplast protein import system: From algae to trees. *Biochim Biophys Acta*. 2013;1833:314-31. <https://doi.org/10.1016/j.bbamcr.2012.10.002>.
 22. Prechtel J, Kneip C, Lockhart P, Wenderoth K, Meier U-G. Intracellular spheroid bodies of *Rhopalodia gibba* have nitrogen-fixing apparatus of cyanobacterial origin. *Mol Biol Evol*. 2004;21:1477-81. <https://doi.org/10.1093/molbev/msh086>.
 23. Stancheva R, Lowe R. Diatom symbioses with other phototroph. In: Seckbach J, Gordon R (eds). *Diatoms: Fundamentals and Applications*. Scrivener Publishing, 2019. p.225-44. <https://doi.org/10.1002/9781119370741.ch10>.
 24. Moulin SLY, Frail S, Braukmann T, Doenier J, Steele-Ogus M, Marks JC, Mills MM, Yeh E. The endosymbiont of *Epithemia clementina* is specialized for nitrogen fixation within a photosynthetic eukaryote. *ISME Commun*.2024;4:ycae055. <https://doi.org/10.1093/ismeco/ycae055>.
 25. Bothe H, Tripp HJ, Zehr JP: Unicellular cyanobacteria with a new mode of life: The lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch Microbiol*. 2010;192:783-90. <https://doi.org/10.1007/s00203-010-0621-5>.
 26. Kamakura S, Mann DG, Nakamura N, Sato S. Inheritance of spheroid body and plastid in the raphid diatom *Epithemia* (Bacillariophyta) during sexual reproduction. *Phycol*.2021;60:265-73. <https://doi.org/10.1080/00318884.2021.1909399>.
 27. Nakayama T, Ikegami Y, Nakayama T, Ishida K-I, Inagaki Y, Inouye I. Spheroid bodies in rhopalodiacean diatoms were derived from a single endosymbiotic cyanobacterium. *J Plant Res*.2011;124:93-97. <https://doi.org/10.1007/s10265-010-0355-0>.
 28. Benson ME, Kociolek P, Spaulding SA, Smith DM. Pre-neogene non-marine diatom biochronology with new data from the late Eocene Florissant Formation of Colorado, USA. *Stratigraphy*. 2012;9:131-52. <https://doi.org/10.29041/strat.09.2.02>.
 29. Cornejo-Castillo FM, Cabello AM, Salazar G, Sánchez-Baracaldo P, Lima-Mendez G, Hingamp P, Alberti A, Sunagawa S, Bork P, de Vargas C, Raes J, Bowler C, Wincker P, Zehr JP, Gasol JM, Massana R, Acinas SG. Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nature Commun*. 2015;7:11071. <https://doi.org/10.1038/ncomms11071>.
 30. Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc Natl Acad Sci USA*. 2017;114:E7737-45. <https://doi.org/10.1073/pnas.1620089114>.
 31. Carpenter EJ, Janson S. Intracellular cyanobacterial symbionts in the marine diatom *Climacodium frauenfeldianum* (bacillariophyceae). *J Phycol*. 2000;36:540-4. <https://doi.org/10.1046/j.1529-8817.2000.99163.x>.
 32. Foster RA, Tienken D, Littmann S, Whitehouse MJ, Kuypers MMMM, White AE. The rate and fate of N₂ and C fixation by marine diatom-diazotroph symbioses. *ISME J*. 2022;16:477-87. <https://doi.org/10.1038/s41396-021-01086-7>.
 33. Nakayama T, Nomura M, Yabuki A, Shiba K, Inaba K, Inagaki Y. Convergent reductive evolution of cyanobacteria in symbiosis with Dinophysiales dinoflagellates. *Sci Reports*. 2024;14:12774. <https://doi.org/10.1038/s41598-024-63502-0>.
 34. Husnik F, Keeling PJ. The fate of obligate endosymbionts: reduction, integration, or extinction. 2019;58-59:1-8. <https://doi.org/10.1016/j.gde.2019.07.014>.
 35. Nakayama T, Kamikawa R, Tanifuji G, Kashiwayama Y, Ohkouchi N, Archibald JM, Inagaki Y. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proc Natl Acad Sci USA*.2014;111:11407-12. <http://www.pnas.org/cgi/doi/10.1073/pnas.1405222111>
 36. Nakayama T, Inagaki Y. Genomic divergence within non-photosynthetic cyanobacterial endosymbionts in rhopalodiacean diatoms. *Sci Rep*. 2017;7:1-8. <http://doi.org/10.1038/s41598-017-13578-8>
 37. Schvarcz CR, Wilson ST, Caffin M, Stancheva R, Li Q, Turk-Kubo KA, White AE, Karl DM, Zehr JP, Steward GF. Overlooked and widespread pennate diatom-diazotroph symbioses in the sea. *Nat Commun*.2022;13:799. <https://doi.org/10.1038/s41467-022-28065-6>.
 38. Abresch H, Bell T, Miller SR. Diurnal transcriptional variation is reduced in a nitrogen-fixing diatom endosymbiont. *ISME J*. 2024;18:wrae064. <https://doi.org/10.1093/ismejo/wrae064>.

-
39. Welsh EA, Liberton M, Stöckel J, Loh T, Elvitagala T, Wang C, Wollam A, Fulton RS, Clifton SW, Jacobs JM, Aurora R, Ghosh BK, Sherman LA, Smith RD, Wilson RK, Pakrasi HB. The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc Natl Acad Sci USA*. 2008;105:15094-9. <https://doi.org/10.1073/pnas.0805418105>.
 40. Mareš J, Johansen JR, Hauer T, Zima J Jr, Ventura S, Cuzman, O, Tiribilli B, Kaštovský J. Taxonomic resolution of the genus *Cyanothece* (Chroococcales, Cyanobacteria), with a treatment on *Gloeotheca* and three new genera, *Crocospaera*, *Rippkaea*, and *Zehria*. *J Phycol*. 2019;55(3):578-610. <https://doi.org/10.1111/jpy.12853>.
 41. Ruck EC, Nakov T, Alverson AJ, Theriot EC. Phylogeny, ecology, morphological evolution, and reclassification of the diatom orders Surirellales and Rhopalodiales. *Mol Phylogenet Evol*. 2016;103:155-71. <https://doi.org/10.1016/j.ympev.2016.07.023>.
 42. Chen M-Y, Teng W-K, Zhao L, Hu C-X, Zhou Y-K, Han B-P, Song L-R, Shu W-S. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J*. 2021;15:211-27. <https://doi.org/10.1038/s41396-020-00775-z>.
 43. Thompson A, Carter BJ, Turk-Kubo K, Malfatti F, Azam F, Zehr JP. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ Microbiol*. 2014;16:3238-49. <https://doi.org/10.1111/1462-2920.12490>.
 44. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nature Rev Microbiol*. 2012;10:13-26. <https://doi.org/10.1038/nrmicro2670>
 45. Cassier-Chauvat C, Veaudor T, Chauvat F. Comparative genomics of DNA recombination and repair in cyanobacteria: biotechnological implications. *Front Microbiol*. 2016. 7:1809. <https://doi.org/10.3389/fmicb.2016.01809>.
 46. Ng W-O, Pakrasi HB. DNA photolyase homologs are the major UV resistance factors in the cyanobacterium *Synechocystis* sp. PCC 6803. *Mol Genet Genom*. 2001;264:924-31. <https://doi.org/10.1007/s004380000383>
 47. Kouzminova EA, Kadyrov FF, Kuzminov A. RNase HIII saves *Escherichia coli* from R-loop-associated chromosomal fragmentation. *J Mol Biol*. 2017;429:2873-94. <http://doi.org/10.1016/j.jmb.2017.08.004>.
 48. Das S, Forrest J, Kuzminov A. Synthetic lethal mutants in *Escherichia coli* define pathways necessary for survival with RNase H deficiency. *J Bacteriol*. 2023;205:e0028023. <https://doi.org/10.1128/jb.00280-23>.
 49. Gil R, Silva FJ, Peretó J, Moya A. Determination of the core of a minimal bacterial gene set. *Microb Molec Biol Rev*. 2004;68:518-37. <https://doi.org/10.1128/MMBR.68.3.518-537.2004>.
 50. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. *Microb Molec Biol Rev*. 2005;69:101-23. <https://doi.org/10.1128/MMBR.69.1.101-123.2005>.
 51. Seefeldt LC, Hoffman BM, Dean DR. Mechanism of Mo-dependent nitrogenase. *Annu Rev Biochem*. 2009;78:701-722. <https://doi.org/10.1146/annurev.biochem.78.070907.103812>.
 52. Fay P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol Rev*. 1992;56:340-73. <https://mmbr.asm.org/content/56/2/340>.
 53. Wolk CP, Ernst A, Elhai J. Heterocyst metabolism and development. In: *The molecular biology of cyanobacteria*. Bryant DA. editor. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994. p.769-823. https://doi.org/10.1007/978-94-011-0227-8_27.
 54. Walsby AE. Cyanobacterial heterocysts: Terminal pores proposed as sites of gas exchange. *Trends Microbiol*. 2007;15:340-3. <https://doi.org/10.1016/j.tim.2007.06.007>.
 55. Valladares A, Maldener I, Muro-Pastor AM, Flores E, Herrero A. Heterocyst development and diazotrophic metabolism in terminal respiratory oxidase mutants of the cyanobacterium *Anabaena* sp. Strain PCC 7120. *J Bacteriol*. 2007;189:4425-30. <https://doi.org/10.1128/JB.00220-07>.
 56. Milligan AJ, Berman-Frank I, Gerchman Y, Dismukes GC, Falkowski PG. Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection. *J Phycol*. 2007;43:845-52. <https://doi.org/10.1111/j.1529-8817.2007.00395.x>.
 57. Ermakova M, Battchikova N, Richaud P, Leino H, Kosourov S, Isojärvi J, Peltier G, Flores E, Cournac L, Allahverdiyeva Y, Aro E-M. Heterocyst-specific flavodiiron protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Proc Natl Acad Sci USA*. 2014;111:11205-10. <https://doi.org/10.1073/pnas.1407327111>.

-
58. Scherer S, Almon H, Böger P. Interaction of photosynthesis, respiration and nitrogen fixation in cyanobacteria. *Photosyn Res.* 1988;15:95-114. <https://doi.org/10.1007/BF00413385>.
 59. Bothe H, Schmitz O, Yates MG, Newton WE. Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiol Molec Biol Rev.* 2010;74:529-551. <https://doi.org/10.1128/MMBR.00033-10>.
 60. Ehira S, Ohmori M. NrrA, a nitrogen-regulated response regulator protein, controls glycogen catabolism in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *J Biol Chem.* 2011;286:38109-14. <https://doi.org/10.1074/jbc.M111.289124>.
 61. Bergman B, Gallon JR, Rai AN, Stal LJ. N₂ fixation by non-heterocystous cyanobacteria. *FEMS Microbiol Rev.* 1997;19:139-85. [https://doi.org/10.1016/S0168-6445\(96\)00028-9](https://doi.org/10.1016/S0168-6445(96)00028-9).
 62. Sherman LA, Meunier P, Colón-López MS. Diurnal rhythms in metabolism: A day in the life of a unicellular, diazotrophic cyanobacterium. *Photosyn Res.* 1998;58:25-42. <https://doi.org/10.1023/A:1006137605802>.
 63. Schneegurt MA, Tucker DL, Ondr JK. Metabolic rhythms of a diazotrophic cyanobacterium, *Cyanothece* sp. strain ATCC 51142, heterotrophically grown in continuous dark. *J Phycol.* 2000;36:107-17. <https://doi.org/10.1046/j.1529-8817.2000.99152.x>.
 64. Magnusen A. Heterocyst thylakoid bioenergetics. *Life.* 2019;9:13. <https://doi.org/10.3390/life9010013>.
 65. Spaans SK, Weusthuis RA, van der Oost J, Kengen SWM. NADPH-generating systems in bacteria and archaea. *Front Microbiol.* 2015;6:742. <https://doi.org/10.3389/fmicb.2015.00742>.
 66. Janaki S, Wolk CP. Synthesis of nitrogenase by isolated heterocysts. *Biochim Biophys Acta.* 1982;698:187-92. [https://doi.org/10.1016/0167-4781\(82\)90027-6](https://doi.org/10.1016/0167-4781(82)90027-6).
 67. Gallon J. N₂ fixation in phototrophs: Adaptation to a specialized way of life. *Plant Soil.* 2001;230:39-48. <https://doi.org/10.1023/A:1004640219659>.
 68. Stöckel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, Pakrasi HB. Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci USA.* 2008;105:6156-61. <https://doi.org/10.1073/pnas.0711068105>.
 69. Muñoz-Marin MC, Shilova IN, Shi T, Farnelid H, Cabello AM, Zehr JP. The transcriptional cycle is suited to daytime N₂ fixation in the unicellular cyanobacterium "*Candidatus Atelocyanobacterium thalassa*" (UCYN-A). *mBio.* 2019;10:e02495. <https://doi.org/10.1128/mBio.02495-18>.
 70. Chen X, Schreiber K, Appel J, Makowka A, Fähnrich B, Roettger M, Hajirezaei MR, Sönnichsen FD, Schönheit P, Martin WF, Gutekunst K. The Entner-Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. *Proc Natl Acad Sci USA.* 2016;113:5441-6. <https://doi.org/10.1073/pnas.1521916113>.
 71. Doello S, Klotz A, Makowka A, Gutekunst K, Forchhammer K. A specific glycogen mobilization strategy enables rapid awakening of dormant cyanobacteria from chlorosis. *Plant Physiol.* 2018;177:594-603. <https://doi.org/10.1104/pp.18.00297>.
 72. Herrero A, Flores E, Imperial J. Nitrogen assimilation in bacteria. In: Schmidt TM, editor. *Encyclopedia of microbiology*, 4th Edition. UK, Elsevier; 2019. Vol. 3:280-300. <https://doi.org/10.1016/B978-0-12-809633-8.20680-8>.
 73. Habenicht A. The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase: Biochemistry, structure, occurrence and evolution. *Biol Chem.* 1997;378:1413-19. <https://doi.org/10.1104/pp.103.030981>.
 74. Elhai J. Engineering of crop plants to facilitate bottom-up innovation: A possible role for broad host-range nitroplasts and neoplasts. *OSF Preprints.* 2023. <https://doi.org/10.31219/osf.io/ny2rc>.
 75. Piattoni CV, Guerrero SA, Iglesias AA. A differential redox regulation of the pathways metabolizing glyceraldehyde-3-phosphate tunes the production of reducing power in the cytosol of plant cells. *Intl J Mol Sci.* 2013;14:8073-92. <https://doi.org/10.3390/ijms14048073>.
 76. Schmetterer G. The respiratory terminal oxidases (RTOs) of cyanobacteria. In: Cramer WA, Kallas T, editors. *Cytochrome complexes: evolution, structures, energy transduction, and signaling*. Springer Science-Business Media, Dordrecht; 2016. p.331-355. https://doi.org/10.1007/978-94-017-7481-9_17.
 77. Valladares A, Herrero A, Pils D, Schmetterer G, Flores E. Cytochrome *c* oxidase genes required for nitrogenase activity and diazotrophic growth in *Anabaena* sp. PCC 7120. *Molec Microbiol.* 2003;47:1239-49. <https://doi.org/10.1046/j.1365-2958.2003.03372.x>.

-
78. Zhang X, Sherman DM, Sherman LA. The uptake hydrogenase in the unicellular diazotrophic cyanobacterium *Cyanothece* sp. strain PCC 7822 protects nitrogenase from oxygen toxicity. *J Bacteriol.* 2014;196:840-9. <https://doi.org/10.1128/JB.01248-13>.
 79. Puggioni V, Tempel S, Latifi. Distribution of hydrogenases in cyanobacteria: A phylum-wide genomic survey. *Front Genet.* 2016;7:1-14. <https://doi.org/10.3389/fgene.2016.00223>.
 80. Allahverdiyeva Y, Isojärvi J, Zhang P, Aro E-M. Cyanobacterial oxygenic photosynthesis is protected by flavodiiron proteins. *Life.* 2015;5:716-43. <https://doi.org/10.3390/life5010716>.
 81. Vavilin DV, Vermaas WFJ. Regulation of the tetrapyrrole biosynthetic pathway leading to heme and chlorophyll in plants and cyanobacteria. *Physiol Plant* 2002;115:9-24. <https://doi.org/10.1034/j.1399-3054.2002.1150102.x>.
 82. Bryant DA, Hunter CN, Warren MJ. Biosynthesis of the modified tetrapyrroles – the pigments of life. *J Biol Chem.* 2020;295:6888-925. <https://doi.org/10.1074/jbc.REV120.006194>.
 83. Gálová E, Šalgovičová I, Demko V, Mikulová K, Ševčovičová A, Slováková L, Kyselá V, Hudák J. A short overview of chlorophyll biosynthesis in algae. *Biologia.* 2008;63:947-51. <https://doi.org/10.2478/s11756-008-0147-3>.
 84. Islam MR, Aikawa S, Midorikawa T, Kashino Y, Satoh K, Koike H. *slr1923* of *Synechocystis* sp. PCC6803 is essential for conversion of 3,8-divinyl(proto)chlorophyll(ide) to 3-monovinyl(proto)chlorophyll(ide). *Plant Physiol.* 2008;148:1068-81. <https://doi.org/10.1104/pp.108.123117>.
 85. Shpilyov AV, Zinchenko VV, Shestakov SV, Grimm B, Lokstein H. Inactivation of the geranylgeranyl reductase (ChlP) gene in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta.* 2005;1706:195-203. <https://doi.org/10.1016/j.bbabi.2004.11.001>.
 86. Funk C, Vermaas W. A cyanobacterial gene family coding for single-helix proteins resembling part of the light-harvesting proteins from higher plants. *Biochem.* 1999;38:9397-404. <https://doi.org/10.1021/bi990545+>.
 87. Sarkar D, Landa M, Bandyopadhyay A, Pakrasi HB, Zehr JP, Maranas CD. Elucidation of trophic interactions in an unusual single-cell nitrogen-fixing symbiosis using metabolic modeling. *PLoS Comput Biol.* 2021;17:e1008983. <https://doi.org/10.1371/journal.pcbi.1008983>.
 88. Leonardi R, Jackowski S. Biosynthesis of pantothenic acid and coenzyme A. *EcoSal Plus.* 2007;2(2). <https://doi.org/10.1128/ecosalplus.3.6.3.4>.
 89. Spry C, Kirk K, Saliba KJ. Coenzyme A biosynthesis: an antimicrobial drug target. *FEMS Microbiol Rev.* 2007;32:56-106. <https://doi.org/10.1111/j.1574-6976.2007.00093.x>.
 90. Pernil R, Picossi S, Mariscal V, Herrero A, Flores E. ABC-type amino acid uptake transporters Bgt and N-II of *Anabaena* sp. strain PCC 7120 share an ATPase subunit and are expressed in vegetative cells and heterocysts. *Mol Microbiol.* 2008;67:1067-80. <https://doi.org/10.1111/j.1365-2958.2008.06107.x>.
 91. Sirithanakorn C, Cronan JE. Biotin, a universal and essential cofactor: synthesis, ligation and regulation. *FEMS Microbiol Rev.* 2021;45:1-18. <https://doi.org/10.1093/femsre/fuab00>.
 92. Hebbeln P, Rodionov DA, Alfandega A, Eitinger T. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. *Proc Natl Acad Sci USA.* 2007;104:2909-14. <https://doi.org/10.1073/pnas.0609905104>.
 93. de Crécy-Lagard V, El Yacoubi B, Díaz de la Garza R, Noiriél A, Hanson AD. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics.* 2007;8:245. <https://doi.org/10.1186/1471-2164-8-245>.
 94. Klaus SMJ, Kunji ERS, Bozzo GG, Noiriél A, Diaz de la Garza R, Basset GJC, Ravanel S, Rébeillé F, Gregory III JF, Hanson AD. Higher plant plastids and cyanobacteria have folate carriers related to those of trypanosomatids. *J Biol Chem.* 2005;280:38457-63. <https://doi.org/10.1074/jbc.M507432200>.
 95. Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, Scanlan DJ, Warren MJ, Smith AG. Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. *Curr Biol.* 2016;26:1-10. <http://dx.doi.org/10.1016/j.cub.2016.02.041>.
 96. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature.* 2005;438:90-3. <https://doi.org/10.1038/nature04056>.
 97. Warren MJ, Bolt EL, Roessner CA, Scott AI, Spencer JB, Woodcock SC. Gene dissection demonstrates that the *Escherichia coli* *cysG* gene encodes a multifunctional protein. *Biochem J.* 1994;302:837-44. <https://doi.org/10.1042/bj3020837>.

-
98. Schubert HL, Rose RS, Leech HK, Brindley AA, Hill CP, Rigby SEJ, Warren MJ. Structure and function of SirC from *Bacillus megaterium*: a metal-binding precorrin-2 dehydrogenase. *Biochem J.* 2008;415:257-63. <https://doi.org/10.1042/BJ20080785>.
 99. Shelton AN, Seth EC, Mok KC, Han AW, Jackson SN, Haft DR, Taga ME. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J.* 2019;13:789-804. <https://doi.org/10.1038/s41396-018-0304-9>.
 100. Heal KR, Qin W, Ribalet F, Bertagnolli AD, Coyote-Maestas W, Hmelo LR, Moffett JW, Devol AH, Armbrust EV, Stahl DA, Ingalls AE. Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proc Natl Acad Sci USA.* 2017;114:364-9. <https://doi.org/10.1073/pnas.1608462114>.
 101. Pérez AA, Rodionov DA, Bryant DA. Identification and regulation of genes for cobalamin transport in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *J Bacteriol.* 2016;198:2753-61. <https://doi.org/10.1128/JB.00476-16>.
 102. Tanioka Y, Miyamoto E, Yabuta Y, Ohnishi K, Fujita T, Yamaji R, Misono H, Shigeoka S, Nakano Y, Inui H, Watanabe F. Methyladeninylcobamide functions as the cofactor of methionine synthase in a cyanobacterium, *Spirulina platensis* NIES-39. *FEBS Lett.* 2010;584:3223-6. <https://doi.org/10.1016/j.febslet.2010.06.013>.
 103. Schneegurt MA, Sherman DM, Sherman LA. Growth, physiology, and ultrastructure of a diazotrophic cyanobacterium, *Cyanothece* sp. strain ATCC 51142, in mixotrophic and chemoheterotrophic cultures. *J Phycol.* 1997;33:632-42. <https://doi.org/10.1111/j.0022-3646.1997.00632.x>.
 104. Feng X, Bandyopadhyay A, Berla B, Page L, Wu B, Pakrasi H, Tang YJ. Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. ATCC 51142 under continuous light. *Microbiol.* 2010;156:2566-74. <https://doi.org/10.1099/mic.0.038232-0>.
 105. Blötz C, Stülke J. Glycerol metabolism and its implication in virulence in *Mycoplasma*. *FEMS Microbiol Rev.* 2017;41:640-52. <https://doi.org/10.1038/ncomms1139>.
 106. Richey D P, Lin, E C C. Importance of facilitated diffusion for effective utilization of glycerol by *Escherichia coli*. *J Bacteriol.* 1972;112:784-90. <https://doi.org/10.1128/jb.112.2.784-790.1972>.
 107. Oren A. Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev.* 1999;63:334-48. <https://doi.org/10.1128/mmb.63.2.334-348.1999>.
 108. Dickson DMJ, Kirst GO. Osmotic adjustment in marine eukaryotic algae: The role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. I. Diatoms and a rhodophyte. *New Phytol.* 1987;106:645-55. <https://doi.org/10.1111/j.1469-8137.1987.tb00165.x>.
 109. Scholz B, Liebezeit G. Compatible solutes in three marine intertidal microphytobenthic Wadden Sea diatoms exposed to different salinities. *Eur J Phycol.* 2012;47:393-407. <https://doi.org/10.1080/09670262.2012.720714>.
 110. Stal LJ, Moezelaar R. Fermentation in cyanobacteria. *FEMS Microbiol Rev.* 1997;21:179-211. [https://doi.org/10.1016/S0168-6445\(97\)00056-9](https://doi.org/10.1016/S0168-6445(97)00056-9).
 111. Wolk CP. Physiology and cytological chemistry of blue-green algae. *Bacteriol Rev.* 1973;37:32-101. <https://doi.org/10.1128/br.37.1.32-101.1973>.
 112. Ekman M, Picossi S, Campbell EL, Meeks JC, Flores E. *Nostoc* sugar transporter necessary to establish a cyanobacterium-plant symbiosis. *Plant Physiol.* 2013;161:1984-92. <https://doi.org/10.1104/pp.112.213116>.
 113. Muñoz-Marin MC, Luque I, Zubkov MV, Díez J, García-Fernández JM. *Prochlorococcus* can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean. *Proc Natl Acad Sci USA.* 2013;110:8597-8602. <https://doi.org/10.1073/pnas.1221775110>.
 114. Verhoeven NM, Jakobs C, Carney G, Somers MP, Wanders RJA, Rizzo WB. Involvement of microsomal fatty aldehyde dehydrogenase in the α -oxidation of phytanic acid. *FEBS Lett.* 1998;429:225-8. [https://doi.org/10.1016/S0014-5793\(98\)00574-2](https://doi.org/10.1016/S0014-5793(98)00574-2).
 115. Muzio G, Maggiora M, Paiuzzi E, Oraldi M, Canuto RA. Aldehyde dehydrogenases and cell proliferation. *Free Radical Biol Med.* 2012;52:735-46. <https://doi.org/10.1016/j.freeradbiomed.2011.11.033>.
 116. Trautmann D, Beyer P, Al-Babili S. The ORF slr0091 of *Synechocystis* sp. PCC6803 encodes a high-light induced aldehyde dehydrogenase converting apocarotenals and alkanals. *FEBS J.* 2013;280:3685-96. <https://doi.org/10.1111/febs.12361>.
 117. Schönert S, Seitz S, Krafft H, Feuerbaum EA, Andernach I, Witz G, Dahl MK. Maltose and maltodextrin utilization by *Bacillus subtilis*. *J Bacteriol.* 2006;188:3911-22. <https://doi.org/10.1128/JB.00213-06>.

-
118. Ojima T, Saburi W, Yamamoto T, Kudo T. Characterization of *Halomonas* sp. strain H11 α -glucosidase activated by monovalent cations and its application for efficient synthesis of α -D-glucosylglycerol. *Appl Environ Microbiol.* 2012;78:1836-45. <https://doi.org/10.1128/AEM.07514-11>.
119. Li W, Xue Y, Li J. A cold-adapted and glucose-stimulated type II α -glucosidase from a deep-sea bacterium *Pseudoalteromonas* sp. K8. *Biotechnol Lett.* 2016;38:345-9. <https://doi.org/10.1007/s10529-015-1987-x>.
120. Okuyama M, Saburi W, Mori H, Kimura A. α -Glucosidases and α -1,4-glucan lyases: structures, functions, and physiological actions. *Cell Mol Life Sci.* 2016;73:2727-51. <https://doi.org/10.1007/s00018-016-2247-5>.
121. Watanabe R, Arimura Y, Ishii Y, Kirimura K. Crystal structure of α -glucosyl transfer enzyme XgtA from *Xanthomonas campestris* WU-9701. *Biochem Biophys Res Commun.* 2020;526:580-5. <https://doi.org/10.1016/j.bbrc.2020.03.109>.
122. Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Schnitzler Parker M, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricomutum* deduced from comparative whole genome analysis. *PLoS ONE.* 2008;3:e1426. <https://doi.org/10.1371/journal.pone.0001426>.
123. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. *Eukaryot Cell.* 2009;8:1038-50. <https://doi.org/10.1128/EC.00079-09>.
124. Gilmore MC, Cava F. Bacterial peptidoglycan recycling. *Trends Microbiol* 2025;33:340-53. <https://doi.org/10.1016/j.tim.2024.11.004>.
125. Jürgens UJ, Drews G, Weckesser J. Primary structure of the peptidoglycan from the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6714. *J Bacteriol.* 1983;154:471-8. <https://doi.org/10.1128/jb.154.1.471-478.1983>.
126. Suzuki S, Kawachi M, Tsukakoshi C, Nakamura A, Hagino K, Inouye I, Ishida K-I. Unstable relationship between *Braarudosphaera bigelowii* (= *Chrysochromulina parkeae*) and its nitrogen-fixing endosymbiont. *Front Plant Sci.* 2021;12:749695. <https://doi.org/10.3389/fpls.2021.749895>.
127. Vollmer W, Joris B, Charlier P, Foster S. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev.* 2008;32:259-86. <https://doi.org/10.1111/j.1574-6976.2007.00099.x>.
128. Mayer C, Kluj RM, Mühleck M, Walter A, Unsleber S, Hoffmann I, Borisova M. Bacteria's different ways to recycle their own cell wall. *Intl J Med Microbiol.* 2019;309:1151326. <https://doi.org/10.1016/j.ijmm.2019.06.006>
129. Cheng Q, Li H, Merdek K, Park JT. Molecular characterization of the β -N-acetylglucosaminidase of *Escherichia coli* and its role in cell wall recycling. *J Bacteriol.* 2000; 182:4836-40. <https://doi.org/10.1128/jb.182.17.4836-4840.2000>.
130. Park JT, Uehara T. How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan). *Microbiol Mol Biol Rev.* 2008;72:211-27. <https://doi.org/10.1128/MMBR.00027-07>.
131. Voadlo DJ, Mayer C, He S, Withers SG. Mechanism of action and identification of Asp242 as the catalytic nucleophile of *Vibrio furnisii* N-acetyl- α -D-glucosaminidase using 2-acetamido-2-deoxy-5-fluoro-R-L-idopyranosyl fluoride. *Biochem.* 2000;39:117-26. <https://doi.org/10.1021/bi991958d>.
132. Tsujibo H, Hatano N, Mikami T, Hirasawa A, Miyamoto K, Inamori Y. A novel β -N-acetylglucosaminidase from *Streptomyces thermoviolaceus* OPC-520: Gene cloning, expression, and assignment to family 3 of the glycosyl hydrolases. *Appl Environ Microbiol.* 1998;64:2920-4. <https://doi.org/10.1128/aem.64.8.2920-2924.1998>.
133. Choi K-H, Seo JY, Park K-M, Park C-S, Cha J. Characterization of glycosyl hydrolase family 3 β -N-acetylglucosaminidases from *Thermotoga maritima* and *Thermotoga neapolitana*. *J Biosci Bioeng.* 2009;108:455-9. <https://doi.org/10.1016/j.jbiosc.2009.06.003>.
134. Reith J, Mayer C. Peptidoglycan turnover and recycling in Gram-positive bacteria. *Appl Microbiol Biotechnol.* 2011;92:1-11. <https://doi.org/10.1007/s00253-011-3486-x>.
135. Saier Jr MH, Reddy VS, Moreno-Hagelsieb G, Handargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, Medrano-Soto A. The transporter classification database (TCDB): 2021 update. *Nucl Acids Res.* 2011;49:D461-7. <https://doi.org/10.1093/nar/gkaa1004>.
136. Kazanov MD, Li X, Gelfand MS, Osterman AL, Rodionov DA. Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermatogae phylum. *Nucl Acids Res.* 2013;41:790-803. <https://doi.org/10.1093/nar/gks1184>.

-
137. Cui J, Davidson AL. ABC solute importers in bacteria. *Essays Biochem.* 2011;50:85-99. <https://doi.org/10.1042/BSE0500085>.
138. Scheepers GH, Lycklama a Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* 2016;590:4393-401. <https://doi.org/10.1002/1873-3468.12445>.
139. Meeks JC, Elhai J. Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol Mol Biol Rev.* 2002;66:94-121. <https://doi.org/10.1128/MMBR.66.1.94>.
140. Herrero A, Stavans J, Flores E. The multicellular nature of filamentous heterocyst-forming cyanobacteria. *FEMS Microbiol Rev.* 2016;40:831-54. <https://doi.org/10.1093/femsre/fuw029>.
141. Burnat M, Herrero A, Flores E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci.* 2014;111:3825-8. <https://doi.org/10.1073/pnas.1318564111>.
142. Flores E, Nieves-Mori3n M, Mullineaux CW. Cyanobacterial septal junctions: Properties and regulation. *Life.* 2018;9:1. <https://doi.org/10.3390/life9010001>.
143. Nieves-Mori3n M, Flores E, Foster RA. Predicting substrate exchange in marine diatom-heterocystous cyanobacteria symbioses. *Environ Microbiol.* 2020;22:2027-52. <https://doi.org/10.1111/1462-2920.15013>.
144. Osborne B, Bergman B. Why does *Gunnera* do it and other angiosperms don't? An evolutionary perspective on the *Gunnera-Nostoc* symbiosis. *Microbiol Monogr.* 2009;8:207-24. https://doi.org/10.1007/7171_2007_116.
145. Lindblad P, Bergman B. Glutamine synthetase: activity and localization in cyanobacteria of the cycads *Cycas revolute* and *Zamia skinneri*. *Planta.* 1986;169:1-7. <https://doi.org/10.1007/BF01369768>.
146. Meeks JC. Physiological adaptations in nitrogen-fixing *Nostoc*-plant symbiotic associations. In: Pawlowski K, editor. *Prokaryotic symbionts in plants*. *Microbiol Monogr.* 2009;8:181-205. Springer, Berlin. https://doi.org/10.1007/7171_2007_101.
147. Flores E, Herrero A. Assimilatory nitrogen metabolism and its regulation. In: Bryant DA, editor. *The molecular biology of cyanobacteria*. Kluwer Academic Publishers, Boston, Mass. 1994. p.487-517. https://doi.org/10.1007/978-94-011-0227-8_16.
148. White J, Prell J, James EK, Poole P. Nutrient sharing between symbionts. *Plant Physiol.* 2007;144:604-14. <https://doi.org/10.1104/pp.107.097741>.
149. Clarke VC, Loughlin PC, Day DA, Smith PMC. Transport processes of the legume symbiosome membrane. *Front Plant Sci.* 2014;5:699. <https://doi.org/10.3389/fpls.2014.00699>.
150. Kleiner D. The transport of NH₃ and NH₄⁺ across biological membranes. *Biochim Biophys Acta.* 1981;639:41-52. [https://doi.org/10.1016/0304-4173\(81\)90004-5](https://doi.org/10.1016/0304-4173(81)90004-5).
151. Rai AN, S3derb3ck E, Bergman B. Cyanobacterium-plant symbioses. *New Phytol.* 2000;147:449-81. <https://doi.org/10.1046/j.1469-8137.2000.00720.x>.
152. Flores E, Romanovicz DK, Nieves-Mori3n M, Foster RA, Villareal TA. Adaptation to an intracellular lifestyle by a nitrogen-fixing, heterocyst-forming cyanobacterial endosymbiont of a diatom. *Front Microbiol.* 2022;13:799362. <https://doi.org/10.3389/fmicb.2022.799362>.
153. Kneip C, Lockhart P, Vo3 C, Maier U-G. Nitrogen fixation in eukaryotes – New models for symbiosis. *BMC Evol Biol.* 2007;7:55. <https://doi.org/10.1186/1471-2148-7-55>.
154. Hagino K, Onuma R, Kawachi M, Horiguchi T. Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLoS ONE.* 2013;8:381749. <https://doi.org/10.1371/journal.pone.0081749>.
155. Garneau AP, Slimani S, Fiola MJ, Tremblay LE, Isenring P. Multiple facets and roles of Na⁺-K⁺-Cl⁻ cotransport: Mechanisms and therapeutic implications. *Physiol.* 2020;35:415-29. <https://doi.org/10.1152/physiol.00012.2020>.
156. Lytle C, McManus TJ, Haas M. A model of Na-K-2Cl cotransport based on ordered ion binding and glide symmetry. *Am J Physiol.* 2008;274:C299-C309. <https://doi.org/10.1152/ajpcell.1998.274.2.C299>.
157. Worrell RT, Merk L, Matthews JB. Ammonium transport in the colonic crypt cell line, T84: role for Rhesus glycoproteins and NKCC1. *Am J Physiol Gastrointest Liver Physiol.* 2008;294:G429-G440. <https://doi.org/10.1152/ajpgi.00251.2006>.

158. Chew TA, Orlando BJ, Zhang J, Latorraca NR, Wang A, Hollingsworth SA, Chen D-H, Dror RO, Liao M, Feng L. Structure and mechanism of the cation-chloride cotransporter NKCC1. *Nature*. 2019;572:488-492. <https://doi.org/10.1038/s41586-019-1438-2>.
159. Rasmussen T, Rasmussen A. Bacterial mechanosensitive channels. In: Harris JR, Boekema EJ, editors. *Membrane protein complexes: structure and function*. Subcellular Biochemistry 87. Springer, Singapore. 2018. p.83-116. https://doi.org/10.1007/978-981-10-7757-9_4.
160. Becker M, Börngen K, Nomura T, Battle AR, Marin K, Martinac B, Kramer R. Glutamate efflux mediated by *Corynebacterium glutamicum* MscCG, *Escherichia coli* MscS, and their derivatives. *Biochim Biophys Acta*. 2013;1828:1230-40. <http://dx.doi.org/10.1016/j.bbamem.2013.01.001>.
161. Wang Y, Cao G, Xu D, Fan L, Wu X, Ni X, Zhao S, Zheng P, Sun J, Ma Y. A novel *Corynebacterium glutamicum* L-glutamate exporter. *Appl Environ Microbiol*. 2018;84:302691-17. <https://doi.org/10.1128/AEM.02691-17>
162. Cox CD, Nakayama Y, Nomura T, Martinac B. The evolutionary 'tinkering' of MscS-like channels: generation of structural and functional diversity. *Pflugers Arch – Eur J Physiol*. 2015;467:3-13. <https://doi.org/10.03.239/s00424-014-1522-2>.
163. Kawasaki H, Martinac B. Mechanosensitive channels of *Corynebacterium glutamicum* functioning as exporters of L-glutamate and other valuable metabolites. *Curr Opin Chem Biol*. 2020;59:77-83. <https://doi.org/10.1016/j.cbpa.2020.05.005>.
164. Miller S, Bartlett W, Chandrasekaran S, Simpson S, Edwards M, Booth IR. Domain organization of the MscS mechanosensitive channel of *Escherichia coli*. *EMBO J*. 2003;22:36-46. <https://doi.org/10.1093/emboj/cdg011>.
165. Nakamura J, Hirano S, Ito H, Wachi M. Mutations of the *Corynebacterium glutamicum* NCgl1221 Gene, encoding a mechanosensitive channel homolog, induce l-glutamic acid production. *Appl Environ Microbiol*. 2007;73:4491-8. <https://doi.org/10.1128/AEM.02446-06>.
166. Nakayama Y, Yoshimura K, Iida H. A Gain-of-function mutation in gating of *Corynebacterium glutamicum* NCgl1221 causes constitutive glutamate secretion. *Appl Environ Microbiol*. 2012;78:5432-4. <https://doi.org/10.1128/AEM.01310-12>.
167. Frail S, Steele-Ogus M, Doenier J, Moulin SLY, Braukmann T, Xu S, Yeh E. Genomes of nitrogen-fixing eukaryotes reveal a non-canonical model of organogenesis. *bioRxiv*. 2025. <https://doi.org/10.1101/2024.08.27.609708>.
168. Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K, Nowack ECM. Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. *Curr Biol*. 2017;27:2763-73. <http://doi.org/10.1016/j.cub.2017.08.010>.
169. Smith AG, Croft MT, Moulin M, Webb ME. Plants need their vitamins too. *Curr Opin Plant Biol*. 2007;10:266-75. <https://doi.org/10.1016/j.pbi.2007.04.009>.
170. Arnal N, Alban C, Quadrado M, Grandjean O, Mireau H. The *Arabidopsis* Bio2 protein requires mitochondrial targeting for activity. *Plant Mol Biol*. 2006;62:471-9. <https://doi.org/10.1007/s11103-006-9034-x>.
171. Rébeillé F, Ravanel S, Jabrin S, Douce R, Storozhenko S, Van Der Straeten D. Foliates in plants: biosynthesis, distribution, and enhancement. *Physiol Plant*. 2006;126:330-42. <https://doi.org/10.1111/j.1399-3054.2005.00587.x>.
172. Esteban R, Ariz I, Cruz C, Moran JF. Review: Mechanisms of ammonium toxicity and the quest for tolerance. *Plant Sci*. 2016;248:92-101. <http://doi.org/10.1016/j.plantsci.2016.04.008>.
173. Alverson AJ, Jansen RK, Theriot EC. Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by Thalassiosiroid diatoms. *Mol Phylog Evol*. 2007;45:193-210. <https://doi.org/10.1016/j.ympev.2007.03.024>.
174. Goodhead I, Darby AC. Taking the pseudo out of pseudogenes. *Curr Opin Microbiol* . 2015;23:102-9. <https://doi.org/10.1016/j.mib.2014.11.012>.
175. Wolfe-Simon F, Grzebyk D, Schofield O. The role and evolution of superoxide dismutase in algae. *J Phycol*. 2005;41:453-65. <https://doi.org/10.1111/j.1529-8817.2005.00086.x>.
176. Dupont CL, Neupane K, Shearer J, Palenik B. Diversity, function and evolution of genes coding for putative Ni-containing superoxide dismutases. *Environ Microbiol*. 2008;10:1831-43. <https://doi.org/10.1111/j.1462-2920.2008.01604.x>.

-
177. Geitler L. Zur Entwicklungsgeschichte der Epithemia den Epithemia, Rhopalodia und Denticula (Diatomophyceae) und ihre vermutlich symbiotischen Sphäroidkörper. *Plant Syst Evol.* 1977;128:259-75. <https://link.springer.com/content/pdf/10.1007/BF00984562.pdf>.
178. Dhaouadi F, Awwad F, Diamond A, Desgagné-Penix. Diatoms' breakthroughs in biotechnology: *Phaeodactylum tricornutum* as a model for producing high-added value molecules. *Am J Plant Sci.* 2020;11:1632-70. <https://doi.org/10.4236/ajps.2020.1110118>.
179. Elhai J, Taton A, Massar J, Myers JK, Travers M, Casey J, Slupesky M, Shrager J. BioBIKE: A web-based, programmable, integrated biological knowledge base. *Nucl Acids Res.* 2009;37:W28-32. <https://doi.org/10.1093/nar/gkp354>.
180. Moore LR, Caspi R, Campbell DA, Casey JR, Crevecoeur S, Lea-Smith DJ, Long B, Omar NM, Paley SM, Schmelling NM, Torrado A, Zehr JP, Karp PD. CyanoCyc cyanobacterial web portal. *Front Microbiol.* 2024;15:1340413. <https://doi.org/10.3389/fmicb.2024.1340413>.
181. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389-3402. <https://doi.org/10.5511/plantbiotechnology.19.145>.
182. Wolf YI, Koonin EV. A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biol Evol.* 2012;4:1286-94. <https://doi.org/10.1093/gbe/evs100>.
183. Li W and 20 others. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucl Acids Res.* 2021; 49:D1020-8. <https://doi.org/10.1093/nar/gkaa1105>.
184. Keseler IM and 20 others. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucl Acids Res.* 2016;D543-550. <https://doi.org/10.1093/nar/gkw1003>.
185. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673-80. <https://doi.org/10.5511/plantbiotechnology.19.145>.
186. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56:564-77. <https://doi.org/10.1080/10635150701472164>.
187. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307-10. <https://doi.org/10.1093/sysbio/syq010>.
188. Lefort V, Longueville J-E, Gascuel O. SMS: Smart model selection in PhyML. *Mol Biol Evol.* 2017;34:2422-4. <https://doi.org/10.1093/molbev/msx149>.
189. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. *Prot Sci.* 2021;31:47-53. <https://doi.org/10.1002/pro.4172>.