

# Genera specific distribution of DEAD-box RNA helicases in cyanobacteria

Denise S. Whitford\*, Brendan T. Whitman and George W. Owttrim

## Abstract

Although RNA helicases are essentially ubiquitous and perform roles in all stages of RNA metabolism, phylogenetic analysis of the DEAD (Asp-Glu-Ala-Asp)-box RNA helicase family in a single phylum has not been performed. Here, we performed a phylogenetic analysis on DEAD-box helicases from all currently available cyanobacterial genomes, comprising a total of 362 helicase protein sequences from 280 strains. DEAD-box helicases belonging to three distinct clades were observed. Two clades, the CsdA (cold shock DEAD-box A)-like and RhIE (RNA helicase E)-like helicases, cluster with the homologous proteins from *Escherichia coli*. The third clade, the CrhR (cyanobacterial RNA helicase Redox)-like helicases, is unique to cyanobacteria and characterized by a conserved sequence motif in the C-terminal extension. Restricted distribution is observed across cyanobacterial diversity with respect to both helicase type and strain. CrhR-like and CsdA-like helicases essentially never occur together, while RhIE always occurs with either a CrhR-like or CsdA-like helicase. CrhR-like and RhIE-like proteins occurred in filamentous cyanobacteria of the orders *Nostocales*, *Oscillatoriales* and *Synechococcales*. Similarly, CsdA- and RhIE-like proteins are restricted to unicellular cyanobacteria of the genera *Cyanobium* and *Synechococcus*. In addition, the unexpected occurrence of RhIE in two *Synechococcus* strains suggests recent acquisition and evolutionary divergence. This study, therefore, raises physiological and evolutionary questions as to why DEAD-box RNA helicases encoded in cyanobacterial lineages display restricted distributions, suggesting niches that require either CrhR or CsdA RNA helicase activity but not both. Extensive conservation of gene synteny surrounding the previously described *rimO*-*crhR* operon is also observed, indicating a role in the maintenance of photosynthesis. The analysis provides insights into the evolution, origin and dissemination of sequences within a single gene family to yield divergent functional roles.

## DATA SUMMARY

All of the protein sequences corresponding to the RNA helicases used in this study are publicly available in the GenBank database (<https://www.ncbi.nlm.nih.gov>). *rpoC2* and most 16S rDNA sequences are also available in the GenBank database. Additional 16S rDNA sequences were obtained from the Joint Genome Institute Integrated Microbial Genomes and Microbiomes (JGI IMG/M) database (<https://img.jgi.doe.gov/>). Accession numbers are provided within the article and the supplementary data.

## INTRODUCTION

Regulation of the expression of gene products occurs at many levels. At the post-transcriptional level, modulation

of RNA structure can regulate gene expression by altering RNA stability or translation efficiency. In many cases, these RNA structural changes require interactions with proteins, such as chaperones [1, 2] or helicases [3–5]. Monomeric and multimeric helicases are classified into two large families: superfamily 1 (SF1) and superfamily 2 (SF2) [3–5]. RNA helicases generally belong to SF2 and function in all aspects of RNA metabolism, regulating a diverse range of cellular processes by unwinding and, more rarely, annealing RNA duplexes, performing local strand separation, and by mediating RNA-protein interactions [4, 6–8]. Generally, RNA helicases share a conserved core structure, with two RecA-like functional domains containing a series of conserved sequence motifs that participate in RNA binding, and ATP binding and hydrolysis (Fig. S1, available with the online version of this

Received 28 July 2020; Accepted 23 December 2020; Published 04 February 2021

**Author affiliations:** <sup>1</sup>Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

**\*Correspondence:** Denise S. Whitford, [dwhitford@ualberta.ca](mailto:dwhitford@ualberta.ca)

**Keywords:** CrhR; cyanobacteria; DEAD-box RNA helicase; phylogenetic analysis.

**Abbreviations:** CrhR, cyanobacterial RNA helicase Redox; CsdA, cold shock DEAD-box A; DbpA, DEAD-box protein A; DEAD, Asp-Glu-Ala-Asp; NCBI, National Center for Biotechnology Information; RhIE, RNA helicase E; *rimO*, ribosomal protein S12 methyltransferase.

**Data statement:** All supporting data and protocols have been provided within the article or through supplementary data files. One supplementary table and four supplementary figures are available with the online version of this article.

000517 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

article) [9–11]. The largest family of RNA helicases, the DEAD (Asp-Glu-Ala-Asp)-box proteins, have 12 highly conserved sequence motifs [7], conferring ATP-dependent local RNA strand separation [12]. Some DEAD-box proteins can also displace proteins from RNA, or clamp to RNA and serve as a nucleation site for the recruitment of other proteins to form functional complexes [6, 7, 13]. In eukaryotes, DEAD-box helicases are essential for survival, as they are intimately involved in crucial cellular processes, including translation, splicing and RNA degradation [7]. Prokaryotic DEAD-box proteins also function in these processes; however, their function is generally only required under non-optimal conditions [14].

Classification of prokaryotic DEAD-box helicases was originally based on the RNA processes in which the protein participates [15, 16], but phylogenetic methods have also been used [14, 17]. Based on conserved sequence domains, López-Ramírez *et al.* [17] classified bacterial DEAD-box proteins into three major groups: (i) proteins that contain a DbpA (DEAD-box protein A) RNA binding domain, including CsdA (cold shock DEAD-box A) and DbpA from *Escherichia coli*, and DeaD/YxiN from *Bacillus subtilis*; (ii) proteins lacking the DbpA domain, including SrmB and RhlB from *E. coli*, and CshA, CshB and YfmL from *B. subtilis*; and (iii) RhlE (RNA helicase E)-like proteins. These classifications, based on sequence and structure, generally relate well to the cellular functions of the DEAD-box protein. For example, the DbpA RNA binding domain facilitates recognition and binding of hairpin 92 of the 23S rRNA by a DEAD-box RNA helicase [18–20]. All of the well-characterized DEAD-box proteins containing this domain, DbpA, CsdA and YxiN, participate in maturation of the 50S ribosomal subunit, and specifically the 23S rRNA [21–23]. These phylogenetic classifications aid in deducing putative functions of DEAD-box helicases from protein sequences as they are annotated in newly sequenced genomes.

*Cyanobacteria* is an ancient phylum of Gram-negative bacteria that evolved as early as 3.5 billion years ago [24]. These bacteria exhibit extreme genetic diversity and are distinguished by being the only known bacteria that perform oxygenic photosynthesis from which they obtain all of their energy [25, 26]. As a group, they have dramatic impacts on cycling of carbon dioxide, fixed carbon, oxygen and nitrogen, and primary productivity on a global scale [27]. In addition, their ability to sequester carbon dioxide makes them ideal platforms for carbon-neutral biotechnological applications [28]. Although phylogenetic analysis of bacterial DEAD-box RNA helicases has been performed [14, 17], extensive analysis within a specific phylum has not been conducted. An initial phylogenetic analysis of cyanobacterial DEAD-box helicases was performed [29]; however, there are a significantly larger number of cyanobacterial genomes now available that makes the analysis more robust. Therefore, it was of interest to evaluate how cyanobacterial DEAD-box RNA helicases relate to those encoded in other bacterial phyla.

### Impact Statement

RNA helicases are enzymes that affect all aspects of RNA metabolism and, thus, are crucial for regulation of gene expression. Here, we applied a phylogenetic approach to analyse DEAD (Asp-Glu-Ala-Asp)-box RNA helicase distribution across all known cyanobacterial genomes. Cyanobacteria perform oxygenic photosynthesis, constituting an extremely diverse and ancient phylum that has dramatic impacts on primary productivity, water quality, and nitrogen, carbon and oxygen cycling worldwide. The analysis indicated that cyanobacteria only possess DEAD-box RNA helicases belonging to three clades, two of which correspond to groups commonly observed in other bacteria. Members of the third clade, the CrhR (cyanobacterial RNA helicase Redox)-like helicases, are only encoded in cyanobacteria, suggesting a unique role related to maintenance of photosynthesis in this phylum. Restricted distribution of the helicase classes is also observed across cyanobacterial diversity with respect to both helicase type and strain, suggesting niche-specific functionality. The phylum-specific clade and restricted distribution of DEAD-box RNA helicases in cyanobacteria provide unique insights into gene evolution and niche-specific functioning.

Only four DEAD-box RNA helicases have previously been experimentally identified in cyanobacteria: CrhA from *Anabaena variabilis* UTCC 387 [30], CrhB and CrhC from *Nostoc* sp. PCC 7120 [31], and CrhR (cyanobacterial RNA helicase Redox) from *Synechocystis* sp. PCC 6803 (from here on *Synechocystis*) [32], with two examples, CrhA and CrhB, receiving only limited study. CrhC is specifically induced at low temperature [31], and interacts with ribosomes and localizes to the cytoplasmic face of the plasma membrane near the cell poles [33]. Expression of CrhR is regulated in response to the redox poise of the plastoquinone pool of the photosynthetic electron transport chain [32, 34]. CrhR was shown to co-sediment with polysomes and RNA degradosome components, and localize to the thylakoid membrane [35]. Consistent with localization to the thylakoid membrane, *crhR* mutation results in disruption of photosynthesis, particularly at low temperatures [36, 37]. As the cyanobacterial DEAD-box proteins, and in particular CrhR, demonstrate localization and functions unique to cyanobacteria, further study of this protein family across the diversity of cyanobacteria was warranted.

Here, we report a phylogenetic analysis of cyanobacterial DEAD-box proteins indicating that they form three major clades, two with homology to the *E. coli* DEAD-box proteins CsdA and RhlE, and a third unique to cyanobacteria, the CrhR-like helicases. The CrhR-like proteins are found throughout cyanobacteria diversity and share a conserved sequence domain in the C-terminal extension that identifies

them as a separate clade within the DEAD-box RNA helicase family. In addition, synteny surrounding the dicistronic *rimO*–*crhR* operon is extensively conserved, suggesting functional conservation and early acquisition of CrhR-like proteins in the cyanobacterial lineage. The three RNA helicase clades are also encoded in specifically restricted patterns throughout the phylum *Cyanobacteria*. CsdA-like helicases, which occur only in *Synechococcales*, have a DbpA RNA binding motif in the C-terminus and may have been acquired following a loss of the CrhR-like helicase in the ancestral lineage. Strains with RhLE-like helicases also encode at least one other DEAD-box protein in the genome. Based on similarities to characterized homologues, functions for CsdA, CrhR and RhLE in ribosome biogenesis, expression and maintenance of photosynthetic systems, and coordination of CsdA-CrhR function, respectively, are proposed.

## METHODS

### Sequence data

Predicted and known amino acid sequences of cyanobacterial DEAD-box RNA helicases were obtained from the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database. The helicase core domain (amino acids F9–R338) of the sole DEAD-box RNA helicase in *Synechocystis*, CrhR, was retrieved from the NCBI database (accession no. WP\_010873784.1) for use as the query sequence. A BLASTP search, restricted to cyanobacteria (taxid: 1117), yielded 461 unique DEAD-box RNA helicase sequences. Following alignment in MEGA7, these sequences were manually curated for full-length sequences that contained the 12 conserved motifs of the DEAD-box helicase protein family (Fig. S1). Curation was performed by initially removing sequences that lacked a start or stop codon, and subsequent deletion of sequences that did not span the full length of the helicase core region, extending from F9 to R338 of the *Synechocystis* sp. PCC 6803 *slr0083*-encoded helicase, CrhR. The remaining sequences were aligned with MUSCLE [38] and those containing motifs in the helicase core characteristic of all SF1 subfamily members or other SF2 helicase clades, other than DEAD-box family members, were removed. Following curation, 362 DEAD-box RNA helicase sequences from 280 strains of cyanobacteria were identified. The sequences of the five DEAD-box proteins from *E. coli* were also retrieved from the NCBI to verify branching and clade assignment.

### Sequence alignments and phylogenetic analysis

The helicase sequences were assembled in MEGA7 v7.0.21 [39], where multiple sequence alignments of the protein sequences were generated using the integrated MUSCLE algorithm [38], using default parameters. This alignment was exported in FASTA format and converted to relaxed PHYLIP format using NCLconverter v2.1 [40], accessed via the CIPRES Science Gateway [41].

The maximum-likelihood tree was created using RAXML-HPC Blackbox v8.2.10 [42], accessed via CIPRES [41] using default settings, with the exception of the amino acid

substitution matrix, which was set to LG [43]. Rapid bootstrapping ran for 252 replicates before auto-termination with MRE-based bootstrapping criteria [44]. The output tree was visualized in iTOL v4.3.2 [45]. Taxa colours correspond to the cyanobacterial order.

The Euler diagram representing the distribution of the helicase families across cyanobacterial strains was created as a three-set Euler diagram in EulerAPE v3.0 [46], consisting of CrhR-like, CsdA-like and RhLE-like helicases. The fourth set, for the unclassified helicases, was added with approximate scaling in Adobe Illustrator to visually demonstrate the intersection of the unclassified helicases with the CrhR-like and RhLE-like proteins.

### Gene context analysis

Gene synteny surrounding each DEAD-box protein in the respective cyanobacterial genome was identified using the NCBI Genome Viewer. Related genes were identified from annotations of conserved motifs.

### C-terminal motif identification

Alignments of the C-terminal domains for sequences from each helicase subfamily were generated in MUSCLE [38]. Alignments of a selection of sequences from the subfamily were used for visualization of the alignments. Alignments including all sequences in the subfamily were used to generate sequence logos in WebLogo 3 [47].

### Pfam domain identification

Protein sequences were searched against the Pfam database [48] using the EMBL-EBI web portal with an *E* value cut-off of 1.0.

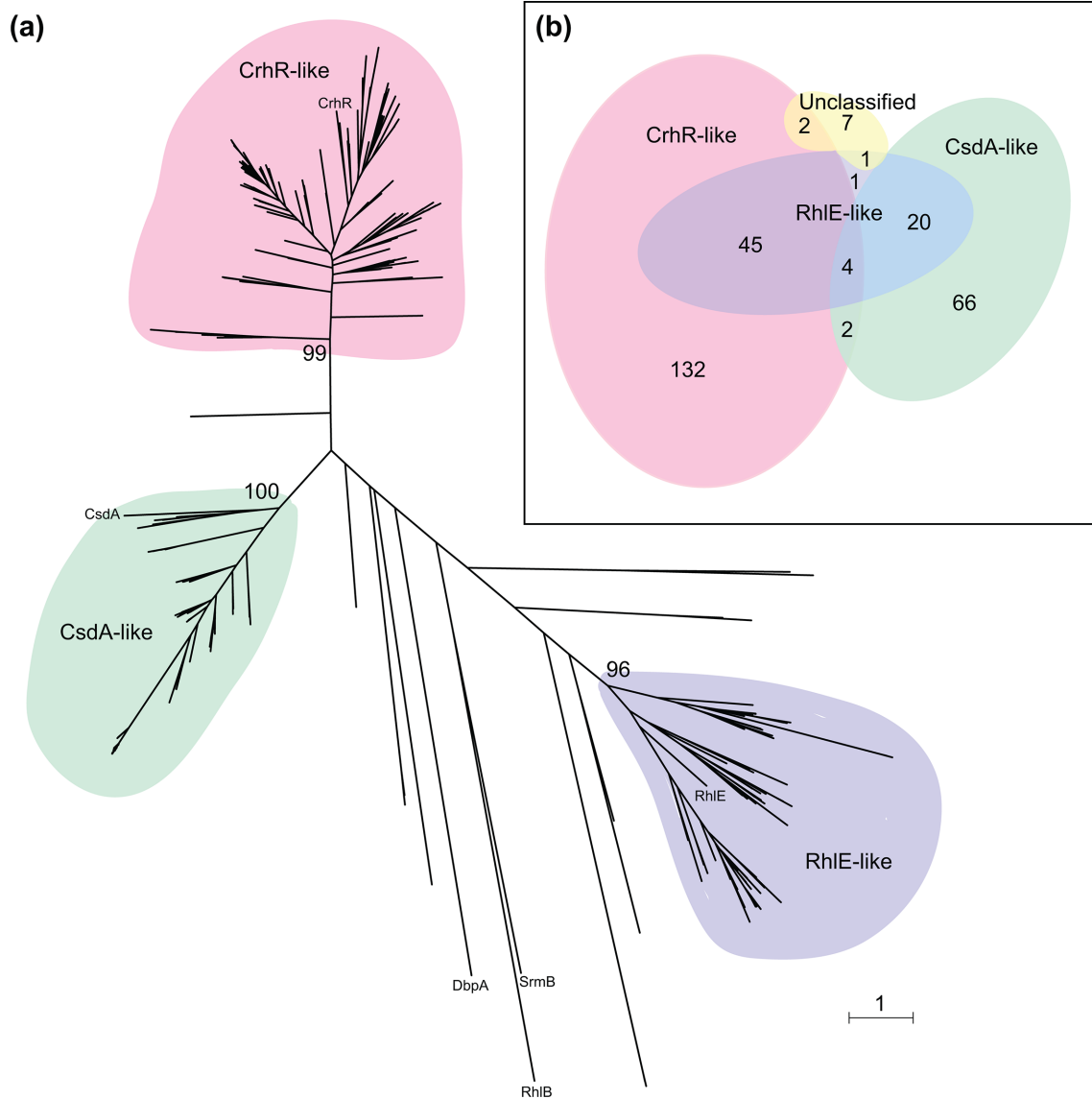
### *rpoC2* neighbour-joining tree

*rpoC2* gene sequences were identified by searching the genome sequences with the RpoC2 protein of *Synechocystis* (BA000022.2: 853497–857450). *rpoC2* nucleotide sequences were used instead of 16S rDNA for tree reconstruction as the genome for *Phormidesmis priestleyi* Ana is incomplete, lacking sequences for the 16S rDNA and several core conserved genes often used in concatenated gene trees. *rpoC2* nucleotide sequences were aligned by MUSCLE [38] with default parameters. A neighbour-joining tree was generated in MEGA7 v7.0.21 [39] and a maximum-likelihood tree was generated using RAXML-HPC Blackbox v8.2.10 [42] using default settings, accessed via the CIPRES [41] Science Gateway.

## RESULTS

### Cyanobacterial DEAD-box RNA helicases form three distinct clades

The amino acid sequence of the *Synechocystis* DEAD-box protein, CrhR, was used to generate a query sequence. The query sequence (F9–R338) was selected as it contained amino acids corresponding to the RNA helicase core, between the conserved phenylalanine upstream of the Q motif [49] and



**Fig. 1.** Cyanobacterial DEAD-box proteins cluster into three clades. The phylogenetic tree (a) shows the 362 cyanobacterial DEAD-box proteins form three major clades with bootstrap values  $\geq 95\%$ , corresponding to CsdA-like and RhIE-like proteins, and a clade unique to cyanobacteria, CrhR-like proteins, named for the sole DEAD-box RNA helicase encoded in the genome of *Synechocystis* sp. PCC 6803. Bootstrap values at the root of each clade, as well as the positions of the five DEAD-box proteins of *E. coli* and CrhR from *Synechocystis* sp. PCC 6803, are indicated. Branch lengths are proportional to the mean number of substitutions per amino acid site as indicated by the scale bar. The distribution of the DEAD-box protein clades identified in the 280 cyanobacterial strains used in the analysis is represented by the Euler diagram shown in the inset (b).

the final arginine of motif VI, HRIGR [50] (Fig. S1), which is highly conserved in all DEAD-box RNA helicases [7]. This sequence was used to search the NCBI non-redundant (nr) protein database, returning 461 unique cyanobacterial records, which were further curated to remove sequences from other SF1 and SF2 helicase families, as well as incomplete sequences. The resulting 362 cyanobacterial DEAD-box protein sequences (Table S1) were aligned, and the phylogenetic relationships were identified by reconstruction of an unrooted maximum-likelihood tree using RAXML [42].

The majority of the cyanobacterial DEAD-box RNA helicase proteins cluster into three clades, each having a bootstrap value  $\geq 95\%$  (Fig. 1a, b). Two of the *E. coli* helicase sequences, RhIE and CsdA, fall within these groups, while the remaining three, DbpA, RhIB and SrmB, do not group with the cyanobacterial helicase sequences. The three clades were designated as CsdA-like and RhIE-like, corresponding to the *E. coli* helicases that cluster with them, and CrhR-like, for the most-characterized protein sequence in this clade, CrhR, from *Synechocystis* [32]. Of the 362 cyanobacterial DEAD-box proteins in this analysis,

185 fall within the CrhR-like helicases, 92 in CsdA-like and 71 in RhIE-like (Fig. 1b). Fourteen of the cyanobacterial helicase proteins do not cluster with any of the three cyanobacterial clades nor with the three additional DEAD-box RNA helicase subfamilies present in *E. coli* (Fig. 1b).

Examination of helicase distribution demonstrates that 199 cyanobacterial strains contain a single member of one of the DEAD-box clades, 67 contain two forms, while only 4 strains contain members of all three clades (Fig. 1a). Of the strains encoding ungrouped DEAD-box proteins, three strains also encode either a CrhR-like or RhIE-like protein, while the remaining seven encode solely DEAD-box proteins that do not cluster with the three protein subfamilies defined in this study. It is important to note that only a few examples of cyanobacterial strains that do not encode DEAD-box RNA helicases were identified. Interestingly, these include the model organism *Synechococcus elongatus*, of which there are several genome sequences available [51]. DEAD-box proteins were also not identified in some early branching cyanobacteria, including *Synechococcus* sp. JA3-3Ab, *Synechococcus* sp. JA2-3B'a, and the proposed order *Gloeomargaritales*, for which only a single genome exists [52]. Based on the currently available sequence data, it can be concluded that cyanobacterial species encode between zero and three DEAD-box proteins that group into separate clades.

### CrhR-like proteins are conserved across cyanobacterial diversity

The maximum-likelihood tree of the cyanobacterial DEAD-box RNA helicases was also visualized, with the species, strain and order of the cyanobacteria encoding each protein indicated (Fig. S2). CrhR-like proteins were detected in every order of cyanobacteria that encode DEAD-box proteins except *Gloeobacterales* (Fig. S2), the oldest branching extant group of cyanobacteria [52]. This order represents a unique branch of cyanobacteria that lack thylakoid membranes, with photosynthetic light harvesting and electron transport located in the cytoplasmic membrane [53]. Interestingly, while *Gloeobacter violaceus* PCC 7421 and *Gloeobacter kilaueensis* JS1 encode DEAD-box proteins, only one of the *Gloeobacter* proteins clusters weakly with any of the three cyanobacterial DEAD-box helicase subfamilies defined in this study. In addition, CrhR-like helicases were not observed in the available sequences of the phylogenetically related non-photosynthetic proposed classes *Melainabacteria* and *Sericytochromatia* [26]. Conservation of the CrhR-like helicases throughout the diversity of cyanobacteria that evolved following branching from the *Gloeobacterales* and *Gloeomargaritales*, and absence of these proteins from organisms other than cyanobacteria, indicates that it is likely that the ancestral CrhR-like protein arose early in the radiation of the cyanobacterial crown group, after evolution of the non-photosynthetic *Melainabacteria* and the more basal *Sericytochromatia*.

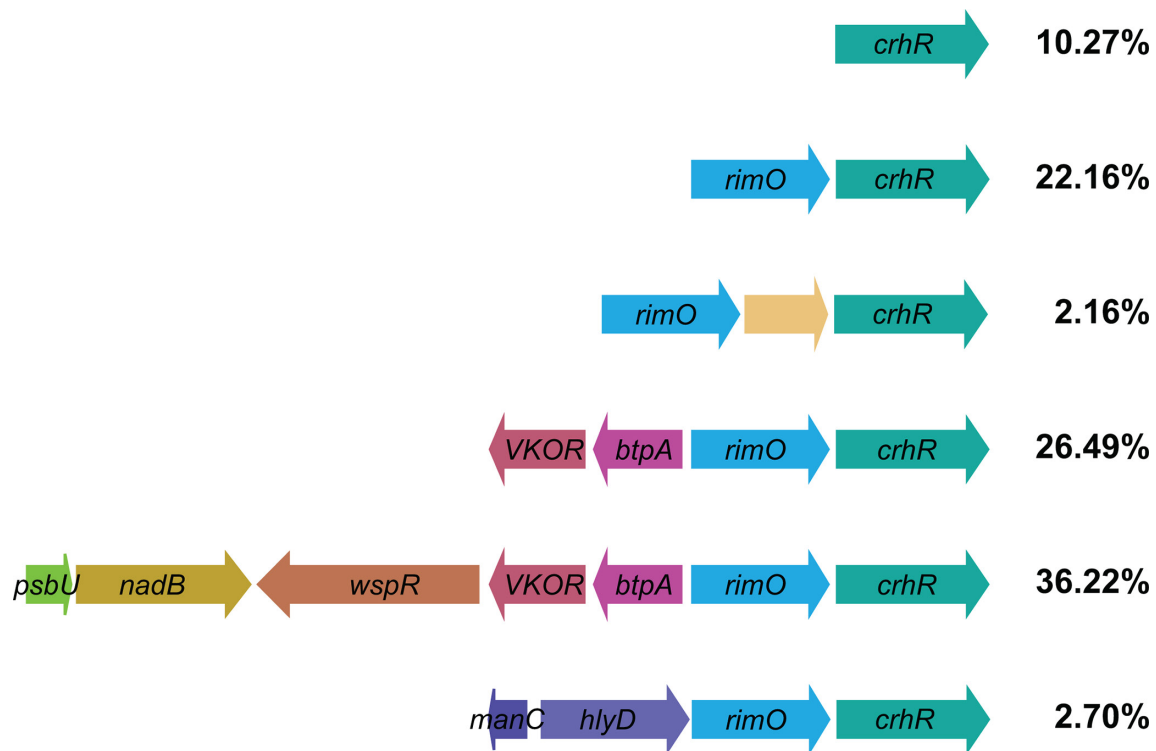
Conservation of the CrhR-like helicases in cyanobacteria was also investigated by examination of gene clustering surrounding the helicase. In *Synechocystis*, *crhR* is expressed

as a dicistronic operon, downstream of the gene encoding the ribosomal protein S12 methylthiotransferase, *rimO* [54]. Of the 185 cyanobacterial strains that encode CrhR-like helicases in this study, only 19 (10.27%) do not preserve the synteny of the *rimO-crhR* operon from *Synechocystis* (Fig. 2). Synteny is extensively maintained outside of the *rimO-crhR* operon with 62.71% of strains encoding a vitamin K epoxide reductase (VKOR) and photosystem I (PSI) stabilizing protein (*btpA*) upstream of the *rimO-crhR* operon (Fig. 2). Of those, 57.76% also encode a diguanylate cyclase (*wspR*), L-aspartate oxidase (*nadB*) and photosystem II (PSII) complex extrinsic protein (*psbU*) further upstream. These proteins perform roles in photosynthesis, membrane protein stability, protein folding and cellular energetics. The strong conservation of the *rimO-crhR* gene pair, as well as upstream genes in a preponderance of strains, supports an early acquisition of the CrhR-like proteins in cyanobacteria. Maintenance of the syntenic gene organization throughout cyanobacterial diversity also intimates that the pathways in which the encoded proteins function, including RimO and CrhR, are likely related, as products of highly conserved gene pairs are typically functionally related and often physically interact [55].

### Restricted distribution of DEAD-box proteins in cyanobacterial strains

Strict boundaries for DEAD-box RNA helicase distribution in cyanobacteria was evidenced by genomes that encode both CrhR- and RhIE-like proteins, which are restricted to filamentous cyanobacteria of the orders *Nostocales*, *Oscillatoriales* and *Synechococcales*. Similarly, co-occurrence of CsdA- and RhIE-like proteins is restricted to unicellular cyanobacteria of the genera *Cyanobium* and *Synechococcus*. The three exceptions are *Synechocystis* sp. PCC 7509, *Synechococcus* sp. NIES-970 and *Synechococcus* sp. NKBG15041c, which are unicellular strains that encode both a CrhR-like and RhIE-like helicase. Polyphyly has long been a known issue in the genera *Synechococcus* [56] and *Synechocystis* [57]. 16S rDNA and concatenated gene trees group *Synechocystis* sp. PCC 7509 near strains from the order *Nostocales* rather than other *Synechococcales* [58], suggesting this unicellular cyanobacterium encodes both a CrhR-like and RhIE-like helicase due to its common lineage with *Nostocales* and should be reclassified. In contrast, the other two unicellular strains that encode CrhR and RhIE, *Synechococcus* sp. NIES-970 and *Synechococcus* sp. NKBG15041c, are closely related to *Synechococcus* sp. PCC 7002 [59, 60], which encodes only a CrhR-like helicase. In a species evolution context, the presence of RhIE in only these two strains of unicellular *Synechococcus* is an important observation, since other closely related *Synechococcus* strains only contain CrhR-like helicases. This unpredicted occurrence implies that RhIE was acquired relatively recently in these two strains. Further study will be required to understand the physiological and evolutionary significance of RhIE helicase acquisition by these *Synechococcus* strains.

Furthermore, proteins belonging to the CsdA-like subfamily are only encoded in one cyanobacterial order, primarily in unicellular strains of the *Synechococcales*. This holds for



**Fig. 2.** Conserved gene sequence surrounding CrhR-like DEAD-box RNA helicases. Gene synteny surrounding the DEAD-box RNA helicase in strains containing CrhR-like proteins. Percentages represent the relative proportion of the 185 CrhR-encoding cyanobacterial strains in this analysis possessing the indicated gene sequence surrounding CrhR. Functionally related genes have similar colours. The cut-off to be considered as a separate example of conserved gene context was conservation in 2% of the strains.

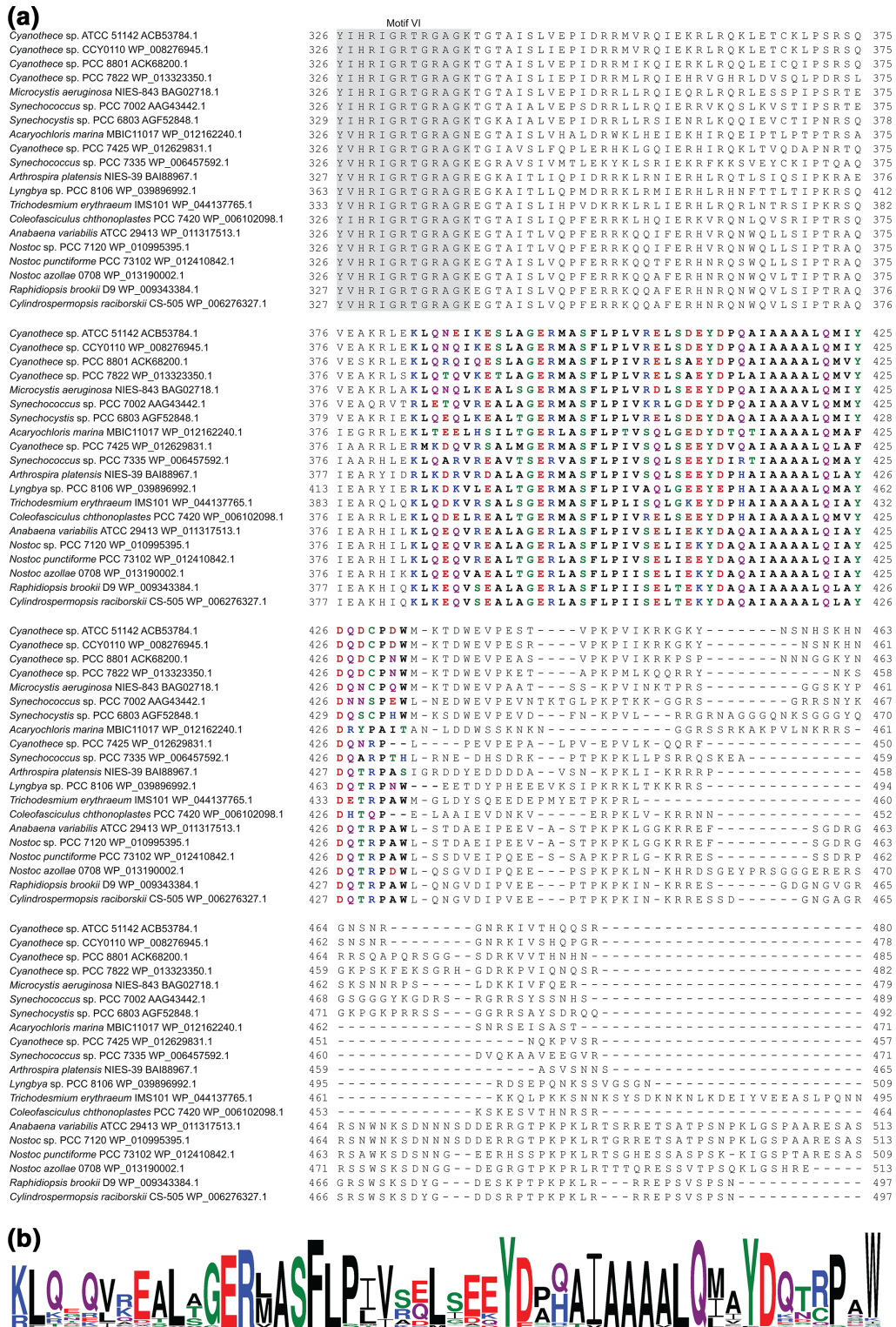
all species examined except for the *Synechococcales* strains *Leptolyngbya* sp. Heron Island J, *Leptolyngbya* sp. PCC 7375 and *Phormidesmis priestleyi* Ana. These three filamentous strains, as well as the unicellular *Synechococcus* sp. PCC 7335, are the only strains that encode DEAD-box proteins from all three subfamilies and are discussed below. Other than the four strains encoding proteins from all three groups, CsdA-like proteins are only encoded in species of *Acaryochloris*, *Cyanobium*, *Prochlorococcus* and *Synechococcus*. The strains from the genera *Cyanobium*, *Prochlorococcus* and *Synechococcus* that encode CsdA are monophyletic, based on both 16S rRNA [60] and concatenated gene [58] phylogenies. The most closely related species of cyanobacteria to these strains is *Synechococcus elongatus* [58], which does not encode a DEAD-box protein.

Co-occurrence of CrhR-like and CsdA-like proteins, excepting the four strains that encode all three subfamilies, is restricted to the genus *Acaryochloris* (Fig. S2). Cyanobacteria belonging to *Acaryochloris*, like *Gloeocapsa*, have divergent photosynthetic pathways, as they are capable of photosynthesis using the far-red spectrum, with chlorophyll *d* as the primary photosynthetic pigment and an altered complement of phycobiliproteins that do not form phycobilisomes on the surface of the thylakoid membrane [61]. As *Acaryochloris* sp. and the four strains that encode all three subfamilies are more distantly related to the other *Cyanobium*, *Prochlorococcus* and

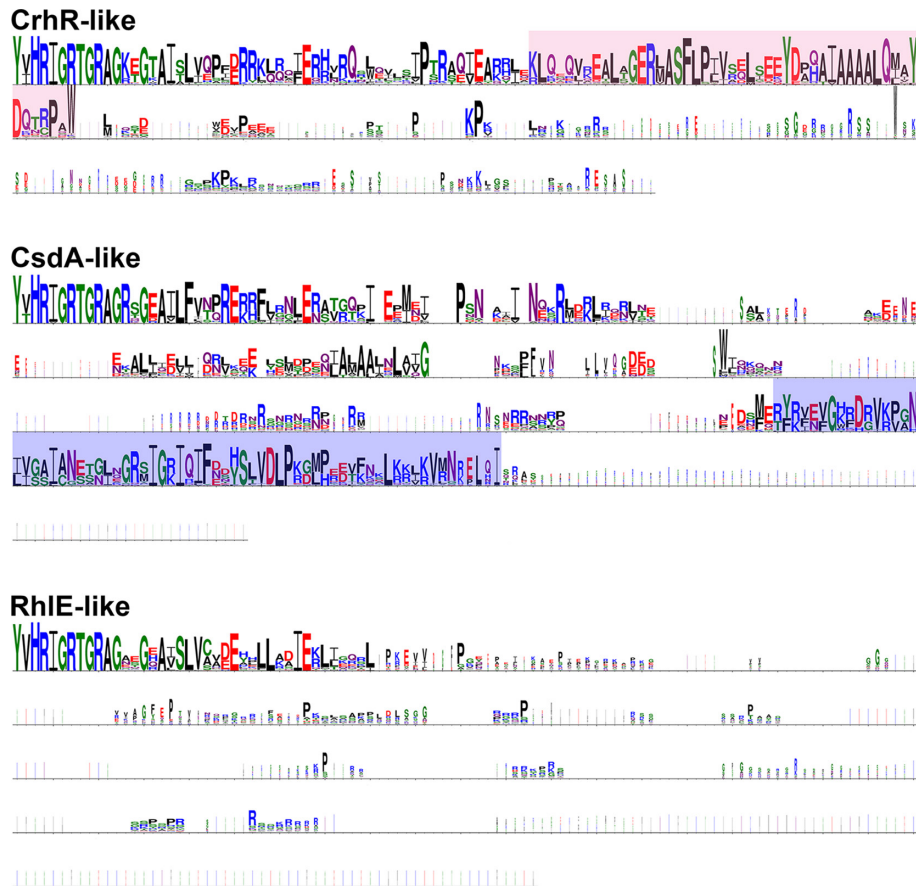
*Synechococcus* strains that encode CsdA [58], further analysis will be required to discern whether the CsdA-like proteins were acquired independently by the different cyanobacterial lineages.

### Conserved sequence motif in the C-terminus of CrhR-like proteins characterizes the CrhR-like clade

The C-terminal extension in CrhR-like helicases extends 150–200 amino acids downstream of the conserved helicase motif VI, HRIGR (Fig. S1). Within the C-terminal extension, a conserved sequence motif of approximately 50 amino acids is present in all CrhR-like helicases (Fig. 3). This motif is unique to and characterizes the CrhR-like RNA helicase clade as it is not found in the other two clades present in cyanobacteria (Fig. 4) and since other sequences containing this motif were not identified in either the Pfam protein families [48] or the NCBI protein databases. The DEAD-box protein from *G. violaceus* PCC 7421 (WP\_011142503.1), the nearest branching protein to the CrhR-like group of DEAD-box proteins, also does not possess the conserved CrhR sequence motif in its C-terminus; however, it does have shorter regions of sequence conservation with the CrhR-like proteins in the C-terminal extension (Fig. S3). This was unexpected as, other than proteins containing the DbpA RNA binding domain, sequence conservation in the C-terminal extension



**Fig. 3.** The C-terminal extension of CrhR-like helicases contains a unique, highly conserved sequence motif. (a) Alignment of the C-terminal extensions of selected CrhR-like DEAD-box proteins generated by MUSCLE. The final conserved motif in the DEAD-box helicase core, motif VI, HRIGRXXR, is indicated with a grey background. The unique conserved sequence motif characteristic of CrhR-like helicases is indicated in bold, with amino acids coloured based on their chemical properties. (b) Sequence logo of the conserved CrhR-specific sequence motif, with the size of the amino acid indicating its conservation. The sequence logo was generated from an alignment of all 185 CrhR-like DEAD-box RNA helicase sequences using WebLogo 3 [47].



**Fig. 4.** Conservation in the C-terminal extension of cyanobacterial DEAD-box proteins. Comparisons of the conservation in the C-terminal extensions of the three clades of cyanobacterial DEAD-box RNA helicases are shown using sequence logos, with the size of the amino acid indicating its conservation. The CrhR-specific sequence motif is indicated in pink and the DbpA RNA binding domain is marked in blue. Sequence logos were generated in WebLogo 3 from alignments of all cyanobacterial DEAD-box proteins in each clade.

of DEAD-box RNA helicases is generally minimal [17]. Therefore, it is likely that both the CrhR-like helicases and the DEAD-box protein from *G. violaceus* PCC 7421 share a common ancestral protein, with evolution of the CrhR-specific C-terminal motif occurring after the divergence of the order *Gloeobacterales*.

#### **CsdA-like proteins are characterized by a DbpA RNA binding domain in the C-terminal extension**

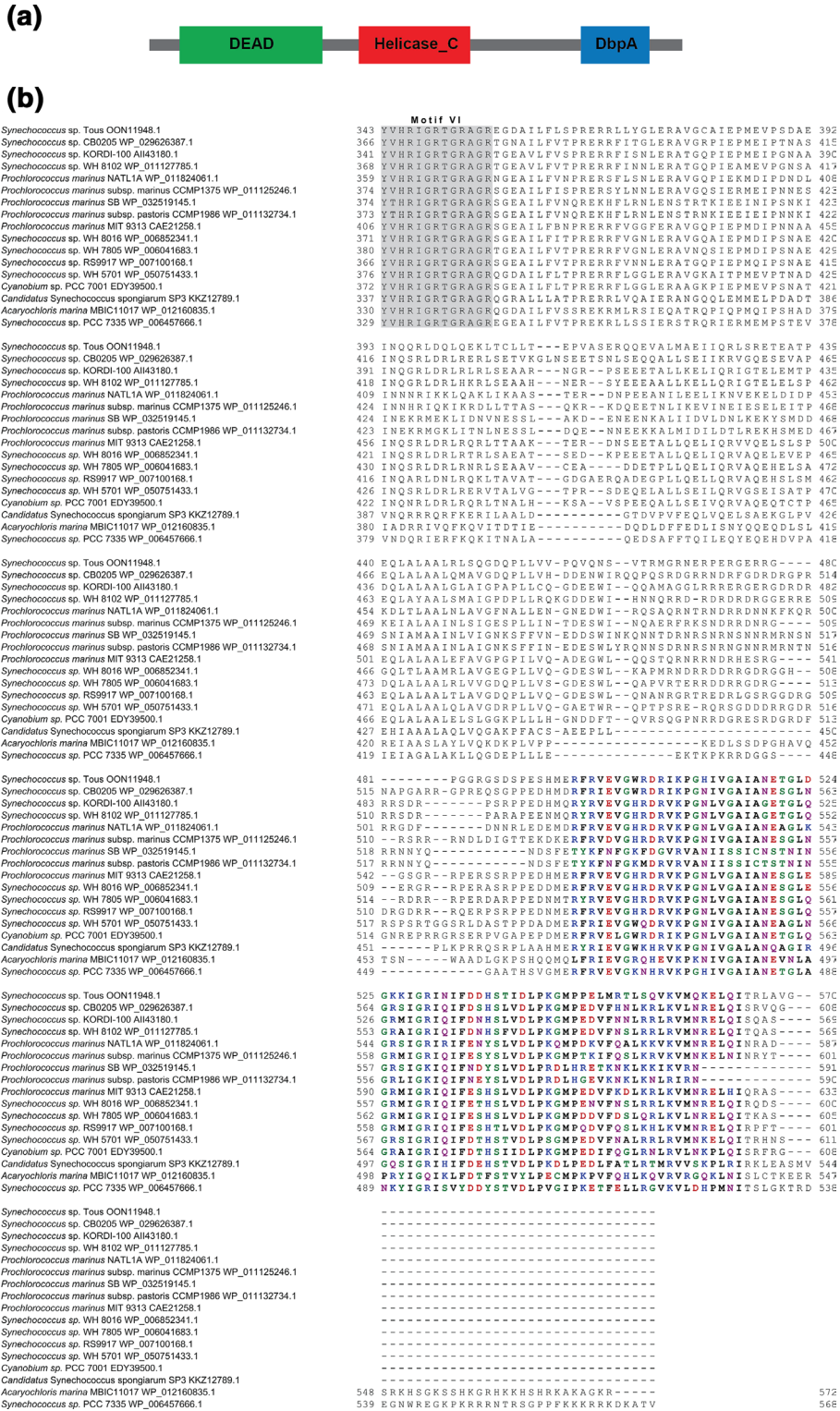
The DbpA RNA binding motif is an arginine-rich protein motif that facilitates recognition and binding of RNA, specifically hairpin 92 of the 23S rRNA [18–20]. As the *E. coli* DEAD-box RNA helicase CsdA, which has a DbpA RNA binding motif in its C-terminal extension (Fig. 4), clustered within the cyanobacterial CsdA-like clade, the C-terminal extension of helicases in this study were also examined for a DbpA RNA binding motif. The C-terminal extension of the CsdA-like DEAD-box RNA helicases extends 200–235 amino acids past the final motif of the helicase core region, the HRIGR motif VI (Fig. S1). It was found that in CsdA-like cyanobacterial helicases, the C-terminal extension contains

a DbpA RNA binding motif, a domain that is not observed in the CrhR-like helicases (Fig. 5). Two other cyanobacterial helicases also contain a DbpA RNA binding motif in their C-terminus: proteins from the unclassified cyanobacteria bacterium strain 13\_1\_20CM\_4\_61\_6 (OLE96526.1) and *Hassallia byssoidea* VB512170 (KIF31317.1). These strains cluster near the CsdA-like proteins and *E. coli* DbpA, respectively, with low bootstrapping support, suggesting a more distant relationship within the same larger group of DbpA-containing DEAD-box proteins.

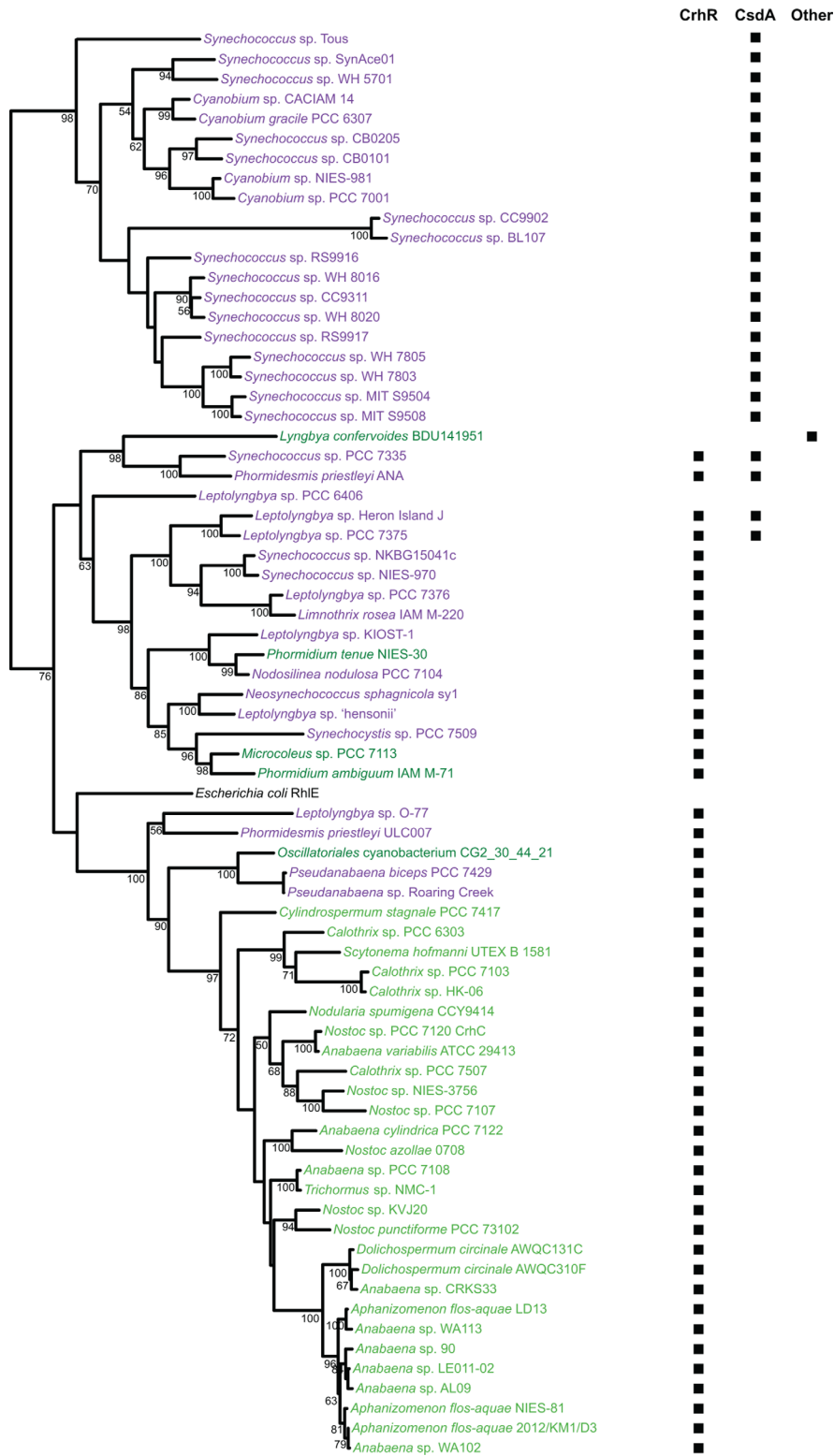
#### **RhIE-like helicases co-occur with another DEAD-box protein**

RhIE-like proteins essentially only occur in strains of cyanobacteria with at least one other DEAD-box protein (Fig. 6). *Leptolyngbya* sp. PCC 6406 is the only cyanobacterium in this study that encodes an RhIE-like helicase as its only identified DEAD-box protein (Table S1); however, as the genome sequence of this organism is a high-quality draft sequence, we predict that an additional DEAD-box RNA helicase, either

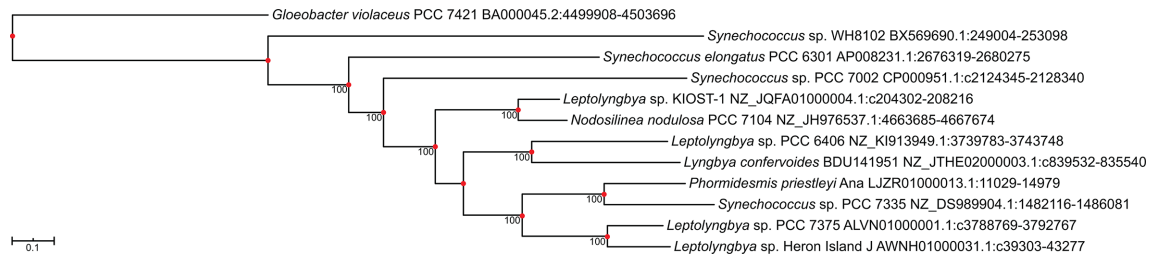




**Fig. 5.** The C-terminal extension of the CsdA-like helicases contains a DbpA RNA binding domain. (a) A schematic showing the Pfam domains present in the CsdA-like helicases: the DEAD/DEAH box helicase domain (DEAD, PF00270), the helicase conserved C-terminal domain (helicase C, PF00271) and the DbpA RNA binding domain (DbpA, PF03880). (b) Alignments generated with MUSCLE of the C-terminal extensions of selected CsdA-like helicases. The final conserved DEAD-box helicase motif, HRIGRXXR, is marked with a grey background on the alignment. The conserved DbpA RNA binding domain is indicated in bold, with amino acids coloured based on their chemical properties. Note that this domain is not present in members of the ChrR-like clade.



**Fig. 6.** Co-occurrence of RhIE-like helicases with CsdA- and CrhR-like proteins. The RhIE-like maximum-likelihood helicase tree was extracted from Fig. S2. Annotations indicate whether the second DEAD-box protein in each strain is from the CrhR-like or CsdA-like groups. Essentially, RhIE is only found in conjunction with either CrhR or CsdA. Bootstrap values  $\geq 50\%$  are shown at the nodes. Taxa colours indicate the cyanobacterial taxonomic order: purple, *Synechococcales*; light green, *Nostocales*; dark green, *Oscillatoriales*. The position of *E. coli* RhIE is provided to support the classification.



**Fig. 7.** *rpoC2* maximum-likelihood tree showing that the four cyanobacterial strains that encode DEAD-box proteins from all three clades of cyanobacterial DEAD-box proteins are closely related. Nucleotide sequences for the RNA polymerase  $\beta'$  subunit, *rpoC2*, were aligned by MUSCLE and used to reconstruct a maximum-likelihood tree with *G. violaceus* PCC 7421 as an outgroup. Several *Leptolyngbya* species and closely related cyanobacteria, as well as several *Synechococcus* species, were included to confirm the clustering of the four strains that encode all three DEAD-box proteins. Bootstrap values  $\geq 50\%$  are shown at the nodes. Branch lengths are proportional to the mean number of substitutions per nucleotide site as indicated by the scale bar. All nodes in the maximum-likelihood tree are duplicated in a neighbour-joining reconstruction, as indicated with a red circle.

CrhR- or CsdA-like, is encoded in the remaining gapped regions.

The RhLE-like proteins cluster into two major related groups that, apart from the four strains that encode all three types of cyanobacterial DEAD-box proteins, correlate with the co-occurrence of CsdA- or CrhR-like helicases (Fig. 6). This suggests that RhLE-like helicases were acquired at least twice in cyanobacterial lineages, once in the lineage with CsdA-like helicases and at least once in strains with CrhR-like proteins.

The RhLE-like proteins have a C-terminal extension of approximately 75–100 amino acids that lacks sequence conservation (Fig. 4). It should also be noted that all RhLE-like proteins from the order *Nostocales* had a substitution of phenylalanine for serine in the conserved motif III [62], resulting in a sequence change of SAT $\rightarrow$ FAT.

### Proteins from all three helicase clades are only encoded in four related strains of cyanobacteria

The only cyanobacterial strains that encode a member of all three clades, a CrhR-, CsdA- and RhLE-like helicase, within the same genome are *Leptolyngbya* sp. PCC 7375, *Leptolyngbya* sp. Heron Island J, *Phormidesmis priestleyi* Ana and *Synechococcus* sp. PCC 7335 (Fig. S2). Although these four strains are annotated as members of three separate genera, 16S rDNA, *rpoC2* and/or concatenated gene phylogenies indicate they are more closely related than another closely related strain, *Leptolyngbya* sp. PCC 6406 [58, 63, 64]. As all four of these strains have not been included in the same phylogenetic analysis to date, an *rpoC2* gene tree was reconstructed (Fig. 7) to confirm that the strains were closely related. As expected, *Leptolyngbya* sp. PCC 7375, *Leptolyngbya* sp. Heron Island J, *Phormidesmis priestleyi* Ana and *Synechococcus* sp. PCC 7335 cluster together (Fig. 7). This clustering was also supported by 16S rDNA analysis in which all nodes were duplicated, excepting the node nearest *Phormidesmis priestleyi* Ana (Fig. S4), for which a 16S rDNA sequence is not available. Clustering of these four strains confirms that, despite the differences in cellular morphology and current taxonomy, the strains of cyanobacteria that encode members of all three

groups of cyanobacterial DEAD-box proteins are closely related and require reclassification.

## DISCUSSION

Cyanobacteria are unique among prokaryotes, as they are the only bacteria that perform oxygenic photosynthesis. In this study, it was shown that the complement of DEAD-box RNA helicases encoded by cyanobacteria is also unique, as proteins with homology to the helicases of *E. coli*, the CsdA-like and RhLE-like protein groups, are only encoded within the genomes of select cyanobacterial taxa. In contrast, the DEAD-box RNA helicase clade that can be found across cyanobacterial diversity, the CrhR-like proteins, is unique to cyanobacteria. The C-terminal extension domain of the DEAD-box RNA helicase is a defining characteristic of this unique clade.

A sequence motif of approximately 50 amino acids that is uniquely observed only in the CrhR-like clade of cyanobacterial DEAD-box RNA helicases is located within the C-terminal extension. This motif is not observed in other organisms, including higher plants and algae, but is widely distributed in cyanobacteria, suggesting that the CrhR-specific domain arose as a specialization of a DEAD-box protein in the ancestral cyanobacterial lineage. This implies that CrhR-like helicases perform a specific function that is unique to cyanobacteria. As CrhR-like helicases are often the sole DEAD-box protein in a cyanobacterial genome, it is likely that they function in multiple RNA processes. We propose that the unique function of CrhR-like helicases likely relates to oxygenic photosynthesis. This conjecture is supported by the co-precipitation of CrhR from *Synechocystis* with polysomes and RNA degradosome components [35], and the altered expression of  $\sim 10\%$  of the *Synechocystis* transcriptome, consisting of both mRNA and non-coding RNAs, in a *crhR* truncation mutant [65]. Physiologically, and consistent with the altered expression profiles, CrhR mutation results in significant disruption to photosynthesis, with cells exhibiting reduced pigment abundance, oxygen evolution rates

and carbon dioxide fixation rates [36]. These physiological changes can also be associated with the regulation of *crhR* expression by a range of abiotic stresses [32, 34, 66, 67]. Taken together, these effects are consistent with CrhR RNA helicase localization to the thylakoid membrane [35], the site of light harvesting and ATP and NADPH formation in *Synechocystis*. Therefore, we propose that CrhR-like RNA helicases perform a unique function, coordinating expression of genes required to maintain oxygenic photosynthesis in response to abiotic stresses in cyanobacteria.

Support for CrhR functioning in maintenance of photosynthesis was also provided by the conserved gene synteny surrounding the CrhR-like proteins. The *rimO*-*crhR* operon of *Synechocystis* is conserved in 89.8% of strains with CrhR-like helicases. Over 60% of strains also have further conservation of synteny, with conserved genes relating to photosynthesis, energy transfer from the electron transport chain and membrane protein stability. In particular, the activities of many of these proteins could protect the photosynthetic electron transport chain from oxidative damage if an excess of reductant was to accumulate. For example, PsbU stabilizes the oxygen-evolving complex of photosystem II (PSII) and protects it from inactivation in response to stress, such as heat [68]. The vitamin K epoxide reductase (VKOR) and L-aspartate oxidase (*nadB*) act as electron sinks, as they both can use quinones from the electron transport chain as the source of energy for their biochemical activity [69, 70]. Expression of these proteins may be regulated under similar conditions as CrhR, which is induced in response to increasing reduction of the photosynthetic electron transport chain [34].

It is interesting that CrhR- and CsdA-like helicases are not observed in the same species, suggesting the requirement for divergent functionality. The CsdA-like helicases are characterized by a DbpA RNA binding domain in the C-terminus, similar to the canonical family member CsdA in *E. coli* [17]. The presence of the DbpA RNA binding domain, which assists DbpA in binding specifically to hairpin 92 of the 23S rRNA [18, 71], may indicate that the cyanobacterial CsdA-like helicases have a role in ribosome biogenesis, and more specifically maturation of the 23S rRNA, similar to DbpA [23, 72]. The domain may also provide more general RNA chaperone properties to the helicase, as some proteins with a DbpA RNA binding domain, including CsdA, do not bind to the 23S rRNA at hairpin 92 [73]. As these proteins have significant homology with CsdA, the cyanobacterial CsdA-like proteins may constitute a multifunctional helicase that can function in ribosome biogenesis, translation initiation and RNA degradation, as observed in *E. coli* [21, 23, 73–76]. The DbpA RNA binding domain is not encoded in CrhR-like helicases, suggesting CrhR is not involved in ribosome biogenesis.

Cyanobacterial RhIE-like helicases lack a conserved C-terminal domain. Interestingly, proteins from this group are essentially only encoded in genomes with at least one other DEAD-box helicase. This is consistent with the distribution of RhIE-like proteins generally in prokaryotes, with only

a limited number of genomes encoding solely an RhIE-like DEAD-box helicase [17]. This also supports conservation of the proposed function of the RhIE helicase in *E. coli* as a helicase that functions cooperatively with other DEAD-box proteins during ribosome biogenesis [77].

Thus, distribution and sequence divergence of the CsdA and RhIE groups of DEAD-box RNA helicases in cyanobacteria suggest they were acquired independently several times. The CrhR-like proteins are widely distributed throughout cyanobacterial diversity and contain a sequence domain unique to the CrhR-like clade of DEAD-box RNA helicases, suggesting they perform a cyanobacterial specific function. Overall, the analysis suggests that the precursor of CrhR-like DEAD-box helicases arose early in cyanobacterial evolution. A common ancestor of the genera *Synechococcus* and *Prochlorococcus* likely lost the CrhR-like helicase. Early branching *Synechococcus* and *Prochlorococcus*, based on 16S rRNA sequence [78], contain CrhR-like helicases; however, CrhR-like proteins are absent from *Synechococcus elongatus* PCC 7942 and the strains with CsdA-like helicases. It is possible that the CsdA-like helicases were acquired by horizontal gene transfer in the ancestral lineage that lacked a CrhR-like helicase and were maintained as they conferred a selective advantage. Acquisition of RhIE-like helicases in cyanobacteria was likely also by horizontal gene transfer, possibly several events, as they are generally only found in proteobacteria [17].

This study raises questions as to why DEAD-box RNA helicases encoded in cyanobacterial genomes display distributions restricted on both the family member type and strain levels. This suggests certain lineages of cyanobacteria require the activity of specific DEAD-box RNA helicases. The specificity is potentially related to habitat niche for marine species, since CsdA-like helicases are principally encoded in *Synechococcus* and *Prochlorococcus* species, which are primarily marine; however, this does not apply to the CrhR-like helicases, which occur in cyanobacteria that occupy a diverse range of habitats. Cyanobacteria primarily encode helicases classified into three DEAD-box RNA helicase clades on the basis of conserved sequences in the C-terminal extension. Two of these clades, the CsdA-like and RhIE-like helicases, have significant similarity to the corresponding *E. coli* proteins. The third clade, the CrhR-like helicases, are unique to cyanobacteria and are characterized by a sequence motif in the C-terminal extension that is only found in CrhR-like helicases. Based on prior characterization of CrhR and the observed syntenic gene conservation, it is likely that CrhR-like proteins are multifunctional helicases, performing roles in RNA metabolism related to maintenance of photosynthesis.

#### Funding information

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, grant number 171 319 to G.W.O.

#### Acknowledgements

We thank R. Glen Uhrig (University of Alberta, Canada) for his advice on molecular phylogenetics.

**Author contributions**

Conceptualization: D. S. W., B. T. W., G. W. O. Methodology: D. S. W., B. T. W., G. W. O. Validation: D. S. W. Formal analysis: D. S. W., G. W. O. Investigation: D. S. W. Resources: D. S. W., G. W. O. Data curation: D. S. W., G. W. O. Writing – original draft preparation: D. S. W. Writing – review and editing: D. S. W., G. W. O. Visualization: D. S. W., G. W. O. Supervision: G. W. O. Funding: G. W. O.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**References**

- Herschlag D. RNA chaperones and the RNA folding problem. *J Biol Chem* 1995;270:20871–20874.
- Woodson SA, Panja S, Santiago-Frangos A. Proteins that chaperone RNA regulation. In: Storz G, Papenfort K (eds). *Regulating with RNA in Bacteria and Archaea*. Washington, DC: American Society for Microbiology; 2019. pp. 385–397.
- Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases. *Gene* 2006;367:17–37.
- Jankowsky E. RNA helicases at work: binding and rearranging. *Trends Biochem Sci* 2011;36:19–29.
- Linder P, Fuller-Pace F. Happy birthday: 25 years of DEAD-box proteins. *Methods Mol Biol* 2015;1259:17–33.
- Jarmoskaite I, Russell R. RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* 2014;83:697–725.
- Linder P, Jankowsky E. From unwinding to clamping – the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 2011;12:505–516.
- Tanner NK, Linder P. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* 2001;8:251–262.
- Schmid SR, Linder P. D-E-A-D protein family of putative RNA helicases. *Mol Microbiol* 1992;6:283–291.
- Schütz P, Karlberg T, van den Berg S, Collins R, Lehtiö L et al. Comparative structural analysis of human DEAD-box RNA helicases. *PLoS One* 2010;5:e12791.
- Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1982;1:945–951.
- Yang Q, Jankowsky E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* 2006;13:981–986.
- Liu F, Putnam AA, Jankowsky E. DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. *Biochemistry* 2014;53:423–433.
- Redder P, Hausmann S, Khemic V, Yasrebi H, Linder P. Bacterial versatility requires DEAD-box RNA helicases. *FEMS Microbiol Rev* 2015;39:392–412.
- Khemic V, Linder P. RNA helicases in bacteria. *Curr Opin Microbiol* 2016;30:58–66.
- Owttrim GW. RNA helicases: diverse roles in prokaryotic response to abiotic stress. *RNA Biol* 2013;10:96–110.
- López-Ramírez V, Alcaraz LD, Moreno-Hagelsieb G, Olmedo-Álvarez G. Phylogenetic distribution and evolutionary history of bacterial DEAD-box proteins. *J Mol Evol* 2011;72:413–431.
- Diges CM, Uhlenbeck OC. *Escherichia coli* DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J* 2001;20:5503–5512.
- Kossen K, Uhlenbeck OC. Cloning and biochemical characterization of *Bacillus subtilis* YxiN, a DEAD protein specifically activated by 23S rRNA: delineation of a novel sub-family of bacterial DEAD proteins. *Nucleic Acids Res* 1999;27:3811–3820.
- Nicol SM, Fuller-Pace FV. The “DEAD box” protein DbpA interacts specifically with the peptidyltransferase center in 23S rRNA. *Proc Natl Acad Sci USA* 1995;92:11681–11685.
- Charollais J, Dreyfus M, Iost I. CsdA, a cold-shock RNA helicase from *Escherichia coli*, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res* 2004;32:2751–2759.
- Lehnik-Habrink M, Rempeters L, Kovács Ákos T, Wrede C, Baierlein C et al. DEAD-box RNA helicases in *Bacillus subtilis* have multiple functions and act independently from each other. *J Bacteriol* 2013;195:534–544.
- Peil L, Virumäe K, Remme J. Ribosome assembly in *Escherichia coli* strains lacking the RNA helicase DeaD/CsdA or DbpA. *FEBS J* 2008;275:3772–3782.
- Baumgartner RJ, Van Kranendonk MJ, Wacey D, Fiorentini ML, Saunders M et al. Nano-porous pyrite and organic matter in 3.5-billion-year-old stromatolites record primordial life. *Geology* 2019;47:1039–1043.
- Beck C, Knoop H, Axmann IM, Steuer R. The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics* 2012;13:56.
- Soo RM, Hemp J, Parks DH, Fischer WW, Hugenholtz P. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* 2017;355:1436–1440.
- Fuchsman CA, Palevsky HI, Widner B, Duffy M, Carlson MCG et al. Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *ISME J* 2019;13:2714–2726.
- Gilmour DJ. Microalgae for biofuel production. *Adv Appl Microbiol* 2019;109:1–30.
- Georg J. A hidden layer of genetic information – regulatory non-protein-coding RNAs in *Synechocystis* PCC6803. PhD thesis, Universität Freiburg, Breisgau, Germany; 2010.
- Magee WC. Characterization of a cyanobacterial RNA helicase gene. MSc thesis, University of Alberta, Edmonton, AB, Canada; 1997.
- Chamot D, Magee WC, Yu E, Owttrim GW. A cold shock-induced cyanobacterial RNA helicase. *J Bacteriol* 1999;181:1728–1732.
- Kujat SL, Owttrim GW. Redox-regulated RNA helicase expression. *Plant Physiol* 2000;124:703–714.
- El-Fahmawi B, Owttrim GW. Polar-biased localization of the cold stress-induced RNA helicase, CrhC, in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol Microbiol* 2003;50:1439–1448.
- Ritter SPA, Lewis AC, Vincent SL, Lo LL, Cunha APA et al. Evidence for convergent sensing of multiple abiotic stresses in cyanobacteria. *Biochim Biophys Acta* 2020;1864:129462.
- Rosana ARR, Whitford DS, Fahlman RP, Owttrim GW. Cyanobacterial RNA helicase CrhR localizes to the thylakoid membrane region and cosediments with degradosome and polysome complexes in *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 2016;198:2089–2099.
- Rosana ARR, Ventakesh M, Chamot D, Patterson-Fortin LM, Tarassova O et al. Inactivation of a low temperature-induced RNA helicase in *Synechocystis* sp. PCC 6803: physiological and morphological consequences. *Plant Cell Physiol* 2012;53:646–658.
- Sireesha K, Radharani B, Krishna PS, Sreedhar N, Subramanyam R et al. RNA helicase, CrhR is indispensable for the energy redistribution and the regulation of photosystem stoichiometry at low temperature in *Synechocystis* sp. PCC6803. *Biochim Biophys Acta* 2012;1817:1525–1536.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
- Lewis PO. NCL: a C++ class library for interpreting data files in NEXUS format. *Bioinformatics* 2003;19:2330–2331.
- Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the 2010 Gateway Computing Environments Workshop (GCE 2010)*. New Orleans, LA, USA: ACM; 2010. pp. 1–8.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.

43. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;25:1307–1320.
44. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *J Comput Biol* 2010;17:337–354.
45. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
46. Micallef L, Rodgers P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* 2014;9:e101717.
47. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–1190.
48. Finn RD, Mistry J, Tate J, Coghill P, Heger A et al. The Pfam protein families database. *Nucleic Acids Res* 2010;38:D211–D222.
49. Cordin O, Tanner NK, Doère M, Linder P, Banroques J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 2004;23:2478–2487.
50. Pause A, Méthot N, Sonenberg N. The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol* 1993;13:6789–6798.
51. Sugita C, Ogata K, Shikata M, Jikuya H, Takano J et al. Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. *Photosynth Res* 2007;93:55–67.
52. Moreira D, Tavera R, Benzerara K, Skouri-Panet F, Couradeau E et al. Description of *Gloeomargarita lithophora* gen. nov., sp. nov., a thylakoid-bearing, basal-branching cyanobacterium with intracellular carbonates, and proposal for *Gloeomargaritales* ord. nov. *Int J Syst Evol Microbiol* 2017;67:653–658.
53. Rippka R, Waterbury J, Cohen-Bazire GA. A cyanobacterium which lacks thylakoids. *Arch Microbiol* 1974;100:419–436.
54. Rosana ARR, Whitford DS, Migur A, Steglich C, Kujat-Choy SL et al. RNA helicase-regulated processing of the *Synechocystis rimO-crhR* operon results in differential cistron expression and accumulation of two sRNAs. *J Biol Chem* 2020;295:6372–6386.
55. Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 1998;23:324–328.
56. Robertson BR, Tezuka N, Watanabe MM. Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *Int J Syst Evol Microbiol* 2001;51:861–871.
57. Juteršek M, Klemenčič M, Dolinar M. Discrimination between *Synechocystis* members (cyanobacteria) based on heterogeneity of their 16S rRNA and ITS regions. *Acta Chim Slov* 2017;64:804–817.
58. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* 2013;110:1053–1058.
59. Shimura Y, Hirose Y, Misawa N, Wakazuki S, Fujisawa T et al. Complete genome sequence of a coastal cyanobacterium, *Synechococcus* sp. strain NIES-970. *Genome Announc* 2017;5:e00139–17.
60. Yoshino T, Honda T, Tanaka M, Tanaka T. Draft genome sequence of marine cyanobacterium *Synechococcus* sp. strain NKBG15041c. *Genome Announc* 2013;1:e00954–13.
61. Swingley WD, Chen M, Cheung PC, Conrad AL, Dejesa LC et al. Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proc Natl Acad Sci USA* 2008;105:2005–2010.
62. Banroques J, Doère M, Dreyfus M, Linder P, Tanner NK. Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol* 2010;396:949–966.
63. Nelson WC, Maezato Y, Wu Y-W, Romine MF, Lindemann SR. Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. *Appl Environ Microbiol* 2015;82:255–267.
64. Paul R. Genome sequencing of *Leptolyngbya* Heron Island, 2a crystal structure of phycoerythrin and spectroscopic investigation of chromatic acclimation. PhD thesis, Arizona State University, Tempe, AZ, USA; 2014.
65. Georg J, Rosana ARR, Chamot D, Migur A, Hess WR et al. Inactivation of the RNA helicase CrhR impacts a specific subset of the transcriptome in the cyanobacterium *Synechocystis* sp. PCC 6803. *RNA Biol* 2019;16:1205–1214.
66. Rosana AR, Chamot D, Owttrim GW. Autoregulation of RNA helicase expression in response to temperature stress in *Synechocystis* sp. PCC 6803. *PLoS One* 2012;7:e48683.
67. Tarassova OS, Chamot D, Owttrim GW. Conditional, temperature-induced proteolytic regulation of cyanobacterial RNA helicase expression. *J Bacteriol* 2014;196:1560–1568.
68. Nishiyama Y, Los DA, Murata N. PsbU, a protein associated with photosystem II, is required for the acquisition of cellular thermotolerance in *Synechococcus* species PCC 7002. *Plant Physiol* 1999;120:301–308.
69. Gerdes SY, Kurnasov OV, Shatalin K, Polanuyer B, Sloutsky R et al. Comparative genomics of NAD biosynthesis in cyanobacteria. *J Bacteriol* 2006;188:3012–3023.
70. Li W, Schulman S, Dutton RJ, Boyd D, Beckwith J et al. Structure of a bacterial homologue of vitamin K epoxide reductase. *Nature* 2010;463:507–512.
71. Wang S, Hu Y, Overgaard MT, Karginov FV, Uhlenbeck OC et al. The domain of the *Bacillus subtilis* DEAD-box helicase YxiN that is responsible for specific binding of 23S rRNA has an RNA recognition motif fold. *RNA* 2006;12:959–967.
72. Sharpe Elles LM, Sykes MT, Williamson JR, Uhlenbeck OC. A dominant negative mutant of the *E. coli* RNA helicase DbpA blocks assembly of the 50S ribosomal subunit. *Nucleic Acids Res* 2009;37:6503–6514.
73. Jones PG, Mitta M, Kim Y, Jiang W, Inouye M. Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in *Escherichia coli*. *Proc Natl Acad Sci USA* 1996;93:76–80.
74. Butland G, Krogan NJ, Xu J, Yang W-H, Aoki H et al. Investigating the *in vivo* activity of the DeaD protein using protein-protein interactions and the translational activity of structured chloramphenicol acetyltransferase mRNAs. *J Cell Biochem* 2007;100:642–652.
75. Prud'homme-Généreux A, Beran RK, Iost I, Ramey CS, Mackie GA et al. Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol* 2004;54:1409–1421.
76. Vakulskas CA, Pannuri A, Cortés-Selva D, Zere TR, Ahmer BM et al. Global effects of the DEAD-box RNA helicase DeaD (CsdA) on gene expression over a broad range of temperatures. *Mol Microbiol* 2014;92:945–958.
77. Jain C. The *E. coli* RhE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA* 2008;14:381–389.
78. Turner S, Pryer KM, Miao VPW, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 1999;46:327–338.