MASTERARBEIT | MASTER'S THESIS

Titel | Title

Tracing Biliverdin Metabolism in Bacteria and Archaea

verfasst von | submitted by

Marta Medic BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 220
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Joint-Masterstudium Evolutionary Genomics and Systems Biology
Betreut von | Supervisor:	Assoz. Prof. Maria Filipa Baltazar de Lima de Sousa PhD

Acknowledgements

Foremost, I would like to thank my two best mentors in the world, M. Filipa Baltazar de Lima de Sousa and Val Karavaeva, for their supervision, tutoring, and support during my thesis. I am grateful to both of you. Dear Filipa, thank you for your support and for giving me the fantastic opportunity to do my research and make my dream come true: to work on a project in evolutionary genomics. Furthermore, thank you, Val, for your guidance and for always caring about how the project is going. This gave me a remarkably comfortable, confidential feeling about working in the awesome GEEG group.

I am deeply grateful to my parents for their love and support, as well as for giving me the opportunity to study abroad. Especially, thanks to my partner for his support throughout my studies and during my master's thesis project.

Abstract

Bilins are linear tetrapyrroles that are essential for cellular antioxidant and light-harvesting functions. Biliverdin, a green tetrapyrrole, is converted to bilirubin by biliverdin reductase in most Bacteria and animals for detoxification. In contrast, in photosynthetic organisms -*Cyanobacteria*, plants, and red algae—biliverdin is reduced by ferredoxin-dependent bilin reductases, a family of enzymes that facilitate the phycobilisome macromolecule, the phycobiliproteins. This includes enzymes, such as phycocyanobilin ferredoxin oxidoreductase and phycoerythrobilin oxidoreductase, in *Cyanobacteria* and red algae. This Master's thesis examines the evolution and functional divergence of biliverdin metabolism, analysing the distribution and evolution of biliverdin reductases, pigment synthesis enzymes, and the phycocyanobilin and phycoerythrobilin ferredoxin oxidoreductases across prokaryotic lineages. This analysis revealed that biliverdin reductase and the pigment-synthesis enzymes of the ferredoxin-dependent bilin reductase family evolved within *Cyanobacteria*. An independent event led to the appearance of F_{420}-BVRs in *Actinobacteria*, where they are involved in protective mechanisms. Moreover, several horizontal gene transfer events involving these proteins were identified, potentially expanding the taxonomic distribution of biliverdin metabolism in prokaryotes.

Zusammenfassung

Biline sind lineare Tetrapyrrole, die für zelluläre antioxidative Funktionen sowie die Lichtabsorption unerlässlich sind. Biliverdin, ein grünes Tetrapyrrol, wird in den meisten Bakterien und Tieren durch Biliverdinreduktase zu Bilirubin umgewandelt, um es zu entgiften. Im Gegensatz dazu wird Biliverdin in photosynthetischen Organismen – *Cyanobakterien*, Pflanzen und Rotalgen – durch Ferredoxin-abhängige Bilinreduktasen reduziert. Diese Enzymfamilie katalysiert die Bildung der Phycobilisom-Makromoleküle, der Phycobiliproteine. Zu dieser Familie gehören Enzyme wie die Phycocyanobilin Ferredoxin Oxidoreduktase und die Phycoerythrobilin Oxidoreduktase in *Cyanobakterien* und Rotalgen. Diese Masterarbeit untersucht die Evolution und die funktionelle Divergenz des Biliverdinstoffwechsels. Analysiert werden die Verteilung und Evolution von Biliverdinreduktasen, Pigmentsyntheseenzymen sowie der Phycocyanobilin- und Phycoerythrobilin-Ferredoxin-Oxidoreduktasen in verschiedenen prokaryotischen Abstammungen. Die Analyse ergab, dass sich Biliverdinreduktase und die Pigmentsyntheseenzyme der Ferredoxin-abhängigen Biliverdinreduktasefamilie innerhalb der *Cyanobakterien* entwickelten. Unabhängig davon sind F_{420}-BVRs in *Actinobakterien entstanden*, wo sie an Schutzmechanismen beteiligt sind. Darüber hinaus wurden mehrere horizontale Gentransferereignisse identifiziert, die diese Proteine betreffen und möglicherweise die taxonomische Verbreitung des Biliverdinstoffwechsels in Prokaryoten erweitern.

Abbreviation table

Bacteriochlorophylls	BChls
Biliverdin	BV
Biliverdin reductase	BVR
Bilirubin	BR
Chlorophylls	Chls
Chlorophyll-binding proteins	CBP
15,16-dihydrobiliverdin	DHBV
Ferredoxin-dependent bilin reductases	FDBR
flavin/deazaflavin oxidoreductase	FDOR
Flavin mononucleotide	FMN
Glucose–fructose oxidoreductase	GFO
Heme oxygenase	HO
Hidden Markov Model	HMM
Horizontal gene transfer	HGT
Inositol 2-dehydrogenase	IDH
Kyoto Encyclopaedia of Genes and Genomes	KEGG
KEGG Orthology	KO
Mycobacterium heme oxygenase[1]	MhuD
Markov Cluster Algorithm	MCL
Minimal ancestor deviation	MAD
Maximum likelihood	ML
Multiple sequence alignment	MSA
Dehydorganase, in rhizopine catabolism	MOCA
β-nicotinamide adenine dinucleotide phosphate	NADPH
Nicotinamide adenine dinucleotide	$NADP^+$
Phthalate dioxygenase system	PDS
Phthalate oxygenase reductase	PDR
Phthalate dioxygenase	PDO
Phycocyanobilin	PCB
Phycocyanobilin: ferredoxin oxidoreductase	PcyA
Phycoerythrobilin	PEB
15,16-dihydrobiliverdin (DHBV): ferredoxin oxidoreductase	PebA
Phycoerythrobilin: ferredoxin oxidoreductase	PebB
Phycoerythrobilin synthase	PebS
Phytochromobilin	PΦB
Phytochromobilin: ferredoxin oxidoreductase	HY2
Phycobilisomes	PBS
Phycobiliproteins	PBP
phycoviolobilin	PVB
phycourobilin	PUB
Photosystem I	PSI
Photosystem II	PSII
pyridoxamine-5-phosphate	PMP
pyridoxine-5-phosphate	PNP
pyridoxal-5-phosphate	PLP
Reaction Centre I	RCI
Reaction Centre II	RCII

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Tetrapyrrole

Tetrapyrroles are molecules that serve as various biological cofactors in biological processes [2]. Heme, siroheme, (bacterio)chlorophyll and cobalamin are examples of tetrapyroles. They play a role in photosynthesis, cellular respiration, energy conversion, signalling, and oxygen transport[3]. This is further evident in plants, where chlorophylls (Chls) and siroheme are involved in processes occurring within plastids, heme is present throughout all cellular compartments, and phytochromobilin (PΦB) accumulates in the cytoplasm as part of the red-light-sensing phytochrome family[4]. These molecules function together to regulate oxygenic photosynthesis, which in turn drives the global oxygen cycle.

Anaerobic photosynthetic bacteria produce several forms of bacteriochlorophylls (BChls), all of which contain a central magnesium ion within the chlorin ring, similar to Chls[4]. These tetrapyrroles contribute to green pigmentation and play a critical role in absorbing light and enabling energy transfer during photosynthesis[3,5]. Furthermore, in higher plants, Chls a and Chls b are the primary forms that perform this function, with each type contributing differently to the overall efficiency of light absorption[6].

Siroheme, one of the simplest modified tetrapyrroles, is the iron-containing cofactor used by sulfite and nitrite reductases that catalyse the six-electron reductions in sulfur and nitrogen metabolism[7]. Another important tetrapyrrole is vitamin B_{12} (cobalamin), which contains a corrin ring coordinated to a central Co ion, rather than a Mg ion[3]. The two ligands of the cobalt centre in B_{12} are located on the upper and lower sides of the cobalt ion. When a ligand on either side is replaced, this ligand switch alters how cobalamin binds to proteins or riboswitches in biological reactions[8].

Heme, also referred to as iron-containing porphyrin, is found in the tree domain of life[9]. It is a cofactor of proteins involved in many different metabolic pathways. Such as electron transport chains, electron-transfer reactions or signalling proteins. The four nitrogen atoms ensure the iron in the middle is held in balance within the heme structure. Each nitrogen atom in the interior orientation is directed towards the centre of the larger ring that they collectively form. The pyrrole units are held together by alternating double and single bonds with equal electron load, which adds to the stability of the entire molecule [10].

Heme can also serve as a starting point for the synthesis of other tetrapyrroles. Heme oxygenase (HO) breaks down heme, an iron-containing porphyrin, by cleaving the α-meso carbon bridge, resulting in the production of carbon monoxide, iron, and bilins, namely biliverdin (BV)[10]. With the help of electrons from NADPH-cytochrome P450 reductase, HO acts on ferric heme and comes in three isoforms: HO-1 (inducible), HO-2 (constitutive), and HO-3 (a pseudogene[11]). During BV biosynthesis, a series of

intermediates, such as verdoheme and α-meso-hydroxyheme, are formed. Ferritin, a common intracellular protein whose expression is upregulated by HO-1, limits the oxidative potential of the released iron and has cytoprotective effects that are on par with or even greater than BV's own antioxidant activity[12].

In addition, iron is crucial for bacterial survival, and many pathogens have their own versions of HO[13]and can obtain iron by degrading heme in various ways. Some bacteria, such as *Neisseria meningitidis*[14], *Corynebacterium diphtheriae*[15], and *Pseudomonas aeruginosa*[16], use an HO with a structure similar to the well-studied human enzyme. A different strategy is found in *Staphylococcus aureus*, where the IsdG enzyme cleaves heme, forming a new chromophore, staphylobilin, and releasing the β- or δ-meso carbon as formaldehyde[17]. In contrast, in *Mycobacterium tuberculosis*, which contains *Mycobacterium* heme oxygenase (MhuD) enzyme, despite its sequence similarity to IsdG, cleaves heme and produces the mycobilin pigment, which retains an aldehyde at the meso carbon, unlike the formaldehyde found in staphylobilin[18].

1.2. Structure and Function of Linear Tetrapyrroles – Billins

The primary focus of the project is on linear tetrapyrroles, bilins. These molecules are found in all domains of life[3] , and bilins perform a variety of functions in nature, including growth control and photoreception [10]. BV, a green bilin found in mammals and some bacterial cells, is primarily produced from heme by two different heme oxygenases[3]: NADP-dependent biliverdin reductase (BVR) in mammals, and ferredoxin-dependent bilin reductases (FDBRs) in algae, *Cyanobacteria*, and plants [13]. In mammals, the BVR converts the γ-methane bridges to bilirubin (BR), an antioxidant[12]. During this catalytic process, β-nicotinamide adenine dinucleotide phosphate (NADPH) and nicotinamide adenine dinucleotide ($NADP^+$) provide the reducing equivalents required for the reduction of biliverdin. The FDBRs found in *Cyanobacteria* are involved in the synthesis of several light-absorbing pigments (see below).

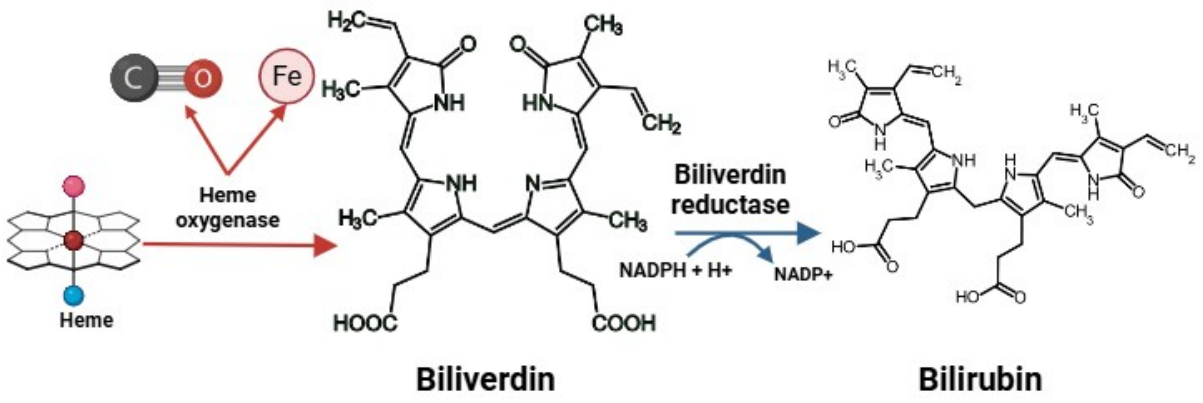

Figure 1. Bilirubin-producing heme degradation pathway: Starting from heme, the heme oxygenase HO-1 releases carbon monoxide and iron to produce biliverdin. Then, biliverdin is converted to the biliverdin reductase. In this reduction, NADPH serves as the electron donor, and a proton (H+) is released, forming NADP+. The image was created in BioRender (Created in https://BioRender.com)[19].

1.2.1. Biliverdin reductase

The mammalian NADP-dependent BVR is a monomeric protein with two structural domains: a C-terminal domain with a six-stranded beta-sheet flanked on one side by alpha-helices, and an N-terminal dinucleotide-binding domain (Rossmann-fold) [20]. Both domains contribute to the development of the active site cleft at their interface. Moreover, the inhibitor and the substrate bind to this single substrate-binding site[20]. The two isoforms of BVR, BVR-A and BVR-B, exhibit distinct developmental expression patterns in humans and substrate specificities[12]. BVR-A efficiently reduces BR IXα, but BVR-B does not affect this isomer. As a result, while BVR-B is the predominant isoform during foetal development, BVRA becomes the primary isoform in adulthood[21]. Interestingly, some *Cyanobacteria* contain distant homologs of the mammalian BVR-A enzyme. Specifically, the mammalian BVR-A and the cyanobacterial BVR-A enzymes share only 20% amino acid sequence identity[22].

1.2.2. Biosynthesis of Phycobiliprotein Pigments by FDBRs in Photosynthetic Organisms

The bilins produced by FDBRs are essential pigments in oxygenic photosynthetic organisms, including *Cyanobacteria*, red algae, and plants[23]. The FDBRs catalyse the formation of phycobilisomes (PBS), macromolecules composed of pigments, phycobiliproteins (PBPs), and colourless proteins. PBPs capture light for the photosynthetic apparatus, while non-pigmented proteins stabilise the complex, attach it to the thylakoid membrane, and position pigments for efficient energy transfer [24]. PBPs are composed of three proteins: at the core is allophycocyanin, and at the rod, phycocyanin and phycoerythrin [25]. These last two proteins are associated with chromophores, specifically phycoerythrobilin (PEB) and phycocyanobilin (PCB). These bilin types, together with phycoviolobilin (PVB) and phycourobilin (PUB), are found only in *Cyanobacteria*, while PΦB is present in both *Cyanobacteria* and plants [26, 27].

FDBRs are enzymes that convert BV into various phycobilins used in PBPs. The phycocyanobilin: ferredoxin oxidoreductase (PcyA) acts as a monomer[28] to catalyse a two-step, four-electron reduction of BV to produce PCB. Spectroscopic analyses show that instead of the stretched linear shape commonly observed in phytochromes and PBPs, BV takes on a characteristic porphyrin-like, cyclic shape when bound to PcyA [29]. In *Cyanobacteria*, FDBRs are used to produce PEB from BV through a two-enzyme pathway: first, 15,16-dihydrobiliverdin (DHBV): ferredoxin oxidoreductase (PebA) enzyme reduces BV to DHBV, in a two-electron step [30], and later, phycoerythrobilin: ferredoxin oxidoreductase (PebB) transforms DHBV into the final PEB pigment in the second two-electron step[28]. Both of these reactions

use ferredoxin as the electron donor [30]. Some marine cyanophages encode a single enzyme, phycoerythrobilin synthase (PebS), which catalyses the complete conversion of BV to PEB, with DHBV serving as a temporary intermediate [31], replacing the host's PebA/PebB pair.

In plants and some algae, the phytochromobilin : ferredoxin oxidoreductase (HY2) enzyme catalyses the conversion of BV to PΦB, being involved in phytochromobilin synthesis. This chromophore, found in plant phytochromes, is covalently bound as a thioether-linked prosthetic group in the homodimer protein, enabling phytochromes to function as red/far-red light photoreceptors. The enzyme HY2 is the sole PΦB synthase in *Arabidopsis* [32].

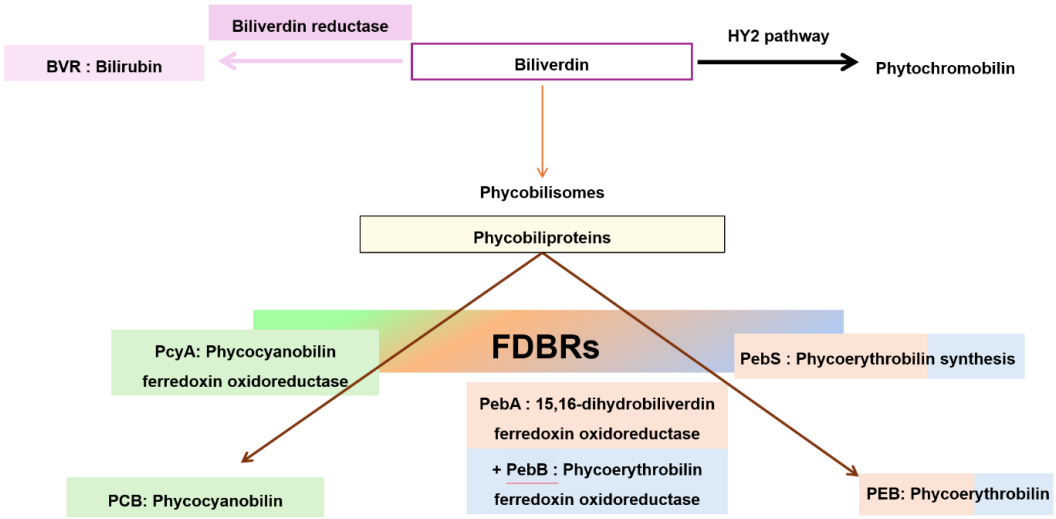

Figure 2.

Biosynthesis pathways of linear tetrapyrrole chromophores derived from biliverdin Ixα, a breakdown product of heme. Different enzymes can reduce biliverdin to produce various light-absorbing pigments: PΦB in plants (via HY2),and phycobiliproteins: phycocyanobilin (via PcyA), and phycoerythrobilin (via PebS or PebA/B) in Cyanobacteria and algae. The colours of enzymes and pigments are consistent with the code shown in Table 1: BVR is marked light pink, the PcyA pathway is marked light green, and PebA/PebB and PebS are coloured according to their reactions.

1.3. Evolutionary Adaptation of Bilin Biosynthesis and Pigment Utilisation

Both biological and geological factors are thought to have influenced the evolution of early photosystem [33] and its associated apparatus. Besides structural modifications within these complexes that lead to the diversification into the current Reaction Centre I(RCI), Reaction Centre II (RCII), Photosystem I (PSI) and Photosystem II (PSII) [34], environmental factors such as oxygenic concentrations and the spectral range of available light also played a role in shaping the evolution of bilin diversity and biosynthesis[23] *Cyanobacteria* are the precursors of chloroplasts [35]. They have established a symbiotic relationship with algae, initially as external partners, and later evolved into internal symbioses [36], giving rise to the primordial chloroplast. Higher plants contain phytochromes that include the chromophore PΦB, while some algae, such as the green algae *Mesotaenium caldariorum*, have the PCB pigment [37]. In contrast to the usual red/far-red absorption of plant phytochromes, the use of PCB produces blue-shifted absorption

spectra, which is consistent with earlier observations of unusual spectral characteristics in green algal phytochromes[37].

In *Cyanobacteria*, light harvesting is most often mediated by PBSs. However, several lineages (*e.g.*, prochlorophytes) have lost PBS and instead contain a membrane-bound Chl *a*/*b* antenna protein, IsiA, which is expressed in response to iron deficiency [38]. The *isiA* gene is structurally homologous to CP43, a chlorophyll *a*-binding protein of photosystem II [39], illustrating how the photosystem components have been used as an alternative antenna complex.

A similar adaptation is found in *Acaryochloris marina*, a cyanobacterium from low-light marine habitats, which lacks PBS but instead has Chl-binding proteins (CBP) [40], related to the prochlorophytes' Chl *a/b*-binding proteins [41]. To compensate for the PBS's absence, it forms its remaining PBPs into rod-shaped structures optimised to capture red-shifted light [42]. Thus, *Acaryochloris marina* distinctly uses bilin pigments, reflecting modifications of the bilin biosynthetic pathway that have evolved to meet a distinct ecological niche [43].

Beyond such species-specific adaptations, molecular phylogenetic analyses of FDBRs proteins reveal lineage-specific diversification[23]. For instance, while PebA and PebB enzymes are limited to lineages that synthesise PEB, PcyA is almost universally present in *Cyanobacteria* and even essential in some species, such as *Synechococcus sp. PCC 7002 and PCC 6803* [26].

Comparative studies of FDBR distribution in non-photosynthetic organisms showed their presence in organisms such as *Mycobacterium tuberculosis* Rv2074, where the structural specialisation of the FDBR active site ensures high affinity for BV IXα and facilitates BR production as an antioxidant during oxidative stress[44]. The adaptation of Rv2074 BVR further demonstrates how different FDBRs have evolved across lineages to regulate specific steps in the bilin biosynthesis pathway.

Moreover, the PEB synthase enzyme, which is actively transcribed during infection and aids in the host tetrapyrrole synthesis, is encoded by the cyanophage *Prochlorococcus P-SSM2*. The phage ensures ongoing chromophore biosynthesis and supports the infected cyanobacterial host's ability to harvest light by providing an alternative pathway for PEB synthesis [31].

Together, these examples show how the structural and functional use of pigments connects ecological adaptation in photosynthetic organisms and physiological adaptation in pathogens.

Aim of the Master's Thesis

The focus of this Master's Thesis is the evolution of the enzymes using biliverdin in Bacteria and Archaea. Therefore, two research questions arise:

1. How diverse is biliverdin metabolism across species, and what are its evolutionary origins?
2. How did PcyA and PebA/PebB enzymes been shaped by the evolution of FDBRs across bacterial lineages?

To investigate these questions and to understand the evolutionary complexity of these enzymes, comparative genomics was used, which can also resolve the evolutionary relationships of BV homologs. The workflow combined the use of online servers, Python programming, and BASH scripts to identify homologous sequences, measure their identity across phyla, and trace the evolutionary origins of biliverdin metabolism in Archaea and Bacteria.

2. Methods

2.1. Software packages

The Life Science Compute Cluster (LiSC) at the University of Vienna was used for all computational steps. This project was developed using custom Python scripts and GitHub resources[45]. Additionally, some coding assistance and text rephrasing were aided via ChatGPT [46].

2.2. Queries and Databases

2.2.1. Query gathering from the BRENDA database

Sequences of enzymes involved in BR and pigment biosynthesis were obtained from the Braunschweig Enzyme Database (BRENDA, version 2024.1) and used as queries. Three enzyme classes were included: biliverdin reductase (EC 1.3.1.24), phycocyanobilin: ferredoxin oxidoreductase (EC 1.3.7.5), and phycoerythrobilin synthase (EC 1.3.7.6). Representative sequences were selected from *Synechococcus sp. (*strain *PCC 6803/Kazusa), Nostoc sp. PCC 7120, Acaryochloris marina (*strain *MBIC 11017), Mycobacterium tuberculosis,* and cyanophage *Prochlorococcus* phage P-SSM2 . For each query, the organism's affiliation, enzyme-catalysed reaction, UniProt accession number, and BRENDA reference entry were summarised in Table 1. Additionally, the colour legend for the upcoming tables is based on the reaction: BVR in two shades of pink, PcyA in light green, and PebA/PebB in combined light orange and blue.

Table 1. Query sequences. Each row displays the UniProt accession number, the organism's affiliation, the enzyme E.C. number, and the enzymatic reaction, including substrates and products involved. Rows are coloured by E.C. number, the BVR in two shades of pink (light shade for Synechococcus sp, and the darker shade for Mycobacterium tuberculosis sequences), the PcyA enzyme in light green, and the PebA/PebB, and PebS in light orange/blue.

Name of the Enzyme	EC Number	Reaction	Organism	UniProt Accession
biliverdin reductase	1.3.1.24	bilirubin + NAD(P)+ = biliverdin + NAD(P)H + H+	*Synechococcus sp . (strain PCC 6803)*	P72782
biliverdin reductase	1.3.1.24	bilirubin + reduced coenzyme F_{420}	*Mycobacterium tuberculosis*	P9WLL7.1
phycocyanobilin:ferredoxin oxidoreductase	1.3.7.5	(3Z)-phycocyanobilin + 4 oxidized ferredoxin = biliverdin IXalpha + 4 reduced ferredoxin	*Acaryochloris marina (strain MBIC 11017)*	B0CAR8
phycocyanobilin:ferredoxin oxidoreductase	1.3.7.5	(3Z)-phycocyanobilin + 4 oxidized ferredoxin = biliverdin IXalpha + 4 reduced ferredoxin	*Synechococcus sp . (strain PCC 6803 / Kazusa)*	55891
phycocyanobilin:ferredoxin oxidoreductase	1.3.7.5	biliverdin IXa + reduced ferredoxin = phycocyanobilin + oxidized ferredoxin	*Nostoc sp. (strain PCC 7120)*	Q93TN0
phycoerythrobilin synthase	1.3.7.6	(3Z)-phycoerythrobilin + 2 oxidized ferredoxin = biliverdin IXalpha + 2 reduced ferredoxin	*Prochlorococcus phage P-SSM2*	Q58MU6

2.3. Similarity searches

The Diamond Blastp program (version 2.1.12) [47] was used for local protein sequence alignment of queries against the in-house database of 3,939 prokaryotic genome assemblies. To ensure high-quality homolog identification, only sequences with an E-value less than or equal to $1e^{-8}$ and a minimum identity of 25% were retained. Sequences that passed the thresholds were extracted from the genome database using the SeqIO module of Biopython (version 1.85).

2.4. All versus all comparison

To ensure that all relationships were significant, independent, an all-versus-all diamond search was also performed on the diamond-hit set. For further analysis, hits that fulfilled the threshold of 25% global and 25% local coverage, and an E-value below 10^-8, were retained. Next, the global pairwise sequence alignments were computed using the Needleman-Wunsch algorithm, implemented in the Needleall tool from the EMBOSS package (version 6.6.0). This algorithm generates end-to-end pairwise sequence alignments[48] and uses scoring systems for matches, mismatches, and gaps to determine the best end-to-end alignment between two sequences. The relationships with identity percentages above or equal to 25% were used for further analysis in this study.

2.5. Taxonomic Distribution

2.5.1. Clustering The Homologs

The Markov Cluster Algorithm, or MCL (version: 22-282) is an effective clustering method used for the detection of clusters within graphs or networks [49]. The main parameter responsible for clustering is the iterative inflation, with the default best value of 1.2. With an increasing inflation parameter, the clusters become more separated; with a decreasing inflation parameter, the clusters become larger. The MCL output data were presented in two columns: the first contained sequencing numbers, and the second contained cluster numbers. For the BVR hits, 11 clusters were identified, of which five were singletons. Among the PcyA and PebA/PebB synthesis proteins two clusters were obtained, with cluster 2 a singleton.

Intra- and intercluster mean, median, maximum, and minimum identities were calculated using the all-vs-all global identity values. This allowed to retrieve information regarding how the proteins from each cluster relate to the ones from the same or a different cluster. Specifically, this percentage measures the difference in amino acid identities between proteins.

A custom Python script was used to generate the presence absence matrix that was subsequently plotted as a heatmap in R Studio (version: 2024.12.1) using the package prettyheatmaps (pheatmap: Pretty Heatmaps. 1.0.12)[50]. In this representation, protein clusters are depicted as columns, and genomes or taxonomic groups as rows. Cells indicate the presence (1) or absence (0) of a protein for each genome-cluster pair, providing a practical overview of protein distribution across the dataset's genomes. The taxonomic distribution was calculated by dividing the percentage of genomes that have a hit in the cluster by the total number of genomes in the phylum and was subsequently plotted as a heatmap. *Candidatus* phyla with hits were retained as separate rows, whereas those without hits were grouped into a single row.

2.6. Functional annotation

2.6.1. KEGG Orthology annotation

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) database systematically connects genes to biological pathways and metabolic systems[51] . An important component is the KEGG Orthology (KO) database [52], which provides functional annotations for protein sequences based on KO numbers. The KO models (version: 2024-12) were created by reconstructing an enhanced metabolic and signalling network using experimental data. A KO-based approach was developed to associate a given gene set with KEGG pathways for its functional annotation[53].

HMMER is a tool used for comparing profile Hidden Marko Models (HMMs) against a sequence database [54]**.** KEGG HMM profiles were compared to our genomic database using HMMER's hmmsearch program. A ranked list of hits was generated for each target sequence after every profile in the HMM file was matched. KO models, with a cut-off score below 100, did not meet the required threshold (Supplementary data 7.3.).

2.6.2. InterProScan functional annotation

InterProScan was used for annotating protein sequences (version: 5.73-104.0), with different databases and methods, such as Pfam, Superfamily, TIGRFAMs, SMART, PROSITE, PRINTS, ProDom, SMART, and PIRSF to identify functional domains, motifs, and signatures in proteins and assigned sequences to certain protein families or domains accordingly[55].

2.6.3. Functional annotation table

The information obtained from functional annotation and clustering was compiled into a single table representing the cluster annotation of protein sequences. It included predicted domains, protein family classifications, and functional sites, each linked to its respective InterPro accession number and summary. KO annotations were integrated to associate individual proteins with specific metabolic and biochemical pathways. In addition, taxonomic information from the genome database (Supplementary data : 7.2.3.), as well as collected protein accessions and genome identifications, were presented in the table.

2.7. Multiple sequence alignments

Multiple sequence alignments (MSA) were performed in Clustal Omega (version 1.2.4), a tool for aligning protein, DNA, or RNA sequences [56]. Clustal Omega applies an HMM-based alignment approach, generating alignments through iterative cycles that estimate sequence distances and reconstruct the guiding tree [57]. To run Clustal Omega, 100 guide-tree iterations and 100 HMM iterations were used, with the output in guide-tree order. Furthermore, alignments were manually inspected and checked for

conserved motifs in Jalview, a bioinformatic tool used for graphical interpretation of MSAs [58]. In addition, the multiple sequence alignments were transformed into a matrix representation, which was then plotted as a heatmap using custom Perl and R scripts. Next, MSAs were trimmed with the tool TrimAl (version:1.5.0)[59],[60] using default parameters and re-plotted as a heatmap for a more precise representation of underlying conserved domains and highlighted sequence features throughout the dataset.

2.8. Phylogenetic reconstructions

Maximum likelihood (ML) phylogenies were constructed for the proteins of interest. The ML was inferred using IQ-Tree (version 3.0.1), a phylogenetic reconstruction software tool. IQ-Tree was used to analyse the aligned sequences and to determine the substitution model based on the Bayesian Information Criterion[61] ,[62]. The resulting ML phylogeny is unrooted, meaning that the direction of the evolution is not specified. Therefore, the phylogeny must be rooted to conclude ancestral relationships among taxa. The two well-known methods, minimal ancestor deviation (MAD) and midpoint rooting, were used to root the phylogenetic tree [61]. The MAD places each possible root along the branches of the unrooted tree by decreasing the deviation of the ancestor-descendant relation. The midpoint, on the other hand, positions the root in the centre of the maximal path between two taxa if all lineages exhibit an equal rate of evolutionary variation. The IQ-TREE output includes a collection of files, three of which are commonly used to interpret the phylogenetic tree: *treefile , iqtree,* and *cntree*. The final inferred tree is saved in Newick format with the extension ".treefile". The main summary file ".iqtree" contains model selection results and statistical details. The output also includes a consensus tree ".contree" generated from bootstrap replicates. Ultrafast bootstrap values above 95% provide strong support for the inferred evolutionary relationships [63]. The rooted phylogenies were analysed using FigTree (version: 2018-11-25 - *v1.4.4)*, a graphical tool used for the analysis, editing, and visualization of phylogenetic reconstructions [64]. This tool offers options for displaying phylogenies, including branch lengths, labels, and bootstrap values, to facilitate easier interpretation of the results.

The methods described above are the analytical and computational pipeline of the study. The analytical steps are presented in Figure 3.

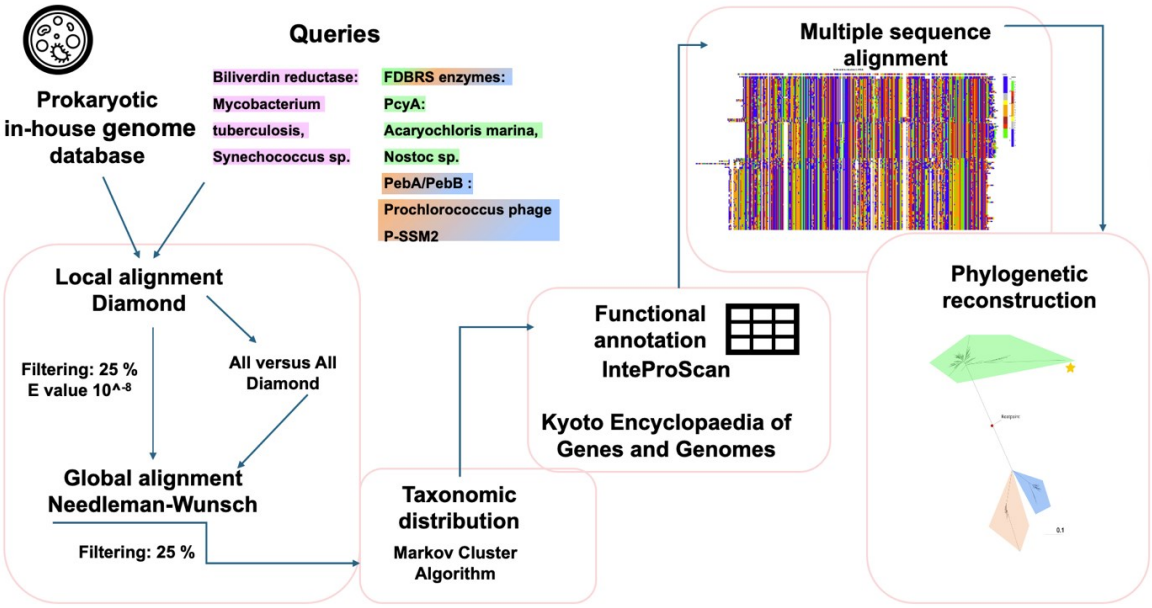

Figure 3. Overview of methods performed in the study.

3. Results

Following the *Methodology* described in Chapter 2, this section presents the results of the data analysis.

3.1. Similarity searches

In total, 2,482 significant homologs were found across 3,939 prokaryotic genomes. Using *Synechococcus sp (strain PCC 6803)* and *Mycobacterium tuberculosis* as queries, 2,299 BVR homologs were identified in the dataset. For pigment biosynthesis pathways, 184 sequences corresponding to the PcyA and PebA/PebB enzymes were identified. PcyA homologs were searched using sequences from *Acaryhochloris marina* (strain *MBIC 11017*), *Synechococcus sp* (strain *PCC 6803 / Kazusa*), and *Nostoc sp.* (strain *PCC 7120*), while *Prochlorococcus phage P-SSM2* sequences were used to conduct PebA/PebB homologs searches.

All-versus-all comparisons were performed using both global (needle) and local (diamond) algorithms. In the all-versus-all diamond comparisons, an E-value of less than or equal to 10^{-8} was used as a cutoff to evaluate alignment accuracy and eliminate potential false positives. Moreover, Diamond results were also filtered using an identity cutoff of 25% to retain only significant homologous sequences. The rationale for combining these approaches is to assess whether the global pairwise sequences are significant. The Needle algorithm does not provide an E-value output. The total number of global pairwise identity relationships that met the cutoffs was 933,797 for BVR homologs and 22,281 for pigment synthesis protein hits.

3.2. Taxonomic distribution

All global pairwise relationships above 25% were used as input for MCL, and a total of 11 clusters were obtained for BVR. In contrast, for the pigment biosynthesis enzymes PcyA and PebA/PebB, only two clusters were obtained. Table 2 presents the total number of protein sequences within the 11 clusters obtained from the BVR homology searches. Cyanobacterial queries were grouped in cluster one, while the BVR sequence from *Mycobacterium* clustered in cluster two.

Table 2. Clusters of BVR homologs.

Cluster number	Total number of sequences
Cluster 1	2101
Cluster 2	151
Cluster 3	18
Cluster 4	17
Cluster 5	4
Cluster 6	2
Cluster 7	2
Cluster 8	1
Cluster 9	1
Cluster 10	1
Cluster 11	1

The MCL clustering of the 184 homologous hits obtained for the PcyA and PebA/PebB enzymes resulted in two clusters, where 183 sequences were grouped into cluster one. In contrast, cluster two contained a single sequence (Table 3).

Table 3. Clusters of PcyA and PebA/PebB homologs.

Cluster number	Total number of sequences
1	183
2	1

The BVR homologous hits were used to generate the taxonomic distribution heatmap (Figure 4). This matrix shows the percentages of genomes within each phylum that contained at least one hit per cluster. Of note, the taxonomic distribution shows that clusters contain sequences from both BVR and its

homologs. By functional annotation, BVR proteins were found only in *Cyanobacteria;* however, within their cluster, several homologous proteins are present. Those homologs are found across 44 different phyla, being absent from the remaining 122. Within the BVR homologous taxonomic distribution, *Cyanobacteria* are highlighted, as the BVR enzyme was found in this phylum, within cluster 1.

Additionally, beyond *Cyanobacteria,* homologous BVR proteins were found among *Actinobacteria*, *Bacteroidetes*, *Firmicute*s, and *Proteobacteria,* with fewer occurrences in other lineages, such as *Chloroflexi*, *Planctomycetes*, and several *Candidatus* phyla. The heatmap is presented in Supplementary data 7.4.5. summarises the number of phyla with and without hits per cluster. Supplementary data 7.4.1. provides Figure 13, which lists the total number of phyla per cluster that had hits.

The total number of hits for the pigment-synthesis proteins PcyA and PebA/PebB was lower. Their taxonomic distribution is limited, as the production of PCB and PEB pigments is mainly restricted to *Cyanobacteria*. Therefore, the taxonomic distribution of the hits was assessed at the genus level within this group. Supplementary data 7.4.2. shows Figure 14 of a table with the total number of genera represented in clusters one and two, while the full genus-level heatmap is given in Supplementary data 7.4.3.

As expected, there were no hits for BVR or pigment synthesis enzymes PcyA and PebA/PebB identified in Archaea.

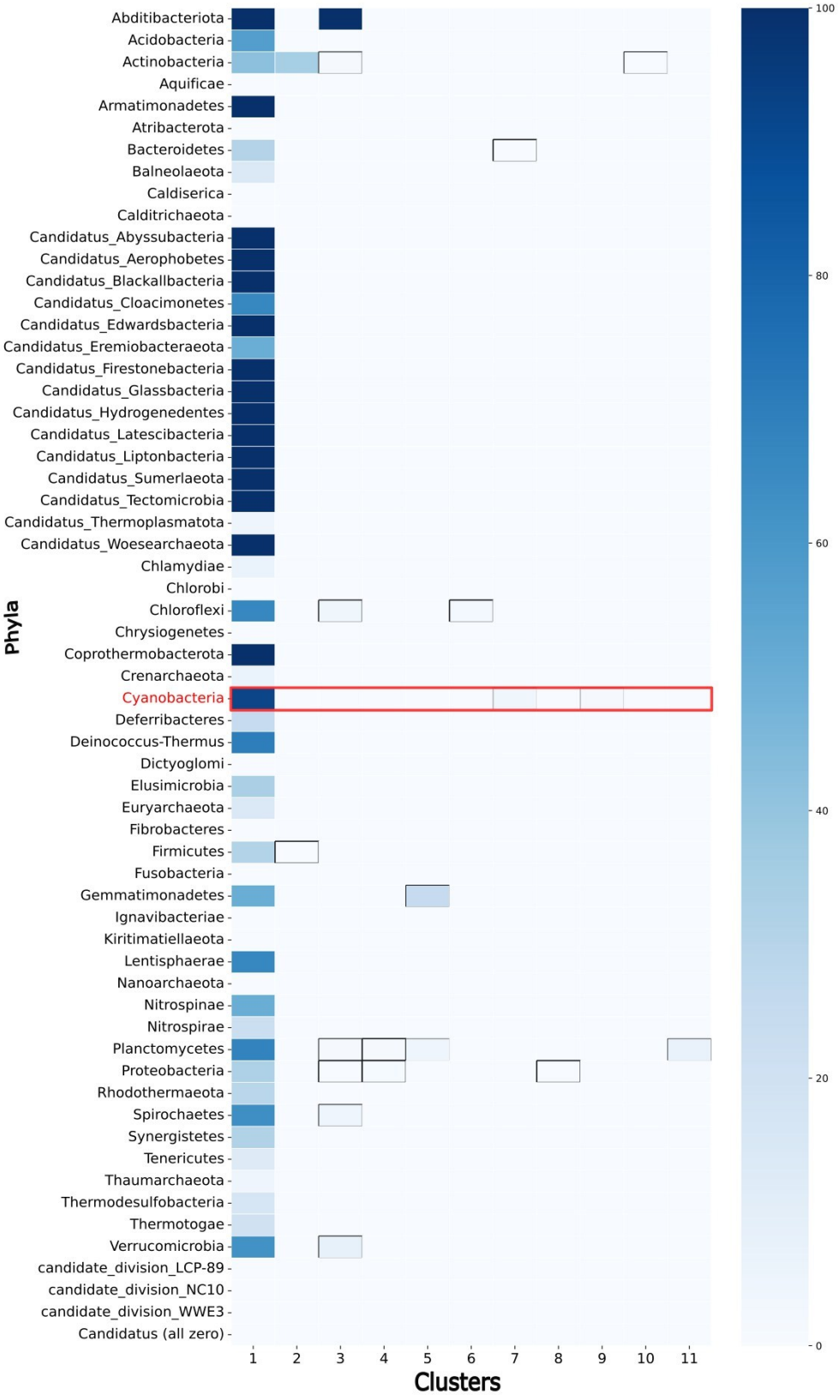

Figure 4. Taxonomic distribution of BVR homologs in bacteria. The x-axis presents different phyla (one per row, in alphabetic order), while the y-axis represents different clusters. Additionally, marked cubes represent the percentage of hits for each genome within each phylum, displayed as a gradient from each genome within phyla, displayed by the gradient marked from white to dark blue.

3.2.1. Inter-cluster relationships

The inter-cluster relationships were analysed based on their pairwise global identity values, as described in the Methods section. The mean intercluster pairwise identity is shown in Figure 5. The complete inter-cluster table (values above 25%) is shown in Supplementary data 7.5.1. Cluster 1 shows inter-cluster relationships with cluster 2, 3, 5 and 6. However, further inter-cluster identity inspection (Supplementary data 7.5.1.) indicated that some of these values derive from single pairwise relationships between clusters. This is not the case for the relationships between clusters 1 and 3, which were considered to be relevant. Since clusters 6 to 11 are singletons, except for cluster 7, their intra-cluster identity is 100% (self-to-self comparison).

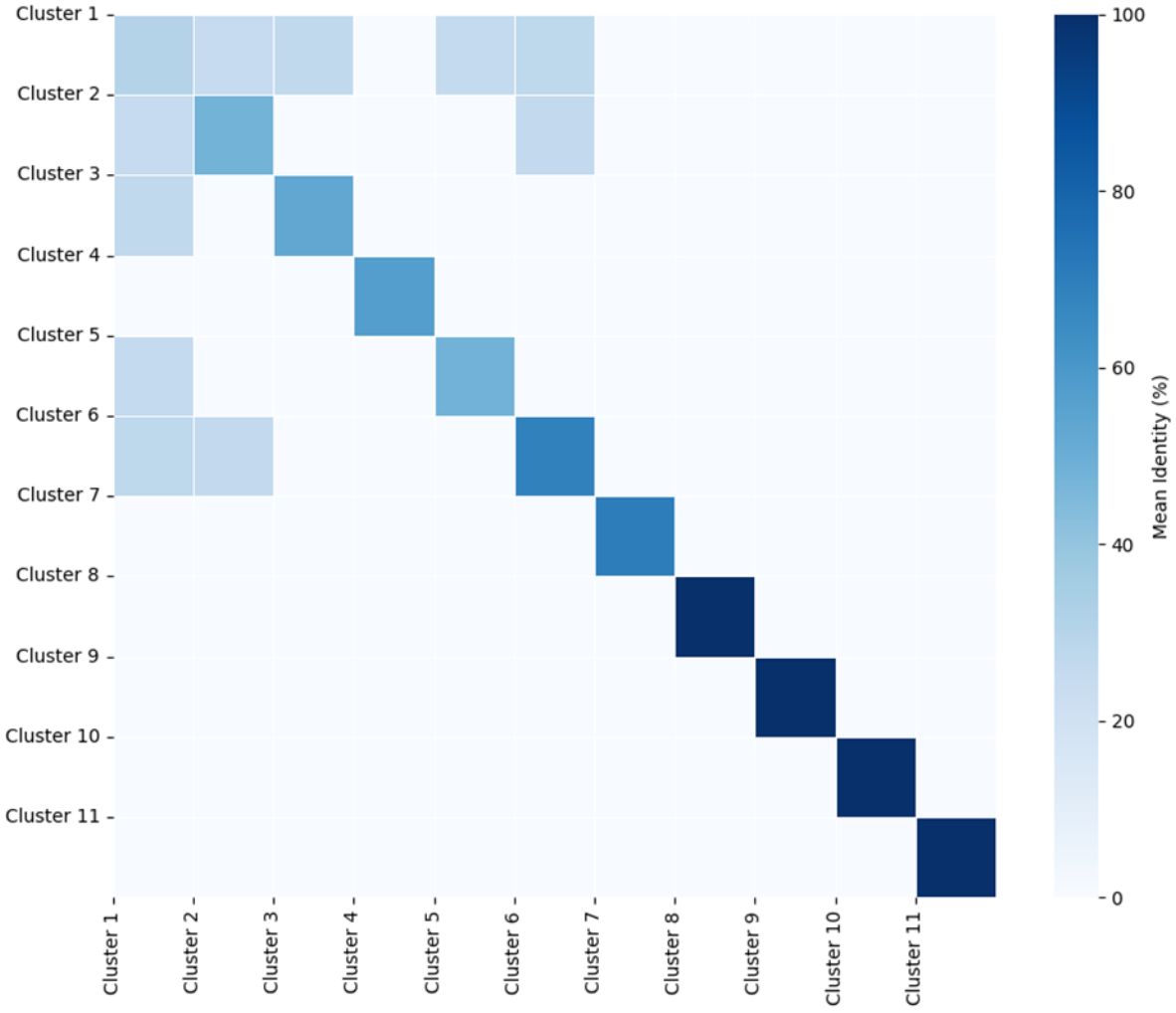

Figure 5. BVR and homologous Intra- and Inter-cluster pairwise mean identity.

Furthermore, the inter-cluster relationships were investigated in more detail based on the global inter-cluster identity values. Figure 6 shows an in-depth examination of the intercluster identities considering all pairs. It is observed that clusters contain sequences of higher identity, indicating stronger relationships. Specifically, clusters 1, 3, and 5 have stronger interconnections among themselves than with the other clusters and were thus joined for further analysis.

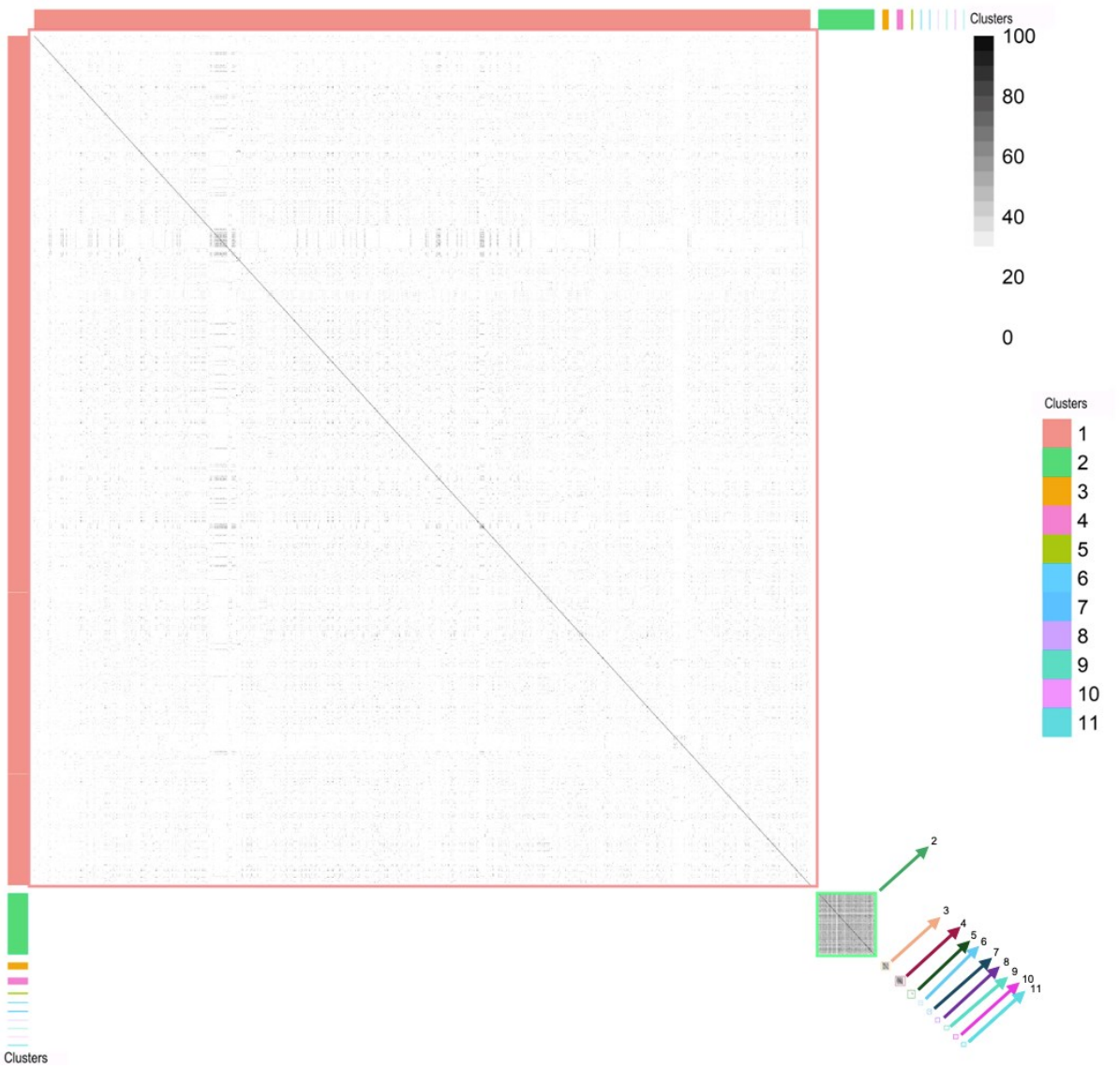

Figure 6. BVR and homologous Intra- and Intercluster global identities. Identity scores below 25% are shown in white. Coloured arrows highlight the small cluster's numbers to make them more visible within the matrix.

3.2.2. Phycocyanobilin- and phycoerythrobilin ferredoxin: oxidoreductase inter- and intra-cluster relationships

Two clusters were obtained from homology searches for PCB- and PEB. Cluster one could be further subdivided into three subgroups, each one corresponding to a pigment synthesis enzyme: PcyA, PebA, and PebB (Figure 8). As expected, sequences within each subgroup have higher pairwise identities. Most members of subgroup one corresponding to PcyA display identities above 70. Although this subgroup shows strong internal similarities, it is less similar to the other subgroups that contain sequences annotated as PebA and PebB. In contrast, subgroups 2 and 3 share higher intergroup identity values (above 40).

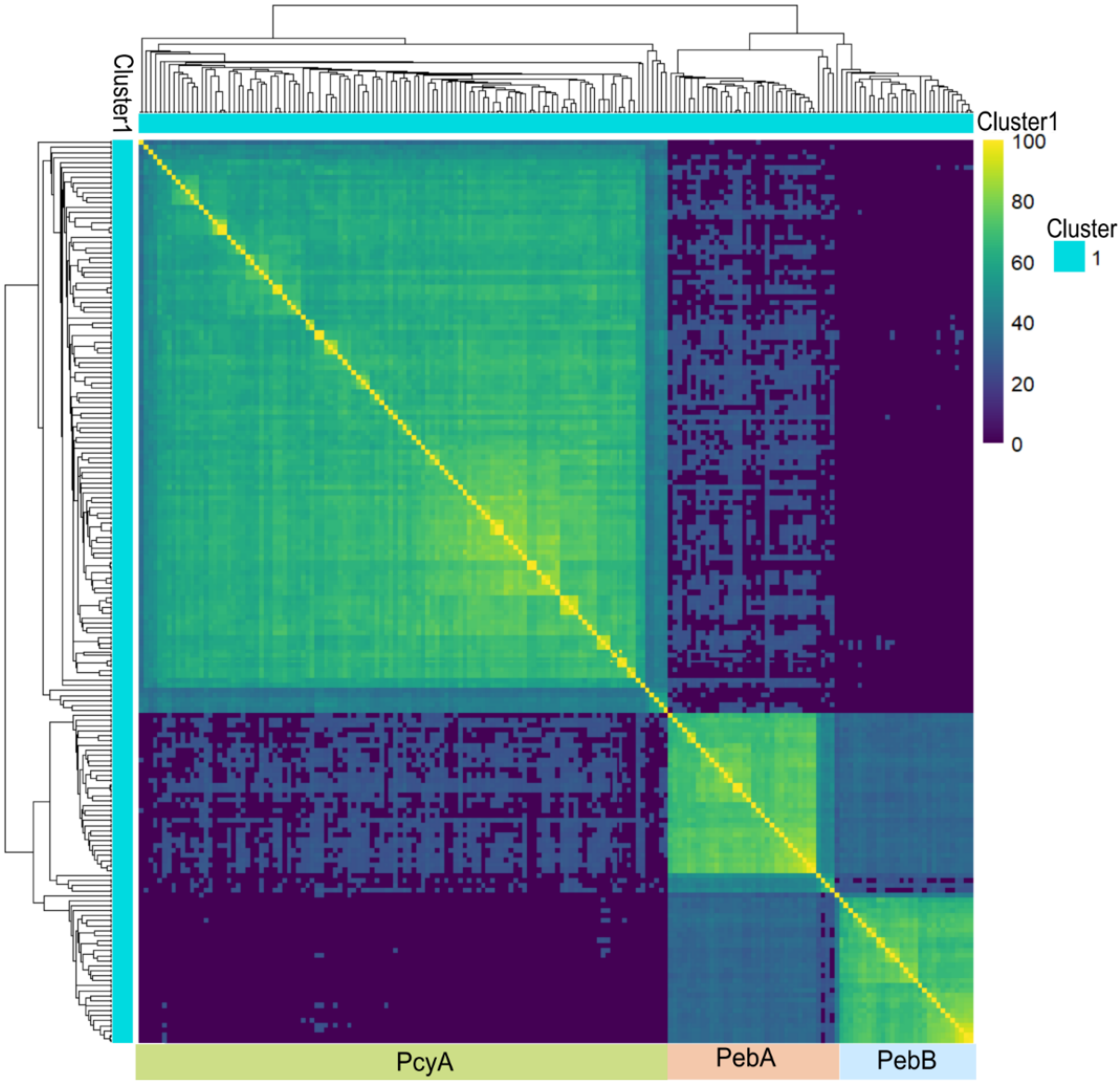

Figure 7. Intracluster identity relationships between homologous sequences of pigment synthesis enzymes, PcyA (highlighted in green) and PebA(marked in light orange), and the PebB enzyme are highlighted in light blue.

3.3. Diversity of PFAM domains

Figure 8 provides an overview of the Pfam domains repeatedly observed among the sequences analysed.

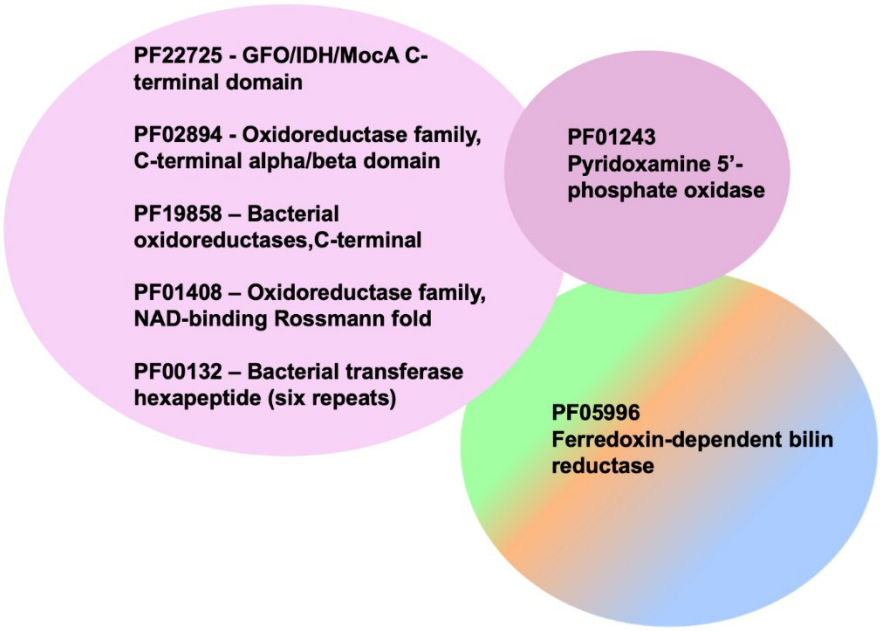

Figure 8. Schematic representation of PFAM domains identified for each protein, following the colour scheme in Table 1. F_{420}-dependent ferredoxin oxidoreductase domains are shown in darker pink, BVR domains in light pink, and enzymes PcyA and PebA/PebB domains highlighted in green/orange/light blue.

Domain analysis of BVR and homologous sequences revealed that five distinct PFAM domains are most frequently retrieved. A closer inspection of their functional annotation revealed that the majority of sequences (83 %) belong to the large clan of C-terminal oxidoreductases, GADPH_aa-bio_dh, which includes numerous enzymes involved in amino acid biosynthesis, as well as GAPDH dehydrogenases involved in glycolysis and gluconeogenesis[65]. These sequences could be further classified according to PFAM annotations: 7 sequences contain the C-terminal domain of bacterial oxidoreductases (PF19858), and 220 have the alpha/beta domain of the GFO/IDH/MOCA enzyme family (PF02894), including 22 BVR homologs. These two protein families are closely related and include enzymes such as glucose–fructose oxidoreductase (GFO), inositol 2-dehydrogenase (IDH), and MOCA, a dehydrogenase involved in rhizopine catabolism, which also contributes to oligomerisation and substrate attachment[66]. In addition, 780 sequences possess the C-terminal domain found in the oxidoreductases (PF22725) from the GFO/IDH family[66] .

In BVR sequences, the NAD-binding Rossmann fold (PF01408) is also found. This domain is part of the NADP_Rossmann clan, which contains a Rossman-fold that adopts an α/β structure for NAD^+ (or FAD) binding[69]. Additionally, the N-terminal domain of the GFO/IDH/MOCA enzymes (PF01408)[70] was identified in 200 enzymes, of which 25 are potential BVRs. This domain corresponds to the catalytic site where the reduction and oxidation of carbohydrates and BV reduction[68] occur. Lastly, one sequence

contained a bacterial transferase hexapeptide (PF00132) domain, belonging to the Hexapeptide clan. This superfamily is characterised by superhelical spirals of three short trans repeats, with each heptapeptide repeat corresponding to a strand [71].

Cluster two, where *Mycobacterium tuberculosis* related sequences are found, includes 151 sequences with the pyridoxamine 5'-phosphate oxidases domain (PF01243), which is part of the large flavin mononucleotide (FMN)-binding clan. This clan comprises several FMN-dependent flavoproteins that catalyse the conversion of pyridoxamine-5-phosphate (PMP) and pyridoxine-5-phosphate (PNP) to pyridoxal-5-phosphate (PLP) [72]. In certain organisms, such as *Mycobacterium tuberculosis* [73]strains, coenzyme F_{420} may replace FMN, explaining why this PFAM was found only among the *actinobacterial* genomes of our study. Furthermore, the 5'-phosphate oxidases (PF01234) domain, found in those sequences, corresponds to the protein's N-terminal segment, which participates in FMN binding during dimerisation.

All 183 homologous sequences of the pigment enzyme synthases, PcyA, PebA, and PebB, belong to the FDBR family , characterised by the FDBR (PF05996) domain. FDBRs represent a family of enzymes – including PcyA, PebA/PebB and HY2- rather than a clan.

3.4. Phylogenetic reconstructions

ML phylogenies were reconstructed for BVR and for homologous sequences of PcyA and PebA/PebB. The MAD rooting algorithm was used to root the phylogenies, as indicated by the arrows in Figures 9, 11, and 12.

3.4.1. The biliverdin reductase phylogeny

The cyanobacterial BVR and its homologous phylogenetic reconstruction are presented in a radial layout in Figure 10. In this large phylogeny, clades were coloured based on the functional annotations of their sequences. Black-marked branches depict no KO assignment.

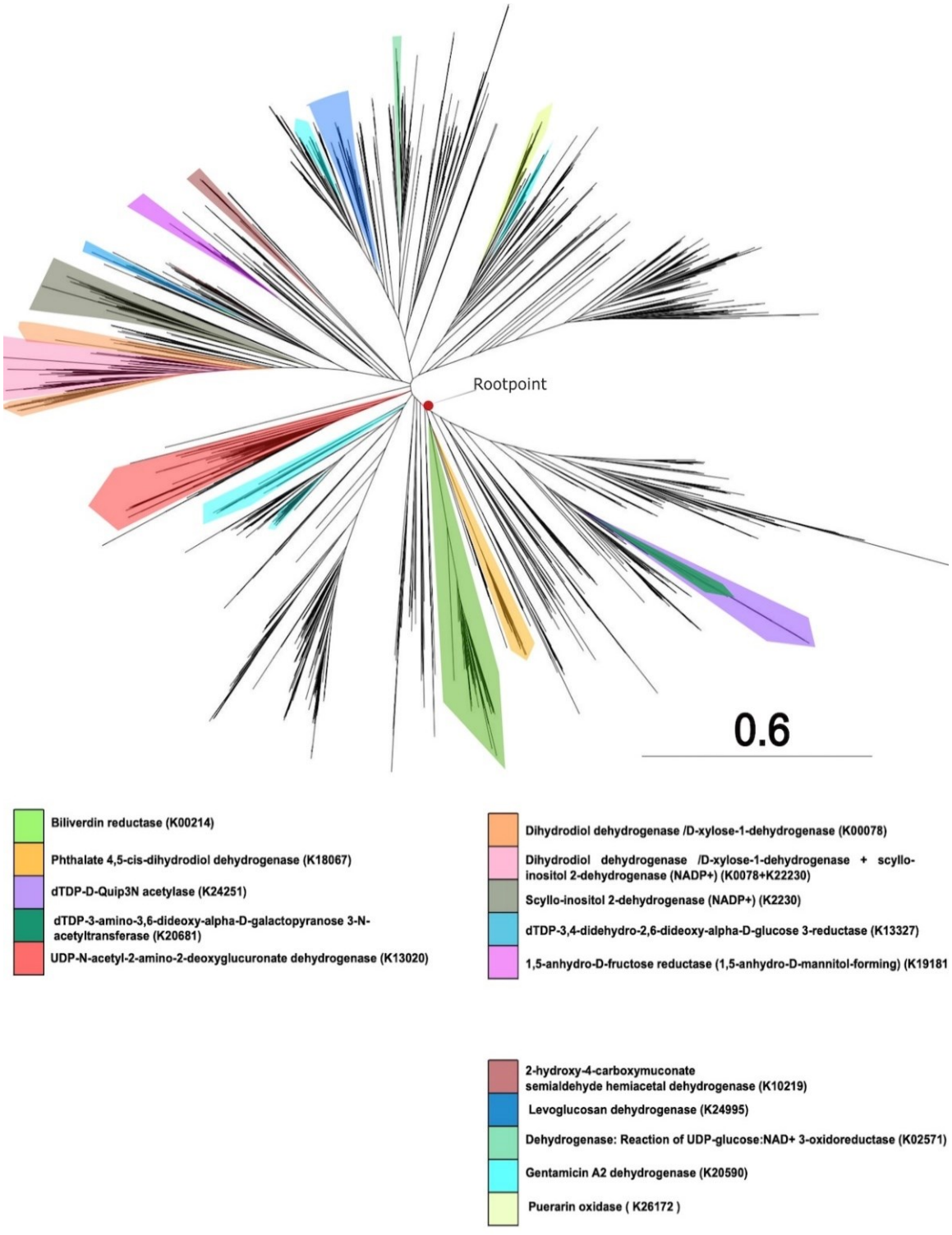

Figure 9. Biliverdin reductase and homologs ML phylogenetic reconstruction, coloured based on functional annotation (KO). Uncoloured branches have no KO assignment.

The potential BVR enzymes form a distinct monophyletic clade, which includes Cyanobacteria sequences and one Acidobacteria sequence. The basal position of the *acidobacterial* sequence could indicate that it belongs to an ancestral lineage containing BVR, as shown in the rectangular phylogenetic representation (Figure 10). However, the presence of a single *Acidobacteria* sequence can also reflect the horizontal gene transfer (HGT) among bacterial taxa, followed by vertical divergence, which would be the most parsimonious scenario. This organism contains a hybrid RCI and is able to perform anoxygenic photosynthesis, further supporting the HGT hypothesis[74]. The remaining BVR sequences form well-supported subclades of *Cyanobacteria*, suggesting these lineages underwent separate diversification, even after the divergence of the basal group. Black-marked *Cyanobacteria* sequences lacked KO annotation but are close homologs of the BVR enzyme and potentially could also be BVRs, given their placement within the BVR clade. Moreover, in Figure 10, the branching point between *Cyanobacteria* and *Proteobacteria* is also shown. The proteobacterial clade contains one *Ca. Tectomicrobia* sequence, indicating HGT between the two taxa. The sequences of these clades are annotated with the same KO model (K18067), which describes a phthalate 4,5-*cis*-dihydrodiol dehydrogenase. However, the branching point between the *Cyanobacteria* and *Proteobacteria* clades has a low bootstrap value, indicating weak support for their divergence. Nevertheless, it might indicate that BVR evolved from existing enzymes, closely related to the 4,5-*cis*-dihydrodiol dehydrogenase ancestor.

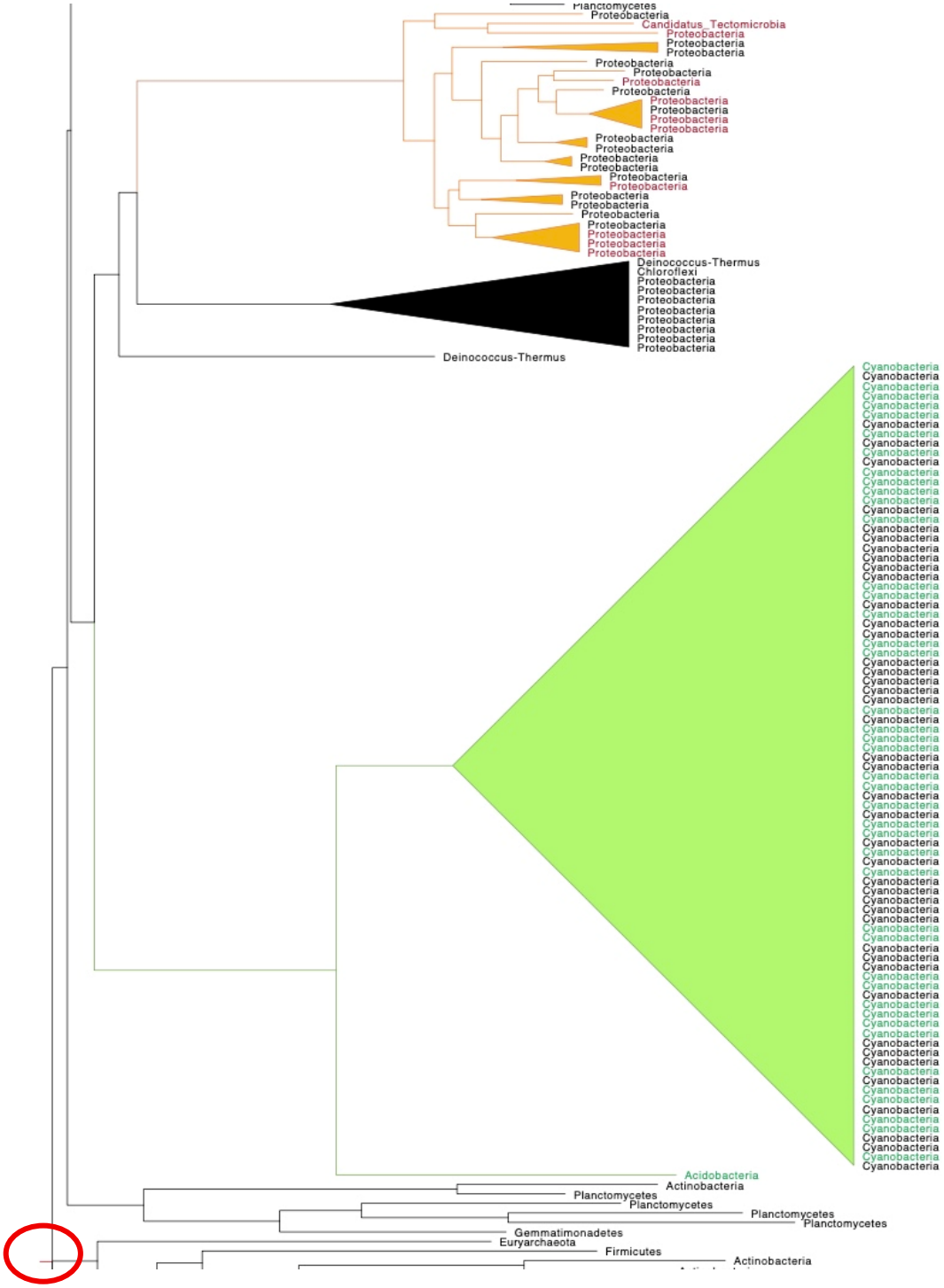

Figure 10. Insert from the BVR ML phylogenetic reconstruction. The rectangular representation highlights the Cyanobacteria clade (K00214 in green), and the Proteobacteria - Ca Tectombirobia clade (K18067 (branches in orange). The red circle presents the root of the phylogeny.

Beyond the *Cyanobacteria* clade, which likely contains *bona fide* BVRs, the rest of the phylogeny includes other homologous BVR sequences. Within the phylogeny, the sequences of gentamicin A2-dehydrogenase enzyme (K02590; Figure 9, navy blue), form four distinct clades, containing *Actinobacteria*, *Bacteroidetes*, and *Firmicutes* sequences. They are non-monophyletic groups dispersed across the phylogeny, suggesting HGT or, in an alternative, the need to fine-tune the model cut-offs, since

some might be mis-annotated. These clades are well-supported, with bootstrap values of 100 for the *Actinobacteria* and *Bacteroidetes* branches and 99 for the Firmicutes branch.

Actinobacteria, *Firmicutes*, and *Proteobacteria* sequences are found in the clade marked as light pink, where a small number of *Chloroflexi* sequences are also present. These sequences are the result of a fusion event that combined scyllo-inositol 2-dehydrogenase ($NADP^+$) (KO: K22230) and dihydrodiol dehydrogenase/D-xylose-1-dehydrogenase ($NADP^+$) (KO: K00078) genes.

The dark grey branches correspond to sequences annotated only as scyllo-inositol 2-dehydrogenase (K22230), primarily from *Proteobacteria*, *Actinobacteria*, and *Firmicutes*. In contrast, the lighter orange branches indicate a clade of *Proteobacteria* sequences annotated as dihydrodiol dehydrogenase/D-xylose-1-dehydrogenase (K00078). Bootstrap values of 100 for both clades indicate strong support for their phylogenetic separation.

An internal node comprising sequences from *Proteobacteria* and *Thermodesulfobacteria* that share the GFO/IDH/MOCA family PFAM domains is represented by the strongly supported branch (bootstrap value of 95, displayed in purple). The two groups have different KO annotations despite having the same domain architecture: the *Proteobacteria* sequences are annotated as dTDP-D-Quip3N acetylases (KO: K24251; purple), while the *Thermodesulfobacteria* sequences are annotated as phosphatase enzymes (KO: K20681; green within purple). This implies that these enzymes have experienced functional diversification across lineages despite sharing a common structural framework. The red clade, consisting of *Proteobacteria* and *Firmicutes*, contains sequences encoding UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase (K13020).

In conclusion, this phylogeny highlights the diversity of functional homologs in the BVR family. Bona-fine BVRs are found only in *Cyanobacteria* and potentially also in the anoxygenic photosynthetic *Chloracidobacterium thermophilum* and might have derived from an ancestor common to the 4,5-cis-dihydrodiol dehydrogenases.

3.4.2. Phylogeny of the F_{420}-dependent biliverdin: ferredoxin oxidoreductase

Figure 11 shows the phylogeny of F_{420}-dependent biliverdin: ferredoxin oxidoreductases. Most of the sequences belong to *Actinobacteria*, while a separate clade, highlighted in pink, contains sequences from the phylum *Firmicutes*. The class *Actinomycetia* is at the root, suggesting the origin of the enzyme family within the phylum *Actinobacteria*. The close association between the *Actinobacteria* and the *Firmicutes* clades suggests an event of HGT, as these taxa are distantly related. Within the phylogeny, three classes of *Actinobacteria*: the *Actinomycetia*, *Thermoleophilia*, and *Acidimicrobiia* are found, while the *Bacilli* are a class of *Firmicutes*. Notably, the sequences belonging to *Bacilli* and *Actinomycetia* form a highly supported group (bootstrap value of 90).

Furthermore, *Thermoleophilia* and *Actinomycetia* classes form distinct clades within the phylogeny. While *Thermophilic* organisms grow in hot spring environments, *Actinomycetia* microorganisms are metabolically more diverse. However, with the bootstrap value of 84, these clades are only moderately supported and may not have a strong evolutionary relationship.

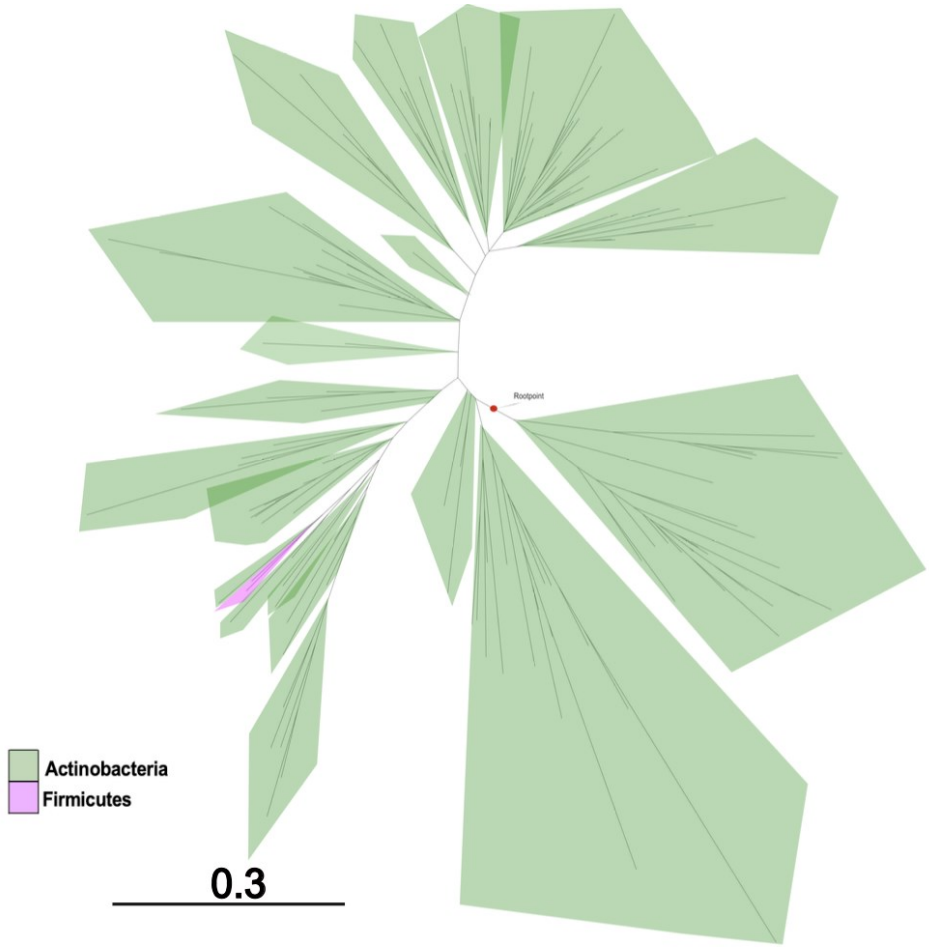

Figure 11. The phylogeny of F_{420}-dependent biliverdin: ferredoxin oxidoreductase with Actinobacteria marked in green and Firmicutes marked in pink.

3.4.3. Phylogeny of pigment synthesis enzymes PcyA and PebA/PebB

The phylogeny of the pigment synthesis enzymes PcyA, PebA, and PebB synthases, shown in Figure 12, is coloured based on the assigned KO models. The PcyA clade (K05371) is highlighted in green, while the sequences annotated as PebA enzyme (K05369) are marked in pink. The blue clade contains PebB enzymes (K03570). The phylogeny reveals that the PcyA and PebA/PebB enzymes are monophyletically organized within the phylogeny, indicating that these enzymes originated from a common ancestral gene. The root separates the PcyA clade from PebA/PebB enzymes, which form a well-supported clade (bootstrap value = 100), indicating their close evolutionary relationship. Based on their position within the FDBR phylogeny, this topology suggests that PebA and PebB share a more recent common ancestor and may have evolved by duplication, after the separation from the ancestor of the PcyA enzymes. The two branches marked by a star belong to non-photosynthetic *Proteobacteria* [76]. These long branches might indicate HGT potential artefacts from metagenomic assembly or functional divergence. Thus, these sequences are good candidates for experimental characterisation.

It's interesting to note that the PebS sequences from the *Prochlorococcus* phage P-SSM2 cluster are located in the PebA clade, suggesting that this cyanophage uses the same pigment synthesis pathway of its cyanobacterial hosts for PEB synthesis pigments during infection using DHBV-related proteins in conjunction with PBS enzymes.

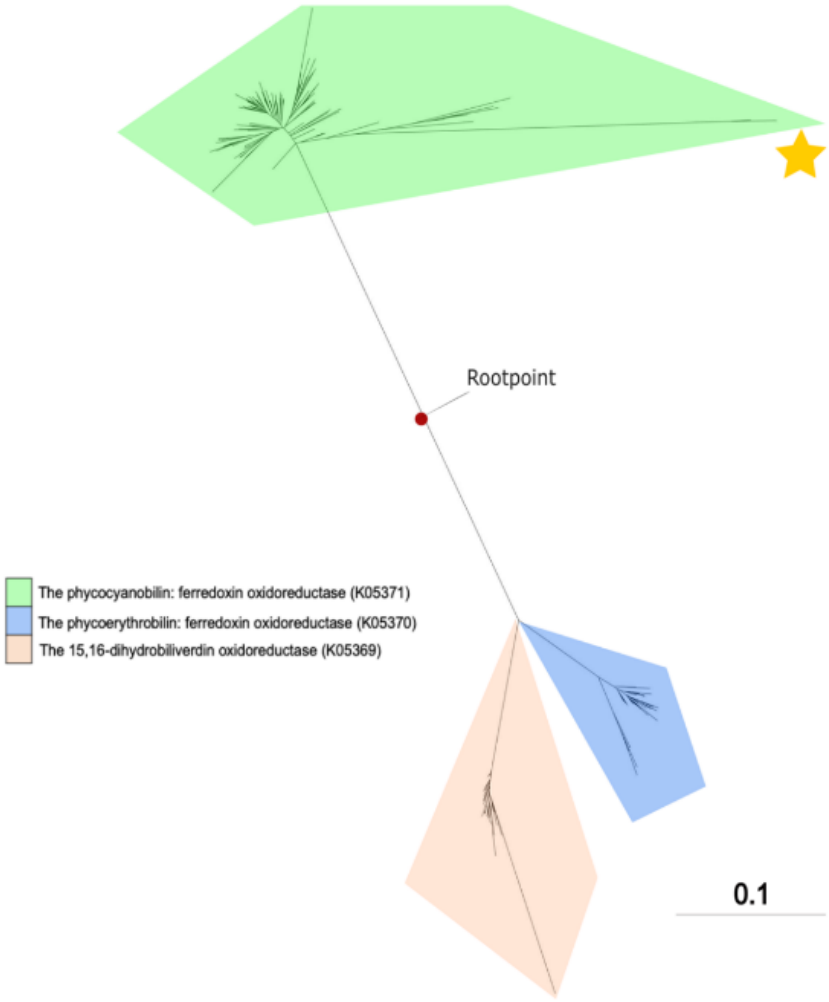

Figure 12. Phylogeny of the PEB and PCB synthesis proteins. The phycocyanobilin: ferredoxin oxidoreductase (K05371) clade is shown in green, the phycoerythrobilin: ferredoxin oxidoreductase (K05370) clade in blue, and the 15,16-dihydrobiliverdin oxidoreductase (K05369) clade in light orange. The star highlights two proteobacterial sequences.

4. Discussion

Homologous searches, domain annotation, and phylogenetic analyses were performed to investigate how biliverdin metabolism differs across species and how it evolved over time. Specifically, two families of BVRs, the cyanobacterial and the F_{420}-dependent BVR present in several pathogenic bacteria, as well as the FDRBS proteins PcyA, PebA and PebB were used in this analysis. The results indicate that enzymes involved in biliverdin metabolism are present in bacterial lineages, and no archaeal sequences were identified as being involved in the enzymatic pathways associated with biliverdin metabolism.

The cyanobacterial BVRs belong to a large family of enzymes with distinct functions, and the retrieved homologs could be further divided into different groups based on sequence similarity and phylogenetic divergence. These homologs share a conserved NAD-binding Rossmann fold PFAM domain (PF01408). In agreement with their functional classification, the sequence division into clades corresponds to their

functional assignments and to different KO models, rather than to their taxonomic origin. For instance, sequences annotated as dihydrodiol dehydrogenase/D-xylose-1-dehydrogenase ($NADP^+$) (KO: K00078), scyllo-inositol 2-dehydrogenase ($NADP^+$) (KO: K22230), and dihydrodiol dehydrogenase/D-xylose-1-dehydrogenase ($NADP^+$) (KO: K00078) split into separate, highly supported clades in the phylogeny. This indicates that the clustering represents evolutionary diversity rather than simple sequence similarity. At the same time, clustering reflects overall sequence similarity. Overall, the high degree of agreement between clustering patterns and phylogenetic topology suggests that the observed clusters among oxidoreductases are actual functional groupings. Consistent KO annotations and shared PFAM domains further support the idea that these enzymes originated from a common ancestral NAD(P)-dependent dehydrogenase. Currently, these homologs are found in *Actinobacteria*, *Firmicutes*, *Proteobacteria*, and, to a lesser extent, *Chloroflexi*. Specifically regarding BVR enzymes, they form a monophyletic and highly supported clade dominated by *Cyanobacteria* sequences (K00214). Interestingly, in this clade, an acidobacterial sequence from one photosynthetic organism is also present. This indicates that this family of BVRs is found exclusively in photosynthetic organisms, further supporting their role in this metabolic process. To our knowledge, this is the first time this enzyme has been identified beyond *Cyanobacteria* or photosynthetic eukaryotes, making it a good candidate for experimental validation. The sister clade of BVRs comprises several proteobacterial phthalate dehydrogenases. Despite the low support of this split, a potential functional correlation for their evolution can be put forward. The phthalate dioxygenase system (PDS) is involved in the aerobic breakdown of aromatic substances. In *Burkholderia cepacia*, PDS operates via an FMN-dependent function within a two-component system, in which the phthalate oxygenase reductase (PDR) possesses FMN and [2Fe-2S] centres that facilitate electron transfer from NADH to the multimeric phthalate oxygenase (PDO)[77]. Since this system depends on oxygen, the functional divergence from the ancestor of both cyanobacterial BVR and PDS might only have occurred after the establishment of a primitive oxygenic photosynthesis system. Thus, cyanobacterial BVR can be seen as a further optimisation of the oxygenic photosynthesis apparatus, possibly occurring in the cyanobacterial ancestor, as seen by its wide distribution within this taxon.

Moreover, BR metabolism and PDS, to which phthalate dioxygenase belongs, use different flavin-based cofactors for electron transfer—F_{420} in BVR and FMN in phthalate dioxygenase. However, their biochemical roles have changed to focus on different metabolic pathways: pigment reduction and aromatic compound degradation, respectively, highlighting the metabolic diversity and functional specialisation of these enzymatic systems.

Outside photosynthetic organisms, BR is produced to shield cells from oxidative stress during infection. The activation of HO in *Actinobacteria*, specifically in *Mycobacterium tuberculosis*-infected macrophages, produces BV, which is then converted to BR[44]. This protects the Bacteria during the early phases of infection. Also, it has been discovered that BR inhibits the synthesis of inducible nitric oxide

synthase, which produces microbicidal nitric oxide in *Mycobacterium tuberculosis* infected human and murine macrophages[78]. The activity of the BVR in the *Actinobacteria* phylum depends on deazaflavin coenzyme F_{420}, a member of the flavin cofactor group a redox cofactor present in the *Actinobacteria* phylum[73]. In addition to serving as an electron carrier in many enzymatic processes, including those catalysed by F_{420}-dependent glucose-6-phosphate dehydrogenases, which keep the cofactor in a reduced form ($F_{420}H_2$), F_{420} is involved in several metabolic and detoxification processes. Furthermore, the unique BVR enzyme Rv2074 belongs to the specific flavin/deazaflavin oxidoreductase (FDOR) superfamily, previously thought to be found only in Actinobacteria [79] and potentially also present in some Firmicutes, as indicated by their taxonomic distribution (Figure 4) and phylogenetic analysis (Figure 11).

$F_{420}(2)$-dependent biliverdin reductases, including the *Mycobacterium tuberculosis* enzyme Rv2074, were used in phylogenetic analyses [44]. These enzymes' wide conservation across *Actinobacteria* and their unique redox function and physiological role in *Mycobacteria,* places them within a well-defined protein family. Along with members of *Acidimicrobiia*, *Thermoleophilia*, and numerous members of the class *Actinomycetia.* a few *Firmicutes* sequences were found in the phylogenetic tree, forming a monophyletic clade. This suggests an HGT event from Actinobacteria to Firmicutes. It would be of interest to investigate their function in these organisms and their potential role in pathogenic metabolism. The short evolutionary divergence within the F_{420}-dependent oxidoreductase family, as indicated by the phylogeny's star-like topology radiating from a central root, might suggest a slow mutational rate, either due to strong selective pressures for sequence conservation or to a short evolutionary time.

A comparative analysis of the pigment synthesis enzymes PcyA, PebA, and PebB reveals essential insights into the evolution and differentiation of the bilin pigment biosynthetic pathway. These enzymes share a conserved (PF05996) domain, belonging to the FDBRs family. The sequences linked to synthesis enzymes PebA and PebB derive from a common internal node in the evolutionary reconstruction, suggesting they have a more recent common ancestor than the PcyA branch. Of note, within the PebA clade, a sequence annotated as PcyA is present (Supplementary data 7.6.1). This most likely represents a case of miss annotation since the root clearly separates PcyA from PebA/PebB major clades, or a secondary event of functional change in which that particular PcyA enzyme derived from a PEB pigment synthesis enzyme.

These proteins are found almost exclusively in *Cyanobacteria*, indicating that the diversification of FDBR enzymes, especially those implicated in PCB and PEB pigment synthesis, occurred within cyanobacterial lineages [80] . However, two *proteobacterial* sequences are in the PcyA clade. These belong to non-photosynthetic organisms, and their in vivo function is not clear. In addition, since they were found in metagenomes, they could indicate assembly contamination. If they indeed are PcyA from p*roteobacteria*, they were acquired via HGT event from cyanobacteria. Altogether, these results highlight the lineage-

specific evolution of bilin pigment production within *Cyanobacteria* and show that the PCB-related enzymes evolved by functional differentiation and duplication[23].

After integration into the cyanobacterial genome, PcyA became crucial for pigment biosynthesis, leading to the formation of complex light-harvesting PBPs and to enzymatic activity in PCB production[81]. The range of chromophores was further broadened by the diversification of PebA and PebB, enabling *Cyanobacteria* and related phototrophs to adapt to diverse light conditions[82]. Interestingly, in the pigment phylogeny (Figure 12), the branch of the sequence PebB enzyme evolves closer to the root and has shorter branch lengths than the branches of sequences of PebA enzyme. However, in both branches, similar taxa indicate gene duplication, suggesting functional diversification in the biosynthesis of biliverdin-derived pigments (Figure 12.) (Supplementary data 7.6.3.).

Synthesis of PEB also occurs in the *Prochlorococcus cyanophage*, where the PebS, which uses DHBV-substrate of PebA is found in cryptophyte biliproteins [83]. Moreover, the *Prochlorococcus* contains genes involved in the biosynthesis of PCB and PEB, with several PBS protein structures [84]. More importantly, evidence of their significance for host fitness comes from the discovery of PBS-related genes in several marine cyanophage genomes. These genes are frequently found in *Myovirus* genomes, which usually encode one to three PBS-related genes, but they are not present in cyanophage podoviruses [85].

5. Conclusion

In conclusion, the comparative and evolutionary analyses performed here emphasise the evolutionary maintenance of the biochemical BV metabolism in Bacteria. Specific phyla, including *Cyanobacteria, Proteobacteria*, *Firmicutes*, and *Actinobacteria*, were consistently represented, suggesting that these phyla share conserved functional domains essential for detoxification and pigmentation, as well as, specifically in *Cyanobacteria*, for photosynthesis. Evolutionarily, BV metabolism has evolved within bacterial lineages, with *Cyanobacteria* showing strong conservation of the BVR pathway and divergence in the synthesis enzyme PcyA.

In contrast, structurally similar domains across different phyla can exhibit lineage-specific functional variation. For instance, in *Actinobacteria* and *Firmicutes*, enzymes involved in BR metabolism use the F_{420} cofactor through a different mechanism. In addition to highlighting structural and functional continuity among these enzymes, a combination of functional annotation and phylogenetic analysis additionally indicates divisions that might be indicative of metabolic specialisations or ecological adaptations. Together, these findings suggest a modular model of pigment evolution, in which cyanobacterial bilin pathways have evolved by adapting to pre-existing bacterial FDBR enzymes.

Collectively, the study results demonstrate the evolutionary complexity and conserved nature of biliverdin metabolism pathways across various lineages.

5.1. Future aspects

In this study, putative BVRs were found in phyla beyond *Cyanobacteria* and Actinobacteria. These enzymes are strong candidates for experimental characterisations, as computational analyses can only offer predictions and propose metabolic potential. Characterising them could broaden our understanding of biliverdin metabolism in prokaryotes and potentially reveal new biological functions.

6. References

1. Matthews, S. J. *et al.* MhuD from Mycobacterium tuberculosis: Probing a Dual Role in Heme Storage and Degradation. *ACS Infect. Dis.* **5**, 1855–1866 (2019).
2. Mochizuki, N. *et al.* The cell biology of tetrapyrroles: a life and death struggle. *Trends in Plant Science* **15**, 488–498 (2010).
3. Bryant, D. A., Hunter, C. N. & Warren, M. J. Biosynthesis of the modified tetrapyrroles—the pigments of life. *Journal of Biological Chemistry* **295**, 6888–6925 (2020).
4. Schlicke, H. *et al.* Function of Tetrapyrroles, Regulation of Tetrapyrrole Metabolism and Methods for Analyses of Tetrapyrroles. *Procedia Chemistry* **14**, 171–175 (2015).
5. Barber, J. Photosynthetic generation of oxygen. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2665–2674 (2008).
6. Tanaka, R. & Tanaka, A. Tetrapyrrole biosynthesis in higher plants. *Annu Rev Plant Biol* **58**, 321–346 (2007).
7. Brânzanic, A. M. V., Ryde, U. & Silaghi-Dumitrescu, R. Why does sulfite reductase employ siroheme? *Chem. Commun.* **55**, 14047–14049 (2019).
8. Mathur, Y. & Hazra, A. B. Methylations in vitamin B12 biosynthesis and catalysis. *Current Opinion in Structural Biology* **77**, 102490 (2022).
9. Choby, J. E. & Skaar, E. P. Heme Synthesis and Acquisition in Bacterial Pathogens. *J Mol Biol* **428**, 3408–3428 (2016).
10. Editors, B. D. Heme - Definition, Structure and Function. *Biology Dictionary* https://biologydictionary.net/heme/ (2018).
11. Hayashi, S. *et al.* Characterization of rat heme oxygenase-3 gene. Implication of processed pseudogenes derived from heme oxygenase-2 gene. *Gene* **336**, 241–250 (2004).
12. Florczyk, U. M., Jozkowicz, A. & Dulak, J. Biliverdin reductase: new features of an old enzyme and its potential therapeutic significance. *Pharmacol Rep* **60**, 38–48 (2008).
13. Wilks, A. & Ikeda-Saito, M. Heme Utilization by Pathogenic Bacteria: Not All Pathways Lead to Biliverdin. *Acc. Chem. Res.* **47**, 2291–2298 (2014).

14. Zhu, W., Wilks, A. & Stojiljkovic, I. Degradation of Heme in Gram-Negative Bacteria: the Product of the hemO Gene of Neisseriae Is a Heme Oxygenase. *J Bacteriol* **182**, 6783–6790 (2000).
15. Wilks, A. & Schmitt, M. P. Expression and Characterization of a Heme Oxygenase (Hmu O) from*Corynebacterium diphtheriae*. *Journal of Biological Chemistry* **273**, 837–841 (1998).
16. Ratliff, M., Zhu, W., Deshmukh, R., Wilks, A. & Stojiljkovic, I. Homologues of Neisserial Heme Oxygenase in Gram-Negative Bacteria: Degradation of Heme by the Product of thepigA Gene of Pseudomonas aeruginosa. *Journal of Bacteriology* **183**, 6394–6403 (2001).
17. Matsui, T. *et al.* Heme Degradation by Staphylococcus aureus IsdG and IsdI Liberates Formaldehyde Rather Than Carbon Monoxide. *Biochemistry* **52**, 3025–3027 (2013).
18. Nambu, S., Matsui, T., Goulding, C. W., Takahashi, S. & Ikeda-Saito, M. A New Way to Degrade Heme. *Journal of Biological Chemistry* **288**, 10101–10109 (2013).
19. Scientific Image and Illustration Software | BioRender. https://www.biorender.com/.
20. Whitby, F. G., Phillips, J. D., Hill, C. P., McCoubrey, W. & Maines, M. D. Crystal Structure of a Biliverdin IXα Reductase Enzyme–Cofactor Complex. *Journal of Molecular Biology* **319**, 1199–1210 (2002).
21. Yamaguchi, T., Komoda, Y. & Nakajima, H. Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from human liver. Purification and characterization. *J Biol Chem* **269**, 24343–24348 (1994).
22. Schluchter, W. M. & Glazer, A. N. Characterization of Cyanobacterial Biliverdin Reductase: CONVERSION OF BILIVERDIN TO BILIRUBIN IS IMPORTANT FOR NORMAL PHYCOBILIPROTEIN BIOSYNTHESIS *. *Journal of Biological Chemistry* **272**, 13562–13569 (1997).
23. Rockwell, N. C., Martin, S. S. & Lagarias, J. C. Elucidating the origins of phycocyanobilin biosynthesis and phycobiliproteins. *Proceedings of the National Academy of Sciences* **120**, e2300770120 (2023).
24. Molecular composition of cyanobacterial phycobilisomes. https://www.pnas.org/doi/epdf/10.1073/pnas.74.4.1635 doi:10.1073/pnas.74.4.1635.
25. Cyanobacterial Phycobiliproteins. https://encyclopedia.pub/entry/53886.
26. Chen, Y.-R., Su, Y. & Tu, S.-L. Distinct phytochrome actions in nonvascular plants revealed by targeted inactivation of phytobilin biosynthesis. *Proceedings of the National Academy of Sciences* **109**, 8310–8315 (2012).
27. Alvey, R. M., Biswas, A., Schluchter, W. M. & Bryant, D. A. Effects of Modified Phycobilin Biosynthesis in the Cyanobacterium Synechococcus sp. Strain PCC 7002. *Journal of Bacteriology* **193**, 1663–1671 (2011).
28. Frankenberg, N. & Lagarias, J. C. Phycocyanobilin:Ferredoxin Oxidoreductase ofAnabaena sp. PCC 7120: BIOCHEMICAL AND SPECTROSCOPIC CHARACTERIZATION *. *Journal of Biological Chemistry* **278**, 9219–9226 (2003).

29. Falk, H. *The Chemistry of Linear Oligopyrroles and Bile Pigments*. (Springer Science & Business Media, 1989).
30. Phycoerythrobilin Synthase (PebS) of a Marine Virus - Journal of Biological Chemistry. https://www.jbc.org/article/S0021-9258(20)57682-5/fulltext.
31. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W. & Frankenberg-Dinkel, N. Efficient Phage-Mediated Pigment Biosynthesis in Oceanic Cyanobacteria. *Current Biology* **18**, 442–448 (2008).
32. Kohchi, T. *et al.* The Arabidopsis HY2 Gene Encodes Phytochromobilin Synthase, a Ferredoxin-Dependent Biliverdin Reductase. *Plant Cell* **13**, 425–436 (2001).
33. Matsuo, T. *et al.* Archaean green-light environments drove the evolution of cyanobacteria's light-harvesting system. *Nat Ecol Evol* **9**, 599–612 (2025).
34. Oliver, T. *et al.* The Evolution and Evolvability of Photosystem II. *Annual Review of Plant Biology* **74**, 225–257 (2023).
35. Zak, E. *et al.* The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. *Proc Natl Acad Sci U S A* **98**, 13443–13448 (2001).
36. Álvarez, C. *et al.* Symbiosis between cyanobacteria and plants: from molecular studies to agronomic applications. *J Exp Bot* **74**, 6145–6157 (2023).
37. Wu, S.-H., McDowell, M. T. & Lagarias, J. C. Phycocyanobilin Is the Natural Precursor of the Phytochrome Chromophore in the Green Alga Mesotaenium caldariorum *. *Journal of Biological Chemistry* **272**, 25700–25705 (1997).
38. Bibby, T. S., Nield, J. & Barber, J. Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**, 743–745 (2001).
39. Falk, S., Samson, G., Bruce, D., Huner, N. P. A. & Laudenbach, D. E. Functional analysis of the iron-stress induced CP 43′ polypeptide of PS II in the cyanobacterium Synechococcus sp. PCC 7942. *Photosynth Res* **45**, 51–60 (1995).
40. Chen, M., Zhang, Y. & Blankenship, R. E. Nomenclature for membrane-bound light-harvesting complexes of cyanobacteria. *Photosynth Res* **95**, 147–154 (2008).
41. Hu, Q. *et al.* Molecular structure, localization and function of biliproteins in the chlorophyll *a*/*d* containing oxygenic photosynthetic prokaryote *Acaryochloris marina*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1412**, 250–261 (1999).
42. Marquardt, J., Senger, H., Miyashita, H., Miyachi, S. & Mörschel, E. Isolation and characterization of biliprotein aggregates from Acaryochloris marina, a Prochloron-like prokaryote containing mainly chlorophyll d. *FEBS Letters* **410**, 428–432 (1997).
43. Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium Acaryochloris marina | PNAS. https://www.pnas.org/doi/10.1073/pnas.0709772105.
44. Ahmed, F. H. *et al.* Rv2074 is a novel F420H2-dependent biliverdin reductase in Mycobacterium tuberculosis. *Protein Science* **25**, 1692–1709 (2016).

45. GitHub · Build and ship software on a single, collaborative platform · GitHub. https://github.com/.
46. ChatGPT. https://chatgpt.com/.
47. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
48. EMBOSS: needle manual. https://www.bioinformatics.nl/cgi-bin/emboss/help/needle.
49. MCL - a cluster algorithm for graphs. https://micans.org/mcl/.
50. Raivo Kolde. pheatmap: Pretty Heatmaps. 1.0.12 https://doi.org/10.32614/CRAN.package.pheatmap (2010).
51. KEGG: new perspectives on genomes, pathways, diseases and drugs | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/45/D1/D353/2605697.
52. KEGG Database. https://www.genome.jp/kegg/kegg1.html.
53. Mao, X., Cai, T., Olyarchuk, J. G. & Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793 (2005).
54. Eddy, S. R. HMMER User's Guide.
55. Mulder, N. J. *et al.* InterPro, progress and status in 2005. *Nucleic Acids Res* **33**, D201–D205 (2005).
56. EMBL-EBI & Institute, E. B. Job Dispatcher homepage | EMBL-EBI. https://www.ebi.ac.uk/jdispatcher/.
57. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
58. Procter, J. B. *et al.* Alignment of Biological Sequences with Jalview. in *Multiple Sequence Alignment: Methods and Protocols* (ed. Katoh, K.) 203–224 (Springer US, New York, NY, 2021). doi:10.1007/978-1-0716-1036-7_13.
59. trimAl — trimAl. https://trimal.readthedocs.io/en/latest/.
60. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
61. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
62. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* **44**, W232–W235 (2016).
63. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* **1**, 1–7 (2017).
64. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **30**, 1188–1195 (2013).
65. GADPH_aa-bio_dh (CL0139) - set - InterPro. https://www.ebi.ac.uk/interpro/set/pfam/CL0139/.

66. Taberman, H., Parkkinen, T. & Rouvinen, J. Structural and functional features of the NAD(P) dependent Gfo/Idh/MocA protein family oxidoreductases. *Protein Sci* **25**, 778–786 (2016).
67. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290–D301 (2012).
68. Kingston, R. L., Scopes, R. K. & Baker, E. N. The structure of glucose-fructose oxidoreductase from Zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure* **4**, 1413–1428 (1996).
69. NADP_Rossmann (CL0063) - set - InterPro. https://www.ebi.ac.uk/interpro/set/pfam/CL0063/.
70. Oxidoreductase family, NAD-binding Rossmann fold (PF01408) - Pfam entry - InterPro. https://www.ebi.ac.uk/interpro/entry/pfam/PF01408/.
71. HEXAPEP (CL0536) - set - InterPro. https://www.ebi.ac.uk/interpro/set/pfam/CL0536/.
72. Safo, M. K. *et al.* X-ray structure of Escherichia coli pyridoxine 5′-phosphate oxidase complexed with FMN at 1.8 Å resolution. *Structure* **8**, 751–762 (2000).
73. Unexpected Abundance of Coenzyme F420-Dependent Enzymes in Mycobacterium tuberculosis and Other Actinobacteria | Journal of Bacteriology. https://journals.asm.org/doi/10.1128/jb.00425-10.
74. Reaction centers of the thermophilic microaerophile, Chloracidobacterium thermophilum (Acidobacteria) I: biochemical and biophysical characterization | Photosynthesis Research. https://link.springer.com/article/10.1007/s11120-019-00650-9.
77. Tarasev, M. & Ballou, D. P. Chemistry of the Catalytic Conversion of Phthalate into Its cis-Dihydrodiol during the Reaction of Oxygen with the Reduced Form of Phthalate Dioxygenase. *Biochemistry* **44**, 6197–6207 (2005).
78. Idelman, G., Smith, D. L. H. & Zucker, S. D. Bilirubin inhibits the up-regulation of inducible nitric oxide synthase by scavenging reactive oxygen species generated by the toll-like receptor 4-dependent activation of NADPH oxidase. *Redox Biology* **5**, 398–408 (2015).
79. Ahmed, F. H. *et al.* Sequence–Structure–Function Classification of a Catalytically Diverse Oxidoreductase Superfamily in Mycobacteria. *Journal of Molecular Biology* **427**, 3554–3571 (2015).
80. Ledermann, B. *et al.* Evolution and molecular mechanism of four-electron reducing ferredoxin-dependent bilin reductases from oceanic phages. *The FEBS Journal* **285**, 339–356 (2018).
81. Jaubert, M. *et al.* A Singular Bacteriophytochrome Acquired by Lateral Gene Transfer *. *Journal of Biological Chemistry* **282**, 7320–7328 (2007).
82. Rockwell, N. C., Lagarias, J. C. & Bhattacharya, D. Primary endosymbiosis and the evolution of light and oxygen sensing in photosynthetic eukaryotes. *Front Ecol Evol* **2**, 10.3389/fevo.2014.00066 (2014).
83. Overkamp, K. E. *et al.* Insights into the Biosynthesis and Assembly of Cryptophycean Phycobiliproteins♦. *J Biol Chem* **289**, 26691–26707 (2014).

84. Coexistence of phycoerythrin and a chlorophyll a/b antenna in a marine prokaryote. | PNAS. https://www.pnas.org/doi/abs/10.1073/pnas.93.20.11126.
85. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* **3**, e144 (2005).

7. Supplementary data

7.1. Annotation supplementary data

https://docs.google.com/spreadsheets/d/1tnQ_h5wAb8L6cks5_RKVeZs6WeAF4u0P/edit?usp=sharing&ouid=102894192486016070981&rtpof=true&sd=true

7.1.1. Table summarising the functional annotations of sequences

KO model	KO functions	Description of function
Annotated functions of Sequences of the BVR homologous		
K00078	Dihydrodiol dehydrogenase /D-xylose-1-dehydrogenase (NADPH)	Participates in detoxification procedures and is an intermediary in the metabolism of polycyclic aromatic hydrocarbons.
K00214	Biliverdin reductase	Heme catabolism
K10219	2-hydroxy-4-carboxymuconate semialdehyde hemiacetal dehydrogenase	Aromatic compound degradation / Xenobiotic metabolism
K13020	UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase	Cell wall biosynthesis / Amino sugar metabolism
K13327	dTDP-3,4-didehydro-2,6-dideoxy-alpha-D-glucose 3-reductase	Sugar nucleotide biosynthesis / Bacterial polysaccharide synthesis
K18067	phthalate 4,5-cis-dihydrodiol dehydrogenase	Aromatic compound degradation Microbial metabolism in diverse environments
K19181	1,5-anhydro-D-fructose reductase (1,5-anhydro-D-mannitol-forming)	Metabolism of specific sugars; possible impact on human health.
K20571	Dehydrogenase : Reaction of UDP-glucose:NAD+ 3-oxidoreductase	Biosynthesis of secondary metabolites Biosynthesis of nucleotide sugars
K20590	Gentamicin A2 dehydrogenase	Neomycin, kanamycin and gentamicin biosynthesis
K20681	dTDP-3-amino-3,6-dideoxy-alpha-D-galactopyranose 3-N-acetyltransferase	Polyketide sugar unit biosynthesis
K22230	scyllo-inositol 2-dehydrogenase (NADP+)	Inositol phosphate metabolism
K24251	dTDP-D-Quip3N acetylase	Sugar nucleotide biosynthesis / Bacterial polysaccharide synthesis
K24995	Levoglucosan dehydrogenase	Carbohydrate derivative / Biomass sugar metabolism
K26172	Puerarin oxidase	Plant secondary metabolism / Flavonoid glycosides
Annotated functions of sequences of PcyA, PebA, and PebB enzymes		
K05371	phycocyanobilin: ferredoxin oxidoreductase phycoerythrobilin: ferredoxin oxidoreductase	
K05370		
K05369	15,16-dihydrobiliverdin:ferredoxin oxidoreductase	
Annotated function of sequences of $F_{420}H(2)$-dependent biliverdin reductase		
K23825	$F_{420}H(2)$-dependent biliverdin reductase	Reductase uses ferredoxin and is dependent on H2

7.2. Supplementary data of KO annotation in clusters

7.2.1. Filtered table based on the KO groups

https://drive.google.com/file/d/1qb5zmku6d3zSe0B6Z_LXJOtSgwzQmmBC/view?usp=sharing

7.2.2. The list of percentages of KOs present across clusters in taxonomic distribution

https://docs.google.com/spreadsheets/d/1i-AIzuTFaJDuFDHXoOIV3YhpjLUkcdfhMLG5RTjgB38/edit?usp=sharing

7.2.3. The taxonomy in-house genome database

https://drive.google.com/file/d/1gbWL_wKA7qUj_zICP15JvOMO8QRAbX-9/view?usp=sharing

7.3. Supplementary data: Bit score model of KO cut-offs

https://drive.google.com/drive/folders/1vPYxP9hmDWNXOtK4ssIDGsvYKd2Xp_Nk?usp=sharing

7.4. Presence Absence supplementary data

7.4.1. The total number of phyla per cluster, and a list of specific phyla in each cluster of BVR.

Number of clusters	Total number of Phyla having a hit in the cluster	Phyla type	
Cluster 1	45	Abditibacteriota, Acidobacteria, Actinobacteria Armatimonadetes, Bacteroidetes, Balneolaeota, Candidatus_Abyssubacteria, Candidatus_Aerophobetes, Candidatus_Blackallbacteria, Candidatus_Cloacimonetes, Candidatus_Edwardsbacteria, Candidatus_Eremiobacteraeota, Candidatus_Firestonebacteria, Candidatus_Glassbacteria, Candidatus_Hydrogenedentes, Candidatus_Latescibacteria, Candidatus_Liptonbacteria, Candidatus_Sumerlaeota, Candidatus_Tectomicrobia, Candidatus_Thermoplasmatota, Candidatus_Woesearchaeota, Chlamydiae, Chloroflexi, Coprothermobacterota, Crenarchaeota, Cyanobacteria,	Deferribacteres, Deinococcus-Thermus, Elusimicrobia, Euryarchaeota, Firmicutes, Gemmatimonadetes Lentisphaerae, Nitrospinae, Nitrospirae, Planctomycetes, Proteobacteria, Rhodothermaeota, Spirochaetes Synergistetes, Tenericutes, Thaumarchaeota, Thermodesulfobacteria, Thermotogae, Verrucomicrobia
Cluster 2	2	Actinobacteria, Firmicutes	
Cluster 3	7	Abditibacteriota, Actinobacteria, Chloroflexi, Planctomycetes, Proteobacteria, Spirochaetes, Verrucomicrobia	
Cluster 4	2	Planctomycetes, Proteobacteria	
Cluster 5	2	Gemmatimonadetes, Planctomycetes	
Cluster 6	1	Chloroflexi	
Cluster 7	2	Bacteroidetes, Cyanobacteria	
Cluster 8	1	Proteobacteria	
Cluster 9	1	Cyanobacteria	
Cluster 10	1	Actinobacteria	
Cluster 11	1	Planctomycetes	

Figure 13. Supplementary data 7.4.1.: total number of BVR homologs sequences within total number of Phyla per clusters, 1-11.

7.4.2. The total number of genera per cluster of enzymes PcyA, and PebA/PebB.

Clusters	Total number of genera with a hit in the cluster	Genera					
Cluster 1	109	Acaryochloris	Crocosphaera	Falsiroseomonas	Mastigocladus	Raphidiopsis	Unclassified genus
		Aliterella	Chondrocystis	Fischerella	Mastigocoleus	Richelia	Unclassified genus
		Alkalinema	Chroogloeocystis	Fortie	Merismopedia	Rippkaea	Unclassified genus
		Allocoleopsis	Chrysosporum	Geitlerinema	Microchaete	Rivularia	Unclassified genus
		Anabaena	Coleofasciculus	Geminocystis	Microcystis	Roseofilum	Unclassified genus
		Anabaenopsis	Crinalium	Gloeobacteria	Microcoleus	Rubidibacter	Vulcanococcus
		Aphanizomenon	Crocosphaera	Gloeocapsa	Moorena	Scytonema	Westiellopsis
		Aphanocapsa	Cyanobacterium	Gloeocapsopsis	Myxosarcina	Shackletoniella	Xenococcus
		Aphanothece	Cyanobium	Gloeomargarita	Neosynechococcus	Snowella	
		Arthrospira	Cyanosarcina	Gloeothece	Nodularia	Sphaerospermopsis	
		Baaleninema	Crocosphaera	Halomicronema	Nodosilinea	Spirulina	
		Candidatus_Atelocyanobacterium	Cylindrospermopsis	Halothece	Nostoc	Stanieria	
		Calothrix	Cylindrospermum	Hormoscilla	Okeania	Stenomitos	
		Chamaesiphon	Cuspidothrix	Hydrococcus	Paracraurococcus	Synechococcus	
		Chlorogloea	Cylindrospermopsis	Hyella	Parasynechococcus	Synechocystis	
		Chlorogloeopsis	Cuspidothrix	Leptolyngbya	Phormidesmis	Thermosynechococcus	
		Chondrocystis	Dactylococcopsis	Limnoraphis	Phormidium	Thermostichus	
		Chroogloeocystis	Desertifilum	Limnospira	Pleurocapsa	Tolypothrix	
		Chrysosporum	Dolichospermum	Limnothrix	Planktothricoides	Trichodesmium	
		Coleofasciculus	Dulcicalothrix	Lyngbya	Prochlorococcus	Trichormus	
		Crinalium	Euhalothece	Mastigocladopsis	Prochlorothrix	Tychonema	
Cluster 2	1	Chondromyces					

Figure 14.Supplementary data 7.4.2.: The total number of PcyA , and PebA/PebB sequence of genera per clusters 1, and 2.

7.4.3. Heatmap genus count

https://drive.google.com/file/d/1FnfIUFD07mHkb5ZxC5ocyzxkpVIXUMCb/view?usp=sharing

7.4.4. Heatmap of homologous sequences of PcyA , and PebA/PebB

https://drive.google.com/file/d/1wc6JqGny0j__NJqHtF-_xcJQtn__SHYs/view?usp=sharing

7.4.5. Heatmap of presence absence of BVR and homologs

https://drive.google.com/file/d/1DpFUDBchlGFMMd18HN4wySDDP0mX0ILX/view?usp=sharing

7.5. Intercluster identity data

7.5.1. Intercluster identity with filter 25

https://docs.google.com/spreadsheets/d/1nfhAd4CYn2ZJoA8vAMdsLV9RfNvEuK7R/edit?usp=sharing&ouid=102894192486016070981&rtpof=true&sd=true

7.6. Supplementary data for phylogenies

7.6.1. Phylogeny of PcyA and PebA/PebB enzyme

https://drive.google.com/file/d/1JfrK1Zc4fDTbde1gylZLrCGr4HUjVC33/view?usp=sharing

7.6.2. Queries of the PcyA, and PebA/PebB enzyme

https://drive.google.com/file/d/1OGPDnXgw5jP3kFuy_a6bEGroXj7LlJlf/view?usp=sharing

7.6.3. Families of the PcyA, and PebA/PebB enzymes

https://drive.google.com/file/d/148HkZt8kwrv7jbZEXlR0pD56Rrn7Z2-S/view?usp=sharing