THE UNIVERSITY *of* EDINBURGH

Computational Comparative Genomics in Cyanobacteria

THE UNIVERSITY *of* EDINBURGH

Maria José Q. Mantas

Submitted for the degree of Doctor of Philosophy
Institute of Ecology and Evolution
University of Edinburgh
2022

Declaration

I declare that this thesis is of my own composition, except where explicitly stated in the text. This work has not been submitted for any other degree or professional qualification.

Maria José Q. Mantas

July, 2022

Acknowledgments

Getting here, to the end of my PhD, was certainly not an easy task, and there were many days where I thought I would not see this day through. It would surely not have been possible without the army of people behind me, who, knowingly or not, pushed me across the finish line. I would like to take this moment to thank them.

To my supervisor, Daniel Barker, thank you for the opportunity to work with you. I think it takes courage to take someone like me in, who knew nothing about Bioinformatics (not even what a Terminal was!), and hope that it would work out fine. From the beginning, I felt like you had more confidence in me than I had for myself, and that was probably what kept me going, especially through the hard times. Many PhDs-to-be have asked me about my doctorate experience and the one advice I always give them is: find a mentor that is a good scientist, but a better person. I was fortunate enough to have had both. Wherever I end up next, they certainly have big shoes to fill!

I want to thank my second supervisor, Catherine, and my committee's chair, Sara, for always supporting me throughout my time as a student. And to Frank, with whom I unfortunately had few meetings with, but who showed me that knowledge serves no purpose if not readily and eagerly shared with others. Frank would often say to me "remember, this is only your PhD!". Him passing away should be a reminder that there are other things in life outside work, and that the relationships we keep are far more valuable than any grade or degree.

I had many collaborators whose input in my research were most valuable. I cannot put in words how grateful I am for all that they taught me, for reviewing my analyses, correcting countless manuscripts and motivating me to always be better. To Dannie, Maureen, Peter and Geoff, thank you for helping me see through the end of the HGT and the cyanotoxins projects, respectively. I have never met a group of scientists so continuously curious and motivated to learn. Thank you for showing me that the interest in science should not be confined to the times when you are a student, but throughout life, even after retiring.

To Stevie, Joe and Eddie, thank you for being the best research group I could have asked for. Not just for always extending a helping hand, but for making this journey so much more fun

than it would have been otherwise. I have missed you and our time in G10 dearly, especially since the pandemic. To the dream-team Jon, Tom, Max, Rory and Surabhi, thank you for babysitting me and for helping me navigate through my first year. It was a pleasure sharing an office with you. A special mention to Stevie, Eddie, Toby, Rory and Lewis: countless times you went out of your ways to help me, wasting precious hours of your very busy days to guide me through all sorts of problems, from code to personal issues. You may not know this, but many times you were the life-vests in a stormy ocean.

I need to dedicate a full paragraph to you, Jack. From a friendly face in Lancaster, to a fellow PhD-sufferer in another University, you grew to be one of my closest friends. Java-jedi, Lord of R, I cannot remember one time when you said you were too busy or too tired to help me. A PhD is an intense experience to go through and I think the bond we share from understanding each other's pain and achievements is special. Thank you for celebrating my milestones, papers and bug-free code as your own.

My time in Edinburgh would not have been nearly as happy (or sane!) without my outside-of-Ashworth friends. To my cousin Diana and my dearest friend Theresa, thank you for showing me that you can be a PhD student and lead a normal and happy life, full of parties, brunches and bad-decisions. To Dani, Nicole, Chloe, Helen, Concha and Robbie: where do I even start! I am so incredibly lucky to have crossed paths with this bunch of talented, beautiful, accomplished, intelligent, fun and diverse group of young women. Thank you for the craziness and the laughter; but most importantly, thank you for your friendship and for loving me for who I am. The Girl Gang is real, and you are living proof that long-distance relationships work.

I want to thank my family, especially my mother, father and brother. Thank you for being a constant in my life, when so many things are changing. For educating me to never give up, to strive for best and to never settle. Thank you for supporting me on this journey since the beginning, and to help cultivate the nerd-bug I had inside of me since very little. You are the reason why I so instinctively throw myself out there and feel confident and comfortable occupying a space at the table.

My mom knows better than anyone else what this journey was and what this means to me. Thank you for the constant presence and support when I am near, but especially when I am far away. Thank you for living my dreams as your own and for being my very best friend.

“She stood in the storm, and when the wind did not blow her way, she adjusted her sails.”

Elizabeth Edwards

I dedicate this thesis to my grandad, Avô Jaime: a man of science, a PhD himself. I am sure you would like to have seen this day through and I hope that, wherever you are, I have made you proud.

Abstract

Cyanobacteria are an ancient clade of photosynthetic prokaryotes, varying in morphology, physiology, biochemistry and habitat. They evolve by typical prokaryotic mechanisms including horizontal gene transfer (HGT). Some species produce toxins (cyanotoxins) that present health hazards to humans and animals, with potential harm to local economies. The biosynthetic pathways and roles of some cyanotoxins are unclear. The rapid increase in high quality publicly available genomes presents opportunities for discovery from comparative genomics in cyanobacteria. The work presented here focuses on three topics in cyanobacteria, using bioinformatics analyses of 130 cyanobacterial genomes. Firstly, I consider hypotheses for the biosynthesis and physiology of the non-encoded neurotoxin 2,4-diaminobutanoic acid (2,4-DAB). Secondly, I consider hypotheses for the biosynthesis and potential roles of its structural analogue, β-*N*-methylaminoalanine (BMAA). These topics use similar methodology: pairwise and multiple sequence alignment, profile hidden Markov models, substrate specificity and active site identification, and the reconstruction of gene phylogenies. We show that some species have genes involved in known biochemical pathways to 2,4-DAB – genes coding for proteins in the aspartate 4-phosphate pathway (including the diaminobutanoate-2-oxo-glutarate transaminase, the downstream decarboxylase, diaminobutanoate decarboxylase, and ectoine synthase) – and BMAA (homologs of the *Staphylococcus aureus* genes *sbnA* and *sbnB*). We highlight the possible involvement of 2,4-DAB and BMAA in the production of siderophores. We show that the biosynthesis of 2,4-DAB and BMAA is likely to be confined to a limited number of species, or to occur via different, unknown, pathways. Thirdly, I investigate hypotheses concerning the association of HGT events with environmental context. I test existing hypotheses claiming that genetic exchanges are more frequent in extreme habitats (*versus* mesophilic) and in terrestrial habitats (*versus* aquatic). My results, based on reconciliation of gene trees with the species tree, do not suggest a link between the prevalence of HGT and extreme or terrestrial environments. I highlight the need for complete descriptions of the isolation source and culture type (axenic, non-axenic monocyanobacterial culture, environmental sample), the need for accurate and robust methods for HGT inference, and for more objective and detailed criteria for environmental classification and of cyanobacterial species. This work contributes to research into cyanobacterial neurotoxins and provides insights into the prevalence and distribution of HGT in cyanobacteria.

Lay Summary

Cyanobacteria are a group of bacteria that have existed for billions of years. Although all species of cyanobacteria are genetically related, they may differ in the way they look, the substances they produce, and the habitats they live in. Some cyanobacteria produce toxins that can be harmful to the health of humans and animals. How and why cyanobacteria produce these toxins is still a mystery. In recent years, scientists across the world have dedicated time and effort to sequencing genomes of cyanobacteria (i.e. reading their DNA). Since DNA specifies the features of the organism, this presents an opportunity to "mine" the DNA data of cyanobacteria for insight into their biology, using computers. The work presented here focuses on three topics in cyanobacteria, using 130 cyanobacterial genomes. I investigate how and why cyanobacteria produce two toxins known by the abbreviations 2,4-DAB and BMAA. We find that some species have the genes for the production of 2,4-DAB and/or BMAA and we believe the production of these toxins is associated with the need to capture iron from the environment. The third topic concerns horizontal gene transfer (HGT). HGT is a process in cyanobacteria and most other bacteria, where genes are passed from organism to another. I investigate whether cyanobacteria experience more or less HGT, depending on the environment they live in. Specifically, I checked whether or not HGT is more frequent in extreme habitats such as hot springs (*versus* benign habitats such as seawater), and in land environments that experience variability in resources such as water (*versus* aquatic environments, which are more stable). My results do not suggest a link between the amount of HGT and either extreme habitats or land habitats. Future work would be assisted by more complete details of the organisms whose genomes were sequenced; and careful use of pure samples for laboratory detection of toxins. These would allow more detailed classification of the environment and make it easier to link toxin production to features of the genome.

Table of Contents

List of Tables and Figures

Abbreviations

16S rRNA	16S small subunit ribosomal RNA gene
2,3-DAP	2,3-diaminopropanoic acid
2,4-DAB	2,4-diaminobutanoic acid
A	alanine
ACEGA	*N*-(1-amino-1-carboxyl-2-ethyl)-S-glutamate
ACI	2-(3-amino-3-carboxypropyl)-isoxazolin-5-one
ACT	Artemis Comparison Tool
AIC	Akaike Information Criterion
ALS-PDC	amyotrophic lateral sclerosis/parkinsonism/dementia complex
antiSMASH	Antibiotics and Secondary Metabolite Analysis Shell
ARG	antibiotic-resistant genes
β-ODAP	β-*N*-oxalyl-2,3-diaminopropanoic acid
BGC	biosynthetic gene cluster
BIA	β-(isoxazolin-5-on-2-yl)-alanine
BMAA	3-*N*-methyl-2,3-diaminopropanoic acid
βUP	β-ureidopropionase
C	cysteine
CASDC	carboxyspermidine decarboxylase
CASDH	carboxyspermidine dehydrogenase
D	aspartate
DAG	S-adenosylmethionine—diacylglycerol
DAPAL	2,3-diaminopropionate ammonia-lyase
DHA	dehydroalanine
DHP	dihydropyrimidinase
DL	duplication-loss
DTL	duplication-transfer-loss
DUD	dihydropyrimidine dehydrogenase
DUF	Domain of Unknown Function
E	glutamic acid
EC	Enzyme Commission
F	phenylalanine
FFT	fast Fourier transform
FR	frames of translation
G	glycine
GFP	green fluorescent protein
GO	Gene Ontology
GOE	Great Oxygenation Event
H	histidine
HCBs	harmful cyanobacterial blooms
HGT	horizontal gene transfer

INSDC	International Nucleotide Sequence Database Collaboration
K	lysine
L	leucine
L-PSP	liver-perchloric acid-soluble protein
LCA	last common ancestor
LGT	lateral gene transfer
LRT	likelihood-ratio test
MAFFT	Multiple Alignment using Fast Fourier Transform
MAM	methylazoxymethanol
MCL	Markov Clustering Algorithm
ML	maximum likelihood
MP	maximum parsimony
MS/MS	tandem mass spectrometry
MSA	multiple sequence alignment
N	asparagine
NCBI	National Centre for Biotechnology Information
NIES	National Institute for Environmental Studies
NIS	NRPS-independent siderophore
NJ	neighbour-joining
NNI	nearest neighbour interchange
NRPS	non-ribosomal peptide synthetase
OAS	*O*-acetyl-L-serine
OASS	*O*-acetyl-S-serine sulfhydrylase
OCD	ornithine cyclodeaminase
OPS	S-*O*-phosphoserine
PCR	polymerase chain reaction
PGAP	Prokaryotic Genome Annotation Pipeline
PHA	polyhydroxy-alkanoates
pHMMs	profile hidden Markov models
PKS	polyketide synthases
PLP	pyridoxal 5'-phosphate
R	arginine
RefSeq	Reference Sequence project
RHAS	rate heterogeneity across sites
RidA	reactive intermediate deaminase A
S	serine
SA	staphyloferrin A
SAM	S-adenosyl-L-methionine
SAMDC	S-adenosylmethionine decarboxylase
SB	staphyloferrin B
smCOGs	specialised metabolism Cluster of Orthologous Groups
SPDS	spermidine synthase
TDH	threonine dehydratase

TIGR	The Institute for Genomic Research
TUB	tuberactinomycin
UFBoot	ultrafast bootstrap
UPGMA	unweighted pair group method with arithmetic mean
WHO	World Health Organisation
Y	tyrosine

Chapter 1

General Introduction

1.1 The 4 W's of Cyanobacteria

1.1.1 Who

Cyanobacteria (from the Greek word κυανος = blue) are an ancient, monophyletic, widespread and diverse group of ecologically important Gram-negative photosynthetic bacteria. They are taxonomically classified within the Kingdom Monera (Prokaryota), Division Eubacteria, class Cyanobacteria and they were originally referred to as "blue-green algae" (initially, cyanobacteria were taxonomically classified as eukaryotic algae; the name is derived from the colour conveyed by the photosynthetic pigment phycocyanin, when present in sufficiently high concentrations) (Whitton and Potts, 2012). According to the current taxonomy, an estimated 150 genera of cyanobacteria, containing over 2000 species, have been identified (Ernst et al., 2006; Hitzfeld et al., 2000). Cyanobacterial taxonomy is an active field of research and their classification, division into orders and number of genera are the subject of recurrent revision (Komarék 2016; Komarék et al., 2014).

Cyanobacteria are morphologically, physiologically and metabolically diverse. They are currently the only recognised prokaryotes capable of oxygenic photosynthesis (although some species can also grow heterotrophically) and, with a global biomass estimated to exceed 10^{15}g, they are of major ecological relevance to the Earth's primary productivity, contributing significantly as a carbon sink (Garcia-Pichel, 2009). A subset of cyanobacterial species fix nitrogen and they play a key role in the global nitrogen budget (Karl et al., 2002). Molecular nitrogen (N^2), although abundant in the environment, is inert and for it to be biologically incorporated it must be reduced through the action of the nitrogenase enzyme complex (Percival and Williams, 2014). Nitrogenase is irreversibly inactivated by free oxygen making this process incompatible with that of photosynthesis, and hence photosynthetic and nitrogen-fixing cyanobacteria had to develop strategies that enabled both processes, in harmony (Postgate, 1998). Cyanobacteria either temporally or spatially separate photosynthesis from nitrogen fixation. Temporal separation strategies entail, for example, that cyanobacteria fix carbon during the day and nitrogen at night (Stal and Krumbein, 1987). Filamentous cyanobacteria undergo a type of spatial separation, where heterocysts - larger, differentiated anaerobic cells, interspersed among vegetative cells or at the end of the trichome – develop in response to a lack of combined nitrogen sources in the environment and are responsible for fixing nitrogen (Wolk et al., 1994). Heterocysts express *nif* genes and synthesise the

nitrogenase enzyme complex (Garcia-Pichel, 2009). Cyanobacteria can be free-living or symbiotic, and symbiotic associations can range from loose mutualistic relationships to obligate endosymbiosis. There are numerous symbiotic partnerships known between cyanobacteria and a wide range of eukaryotic hosts, including diatoms, dinoflagellates, fungi (e.g. lichens), plants (e.g. Cycads, *Azolla*, rice plants, duckweed, mosses) and animals (e.g. sponges, ascidians, corals). These associations are beneficial because cyanobacteria provide the host with fixed nitrogen and/or fixed carbon, receiving protection (i.e. from predators and environmental extremes) in return (Adams et al., 2012).

The genomes of free-living cyanobacteria vary widely in GC content (from 32 to 71%) and in size (from 1.6 to 14 x 10^6 base pairs; 1700 to 10 000 genes), with the smallest cyanobacterial genome sizes equalling those of most bacteria, and the largest matching those of some eukaryotic fungal genomes (Garcia-Pichel, 2009). Cyanobacterial cells are also larger than most other bacteria and archaea cells (which are, on average, 4μm), ranging in size from <1μm for unicellular species (e.g. *Prochlorococcus*), to over 50μm for multicellular types (e.g. *Cyanothece*). They mostly occur in three unicellular morphological shapes: spherical, rod and spiral (e.g. *Synechococcus*, *Chroococcus*), but can also grow in colonies (e.g. *Microcystis*) or filaments (e.g. *Oscillatoria*, *Aphanizomenon*) (Singh and Montgomery, 2011). Apart from heterocysts, cyanobacteria can have other specialist (i.e. differentiated or morphology-specific) cell types, including trichomes, hormogonia and akinetes (Fig. 1.1). In all representatives of the filamentous orders, cells are arranged in trichomes. The number and position of cells within each trichome is characteristic of the genus (Komárek et al., 2003). Hormogonia, distinct reproductive segments of the trichomes, can form upon trichome fragmentation. These short filaments (5-25 cells) exhibit active gliding motility (cyanobacteria do not possess flagella) and are important in symbiotic colonisation (Rippka et al., 1979; Wong and Meeks, 2002). Akinetes are large, thick-walled, spore-like cells that allow filamentous cyanobacterial survival under unfavourable conditions (Flores and Herrero, 2010). They are resistant to desiccation, extreme temperatures, and phosphate-deprivation and, in natural planktonic populations, massive akinete formation occurs at the end of the growth season (Kaplan-Levy et al., 2010). Cyanobacteria can grow as biofilms and some species can proliferate and accumulate at the surface of water producing macroscopic mats and blooms (Rastogi et al., 2015; Rossi et al., 2012).

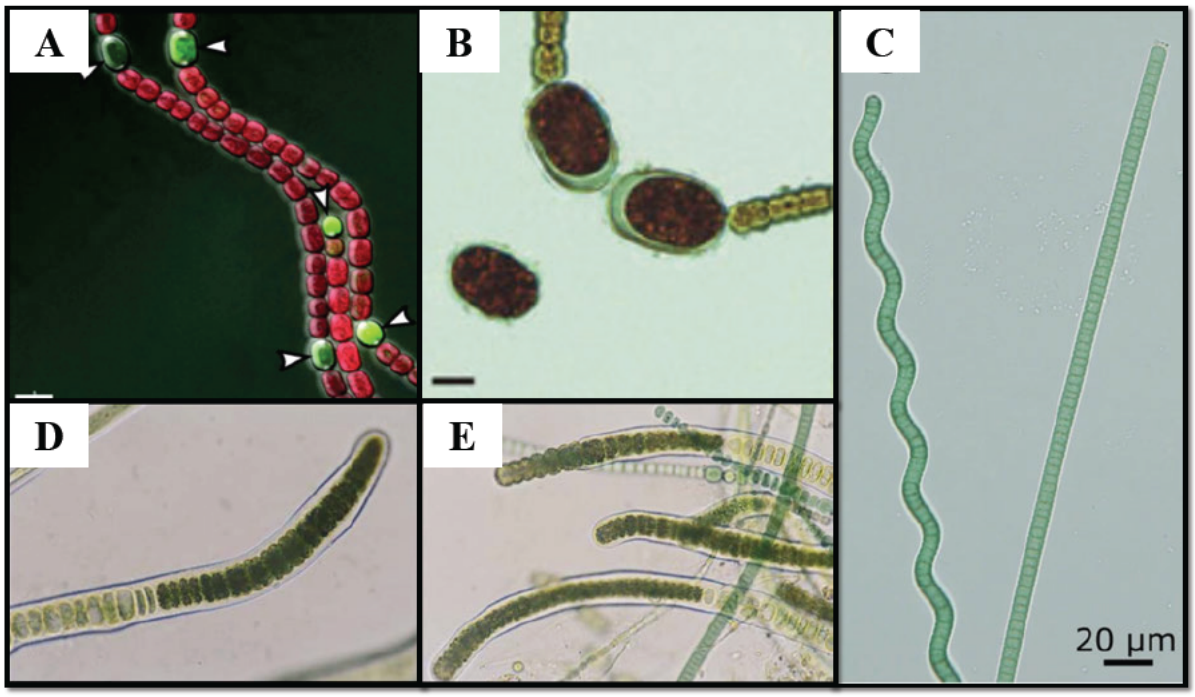


Figure 1.1. Specialised cells in cyanobacteria. A) Arrows point to heterocysts of *Anabaena* PCC 7120. Note that the heterocysts carry a green fluorescent protein (GFP) reporter. Image from Kumar et al. (2010). **B)** A free akinete and two connected akinetes within a filament of vegetative cells of *Aphanizomenon ovalisporum*. Image from Kaplan-Levy et al. (2010). **C)** Morphology of straight and spiral trichomes of *Arthrospira jenneri*. Image from Nowicka-Krawczyk et al. (2019). **D)** and **E)** Apical hormogonia within trichome and release of hormogonium of *Fischerella* sp. Image from Hindák (2012).

Traditionally, cyanobacteria have been classified based on morphology into five orders (also called sections) (Table 1). This allocation into taxonomic groups is not fully supported by phylogenetic studies, i.e. unicellular and simple filamentous forms are polyphyletic (Hayes et al., 2007; Rippka et al., 1979). A new order (order VI, Prochlorophytes) has been recently described, comprising marine species with coccoid shapes (e.g. *Prochloron*, *Prochlorococcus*) (Sukenik et al., 2009).

Table 1.1 Classification of cyanobacteria into five orders. Orders IV and V are monophyletic. Compiled from Sukenik et al. (2009) and Tomitani et al. (2006).

Name	Morphology	Type of asexual reproduction	Examples
Chlorococcales (I)	Unicellular or non-filamentous aggregates of cells. Non-heterocystic.	Binary fission in 1, 2 or 3 symmetric or asymmetric planes, or by budding.	*Cyanobium, Gloeobacter, Gloeocapsa, Synechococcus*
Pleurocapsales (II)	Unicellular or non-filamentous aggregates of cells. Non-heterocystic. Rarely form akinetes.	Internal multiple fission, yielding baeocysts, or binary fission.	*Stanieria, Chroococcidiopsis, Myxosarcina, Pleurocapsa*
Oscillatoriales (III)	Filamentous and non-heterocystic. Forms uniseriate (one cell) trichomes, without true branching (unidirectional filaments in one plane). Does not form akinetes.	Binary fission in one plane.	*Arthrospira, Leptolyngbya, Oscillatoria, Spirulina*
Nostocales (IV)	Filamentous and heterocystic. Forms trichomes with one or more cells, without true branching. Some species produce akinetes.	Binary fission in one plane.	*Calothrix, Nodularia, Nostoc, Tolypothrix*
Stigonematales (V)	Filamentous and heterocystic. Forms multiseriate trichomes in several planes (truly branched). Some species produce akinetes.	Binary fission in more than one plane.	*Hapalosiphone, Fischerella, Geitleria, Stigonema*

1.1.2 When

The ancient ancestors of today's cyanobacteria are amongst the oldest organisms on Earth and are believed to have originated between 2500 and 3500 million years ago (Schirrmeister et al., 2016; Schopf and Packer, 1987). Uncertainty still exists regarding time of origin and of early speciation events due to the scarceness of Precambrian fossil deposits (i.e. most Precambrian sedimentary rocks have been subducted or have been heavily metamorphosed by igneous activity) and limited characteristics for the identification of taxa (Schirrmeister et

al., 2015). Nevertheless, fossil records of cyanobacteria still contain the oldest entries that can be confidently assigned to any extant group of organisms, including microfossil traces from Archaean sedimentary rocks bearing indisputable cyanobacterial morphologies (Furnes et al., 2004) and carbon isotopic data consistent with Rubisco-mediated CO_2-fixation (Schopf, 2012). The abundant organo-sedimentary laminated macrofossils from the Phanerozoic, known as stromatolites, offer convincing evidence that cyanobacteria originated between 2700 and 3200 million years ago and inhabited shallow seas and lakes (Flannery and Walter, 2012; Homann et al., 2015). Stromatolites - biogenic, laminated, biochemical accretionary structures formed in shallow waters by the trapping, binding and precipitating activity of microorganisms - occur nowadays and often in extreme environments (i.e. high temperature, salinity and sediment influx), where grazing by metazoans, algae and protists is considered to be low (Riding, 1999; Taylor et al., 2009). Examples of these structures exist in the Caribbean (Highborne Cay, Bahamas) and in Western Australia (Shark Bay) and are viewed as environmental equivalents to those inhabited by cyanobacteria in the early Earth (Allwood et al., 2006; van Kranendonk, 2006).

Cyanobacteria have played a key role in shaping the Earth's environment and history. As the first photosynthetic organisms to evolve, they are believed to have been responsible for raising the levels of free oxygen in the atmosphere during the Archaean and Proterozoic Eras – an event referred to as the Great Oxygenation Event (GOE), ca. 2.4 billion years ago (Bekker et al., 2004) –, which enabled the oxygenation of Earth's atmosphere, the formation of the protective ozone layer, as well as the development of aerobic respiration and complex life (Schirrmeister et al., 2015). Nevertheless, the timing of emergence of O_2-producing photosynthesis and the extent of cyanobacteria abundance prior to the GOE is still a subject of debate (Lyons et al., 2014). Evidence of transient "oxygen oases" suggests that oxygenic photosynthesis pre-dated the GOE (Ligrone, 2019), but it is known that most major taxonomic groups of cyanobacteria, including the lineage that led to chloroplasts, appeared after this event (Sánchez-Baracaldo and Cardona, 2020). Cyanobacteria are implicated in the evolution of photosynthetic plants and algae. It is believed that, in the late Proterozoic/early Cambrian, events occurred where a cyanobacterial cell was engulfed by a eukaryotic cell. In this mutualistic relationship, cyanobacteria provided organic compounds to the eukaryotic host via photosynthesis and received the protection of the host in return (Sukenik et al., 2009). This endosymbiotic event is thought to have occurred recursively and independently

during the evolution of protists and higher plants, and is the origin of the eukaryotic chloroplast (Falcón et al., 2010; Raven and Allen, 2003; Sagan, 1967).

1.1.3 Where

Cyanobacteria are widely distributed throughout many different environments and can be found in almost every habitat on Earth. They are present in aquatic environments including freshwater, brackish and marine habitats, and also in soils (Chen et al., 2021; Sukenik et al., 2009). Cyanobacteria are prolific in freshwater environments such as tropical, subtropical and temperate wetlands (e.g. Pantanal, the marshes in Florida and Belize and the prairie marshes of North America) (Mayer and Galatowitsch 2001; McCormick and O'Dell, 1996), rice fields (Roger and Ladha, 1992), lake benthos (including spring-fed ponds and subalpine, tropical and subtropical lakes) (Carrick and Steinman, 2001; Dodds and Castenholz, 1987; Higgings et al., 2003; Loeb and Reuter, 1981) and lotic freshwaters (e.g. streams in tropical, tundra, forested, desert and coastal plain regions of North America, Asia, Australia, Africa and Europe) (Grimm and Petrone, 1997; Sheath and Cole, 1992; Ward, 1985). Despite the fact that much of the world's oceans is destitute of life due to their ultra-oligotrophic nature (Holland, 1978), the marine environment is rich in planktonic cyanobacteria, and all major taxonomical groups are presented in the marine plankton. It is estimated that cyanobacteria encompass <10% of the oceanic phytoplankton in high latitudes and up to >50% in tropical oligotrophic waters (Paerl, 2000). They are also present in coastal and estuarine waters, such as the Baltic Sea (Kononen et al., 1996), Peel-Harvey Estuary in Australia (Huber, 1986), Lake Ponchartrain in the USA (Dortch et al., 1999) and coastal lagoons of South America (Paerl and Fulton, 2006). Cyanobacteria are the oldest photoautotrophic component of biological soil crust communities, and are often their most important primary producers (Rivera-Aguilar et al., 2006). They can be encountered across the globe, from the dryland areas of Africa (Büdel et al., 2016) to the mountains at the Northern Urals (Novakovskaya et al., 2022). Cyanobacteria can also populate what are considered extreme habitats. They can survive the extreme cold and near absence of liquid water, characteristic of the cryosphere (Vincent, 2007; Wynn-Williams, 2000). There have been reports of cyanobacteria in snow (e.g. Alaska) (Takeuchi, 2001), glacier ice (e.g. Svalbard) (Stibal et al., 2006), lake ice and ice-shelves (e.g. McMurdo ice shelf) (de los Ríos et al., 2005), and cold desert ecosystems in the Artic, Antarctica and alpine regions. They can be found in many geothermal habitats worldwide, with exemplars of well-studied extremophilic/extremotolerant cyanobacterial

communities existing at the Yellowstone National Park (e.g. White Creek, Norris Geyser Basin, Lower Geyser Basin, Octopus, Mushroom and Clearwater springs) (Miller et al., 2006; Ward and Castenholz, 2000; Ward et al., 2006; 2012). Cyanobacteria can withstand living at high salinities, including in salt lakes, solar salterns, hypersaline lagoons, halite, hypersaline sulphur springs and evaporite crusts of gypsums (Javor, 1989; Oren, 2006). Such habitats are widespread across the world and comprise the Great Salk Lake, Utah (Brock, 1976), the Abu Gabara Lake in Wadi Natrun, Egypt (Imhoff et al., 1979), salterns in Queensland, Australia (Coleman and White, 1993), the Solar Lake, Sinai (Potts, 1980), the Dead Sea (Volcani, 1944), among others. Many cyanobacteria can grow at a pH range from 7 to 11, and they frequently dominate alkaline environments, such as soda lakes across Africa, Europe, Asia, Australia, North, Central and South America (Andreote et al., 2018). Cyanobacteria are present in environments exposed to high levels of UV radiation, spanning urban (e.g. walls and pavements), intertidal (e.g. mudflats and salt water marshes) and terrestrial habitats (e.g. mountainous areas) (Garcia-Pichel, 2009; Seckbach and Oren, 2007), and thrive in contaminated environments with high metal concentrations (Huertas et al., 2014). They are important components of crust and soil structure, especially when subject to water stress, which is the case of many desert regions (Alwathnani and Johansen, 2011; Hagemann et al., 2017). The Atacama Desert, known for its wide range of environmental extremes, including the fact that it is the driest non-polar habitat on Earth, is a prolific site for the study of xerophilic cyanobacteria (Lacap et al., 2011). Overlap exists between environmental conditions in desert regions on Earth and on Mars, which have fueled research concerning the capability of cyanobacteria to survive extraterrestrial conditions (Mapstone et al., 2022). Despite the adaptability of cyanobacteria to life in different environments, bona fide reports of cyanobacterial growth in highly acidic environments, with a $pH < 3$, or under very high or very low pressures (i.e. barophiles) are lacking (Brock, 1973; Paerl et al., 2000).

1.1.4 Why

Research in cyanobacteria is prolific and worthwhile, and here I will focus on two reasons why investigation in this field is important: biotechnological applications of cyanobacteria are numerous and relevant in many different fields, and the hazardous consequences to animal and human health derived from the production of a myriad of different cyanobacterial toxins (cyanotoxins).

Cyanobacteria are rich sources of bioactive compounds and specialised metabolites (syn: secondary metabolites) with potential uses for mankind. Multiple applications have been described in the food, agriculture, cosmetics, pharmaceutical, biotechnology and fuel industries (see below). Regarding the first, some species (e.g. *Arthrospira* and *Aphanizomenon flos-aquae,* generally known as "Spirulina") are used as food by indigenous populations in Africa and Central America, and their commercial importance as a dietary supplement has risen in recent years due to their alleged protein-rich, low-calorie, cholesterol-free, vitamin-loaded properties (Sotiroudis and Sotiroudis, 2013). Compounds derived from the shikimate pathway are of interest to the cosmetic industry as they could potentially be used as natural UV blockers in the formulations of creams and varnishes (Soule et al., 2009) and novel gel extracts from *Aphanothece sacrum* could be useful as moisturisers and anti-inflammatory agents (Okeyoshi et al., 2021). In the pharmaceutical industry, synthesis of cyanobacterial fatty acids is currently being investigated for anticarcinogenic, antibiotic, antifungal and antiviral properties (El-Baz et al., 2013) and Astaxanthin from *Haematococcus pluvialis* is being study as a potential drug therapy in HIV (Singh et al., 2017). There are many other cyanobacterial metabolites with possible antibacterial (Jaki et al., 2000), antitumorigenic (Gerwick et al., 1994), antifungal (Kajiyama et al., 1998) and antimalarial (Papendorf et al., 1998) properties. Concerning agriculture, nitrogen-fixing cyanobacterial species are currently used as a low-cost bio-fertiliser for paddy fields used in rice cultivation in various geographical locations worldwide (Watanabe and Rockwood, 1979). The symbiotic association of *Anabaena* and the water fern *Azolla* is also cultivated for use as green manure and as food for poultry and swine (Kaushik and Venkataraman, 1979). Cyanobacteria have environmental biotechnology applications, such as wastewater purification and bio-remediation as they are capable of oxidising oil components (Martins et al., 2011). Recent studies on polyhydroxy-alkanoate (PHA), which accumulates intracellularly in some cyanobacteria, indicate that this compound may be used in the production of bio-plastics (Steinbüchel et al., 1998). Numerous academic and industrial ventures are investigating the potential use of cyanobacteria as a bio-fuel, since many species, in certain conditions, produce molecular hydrogen via the activity of the nitrogenase and hydrogenase enzyme complexes (Doamekpor et al., 2019; Dutta et al., 2005). Recruiting the efficient carbon assimilation process of cyanobacteria for atmospheric CO_2 sequestering has also been discussed as a potential solution for reducing global levels of carbon dioxide (Sukenik et al., 2009). Efforts to study the largely unexplored cyanobacterial inventory of specialised metabolites are likely to increase in future years.

Although cyanobacterial species are not infectious agents, nor are classified as true pathogens, they can produce various low molecular weight toxins (usually alkaloids and cyclic peptides). The majority of these are potent substances, with LD_{50} in the range between 10 and 200μg/kg (Chorus, 2001). Cyanotoxins are usually named after the original producer organism and are often classified regarding their biological effects on human organs (Codd et al., 2005). They include (I) neurotoxins (e.g. anatoxin-a, anatoxin a-(s), homoanatoxin-a, saxitoxin, neosaxitoxin, 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA) and 2,4-diaminobutanoic acid (2,4-DAB)) (Colas et al., 2021; Llewellyn, 2006; Nunn and Codd, 2017; Vega and Bell, 1967); (II) hepatotoxins (e.g. microcystin and nodularin) (Labine and Minuk, 2009); (III) cytotoxins (e.g. cylindrospermopsin) (Armah et al., 2013); (IV) tumorigenic toxins (e.g. microcystins) (Žegura et al., 2011); and (V) irritants and the gastrointestinal toxins (e.g. lipopolysaccharides, aplysiatoxins, lyngbyatoxins, debromoaplysiatoxin) (Metcalf and Codd, 2012; Stewart et al., 2006). There are other cyanotoxins, whose toxicological profiles are still poorly understood (e.g. microviridin) (Funari and Testai, 2008). Although they have been found in most environments and are considered an ancient, widespread phenomenon, the physiological roles of most cyanotoxins are unknown.

The main exposure route to cyanotoxins is through the ingestion of contaminated water or food (including supplements), although contact-exposure (e.g. in recreational waters) and aerosolization (i.e. through inhalation) can also lead to significant illness (Carmichael et al., 1985). Depending on the toxin, the concentration and the route to exposure, acute symptoms can range from sore throat, pneumonia, earache, allergic reactions and gastroenteritis to liver damage, nervous system drainage and even death (Metcalf and Codd, 2012). The consequences of chronic exposure to cyanotoxins are less clear. However, studies have suggested that higher rates of liver cancer in some parts of China and Serbia may be associated with exposure to hepatotoxins in drinking water (Carmichael, 1994; Svircev et al., 2009).

The current understanding of the ecotoxicology of cyanotoxins is largely derived from investigations of animal and human health incidents. Many animal deaths have been attributed by cyanotoxins, including those of sheep, cattle, horses, pigs, dogs, bats, birds, fish and primates (Codd et al., 2005). Cattle are believed to be at increased risk, given their large

water consumption from (usually) nutrient-enriched water sources, prone to contamination by cyanobacteria (Metcalf and Codd, 2012). Multiple events of cattle deaths by cyanotoxin ingestion have been described (Mez et al., 1997; Saker et al., 1999). Although less frequent, human health incidents, including illness and death, are also continuously reported. Among known events are outbreaks of gastroenteritis from exposure to LPS (Teixeira et al., 1993), hepatomegaly from exposure to cylindrospermopsins (Griffiths and Saker, 2003) and pneumonia-like symptoms from exposure to *Microcystis* scum (Turner et al., 1990). Perhaps the most well-known incident occurred in a haemodialysis clinic in Caruaru, Brazil, where the water to be used in treatment was contaminated with microcystins and cylindrospermopsin, causing the death of 100 patients (Pouria et al., 1998; Carmichael et al., 2001). Widespread fish-kills may occur due to contamination by cyanotoxins or may follow the natural life cycle of cyanobacterial blooms, i.e. bloom decomposition leads to a decline in the available oxygen, resulting in the death of fish by asphyxiation (Metcalf and Codd, 2012).

The production of cyanotoxins is thought to occur in greater quantities during periods of algal blooms. Blooms are most frequent when water temperature is higher (i.e. during summer months in temperate zones and year-round in tropical and sub-tropical regions) and in conditions of elevated inorganic nitrogen and phosphorus (i.e. in eutrophic lakes and reservoirs) (Kosten et al., 2012). The presence of blooms is generally undesirable, due to the accumulation and rotting of stagnant scums, which are accompanied by taste and odour problems (Sukenik et al., 2009). They can have detrimental effects on human and animal health, and can also severely impact local economies (Steffensen, 2008). Although not all cyanobacterial species appear to produce cyanotoxins, at least 25 species have been associated with adverse health effects, and toxic algal blooms have been reported worldwide, including many locations in Europe, USA, Australia, Africa, Asia and New Zealand (Percival and Williams, 2014).

Risk management of cyanobacterial blooms is a complex task. It is thought that most standard drinking water treatments (e.g. coagulation, sedimentation and filtration) are effective only at removing large concentrations of cyanotoxins (Percival and Williams, 2014). The rehabilitation of waters affected by cyanobacterial blooms is difficult as most toxins can remain potent for long periods of time, even after the death of cyanobacterial cells (Saadi et al., 1995). Algal blooms are expected to increase in quantity, extension and duration, following anthropogenic activities and global warming, and the prevention of

bloom formation is believed to be the best way to ensure cyanotoxin-free water (Paerl and Otten, 2013).

1.2 Comparative Genomics

1.2.1 Definition and importance

Cross-species comparative genomics is a field of biology dedicated to the comparison of genome sequences from different species and it is a powerful tool to study evolution. Comparing the characteristics that define different organisms enables the establishment of kinship relationships (i.e. phylogenetic trees) and the discovery of regions of similarity and difference between species. This analysis of evolutionary relationships can provide valuable information on the organisms' genetic makeup, development, phenotype and behaviour and how they change over time. The first two whole bacterial genomes (that of *Haemophilus influenzae* Rd and *Mycoplasma genitalium*) were published in 1995 (Fleischmann et al., 1995; Fraser et al., 1995) and since then, comparative genomics has proved useful in understanding the genetic basis for bacterial diversity and speciation. Generally, in the biomedical field, comparative genomics has allowed breakthroughs in the identification of pathogenesis-related genes/species, susceptibility to disease, as well as the development of new drug targets (Frishman et al., 2003). Ecological applications include strategies for conservation of rare, endangered and even extinct species (Zoonomia Consortium, 2020).

Genome comparisons have revealed an astonishing unity amongst all life (e.g. in microbial genomes, approximately 70% of the genes are highly conserved, over long phylogenetic distances (Koonin et al., 1997; Tatusov et al., 2000)). Nevertheless, more than a third of all genes in sequenced genomes lack known functions and there are over 2000 characterised enzymes that do not have an associated gene (de Crécy-Lagard and Hanson, 2013). One of the biggest challenges in comparative genomics is the accurate functional annotation of genomes. Function prediction is often achieved via homology-based (i.e. BLAST) or context-based methods (i.e. the "guilt by association" principle (Aravind, 2000)), whereby known and unknown genes are grouped through various different and often complementary approaches, and the function of the unknown gene is inferred by the function of the known gene. Some of the strategies often used include comparison of gene location, relative gene order and regulation and the examination of gene loss, duplication and horizontal gene transfer events.

1.2.2 RefSeq: the wonders of publicly-available data

All research that contributed to this thesis was done computationally (i.e. no laboratory experiments were conducted) by the analyses of 130 cyanobacterial genomes downloaded from RefSeq (the Reference Sequence project) (i.e. in 2018, there were 180 cyanobacterial genomes available on RefSeq, however only 130 met our quality criteria, see Chapter 2). If it was not for the ability to access high quality genomic data, for free, sampled from multiple locations across the world (Fig. 1.2), the research presented here would not have been possible. Since 1999, RefSeq has served as an essential resource for genomic, genetic and proteomic research and is internationally recognised as the "gold standard" for genome annotation (Maglott et al., 2000).

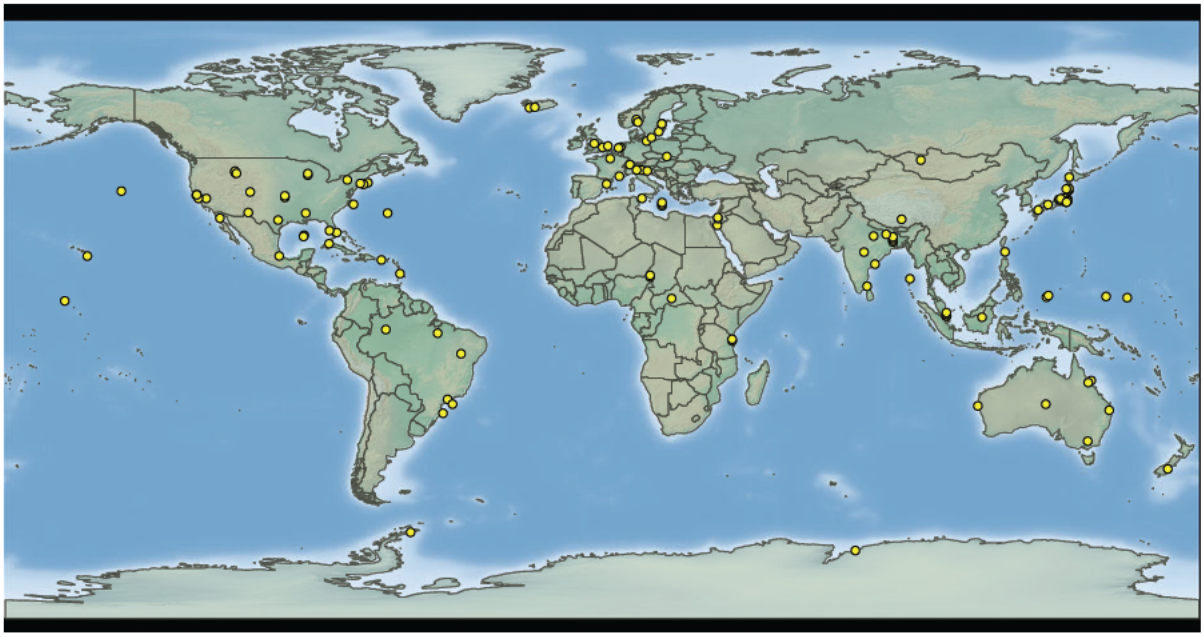


Figure 1.2. Geographic distribution of the 130 cyanobacterial species included in the dataset. There were two instances (*Oscillatoria* sp. PCC 6506 and *"Nostoc azollae"* 0708) where geographic location was not available.

The RefSeq project at the National Centre for Biotechnology Information (NCBI) maintains and curates, at no cost to the user, a publicly available database of annotated genomic, transcript, and protein sequence records for numerous species, from viruses to eukaryotic organisms (http://www.ncbi.nlm.nih.gov/refseq/). The RefSeq project pulls data submitted to the International Nucleotide Sequence Database Collaboration (INSDC) (which includes GenBank, the European Nucleotide Archive and the DNA Data Bank of Japan (Cochrane et al., 2016)) against a combination of computational and manual curation (i.e. using knowledge from publications, users feedback, standardised functional features and informative

nomenclature) in order to produce a standard set of stable, non-redundant reference sequences (O'Leary et al., 2016). RefSeq collaborates with multiple expert groups, including UniProtKB and the Institute for Genomic Research (TIGR) to help maintain and improve data quality through the exchange of sequence information and functional annotation (Pruitt et al., 2005). All RefSeq data has been subjected to quality assurance checks and undergoes regular updates of functional annotation and taxonomy attributions, excluding genomes with poor sequencing and assembly quality and/or with significant contamination (Haft et al., 2018). Among RefSeq's goals are reducing the number of proteins annotated as "hypothetical protein" or "Domain of Unknown Function (DUF)", associating Enzyme Commission (EC) numbers, Gene Ontology (GO) terms, gene symbols, literature references and explanatory comments for every entry. Unlike GenBank, a publicly available archival database of redundant genomic sequence records encompassing several alternate information on protein annotation and name, RefSeq allows researchers to focus on the best representative sequence data and use a stable and consistent coordinate system that can be used to unambiguously reference specific genetic sequences, allowing accurate reporting and reproducible results (Nosek et al., 2015).

The RefSeq prokaryotic genome collection includes bacterial and archaeal genomes with different levels of quality and sampling density. All prokaryotic genomes are consistently annotated via the Prokaryotic Genome Annotation Pipeline (PGAP), which has been continuously upgraded (Tatusova et al., 2016). In response to the growing number of identical proteins shared by many genomes, each with a distinct accession number, NCBI introduced, in 2013, a non-redundant new protein data model and accession prefix (WP) for the RefSeq collection, which drastically reduced the redundancy in prokaryotic proteins (Tatusova et al., 2014). As of this date, the RefSeq collection of prokaryotes has nearly 200 000 genomes and 150 million non-redundant proteins (Li et al., 2021).

1.2.3 Some tools used: understanding what they are

1.2.3.1 Orthogroup prediction using OrthoFinder

The inference of homology relationships (i.e. the relationship between two genes/proteins sequences that have descended from a single last common ancestor (LCA)) is fundamental to our understanding of evolution and species diversity and most comparative genomic applications, such as phylogenetic and metabolic reconstructions, require proteins from

different species to be grouped in orthologous gene families. Homologs (i.e. genes/proteins with a common origin) can be classified as orthologs – homologs derived from a single gene in the LCA by speciation –, paralogs – homologs that are related by gene duplication – or xenologs – homologs derived from horizontal gene transfer events (Fitch, 1970, 2000; Gray and Fitch, 1983). The rationale behind grouping gene families into orthogroups (i.e. a set of genes that descended from a single gene in the LCA of a clade of species) is that they often have the same or similar functions. However, orthogroup prediction is not a trivial task as it is complicated by, among other things, the presence of paralogs and xenologs, i.e. an orthogroup can encompass orthologous, paralogous and xenologous sequences (de Crécy-Lagard and Hanson, 2013).

Ziying Ke grouped the 130 cyanobacterial genomes used in the analyses depicted in Chapter 2, 3 and 4 into orthogroups using OrthoFinder (v.2.1.2) (Emms and Kelly, 2015). OrthoFinder is an orthogroup inference algorithm for protein coding genes across multiple species that has been shown to be more accurate at predicting orthogroups than the previously available software (i.e. MultiParanoic (Alexeyenko et al., 2006), OMA (Altenhoff et al., 2010), TreeFam (Li et al., 2006), eggNOG (Jensen et al., 2007) and the widely used OrthoMCL (Li et al., 2003)) by between 8 and 33%. OrthoFinder addresses gene length bias in orthogroup inference. Other advantages of using this software is that it is fast, freely available, easy to use, robust to the missing genes typical of incomplete genomes and *de novo* transcriptome assemblies, and it is scalable to thousands of genomes (Emms and Kelly, 2015).

One parameter that can be manually chosen when running OrthoFinder is the inflation value for the Markov Clustering Algorithm (MCL) (van Dongen, 2000). For each inflation value (default parameter = 1.5), a different set of orthogroups is generated. Increasing the inflation parameter leads to higher precision at the cost of lower recall (i.e. more orthogroups with fewer sequences each) and decreasing the inflation value leads to lower precision, but higher recall (i.e. fewer, but larger orthogroups).

1.2.3.2 Multiple sequence alignment using MAFFT

Alignment of DNA/protein sequences is one of the core processes in comparative genomics. A multiple sequence alignment (MSA) is the mapping of the nucleotides/amino acids in one sequence onto the nucleotides/amino acids in another sequence, with gaps (representing insertions or deletions) introduced to optimise the number of matching positions (Hardison, 2003). It is implied that all input sequences on a MSA are homologous (i.e. derived from a single LCA) and aligned residues are usually interpreted to share an evolutionary origin and/or to be functionally related (Katoh et al., 2009). The information provided on sequence similarity is crucial for predicting function and for phylogenetic inference (Flicek et al., 2003; Woese and Fox, 1977).

MAFFT (Multiple Alignment using Fast Fourier Transform) is a similarity-based MSA software, first released in 2002 (Katoh et al., 2002), that has become increasingly popular due to its high performance, speed and the few computational resources required (Golubchik et al., 2007; Nuin et al., 2006). MAFFT has shown higher accuracy than previously available methods, including TCoffee and Clustal W, and is currently used in several highly-regarded projects, such as Pfam (Bateman et al., 2004; Katoh et al., 2005). For amino acid alignment, MAFFT uses the BLOSUM62 substitution matrix by default, suitable for aligning distantly related sequences (Eddy, 2004). In our analyses, MAFFT (v.7.307) was run with the E-ins-I algorithm, which means that all pairwise alignments were computed with a local alignment algorithm with generalised affine gap costs. The rationale behind this is that, in an MSA, conserved residues frequently fall into un-gapped blocks separated by relatively non-conserved regions and the generalised affined gap costs allow non-conserved regions to be effectively ignored (i.e. unaligned regions are left unaligned at the pairwise alignment stage) (Altschul, 1998). In order to solve incorrectly introduced gaps, MAFFT was run with the iterative refinement method (10000 cycles of iterative refinement), which means that the MSA was partitioned in two groups that were each realigned using a group-to-group alignment algorithm. This was repeated 10000 times or until no more improvements were possible (Katoh et al., 2009).

1.2.3.3 Analyses of specialised metabolite clusters using antiSMASH

Specialised metabolites (syn: secondary metabolites; "secondary" because they do not directly contribute to the organisms' growth or reproduction, but may confer a selective advantage to the organism) produced by prokaryotes are a rich and important source of antimicrobials and other bioactive compounds. Although these compounds have diverse chemical structures, their biosynthetic pathways, organised in biosynthetic gene clusters (BGCs), are often conserved (Blin et al., 2019). Usually, all the genes required for precursor biosynthesis, assembly and modification of the compound scaffold, as well as genes for resistance, export and regulation are clustered together in BGCs. Gene clustering (i.e. the proximity on the genome) is one of the most useful clues to gene function, especially in prokaryotes, where functionally related genes are often arranged in operons (de Crécy-Lagard and Hanson, 2013). With the rapid increase in sequenced genomes, fast and reliable *in silico* approaches for the identification and analysis of specialised metabolites are essential to complement chemical/analytical strategies.

AntiSMASH (Antibiotics and Secondary Metabolite Analysis Shell) (https://antismash.secondarymetabolites.org) is currently considered to be the "gold standard" tool for the identification, annotation and analysis of BCGs in bacterial and fungal genomes (Blin et al., 2021). Prior to antiSMASH, there were many available tools for the analysis of specialised metabolism: BAGEL3 (van Heel et al., 2013), CLUSEAN (Weber et al., 2009), ClustScan (Starcevic et al., 2008), NP.searcher (Li et al., 2009), NRPSPredictor2 (Rausch et al., 2005), PRISM (Skinnider et al., 2020), RiPPER (Santos-Aberturas et al., 2019), SBSPKS (Anand et al., 2010), SMURF (Khaldi et al., 2010) and TOUCAN (Almeida et al., 2020). However, these approaches were usually specific to the characterisation of particular classes of clusters. AntiSMASH provides a successful integration of the previously available tools, whilst ensuring high(er) accuracy in cluster annotation (Blin et al., 2021). The software uses many techniques for the prediction and annotation of BGCs, including the prediction of the chemical structures of NRPS/PKS products using profile hidden Markov models (pHMMs), BLAST, substrate specificity/active site searches, as well as manually curated BGC cluster rules, similar to those used by PRISM (Craig et al., 2011), SMURF (Bauer et al., 2015) and BAGEL (de Jong et al., 2010). AntiSMASH is also equipped with a tool for comparative gene cluster analysis, allowing for rapid inference of genes and operons based on homology, and with a database for specialised metabolism Clusters of Orthologous Groups (smCOGs),

which enables the prediction and categorisation of accessory gene function (Medema et al., 2011). There are a multitude of independent tools incorporating or utilising antiSMASH, including the antibiotic resistance target seeker ARTS (Mungan et al., 2020), the mass-spectronomy-guided peptide mining tool Pep2Path (Medema et al., 2014), the sgRNA design tool CRISPY-web (Blin et al., 2020) and the BGC classification and clustering platform BiG-SCAPE (Navarro-Muñoz et al., 2020).

1.2.3.4 Phylogeny reconstruction using IQ-Tree

Phylogenetic trees (syn: phylogenies) are diagrams representing the evolutionary history of different species or genes from a common ancestor (Fig. 1.3). In a phylogeny, nodes are connected by branches. Each branch corresponds to the persistence of a genetic lineage/gene, and each node represents the origin of a new lineage/gene (i.e. in the case of a lineage, this is called a speciation event) (Yang and Rannala, 2012). Phylogenetic studies can be used to help answer a myriad of different biological questions, stemming from gene prediction (Kellis et al., 2003) and reconstruction of ancestral genomes (Latysheva et al., 2012) to understanding the origin and spread of viral infection (Li et al., 2020) and migration patterns (Drummond et al., 2002). Various different methods can be used to generate phylogenies and which approach is best is still the subject of debate. The four main methods for inferring molecular phylogenies are pairwise distances (i.e. neighbour-joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA)), maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference methods (Whelan et al., 2001).

In character-based methods (i.e. MP, ML and Bayesian inference methods), all sequences in the alignment are simultaneously compared, considering one character (i.e. a site in the alignment) at a time so as to calculate a score for each tree (Yang and Rannala, 2012). The tree score corresponds to the minimum number of changes in MP methods, the log-likelihood value for ML approaches and the posterior probability for Bayesian inference (Whelan et al., 2001). In ML and Bayesian inference approaches a substitution model of evolution is used (i.e. model-based methods). Nowadays, ML and Bayesian-based inferential techniques are believed to be the most powerful and robust for phylogeny reconstruction, especially when inferring deep phylogenies (Yang and Rannala, 2012). ML approaches have been shown to outperform distance and parsimony methods over a wide range of realistic conditions (Gaut and Lewis, 1995; Guindon and Gascuel, 2003; Huelsenbeck, 1995; Kuhner and Felsenstein,

1994). They also allow for robust statistical assessment and provide the means for comparing competing trees and models (Whelan et al., 2001). One drawback of ML inference is the fact that, except for very small data sets, there are an immense number of possible trees and an exhaustive search is often a computational bottleneck (Felsenstein, 1978). To circumvent this issue, heuristic searches (i.e. "best guess") are often used. These approaches usually employ a "hill climbing" technique, where a starting tree is generated using a fast algorithm and then local rearrangements are performed to improve the tree score. Heuristic searches provide no guarantee of finding the optimal tree (i.e. the global optimum/the summit) (Swofford, 1996).

IQ-TREE is currently considered the "gold-standard" software for ML phylogeny inference, often performing better than (i.e. achieving higher likelihoods) RAxML (Stamatakis, 2006) and PhyML (Guindon et al., 2010), in a time-efficient manner (Nguyen et al., 2014). IQ-TREE employs a stochastic algorithm, developed to attempt to escape local optima trees, which is a recurrent problem in hill-climbing approaches. The success of IQ-TREE stems mainly from two factors: (1) combining elements of hill-climbing algorithms, random perturbation of current best trees (stochastic nearest neighbour interchange (NNI)), and a broad sampling of initial starting trees (i.e. a small population of locally optimal trees) result in high(er) likelihood and (2) the implementation of a phylogenetic likelihood library (Flouri et al., 2015) – an optimised application programming interface that allow users to quickly implement likelihood calculations, model parameter, branch length optimisation and tree space exploration – helps reduce time for likelihood computation (Nguyen et al., 2014). IQ-TREE also incorporates ModelFinder (Kalyaanamoorthy et al., 2017), a model-selection method that combines substitution models with a flexible rate heterogeneity across sites (RHAS) model, which allows users to identify the best-fitting model of sequence evolution (i.e. the best-fitting substitution model and the best-fitting RHAS model) that led to the data. ModelFinder is often fast(er), more flexible and accurate than other available model-selection methods (Darriba et al., 2011, 2012; Lanfear et al., 2012).

Assessing phylogenetic uncertainty remains a challenging statistical problem. Non-parametric bootstrapping is a widely used technique for estimating the phylogenetic support of certain clades/splits and it is crucial for assessing the robustness of a tree (Felsenstein, 1985; Swofford, 1996). It consists of a numerical re-sampling approach where a subset of a multiple sequence alignment is generated (i.e. pseudo-replicates) from the data by random sampling with replacement, repeated several times (usually 1000 replicates) and compared

(Sleator, 2011). Bootstrap values, usually represented in percentages within each branch, are then placed on the reconstructed tree (Fig. 1.3). Bootstrap values of > 70% are thought to represent "true" clades in phylogenetic trees (Hillis and Bull, 1993; Hillis et al., 1992). In this work, ultrafast bootstrap (UFBoot2, implemented in the IQ-TREE software (Hoang et al., 2017)) was used when reconstructing phylogenies as it is less computationally intensive, is relatively robust against moderate model violations and has been proved to achieve more unbiased branch support than standard bootstrap (Efron, 1992; Felsenstein, 1985).

The most common and widely accepted reconstructions of cyanobacterial phylogenies are those based on comparisons of the 16S small subunit ribosomal RNA gene (16S rRNA) sequences (Bernroitner et al., 2009; Garcia-Pichel et al., 1998; Giovannoni et al., 1988; Gugger and Hoffmann, 2004; Honda et al., 1999; Ishida et al., 2001; Iteman et al., 2000; Litvaitis, 2002; Moore et al., 2019; Neilan et al., 1997; Turner et al., 1999). Most of these cyanobacterial phylogenies were reconstructed using distance-based methods (NJ approaches) and comprised a relatively small number of species. The accurate structure of cyanobacterial phylogenies cannot be easily resolved as most of their diversity originates in an explosive radiation that took place early during evolutionary history (Garcia-Pichel, 2009). Given the divergence between the current taxonomic treatment of cyanobacteria and the natural system that reflects their evolutionary relationships, ecology and physiology are important parameters in understanding cyanobacterial phylogenetic relationships (Garcia-Pichel, 2009).

Although the most prominent traits of cyanobacterial ribosomal phylogeny have found support when other genes or multi-locus analyses were performed, conflicting results of phylogenies based only on 16S rRNA sequences suggest that trees based on a single gene are not sufficient to resolve relationships among very closely or very distantly related organisms (Casamatta et al., 2005; Fox et al., 1992). With the rapid increase in sequenced genomes, where most/all genes from a multitude of species is available, it is now possible to reconstruct phylogenetic trees from concatenated methods, thought to converge faster on the true topology (Gadagkar et al., 2005) (Fig. 1.3). Combined analyses taking into account several single gene data sets concatenated into one larger data set have rarely been applied to cyanobacterial phylogenies (Kauff and Büdel, 2010).

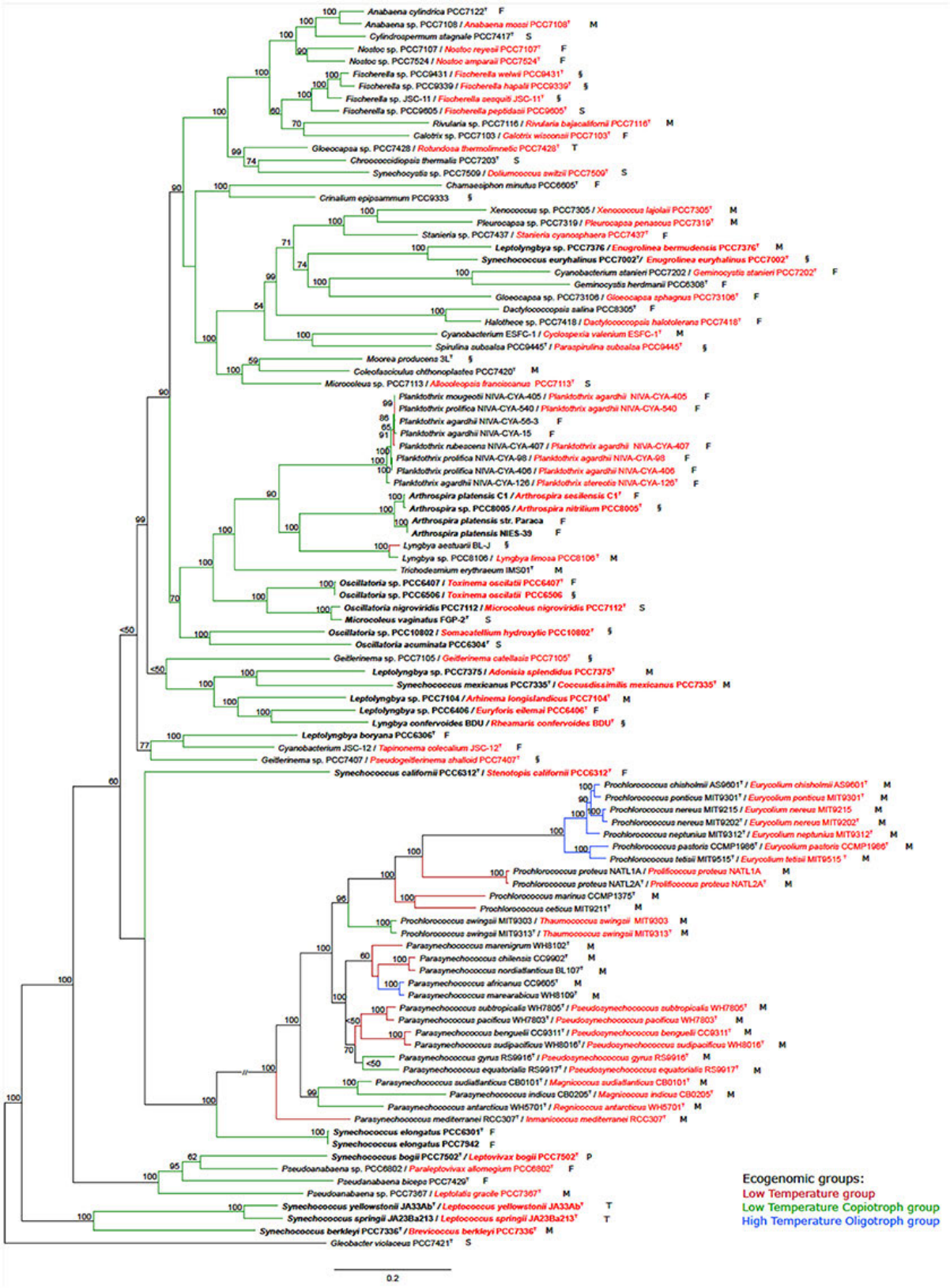


Figure 1. 3. State-of-the-art cyanobacterial species phylogeny prior to our phylogeny reconstruction. The phylogeny from Walter et al. (2017) was estimated using a similar approach with AMPHORA2 and ML (using MEGA) for 100 genomes. The placement of well-described cyanobacterial species/orders is comparable to our phylogeny.

Whilst most phylogenetic analysis to date only consider the vertical mode of inheritance (parent – offspring), recent evidence shows a significant degree of horizontal gene transfer (HGT) events, especially in prokaryotic species (i.e. in certain species the amount of HGT is comparable or higher than the rate of spontaneous mutation (Lawrence, 2002)), and must be taken into account when constructing species phylogenies. There are ways to address and potentially minimise the detrimental effects of HGT on phylogenetic inference. Ziying Ke used AMPHORA's protein phylogenetic marker database (Wu and Eisen, 2008), which consists on a set of curated protein sequence alignments and corresponding profile hidden Markov models (pHMMs, see below) from representatives of complete bacterial genomes. These markers are universally distributed in bacteria, are usually present as a single copy within the genome, and are believed to be somewhat resistant to HGT (i.e. are thought to not have been subject to extensive HGT). Due to the substantial influence of HGT on prokaryotic evolution, phylogenetic networks, which consider genetic exchanges between species (instead of phylogenetic trees, which only consider vertical descent) have been used to more correctly attempt to depict evolutionary relationships (Kunin et al., 2005).

1.2.3.5 Profile hidden Markov models using HMMER

Functionally related proteins frequently share common patterns of amino acids. Different residues within a sequence are subject to different selective pressures (i.e. some positions may be conserved; some may tolerate certain substitutions whilst conserving physiochemical properties, such as hydrophobicity, charge or size; and others may be evolutionary neutral and/or variable) and multiple sequence alignments can be used to understand conservation patterns.

Profile HMMs (pHMMs) are statistical models that convert information from multiple sequence alignments into a set of probability values that reflect position-specific variation in evolutionary related sequences (Eddy, 1998; Reyes et al., 2017). pHMMs are designed to represent multiple sequence alignments and they characterise precisely each position of a sequence, including matches, insertions and deletions (Fig. 1.4). Since pHMMs model the diversity of a set of homologs, they show higher sensitivity than position-independent (i.e. position invariant for both similarity and gap evaluation) similarity-based methods (i.e. BLAST (Altschul et al., 1990)) for the detection of remote (i.e. distant) homologs.

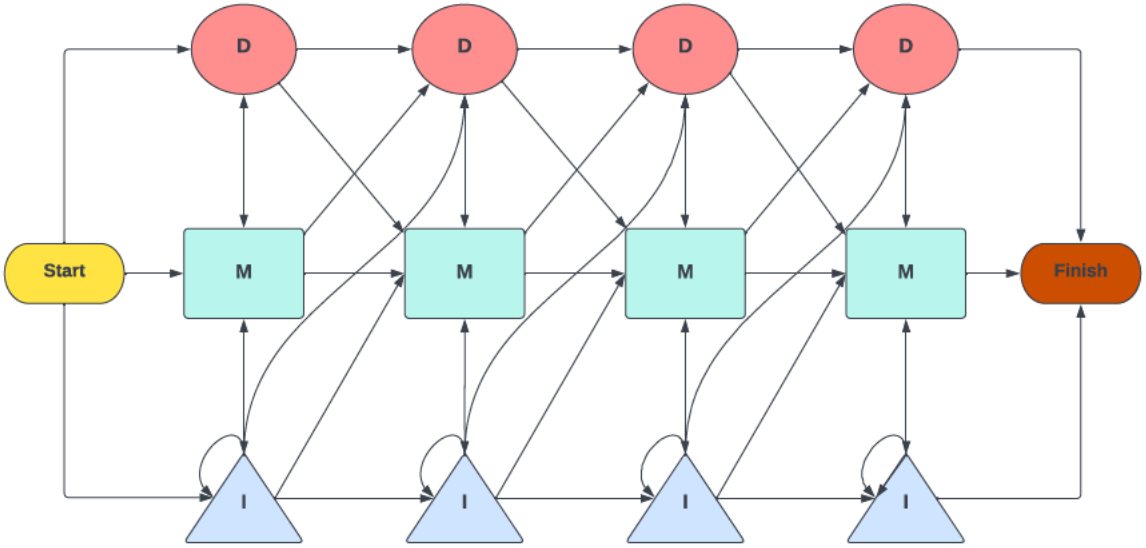


Figure 1.4. Diagram representing the structure of a profile hidden Markov model (pHMM). The model has a beginning and an end. The rectangles correspond to match states (M), the circles correspond to deletion states (D) and the triangles correspond to insertion states (I). Insertion and deletion states mean that sequences of different lengths can be used in generating the model, as well as testing it. The arrows between states represent transition probabilities.

Biological applications of pHMMs include modelling the characteristics of protein families, protein prediction/classification and motif detection (Yoon, 2009). Although many pHMM tools are available (SAM (Hughey and Krogh, 1996), PFTOOLS (Bucher et al., 1996), HMMpro (Baldi et al., 1994)), the most commonly used in computational biology is HMMER (hmmer.org). PFAM (Mistry et al., 2021), PROSITE (Hulo et al., 2006) and TIGRFAM (Haft et al., 2003) are some of the currently available large collections of annotated pHMMs. PFAM and TIGRFAM use HMMER, whilst PROSITE used PFTOOLS.

1.2.3.6 Inference of horizontal gene transfer using Notung

Horizontal gene transfer (HGT, also known as lateral gene transfer, LGT) is the acquisition and incorporation of foreign genes by a species and can occur through three main mechanisms: conjugation, transformation and transduction (Griffith, 1928; Lederberg and Tatum, 1946; Zinder and Lederberg, 1952). It is now known that HGT is major force in prokaryotic evolution (Koonin et al., 2001), including cyanobacteria (Garcia-Pichel, 2009). The mechanisms and frequency of HGT events are largely unclear, however they are

believed to be dependent on specific cellular systems/factors, namely gene content, donor and recipient species, cell physiology and environmental conditions (Hardiman et al. 2016; Sheppard et al., 2020).

HGT has been shown to occur frequently across cyanobacterial evolutionary history. In some species, it may have contributed to the acquisition of advantageous traits, enabled niche adaptation and led to the emergence of new phenotypes. In cyanobacterial genomes, between 9.5% and 16.6% of genes are thought to have been acquired through HGT (e.g. 11% in *Synechocystis*) (Ochman et al., 2000; Nakamura et al., 2004). There are many examples of HGT events reported in cyanobacteria, including (i) the transfer of the *nif* operon to *Microcoleus chthonoplastes* from *Desulfovibrio* (Bolhuis et al., 2010), (ii) the transfer of citrullinating enzymes from cyanobacteria into animals (Cummings et al., 2021), (iii) the transfer of ATP sulfurylases and class I fructose biphosphate adolases to cyanobacteria from eukaryotic green and red algae (Godde et al., 2018), (iv) the transfer of components of the eukaryotic actin cytoskeleton – actin and profilin – into *Microcystis* (Guljamow et al., 2007), and (v) the transfer of the cluster coding for the pigment phycoerythrin in the filamentous cyanobacteria *Planktothrix* from closely-related strains (Tooming-Klunderud et al., 2013). All mechanisms of HGT (i.e. transformation, conjugation and transduction) are known in cyanobacteria. Natural transformation has been documented in some cyanobacterial species, including *Synechococcus* sp. PCC 7942 (Nakasugi et al., 2006), *Synechocystis* (Grigorieva and Shestakov, 1982; Stevens-Jr and Porter, 1986), *Anabaena* sp. PCC 7120 (Trehan and Sinha, 1981) and *Thermosynechococcus elongatus* BP-1 (Iwai et all., 2004; Onai et al., 2004). Although no special requirements are known to induce competency in transformable cyanobacteria, light has been shown to positively influence transformation in *Synechococcus* and *Synechocystis* spp. (Morrison et al., 2005). Despite less documented, conjugation is known in several *Anabaena* strain (Muro-Pastor et al., 1994; Wolk et al., 1984). Cyanophages are frequently encountered in marine cyanobacteria and are known to possess several important photosynthesis genes, as well as other genes from both archaea and bacteria. Gene transfer and recombination in marine cyanobacteria is believed to be mediated by these cyanophages (Zeidner et al., 2005). The importance of transduction in freshwater species is still unclear.

There are three different strategies to identify and quantify HGT events: *in vitro, in vivo* and *in-silico* approaches. *In vitro* analyses include laboratory experiments conducted under highly

controlled conditions and rely on quantitative techniques such as microfluidics (Lopatkin et al., 2016), qPCR (Wan et al., 2011) and flow cytometry (Loftie-Eaton et al., 2014). *In vivo* approaches often consist of experiments using well-characterised strains/representative bacterial communities in native/natural environments (Klümper et al., 2014; Li et al., 2018). *In-silico* techniques, used in this work, are often based on analyses of genomic sequence data and focus on computational techniques to identify HGT events and how they affect population structure or phylogenies. *In-silico* methods are largely based on the fact that when a gene/protein sequence from a particular organism has a strong similarity to a homolog from a distant organism, then it is indicative of an HGT event. These approaches can be broadly separated into two groups: parametric and non-parametric (i.e. phylogenetic) (Ravenhall et al., 2015). Parametric approaches include strategies that attempt to find genes that significantly differ from the genomic average (i.e. deviant GC content, trinucleotide analysis, or codon usage) (Lawrence and Ochman, 2002). Non-parametric methods analyse incongruences between gene and species tree and flag discrepancies as potential gene duplication, gene loss or HGT events (Ravenhall et al., 2015) (Figure 1.5). Phylogenetic inference is currently believed to be the best way to detect HGT events (Husnik and McCutcheon, 2018; Syvanen, 1994). These analyses also allow the inference of the donor and recipient of a particular transfer event, as well as the approximate time of transfer (Yerrapragada et al., 2009).

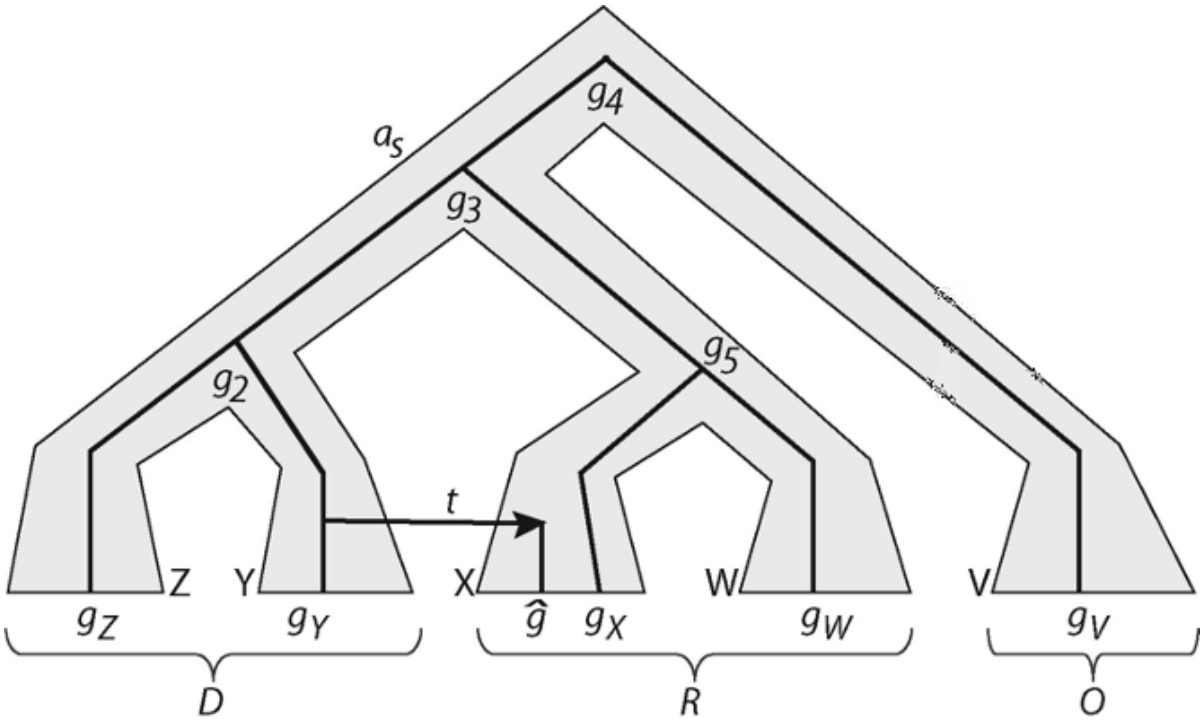


Figure 1.5. Diagram representing a gene phylogeny with one horizontal gene transfer (HGT). The transfer (*t*) has a donor in species D and a recipient in species R. Figure from Darby et al. (2017).

I used the NOTUNG software for HGT inference (Stolzer et al., 2012). NOTUNG is the first parsimony-based reconciliation algorithm for a duplication-transfer-loss (DTL) event model. Some important features that differentiate NOTUNG from other reconciliation programs (reviewed in Doyon et al., 2011) are (1) the fact that it reports on all optimal event histories that are temporally feasible (i.e. transfers from ancestral nodes to the nodes that descended from them; transfers in the opposite direction - from descendants to ancestors - are classified as temporally infeasible), (2) it places no restriction on speciation times, (3) it does not require a species tree with branch lengths, (4) it allows for the rearrangement of gene trees with weakly-supported edges (i.e. poorly supported nodes in the gene phylogeny can be rearranged to match the species phylogeny's topology) and (5) it allows for the resolution of non-binary gene trees. NOTUNG also comprises a phylogenomic analysis function that aggregates and summarises results across all reconciled gene trees and all nodes/branches of the species tree, providing useful information on ancestral gene content, HGT traffic between species and the timings of genome expansion/contractions (Stolzer et al., 2012).

1.3 Thesis outline

In this thesis, I present the results of several comparative genomic analyses of 130 cyanobacterial species. Through the use of computational techniques, such as phylogeny reconstruction, multiple sequence alignments, profile hidden Markov models (pHMMs), sequence similarity searches, specialised metabolite cluster identification and characterisation, and active site specification, I report on cyanobacterial species capable of neurotoxin production, offering plausible physiological explanations for their biosynthesis. I present my work on the development of a pipeline for finding horizontal gene transfer (HGT) events in cyanobacterial species using the NOTUNG software. This data allowed my scientific contribution to the current debate on whether or not the occurrence of HGT events is associated with the environmental context experienced by a species. The work presented here substantially improves our understanding of 2,4-DAB and BMAA production in cyanobacteria and the extent to which HGT events have impacted cyanobacterial evolution. My work comprises an important methodological resource for large scale *in-silico* cyanotoxin discovery and HGT inference, which could be applied to species other than cyanobacteria, and improves the current understanding of cyanobacterial diversity, biochemistry and physiology.

1.4 Research questions

1	Chapters 2 & 3	Given the known pathways for the biosynthesis of 2,4-DAB, BMAA and their precursors in other species, is the production of these neurotoxins possible in cyanobacteria?
2	Chapters 2 & 3	Are there cyanobacterial species that encode all enzymes in a known pathway for the biosynthesis of 2,4-DAB and/or BMAA?
3	Chapters 2 & 3	Are key enzymes in pathways leading to 2,4-DAB and/or BMAA co-localised within specialised metabolite clusters?
4	Chapters 2 & 3	Which cyanobacterial species are potentially capable of biosynthesising 2,4-DAB and/or BMAA?
5	Chapters 2 & 3	Are the neurotoxins 2,4-DAB and BMAA widespread in cyanobacteria?
6	Chapters 2 & 3	Is there a phylogenetic, biochemical, physiological or morphological relationship between 2,4-DAB and/or BMAA producers?
7	Chapters 2 & 3	Can a physiological role be predicted for the biosynthesis of 2,4-DAB and/or BMAA in cyanobacteria? (i.e. why do cyanobacteria produce these neurotoxins?)
8	Chapter 3	Do cyanobacterial species synthesise a bound form of BMAA via the *pam* gene cluster from *Paenibacillus larvae*?
9	Chapter 3	Is there an explanation for the fact that the precursor of BMAA – 2,3-DAP – has never been detected in analytical studies to date?
10	Chapter 4	Is HGT common in cyanobacteria?
11	Chapter 4	Is HGT in cyanobacteria more common between closely-related species or species that have diverged a long time ago?
12	Chapter 4	Is the prevalence of HGT in cyanobacteria associated with the environment?
13	Chapter 4	Is HGT more prevalent in extreme environments, as opposed to more benign habitats?
14	Chapter 4	Is HGT more prevalent in terrestrial environments, as opposed to aquatic niches?

1.5 Chapter summaries

1.5.1 Chapter 2

Cyanobacteria can present health hazards to humans and animals due to the production of a wide range of toxins (cyanotoxins), which include the sometimes co-occurring, non-encoded diaminoacid neurotoxins 2,4-diaminobutanoic acid (2,4-DAB) and its structural analogue β-*N*-methylaminoalanine (BMAA). Knowledge of the biosynthetic pathway for 2,4-DAB, and its role in cyanobacteria, is lacking. The aspartate 4-phosphate pathway is a known route of 2,4-DAB biosynthesis in other bacteria and in some plant species. Another pathway to 2,4-DAB has been described in *Lathyrus* species. I used bioinformatics analyses to investigate hypotheses concerning 2,4-DAB biosynthesis in cyanobacteria. I assessed the presence or absence of each enzyme in candidate biosynthesis routes, the aspartate 4-phosphate pathway and a pathway to 2,4-DAB derived from S-adenosyl-L-methionine (SAM), in 130 cyanobacterial genomes using sequence alignment, profile hidden Markov models, substrate specificity/active site identification and the reconstruction of gene phylogenies. In the aspartate 4-phosphate pathway, for the 18 species encoding diaminobutanoate-2-oxo-glutarate transaminase, the co-localisation of genes encoding the transaminase with the downstream decarboxylase or ectoine synthase - often within hybrid non-ribosomal peptide synthetase (NRPS)-polyketide synthases (PKS) clusters, NRPS-independent siderophore (NIS) clusters and incomplete ectoine clusters - is compatible with the hypothesis that some cyanobacteria use the aspartate 4-phosphate pathway for 2,4-DAB production. Through this route, in cyanobacteria, 2,4-DAB may be functionally associated with environmental iron-scavenging, via the production of siderophores of the schizokinen/synechobactin type and of some polyamines. In the pathway to 2,4-DAB derived from SAM, eight cyanobacterial species encode homologs of SAM-dependent 3-amino-3-carboxypropyl transferases. Other enzymes in this pathway have not yet been purified or sequenced. Ultimately, the biosynthesis of 2,4-DAB appears to be either restricted to some cyanobacterial species, or there may be multiple and additional routes, and roles, for the synthesis of this neurotoxin.

1.5.2 Chapter 3

Cyanobacterial toxins (cyanotoxins) include the diaminoacid neurotoxin 3-*N*-methyl-2,3-diaminopropanoic acid (β-*N*-methylaminoalanine, BMAA). Knowledge of the biosynthetic pathway for BMAA, and its role in cyanobacteria, is lacking. Present evidence suggests that BMAA is derived by 3-*N* methylation of 2,3-diaminopropanoic acid (2,3-DAP) and, although the latter has never been reported in cyanobacteria, there are multiple pathways to its biosynthesis known in other bacteria and in plants. I used bioinformatics analyses to investigate hypotheses concerning 2,3-DAP and BMAA biosynthesis in cyanobacteria. I assessed the potential presence or absence of each enzyme in candidate biosynthetic routes known in *Albizia julibrissin*, *Lathyrus sativus* seedlings, *Streptomyces*, *Clostridium*, *Staphylococcus aureus*, *Pantoea agglomerans*, and *Paenibacillus larvae*, in 130 cyanobacterial genomes using similar methodology to that described in Chapter 2 (i.e. sequence alignment, profile hidden Markov models, substrate specificity/active site identification and the reconstruction of gene phylogenies). Most enzymes involved in pathways leading to 2,3-DAP in other species were not found in the cyanobacteria analysed. Nevertheless, two species appear to have the genes *sbnA* and *sbnB*, responsible for forming the 2,3-DAP constituent in staphyloferrin B, a siderophore from *Staphylococcus aureus*. It is currently undetermined whether these species are also capable of biosynthesising BMAA. It is possible that, in some cyanobacteria, the formation of 2,3-DAP and/or BMAA is associated with environmental iron-scavenging. The *pam* gene cluster, responsible for the biosynthesis of the BMAA-containing peptide, paenilamicin, so far appears to be restricted to *Paenibacillus larvae*. It was not detected in any of the cyanobacterial genomes analysed, nor was it found in 93 other *Paenibacillus* genomes or in the genomes of two BMAA-producing diatom species. I hypothesise that the presence, in some cyanobacterial species, of the enzymes 2,3-diaminopropionate ammonia-lyase (DAPAL) and reactive intermediate deaminase A (RidA) may explain the failure to detect 2,3-DAP in analytical studies. Overall, the taxonomic distribution of 2,3-DAP and BMAA in cyanobacteria is unclear; there may be multiple and additional routes, and roles, for the biosynthesis of 2,3-DAP and BMAA in these organisms.

1.5.3 Chapter 4

Horizontal gene transfer (HGT) plays an important role in the evolution of prokaryotes, including cyanobacteria. A key unresolved question is, the extent to which HGT is associated with environmental conditions. It has been proposed that an unusually high prevalence of HGT may occur between prokaryotes in physicochemically extreme environments, although this is not supported by all studies. It has also been suggested that soil may be an extreme environment from the organism's point of view, hence a terrestrial environment may be associated with high HGT. We investigate these hypotheses using reconciliation of gene family phylogenies with the species phylogeny for 130 cyanobacteria, followed by phylogenetically corrected statistical comparisons across species. We find no association between the prevalence of HGT and either an extreme or a terrestrial environment. Future studies could benefit from more detailed environmental information (e.g. free-living *versus* biofilm; bulk soil *versus* rhizophere) and optimal growth conditions (e.g. temperature and pH).

1.5.4 Chapter 5

I summarise the main findings of my work, highlight the limitations of my studies and present thoughts on future directions.

Chapter 2

Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 2,4-diaminobutanoic acid (2,4-DAB)

2.1 Preface

The work in this chapter has been published as a manuscript in *Phytochemistry* and the first-person plural is used throughout to maintain consistency. Minor changes have been made to the published version to preserve formatting across the thesis. I performed all analyses, wrote the first draft of the manuscript and supplementary materials, and produced table 2.1 and figures 2.7-2.9.

Peter B. Nunn produced all figures depicting biochemical pathways/reactions and molecules (figures 2.1-2.6). Ziying Ke performed the reconstruction of the cyanobacterial species phylogeny and orthogroup prediction.

Citation:

Mantas, M. J. Q., Nunn, P. B., Ke, Z., Codd, G. A., Barker, D., 2021. Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 2,4-diaminobutanoic acid (2,4-DAB). *Phytochemistry* 192, 112953.

2.2 Graphical abstract

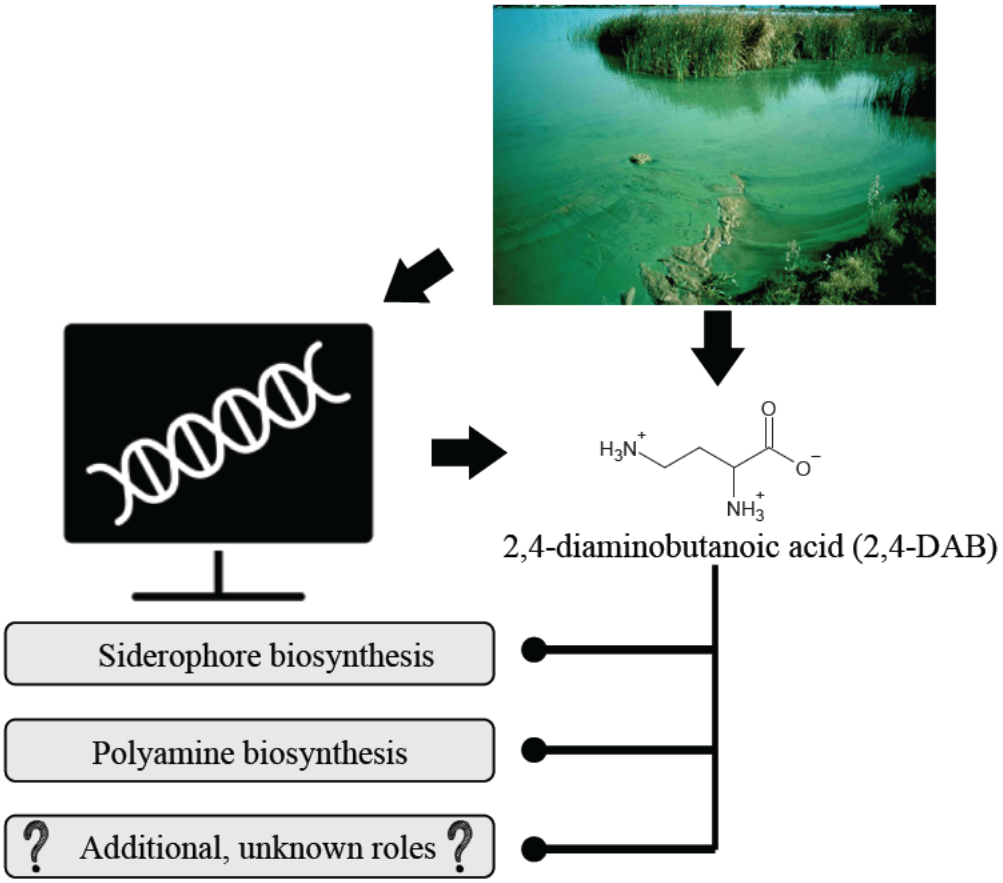


Photo: Geoffrey A. Codd; used with permission. Freshwater lake-side scum of toxin-producing cyanobacteria (*Nodularia spumigena*), South Australia.

2.3 Introduction

The non-encoded diaminoacid neurotoxin 2,4-diaminobutanoic acid (2,4-DAB) (syn: 2,4-diaminobutyric acid; α,γ-diaminobutyric acid; γ-aminobutyrine) (Fig. 2.1) was first discovered in acid hydrolysates of polymyxin antibiotics (Catch and Jones, 1948). It is widely distributed and occurs in the cell wall of Gram-positive bacteria (Perkins and Cummins, 1964), in the seeds of leguminous plants (Bell, 1962; Nigam and Ressler, 1966) and as a homopolymer in *Streptomyces celluloflavus* USE32 and *Streptoalloteichus hindustanus* NBRC15115 (Yamanaka et al., 2020). Both 2,4-DAB and its structural analogue β-*N*-methylaminoalanine (BMAA) are neurotoxic (Chen et al., 1972; Tan et al., 2018). These neurotoxins have also been found to be co-distributed in some cyanobacterial species (Al-Sammak et al., 2014; Rosén and Hellenäs, 2008), but may also occur alone (Krüger et al., 2010; Violi et al., 2019).

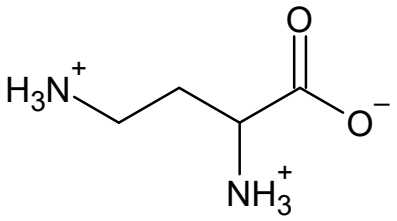


Figure 2.1. Structure of 2,4-diaminobutanoic acid in the ionised form (2,4-diaminobutanoate). This form is present at physiological pH values. The carboxyl group is then completely ionised and positive charge is shared between the two amino groups. Figure: Peter B. Nunn; used with permission.

The biosynthetic pathway to 2,4-DAB is well documented in other bacteria (Yamamoto et al., 1992; Yamanaka et al., 2020) and in the legume *Lathyrus sylvestris* (Nigam and Ressler, 1966), but it has not been established in cyanobacteria. Also, the extent of 2,4-DAB biosynthesis among cyanobacteria has not been widely determined.

There are at least four factors contributing to the lack of knowledge concerning the presence of 2,4-DAB in cyanobacteria: (i) the biosynthetic pathway(s) in cyanobacteria are unknown; (ii) some early analytical techniques to analyse cyanobacterial extracts for the presence of 2,4-DAB and its analogues have lacked sensitivity and specificity (Cohen, 2012; Faassen et al., 2016; Jiang et al., 2013; Lage et al., 2016; Rosén et al., 2016); (iii) whilst adequate

analytical methods can unambiguously assign 2,4-DAB biosynthesis to a monocyanobacterial axenic culture, ambiguity can arise regarding toxin origin when non-axenic cyanobacterial environmental samples and laboratory cultures, and mixtures of cyanobacterial species/strains are examined; (iv) the facultative expression of 2,4-DAB biosynthesis means that lack of evidence for this toxin is not evidence of its absence under all conditions. Some of these problems are beginning to be resolved. For example, although a consensual standard method of sample preparation and analytical techniques for the identification of 2,4-DAB and BMAA has been lacking, it is now widely agreed that tandem mass spectrometry (MS/MS) is the best currently available tool for the correct identification and quantification of these neurotoxins, as it relies on four identification criteria, whilst resolving structural isomers, improving detection limits and reducing background interference (Faassen et al. 2016; Faassen, 2017; Metcalf et al., 2017).

The environmental and toxicological monitoring of cyanobacterial mass populations, for example by microscopy and by chemical, immunological, toxicological and remote sensing methods (Meriluoto et al., 2017), is being successfully complemented by molecular genetic methods (Kurmayer et al., 2017; Pan et al., 2002). The latter include the targeting of genes involved in the biosynthesis of specific cyanotoxins by polymerase chain reaction (PCR)-based methods (Humbert, 2017; Kurmayer et al., 2017). The contribution of molecular genetic methods to the early detection of cyanobacteria with the potential for the production of 2,4-DAB and its isomers, would similarly contribute to the overall risk management of the production of these neurotoxins. To enable this, knowledge of the biosynthetic pathway(s) for 2,4-DAB formation, with gene identification is necessary. These needs are addressed here by our bioinformatics study to test hypotheses concerning the biosynthetic pathway or pathways for 2,4-DAB in cyanobacteria.

As suggested by Nunn and Codd (2017), a possible route for the biosynthesis of 2,4-DAB is through the aspartate 4-phosphate pathway (Fig. 2.2). This pathway is utilised by bacteria, fungi and higher plants to produce four encoded amino acids; methionine, threonine, isoleucine and lysine, at various branchpoints (Jander and Joshi, 2010). The pathway starts at aspartate transaminase (EC 2.6.1.1), which catalyses a reversible transamination between glutamate and oxaloacetate to yield aspartate and 2-oxo-glutarate (Jansonius and Vincent, 1987). Aspartate then serves as substrate for aspartate kinase (EC 2.7.2.4), which catalyses the formation of aspartate 4-phosphate. Aspartate 4-semialdehyde is then synthesised from

aspartate 4-phosphate via the action of aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (Jander and Joshi, 2010). These three enzymes are widespread in prokaryotes and are essential for their viability (Cohen, 1987; Muriana et al., 1991).

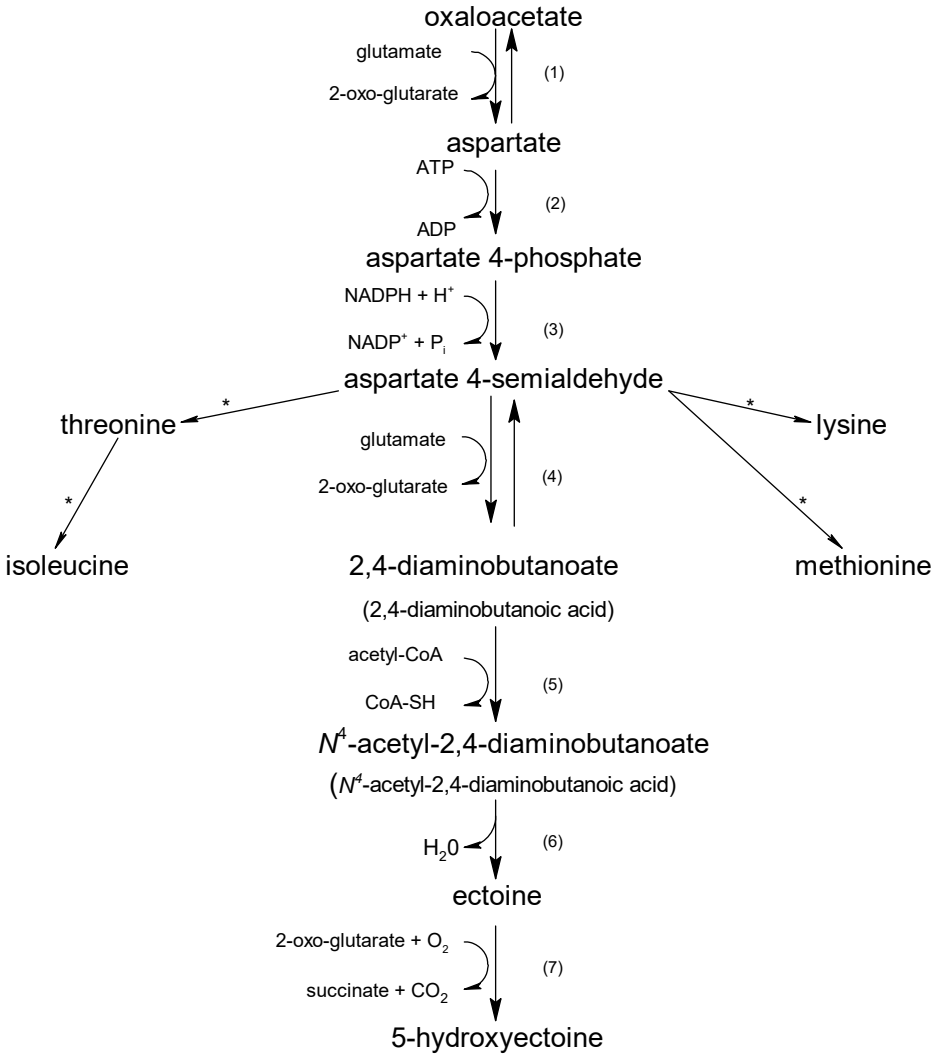


Figure 2.2. The aspartate 4-phosphate pathway. Established routes to 2,4-diaminobutanoate (2,4-diaminobutanoic acid) and derivatives. The nomenclature is from **www.brenda-enzymes.org**; the Enzyme Commission numbers are as follows: (1) aspartate transaminase: EC 2.6.1.1; (2) aspartate kinase: EC 2.7.2.4; (3) aspartate-semialdehyde dehydrogenase: EC 1.2.1.11; (4) diaminobutanoate-2-oxo-glutarate transaminase: EC 2.6.1.76; (5) diaminobutanoate acetyltransferase: EC 2.3.1.178; (6) ectoine synthase: EC 4.2.1.108; and (7) ectoine hydroxylase: EC 1.14.11.55. Multiple-step pathways are indicated by *. Figure: Peter B. Nunn; used with permission.

2.3.1 The aspartate 4-phosphate pathway and the biosynthesis of 2,4-DAB

An additional branch-point from this pathway at aspartate 4-semialdehyde begins with diaminobutanoate-2-oxo-glutarate transaminase (EC 2.6.1.76), which catalyses the biosynthesis of 2,4-DAB. 2,4-DAB has been described as a precursor for the biosynthesis of 1,3-diaminopropane (Ikai and Yamamoto, 1997; Yamamoto et al., 1992) (Fig. 2.3), and the compatible solutes ectoine and hydroxyectoine (Louis and Galinski, 1997; Reuter et al., 2010) (Fig. 2.2). In *Acinetobacter baumannii*, 2,4-DAB is also biosynthesised by the activity of diaminobutanoate-2-oxo-glutarate transaminase, utilising aspartate 4-semialdehyde as substrate (Ikai and Yamamoto, 1997). This diaminoacid is subject to decarboxylation by the activity of 2,4-DAB decarboxylase (EC 4.1.1.86) to yield 1,3-diaminopropane (Yamamoto et al., 2000; Yamamoto et al., 1992). Transaminases can exhibit multiple substrate specificity, and the substrate for the 2,4-DAB transaminase may differ. In a species of *Xanthomonas*, 2,4-DAB has been shown to be synthesised by a transaminase for 2,4-DAB, diaminobutanoate-pyruvate transaminase (EC 2.6.1.46), which uses pyruvate, instead of 2-oxo-glutarate, as the oxo-acid (Rao et al., 1969). In organisms that produce the hydrogenated pyrimidine derivative ectoine, an *ect* operon is usually found, encompassing the genes *ectB* (encoding diaminobutanoate-2-oxo-glutarate transaminase), *ectA* (encoding 2,4-diaminobutanoate acetyltransferase, EC 2.3.1.178) and *ectC* (encoding ectoine synthase, EC 4.2.1.108). EctC can be regarded as a marker enzyme for ectoine-producers (Widderich et al., 2014a). The route to ectoine biosynthesis was elucidated by Peters et al. (1990) in cell-free extracts of *Ectothiorhodospira halochloris* and *Halomonas elongata*. It begins with the transamination of aspartate 4-semialdehyde, catalysed by diaminobutanoate-2-oxo-glutarate transaminase (enzyme 4, Fig. 2.2), to yield 2,4-DAB, which serves as substrate for diaminobutanoate acetyltransferase (enzyme 5, Fig. 2.2) to produce N^4-acetyl-2,4-diaminobutanoic acid. Ectoine synthase (enzyme 6, Fig. 2.2) catalyses the ring closure reaction, leading to the end product, ectoine. A fraction of the organisms that encode the ectoine operon are capable of producing 5-hydroxyectoine through a position- and stereo-specific hydroxylation of ectoine, via the activity of ectoine hydroxylase (*ectD,* EC 1.14.11.55) (Bursy et al., 2007). Although compatible solutes are known in many cyanobacteria (Klähn and Hagemann, 2011), ectoine and hydroxyectoine have not been identified in these organisms (Nunn and Codd, 2017). Until this present investigation, no

genetic investigations have attempted to establish whether the biosynthesis of 2,4-DAB in cyanobacteria occurs through the aspartate 4-phosphate pathway.

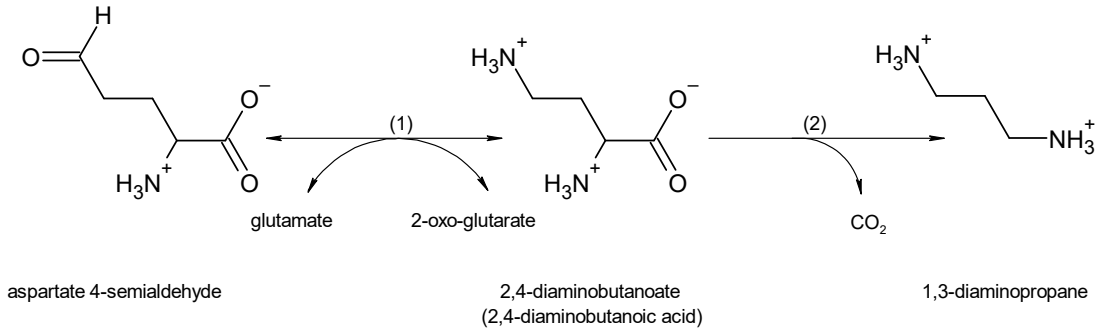


Figure 2.3. The synthesis of 1,3-diaminopropane from 2,4-DAB. Established route to 1,3-diaminopropane. The nomenclature is from www.brenda-enzymes.org; the Enzyme Commission numbers are as follows: (1) diaminobutanoate-2-oxo-glutarate transaminase: EC 2.6.1.76; (2) diaminobutanoate decarboxylase: EC 4.1.1.86. Figure: Peter B. Nunn; used with permission.

In prokaryotes, toxicity-related genes are often found within specialised metabolite clusters (secondary metabolite clusters), which are synthesised by a process external to the default ribosomal system, via a non-ribosomal synthetic mechanism (NRPS) (Jones et al., 2010), which involves a multienzyme complex (Miller and Gulick, 2016). Compounds that, unlike peptides, do not contain nitrogen, may be biosynthesised by similar enzyme assemblies, named polyketide synthetases (PKS) (Robbins et al., 2016), and others can be synthesised through hybrid NRPS/PKS complexes (Fisch, 2013). Many classes of cyanotoxins, including cylindrospermopsins, saxitoxins, nodularins and microcystins are known to be produced via these routes (Kellmann et al., 2008; Mbedi et al., 2005; Mihali et al., 2008; Moffitt and Neilan, 2004). A third and less widespread category of multienzyme complexes exists in cyanobacteria, namely the NRPS-independent siderophore biosynthesis systems (NIS). NIS clusters are responsible for the production of siderophores, including rhizobactin 1021, schizokinen and synechobactin (Challis, 2005; Årstøl and Hohmann-Marriott, 2019).

2.3.2 Siderophores

Siderophores are low molecular weight (400-1000kDa) carrier molecules that function in microbial iron uptake (Simpson and Neilands, 1976). These iron-chelators are usually Fe(III)-

specific, and bind Fe(III) with an affinity of the order of 10^{30}. Of the compounds that are used as siderophores by cyanobacteria, two - schizokinen and synechobactin - have been shown to be synthesised following the decarboxylation of 2,4-DAB (Balasubramanian et al., 2006; Nicolaisen et al., 2008; Årstøl and Hohmann-Marriott, 2019).

The low concentration of iron (picomolar to nanomolar range) in the environment is typically rate-limiting for the growth of cyanobacteria and siderophores are essential to reconcile the profound insolubility of Fe(III) with the vital need of iron in cellular activities (Cunningham and John, 2017; Sunda and Huntsman, 2015). Cyanobacteria have larger iron requirements than non-photosynthetic bacteria (Shcolnick and Keren, 2006), as the processes of photosynthesis and, in some species, of N_2 fixation, require Fe-containing ferredoxin and nitrogenase, respectively (Simpson and Neilands, 1976). Cyanobacteria capable of producing siderophores under Fe-limiting conditions are probably at a selective advantage in microbial competition, protection from heavy-metal toxicity and bloom formation (Simpson and Neilands, 1976; Årstøl and Hohmann-Marriott, 2019).

Schizokinen, first discovered in *Bacillus megaterium* ATCC 19213 and isolated from *Anabaena* sp. strains PCC 7120 and PCC 6411 (Goldman et al., 1983; Simpson and Neilands, 1976), contains a derivative of 1,3-diaminopropane (Nunn and Codd, 2017). Synechobactins, amphiphilic cyanobacterial siderophores from *Synechococcus* sp. PCC 7002 (Armstrong and Van Baalen, 1979) are identical in structure to schizokinen, except for the fully saturated fatty acid tail on one of the two α-hydroxamate groups (Ito and Butler, 2005), but also contain the same derivative of 1,3-diaminopropane (Nunn and Codd, 2017). There is currently no biochemically substantiated biosynthetic pathway established for the cyanobacterial siderophores schizokinen and synechobactin. However, it is known that citrate-based siderophores are produced by NRPS-independent synthases (NIS) in cyanobacteria (Årstøl and Hohmann-Marriott, 2019), and that the two siderophores are structurally similar to other dihydroxamate-type siderophores, such as rhizobactin 1021 (Persmark et al., 1993) and aerobactin (de Lorenzo et al., 1986). *A priori*, we hypothesise that cyanobacteria use the same metabolic route. Rhizobactin 1021 is synthesised from the hydroxylation and acetylation of two molecules of 1,3-diaminopropane, which combined with citrate by a NIS synthetase, yields schizokinen (Lynch et al., 2001) (Fig. 2.4).

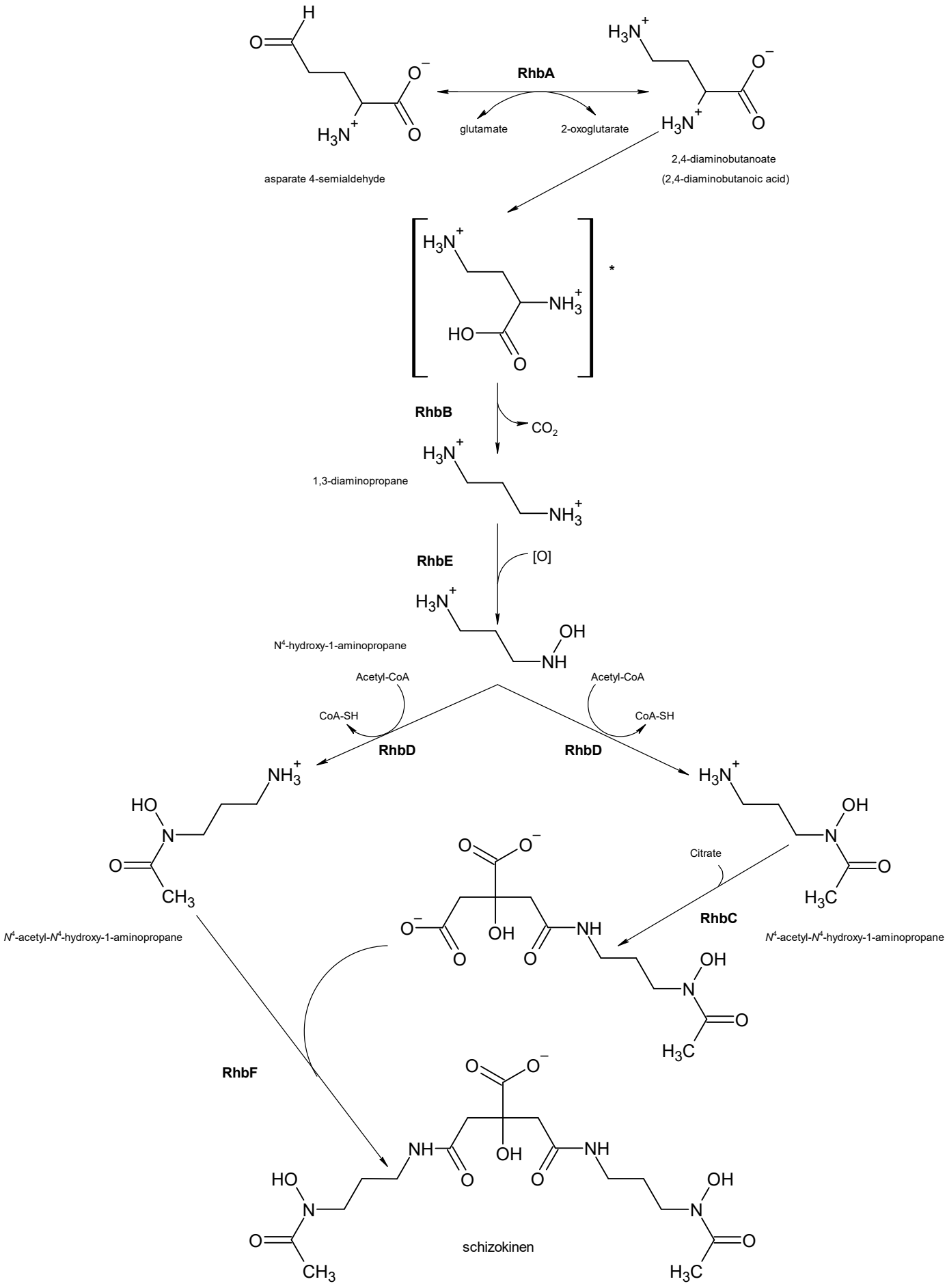

RhbA
glutamate
2-oxoglutarate
asparate 4-semialdehyde
2,4-diaminobutanoate
(2,4-diaminobutanoic acid)
RhbB
CO2
1,3-diaminopropane
RhbE
[O]
N4-hydroxy-1-aminopropane
Acetyl-CoA
CoA-SH
RhbD
RhbD
Acetyl-CoA
CoA-SH
N4-acetyl-N4-hydroxy-1-aminopropane
Citrate
RhbC
N4-acetyl-N4-hydroxy-1-aminopropane
RhbF
schizokinen

Figure 2.4. The biochemical pathway to schizokinen (Previous page). The established route to rhizobactin 1021 in *Sinorhizobium meliloti* (Lynch et al., 2001) includes schizokinen as its immediate precursor. (The enzyme responsible for the biosynthesis of rhizobactin 1021 from schizokinen is currently unknown). Nomenclature is from **www.brenda-enzymes.org**. The genes *rhbA* and *rhbB* (Lynch et al., 2001), correspond to the enzymes diaminobutanoate-2-oxo-glutarate transaminase (RhbA, EC 2.6.1.76) and diaminobutanoate decarboxylase (RhbB, EC 4.1.1.86). The enzymes RhbD (an acetylase), RhbE (which catalyses the oxidation of a single amino group of 1,3-diaminopropane), RhbC and RhbF (which catalyse condensation reactions) are unclassified. Redrawn from Lynch et al., (2001). The molecule in square brackets* represents 2,4-diaminobutanoate (2,4-diaminobutanoic acid) after rotation of the amino and carboxylate functions about carbon-2 to allow visual alignment of the amino groups in 1,3-diaminopropane and subsequent derivatives. Figure: Peter B. Nunn; used with permission.

2.3.3 Polyamines

Polyamines, polycationic hydrocarbon molecules with multiple amino groups, have important roles in prokaryotic protein synthesis (Igarashi and Kashiwagi, 2000; Shah and Swiatlo, 2008), also contributing to cyanobacterial bloom formation and siderophore production (Keating et al., 2002). Spermidine, a widespread polyamine in all groups of cyanobacteria (Hosoya et al., 2005), can be synthesised from putrescine via two different routes (Fig. 2.5). Putrescine biosynthesis is shared by both pathways and may occur via ornithine, through the action of ornithine decarboxylase (EC 4.1.1.17) or via arginine, which serves as the initial substrate for arginine decarboxylase (EC 4.1.1.19), agmatine iminohydrolase (EC 3.5.3.12) and *N*-carbamoylputrescine amidohydrolase (EC 3.5.1.53) (Zhu et al., 2015). The classical pathway begins with the decarboxylation of SAM by S-adenosylmethionine decarboxylase (SAMDC, EC 4.1.1.50). The decarboxylated SAM and putrescine serve as substrates for the biosynthesis of spermidine, in a reaction catalysed by spermidine synthase (SPDS, EC 2.5.1.16) (Lee et al., 2009; Shah and Swiatlo, 2008). In the alternative pathway, carboxyspermidine dehydrogenase (CASDH, EC 1.5.1.43) catalyses the reaction between aspartate 4-semialdehyde and putrescine that leads to carboxyspermidine, which is then decarboxylated by carboxyspermidine decarboxylase (CASDC, EC 4.1.1.96) to yield spermidine (Hanfrey et al., 2011; Lee et al., 2009). In organisms encoding diaminobutanoate-2-oxo-glutarate transaminase and diaminobutanoate decarboxylase, 1,3-diaminopropane is produced and can serve as a substrate for CASDH, yielding carboxynorspermidine, which in turn is used by CASDC to produce norspermidine (Yamamoto et al., 1986; Zhu et al., 2015). Both the putative and alternative pathways for spermidine biosynthesis have been described

in cyanobacteria (Zhu et al., 2015). In these species, the route to norspermidine via the alternative pathway could represent another physiological role for the production of 2,4-DAB via the aspartate 4-phosphate pathway.

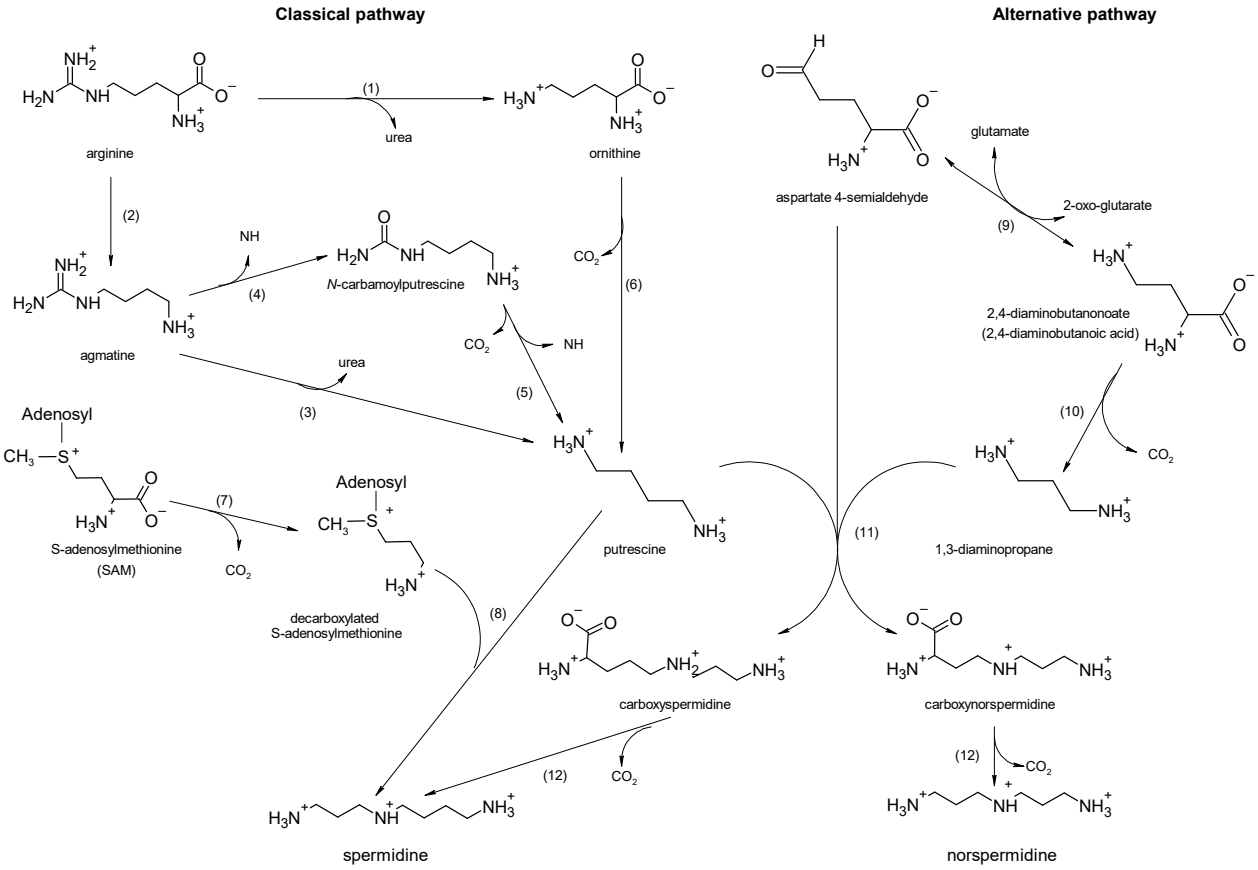


Figure 2.5. The biochemical pathways to spermidine and norspermidine in cyanobacteria. The established route to spermidine and norspermidine is adapted from Zhu et al. (2015) and the figure of Lee et al. (2009) has been redrawn and updated.
The nomenclature is from **www.brenda-enzymes.org**; the Enzyme Commission numbers are as follows: (1) arginase: EC 3.5.3.1; (2) arginine decarboxylase: EC 4.1.1.19; (3) agmatine ureohydrolase: EC 3.5.3.11; (4) agmatine iminohydrolase: EC 3.5.3.12; (5) *N*-carbamoylputrescine amidohydrolase: EC 3.5.1.53; (6) ornithine decarboxylase: EC 4.1.1.17; (7) S-adenosylmethionine decarboxylase: EC 4.1.1.50; (8) spermidine synthase: EC 2.5.1.16; (9) diaminobutanoate-2-oxo-glutarate transaminase: EC 2.6.1.76; (10) diaminobutanoate decarboxylase: EC 4.1.1.86; (11) carboxyspermidine dehydrogenase: EC 1.5.1.43; (12) carboxyspermidine decarboxylase: EC 4.1.1.96. Reactions (7) and (8) are involved in the biosynthesis of spermidine from putrescine *via* the classical pathway. Reactions (11) and (12) are involved in the biosynthesis of spermidine (in species without 1,3-diaminopropane) or norspermidine *via* the alternative pathway (where 1,3-diaminopropane is available). Figure: Peter B. Nunn; used with permission.

2.3.4 The biosynthesis of 2,4-DAB derived from SAM and other pathways to 2,4-DAB

There is another known mechanism to 2,4-DAB biosynthesis. In *Lathyrus* species, isoxazolinones are used for the biosynthesis of both 2,3-diaminopropanoic acid (2,3-DAP) and 2,4-DAB. In *Pisum sativum* and *Lathyrus sativus* seedlings, β-(isoxazolin-5-on-2-yl)-alanine (BIA) was shown to be synthesised from *O*-acetyl-L-serine (OAS) and isoxazolin-5-one by cysteine synthase (EC 2.5.1.47). 2,3-DAP can be released after enzymatic cleavage of BIA (Ikegami et al., 1991). In *Lathyrus odoratus* seedlings, 2-(3-amino-3-carboxypropyl)-isoxazolin-5-one (ACI), the higher homologue of BIA, is present and causes neurotoxic symptoms similar to 2,4-DAB in experimental animals (Lambein, 1981). This amino acid was shown to be synthesised from SAM and isoxazolin-5-one (Callebaut and Lambein, 1977; Ikegami et al., 1993). A suitable route to 2,4-DAB would be through enzymatic transfer of the 3-amino-3-carboxypropyl moiety from SAM to an intermediate, which could be cleaved to yield 2,4-DAB (Ikegami and Murakoshi, 1994) (Fig. 2.6). Although the transferase from *Lathyrus odoratus* has not been purified or sequenced, SAM-dependent 3-amino-3-carboxypropyl transferases have been described that can catalyse this transfer reaction (Meyer et al., 2020; Reeve et al., 1998; Riekhof et al., 2005). However, no enzyme has been shown to be able to cause the cleavage of such a complex. Additionally, although there are no reports of isoxazolinones having been found in cyanobacteria (Nunn and Codd, 2017), the four-carbon chain of methionine has been shown to be donated from SAM to a variety of structures, other than isoxazolinones. Previously characterised naturally occurring 3-amino-3-carboxypropyl-substituted heterocycles shown to derive from SAM as the donor of the 3-amino-3-carboxypropyl side chain include discadenine in *Dictyostelium discodeideum* (Taya et al., 1978), a modified nucleotide in *Escherichia coli* tRNA (Nishimura et al., 1974), plant siderophores of the mugineic acid family (Shojima et al., 1990; Shojima et al., 1989), the post-translationally modified histidine of elongation factor 2, diphthamide (Chen and Bodley, 1988), and the polyamine, spermidine (Zhu et al., 2015).

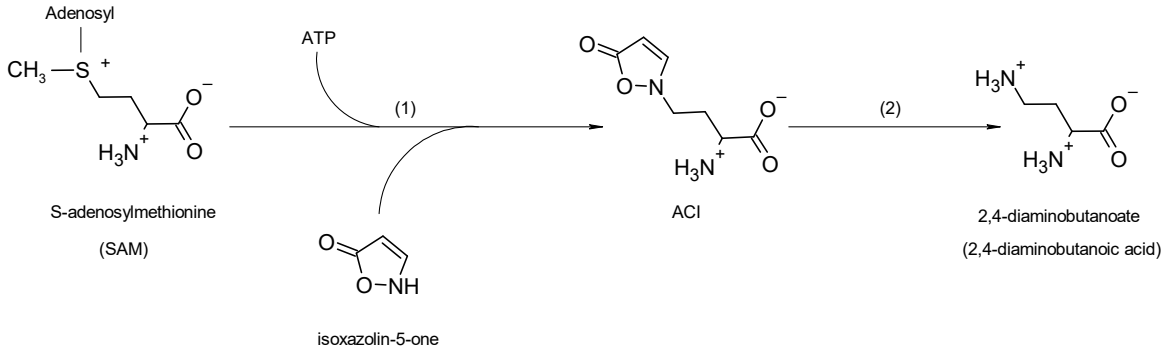


Figure 2.6. The pathway to 2,4-DAB derived from S-adenosylmethionine (SAM). The formation of 2,4-diaminobutanoate (2,4-diaminobutanoic acid) from S-adenosylmethionine (SAM) and isoxazolin-5-one in *Lathyrus sylvestris* (Callebaut and Lambein, 1977; Ikegami and Murakoshi, 1994). ACI: 2-(3-amino-3-carboxypropyl)-isoxazolin-5-one (Kuo et al., 1982). The enzymes catalysing the two reactions are unclassified. Figure: Peter B. Nunn; used with permission.

2,4-DAB could be also formed non-enzymatically. In the protein gliadin, the neurotoxin is formed chemically from glutamine following Hoffman degradation (Synge, 1939), and from asparagine during peptide synthesis (Ressler, 1956). There appear to be no biological equivalents to these reactions.

2.3.5 Our contribution

In this study, we use bioinformatics tools to investigate hypotheses concerning 2,4-DAB biosynthesis in cyanobacteria through the assessment of the presence or absence of each enzyme in the aspartate 4-phosphate pathway and a pathway to 2,4-DAB derived from SAM, across 130 cyanobacterial genomes. We propose that, in cyanobacteria, the biosynthesis of 2,4-DAB through the aspartate 4-phosphate pathway occurs but is limited to a subset of species. We highlight its potential physiological roles in siderophore and polyamine biosynthesis. We also suggest that 2,4-DAB production from SAM may be possible in some cyanobacterial species, however in-depth analyses are hindered, as candidate enzymes and intermediates participating in this pathway are lacking. Ultimately, the biosynthesis of 2,4-DAB appears to be either restricted to some cyanobacterial species, or there may be multiple and additional routes, and roles, for the synthesis of this neurotoxin.

2.4 Methodology

2.4.1 Phylogenetic reconstruction

130 high-quality cyanobacterial genomes and three outgroup genomes (one Proteobacteria, one Actinobacteria, and one Chloroflexi) were selected from RefSeq (Supplementary Material Section S4).

The genomes used were sequenced from species cultured under different conditions. The majority (53-54%) of genomes were extracted from species grown in axenic culture, whilst the remaining genomes were derived from species in non-axenic monocyanobacterial cultures (21-22%), or consisted of metagenome-assembled genomes (15-16%). The culture types of the organisms yielding the remaining 8-9% of the genomes are unknown (i.e. were not disclosed/described by authors/databases) (Supplementary Table S6).

Genome-wide protein sets were compared with BLAST (Altschul et al., 1997) and clustered into orthogroups using OrthoFinder (v.2.1.2) (Emms and Kelly, 2015) with a permissive MCL (van Dongen, 2000) inflation value of 1.1 for species phylogeny reconstruction. This resulted in 9591 orthogroups of two or more sequences each. Multiple sequence alignments were generated for all orthogroups using MAFFT (v.7.307) (Katoh and Standley, 2013). Orthogroups were mapped to pre-defined reliable AMPHORA2 markers for bacteria (Wu and Eisen, 2008). 29 orthogroups showed no strong evidence of horizontal gene transfer, or systematic deviation from the single maximum likelihood species tree, which was reconstructed from the concatenation of these 29 orthogroups using the LG+C60+F+R substitution model in IQ-TREE (v.1.6.10) (Nguyen et al., 2015) (Supplementary Material Section S6). Branch support was estimated using ultrafast bootstrap with 1000 replicates (Hoang et al., 2017), the tree was rooted between the ingroup and outgroup, and the outgroup was removed. The tree was annotated using the iTOL software (Letunic and Bork, 2016).

2.4.2 Profile hidden Markov models

From the protein sequence database at the NCBI (https://www.ncbi.nlm.nih.gov/), amino acid sequences with the same enzyme nomenclature as given in the BRENDA database (Chang et al., 2021) were selected for each enzyme in the aspartate 4-phosphate pathway (Fig. 2.2) and

for SAM-dependent 3-amino-3-carboxypropyl transferase (Fig. 2.6). Sequence choice prioritised cyanobacterial proteins and, where absent/scarce, bacterial proteins. The selection of multiple sequences from the same species was avoided. The number of sequences selected for the construction of each profile hidden Markov model (pHMM) ranged from 15 to 62 (Supplementary Table S8), comparable with an average of 22 sequences used for seed alignments in the Pfam database (Sonnhammer et al., 1998). For each enzyme, multiple sequence alignments were generated using MAFFT (v.7.307), with the E-ins-I algorithm, and pHMMs were generated using HMMER (v. 3.1b2) (hmmer.org).

Quality assessments were performed to analyse how accurate the generated pHMMs were at finding homologues, using cytochrome C oxidase subunit III, phycobilisome protein, photosystem II protein and photosystem II reaction centre M protein as positive controls (Supplementary Material Section S7.1).

The pHMMs for enzymes in the aspartate 4-phosphate pathway and for SAM-dependent 3-amino-3-carboxypropyl transferase were used to search proteins of the 130 cyanobacterial genomes for homologues using HMM search in the HMMER package, with the default threshold (E-value ≤ 0.01).

2.4.3 Online database searches

Where necessary to confirm absence, enzymes in the aspartate 4-phosphate pathway (Fig. 2.2) and SAM-dependent 3-amino-3-carboxypropyl transferase (Fig. 2.6) were also used as queries in BLAST searches against all cyanobacterial proteomes available at NCBI.

2.4.4 Substrate specificity and active site identification

In order to differentiate between the enzymes present in the aspartate 4-phosphate pathway (Fig. 2.2) and the pathway to 2,4-DAB from SAM (Fig. 2.6) from homologous enzymes that carry out different functions, active site and substrate specificity searches were conducted for enzymes downstream of- and including diaminobutanoate-2-oxo-glutarate transaminase, and the SAM-dependent 3-amino-3-carboxypropyl transferase, based on knowledge from the literature, indicated below.

Sequence alignments, including information about strictly conserved sites, residues involved in ligand-binding, pyridoxal 5'-phosphate (PLP)-binding, iron-binding, SAM-binding, and stabilisation of loop-architectures were available for diaminobutanoate-2-oxo-glutarate transaminase (Richter et al., 2019), diaminobutanoate acetyltransferase (Richter et al., 2020), ectoine synthase (Widderich et al., 2016), ectoine hydroxylase (Reuter et al., 2010; Widderich et al., 2014a; Widderich et al., 2014b), SAM-dependent 3-amino-3-carboxypropyl transferase (Reeve et al., 1998; Riekhof et al., 2005), and to some extent for diaminobutanoate decarboxylase (Yamamoto et al., 2000). No information is currently available regarding the active site of diaminobutanoate-pyruvate transaminase. Where data were available, multiple sequence alignments were constructed for each of these enzymes using MAFFT, which comprised the significant pHMM search hits from the cyanobacterial species in the dataset, as well as sequences with known active sites and substrate specificity from the literature, listed above. Multiple alignments were visualised using Jalview (v.2.11.1.0) (Waterhouse et al., 2009). Amino acid differences between sequences with known active sites and the significant pHMM search hits in cyanobacteria were counted, giving a Hamming distance (Supplementary Tables S12-S17). Low Hamming distance indicates well-conserved sequence at the active site. This serves as supporting evidence that the pHMM search hit in question corresponds to a true enzyme in the aspartate 4-phosphate pathway or in the pathway to 2,4-DAB derived from SAM.

2.4.5 Gene family phylogeny reconstruction

To further aid in the differentiation between enzymes belonging to the aspartate 4-phosphate pathway (Fig. 2.2) and the pathway to 2,4-DAB from SAM (Fig. 2.6) from homologs with different functions, gene trees were reconstructed for enzymes downstream of- and including diaminobutanoate-2-oxo-glutarate transaminase, and the SAM-dependent 3-amino-3-carboxypropyl transferase, using IQ-TREE with the protein substitution model chosen according to BIC at the ModelFinder step (Kalyaanamoorthy et al., 2017). Branch support was estimated with ultrafast bootstrap for 1000 replicates (Supplementary Fig. S3-S9). The grouping of sequences with different functional annotations was assessed. For example, in a gene phylogeny, if one sequence is included within a group of sequences with a different and consistent functional annotation, it is plausible to assume that the first sequence may have been mis-annotated and should adopt the annotation of the neighbouring branches.

2.4.6 Analysis of specialised metabolite clusters

AntiSMASH (v.5.1.2) (Blin et al., 2019) was used for the identification and annotation of specialised metabolite clusters in the cyanobacterial species in the dataset. Diaminobutanoate-2-oxo-glutarate transaminase and diaminobutanoate decarboxylase were checked for their co-localisation within specialised metabolite clusters and their proximity to each other. In species where these enzymes were not inserted in specialised metabolite clusters, the Artemis Comparison Tool (ACT) (Carver et al., 2005) was used for the localisation of the proteins in the genome, their juxtaposition, and assessment of the neighbouring genes.

2.4.7 Compilation of evidence of 2,4-DAB-producing cyanobacteria from the literature

Supplementary Table S5 presents a compilation of 13 studies which have reported evidence from analytical chemistry of cyanobacterial species capable of producing 2,4-DAB or 2,4-DAB and BMAA. Information regarding the analytical method used to identify and quantify the neurotoxins, as well as the culture type of the cyanobacterial samples are included, if available.

2.5 Results and discussion

2.5.1 Cyanobacterial species encoding enzymes in the aspartate 4-phosphate pathway

The first three enzymes in the aspartate 4-phosphate pathway (Fig. 2.2), aspartate transaminase, aspartate kinase and aspartate-semialdehyde dehydrogenase, are widespread in cyanobacteria. These enzymes are responsible for amino acid biosynthesis and metabolism and, in the few genomes where genetic evidence for one or more of these enzymes was lacking (one instance for aspartate transaminase, five instances for aspartate kinase and four instances for aspartate-semialdehyde dehydrogenase), we propose errors in sequencing, assembly, genome annotation or functional annotation as explanations (Supplementary Tables S1-S13).

Based on knowledge from active site and substrate specificity searches (Supplementary Tables S12-S17), from the 130 cyanobacterial species present in our dataset, diaminobutanoate-2-oxo-glutarate transaminase is believed to be present in 18 species, out of which only seven also encode diaminobutanoate decarboxylase (Table 2.1). *Xanthomonas* sp. encodes a transaminase that uses pyruvate, instead of 2-oxo-glutarate, as the oxo-acid (Rao et al., 1969). In cyanobacteria a gene for diaminobutanoate-pyruvate transaminase (EC 2.6.1.46) was not found in the genome of any species.

Table 2.1. Localisation in the genome of the enzymes diaminobutanoate-2-oxo-glutarate transaminase and diaminobutanoate decarboxylase. On the left are the cyanobacterial species encoding both enzymes. AntiSMASH was used to identify and annotate specialised metabolite clusters containing both 2,4-DAB transaminase (EC 2.6.1.76) and decarboxylase (EC 4.1.1.86). When these enzymes where not included in specialised metabolite clusters, the Artemis Comparison Tool (ACT) was used to locate the proteins and characterise the neighbouring genes.

	Proximity on genome	**Inclusion in specialised metabolite cluster**	**Specialised cluster type**
Chroococcidiopsis thermalis PCC 7203	Adjacent	No	Not applicable
Cyanobacterium aponinum PCC 10605	Adjacent	Yes	Siderophore/NIS cluster
Gloeobacter violaceus PCC 7421	Adjacent	No	Not applicable
Hapalosiphon sp. MRB220	Not adjacent	Yes (transaminase), No (decarboxylase)	NRPS/PKS cluster (transaminase)
Moorea bouillonii PNG	Not adjacent	No	Not applicable
Phormidium ambiguum NIES-2119	Adjacent	Yes	Siderophore/NIS cluster
Trichormus variabilis ATCC 29413	Adjacent	Yes	Siderophore/NIS cluster

No cyanobacterial genomes appeared to encode diaminobutanoate acetyltransferase or ectoine hydroxylase, and only three species appear to possess the gene coding for ectoine synthase: *Calothrix rhizosoleniae* SC01*, Hydrocoleum* sp. CS-953 and *Planktothricoides* sp. SR001.

Some species were shown to encode diaminobutanoate-2-oxo-glutarate transaminase but not diaminobutanoate decarboxylase (e.g. *Nodularia* sp. NIES-3585), and *vice-versa* (e.g. *Phormidium tenue* NIES-30). There was also variation in the range of specialised metabolite clusters in which these enzymes were co-localised (Fig. 2.7).

Single cyanobacterial species derived from axenic cultures that encode diaminobutanoate-2-oxo-glutarate transaminase (*Calothrix brevissima* NIES-22*, Chroococcidiopsis thermalis* PCC 7203*, Cyanobacterium aponinum* PCC 10605*, Cyanobacterium* PCC 7702*, Cylindrospermum stagnale* PCC 7417*, Gloeobacter violaceus* PCC 7421*, Myxosarcina* sp. GI1*, Nodularia* sp. NIES-3585*, Stanieria cyanosphaera* PCC 7437 and *Trichormus variabilis* ATCC 29413), are believed to be able to biosynthesise 2,4-DAB through the aspartate 4-phosphate pathway, since contamination by other 2,4-DAB-producing bacteria can be excluded. It is possible that other cyanobacterial species encoding the transaminase that were not cultured in axenic conditions are also 2,4-DAB-producers (namely *Acaryochloris marina* MBIC11017*, Hapalosiphon* sp. MRB220*, Hydrocoleum* sp. CS-953*, Moorea bouillonii* PNG*, Nostoc calcicola* FACHB-389*, Nostoc linckia* z2*, Phormidium ambiguum* NIES-2119 and *Tolypothrix bouteillei* VB521301). However, this would have to be verified in genomes sequenced from axenic cultures. There is no apparent correlation between species encoding genes for 2,4-DAB transaminase/decarboxylase and a specific cyanobacterial morphology, biochemistry or habitat (Fig. 2.7).

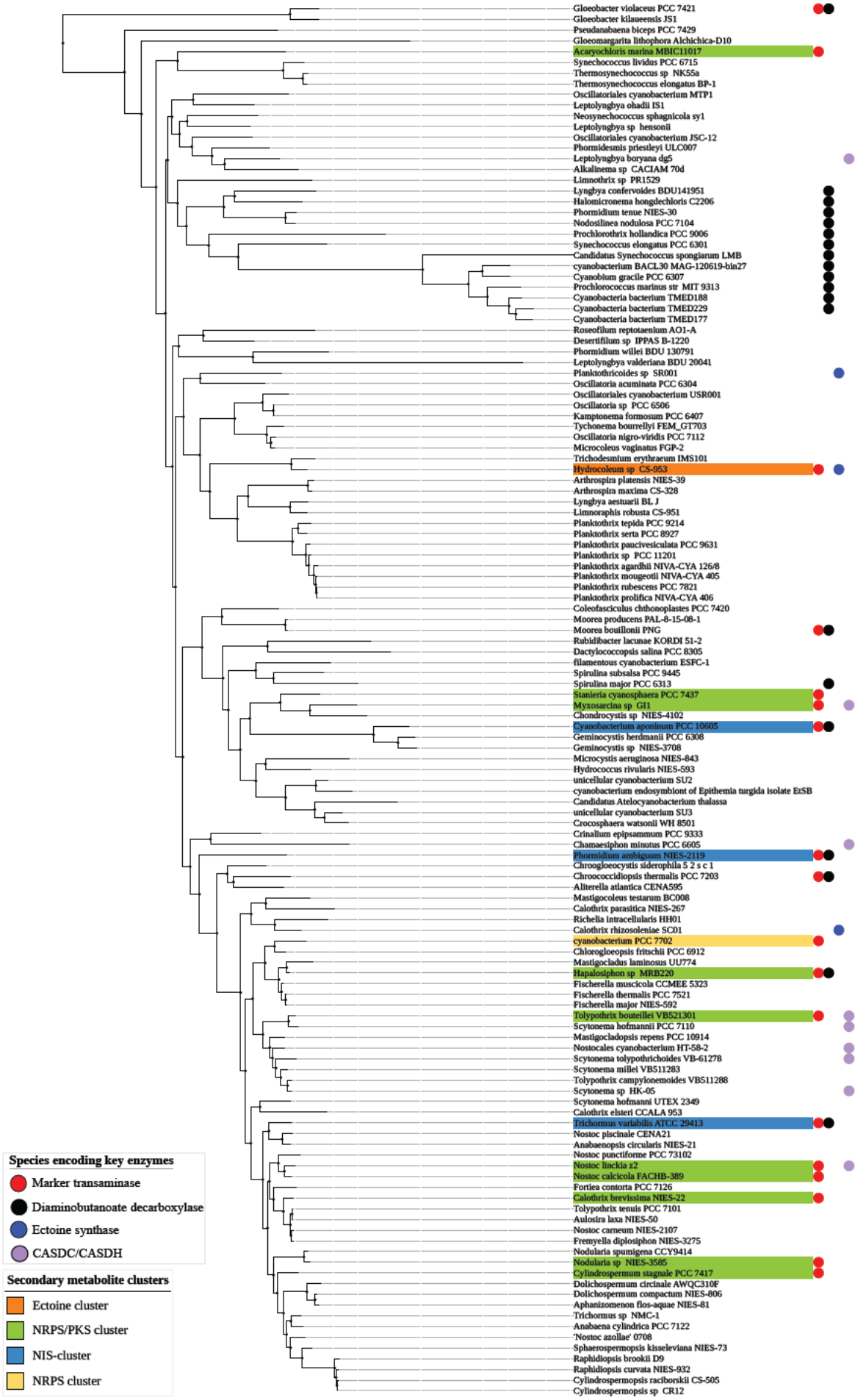

Gloeobacter violaceus PCC 7421
Gloeobacter kilaueensis JS1
Pseudanabaena biceps PCC 7429
Gloeomargarita lithophora Alchichica-D10
Acaryochloris marina MBIC11017
Synechococcus lividus PCC 6715
Thermosynechococcus sp NK55a
Thermosynechococcus elongatus BP-1
Oscillatoriales cyanobacterium MTP1
Leptolyngbya ohadii IS1
Neosynechococcus sphagnicola sy1
Leptolyngbya sp hensonii
Oscillatoriales cyanobacterium JSC-12
Phormidesmis priestleyi ULC007
Leptolyngbya boryana dg5
Alkalinema sp CACIAM 70d
Limnothrix sp PR1529
Lyngbya confervoides BDU141951
Halomicronema hongdechloris C2206
Phormidium tenue NIES-30
Nodosilinea nodulosa PCC 7104
Prochlorothrix hollandica PCC 9006
Synechococcus elongatus PCC 6301
Candidatus Synechococcus spongiarum LMB
cyanobacterium BACL30 MAG-120619-bin27
Cyanobium gracile PCC 6307
Prochlorococcus marinus str MIT 9313
Cyanobacteria bacterium TMED188
Cyanobacteria bacterium TMED229
Cyanobacteria bacterium TMED177
Roseofilum reptotaenium AO1-A
Desertifilum sp IPPAS B-1220
Phormidium willei BDU 130791
Leptolyngbya valderiana BDU 20041
Planktothricoides sp SR001
Oscillatoria acuminata PCC 6304
Oscillatoriales cyanobacterium USR001
Oscillatoria sp PCC 6506
Kamptonema formosum PCC 6407
Tychonema bourrellyi FEM_GT703
Oscillatoria nigro-viridis PCC 7112
Microcoleus vaginatus FGP-2
Trichodesmium erythraeum IMS101
Hydrocoleum sp CS-953
Arthrospira platensis NIES-39
Arthrospira maxima CS-328
Lyngbya aestuarii BL J
Limnoraphis robusta CS-951
Planktothrix tepida PCC 9214
Planktothrix serta PCC 8927
Planktothrix paucivesiculata PCC 9631
Planktothrix sp PCC 11201
Planktothrix agardhii NIVA-CYA 126/8
Planktothrix mougeotii NIVA-CYA 405
Planktothrix rubescens PCC 7821
Planktothrix prolifica NIVA-CYA 406
Coleofasciculus chthonoplastes PCC 7420
Moorea producens PAL-8-15-08-1
Moorea bouillonii PNG
Rubidibacter lacunae KORDI 51-2
Dactylococcopsis salina PCC 8305
filamentous cyanobacterium ESFC-1
Spirulina subsalsa PCC 9445
Spirulina major PCC 6313
Stanieria cyanosphaera PCC 7437
Myxosarcina sp GI1
Chondrocystis sp NIES-4102
Cyanobacterium aponinum PCC 10605
Geminocystis herdmanii PCC 6308
Geminocystis sp NIES-3708
Microcystis aeruginosa NIES-843
Hydrococcus rivularis NIES-593
unicellular cyanobacterium SU2
cyanobacterium endosymbiont of Epithemia turgida isolate EtSB
Candidatus Atelocyanobacterium thalassa
unicellular cyanobacterium SU3
Crocosphaera watsonii WH 8501
Crinalium epipsammum PCC 9333
Chamaesiphon minutus PCC 6605
Phormidium ambiguum NIES-2119
Chroogloeocystis siderophila 5 2 s c 1
Chroococcidiopsis thermalis PCC 7203
Aliterella atlantica CENA595
Mastigocoleus testarum BC008
Calothrix parasitica NIES-267
Richelia intracellularis HH01
Calothrix rhizosoleniae SC01
cyanobacterium PCC 7702
Chlorogloeopsis fritschii PCC 6912
Mastigocladus laminosus UU774
Hapalosiphon sp MRB220
Fischerella muscicola CCMEE 5323
Fischerella thermalis PCC 7521
Fischerella major NIES-592
Tolypothrix bouteillei VB521301
Scytonema hofmannii PCC 7110
Mastigocladopsis repens PCC 10914
Nostocales cyanobacterium HT-58-2
Scytonema tolypothrichoides VB-61278
Scytonema millei VB511283
Tolypothrix campylonemoides VB511288
Scytonema sp HK-05
Scytonema hofmanni UTEX 2349
Calothrix elsteri CCALA 953
Trichormus variabilis ATCC 29413
Nostoc piscinale CENA21
Anabaenopsis circularis NIES-21
Nostoc punctiforme PCC 73102
Nostoc linckia z2
Nostoc calcicola FACHB-389
Fortiea contorta PCC 7126
Calothrix brevissima NIES-22
Tolypothrix tenuis PCC 7101
Aulosira laxa NIES-50
Nostoc carneum NIES-2107
Fremyella diplosiphon NIES-3275
Nodularia spumigena CCY9414
Nodularia sp NIES-3585
Cylindrospermum stagnale PCC 7417
Dolichospermum circinale AWQC310F
Dolichospermum compactum NIES-806
Aphanizomenon flos-aquae NIES-81
Trichormus sp NMC-1
Anabaena cylindrica PCC 7122
'Nostoc azollae' 0708
Sphaerospermopsis kisseleviana NIES-73
Raphidiopsis brookii D9
Raphidiopsis curvata NIES-932
Cylindrospermopsis raciborskii CS-505
Cylindrospermopsis sp CR12
Species encoding key enzymes
Marker transaminase
Diaminobutanoate decarboxylase
Ectoine synthase
CASDC/CASDH
Secondary metabolite clusters
Ectoine cluster
NRPS/PKS cluster
NIS-cluster
NRPS cluster

Figure 2.7. Species phylogeny showing the cross-species distribution of specialised metabolite clusters containing a gene coding for diaminobutanoate-2-oxo-glutarate transaminase (EC 2.6.1.76) (Previous page). The enzymes encoded by each species are indicated by coloured circles, next to the species name. For enzymes' accession numbers, see Supplementary Tables S12-S17. See Supplementary Fig. S2 for bootstrap support of each branch.

2.5.2 Where the gene coding for 2,4-DAB transaminase is located within a NIS-cluster, it probably plays a physiological role in environmental iron-scavenging

From the seven cyanobacterial species encoding both diaminobutanoate-2-oxo-glutarate transaminase and diaminobutanoate decarboxylase, three (*Cyanobacterium aponinum* PCC 10605*, Phormidium ambiguum* NIES-2119 and *Trichormus variabilis* ATCC 29413) were shown to have both enzymes, co-localised within NIS clusters, as predicted by antiSMASH (Fig. 2.8).

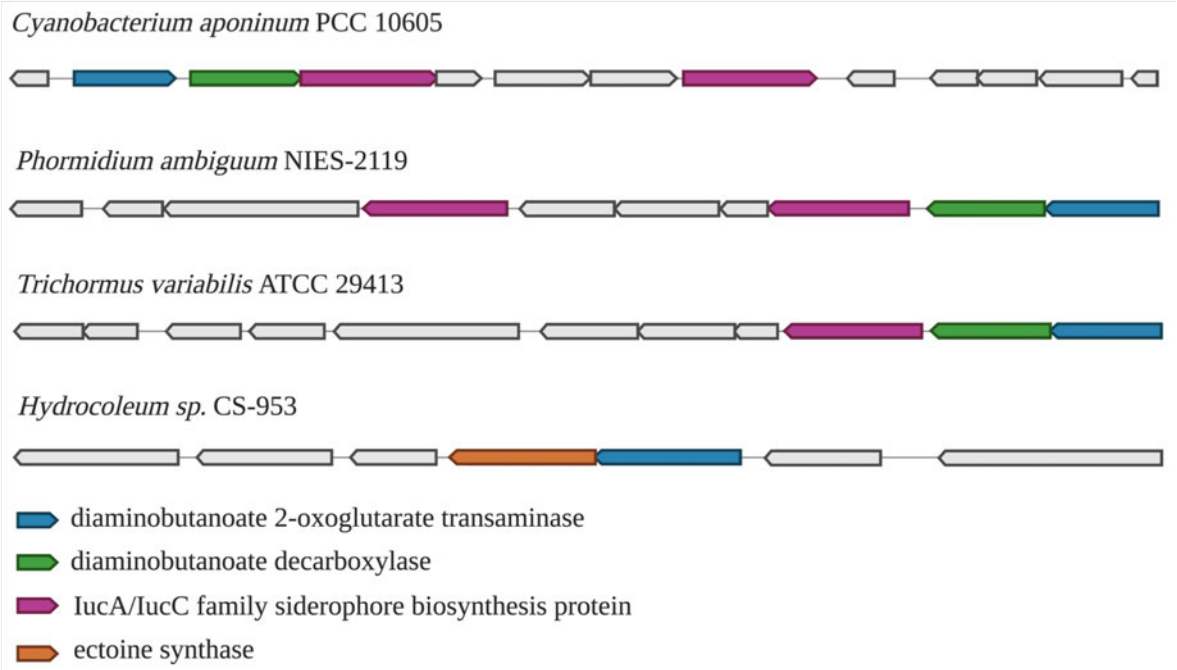


Figure 2.8. The localisation of the enzymes diaminobutanoate-2-oxo-glutarate transaminase, the downstream decarboxylase, and ectoine synthase within siderophore clusters and incomplete ectoine clusters. The top three clusters correspond to NIS clusters and the bottom cluster corresponds to an incomplete ectoine cluster. The organisation of NIS clusters follows the same pattern: diaminobutanoate-2-oxo-glutarate transaminase (EC 2.6.1.76), followed by diaminobutanoate decarboxylase (EC 4.1.1.86) and IucA/IucC (EC 6.3.2.38, EC 6.3.2.39) family siderophore biosynthesis protein. Arrows point to the orientation of transcription. AntiSMASH was used to identify and annotate specialised metabolite clusters containing both 2,4-DAB transaminase and decarboxylase or 2,4-DAB transaminase and ectoine synthase (EC 4.2.1.108).

Lynch et al. (2001), proposed a biosynthetic route to rhizobactin 1021 in the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti*, where schizokinen was indicated as the preceding compound to rhizobactin 1021. Since schizokinen is an intermediary product of the siderophore rhizobactin 1021, it is possible that these siderophores share an enzymatic pathway (Årstøl and Hohmann-Marriott, 2019). The biosynthesis of synechobactin is probably an extension of the pathway to schizokinen (Årstøl and Hohmann-Marriott, 2019). Evidence supporting these statements has been provided in cyanobacteria: Nicolaisen et al. (2008) identified a cluster of genes in *Anabaena* sp. PCC 7120 homologous to the biosynthesis-genes for rhizobactin 1021 and Balasubramanian et al. (2006) identified the siderophore characteristic IucABC-family proteins in an eight-gene operon, in *Synechococcus* sp. PCC 7002.

There are eight genes involved in the regulation, biosynthesis and transport of rhizobactin 1021 from 1,3-diaminopropane (Lynch et al., 2001). The genes *rhbA*, *rhbB*, *rhbC*, *rhbD*, *rhbE* and *rhbF* are part of a single operon and function in the biosynthesis of the siderophore. The product of *rhbA* and *rhbB* shows similarity to the proteins diaminobutanoate-2oxo-glutarate transaminase and diaminobutanoate decarboxylase (involved in 1,3-diaminopropane synthesis), whilst the products of *rhbC*, *rhbD*, *rhbF* and *rhbE* are homologous to the iron uptake chelate domains IucA, IucB, IucC and IucD, involved in the production of aerobactin (Lynch et al., 2001). These genes are characteristic of siderophore clusters (i.e. NIS biosynthesis pathways contain at least one enzyme with conserved N-terminal iron uptake chelate domains (Årstøl and Hohmann-Marriott, 2019)), and all characterised NIS clusters known to date encode either one or the two homologous synthetases IucA/IucC (de Lorenzo et al., 1986; de Lorenzo and Neilands, 1986). These synthetases define the NIS pathway and are present in the siderophore clusters of three cyanobacterial species in the dataset (Fig. 2.8), which we propose are able to synthesise iron chelators of the schizokinen/synechobactin type. Additional proteins, such as (GNAT family) N-acetyltransferase, lysine 6-monooxygenase, α/β hydrolase and transport proteins (e.g. iron-siderophore ABC transporter substrate-binding protein, ABC transporter ATP-binding protein, MFS transporter, and the TonB-dependent receptor) found in the cyanobacterial NIS clusters of Fig. 2.8, also have similar functional annotations to the proteins in the NIS clusters of other species (Challis, 2005; Lynch et al., 2001).

The synthesis of rhizobactin 1021 is similar to that of aerobactin (Årstøl and Hohmann-Marriott, 2019). Both iron-chelators are citrate-based hydroxamate siderophores with similar structures, differing only in the presence of a fatty acid moiety in rhizobactin 1021 and an additional carboxylic acid moiety and two saturated carbon atoms in each of the symmetrical arms of aerobactin (Lynch et al., 2001). It has been argued that, if aerobactin was derived from the same substrate as rhizobactin 1021 (i.e. 1,3-diaminopropane instead of lysine), the resulting product would have the same structure as rhizobactin 1021 (Lynch et al., 2001). Given this, Lynch et al. (2001) proposed that 1,3-diaminopropane is incorporated into rhizobactin 1021 by steps similar to those involved in the biosynthesis of aerobactin. However, since there is no evidence in the literature for the production of these siderophores in cyanobacteria, we hypothesise that the enzymes used for the biosynthesis of

schizokinen/synechobactin in cyanobacteria also correspond to those used for the biosynthesis of aerobactin and rhizobactin 1021.

The co-localisation of the 2,4-DAB transaminase and decarboxylase with the IucA/IucC synthetase genes, inside an NIS cluster, suggests that these enzymes provide the catalysis to form siderophores of the schizokinen/synechobactin type (Årstøl and Hohmann-Marriott, 2019). Hence, we propose that, in some cyanobacteria, 2,4-DAB can be biosynthesised through the aspartate 4-phosphate pathway, which is likely to play a physiological role in environmental iron-scavenging.

Since hydroxamate siderophores have the capability of complexing with metals in addition to iron (i.e. copper, uranium) (McKnight and Morel, 1980; Rashmi et al., 2013), additional physiological roles of 2,4-DAB as an antimicrobial agent (Matz et al., 2004) or in the detoxification of heavy metals (Årstøl and Hohmann-Marriott, 2019), respectively, may exist.

The genomes of *Cyanobacterium aponinum* PCC 10605 and *Trichormus variabilis* ATCC 29413 were both derived from axenic cultures, and therefore, based on our genomic analyses, it seems probable that these species are able to synthesise 2,4-DAB through the aspartate 4-phosphate pathway, possibly as a by-product of the production of siderophores of the schizokinen/synechobactin type. Although *Cyanobacterium aponinum* PCC 10605*, Trichormus variabilis* ATCC 29413 and *Phormidium ambiguum* NIES-2119 were all isolated from freshwater environments, siderophores are also known to be produced in terrestrial and marine environments (Årstøl and Hohmann-Marriott, 2019), possibly via the same pathways.

Since no cyanobacterial species in the dataset encode a gene for diaminobutanoate-pyruvate transaminase, the synthesis of 2,4-DAB, where present in cyanobacteria, appears to occur via a 2,4-DAB transaminase that uses 2-oxo-glutarate, and not pyruvate, as the oxo-acid. However, the possibility of uncharacterised diaminobutanoate-pyruvate transaminase enzymes, non-homologous to those that are known, cannot be disregarded. Without further knowledge of the relevant metabolite pool sizes in cyanobacteria, particularly of pyruvate, and the kinetic properties of the transaminase, the possibility of 2,4-DAB biosynthesis in cyanobacteria via diaminobutanoate-pyruvate transaminase cannot be excluded.

2.5.3 Co-localisation of 2,4-DAB transaminase within hybrid NRPS/PKS and unanswered questions

Many siderophores are polypeptides that are biosynthesised by members of the NRPS multienzyme family (Challis, 2005). These siderophores are different in structure from schizokinen (Jeanjean et al., 2008). Eleven species of cyanobacteria were found to encode 2,4-DAB transaminase within hybrid NRPS/PKS clusters. The product of these specialised metabolite clusters and the role played by 2,4-DAB transaminase is unknown. However, some similarity to pyoverdine clusters was found (Supplementary Table S4), possibly suggesting still unknown/uncharacterised siderophore pathways in cyanobacteria.

Although siderophores of the schizokinen type have been described in *Anabaena* sp. (Armstrong and Van Baalen, 1979; Simpson and Neilands, 1976), we did not find the genes for diaminobutanoate-2-oxo-glutarate transaminase or diaminobutanoate decarboxylase in the genome of *Anabaena cylindrica* PCC 7122. This is suggestive of a different pathway to siderophore production (i.e. not via the aspartate 4-phosphate pathway) and/or a yet-uncharacterised siderophore in this species. Our results are compatible with those of Goldman et al. (1983), who showed that *Anabaena cylindrica* PCC 7122 is able to produce hydroxamate siderophores with different characteristics compared to schizokinen.

Gloeobacter violaceus PCC 7421 was shown to encode both 2,4-DAB transaminase and decarboxylase, via adjacent genes on the genome, but not within a specialised metabolite cluster. Given that in cyanobacteria, siderophores were found to be produced in all but the most early-branching clades (Årstøl and Hohmann-Marriott, 2019), this finding may point to potentially additional and unknown physiological roles for the production of 2,4-DAB; for example, to enable this species to synthesise siderophores from 1,3-diaminopropane via a yet uncharacterised route; or to circumvent the loss of key genes essential for the production of iron chelators of the schizokinen/synechobactin type in this species.

2.5.4 An incomplete ectoine cluster as supporting evidence for the use of the aspartate 4-phosphate pathway for the production of 2,4-DAB in cyanobacteria

From the three cyanobacterial species that encode ectoine synthase (Fig. 2.7), *Calothrix rhizosoleniae* SC01 is the only one that also possesses diaminobutanoate-2-oxo-glutarate

transaminase. The genes coding for these enzymes appear adjacent on the genome and are predicted to be co-localised within an ectoine cluster (Fig. 2.8).

The fact that the genes for diaminobutanoate-2-oxo-glutarate transaminase and ectoine synthase appear adjacent on the genome of this species is compatible with our hypothesis that cyanobacteria use the aspartate 4-phosphate pathway for the production of 2,4-DAB from aspartate-4-semialdehyde.

Our results are supported by the findings of Czech et al. (2018) and Widderich et al. (2014a), who showed that, when present in cyanobacteria, ectoine synthase genes appear alone (i.e. as orphan genes), and not as part of complete ectoine clusters. It is currently not clear whether the solitary ectoine synthases are remnants of previously intact ectoine biosynthetic routes, whether they were recruited by existent genes in the ectoine pathway, whether they rely on an environmental provision of substrate to ectoine synthesis, or whether they have evolved new enzymatic functions, while possibly still retaining the capability of converting N^4-acetyl-2,4-diaminobutanoic acid to ectoine (Czech et al., 2018; Widderich et al., 2014a).

The cyanobacterium *Calothrix rhizosoleniae* SC01 was isolated from a symbiotic diatom of the genus *Chaetoceros*. In resemblance to what occurs in *Pseudomonas syringae* pv. syringae B728a (Kurz et al., 2010), we speculate that the diatom could provide the cyanobacterial symbiont with the substrate for the synthesis of ectoine. Genome-driven investigations of compatible solutes in the plant pathogen *Pseudomonas syringae* pv. syringae B728a indicated that this bacterium does not produce ectoine naturally under laboratory conditions (i.e. as it lacks the *ectAB* genes), but when surface-sterilised leaves of its host plant *Syringa vulgaris* were subjected to osmotic stress, ectoine production was observed (Kurz et al., 2010). The clustering of cyanobacteria in the phylogenomic studies of Czech et al. (2018) suggest that, when present, ectoine synthases in these species are probably catalytically competent.

2.5.5 Polyamine biosynthesis in cyanobacteria and additional potential physiological roles of 2,4-DAB

The co-localisation of the 2,4-DAB transaminase with CASDH and CASDC could indicate another physiological role for the biosynthesis of 2,4-DAB in cyanobacteria: that of polyamine (e.g. spermidine and norspermidine) biosynthesis (Fig. 2.5).

The gene pair CASDH/CASDC is present in nine cyanobacterial species in the dataset: *Scytonema* sp. HK-05, *Tolypothrix bouteillei* VB521301*, Chamaesiphon minutus* PCC 6605*, Scytonema hofmannii* PCC 7110*, Nostocales cyanobacterium* HT-58-2*, Nostoc linckia* z2*, Leptolyngbya boryana* dg5*, Scytonema tolypothrichoides* VB-61278 and *Myxosarcina* sp. GI1 (Fig. 2.7), and are only adjacent in the genome of the latter.

The presence, in a species, of diaminobutanoate-2-oxo-glutarate transaminase and diaminobutanoate decarboxylase with CASDH and CASDC is indicative of the ability to produce norspermidine through the alternative pathway for polyamine biosynthesis (Lee et al., 2009) (Fig. 2.5). However, the gene pair CASDH/CASDC was not found to be adjacent to, nor included in the same specialised metabolite cluster of diaminobutanoate-2-oxo-glutarate transaminase in any cyanobacterial species in the dataset. This suggests that, in cyanobacteria, the alternative pathway for polyamine biosynthesis is largely used for spermidine biosynthesis rather than norspermidine production, consistent with the findings of Lee et al. (2009).

However, in *Tolypothrix bouteillei* VB521301*,* the enzymes participating in the synthesis of spermidine via the classical pathway, SAMDC and SPDS, appear co-localised with diaminobutanoate-2-oxo-glutarate transaminase, within the same hybrid NRPS/PKS cluster (Fig. 2.9). We suggest that, in some cyanobacteria, spermidine synthesis may be linked to 2,4-DAB production, which is itself associated with the biosynthesis of yet unknown specialised metabolites, probably siderophores (i.e. the product of the NRPS/PKS cluster is currently undetermined). This would not be unprecedented as, in addition to their core physiological roles (e.g. in cell proliferation, growth and development), polyamines have been shown to be involved in microbial swarming (Sturgill and Rather, 2004), biofilm formation (Lee et al., 2009; McGinnis et al., 2009) and siderophore production (Brickman and Armstrong, 1996; Griffiths et al., 1984; Pfleger et al., 2007).

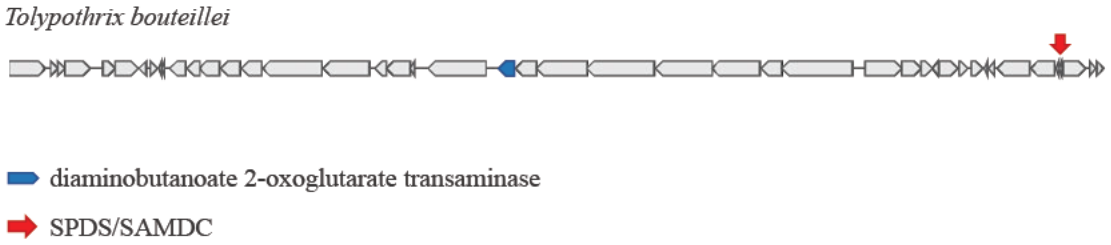


Figure 2.9. The localisation of the enzymes diaminobutanoate-2-oxo-glutarate transaminase, S-adenosylmethionine decarboxylase and spermidine synthase within a hybrid NRPS/PKS cluster. The product of the NRPS/PKS cluster is unknown. Arrows point to the orientation of transcription. The red arrow indicates the location of the gene pair SPDS/SAMDC (EC 2.5.1.16, EC 4.1.1.50), which is transcribed in the same direction as diaminobutanoate-2-oxo-glutarate transaminase (EC 2.6.1.76).

2.5.6 The route to 2,4-DAB derived from SAM: cyanobacterial species encoding homologs of SAM-dependent 3-amino-3-carboxypropyl transferase and unanswered questions

In *Lathyrus odoratus*, ACI was shown to be synthesised from isoxazolin-5-one and SAM (Ikegami et al., 1993). The enzymatic cleavage of ACI is a possible route to 2,4-DAB (Ikegami and Murakoshi, 1994) (Fig. 2.6). Bioinformatics analysis of the feasibility of this pathway in cyanobacteria is hindered as the transferase catalysing the donation of the four-carbon chain of methionine from SAM to isoxazolin-5-one has not been purified, and an enzyme capable of cleaving this complex is currently unknown. Additionally, there are no reports of isoxazolinones having been found in cyanobacteria (Nunn and Codd, 2017), and so other candidate acceptors would be needed to test this hypothesis.

However, SAM-dependent 3-amino-3-carboxypropyl transferases, shown to catalyse the transfer reaction from SAM, have been described (Meyer et al., 2020; Reeve et al., 1998; Riekhof et al., 2005), which enabled the search for homologous enzymes in cyanobacterial genomes. Homologs of SAM-dependent 3-amino-3-carboxypropyl transferases proteins, encompassing a methyltransferase fold (Fauman et al., 1999; Martin and McMillan, 2002) with a conserved SAM-binding pocket, were found in eight cyanobacterial species in the dataset: *Oscillatoriales cyanobacterium* MTP1, *Roseofilum reptotaenium* AO1-A, *Spirulina subsalsa*, *Spirulina major*, *Trichormus* sp. NMC-1, *Desertifilum* sp. IPPAS B-1220 and *Oscillatoria acuminata*. In all but one case, the homolog was functionally annotated as

Domain of Unknown Function (DUF) 3419 domain-containing protein (Supplementary Table S17). In *Roseofilum reptotaenium* AO1-A, the protein found was annotated as S-adenosylmethionine—diacylglycerol (DAG) 3-amino-3-carboxypropyl transferase, suggesting in this species, a reaction with a diacylglycerol acceptor. Complexes resulting from the transfer of the four-carbon chain of methionine from SAM to DAG have been described in bacteria (Riekhof et al., 2005), plants and fungi (Künzler and Eichenberger, 1997), and are involved in the replacement of membrane phospholipids with non-phosphorus lipids, when phosphate becomes limited in the environment (Benning et al., 1995; Geiger et al., 1999).

Cyanobacterial species encoding SAM-dependent transferases are candidates for 2,4-DAB production derived from SAM. The purification and characterisation of the enzymes in this pathway and their specificity, as well as knowledge regarding possible acceptors for the transfer reaction are needed to further analyse the possibility of 2,4-DAB production in cyanobacteria through this route.

2.5.7 Comparison between bioinformatics results and the analytical chemistry data from the literature

In an attempt to support our bioinformatics findings with analytical data from the literature, we compared our results with those of 13 studies where evidence for 2,4-DAB- or 2,4-DAB- and BMAA-producing cyanobacteria was provided (Supplementary Table S5). A scenario where a particular species is shown to encode the necessary genes needed for the biosynthesis of 2,4-DAB, and is proven to produce the same compound via analytical studies, serves as a strong indicator that this species is a 2,4-DAB-producer. On the other hand, a species that lacks the genes responsible for 2,4-DAB production, but is shown by analytical studies to produce the neurotoxin could be indicative of a false-positive result derived from a failed analytical method (especially probable if non-axenic cultures and/or low-specificity methods were employed), or suggests the existence of other, still undescribed routes to the biosynthesis of 2,4-DAB. All studies used variations of MS/MS for 2,4-DAB/BMAA identification and quantification, however only four analyses were derived from axenic cultures. Our 130 cyanobacterial species dataset and the compilation from the literature overlapped in six species: *Aphanizomenon flos-aquae*, *Cylindrospermopsis raciborskii*, *Microcystis aeruginosa*, *Nodularia spumigena*, *Planktothrix agardhii* and *Synechococcus*

elongatus. All species, apart from the latter, do not appear to have genes coding for 2,4-DAB transaminase, 2,4-DAB decarboxylase, and ectoine genes. *Synechococcus elongatus*, derived from an axenic culture, has a gene coding for 2,4-DAB decarboxylase, which is not included in a specialised metabolite cluster, and is lacking a gene for 2,4-DAB transaminase. This observation may suggest different routes for 2,4-DAB biosynthesis, or that there are unknown 2,4-DAB transaminase enzymes, non-homologous to those that have been characterised. The scarce amount of analytical studies using axenic cultures of clearly defined cyanobacterial species (most studies only include the cyanobacterial genus) does not allow for further comparisons.

2.6 Summary and conclusions

We have provided an initial bioinformatics insight into potential pathways of 2,4-DAB biosynthesis in cyanobacteria, which can lead to future experimental investigations.

The bioinformatics results point towards 2,4-DAB biosynthesis in some cyanobacterial species via the aspartate 4-phosphate pathway. In some cyanobacteria, the enzymes 2,4-DAB transaminase/decarboxylase are co-localised within NIS clusters, suggesting a functional association with siderophore biosynthesis. The fact that 2,4-DAB transaminase is found adjacent to an *ectC* gene in one cyanobacterial species serves as supporting evidence that cyanobacteria can produce 2,4-DAB via the aspartate 4-phosphate pathway, where it can serve as a substrate for either 1,3-diaminopropane or ectoine synthesis.

Due to the occurrence of 2,4-DAB transaminase/decarboxylase within hybrid NRPS/PKS clusters, we suggest that 2,4-DAB production may also be associated with polyamine synthesis, which itself is associated with biofilm formation and siderophore production. Species in which these enzymes occur outside specialised metabolite clusters may have developed new and, as yet, unknown physiological roles for 2,4-DAB biosynthesis, or may have lost genes needed for the production of siderophores, ectoine or polyamines.

In some cyanobacterial species, 2,4-DAB may be derived from SAM. Further information regarding the enzymes in this pathway, their specificity and possible acceptors is needed to further test this hypothesis.

Ultimately, the ability to synthesise 2,4-DAB via the aspartate 4-phosphate pathway and the pathway derived from SAM does not appear to be universal among cyanobacteria. This means that either 2,4-DAB is not a widespread neurotoxin in these species, or that there are additional and yet unknown pathways for the synthesis of 2,4-DAB. The presence of NIS clusters is also limited, appearing scattered in the species phylogeny. 2,4-DAB production and the presence of NIS clusters is not linked to specific environments, biochemistry or morphology.

Chapter 3

Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA)

3.1 Preface

The work in this chapter has been published as a manuscript in *Phytochemistry* and the first-person plural is used throughout to maintain consistency. Some changes have been made to the published version to preserve formatting across the thesis, and the introduction has been extended. I performed all analyses, wrote the first draft of the manuscript and supplementary materials, and produced table 3.1 and figures 3.6-3.8.

Peter B. Nunn produced all figures depicting biochemical pathways/reactions and molecules (figures 3.1-3.5 and 3.9).

Citation:

Mantas, M. J. Q., Nunn, P. B., Codd, G. A., Barker, D., 2022. Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA). *Phytochemistry* 200, 113198.

3.2 Graphical abstract

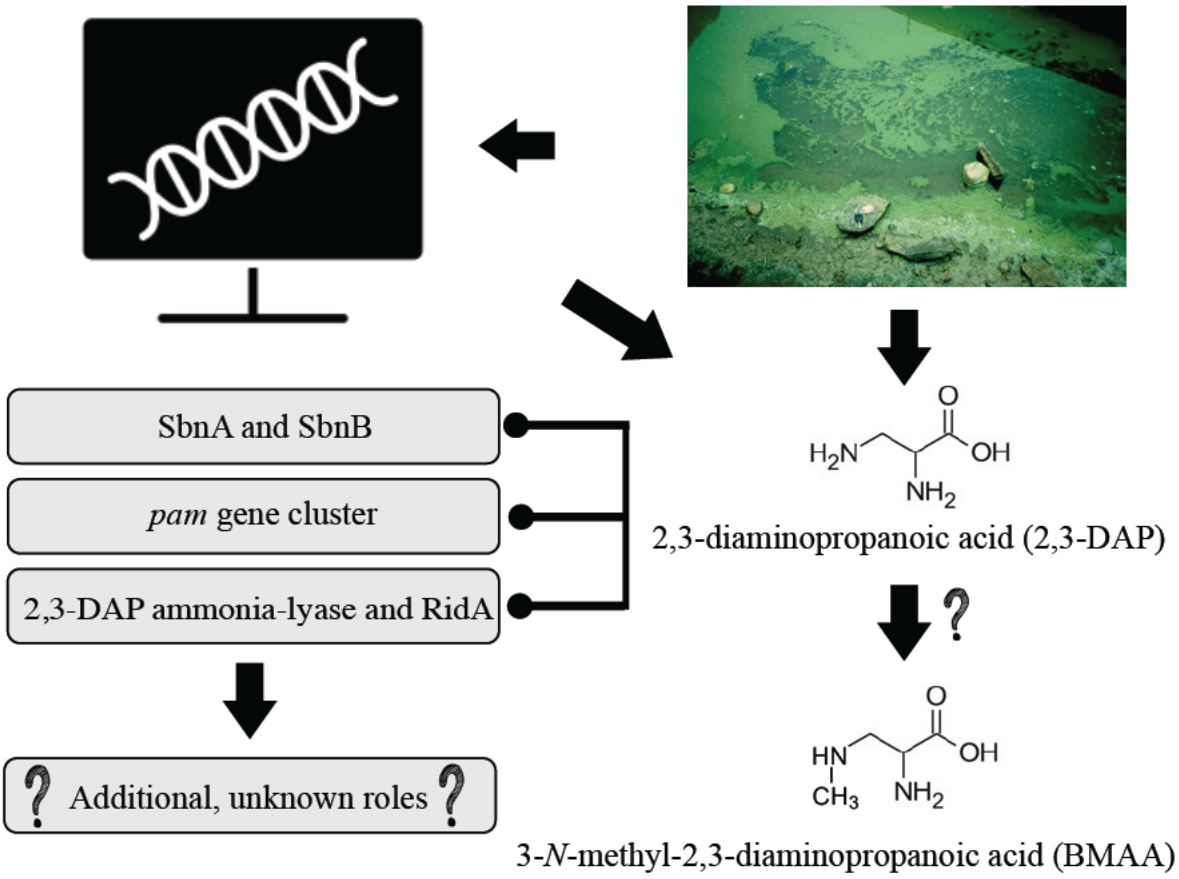


Photo: Geoffrey A. Codd; used with permission. Shore-line accumulation of toxin-producing cyanobacteria (*Microcystis aeruginosa*); freshwater loch, Scotland.

3.3 Introduction

3-*N*-methyl-2,3-diaminopropanoic acid (syn: α-amino-β-methylaminopropionic acid, MeDAP; β-*N*-methylaminoalanine, BMAA) is a neurotoxin that was first isolated from seed of *Cycas micronesica* (Vega and Bell, 1967; Nunn, 2017). BMAA is a non-encoded amino acid (i.e. not coded for by a codon in the genetic code). It has since been detected in a range of organisms, including some species of cyanobacteria (Cox et al., 2003; Cox et al., 2005; Banack et al., 2007; Downing et al., 2011; Downing and Downing, 2016). Ecological and physiological functions for BMAA have not been determined and metabolic pathways for the biosynthesis of BMAA in cyanobacteria are currently unknown. Among other benefits, knowledge of such pathways would allow targeting of specific biosynthesis genes by polymerase chain reaction (PCR)-based methods, which could be used as an early-warning system for the potential presence of cyanotoxins in water resources and other environments (Humbert, 2017; Kurmayer et al., 2017).

Interest in BMAA was generated by the possibility that it may be a contributory causative agent of the chronic neurological complex, amyotrophic lateral sclerosis/parkinsonism/dementia (ALS-PDC) of Guam (Nunn, 2017), but this remains a controversial matter (Chernoff et al., 2017; Dunlop et al., 2021). It has been show that the dietary habits of the indigenous Chamorro people of Guam, known to be rich in food containing BMAA (i.e. cycad seeds, cycad flour, and fruit bats that feed on the sarcotesta of the same seeds (Banack et al., 2006; Cox and Sacks, 2002)), associated with biomagnification of this neurotoxin in the food chain (i.e. fruit bats were reported to suffer a biomagnification of up to 100x compared with the concentration of BMAA in the coralloid roots of *Cycas* trees (Banack and Cox, 2003; Cox et al., 2003)), were concomitant with higher incidences of neurodegenerative diseases (Kurland and Mulder, 1954; Spencer et al., 1986). Epidemiological evidence for an environmental risk factor(s) in the causation of ALS-PDC exists with the possibility that BMAA, alongside other cycad toxins/carcinogens (i.e. methylazoxymethanol (MAM)-β-D-glucoside, formaldehyde-responsive miRNAs (Spencer et al., 2012; Spencer, 2019)) play a role in the incidence of sporadic neurodegenerative diseases in genetically vulnerable individuals (Bradley et al., 2013).

BMAA occurs free and as two bound forms (Vega and Bell, 1967; Polsky et al., 1972; Dossaji and Bell, 1973; Murch et al., 2004; Faassen et al., 2016) but the chemical complexes

in which BMAA is incorporated (corresponding to the soluble bound form and TCA-precipitated bound form) have not been elucidated. All three fractions of BMAA may or may not be present in organisms shown to produce BMAA (Rosén et al., 2016) and it cannot be dismissed that free and bound BMAA moieties may be derived from different metabolic precursors (Nunn and Codd, 2017).

The extent of BMAA biosynthesis throughout the Cyanobacteria has not been systematically determined due, to some extent, to analytical limitations (Faassen, 2014; Mantas et al., 2021). It is widely accepted that analytical methods using tandem mass spectrometry (MS/MS) methods are the most suitable for the identification and quantification of BMAA (Cohen, 2012; Faassen et al., 2012; Faassen, 2014). However, variation still exists regarding sample processing and separation methods (Faassen, 2014) and inter-laboratory comparisons are scarce (Faassen et al., 2016). Despite such problems, the ability of axenic, single strain cultures of some cyanobacteria to biosynthesise BMAA has been unequivocally confirmed (Eriksson et al., 2009; Downing et al., 2011; Downing and Downing, 2016).

In this study, we used bioinformatics tools to investigate hypotheses concerning BMAA biosynthesis in cyanobacteria through an assessment of the presence or absence of enzymes in six known potential metabolic pathways, across 130 cyanobacterial genomes. We show that most enzymes involved in pathways leading to the putative precursor of BMAA (2,3-diaminopropanoic acid, 2,3-DAP) in other species, were not found in cyanobacteria. Genes coding for SbnA and SbnB, by whose concerted action the biosynthesis of 2,3-DAP is known to occur in *Staphylococcus aureus*, were found, though limited to a subset of cyanobacterial species. We highlight the potential physiological role of 2,3-DAP in siderophore formation in some cyanobacterial species and show that the *pam* gene cluster, responsible for directing the biosynthesis of peptide-bound BMAA in *Paenibacillus larvae*, was not detected in 130 cyanobacterial species, nor was it found in 93 genomes of *Paenibacillus* (other than *P. larvae*) or two diatom species. We also show that the presence, in some cyanobacterial species, of genes putatively encoding the enzymes 2,3-diaminopropionate ammonia-lyase (DAPAL, EC 4.3.1.15) and reactive intermediate deaminase A (RidA, EC 3.5.99.10) could explain the failure to detect 2,3-DAP in analytical studies. The biosynthesis of 2,3-DAP in cyanobacteria appears to be either restricted to a

small subset of cyanobacterial species, or there may be multiple, additional, routes for the biosynthesis of this amino acid.

3.4 Routes to the biosynthesis of 2,3-DAP and BMAA in other taxa and potential relevance in cyanobacteria

3.4.1 The biosynthesis of BMAA from 2,3-diaminopropanoic acid (2,3-DAP)

The simplest explanation for the biosynthesis of BMAA is through the 3-*N* methylation of 2,3-DAP, found free and as simple derivative forms in plant species (Gmelin, 1959; Nunn and Codd, 2017) (Fig. 3.1).

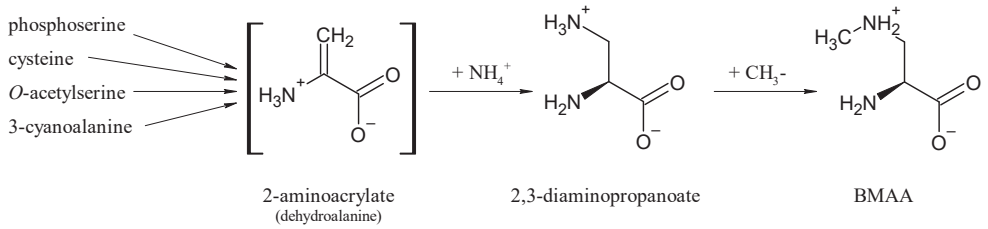


Figure 3.1. Biochemical pathway to free 2,3-diaminopropanoate and BMAA as proposed by Brenner et al. (2003). A cysteine synthase-like enzyme catalyses the synthesis of 2,3-diaminopropanoate from S-*O*-acetylserine, phosphoserine, cysteine or 3-cyanoalanine, and ammonium. A likely intermediate in this reaction is 2-aminoacrylate (dehydroalanine). S-adenosyl-S-methionine (SAM) donates the 3-*N*-methyl group to 2,3-diaminopropanoate, yielding BMAA. Figure: Peter B. Nunn; used with permission.

However, mechanistic support for this pathway (Brenner et al., 2003) is lacking in cyanobacteria. First, although the possible genes encoding cysteine synthase-like and methyltransferase enzymes occur in cyanobacteria (Aráoz et al., 2010), the existence of such enzymes with specificity for the proposed substrates remains speculative. Second, since this pathway requires ammonium, it may be difficult to reconcile with the results of Downing et al. (2011), who showed that BMAA levels in the non-nitrogen-fixing cyanobacterium *Microcystis* PCC 7806 increase under conditions of nitrogen (ammonium/nitrate) starvation and decrease when ammonium is added. Finally, this pathway assumes that BMAA is

biosynthesised by direct methylation of free 2,3-DAP, which has not been found in any cyanobacterial species or in cycads, which can accommodate cyanobacteria, to date.

There are at least four known biosynthetic routes to 2,3-DAP. Two are complex pathways in plants, and the others are simpler mechanisms found in *Streptomyces* and *Staphylococcus*.

3.4.2 Biosynthesis of 2,3-DAP from uracil in *Albizia julibrissin*

2,3-DAP biosynthesis from uracil occurs in plants such as *Albizia julibrissin*, which produce albizziine (3-*N*-ureido-2,3-diaminopropanoic acid) from uracil as part of the pyrimidine degradation pathway (Brown and Turan, 1995, 1996). This pathway encompasses five steps and the enzymes dihydrouracil dehydrogenase (syn: dihydropyrimidine dehydrogenase, DUD, EC 1.3.1.2), dihydropyrimidinase (DHP, EC 3.5.2.2), and β-ureidopropionase (syn: β-alanine synthase, *N*-carbamoyl-β-alanine amidohydrolase, βUP, EC 3.5.1.6) (Fig. 3.2). In theory, the methylation reaction resulting in the formation of BMAA could occur at any step in the metabolic pathway; 1-methyluracil, a potential methylated primary precursor, does not appear to be a natural product. Albizziine formation is almost entirely confined to the Mimosoideae, and there is no indication of its presence in cyanobacteria.

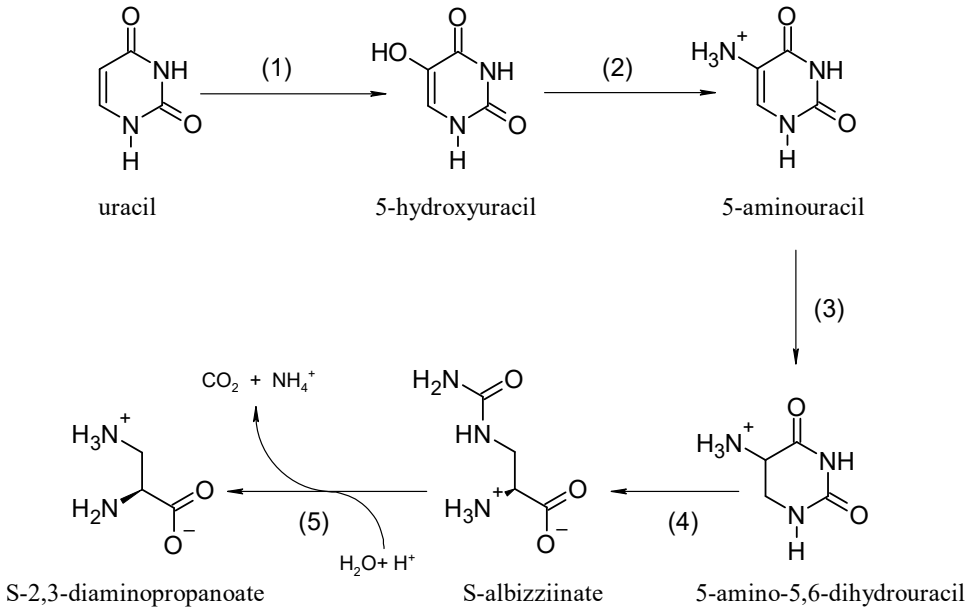


Figure 3.2. Biochemical pathway to S-2,3-diaminopropanoate from uracil in *Albizia julibrissin* (Previous page). Redrawn from Brown and Turan (1996). Uracil serves as a precursor for 5-hydroxyuracil formation through (1) unspecified hydroxylase activity and (2) an amination reaction, forming 5-aminouracil. (3) Dihydrouracil dehydrogenase (syn: dihydropyrimidine dehydrogenase, DUD, EC 1.3.1.2) reduces 5-aminouracil, in an NADPH-dependent reaction, to 5-amino-5,6-dihydrouracil, which serves as substrate for (4) dihydropyrimidinase (DHP, EC 3.5.2.2). Following cleavage of the pyrimidine ring, catalysed by DHP, S-2,3-diaminopropanoate is formed from S-albizziinate by hydrolysis, catalysed by (5) β-ureidopropionase (syn: β-alanine synthase, *N*-carbamoyl-β-alanine amidohydrolase, βUP, EC 3.5.1.6). The nomenclature is from www.brenda-enzymes.org . The charged forms of the amino acids are those predominating at physiological pH values. Figure: Peter B. Nunn; used with permission.

3.4.3 Biosynthesis of 2,3-DAP from β-(isoxazolin-5-on-2-yl) in *Lathyrus sativus* seedlings

Lathyrus sativus biosynthesises the amino acid β-*N*-oxalyl-2,3-diaminopropanoic acid (β-ODAP) (Rao et al., 1964). In the pathway to β-ODAP, β-(isoxazolin-5-on-2-yl)-S-alanine (BIA), identified by Kuo et al. (1998) in germinating seedlings of this species, is synthesised from isoxazolin-5-one and S-*O*-acetylserine via cysteine synthase (EC 2.5.1.47). Although BIA, and not 2,3-DAP, was detected in *L. sativus* seedlings, convincing evidence exists supporting the hypothesis that 2,3-DAP is the immediate precursor of β-ODAP (Malathi et al., 1970; Ikegami et al., 1999). Hence, it is suggested that 2,3-DAP, derived from BIA, is used to synthesise β-ODAP via S-2,3-diaminopropionate *N*-oxalyltransferase (EC 2.3.1.58) (Nunn and Codd, 2017) (Fig. 3.3). This enzyme has not been purified or sequenced. Since, in this pathway, 2,3-DAP does not appear free in plant seedlings, it is hypothesised that substrate channelling may occur, i.e. that the 2,3-DAP intermediate remains bound to the catalytic site of the enzyme, until the following reaction coverts it to the final product (Nunn and Codd, 2017). This process occurs frequently in a broad spectrum of metabolic reactions (Agius, 1997; Jørgensen et al., 2005). Although there are no reports of 2,3-DAP released from a bound form after hydrolysis of cyanobacterial cells, the fact that a 3-*N*-derivative of 2,3-DAP could be formed whilst the amino acid remains bound to the catalytic site of an enzyme could explain the failure to detect free 2,3-DAP in BMAA-producing cyanobacterial species (Nunn and Codd, 2017). Nevertheless, although isoxazolinones are widespread in plants (Lambein et al., 1986) and bacteria (Becker et al., 2017), there are no reports of their presence in cyanobacteria.

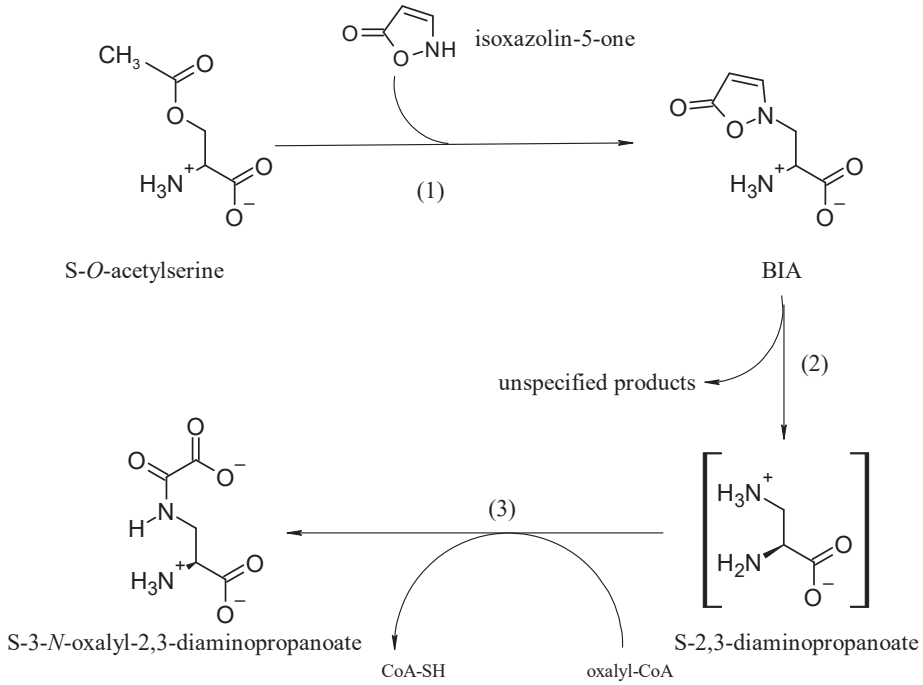


Figure 3.3. Biochemical pathway to S-2,3-diaminopropanoate from β-(isoxazolin-5-on-2-yl) in *Lathyrus sativus* seedlings. Redrawn from Ikegami et al. (1999). The nomenclature is from www.brenda-enzymes.org; Enzyme Commission numbers are: (1) cysteine synthase: EC 2.5.1.47; (2) S-2,3-diaminopropionate *N*-oxalyltransferase: EC 2.3.1.58. The charged forms of the amino acids are those predominating at physiological pH values. S-2,3-diaminopropanoic acid is shown in square brackets as it is not released free from the enzyme during the reaction. Figure: Peter B. Nunn; used with permission.

3.4.4 Biosynthesis of 2,3-DAP from serine in *Streptomyces*, *Clostridium* and plants

2,3-DAP biosynthesis from serine has been unequivocally shown to occur during the production of the antibiotics tuberactinomycin (TUB) and viomycin, in *Streptomyces* (Carter et al., 1974), and also in the biosynthesis of the similar peptide antibiotic, capreomycin, in *Streptomyces capreolus* A250 (Wang and Gould, 1993; Felnagle et al., 2007), although this was recently contested by Hsu et al. (2020). TUBs contain several unusual amino acids including β-ureidodehydroalanine, β-lysine, S-capreomycidine and 2,3-DAP, and are believed to be biosynthesised by a non-ribosomal peptide synthesis (NRPS) mechanism (Thomas et al., 2003). Following the identification of the viomycin biosynthetic gene cluster from *Streptomyces* sp. ATCC 11861, it became apparent that the biosynthesis of 2,3-DAP relied on the concerted action of two enzymes: VioK and VioB (Thomas et al., 2003). VioK

is an ornithine cyclodeaminase (OCD) (EC 4.3.1.12) and VioB (syn: 2,3-diaminopropanoic acid synthetase) is a putative *O*-acetyl-S-serine sulfhydrylase (OASS), homologous to serine dehydratases and cysteine synthases. However, while cysteine synthases use sulfur from sulfide in the addition reaction, VioB is thought to use the nitrogen from ammonium, released from S-ornithine by VioK (Thomas et al., 2003). The combined action of VioK and VioB enables the direct transfer of ammonium needed to form the 3-*N* amino group of 2,3-DAP, by-passing the generally low intracellular concentration of ammonium and driving the reaction towards 2,3-DAP formation (Fig. 3.4). Similar mechanisms operate in *Bacillus cereus* and *Bacillus thuringiensis* subsp. *kurstaki* YTP-1520 for the biosynthesis of the 2,3-DAP-containing peptide antibiotic, zwittermicin A (Emmert et al., 2004; Zhao et al., 2008). However, although multiple enzymes can catalyse β-replacement reactions involving ammonium (Parmeggiani et al., 2018), the rate of diffusion of ammonia inside a cell can be a limiting factor (Beasley et al., 2011).

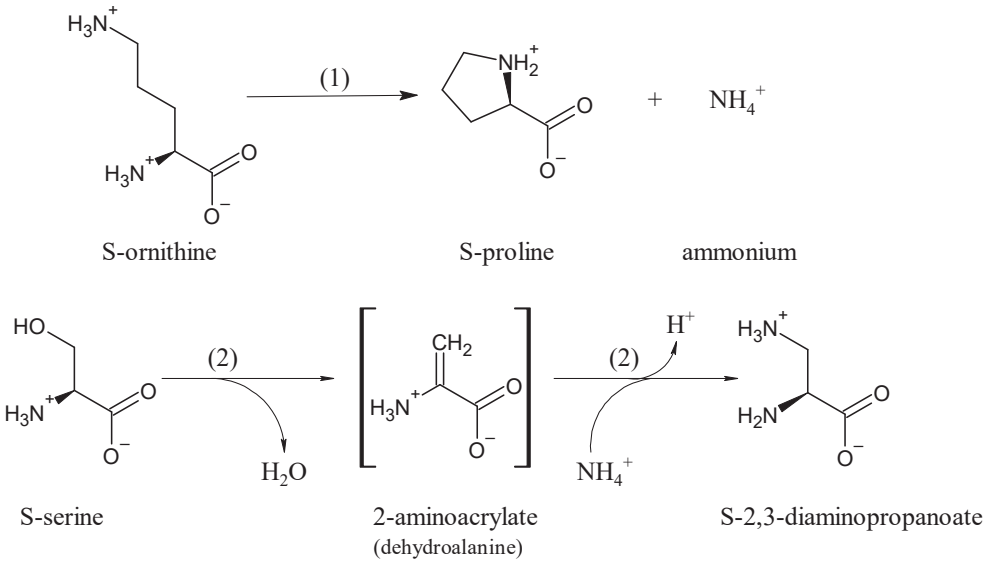


Figure 3.4. Biochemical pathway to S-2,3-diaminopropionate from S-serine in *Streptomyces*, *Clostridium* and some plants. Redrawn from Thomas et al. (2003).
(1) ornithine cyclodeaminase (syn: OCD, VioK, EC 4.2.1.12) catalyses the cyclisation of S-ornithine to S-proline, releasing ammonium, using NAD^+ as cofactor in *Clostridium* sp. PA 3679, *Nicotiana tabacum*, *Datura stramonium* and in *Lupinus angustifolius* (Costilow and Laycock, 1971; Mestichelli et al., 1979; Muth and Costilow, 1974).
(2) S-2,3-diaminopropionate synthetase (VioB, unclassified) catalyses the biosynthesis of a PLP-bound 2-aminoacrylate intermediate from S-serine or *O*-acetyl-S-serine (OAS) (Carter et al., 1974; Wang and Gould, 1993) and subsequently catalyses a β-substituent replacement reaction, similar to that observed with cysteine synthases. The enzyme nomenclature is from www.brenda-enzymes.org. The charged forms of the amino acids are those predominating at physiological pH values. Figure: Peter B. Nunn; used with permission.

3.4.5 Biosynthesis of 2,3-DAP in *Staphylococcus aureus*

Under iron-limiting conditions, *Staphylococcus aureus* synthesises two carboxylate-type siderophores, staphyloferrin A (SA) (Konetschny-Rapp et al., 1990) and staphyloferrin B (SB) (Beasley et al., 2011). To date, SB is the only iron-chelator known to contain 2,3-DAP (Cheung et al., 2009). *S. aureus* contains a nine-gene *sbn* siderophore biosynthesis operon (*sbnA-I*) in which SbnA and SbnB are responsible for 2,3-DAP biosynthesis (Thomas et al., 2003; Beasley et al., 2011; Kobylarz et al., 2014) (Fig. 3.5). SbnA (*N*-(2-amino-2-carboxyethyl)-S-glutamate synthase, EC 2.5.1.140) is a pyridoxal 5'-phosphate (PLP)-dependent enzyme similar to VioB (Heine et al., 2004) and commonly annotated as cysteine synthase (Kobylarz et al., 2014). However, unlike all previously known sulfhydrylases, SbnA uses S-*O*-phosphoserine (OPS) instead of OAS as substrate, with S-glutamate as nitrogen donor, instead of sulfur (Kobylarz et al., 2016). SbnB (*N*-[(2S)-2-amino-2-carboxyethyl]-S-glutamate dehydrogenase, EC 1.5.1.51) is homologous to the OCD protein family (including VioK) and amino acid dehydrogenases (Beasley et al., 2011). Nevertheless, Kobylarz et al. (2014) showed that SbnB uses NAD^+ as substrate, rendering it a closer relative of NAD^+-dependent amino acid dehydrogenases than to ornithine cyclodeaminase (OCD).

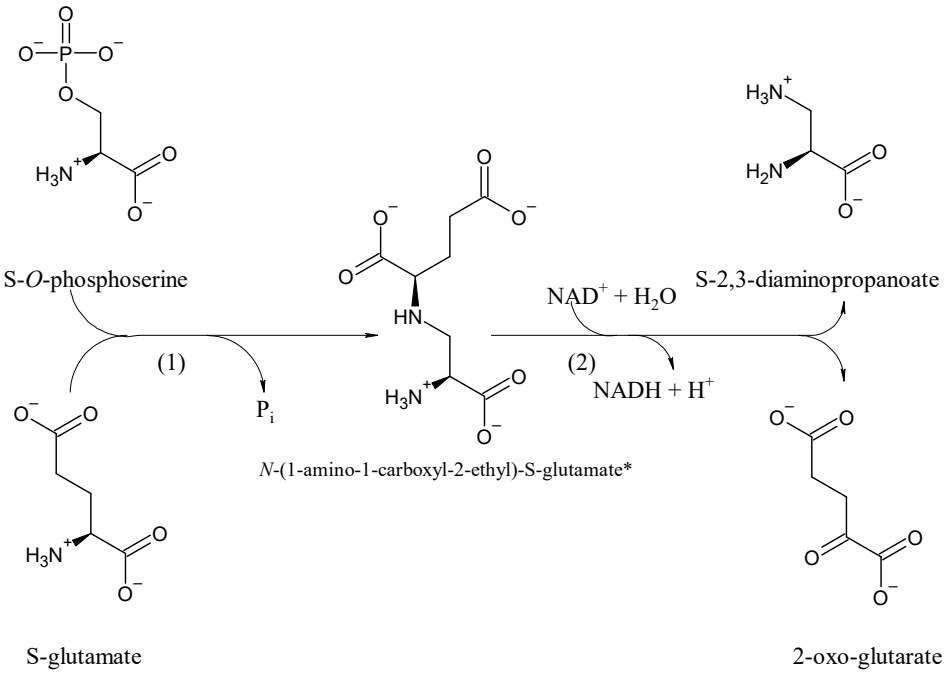


Figure 3.5. Biochemical pathway to 2,3-diaminopropanoate in *Staphylococcus aureus.* Redrawn from Kobylarz et al. (2014). (1) SbnA (*N*-(2-amino-2-carboxyethyl)-S-glutamate synthase, EC 2.5.1.140) uses PLP and the substrates S-*O*-phosphoserine (OPS) and S-glutamate to form the serine-glutamate conjugate: *N*-(1-amino-1-carboxyl-2-ethyl)-S-glutamate (Kobylarz et al. (2014); syn *N*-[(2S)-2-amino-2-carboxylethyl]-S-glutamate, ACEGA*). (2) SbnB (*N*-[(2S)-2-amino-2-carboxyethyl]-S-glutamate dehydrogenase, EC 1.5.1.51) oxidatively hydrolyses ACEGA, in an NAD^+-dependent reaction, to yield 2,3-diaminopropionate, 2-oxo-glutarate and NADH (Kobylarz et al., 2014). The nomenclature is from www.brenda-enzymes.org. The charged forms of the amino acids are those predominating at physiological pH values. The intermediate may or may not be released during the reaction. It has been chemically synthesised (Hsu et al., 2020). Figure: Peter B. Nunn; used with permission.

3.4.6 Biosynthesis of peptide-bound BMAA

Although free forms of BMAA in cyanobacteria could be biosynthesised directly by a pathway analogous to that of Brenner et al. (2003), another possibility is that free BMAA is released from a polymeric structure by metabolic turnover of small molecules or macromolecule assemblies (Tripathi and Gottesman, 2016). Galantins and paenilamicins are examples of peptide-bound BMAA that might act as reservoirs for the neurotoxin and release it upon metabolic turnover (Nunn and Codd, 2017). These peptides are either hypothesised (galantins) or known (paenilamicins) to be synthesised via a NRPS mechanism (Müller et al.,

2014). Several classes of other cyanotoxins are produced via the same mechanisms (Moffitt and Neilan, 2004; Mbedi et al., 2005; Kellmann et al., 2008; Mihali et al., 2008) and it is possible that an as yet unidentified specialised metabolite cluster exists for cyanobacterial BMAA biosynthesis.

Peptide-bound BMAA occurs in the bacterium *Paenibacillus larvae* (Müller et al., 2014). Complex antibacterial, antifungal and cytotoxic peptide paenilamicins are biosynthesised by an elaborate hybrid NRPS/polyketide synthetase (PKS) system (*pam*), in which 2,3-DAP is methylated within the specialised cluster to form BMAA (Garcia-Gonzalez et al., 2014; Müller et al., 2014). The *pam* gene cluster has been found only in *P. larvae*. There are more than 200 characterised *Paenibacillus* species occurring naturally in the environment (Nunn and Codd, 2019). Paenilamicins are similar in structure to the antibiotic peptide galantin 1 (Shoji et al., 1975; Sakai and Ohfune, 1992), also isolated from *P. larvae* subsp. *pulvifaciens* (Sakai and Ohfune, 1990). It is hypothesised that both galantins and paenilamicins are biosynthesised by similar NRPS mechanisms, comprising a transmethylase, which produce peptide-bound BMAA (Nunn and Codd, 2017). However, the biosynthetic route to this peptide has not been investigated biochemically. Peptides larger than paenilamicins and galantins and insoluble in protein precipitants may exist in cyanobacteria, which would explain the bound BMAA found in pellet fractions (Nunn and Codd, 2017). To date, galantins and paenilamicins are the only peptides known to contain this neurotoxin.

Enzyme-bound 2,3-DAP occurs in the peptide antibiotic dapdiamides from *Pantoea agglomerans* CU0119 (Dawlaty et al., 2010). In *P. agglomerans*, dapdiamides are biosynthesised by an NRPS module comprising two genes for the production of 2,3-DAP (*ddaA* and *ddaB*) (Hollenhorst et al., 2010).

Another possibility is that BMAA biosynthesis could occur via the dehydration or desulfydration of a serine or cysteine residue, respectively, after insertion into a peptide chain by an NRPS module (Nunn and Codd, 2017). This would generate a dehydroalanine (DHA) residue, enclosed within a protein or peptide. The addition of ammonium to the double bond of the DHA residue would generate 2,3-DAP (Eiger and Greenstein, 1948), which, if *N*-methylated, could yield peptide-bound BMAA. If these hypothetical peptides were subject to metabolic turnover, free BMAA would be released intracellularly. Dehydroamino acids are abundant in cyanobacteria (Siodłak, 2015) and are chemically reactive (Humphrey and

Chamberlin, 1997). Available methods for obtaining peptide-bound DHA enzymatically have not been described, and it is currently not possible to bioinformatically test the feasibility of this pathway for the production of 2,3-DAP and/or BMAA in cyanobacteria.

3.5 Methodology

Methods are similar to/based on those of Chapter 2. A summary, highlighting the main differences/changes to the methodology already described, is presented here.

iTOL (Letunic and Bork, 2016) was used to re-annotate a previously published phylogeny of 130 high-quality cyanobacterial genomes of various culture types (Chapter 2).

From the protein sequence database at the NCBI (https://www.ncbi.nlm.nih.gov), amino acid sequences with the same enzyme nomenclature as in the BRENDA database were selected for each enzyme in the known pathways of 2,3-DAP and/or BMAA biosynthesis (Fig. 3.2-3.5) to build profile hidden Markov models (pHMMs) (Supplementary Section 7.1). The pHMMs for each enzyme were used to search proteins of the 130 cyanobacterial genomes for homologs using HMM search (HMMER package v.3.1b2, hmmer.org), with a default threshold (E-value ≤ 0.01). Where necessary to confirm absence, protein sequences for enzymes in the known pathways for 2,3-DAP and/or BMAA biosynthesis were also used as queries in BLAST searches (Altschul et al., 1997) against all cyanobacterial proteomes available at the NCBI.

To differentiate between enzymes present in the known pathways for 2,3-DAP and/or BMAA biosynthesis and homologous enzymes carrying out different functions, active site and substrate specificity searches were conducted for DUD, DHP, βUP, cysteine synthase, SbnA, SbnB, VioK, DAPAL, and RidA based on knowledge from the literature (Supplementary Section 7.2). To further aid in differentiation, gene trees were reconstructed for DUD, DHP, βUP, cysteine synthase, SbnA, SbnB, VioK, and DAPAL using IQ-TREE (Nguyen et al., 2014) (Supplementary Fig. S1-S8).

AntiSMASH (bacterial version, v.5.1.2) (Blin et al., 2019) was used to search for the *pam* gene cluster of *P. larvae* (Supplementary Table 19) in the 130 cyanobacterial species in the

dataset and in an additional 93 *Paenibacillus* spp. genomes (Supplementary Table S20). In two diatom genomes (Supplementary Table S21), sequenced from species previously shown to produce BMAA (Jiang et al., 2014; Réveillon et al., 2016), a combination of antiSMASH bacterial and plant (plantiSMASH) options were used to search for the *pam* gene cluster (Oliver et al., 2021).

3.6 Results and discussion

3.6.1 Cyanobacterial species encoding enzymes in the pathway to 2,3-DAP from uracil in *Albizia julibrissin*

Several enzymes can catalyse the first two reactions in the pathway to 2,3-DAP from uracil (Fig. 3.2) and further information is needed regarding their substrate specificity in order to test the feasibility of these reactions in cyanobacteria.

Our results indicate that NADPH-dependent DUD (EC 1.3.1.2) was not found in any of the 130 cyanobacteria in our dataset. Protein sequences with some similarity to DUD are present in 125 cyanobacterial species, however they are functionally annotated to enzymes other than DUD and similarity to a sequence model of known DUD is low (Supplementary Table S1). It is currently not possible to discern if these protein sequences have the same substrate specificity and/or are able to catalyse the synthesis of 5-amino-5,6-dihydrouracil from 5-aminouracil.

DHP belongs to the cyclic amidohydrolase family of enzymes, which also include allantoinase, dihydroorotase, hydantoinase, and imidase (Holm and Sander, 1997). Although these metalloenzymes possess similar active sites and may use analogous catalysis mechanisms (Huang, 2015), they have different substrate specificities and relatively low amino acid sequence identity (Hsu et al., 2010; Peng and Huang, 2014). Site-directed mutagenesis studies on a DHP from *Pseudomonas aeruginosa* have shown that histidine residues at positions 59, 61, 183, and 239 (H59, H61, H183, H239) and the aspartate residue at position 316 (D316) are essential for the assembly of the binuclear metal centre of the active site, whereas the tyrosine residue at position 155 (Y155), the serine residue at position

289 (S289), and the asparagine residue at position 337 (N337) are necessary for substrate-binding. These residues are conserved in all DHPs (Huang, 2015).

In our study, six cyanobacterial species appear to possess a gene putatively coding for DHP (EC 3.5.2.2): *Nodosilinea nodulosa*, *Lyngbya confervoides, Leptolyngbya ohadii* IS1, *Spirulina major*, *Desertifilum* sp. IPPAS B-1220 and *Chroogloeocystis siderophila* (Fig. 3.6). The catalytic site described is conserved in all amino acid sequences (Supplementary Table S2), except for the conserved serine residue that is substituted by threonine in five out of the six sequences functionally annotated as DHP.

Similarly to DHP, βUP (EC 3.5.1.6) belongs to a large class of amidohydrolases, including nitrilases, cyanide hydratases, aliphatic amidases and ureidohydrolases, that, although catalysing different reactions, have relatively low, but significant, amino acid sequence identity (Bork and Koonin, 1994; Novo et al., 1995). Studies on the crystal structure of *N*-carbamyl-R-amino acid amidohydrolase have indicated a conserved cysteine (C), glutamic acid (E), and lysine (K) residue that form a catalytic triad in the active site (Novo et al., 1995; Nakai et al., 2000; Walsh et al., 2001). βUP hydrolases require a Zn^{2+} ion as a catalytic cofactor (Walsh et al., 2001).

34 cyanobacterial species appear to have a putative gene for βUP (Table 3.1). All amino acid sequences from these species were functionally annotated as Zn-dependent hydrolases (Supplementary Table S3).

Table 3.1. Cyanobacterial species with a putative gene for β-ureidopropionase (βUP, EC 3.5.1.6). *Lyngbya confervoides* and *Leptolyngbya ohadii* were found to have two copies.

Species with a putative gene for β-ureidopropionase
Lyngbya confervoides
Chroococcidiopsis thermalis
Phormidium ambiguum
Leptolyngbya ohadii
Desertifilum sp. IPPAS B-1220
Planktothrix tepida
Cyanobacteria bacterium TMED229
Planktothrix serta
Spirulina major
Leptolyngbya sp. “hensonii”
Chroogloeocystis siderophila
Leptolyngbya boryana
Phormidesmis priestleyi
Spirulina subsalsa
Dactylococcopsis salina
Oscillatoriales cyanobacterium JSC-12
Phormidium tenue
Myxosarcina sp. GI1
Planktothrix paucivesiculata
Chamaesiphon minutus
Planktothrix sp. PCC 11201
Kamptonema formosum
Oscillatoria sp. PCC 6506
Roseofilum reptotaenium AO1-A
Nodosilinea nodulosa
Pseudanabaena biceps
Alkalinema sp. CACIAM 70d
Planktothrix agardhii
Planktothrix prolifica
Planktothrix rubescens
Planktothrix mougeotii
filamentous cyanobacterium ESFC-1
Aliterella atlantica
Tychonema bourrellyi

Six cyanobacterial species appear to have putative genes for both DHP and βUP. These are the same species, listed above, that have putative genes for DHP (Fig. 3.6). In the apparent absence of a gene for NADP-dependent DUD, this thus appears to be an unlikely route to BMAA in cyanobacterial species. Nevertheless, given the presence of many cyanobacterial protein sequences similar to DUD, biochemical studies on the feasibility of this pathway for BMAA biosynthesis in these six cyanobacterial species are merited.

3.6.2 Cyanobacterial species encoding enzymes in the pathway to 2,3-DAP from β-(isoxazolin-5-on-2-yl) in *Lathyrus sativus* seedlings

Cysteine synthase (EC 2.5.1.47) appears to be ubiquitous in cyanobacteria with 127 of the 130 species examined having two or more copies of a putative gene coding for the enzyme. In *Hydrocoleum* sp. CS-953, genetic evidence for the presence of this enzyme is lacking; this may be due to errors in sequencing, assembly, genome annotation and/or functional annotation. *O*-acetyl-S-serine-dependent enzymes, such as cysteine synthase, have conserved alanine, phenylalanine and glycine residues at positions 132 (A132), 152 (F152), and 185 (G185), respectively (Kobylarz et al., 2016). These residues are conserved in most cyanobacterial cysteine synthases (Supplementary Table S4).

Although S-2,3-diaminopropionate *N*-oxalyltransferase (EC 2.3.1.58) is included in enzyme databases such as BRENDA (Chang et al., 2021), no corresponding amino acid or nucleotide sequence is available. In the protein sequence database at NCBI (https://www.ncbi.nlm.nih.gov) four sequences are associated with EC 2.3.1.58, which are all functionally annotated as cysteine synthase A. S-2,3-diaminopropionate *N*-oxalyltransferase appears to have been described once in *Lathyrus sativus* (Malathi et al., 1970), but since then has not been purified or sequenced. Whether the enzyme is specific to *L. sativus*, or was mistakenly characterised, is undetermined. Nevertheless, until further data become available on the enzyme's substrate specificity, catalytic site and taxonomic distribution, no conclusion can be advanced regarding the use of this pathway (Fig. 3.3) for BMAA formation in cyanobacteria.

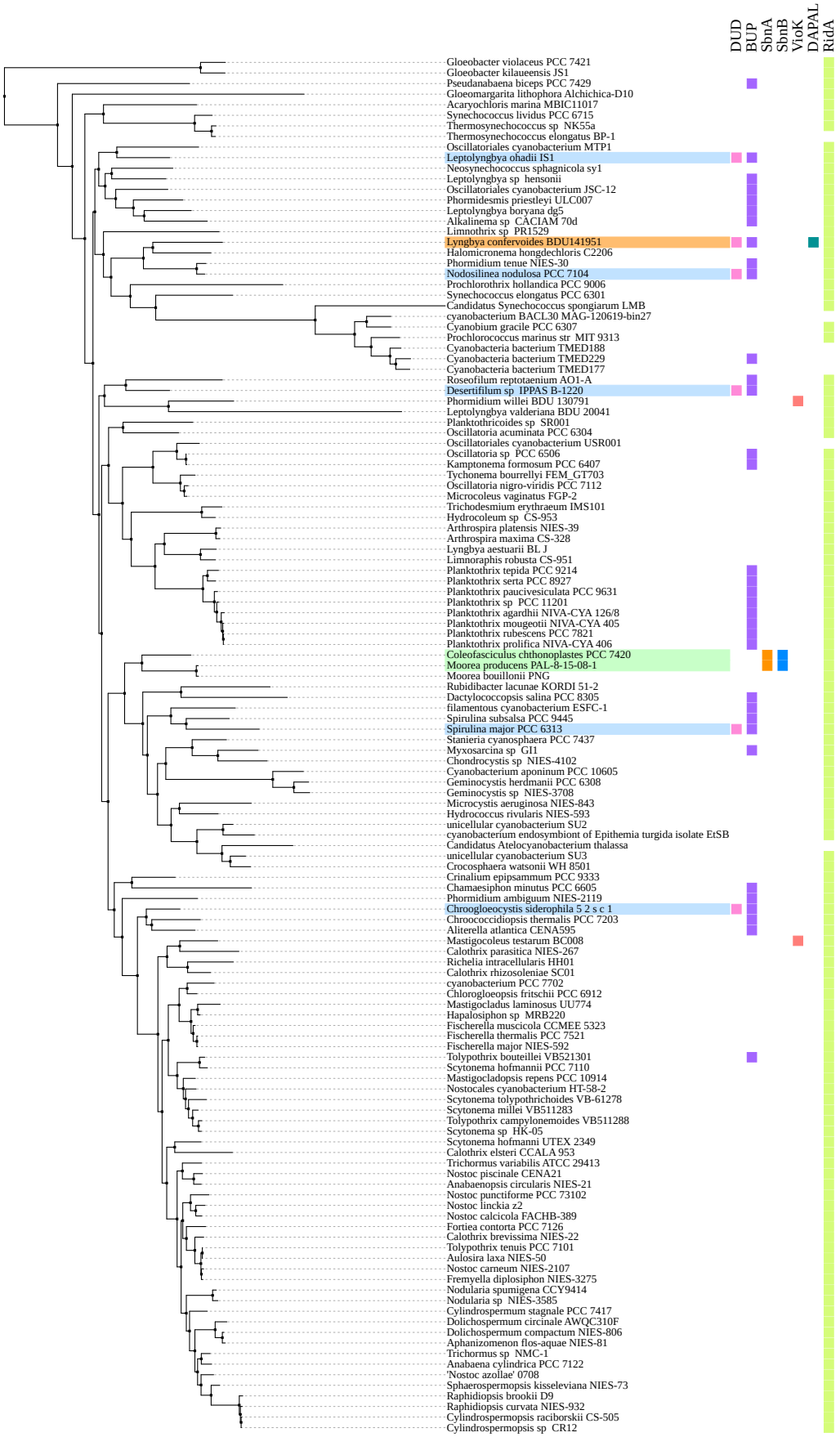
3.6.3 Cyanobacterial species encoding genes for 2,3-DAP biosynthesis via SbnA/SbnB and VioK/VioB

SbnA has been shown to be homologous to VioB and cysteine synthase, and SbnB to be homologous to VioK (Beasley et al., 2011). Despite their structural similarity, these enzymes catalyse different reactions. Bioinformatically, in order to understand whether cyanobacteria possess either gene (i.e. *sbnA/sbnB* and/or *vioK/vioB*), it is necessary to be able to distinguish between the pairs of homologous enzymes (i.e. SbnA/VioB/cysteine synthase and SbnB/VioK). Analysis of SbnA crystals from *Staphylococcus aureus* incubated with OPS revealed the structure of the intermediate *N*-(1-amino-1-carboxyl-2-ethyl)-S-glutamate (syn: *N*-(1-amino-1-carboxyl-2-ethyl)glutamic acid, ACEGA), allowing the identification of three conserved active site residues: arginine at position 132 (R132), tyrosine at position 152 (Y152) and serine at position (S185), essential for the highly specific OPS recognition and turnover and intermediate formation (Kobylarz et al., 2016).

Site-directed mutagenesis studies by Kobylarz et al. (2016) showed that, in addition to R132, Y152 and S185, substrate specificity also required the positively-charged residues lysine at position 100 (K100), and arginine at position 224 (R224), thought to be responsible for S-glutamate binding. A four amino acid insertion: glycine at position 126 and 127 (G126, G127), tyrosine at position 128 (Y128), and leucine at position 129 (L129) was also shown to be moderately conserved in SbnA and in homologs that biosynthesise 2,3-DAP (Kobylarz et al., 2016). The substrate specificity of SbnB is still uncertain (Beasley et al., 2011). Nevertheless, it was shown to differ from ornithine cyclodeaminases (including VioK) and alanine dehydrogenases as its active site has been expanded to accommodate a larger substrate than ornithine or alanine (Kobylarz et al., 2014). This is evident by the presence of an arginine at position 94 (R94), forming a salt bridge to the ACEGA terminal carboxylate (Kobylarz et al., 2014).

Two cyanobacterial species appear to encode both SbnA and SbnB: *Coleofasciculus chthonoplastes* and *Moorea producens* (Fig. 3.6). In these species, the corresponding catalytic residues characteristic of SbnA (i.e. K100, G126, L129, R132, Y152, S185, and R224) and SbnB (i.e. R94) are conserved (Supplementary Table S5-S6). Other protein sequences, functionally annotated as cysteine synthase family proteins, include the corresponding residues K100, G126, R132, and S185 (Supplementary Table S5). The

possibility exists that these enzymes perform the same catalytic functions as SbnA. No species appear to have orphan copies of *sbnA* or *sbnB* (i.e. encoding SbnA but not SbnB, or *vice-versa*).



Species encoding two or more enzymes from a pathway to 2,3-DAP/BMAA

- Biochemical pathway to 2,3-DAP from uracil in Albizia julibrissin/Biochemical pathway to the degradation of 2,3-DAP
- Biochemical pathway to 2,3-DAP from uracil in Albizia julibrissin
- Biochemical pathway to 2,3-DAP in Staphylococcus aureus

Figure 3.6. Species phylogeny showing the cross-species distribution of key enzymes involved in the biosynthesis of S-2,3-diaminopropanoate and BMAA (Previous page). The enzymes encoded by each species are indicated by coloured rectangles, next to the species name. DUD: dihydropyrimidine dehydrogenase; βUP: β-ureidopropionase; DAPAL: 2,3-diaminopropionate ammonia-lyase; RidA: reactive intermediate deaminase A. For the accession numbers of the enzymes, see Supplementary Tables S1-S6, S9, S15-S16. This is a re-annotation of the phylogeny of Chapter 2, who also provide bootstrap support (Appendix A, Supplementary Fig. S2).

In the cyanobacterium *Coleofasciculus chthonoplastes,* putative *sbnAB* appear together and are co-localised within an NRPS/PKS1 cluster (Fig. 3.7). The product of this cluster is unknown, but the complex resembles the NPRS assembly responsible for synthesising the siderophore amonabactin P 750 in *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966 (Barghouthi et al., 1989). Given that *sbnAB* are involved in the biosynthesis of staphyloferrin B in *Staphylococcus aureus*, and that the NRPS cluster containing these genes in cyanobacteria is similar to that of a siderophore-producing cluster in *Aeromonas hydrophila*, we hypothesise that, in some cyanobacteria, 2,3-DAP is probably involved in the production of iron-chelators. Our previous study indicated the same physiological role for 2,4-DAB, a structural isomer of BMAA (Chapter 2). Nevertheless, in the absence of a nearby methyltransferase, a feasible route to BMAA is lacking in *Coleofasciculus chthonoplastes.* In the cyanobacterium *Moorea producens,* putative *sbnAB* appear adjacent on the genome, but are not co-localised within a specialised metabolite cluster, and do not have an adjoining methyltransferase (Fig. 3.7). Nearby genes are not suggestive of any specific biosynthetic pathway, indicating multiple and different roles of 2,3-DAP in cyanobacteria.

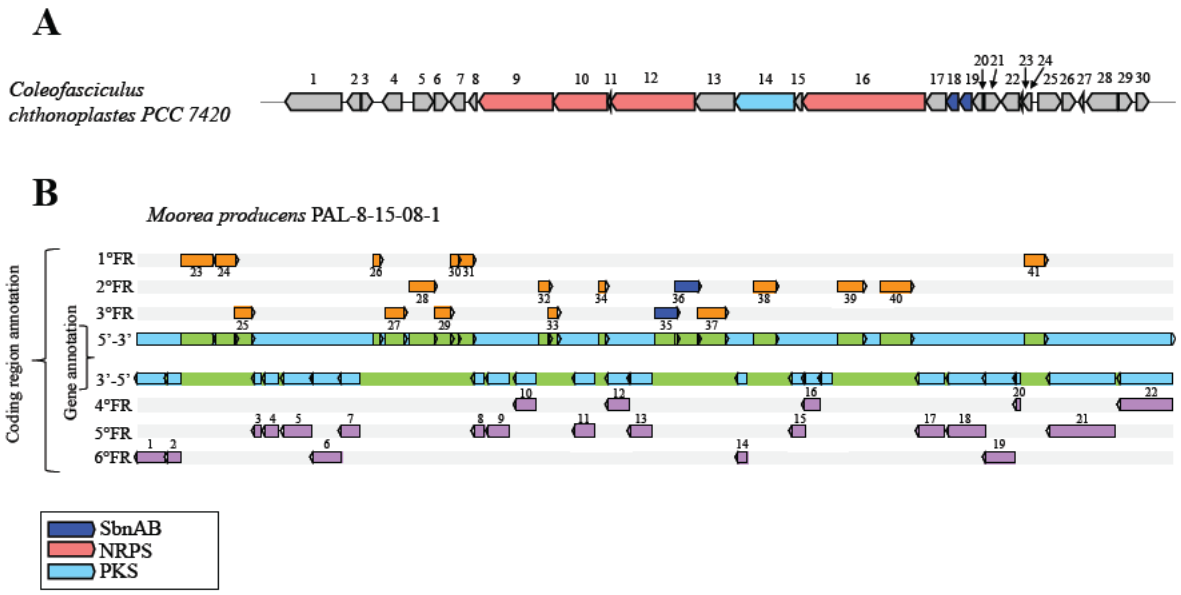


Figure 3.7. Localisation and characterisation of the genomic neighbourhood surrounding the enzymes SbnA and SbnB in two cyanobacterial species. (A) In *Coleofasciculus chthonoplastes* PCC 7420, SbnAB (EC 2.5.1.140, 1.5.1.51) (in dark blue) are co-localised within a hybrid NRPS/PKS cluster, the product of which is currently unknown. The gene cluster was adapted from antiSMASH (v.5.1.2). (B) In *Moorea producens* PAL-8-15-08-1, SbnAB are found adjacent to each other on the genome, but are not co-localised within a specialised metabolite cluster. SbnA and SbnB appear on the third and second frames of translation (FR), respectively. The genomic neighbourhood was adapted from Artemis Comparison Tool (ACT; Carver et al, 2005). For protein accession numbers and functional annotations, see Supplementary Table S22. Arrows indicate the orientation of transcription.

Moorea producens and *Coleofasciculus chthonoplastes* are in the same clade of the cyanobacterial species phylogeny (Fig. 3.6). Both are marine species, prone to forming thick mats and blooms (Siegesmund et al., 2008; Engene et al., 2012). Our results suggest that, although never detected/quantified, 2,3-DAP can potentially be biosynthesised in cyanobacteria through a pathway similar to that used by *Staphylococcus aureus* for the synthesis of SB. This pathway to 2,3-DAP does not appear to be widespread in cyanobacteria.

Homologs of SbnA and SbnB have also been described in a novel NRPS cluster responsible for poly(S-diaminopropanoic acid) synthesis in *Streptomyces albulus* PD-1 (NjxA and NjxB) (Xu et al., 2015), and in the gene cluster for sulfazecin in *Pseudomonas acidophila* (SulG and SulH) (Li et al., 2017) (note, diaminopropionic acid is a valid synonym of diaminopropanoic

acid). In *Streptomyces albulus* PD-1, NjxA and NjxB, homologous to cysteine synthase/serine dehydratase and OCD respectively, use S-serine and S-ornithine as substrates and lack the conserved residues indicated in SbnA and SbnB (Supplementary Table S7 – S8). From the description of Xu et al. (2015), it appears that NjxA and NjxB are more structurally similar to VioK and VioB, than to SbnA and SbnB. The substrate specificities of SulG, homologous to OCD, and SulH, homologous to cysteine synthase, have not been characterised. However, they too lack the conserved residues encoded in SbnA and SbnB (Supplementary Table S7 – S8). It is possible that cyanobacterial homologs of SbnA and SbnB, with distinct amino acid residues at the catalytic site, are still capable of synthesising 2,3-DAP.

In *S. albulus* PD-1, the genes responsible for 2,3-DAP biosynthesis (NjxA nd NjxB) were found to be adjacent on the genome, but not co-localised within poly(S-diaminopropionic acid) synthetase gene clusters (Xu et al., 2015). The possibility that putative SbnA and SbnB in the cyanobacterium *Moorea producens* are responsible for the biosynthesis of 2,3-DAP for use in an unidentified, specialised metabolite cluster cannot be disregarded.

In contrast to SbnA and SbnB, to date, the active sites of VioK and VioB have not been characterised. Provisionally, VioB and VioK may be distinguished from their homologs SbnA and SbnB by functional annotation in bioinformatics databases. VioK can also be differentiated from SbnB by excluding homologous sequences encompassing active site residues characteristic of SbnB (Kobylarz et al., 2014). Given that VioB uses nitrogen atoms from ammonium as the nucleophile, while cysteine synthase uses sulfur atoms from hydrogen sulfide (Thomas et al., 2003), and that VioB uses OAS as substrate whilst SbnA uses OPS, it is not expected that these enzymes (i.e. VioB/cysteine synthase and VioB/SbnA) share identical catalytic sites (i.e. VioB sequences are not expected to contain residues A132, F152 and G185, characteristic of cysteine synthases, nor residues R132, Y152, and S185, characteristic of SbnA). The gene product of *vioB*, S-2,3-diaminopropionate synthetase, does not have an assigned enzyme commission (EC) number and searches in protein sequence databases yield no results with the same name. Until the active sites of VioK and VioB are described, the possibility that these enzymes share the same catalytic residues as SbnB and SbnA/cysteine synthase, respectively, cannot be excluded.

Two cyanobacterial species appear to have *vioK* genes, coding for ornithine cyclodeaminase: *Phormidium willei* and *Mastigocoleus testarum* (Supplementary Table S9). The putative gene was not included in a specialised metabolite cluster, and no genes functionally annotated to a methyltransferase were found in its genetic neighbourhood (Fig. 3.8). Although *vioK* was found in two cyanobacteria, it is not accompanied by *vioB*. It is unclear whether the species encoding a gene for OCD have developed new physiological roles for the protein, or if a gene coding for S-2,3-diaminopropionate synthetase was lost from these species. In the apparent absence of VioB and given the seeming restricted distribution of OCD, the production of 2,3-DAP, and hence, of BMAA in cyanobacteria is unlikely to occur via a pathway analogous to that present in *Streptomyces* and *Clostridium*.

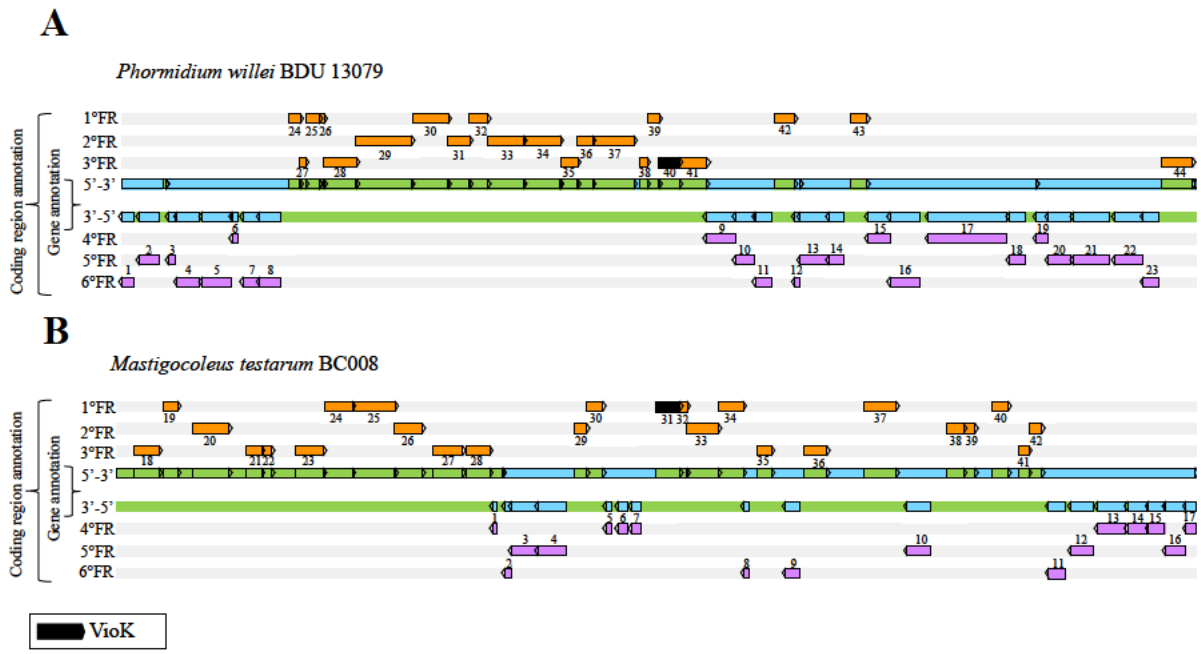


Figure 3.8. Localisation and characterisation of the genomic neighbourhood surrounding the enzyme VioK in two cyanobacterial species. VioK is present in both *Phormidium willei* BDU 13079 (A) and *Mastigocoleus testarum* BC008 (B), but is not co-localised within a specialised metabolite cluster. (A) VioK appears on the third frame of translation (FR), adjacent to a protein functionally annotated as FAD-dependent oxidoreductase. (B) VioK appears on the first frame of translation (FR), adjacent to a protein functionally annotated as (2Fe-2S)-binding protein. The genomic neighbourhood was adapted from Artemis Comparison Tool (ACT). For protein accession numbers and functional annotations, see Supplementary Table S23. Arrows indicate the orientation of transcription.

From the cyanobacterial species that appear to have either *sbnAB* or *vioK* (Fig. 3.6), only *Coleofasciculus chthonoplastes* was sequenced from an axenic monocyanobacterial culture (Chapter 2, Supplementary Table S6). It is believed that this species can biosynthesise 2,3-DAP through an analogous pathway to that of *Staphylococcus aureus*, since contamination by other potential 2,3-DAP-producing bacteria can be excluded. Results from the remaining species would have to be verified in genomes sequenced from axenic cultures.

3.6.4 The search for the *pam* gene cluster in cyanobacterial, *Paenibacillus*, and diatom species

The *pam* gene cluster, encoding a bound form of BMAA, has been found in *Paenibacillus larvae*. The gene products of *pamS* and *pamR* are responsible for the biosynthesis of 2,3-DAP, the methylation of which within the metabolite cluster could form BMAA (Müller et al., 2014). Since in the similar complex peptide galantin 1, only one of the two 2,3-DAP residues is methylated, and in the paenilamicins, both 2,3-DAP residues are methylated, it is possible that the *pam* gene cluster first assembles 2,3-DAP and then methylates it within the complex.

Although it was not specified by Kobylarz et al. (2014), our results show that the 2,3-DAP constituent of paenilamicin is probably derived from the concerted action of genes homologous to *sbnAB*, as the active sites of both SbnA and SbnB are conserved in PamR and PamS (Supplementary Table S7-S8). In the peptide antibiotic dapdiamide from *Pantoea agglomerans*, the genes responsible for 2,3-DAP production (*ddaA* and *ddaB*) are also homologs of *sbnA*/*sbnB*, with conservation of the enzyme's catalytic site (Supplementary Table S7-8). It appears that, when present, 2,3-DAP, and possibly BMAA, are most likely derived from the action of *sbnAB* rather than of *vioB/vioK*.

Our results confirm that the paenilamicin gene cluster occurs in 10 *Paenibacillus larvae* genomes (Supplementary Table S10). However, evidence for the presence of the hybrid *pam* gene cluster was not found in any of the 130 cyanobacterial genomes included in our dataset (Supplementary Table S11), or in an additional 93 *Paenibacillus* genomes other than *P. larvae* (Supplementary Table S12), or in the genomes of two BMAA-producing diatom species, *Thalassiosira pseudonana* CCMP1335 and *Phaeodactylum tricornutum* CCAP 1055/1 (Jiang et al., 2014; Réveillon et al., 2016) (Supplementary Tables S13-S14).

In the cyanobacterial species examined, the *pam* gene cluster does not appear to be a viable route of BMAA biosynthesis. So far, this specialised cluster appears to be restricted to *Paenibacillus larvae*, suggesting that paenilamicin production is likely not a widespread environmental source of the neurotoxin BMAA. Given the limited number of published diatom genomes, it cannot be concluded that the *pam* gene cluster is absent in all such species. However, our results indicate that, when BMAA is present in diatoms, it is probably not derived from paenilamicin-like peptides. The reduced number of specialised secondary clusters in the diatom species investigated (Supplementary Table S13-S14), and the absence of genes homologous to those implicated in 2,3-DAP synthesis suggest that, in these species, BMAA is probably either acquired from other sources (i.e. not by direct biosynthesis by the diatoms themselves), or that additional unknown routes to BMAA biosynthesis exist.

It is intriguing that a metabolically expensive and complex peptide such as paenilamicin can apparently only be found in *P. larvae*. It may be that the *pam* gene cluster represents a specialised version of a more general mechanism. However, although *Paenibacillus* spp. other than *P. larvae* do not appear to synthesise BMAA via the *pam* gene cluster, many species, including those isolated from human cerebrospinal fluid (Hehnly et al., 2020), appear to be able to synthesise other compounds, such as ectoines, derived from the BMAA structural isomer 2,4-DAB (Nunn and Codd, 2017; Mantas et al., 2021; Supplementary Table S12). The neurological significance of the widespread occurrence of *Paenibacillus* spp. in milk and prepared foods merits investigation (Nunn and Codd, 2019).

3.6.5 A potential explanation for the apparent absence of 2,3-DAP in cyanobacteria

2,3-DAP expresses severe cytotoxicity in some cell types, and many organisms have developed strategies that prevent endogenous accumulation of specific amino acids (Ernst et al., 2016). DAPAL (EC 4.3.1.15) is a prokaryotic type II PLP-dependent enzyme that catalyses the degradation of R- and S-forms of 2,3-DAP to 2-aminoacrylate and ammonium (Bisht et al., 2012) (Fig. 3.9). 2-aminoacrylate is a three-carbon reactive enamine intermediate synthesised by several PLP-dependent enzymes (Downs and Ernst, 2015). This compound is formed in the reaction catalysed by VioK/VioB; however, there it is contained, bound to a specific position within the enzyme's catalytic site. 2-aminoacrylate is susceptible

to tautomerisation and spontaneous hydrolysis in solvent water, releasing ammonium and pyruvate (Datta and Bhadra, 1978). It can also react with other compounds, leading to covalent modification and the inactivation of enzymes involved in essential metabolic processes (Flynn and Downs, 2013; Downs and Ernst, 2015) (Fig. 3.9).

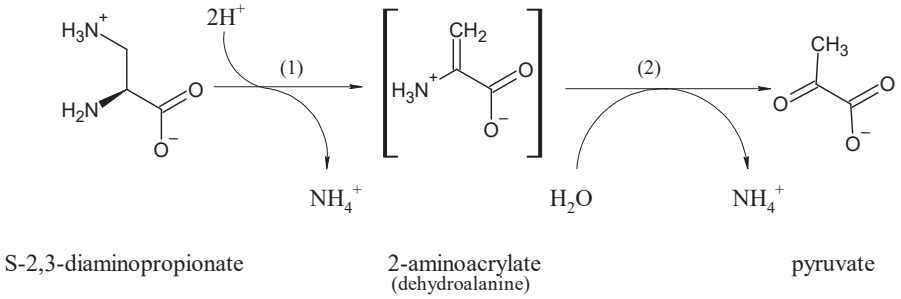


Figure 3.9. Pathway of the degradation of 2,3-diaminopropanoate to pyruvate and ammonium. Redrawn from Bisht et al. (2012) and Lambrecht et al. (2012). The nomenclature is from www.brenda-enzymes.org; Enzyme Commission numbers are as follows: (1) 2,3-diaminopropionate ammonia-lyase (DAPAL): EC 4.3.1.15; (2) reactive intermediate deaminase A (RidA): EC 3.5.99.10. The charged forms of the amino acids are those predominating at physiological pH values. Figure: Peter B. Nunn; used with permission.

DAPAL catalyses analogous reactions and is known to share significant amino acid sequence identity with type II PLP-dependent enzymes (Bisht et al., 2012), especially with the biosynthetic serine/threonine dehydratase/deaminase (TDH, EC 4.3.1.19), where sequence similarity can be up to 55% (Khan et al., 2003). The active site of DAPAL has been described in *Escherichia coli* K-12 (*Ec*DAPAL) (Bisht et al., 2012) and *Salmonella enterica* serovar Typhimurium LT2 (*St*DAPAL) (Nagasawa et al., 1988). In *Escherichia coli*, the residues lysine at position 77 (K77), aspartic acid at positions 120 and 189 (D120, D189), and tyrosine at position 168 (Y168) are important for catalytic function (Bisht et al., 2012). K77 and D120 are potentially responsible for the abstraction of protons from the R- and S-isoforms of 2,3-DAP, respectively (Bisht et al., 2012). All residues contained within the catalytic domain of TDH are conserved in *Ec*DAPAL (Uo et al., 2002).

The similarities between DAPAL and other PLP-dependent enzymes, the fact that DAPAL exhibits amino acid sequence dissimilarity in closely related species (Khan et al., 2003; Kalyani et al., 2012; Deka et al., 2018), and that the structure of DAPAL has not been

characterised in cyanobacteria, hinder the search and distinction of DAPAL homologs in the 130 cyanobacterial genomes analysed. However, an enzyme functionally annotated as DAPAL and exhibiting amino acid residues similar to *Ec*DAPAL is present in the cyanobacterium *Lyngbya confervoides* (Fig. 3.6). Additionally, 108 cyanobacterial species were found to encode enzymes functionally annotated to biosynthetic threonine ammonia-lyases (Supplementary Table S15). Residue D120 is not conserved in any of the cyanobacterial homologs evaluated, including in *L. confervoides,* suggesting a different mode of abstraction of C1 protons from the S-DAP isoform, or loss of function. Given the observed similarities between the catalytic cleft of *Ec/St*DAPAL and cyanobacterial biosynthetic threonine ammonia-lyases (Supplementary Table S15), it is possible that the ability to degrade 2,3-DAP into 2-aminoacrylate and ammonium is widespread in cyanobacteria. Purification and molecular characterisation studies of cyanobacterial DAPAL are needed to clarify the prevalence and action of the enzyme in these species.

In this pathway (Fig. 3.9), RidA (EC 3.5.99.10) is required to quench 2-aminoacrylate, preventing enamine stress (Downs and Ernst, 2015). RidA is apparently present across all life, with the vast majority of free-living organisms encoding at least one RidA homolog (Irons et al., 2020). It has a defined role as a 2-aminoacrylate-stress modulator, responding to endogenously generated reactive metabolites, by facilitating and enhancing the rate of hydrolysis of the enamine or its tautomer to pyruvate and ammonium, thereby detoxifying 2-aminoacrylate and averting metabolic imbalance (Lambrecht et al., 2012; Lambrecht et al., 2013; Downs and Ernst, 2015). RidA proteins appear to play diverse, but important, molecular functions, many of which are still poorly understood (Oka et al., 1995; Goupil-Feuillerat et al., 1997; Asagi et al., 1998). Crystal structure studies in *Escherichia coli* identified a signature of five conserved amino acids (Liu et al., 2016), although only a highly conserved arginine at position 105 (R105) was shown to be strictly necessary and sufficient for biochemical activity (Burman et al., 2007; Lambrecht et al., 2012). The residues glutamic acid at position 120 (E120), tyrosine at position 17 (Y17), serine at position 30 (S30), and asparagine at position 88 (N88) were also shown to be important for RidA catalytic activity (Burman et al., 2007; Lambrecht et al., 2012; Degani et al., 2018).

Putative RidA appears to be omnipresent in cyanobacteria. 4275 homologs of Rid family proteins were found, 611 of which have a conserved arginine residue. Only seven cyanobacterial species appear to lack a RidA homolog with a conserved arginine at the

catalytic site (Supplementary Table S16) (Fig. 3.6). There are several species encoding homologs with full amino acid sequence conservation at the active site. Many of these homologs are annotated as hypothetical proteins, and less frequently as endoribonuclease L-PSP (liver-perchloric acid-soluble protein). Apart from the active site, amino acid sequence identity was low between cyanobacterial homologs and those from other bacteria, yeast and mammals. This is consistent with the data from Mistiniene et al. (2003) and Parsons et al. (2003) who found amino acid sequence identity to vary between RidA homologs from different origins, with some members sharing <8% sequence similarity. The observed sequence diversity and variation in functional annotation corroborates the premise that RidA may execute myriad different roles (Niehaus et al., 2015).

Our results show that the failure to detect free 2,3-DAP in cyanobacteria may be explained by the presence of DAPAL and RidA. DAPAL would catalyse the synthesis of 2-aminoacrylate and ammonium from 2,3-DAP (Fig. 3.9). If free 2,3-DAP and the ammonia-lyase were present in cyanobacteria, RidA would be required to quench 2-aminoacrylate, preventing disruption of cellular functions (Downs and Ernst, 2015). The expression of RidA has been shown to be inversely correlated with iron availability (Irons et al., 2020). In conditions of iron limitation, cyanobacteria produce siderophores (Whitton, 2012). If cyanobacteria can synthesise 2,3-DAP-based iron-chelators, this would corroborate our hypothesis that, in some species, 2,3-DAP and hence BMAA, may be involved in siderophore formation.

3.6.6 Comparison between bioinformatics results and the analytical chemistry data from the literature

Following the methodology of Chapter 2, our bioinformatics findings are compared with those of 28 biochemical studies where evidence for BMAA-, or BMAA- and 2-4-DAB-producing cyanobacteria has been provided (Supplementary Table S17). A scenario where a particular species is shown to possess the necessary genes for BMAA biosynthesis, and is proven to produce the same compound via analytical studies, serves as a strong indicator that this species is a BMAA-producer. However, a species that lacks known genes required for BMAA formation, but is shown by analytical studies to produce the neurotoxin could be indicative of a false-positive result derived from an ambiguous analytical approach (especially probable if non-axenic cyanobacterial cultures and/or low-specificity analytical methods were employed), or suggests the existence of other, still undescribed routes for

BMAA biosynthesis. 24 of the 28 studies used variations of MS/MS for BMAA/2,4-DAB identification and quantification, and more than one analytical method was commonly used (either to test the accuracy and specificity of one analytical method compared to others, or to cross-check results). Only seven analyses for the neurotoxin itself were derived from all-axenic cultures, and some surveys involved a combination of axenic cultures, non-axenic monocyanobacterial cultures, and environmental samples. Others did not report on the culture type of the isolates studied. Our 130 cyanobacterial species genomic dataset and the compilation of BMAA production from the literature overlapped in seven species: *Prochlorococcus marinus, Planktothrix agardhii, Aphanizomenon flos-aquae, Cylindrospermopsis raciborskii, Nodularia spumigena, Microcystis aeruginosa* and *Microcoleus vaginatus.* None of these common bloom-forming species (Whitton, 2012) appear to have genes coding for relevant enzymes in the pathways to 2,3-DAP/BMAA described above (Fig. 3.2-3.5), or specialised gene clusters analogous to those of dapdiamides, galatins, or paenilamicin peptides. Apart from the *Microcystis aeruginosa* used by Downing et al. (2011), the studies that have used these species for the identification and quantification of BMAA employed analytical techniques other than MS/MS, and/or used non-axenic cyanobacterial cultures (i.e. non-axenic monocyanobacterial cultures and/or environmental samples) in their analyses. It cannot be excluded that these correspond to false-positive results as a consequence of the use of low-specificity analytical techniques and non-axenic cultures. However, it is also possible that cyanobacteria synthesise BMAA through a yet-undiscovered route or routes.

3.7 Summary and conclusions

Our bioinformatics survey has provided insight into potential pathways for the biosynthesis of 2,3-DAP/BMAA in cyanobacteria, which can lead to future experimental investigations. The bioinformatics results indicate potential 2,3-DAP biosynthesis in some cyanobacterial species via the staphyloferrin B pathway. The fact that, in *Staphylococcus aureus*, 2,3-DAP is synthesised via *sbnAB* for inclusion in siderophores, and that in one cyanobacterial species (*Coleofasciculus chthonoplastes*) putative *sbnA* and *sbnB* were found to be co-localised within an NRPS/PKS1 cluster similar to amonabactin P 750 from *Aeromonas hydrophila*, suggests a functional association with siderophore biosynthesis. 2,3-DAP also appears to be synthesised by homologs of *sbnAB* in the chemoheterotrophic bacteria *Paenibacillus larvae*

(*pamRS*) and *Pantoea agglomerans* (*ddaAB*), indicating that this is a more common route to free and bound 2,3-DAP, and potentially to BMAA, in prokaryotes (compared to the homologous *vioKB*).

A homolog of the *pam* gene cluster from *P. larvae* was not detected in any of the 130 cyanobacterial genomes in our analysis, two diatom species or *Paenibacillus* other than *P. larvae*. Other mechanisms may exist that remain to be discovered.

2,3-DAP has not been reported in cyanobacteria. We have shown that, in some species, the presence of putative DAPAL and RidA could explain the failure to detect this compound from cyanobacterial analytical results. RidA expression is upregulated during periods of iron scarcity, which corroborates our hypothesis that, in some cyanobacterial species, 2,3-DAP may be implicated in environmental iron-scavenging.

The existence of many homologs of enzymes potentially leading to 2,3-DAP and BMAA biosynthesis, and the presence of uncharacterised enzymes, with amino acid sequences and structural characterisation not yet available, hinder the search for possible pathways to 2,3-DAP and BMAA in cyanobacteria. Further information regarding the enzymes in the described pathways and their specificity is needed to further test our hypotheses. Also, analytical studies using MS/MS techniques and axenic cyanobacterial cultures would be helpful to ensure accurate and consistent results concerning the identification, origin(s) and quantification of BMAA. Ultimately, the ability to synthesise 2,3-DAP and BMAA does not appear to be universal among cyanobacteria. This suggests that either this compound is not widespread in cyanobacteria, or that there are additional and as yet unknown pathways for the synthesis of 2,3-DAP and BMAA.

Chapter 4

In cyanobacteria, is horizontal gene transfer more prevalent in extreme environments?

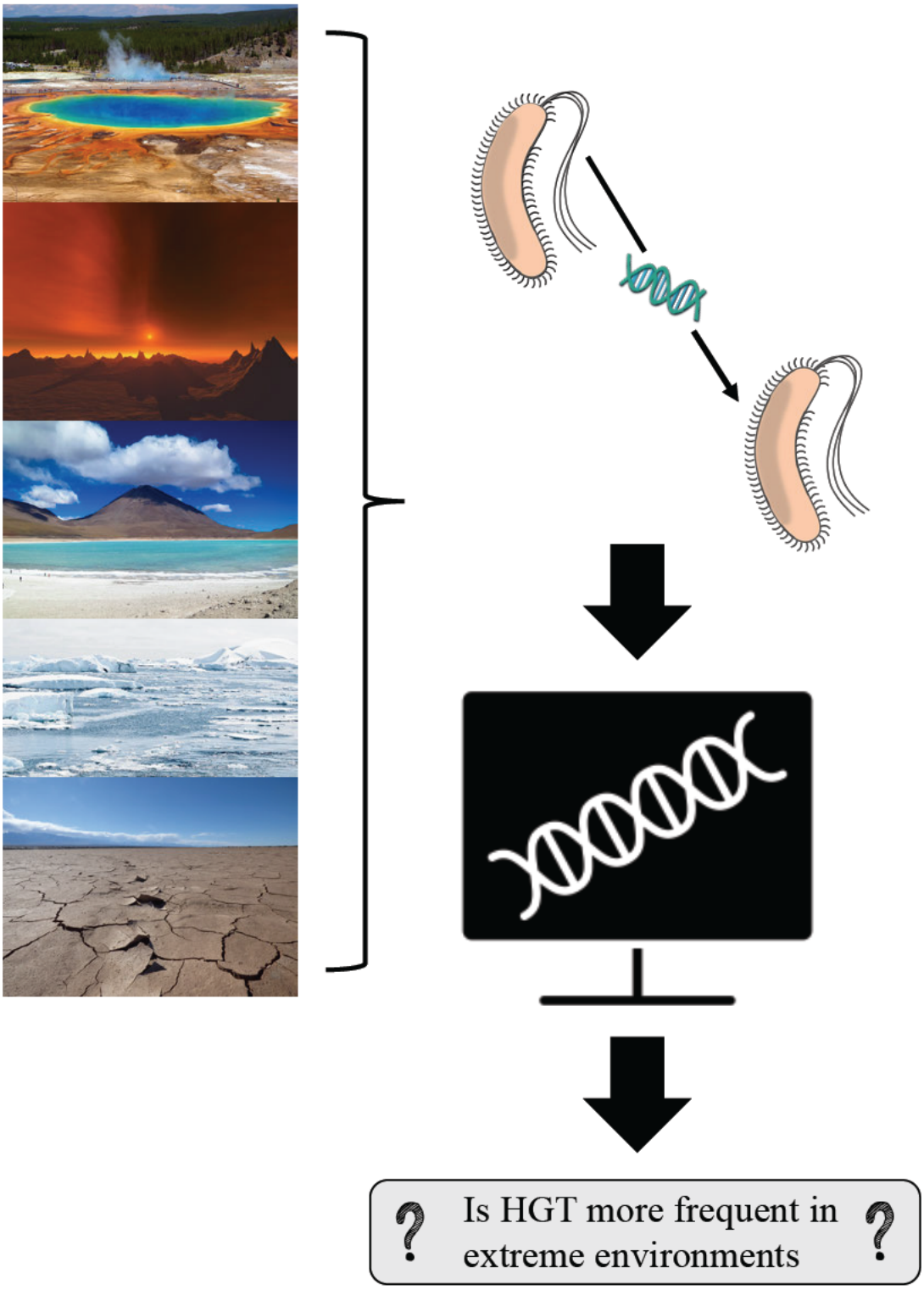
4.1 Preface

The work in this chapter is currently a manuscript potentially intended for publication in *Molecular Phylogenetics and Evolution* and the first-person plural is used throughout to maintain consistency. Ziying Ke performed the orthogroup delimitation. I performed all other analyses, wrote the first draft of the manuscript and supplementary materials, and produced all figures and tables.

Citation:

Mantas, M. J. Q., Maureen Stolzer, Ziying Ke, Catherine Kidner, Geoffrey A. Codd, Dannie Durand, Daniel Barker. In cyanobacteria, is horizontal gene transfer more prevalent in extreme environments? *In preparation*.

4.2 Graphical abstract



4.3 Introduction

HGT (also known as lateral gene transfer, LGT) is the acquisition and incorporation of foreign genes by a species. It can occur through three mechanisms: transformation, transduction and conjugation. Transformation is the process whereby prokaryotes take up free DNA from the surrounding environment. Transduction encompasses the transfer of genes between prokaryotic species via viruses. Conjugation occurs when plasmids are passed from one prokaryote to another through a tube-like structure called the pilus (Jain et al., 2002). HGT has been indisputably important in shaping prokaryotic evolution (Koonin et al., 2001; Wiedenbeck and Cohan, 2011), including in cyanobacteria (Kellmann et al., 2008; Leikoski et al., 2009; Bolhuis et al., 2010; Yerrapragada et al., 2009).

HGT rates may vary across environments, both in the frequency of HGT events and the frequency of fixation of those events. This potential association of HGT events and environmental context has been discussed in the literature (Schönknecht et al., 2013). Although not all studies classify environments in the same way, and our view of an environmentally cosmopolitan species' habitat may depend on the collection site of an isolate, the broad characteristics of an extreme environment are generally agreed. Extreme environments are recognised as hostile, unfavourable and even lethal for most organisms (Kristjansson and Hreggvidsson, 1995). They include environments with extremes of temperature, pH, salt concentration, pressure, radiation, desiccation, oxygen availability, as well as those where toxic waste, organic solvents and/or heavy metals are present (Seckbach and Oren, 2007). It is currently unclear whether the mechanisms and frequency of HGT vary depending on the environmental setting and research on this topic is scarce. It has been proposed that HGT is elevated in extreme environments (Chen et al., 2021; Li et al., 2014). Others studies suggest that the amount of HGT is negatively correlated with aspects characteristic of more extreme habitats (e.g. temperature) (Gophna, et al., 2015) and that HGT is less frequent in fluctuating environments (e.g. soil) (Goyal, et al., 2020). Makarova et al. (2011) implied a link between the abundance of "mobilome genes" (i.e. mobile elements, such as viral and transposon genes, within a genome and across genomes, by HGT) and specific environmental factors (i.e. temperature), without specifying a direct/broader link between HGT and environmental harshness.

Archaea and bacteria living in anaerobic and/or high-temperature conditions have been shown to share large numbers of genes, probably due to HGT (Fuchsman et al., 2017). Some traits important for extremophile survival (e.g. metal resistance genes, transport proteins, nutrient acquisition genes and symbiosis islands) are known to have been acquired by HGT (Bakermans, 2015). Habitats such as hot springs (Li et al., 2014), acid mine drainage (Guo et al., 2015), saline lakes (Li et al., 2014), Antarctic ice sheaths and lakes (Giménez et al., 2019), high mercury (Barkay et al., 2010) and oil wells (Fuchsman et al., 2017) have been proposed as potential hotspots for HGT. In dilute, aerobic, mesophilic environments, such as marine and freshwater surface waters, it has been suggested that HGT is a rarer event (Fuchsman et al., 2017; Li et al., 2014).

We investigate the association of HGT with environment among cyanobacteria. Cyanobacteria are a clade of photosynthetic bacteria found in a wide range of environments, including some of the most physically extreme habitats on Earth (Abed et al., 2009; Rampelotto, 2013; Seckbach, 2007; Whitton, 2012). Cyanobacteria in extreme environments are considered to be mostly extremotolerant (i.e. able to withstand extreme conditions, but growing optimally under more mesophilic niches) as opposed to extremophilic (i.e. reliant on extreme habitats to thrive) (Seckbach and Oren, 2007). Cyanobacteria occur in microbial mats in permafrost regions of Siberia and Antarctica (Vincent, 2007), as well as in continental hot springs (Brock, 2012; Ward et al., 2012) and other geothermal areas worldwide. They are found in hypersaline environments (Garcia-Pichel et al., 1998; Golubic, 1980) and in alkaline habitats, including soda lakes (Boussiba et al., 2000; Grant, 1986). Cyanobacteria are present in environments exposed to high levels of UV radiation (e.g. walls and pavements) (Seckbach and Oren, 2007), to high metal concentrations (Huertas et al., 2014) and in xerophilic conditions (Alwathnani and Johansen, 2011; Hagemann et al., 2017). Cyanobacteria are rarely found in acidic environments with a pH below 3 (Brock, 1973; Paerl et al., 2000) and barophiles have not been reported.

Any terrestrial environment may be classed as extreme for cyanobacteria. Compared to aquatic environments, temperatures in terrestrial habitats fluctuate more widely, the availability of water and nutrients is less stable, and exposure to radiation on rock and soil surfaces can be more intense (Kashyap et al., 2022). In line with this, it has been proposed that terrestrial cyanobacteria are subject to more HGT than their aquatic counterparts (Chen et al., 2021).

It is unclear whether HGT events are more or less frequent in extreme (or terrestrial) environments. This hypothesis has been tested in metagenomic samples (Li et al., 2014) and more recently in cyanobacteria (Chen et al., 2021). Li et al. (2014) concluded that microbial communities evolve faster in extreme environments, using, among other factors, the apparent higher frequency of HGT in extreme habitats as supporting evidence. Their study was based on 40 metagenomes from microbial communities sampled from six different environments ("saline lake", "acid mine drainage", "hot spring", "surface ocean", "freshwater" and "soil"). The relative abundance of transposase encoding genes was used as a proxy for HGT frequency and statistical evidence for varying transposase number in extreme *versus* benign environments was evaluated via a Mann-Whitney U-test. Chen et al. (2021), in a phylogenetic study of 650 Cyanobacteria and Melainabacteria (491 of which were classified as inhabiting "thermal springs", "terrestrial", "marine", "freshwater", "host-associated" or "others") concluded that HGT events were more frequent in terrestrial environments compared to their aquatic counterparts, proposing that these aid survival in fluctuating habitats. HGT events were inferred via sequence similarity-based HGT detection and COUNT software (Csűös, 2010; Zhu et al., 2014) and the association between HGT count and the environment was assessed using t-tests (Chen et al., 2021). Makarova et al. (2011), in a study of 383 bacterial and archaeal taxa, focused on defence systems, found a higher number of viral defence systems in thermophiles compared to mesophiles, which could imply that the frequency of HGT varies depending on environmental temperature.

Different insights were provided by Gophna et al. (2015) and Goyal et al. (2020). Gophna et al. (2015) analysed the activity of CRISPR-Cas in microbes and its link to recent HGT events and found a negative correlation between growth temperature and gene transfer, which they attributed to the lower genetic diversity in warmer environments. Gene acquisition via HGT was assessed in three different ways: as a fraction of prophage genes in bacteria, as a fraction of singletons in the ATGC database cluster of orthologous groups, and as a fraction of the recently acquired genes inferred on the basis of dinucleotide composition. Data from 1399 prokaryotic genomes with at least one proxy measure for HGT were available and statistical tests were performed using linear models. Goyal et al. (2020), in their theoretical study of a vector-based minimal model of a population evolving via mutation, HGT, selection and drift, concluded that HGT rates in nature are linked with environmental dynamics, being high in

"static environments" (e.g. marine habitats) and low in "fluctuating environments" (e.g. terrestrial habitats).

These existing studies diverge widely in (i) the taxa and number of taxa analysed; (ii) the methodology used; (iii) whether different types of HGT (i.e. transformation, conjugation or transduction) were prioritised (e.g. thermophilic bacteria belonging to the genus *Thermus* exhibit high competence for natural transformation (Averhoff et al., 2021)); and (iv) which environments were classified as extreme. Other factors, such as exposure to high levels of radiation, relevant to both extreme and non-extreme environments, have been shown to affect HGT (Levy et al., 1993). Population sizes (Edwards et al., 1999), geographic isolation (Ramette and Tiedje, 2007), genome composition (Azad and Lawrence, 2011), the presence of extracellular vesicles (Forterre, 2016) and methylation patterns (Beaulaurier et al., 2018) may also be relevant.

4.3.1 Our contribution

The aim of this study is to determine whether the number of HGT events differs among cyanobacterial species, depending on broad characteristics of the environments which the species inhabit. Here, we perform new tests of existing hypotheses using phylogenetic analyses of 130 cyanobacterial genomes. Central to our approach is gene tree-species tree reconciliation, for which we use Notung software (Stolzer et al., 2012). Notung incorporates costs for gene loss, duplication and horizontal transfer events, and seeks the reconciliation of overall minimum cost (i.e. maximum parsimony). Notung assumes HGT to be relatively rare (depending on HGT cost compared to duplication and loss costs), rejects reconciliation involving HGT events that cannot be temporally feasible, and can rearrange poorly supported nodes in the gene phylogeny to match the species phylogeny's topology. Notung has been widely used to infer HGT events in a wide range of taxa (Misner et al., 2015; Monteil et al., 2020; Ropars et al., 2015). Because of the difficulty of determining precisely which biotic and physicochemical aspects of the environment are relevant, we have used four different binary classifications. To correct for potential non-independence in cross-species analyses, we used phylogenetically corrected t-tests implemented in BayesTraits (Pagel et al., 2004) to test for a different per-species average number of HGT events across the categories of each classification. Our results do not support the hypotheses of increased HGT in extreme or terrestrial environments in cyanobacteria.

4.4 Methodology

4.4.1 Gene family phylogeny reconstruction

Genome-wide protein sets from 130 high-quality cyanobacterial genomes, whose species phylogeny was previously published (Mantas et al., 2021; Chapter 2), were compared using BLAST (Altschul et al., 1997) and clustered into orthogroups using OrthoFinder (v.2.1.2) (Emms and Kelly, 2015) with the Markov cluster algorithm (MCL) (van Dongen, 2000). An MCL inflation value of 1.6 was chosen on the basis of its performance on grouping members of a known gene family (Supplementary Section S1). This resulted in the identification of 17198 orthogroups, for each of which a protein multiple sequence alignment was generated using MAFFT (v.7.307) with the E-ins-I algorithm (Katoh and Standley, 2013). Gene trees were reconstructed using IQ-TREE (v.1.6.10) with the substitution model chosen according to Bayesian Information Criterion at the ModelFinder step (Kalyaanamoorthy et al., 2017). Branch support was estimated with ultrafast bootstrapping with 1000 replicates (Hoang et al., 2017).

4.4.2 Inference of HGT events

Reconstruction of duplication, loss and HGT events for each gene phylogeny with a minimum of 4 and a maximum of 2988 sequences was carried out using Notung (v.3.0_21-beta) (Stolzer et al., 2012). Given the rooted species tree (Mantas et al., 2021; Chapter 2), gene phylogenies were rooted with a duplication-loss (DL) model using default parameters (duplication cost of 1.5 and loss cost of 1.0). Rooted gene trees with weakly supported edges (i.e. bootstrap support below 75%) were rearranged with a DL model. Phylogenies were re-rooted as before and outgroup taxa were pruned (i.e. removed). Then reconciliation between gene and species trees was performed using a duplication-transfer-loss (DTL) model. The reconciliation was carried out three times in a grid search for the optimal horizontal transfer cost, with a transfer cost of 4, 8 and 15 and duplication and loss costs kept as default. The results were averaged over all equally parsimonious optimal solutions for each transfer cost. For downstream analyses, we chose the transfer cost that minimised genome size flux (i.e. difference in genome size between parent and child nodes (David and Alm, 2011)), giving a transfer cost of 4 (Supplementary Fig. S1-S3).

Orthogroups containing 2 and 3 sequences, exceptionally large gene families (our orthogroups 0, 1 and 2 with 6622, 3334, and 3149 sequences each) and those for which inference of HGT events is temporally infeasible (i.e. when the timing constraints imposed by the species tree cannot be compatible with the inferred transfers; e.g. transfers from descendant to ancestral nodes) were excluded from the analysis (Supplementary Tables S2-S4).

A heat map summarising the total number of transfers, summed over all gene families, between pairs of species was generated using the seaborn package in Python (Waskom, 2021) (Supplementary Fig. S4-S6).

4.4.3 Environmental classification

We devised four binary ecological classifications for the 130 cyanobacterial species, based on information present in the literature and reliable online sources (e.g. NCBI's BioSample database (https://www.ncbi.nlm.nih.gov/biosample), Catalogue of Microorganisms of the Biological Resource Centre of Institute Pasteur (https://catalogue-crbip.pasteur.fr) and Microbial Culture Collection at the National Institute for Environmental Studies (NIES Collection) (https://mcc.nies.go.jp) (Supplementary Table S5)).

We classified species by whether they inhabited "extreme" *versus* "non-extreme" habitats, in three different ways (classifications available in Supplementary Table S6); and whether they inhabit "aquatic" *versus* "terrestrial/variable" habitats (Supplementary Table S7). Given the difficulty in decisively assigning species to "extreme" *versus* "non-extreme" environments (Horikoshi et al., 2010), three binary classifications were created: "clearly extreme" *versus* "non-extreme"; "likely extreme" *versus* "non-extreme"; and "ambiguously extreme" *versus* "non-extreme". "Clearly extreme" includes species inhabiting indisputably extreme environments, such as hot springs, desert crusts, alkaline lakes and surface rocks in Antarctica. "Likely extreme" is a superset of "clearly extreme", also including species inhabiting environments including intertidal regions, soil samples and other habitats subject to diurnal, wide variations in temperature, UV radiation, humidity and salinity. "Ambiguously extreme" is a superset of "likely extreme", also including species inhabiting environments that could be considered extreme but where details are lacking in order to make a clearer decision (e.g. the gut of *Culex decens* and samples from sewage plants). Within each

of these three classifications, “non-extreme” encompassed the remaining species, hence the size of this category varied. However, in all classifications “non-extreme” included species inhabiting mesophilic, benign environments, such as freshwater lakes and marine habitats.

For the classification “aquatic” *versus* “terrestrial/variable”, species taken from water samples were classified as “aquatic”; species living in soil or heterogenous environments (including intertidal regions, bogs, marshes) were classified as “terrestrial/variable” (Table 1).

Table 4.1. Cyanobacterial species classification based on isolation source. 130 cyanobacterial species were classified according to the environment inhabited: extreme *versus* non-extreme and terrestrial/variable *versus* aquatic. See Supplementary Table S5 and Chapter 1, Fig. 1 for details regarding each species isolation source and geographic location.

Species counts: extreme *versus* non-extreme environments		
	Extreme	**Non-extreme**
Clearly extreme	25	105
Likely extreme	55	75
Ambiguously extreme	60	70
Species counts: terrestrial/variable *versus* aquatic environments		
	Aquatic	**Terrestrial/Variable**
	90	40

4.4.4 Statistical analysis

For the comparison of HGT events, for each species the sum of gene transfer events to the species (recipient events) and from the species (donor events) was used (i.e. total events) (preliminary statistical analyses conducted on recipient events only showed similar trends and statistical significance). After transformation (see below), this formed the dependent variable in statistical analyses, in which the binary classification of species was the independent variable. Rather than explicitly reconstruct ancestral habitats, we used a phylogenetic cross-species comparison of extant species.

Preliminary analyses were conducted using nonphylogenetic linear models implemented using the lm function in R, to select a transformation (logarithmic, square-root, cube-root or Box-Cox) of total HGT event data to increase the normality of residuals and homogeneity of variance for downstream analyses (Supplementary Fig. S7-S9, Table S8). Subsequent tests for statistical significance were performed on the square-root transformed data using the continuous random walk models of BayesTraits (v.3.0.1) (Pagel et al., 2004) to implement phylogenetically corrected t-tests for the four species classifications (Gardiner et al., 2008; Organ et al., 2007).

For each classification (i.e. "clearly extreme" *versus* "non-extreme"; "likely extreme" *versus* "non-extreme"; "ambiguously extreme" *versus* "non-extreme"; and "aquatic" *versus* "terrestrial/variable"), initial BayesTraits analyses were performed to select the parameters to include in the model. Uncorrelated and correlated models were fitted to the species phylogeny and total HGT data for extant species, differing in one free parameter: whether the correlation is constrained to zero (uncorrelated) or fitted to its maximum likelihood value (correlated). In each case (uncorrelated and correlated), all possible sets of tree transformation parameters κ, λ and δ (Pagel, 1994, 1997, 1999a) were used – that is, no transformation; κ alone; λ alone; δ alone; κ and λ simultaneously; κ and δ; λ and δ; and κ, λ and δ simultaneously. Although κ, λ and δ have interpretations relating to the tempo, mode, and phylogenetic signal of trait evolution, respectively, incomplete taxon sampling can affect their fitted values (Webster et al., 2003). Hence, we do not apply any biological interpretation. From these fitted models, the appropriate parameterisation was selected by Akaike Information Criterion (AIC). For "clearly extreme" *versus* "non-extreme", the κ transformation yielded the lowest AIC for both correlated and uncorrelated models; for "likely extreme" and "ambiguously extreme" *versus* "non-extreme" the κ, λ and δ transformation yielded the lowest AIC for both correlated and uncorrelated models; and for "aquatic" *versus* "terrestrial/variable", the κ and λ transformation yielded the lowest AIC for both correlated and uncorrelated models. Then for each of the four comparisons, BayesTraits analyses were performed to detect the direction and statistical significance of difference in overall HGT between the two groups. For each comparison, to ensure that the models to be compared were nested, the value of any included transformation parameter was restricted to a single value, the average of the maximum likelihood (ML) estimates from the initial analyses with and without correlation. For each of the four comparisons, a likelihood-ratio test (LRT)

(Wilks, 1938) was performed between correlated and uncorrelated models, comparing the likelihood ratio statistic $2\Delta\ell$ to a $\chi^2(1)$ distribution to obtain a *p*-value (Supplementary Table S9-S10).

4.5 Results and discussion

As known from previous studies (Zhaxybayeva et al., 2006), our results confirm that HGT occurs across the cyanobacteria (Supplementary Fig. S4-S6). As has been described in other bacterial species, HGT events in cyanobacteria appear to be most common among closely related species (Kloesges et al., 2011) (Fig. 1; Supplementary Fig. S10).

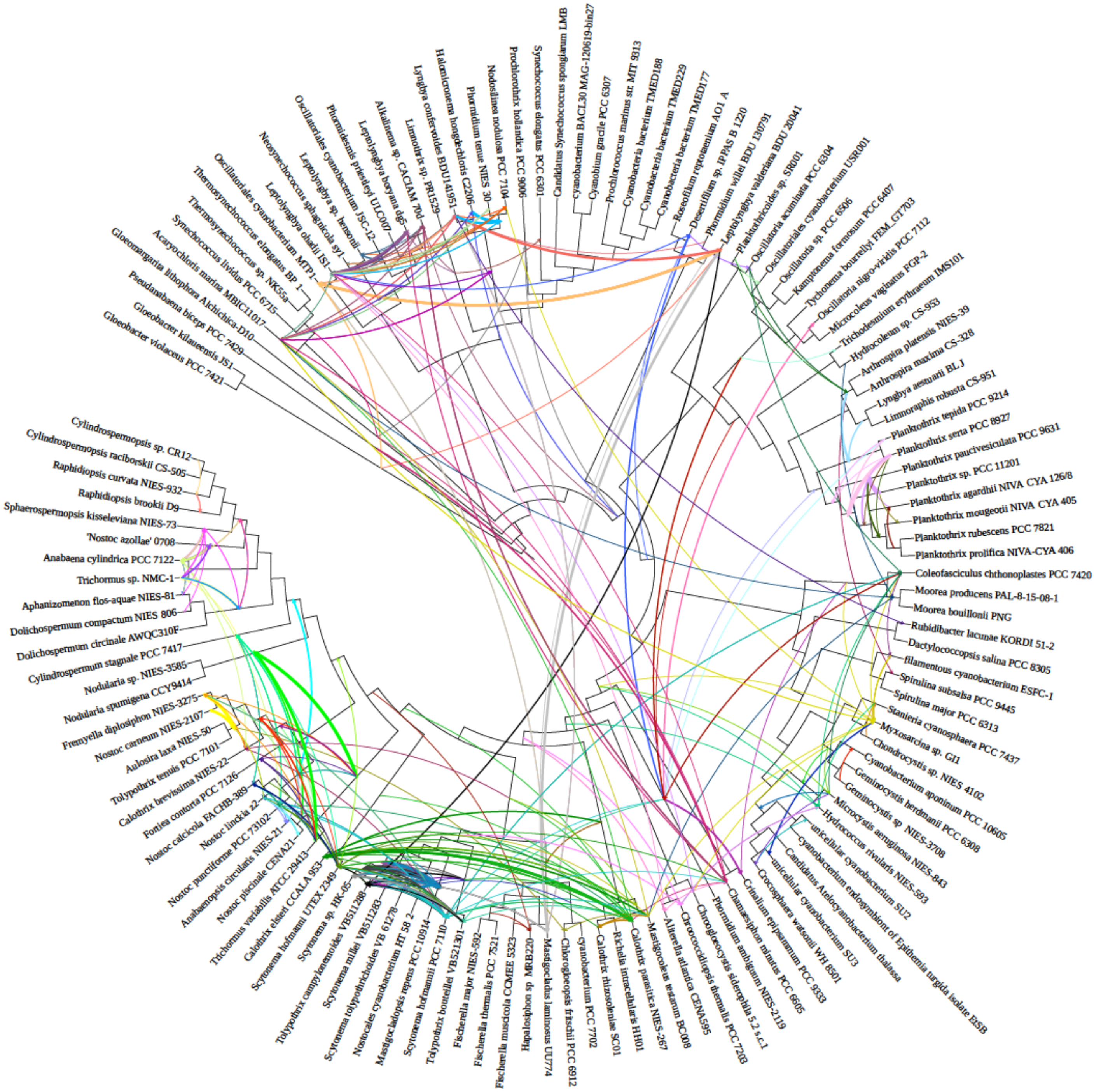


Figure 4.1. Cyanobacterial species phylogeny showing instances of horizontal gene transfer (HGT). Results pertain to gene tree-species tree reconciliation with a duplication-transfer-loss (DTL) model, with the parameter for cost of transfer set to 4. Only HGT events where ≥ 10 gene families were transferred are depicted. The width of the arrow corresponds to the number of orthogroups transferred. See Mantas et al. (2021) Supplementary Fig. S2 for bootstrap support of each branch.

Overall, we find no association between the prevalence of HGT and extreme or terrestrial habitats. Phylogenetically-corrected statistical tests of extant species suggest that HGT is *less* frequent in extreme environments, irrespective of the classification details, though with little or no statistical significance (Table 2; Supplementary Table S10). The only statistically significant result at the traditional cut-off ($p < 0.05$) is for the test of "clearly extreme" *versus* "non-extreme". However, the statistical significance is borderline ($p = 0.021$) and becomes non-significant if a Bonferroni correction is applied for running four statistical tests (i.e. $p > 0.013$). Terrestrial species or species living in variable habitats did experience more HGT than aquatic species, but without statistical significance ($p > 0.05$), even before adjustments for multiple testing (i.e. Bonferroni correction) (Table 2).

Table 4.2. Phylogenetically corrected statistical tests for non-equality of HGT extent between two groups. The environments compared are shown, along with the numerical encoding of the two categories. *r* is the correlation coefficient estimated using the correlated model in BayesTraits. The likelihood ratio statistic, $2\Delta\ell$, was converted to a *p*-value by comparison with a χ^2distribution (1 d.f.).

Comparison	***r***	$\boldsymbol{2\Delta\ell}$	***p***
Clearly extreme (1) *versus* non-extreme (0)	–0.120	5.286	0.021
Likely extreme (1) *versus* non-extreme (0)	–0.057	0.429	0.512
Ambiguously extreme (1) *versus* non-extreme (0)	–0.027	0.091	0.762
Terrestrial/variable (1) *versus* aquatic (0)	+0.038	0.190	0.663

Our findings are in contrast with previous studies that reported more frequent HGT in extreme/terrestrial environments (Chen et al., 2021; Li et al., 2014). Our analyses were carried out in cyanobacteria, allowing for direct comparison with the recent study of Chen et al. (2021). One possible reason for the difference in results may be statistical methodology. In the case of our HGT counts from Notung, the same data give different conclusions depending on whether the phylogeny is incorporated into the statistical test or not. When conducting our preliminary analyses of normality of residuals using linear models in R, we did not use the *p*-values to draw biological conclusions, because these ignore non-independence among related species. However, we note that these preliminary results (ignoring the lack of independence among closely related species) contradict our final results, which do take the phylogeny into consideration (Supplementary Fig. S7-S9, Table S8). This shows that the choice of statistical methodology for cross-species comparisons is crucial. In general, it is known that

nonphylogenetic statistical methods for cross-species comparison can lead to misleading claims of statistical significance (Freckleton, 2002; Hernandez et al., 2013).

One argument in favour of high HGT in extreme or terrestrial environments is that extremophiles tend to form densely-packed biofilm communities (Marks et al., 2012), known to facilitate HGT (Molin and Tolker-Nielsen, 2003). However, biofilms also occur in mesophilic, aquatic niches (Aminov, 2011; Whitton, 2012). Among extremophiles, thermophiles are known to employ various mechanisms that prevent DNA degradation at high temperatures (e.g. changes in the chain length of carbon atoms present in the fatty acid cell membrane, higher GC content of genetic constitutes, such as rRNA and tRNA, and changes in protein structure, which include for example more salt bridges) and DNA from these species may be more likely to persist outside of the cell for longer periods of time (Fuchsman et al., 2017; Patel et al., 2019). Evidence suggests that HGT has been important in the adaptation of the cyanobacterium *Thermosynechococcus* to extreme environments (Cheng et al., 2020). However, this does not preclude the importance of HGT to adaptation for non-extreme environments as well. Far larger numbers of species can live in non-extreme environments, where competition is likely to be more intense and chances for HGT are probably higher. Successful HGT events depend on the transfer of genetic material, the survival of foreign DNA in the cell and their integration via recombination, as well as their fixation in the population or at least their persistence over evolutionary time. Transferred genes can be neutral, deleterious or advantageous to the recipient. As with any changes to the genome, fixation is more probable when the HGT event is advantageous (Tooming-Klunderud et al., 2013).

Another challenge is determining what constitutes an extreme environment. Studies have varied in their definition of “extreme”. For example, terrestrial cyanobacteria have been classified as mesophilic by Li et al. (2014) and extremophilic by Chen et al. (2021). In fact, soil ecosystems are vastly heterogenous and microbiota diversity has been shown to vary depending on the site’s physicochemical conditions (Smiles, 1988). The rate of conjugal plasmid transfer in soil is contingent on abiotic factors such as soil type, temperature, pH and moisture (Richaume et al., 1989; Richaume et al., 1992) and biotic factors, such as the size and presence of microbial and eukaryotic regulatory networks (Daane et al., 1996; Parter et al., 2007). Whereas bulk soil (generally nutrient-poor) may impose barriers to HGT, nutrient-rich areas (e.g. rhizosphere, phyllosphere, decaying plant/animal tissues and manure-applied

soil) are known as hotspots for HGT (Bjrklf et al., 1995; Lilley and Bailey, 1997; Mølbak and Kroer, 2007). Similarly, Fuchsman et al. (2017) have shown that species inhabiting anaerobic and/or high temperature environments share a large number of genes (likely due to HGT), however the same was not observed for species living in high salt conditions. It could be that HGT incidence is higher is some extreme environments, but not in others. The classification used in this and other studies may conceal such distinctions.

In marine bacteria, the presence of plasmids has been shown to correlate with the degree of water pollution (Baya et al., 1986). Hence regions with higher propensity for HGT may exist within specific aquatic habitats, such as polluted coastal waters. Also, some organisms are known to be more prone to HGT than others, for example with varying levels of competence for transformation (Johnsborg et al. 2007). Although the importance of transduction in freshwater cyanobacteria remains speculative, cyanophages in marine environments are believed to play a key role in HGT (Zeidner et al., 2005). The most relevant types of HGT in cyanobacteria, whether different HGT types are preferred in extreme *versus* mesophilic environments, and which cyanobacteria are prone to which type of HGT mechanism, are all poorly understood. Nevertheless, to our knowledge, no studies have correlated natural competence with extremotolerance in cyanobacteria.

Previous studies have used various methods to quantify HGT. For example, Li et al. (2014), dealing with metagenomic data, used transposase number as a proxy for HGT, which has been suggested to correlate with the frequency of transfer events (Brazelton and Baross, 2009; Hemme et al., 2010; Li et al., 2014; Xie et al., 2011). The results of gene exchange studies depend crucially on the accuracy of HGT inference. HGT events can generate gene families with patchy phylogenetic distributions, which is the basis of a number of HGT inference methods in widespread use. Parallel gains and losses can also generate incongruent phylogenetic distributions and methods, like Notung and unlike BLAST-based approaches, which can not only infer, but distinguish between these events and HGT are indispensable (e.g. Notung can infer gene gains and gene transfers separately, whereas BLAST-based approaches will inform of a relationship (i.e. homology) between two genes/proteins, based on sequence similarity, but will not necessarily be able to discern between a gain and a transfer). Our study of HGT applies the explicitly phylogenetic approach of reconciling the gene tree with the species tree. This is an attractive strategy because the proxy or operational definition of HGT – a mismatch between the gene phylogeny and species phylogeny which is

most parsimoniously explained by transfer from one species to another – is close to the biological definition. It has also been argued that phylogenetic inference is the best approach to detect HGT events (Husnik and McCutcheon, 2018; Syvanen, 1994). On the other hand, our method requires a donor and a recipient within the study group (in contrast to methods which, for example, infer HGT from locally anomalous DNA summary statistics, such as dinucleotide frequencies). Reconciliation cannot detect HGT involving other taxa (i.e. HGT between a cyanobacterium and archaea/bacterial species).

Notung infers HGT counts for ancestral states with an explicit model of gene content evolution (DTL), however only data pertaining to extant species were included in our phylogenetically corrected t-tests implemented in BayesTraits. BayesTraits implicitly makes its own ancestral state reconstructions – e.g. mean and variance – on the basis of extant species and the species tree, without an explicit model of gene content evolution. Whether better approaches exist for the incorporation of ancestral HGT rates merits investigation. Empirical Bayes ancestral state reconstructions (Pagel, 1999b; Yang, 2006) of the habitat could improve visualisation of the environment inhabited by ancestral species, and hence would make it easier to locate potential associations between the extent of HGT and the habitat. However, the presiding ambiguity and uncertainty regarding extant species' isolation source and classification, and the fact that there are multiple different classifications of the environment would add to the complexity, feasibility and potential usefulness of this analysis. Regarding the statistical comparison of HGT events (Section 4.4.4), there is an argument that the use of recipient events only, instead of the total events, is more appropriate. This is because, in total events, each transfer is being counted twice (i.e. once for the recipient and once for the donor). Although this has been done for all species and environments (and hence does not introduce bias) and preliminary statistical analyses conducted on recipient events only showed similar trends and statistical significance, it makes the overall interpretation less clear. Analyses on the acquisition of new genes is also potentially more biologically relevant.

Complementing our study with similar analyses in other clades, which also include a mix of extreme and non-extreme environments, would be an interesting future direction. Further research in cyanobacteria would benefit from more focused analysis of HGT, e.g. investigations of specific gene families and/or in a limited number of species, so that more precise conclusions can be drawn. Although broad analyses of 130 cyanobacterial genomes

and thousands of gene families can provide valuable information about HGT at the phylum level, smaller changes at the species/gene level may go unnoticed. Moreover, given the dissimilarities between cyanobacterial species, both in biochemistry, physiology and morphology, it is likely that HGT forces act differently depending on the species involved. A more focused analysis, with fewer computational resources, would also enable the inclusion of additional (non-parametric) methods for the identification of HGT events, which would serve as a comparison and as an extra layer of evidence for the transfer events encountered. Irrespective of the methodology applied and the size of the dataset analysed, a more precise description of the isolate's biotic environment (e.g. biofilm formation, phage prevalence), physicochemical environment (e.g. temperature, pH, UV exposure) and its optimal growth conditions in the laboratory are necessary for studies of this kind. Without the appropriate and metadata, the allocation of cyanobacterial species into discrete groups/classifications cannot be accomplished.

4.6 Summary and conclusions

We have used a bioinformatics approach to test existing hypotheses on the amount of HGT activity in cyanobacteria. We addressed the question of whether the quantity of HGT events is positively correlated with extreme/terrestrial environments by using phylogenetic reconciliation, which can distinguish between HGT events and parallel, loss or reversal events. We used a phylogenetic statistical method to test our hypotheses and we considered several different definitions of extreme environments. We did not find evidence of higher prevalence of HGT among species from extreme environments, or among species from terrestrial environments. We believe that testing these associations without correcting for phylogenetic relationships can overestimate the significance of such associations. As far as we are aware, phylogenetic statistical corrections were not employed in previous studies. Our study presents a further contribution to the current knowledge of HGT and its link to environmental conditions. Although this study focuses on cyanobacteria only, it would be interesting to perform similar studies on other prokaryotic clades. Future research will benefit from a more precise description of the biotic and abiotic environment of isolates and their optimal growth conditions.

Chapter 5

General discussion

5.1 Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxins 2,4-DAB and BMAA

5.1.1 Key findings and implications of the work

With the rising threat of global warming and the continuous anthropogenic damage to the environment, especially when it comes to the eutrophication of water systems, the increased frequency, duration and extent of harmful cyanobacterial blooms (HCBs) are believed to become an ever-growing concern in future years (Gobler, 2020; Zhang et al., 2022). HCBs can be caused by the uncontrolled proliferation of toxic cyanobacterial species, which concentration/abundance can have detrimental effects to human and animal health, also often causing a negative impact on regional economies (Metcalf et al., 2018; Schaefer et al., 2020). Nowadays, food supplements derived from cyanobacteria (commonly referred to as "Spirulina") have become popular as a heath food, due to its multitude and diverse potential benefits. For a review of bioactive peptides from "Spirulina" see Ovando et al. (2018).

Analytical studies on cyanobacteria showed that the neurotoxins BMAA and 2,4-DAB have been found (either together or alone) in some species (Al-Sammak et al., 2014; Krüger et al., 2010; Rosén and Hellenäs, 2008; Violi et al., 2019). The metabolic pathways to the biosynthesis of these neurotoxins in cyanobacteria are currently unknown. In cyanobacteria, the ecological and physiological functions for BMAA and 2,4-DAB remain unspecified and the extent to which they are biosynthesised in these species has not been widely determined.

BMAA and potentially also 2,4-DAB, have been associated with the development of neurodegenerative diseases, including amyotrophic lateral sclerosis, Parkinson's and Alzheimer's disease (Banack and Murch, 2009; Cox et al., 2016; Murch et al., 2004), however this remains controversial (Chernoff et al., 2017; Dunlop et al., 2021). If 2,4-DAB and/or BMAA occur at high concentrations in HCBs and in cyanobacteria-concentrated food supplements, people and animals living near areas of intense and recurring HCBs and those ingesting concentrated cyanobacterial supplements may be at higher risk of the aforementioned neurodegenerative illnesses. As of today, drinking water systems are not systematically tested for these neurotoxins and "Spirulina" producers are not legally required to carry out neurotoxin screening (Chorus and Welker, 2021; Holtcamp, 2012).

Regarding 2,4-diaminobutanoic acid (2,4-DAB), our bioinformatics results point towards its biosynthesis in some cyanobacterial species via the aspartate 4-phosphate pathway. In some cyanobacteria, the presence of 2,4-DAB transaminase and decarboxylase within NIS clusters suggests a physiological association with iron scavenging. Because 2,4-DAB also occurs within hybrid NRPS/PKS clusters, we propose that 2,4-DAB production may also be associated with polyamine synthesis, which itself is associated with biofilm formation and siderophore production. There are species in which 2,4-DAB transaminase and decarboxylase occur outside specialised metabolite clusters. In these cases, we hypothesise that there are other and, as yet, unknown physiological roles for 2,4-DAB biosynthesis. Concerning 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA) our results indicate that the precursor 2,3-DAP is biosynthesised in some species via the staphyloferrin B pathway, which is also suggestive of a functional role in siderophore biosynthesis. The *pam* gene cluster from *Paenibacillus larvae*, synthesising a bound form of BMAA (paenilamicins) was not detected in any cyanobacterial genomes in our analysis. The *pam* complex was also not detected in searches of *Paenibacillus* species (other than *P. larvae*) and diatoms. Finally, the ability to synthesise 2,4-DAB and/or BMAA does not appear to be universal among cyanobacteria. This means that either these neurotoxins are not widespread in cyanobacteria, or that there are additional and yet unknown pathways for their biosynthesis. In species where there is bioinformatics support for 2,4-DAB and/or BMAA production, a link to specific environments, biochemistry or morphology is not clear.

Although our bioinformatics analyses appear to rule out a pathway to paenilamicin in 93 *Paenibacillus* genomes (other than *P. larvae*), a route to 2,4-DAB via the aspartate 4-phosphate pathway or the ectoine pathway was evident in some *Paenibacillus* species. *Paenibacillus* spp. are widespread in many ecosystems, including soil and marine sediment, and many have been isolated from human food, such as pasteurised milk, cheese and fruit, and from human samples (e.g. faecal and ventricular cerebrospinal fluid samples) (Beno et al., 2020; Hehnly et al., 2020; Khalaf and Raizada, 2020). If 2,4-DAB was found to be a common neurotoxin in *Paenibacillus*, then *in-silico* identification, followed by laboratory confirmation, would be essential in order to determine which species are capable of biosynthesising this neurotoxin and why.

Our bioinformatics results regarding 2,4-DAB and BMAA biosynthesis in cyanobacteria point to a physiological role associated with siderophore production. This is suggestive of a potential link between neurotoxin biosynthesis and the need to capture iron from the environment, which is known to be vital in many cellular activities (Cunningham and John, 2017; Sunda and Huntsman, 2015). Associations between other cyanotoxins, namely microcystins, and iron-scavenging have been described (Lyck et al., 1996; Utkilen and Gjølme, 1995). Fe(III) is highly insoluble and is typically rate-limiting for the growth of cyanobacteria, with many studies showing a positive correlation between cyanobacterial density and siderophore production (Harland et al., 2013). Although chemically and ecologically a difficult problem to resolve, it could be that a mechanism for providing/adding iron to the environment (or to cyanobacteria themselves) could lead to the cessation/reversal of the biosynthesis of 2,4-DAB, BMAA and perhaps of other cyanotoxins in cyanobacteria.

Bioinformatically screening cyanobacterial species for 2,4-DAB, BMAA and other cyanotoxins would likely be of wide interest to the research community (Cordeiro et al., 2021; Dulić et al., 2022). There are many benefits of checking for the presence/absence of genetic evidence for cyanotoxin production in multiple different fields. In the biotechnology, pharmaceutical and cosmetic sectors, bio-engineered/modified strains should be tested for cyanotoxin production to rule out potential health complications derived from human consumption or topic use of toxic products. Regarding the ecological importance of cyanotoxin screening, there is a growing interest in the early detection and monitoring of cyanobacterial species with the potential for cyanotoxin production in reservoirs for drinking and recreational waters, prior to toxin accumulation in the environment. This would enable the incorporation of preventive measures as opposed to corrective actions.

5.1.2 Study limitations

As with most scientific research, there are some limitations to our studies. Here, the limitations can be classified into (i) those derived from the data itself, (ii) those pertaining to the methodology used, (iii) and those related to the knowledge available.

Since all analyses performed were conducted *in-silico* (i.e. computationally; no laboratory research was carried out), we were totally reliant on the quality of publicly available data. Up to 2018, 180 characterised cyanobacterial genomes were available in the Genome database at

the NCBI (https://www.ncbi.nlm.nih.gov/genome). Prior to my research, Ziying Ke manually selected a subset of 130 genomes based on the quality of genome assembly annotation (e.g. number of proteins, RNAs, genes and pseudogenes, as well as completeness of assemblies). Although the inclusion criteria applied was rigorous, the failure to detect some enzymes (e.g. aspartate transaminase: EC 2.6.1.1, aspartate kinase: EC 2.7.2.4, and aspartate-semialdehyde dehydrogenase: EC 1.2.1.11, cysteine synthase: EC 2.5.1.47), thought to be essential/ubiquitous, in some cyanobacterial species, may suggest errors in sequencing, assembly, genome annotation or functional annotation. This could have led to false-negative results, i.e. the failure to detect an enzyme relevant to 2,4-DAB or BMAA biosynthesis and the consequential rejection of that species as a potential neurotoxin producer. As of June 2022, there are 705 cyanobacterial genomes in the Genome database at the NCBI. Although it is not expected that all genomes meet our inclusion criteria, it would be worthwhile repeating the analyses presented here, including the species phylogeny, so as to incorporate new information/species and to improve statistical power (i.e. likely more relevant in Chapter 4).

Still concerning the quality of the starting data, one issue substantially affecting our ability to derive clear conclusions is the widespread use of non-axenic cyanobacterial cultures. When non-axenic monocyanobacterial cultures or environmental samples are used for sequencing efforts, or for identifying/quantifying toxins, it is not possible to rule out contamination or that toxin production is the result of the metabolic activity of other bacteria, respectively. For cyanobacterial genomes derived from non-axenic cultures where bioinformatics support exists for either 2,4-DAB/BMAA biosynthesis, analytical tests for toxin production must be repeated in axenic cultures. Nevertheless, practical complications exist when using axenic cultures in a laboratory setting. Aside from the difficulty of obtaining pure cyanobacterial cultures, the deprivation of cyanobacteria from their "natural" environment (i.e. where more species are present) often imposes stresses on cultured isolates. These stress factors can complicate some molecular analyses (i.e. meaningful gene expression analyses are difficult to obtain since multiple stress pathways are continuously activated) and can limit culture viability (S. Collins, personal communication, 5 July). Despite these problems, we have highlighted (Chapter 2; Chapter 3; Mantas et al., 2021, 2022) the importance of research on axenic cultures, and urged the scientific community to only use non-axenic samples when all other alternatives were exhausted/unavailable (e.g. metagenomic studies).

The methodology employed in this study included sequence alignment, profile hidden Markov models (pHMMs), substrate specificity/active site identification, the reconstruction of gene phylogenies, among other techniques. Although rigorous quality criteria were devised at every step and multiple positive and negative controls were included to ensure the results matched our expectations (i.e. we expected to find particular amino acids at the active site of an enzyme, if the active site of that enzyme had been previously described to occur in that species), it is in the nature of pHMMs to be as good at finding remote homologs as the sequences included to generate the probabilistic model. To generate the best pHMMs possible, I have carefully and methodically selected cyanobacterial or bacterial amino acid sequences (in instances where cyanobacterial proteins were absent or scarce), avoided including multiple sequences from the same species and attempted to include, at least, as many sequences, on average, as PFAM (Sonnhammer et al., 1998). However, it is my understanding that an equally good choice of amino acid sequences could potentially lead to a different pHMM, which would consequentially find dissimilar homologs, with different statistical confidence (i.e. E-value, bit score). The same rationale can be applied to the reconstruction of (gene family) phylogenies (i.e. different tree topologies will be obtained when using different amino acid sequences). Another issue stemming from this is the fact that many homologs of enzymes potentially leading to the biosynthesis of 2,4-DAB and BMAA are very similar, and yet catalyse distinct chemical reactions. Distinguishing between such enzymes through bioinformatics techniques is extremely difficult and error-prone, and often not possible.

Although the results presented here provide valuable information regarding 2,4-DAB and BMAA biosynthesis, testing multiple hypothesis for their production in cyanobacteria and alluding to their potential physiological role in these species, experimental verification (in a wet laboratory setting) is still crucial to derive solid conclusions. The aim of these studies was to find potential genetic evidence for 2,4-DAB and/or BMAA biosynthesis in cyanobacteria and to focus the efforts (both financially and in time) of researchers to a few cyanobacterial genomes only (i.e. where genetic evidence for the neurotoxins was found). The bioinformatics insights presented here do not replace the need for analytical confirmation.

Our studies only found genetic evidence for the biosynthesis of 2,4-DAB and BMAA in a limited number of cyanobacterial species. While it could be true that these neurotoxins are

not produced by the majority of cyanobacteria, it is possible that there are other biochemical routes to 2,4-DAB and BMAA that we are currently unaware of. Although we tested an extensive number of genomes, not all cyanobacterial species were included in the study (e.g. *Okeania*, *Brasilonema*, *Cuspidothrix*, for example, were not included), either because there were no sequenced genomes available at the time or because the available genomes did not meet our inclusion criteria. Conclusions about the toxicologic capabilities of these species cannot be made on the basis of our research.

5.1.3 Future directions

We have provided an initial bioinformatics insight into potential pathways to 2,4-DAB and BMAA biosynthesis in cyanobacteria, which can lead to future experimental investigations. Laboratory confirmation is still crucial to confirm or disprove the genetic evidence presented here. Among useful laboratory experiments that should be performed, I highlight the need for 1) enzyme expression assays to check for the activity of key enzymes involved in biochemical pathways potentially leading to 2,4-DAB and/or BMAA in cyanobacteria; 2) analytical tests for the identification and quantification of 2,4-DAB and/or BMAA in species shown to possess the genes needed for the biosynthesis of these neurotoxins, preferably through MS/MS techniques; 3) assays to check for the production of siderophores in species shown to be able to biosynthesise 2,4-DAB and/or BMAA; and 4) assays to check whether the production of such siderophores, if they exist, is correlated with iron availability and/or with 2,4-DAB/BMAA production in cyanobacteria. Concerning further *in-silico* studies, I believe that a tool for the genomic detection of all/most cyanotoxins is warranted and would meet the needs of a wide audience in the field of cyanobacterial research.

5.2 In cyanobacteria, is horizontal gene transfer more prevalent in extreme environments?

5.2.1 Key findings and implications of the work

We confirmed that HGT has played an important role in the evolution of cyanobacteria. Similarly to what has been previously found in other bacterial species, HGT in cyanobacteria appears to be more common among closely related species, compared to species that have diverged early in the course of evolution. In our study, we found no evidence of higher

prevalence of HGT among species from extreme environments, or among species from terrestrial environments. For HGT inference, we have used the Notung software (Stolzer et al., 2012), which performs parsimonious reconciliation of gene trees and the species tree with explicit costs associated with gene gain, loss and transfer. Objectively defining extreme habitats is difficult and to mitigate this issue we have considered several different definitions of extreme environments. Unlike other studies investigating the same question, we adopted a phylogenetic statistical method, BayesTraits, to test our hypotheses, as we believe that testing these associations without correcting for phylogenetic relationships can lead to an overestimation of the significance of such associations.

HGT is one of the main mechanisms contributing to genetic variation and evolution in prokaryotes (Hall et al., 2020), however knowledge regarding its frequency and mechanisms is still at its infancy. There is little information concerning (i) how much HGT occurs in a natural setting; (ii) what are the main drivers of genetic exchange; (iii) whether some species are more prone to (some mechanisms of) HGT and why; and (iv) the importance/impact that HGT has played so far in the evolution of species.

Currently, many studies on prokaryotic HGT are either focused on biotechnological applications or the public health sector. Regarding the first, a deeper understanding of HGT would allow, for example, the utilisation of a species' natural competence for transformation, which would enable faster, easier and cheaper exploitation for biotechnological applications, compared with the currently used cell-engineering efforts (Schirmacher et al., 2020). Nevertheless, the most well-known studies on HGT relate to the dangers of the appearance of multi-resistant bacteria, also called "superbugs". "Superbugs" arise when genetic exchanges of antibiotic-resistant genes (ARG), usually within plasmids, occur between antibiotic-resistance bacteria and antibiotic-susceptible bacteria, resulting in species that are resistant to multiple antibiotics (Malhotra-Kumar et al., 2016; Mathers et al., 2015). The rate of the spread of ARG by HGT, the widespread mis-use of antibiotics combined with the slow-pace at which new antibiotics are being discovered/manufactured have become a global challenge (Sun et al., 2019). It is estimated that the number of human deaths accounted for by multidrug resistance will increase to 10 million by 2050, surpassing the number of deaths arising from cancer (WHO, 2014).

We have introduced a pipeline for HGT inference using Notung, which performs parsimonious reconciliation between gene trees and the species tree, with costs associated with gene loss, gene duplication and gene transfer (Stolzer et al., 2012). I believe the current work improves our understanding of HGT events in cyanobacteria and its relation to the environment, and contributes to the narrative of whether or not HGT is more frequent in extreme habitats. The same methodological approach can be applied to other groups of species. It would be interesting to compare the results obtained here with those of other prokaryotic clades. I believe analyses in pathogenic bacteria would be especially worthwhile, given their relevance in multidrug resistance and global health.

5.2.2 Study limitations

One of the main limitations concerning this work is the patchy or complete lack of metadata concerning each cyanobacterial species' isolation source, geographic location and optimal growth conditions (i.e. temperature, pH). This hindered the classification of species into those inhabiting extreme environments *versus* mesophilic niches and those living in aquatic *versus* terrestrial/variable habitats. It would have been interesting to investigate the differences in HGT amount between facultative extremophiles (i.e. species thriving in mesophilic niches, but that are also able to survive under extreme conditions at little or no growth cost) (Low-Décarie et al., 2016) and extremophilic cyanobacteria, but the metadata available would often not allow for a clear distinction between the two groups.

There are many methodological approaches for the inference of HGT events (see Chapter 1, Section 2.3.6), with different strengths and weaknesses. I believe that, currently, there is no unassailable/bullet-proof method for the identification and quantification of HGT, however phylogenetic approaches are believed to be the best methods available (Husnik and McCutcheon, 2018; Syvanen, 1994). Although the pipeline using Notung described here represents a robust (i.e. well-rounded and complete) technique for measuring HGT, there are some limitations. Notung assigns specific costs associated with gene loss (default cost = 1.0), duplication (default cost = 2.0) and transfer (default cost = 3.5), which can be modified by the user. The optimal choice of costs is a non-trivial decision as it is often not known which parameters are most biologically relevant. The best combination of costs may differ depending on the data being analysed: eukaryotic *versus* prokaryotic, symbiotic *versus* free-living, known evolutionary history of genome expansion *versus* genome contraction, and

testing multiple parameters can be a computationally expensive task, especially for large datasets and low cost values. Here, we choose the cost of transfer that minimised genome size flux (i.e. difference in genome size between parent and child nodes (David and Alm, 2011)), but we do not exclude the possibility that there are other equally adequate approaches to find the optimal horizontal transfer cost. Moreover, when running a reconciliation in Notung with a duplication-transfer-loss (DTL) model, temporally infeasible trees (i.e. when the timing constraints imposed by the species tree cannot be compatible with the inferred transfers) are identified and excluded from the analysis. This could amount to a substantial portion of the data, which could decrease statistical power and potentially impact the ability to derive solid conclusions. In our analyses, increasing the cost of transfer led to a decrease in the number of temporally infeasible trees, however high(er) transfer costs may not always be biologically relevant, especially when dealing with prokaryotic data.

Because Notung performs parsimonious reconciliation between gene tree-species tree, HGT events can only be inferred for species contained in the species phylogeny. Thus, this study provides no information regarding HGT events that may have occurred between other prokaryotic clades (and even eukaryotes) and cyanobacteria. Although such information would undoubtedly be useful, this would certainly entail much larger datasets and computational resources.

Finally, there is limited ability to cross check our results with those of different research groups. As previously mentioned, studies analysing the links between the frequency of HGT and environmental factors are scarce (Chen et al., 2021; Gophna et al., 2015; Goyal et al., 2020; Li et al., 2014). Among the available studies, there is a wide discrepancy between (i) the type of data being analysed (e.g. metagenomics data *versus* whole-genome sequencing data); (ii) the methodological approaches employed (e.g. theoretical models, BLAST-based approaches, among others); (iii) the statistical methods used to test for correlation (e.g. non-phylogenetic *versus* phylogenetic corrected tests); (iv) the taxa/number of taxa included in the study; and (v) how species are classified according to the environment inhabited (e.g. disagreements between terrestrial habitats being considered extreme *versus* benign).

5.2.3 Future directions

In this study, I have analysed the frequency of HGT events within the cyanobacteria phylum. This is a novel and interesting approach, as it enables a broad understanding of the extent of genetic exchanges in cyanobacteria, and the role that it played in shaping cyanobacterial evolution. However, given that the whole genomes of 130 cyanobacterial species were analysed, specific details about HGT events in a particular species or in a gene family may have gone unnoticed. Most studies concerning HGT in cyanobacteria are focused on the analysis of a smaller group of species and/or genes (Bolhuis et al., 2010; Tooming-Klunderud et al., 2013; Watanabe and Horiike, 2021). I believe that a more targeted approach (i.e. analysing a few gene families or species of interest) in cyanobacteria would be valuable, with the added benefit of potentially faster analyses and fewer computational resources needed.

5.3 Concluding remarks

This thesis presents the results of multiple comparative genomics analyses in 130 cyanobacterial genomes, applied to neurotoxin biosynthesis prediction and HGT inference. The bioinformatic techniques applied here enabled conclusions regarding BMAA and 2,4-DAB biosynthesis in cyanobacteria, and should serve as a starting-point for future analytical experiments concerning the production of these neurotoxins by cyanobacterial species. We present a pipeline for HGT inference that has contributed to the current knowledge of the association between genetic exchange and the environment. This pipeline can be applied for different taxa, including pathogenic bacteria.

I believe the use of comparative genomics techniques allow for powerful analyses on large datasets that enable conclusions concerning evolution, whether at the phylum, species and gene level. Comparative genomics tools allow investigation into gene function/annotation and adaptations over time, essential to understand the complexity of life on Earth. There is an ever-increasing number of sequenced genomes and databases that harbour genomic data are becoming progressively larger. However, generating data is not sufficient. There is a pressing need for people and tools to make sense of this information. The work presented here is the result of a collaboration between bioinformaticians, microbiologists and chemists, and I

believe the interdisciplinarity nature of these projects is what allowed the advancement in the field and what made this research so original and special.

References

Abed, R. M. M., Dobretsov, S., Sudesh, K., 2009. Applications of cyanobacteria in biotechnology. J. Appl. Microbiol. 106, 1-12. https://doi.org/10.1111/j.1365-2672.2008.03918.x.

Adams, D. G., Duggan, P. S., Jackson, O., 2012. Cyanobacterial Symbioses. In: Whitton, B. A. (Ed.), Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands, Dordrecht, pp. 593-647.

Agius, L., 1997. Channelling in intermediary metabolism. Portland Press, London, Chapel Hill.

Al-Sammak, M. A., Hoagland, K. D., Cassada, D., Snow, D. D., 2014. Co-occurrence of the cyanotoxins BMAA, DABA and anatoxin-a in Nebraska reservoirs, fish, and aquatic plants. Toxins 6(2), 488-508. https://doi.org/10.3390/toxins6020488.

Alexeyenko, A., Tamas, I., Liu, G., Sonnhammer, E. L. L., 2006. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics 22, e9-e15. https://doi.org/10.1093/bioinformatics/btl213.

Allwood, A. C., Walter, M. R., Kamber, B. S., Marshall, C. P., Burch, I. W., 2006. Stromatolite reef from the Early Archaean era of Australia. Nature 441, 714-718. https://doi.org/10.1038/nature04764.

Almeida, H., Palys, S., Tsang, A., Diallo, A. B., 2020. TOUCAN: a framework for fungal biosynthetic gene cluster discovery. NAR Genom. Bioinform. 2, lqaa098. https://doi.org/10.1093/nargab/lqaa098.

Altenhoff, A. M., Schneider, A., Gonnet, G. H., Dessimoz, C., 2010. OMA 2011: orthology inference among 1000 complete genomes. Nucleic Acids Res. 39, D289-D294. https://doi.org/10.1093/nar/gkq1238.

Altschul, S. F., 1998. Generalized affine gap costs for protein sequence alignment. Proteins 32, 88-96. https://doi.org/10.1002/(SICI)1097-0134(19980701)32:1<88::AID-PROT10>3.0.CO;2-J.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402. https://doi.org/10.1093/nar/25.17.3389.

Alwathnani, H., Johansen, J. R., 2011. Cyanobacteria in soils from a Mojave Desert ecosystem. Monogr. West N. Am. Nat. 5, 71-89. https://doi.org/10.3398/042.005.0103.

Aminov, R. I., 2011. Horizontal gene exchange in environmental microbiota. Front. Microbiol. 2, 158. https://doi.org/10.3389/fmicb.2011.00158.

Anand, S., Prasad, M. V. R., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z., Mohanty, D., 2010. SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res. 38, W487-W496. https://doi.org/10.1093/nar/gkq340.

Andreote, A. P. D., Dini-Andreote, F., Rigonato, J., Machineski, G. S., Souza, B. C. E., Barbiero, L., Rezende-Filho, A. T., Fiore, M. F., 2018. Contrasting the genetic patterns of microbial communities in soda lakes with and without cyanobacterial bloom. Front. Microbiol. 9, 244. https://doi.org/10.3389/fmicb.2018.00244.

Aráoz, R., Molgó, J., de Marsac, N. T., 2010. Neurotoxic cyanobacterial toxins. Toxicon 56, 813-828. https://doi.org/10.1016/j.toxicon.2009.07.036.

Aravind, L., 2000. Guilt by association: contextual information in genome analysis. Genome Res. 10, 1074-1077. https://doi.org/10.1101/gr.10.8.1074.

Armah, A., Hiskia, A., Kaloudis, T., Chernoff, N., Hill, D., Antoniou, M. G., He, X., Loftin, K., O'Shea, K., Zhao, C., 2013. A review on cylindrospermopsin: the global occurrence, detection, toxicity and degradation of a potent cyanotoxin. Environ. Sci. Process Impacts 15, 1979-2003. https://doi.org/10.1039/C3EM00353A.

Armstrong, J. E., Van Baalen, C., 1979. Iron transport in microalgae: the isolation and biological activity of a hydroxamate siderophore from the blue-green alga *Agmenellum quadruplicatum*. J. Gen. Microbiol. 111(2), 253-262. https://doi.org/10.1099/00221287-111-2-253.

Årstøl, E., Hohmann-Marriott, M. F., 2019. Cyanobacterial siderophores—Physiology, structure, biosynthesis, and applications. Mar. Drugs 17(5), 281. https://doi.org/10.3390/md17050281.

Asagi, K., Oka, T., Arao, K., Suzuki, I., Thakur, M. K., Izumi, K., Natori, Y., 1998. Purification, characterization and differentiation-dependent expression of a perchloric acid soluble protein from rat kidney. Nephron 79, 80-90. https://doi.org/10.1159/000044996.

Averhoff, B., Kirchner, L., Pfefferle, K., Yaman, D., 2021. Natural transformation in Gram-negative bacteria thriving in extreme environments: from genes and genomes to proteins, structures and regulation. Extremophiles 25, 425-436. https://doi.org/10.1007/s00792-021-01242-z.

Azad, R. K., Lawrence, J. G., 2011. Towards more robust methods of alien gene detection. Nucleic Acids Res. 39, e56-e56. https://doi.org/10.1093/nar/gkr059.

Bakermans, C., 2015. Microbial evolution under extreme conditions. Walter de Gruyter GmbH & Co, KG, Berlin. https://doi.org/10.1515/9783110340716.

Balasubramanian, R., Shen, G., Bryant, D. A., Golbeck, J. H., 2006. Regulatory roles for IscA and SufA in iron homeostasis and redox stress responses in the cyanobacterium *Synechococcus* sp. strain PCC 7002. J. Bacteriol. 188(9), 3182-3191. https://doi.org/10.1128/JB.188.9.3182-3191.2006.

Baldi, P., Chauvin, Y., Hunkapiller, T., McClure, M. A., 1994. Hidden Markov models of biological primary sequence information. Proc. Natl. Acad. Sci. 91, 1059-1063. https://doi.org/10.1073/pnas.91.3.1059.

Banack, A. S., Johnson, E. H., Cheng, R., Cox, A. P., 2007. Production of the neurotoxin BMAA by a marine cyanobacterium. Mar. Drugs 5, 180-196. https://doi.org/10.3390/md504180.

Banack, S. A., Cox, P. A., 2003. Biomagnification of cycad neurotoxins in flying foxes: implications for ALS-PDC in Guam. Neurology 61, 387-389. https://doi.org/10.1212/01.WNL.0000078320.18564.9F.

Banack, S. A., Murch, S. J., 2009. Multiple neurotoxic items in the Chamorro diet link BMAA with ALS/PDC. Amyotroph. Lateral Scler. 10, 34-40. https://doi.org/10.3109/17482960903278451.

Banack, S. A., Murch, S. J., Cox, P. A., 2006. Neurotoxic flying foxes as dietary items for the Chamorro people, Marianas Islands. J. Ethnopharmacol. 106, 97-104. https://doi.org/10.1016/j.jep.2005.12.032.

Barghouthi, S., Young, R., Olson, M. O., Arceneaux, J. E., Clem, L. W., Byers, B. R., 1989. Amonabactin, a novel tryptophan-or phenylalanine-containing phenolate siderophore in *Aeromonas hydrophila*. J. Bacteriol. 171, 1811-1816. https://doi.org/10.1128/jb.171.4.1811-1816.1989.

Barkay, T., Kritee, K., Boyd, E., Geesey, G., 2010. A thermophilic bacterial origin and subsequent constraints by redox, light and salinity on the evolution of the microbial mercuric reductase. Environ. Microbiol. 12, 2904-2917. https://doi.org/10.1111/j.1462-2920.2010.02260.x.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., 2004. The Pfam protein families database. Nucleic Acids Res. 32, D138-D141. https://doi.org/10.1093/nar/gkh121.

Bauer, J. S., Ghequire, M. G. K., Nett, M., Josten, M., Sahl, H. G., De Mot, R., Gross, H., 2015. Biosynthetic origin of the antibiotic pseudopyronines A and B in *Pseudomonas putida* BW11M1. ChemBioChem 16, 2491-2497. https://doi.org/10.1002/cbic.201500413.

Baya, A. M., Brayton, P. R., Brown, V. L., Grimes, D. J., Russek-Cohen, E., Colwell, R. R., 1986. Coincident plasmids and antimicrobial resistance in marine bacteria isolated from polluted and unpolluted Atlantic Ocean samples. Appl. Environ. Microbiol. 51, 1285-1292. https://doi.org/10.1128/aem.51.6.1285-1292.1986.

Beasley, F. C., Cheung, J., Heinrichs, D. E., 2011. Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*. BMC Microbiol. 11, 199. https://doi.org/10.1186/1471-2180-11-199.

Beaulaurier, J., Zhu, S., Deikus, G., Mogno, I., Zhang, X.-S., Davis-Richardson, A., Canepa, R., Triplett, E. W., Faith, J. J., Sebra, R., 2018. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. Nat. Biotechnol. 36, 61-69. https://doi.org/10.1038/nbt.4037.

Becker, T., Pasteels, J., Weigel, C., Dahse, H.-M., Voigt, K., Boland, W., 2017. A tale of four kingdoms–isoxazolin-5-one-and 3-nitropropanoic acid-derived natural products. Nat. Prod. Rep. 34, 343-360. https://doi.org/10.1039/c6np00122j.

Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., Stein, H. J., Hannah, J. L., Coetzee, L. L., Beukes, N. J., 2004. Dating the rise of atmospheric oxygen. Nature 427, 117-120. https://doi.org/10.1038/nature02260.

Bell, E. A., 1962. Associations of ninhydrin-reacting compounds in the seeds of 49 species of *Lathyrus*. Biochem. J. 83(2), 225-229. https://doi.org/10.1042/bj0830225.

Benning, C., Huang, Z.-H., Gage, D. A., 1995. Accumulation of a novel glycolipid and a betaine lipid in cells of *Rhodobacter sphaeroides* grown under phosphate limitation. Arch. Biochem. Biophys. 317(1), 103-111. https://doi.org/10.1006/abbi.1995.1141.

Beno, S. M., Cheng, R. A., Orsi, R. H., Duncan, D. R., Guo, X., Kovac, J., Carroll, L. M., Martin, N. H., Wiedmann, M., 2020. *Paenibacillus odorifer*, the predominant *Paenibacillus* species isolated from milk in the United States, demonstrates genetic and phenotypic conservation of psychrotolerance but clade-associated differences in nitrogen metabolic pathways. Msphere 5, e00739-00719. https://doi.org/10.1128/mSphere.00739-19.

Bernroitner, M., Zamocky, M., Furtmüller, P. G., Peschek, G. A., Obinger, C., 2009. Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. J. Ex. Bot. 60, 423-440. https://doi.org/10.1093/jxb/ern309.

Bisht, S., Rajaram, V., Bharath, S. R., Kalyani, J. N., Khan, F., Rao, A. N., Savithri, H. S., Murthy, M. R. N., 2012. Crystal structure of *Escherichia coli* diaminopropionate ammonia-lyase reveals mechanism of enzyme activation and catalysis. J. Biol. Chem. 287, 20369-20381. https://doi.org/10.1074/jbc.M112.351809.

Bjrklf, K., Suoniemi, A., Haahtela, K., Romantschuk, M., 1995. High frequency of conjugation versus plasmid segregation of RP1 in epiphytic *Pseudomonas syringae* populations. Microbiology 141, 2719-2727. https://doi.org/10.1099/13500872-141-10-2719.

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49, W29-W35. https://doi.org/10.1093/nar/gkab335.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47, W81-W87. https://doi.org/10.1093/nar/gkz310.

Blin, K., Shaw, S., Tong, Y., Weber, T., 2020. Designing sgRNAs for CRISPR-BEST base editing applications with CRISPy-web 2.0. Synth. Syst. Biotechnol. 5, 99-102. https://doi.org/10.1016/j.synbio.2020.05.005.

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U. I. A., Stal, L. J., 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus chthonoplastes*. ISME J. 4, 121-130. https://doi.org/10.1038/ismej.2009.99.

Bork, P., Koonin, E. V., 1994. A new family of carbon-nitrogen hydrolases. Protein Sci. 3, 1344-1346. https://doi.org/10.1002/pro.5560030821.

Boussiba, S., Wu, X., Zarka, A., 2000. Alkaliphilic cyanobacteria. In: Seckbach, J. (eds) Journey to Diverse Microbial Worlds. Cellular Origin and Life in Extreme Habitats, vol 2. Springer, Dordrecht, pp. 209-224. https://doi.org/10.1007/978-94-011-4269-4_15.

Bradley, W. G., Borenstein, A. R., Nelson, L. M., Codd, G. A., Rosen, B. H., Stommel, E. W., Cox, P. A., 2013. Is exposure to cyanobacteria an environmental risk factor for amyotrophic lateral sclerosis and other neurodegenerative diseases? Amyotroph. Lateral Scler. Frontotemporal Degener. 14, 325-333. https://doi.org/10.3109/21678421.2012.750364.

Brazelton, W. J., Baross, J. A., 2009. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. ISME J. 3, 1420-1424. https://doi.org/10.1038/ismej.2009.79.

Brenner, E. D., Stevenson, D. W., McCombie, R. W., Katari, M. S., Rudd, S. A., Mayer, K. F., Palenchar, P. M., Runko, S. J., Twigg, R. W., Dai, G., Martienssen, R. A., Benfey, P. N., Coruzzi, G. M., 2003. Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. Genome Biol. 4, R78. https://doi.org/10.1186/gb-2003-4-12-r78.

Brickman, T. J., Armstrong, S. K., 1996. The ornithine decarboxylase gene odc is required for alcaligin siderophore biosynthesis in *Bordetella* spp.: putrescine is a precursor of alcaligin. J. Bacteriol. 178(1), 54-60. https://doi.org/10.1128/jb.178.1.54-60.1996.

Brock, T. D., 1973. Lower pH limit for the existence of blue-green algae: evolutionary and ecological implications. Science 179, 480-483. https://doi.org/10.1126/science.179.4072.480.

Brock, T. D., 1976. Halophilic blue-green algae. Arch. Microbiol. 107, 109-111. https://doi.org/10.1007/BF00427875.

Brock, T. D., 2012. Thermophilic microorganisms and life at high temperatures. Springer Science & Business Media, pp. 132-138.

Brown, E. G., Turan, Y., 1995. Pyrimidine metabolism and secondary product formation; biogenesis of albizziine, 4-hydroxyhomoarginine and 2,3-diaminopropanoic acid. Phytochemistry 40, 763-771. https://doi.org/10.1016/0031-9422(95)00317-Z.

Brown, E. G., Turan, Y., 1996. Formation of albizziine and 2,3-diaminopropanoic acid from uracil in *Albizia* seedlings. Phytochemistry 41, 1491-1495. https://doi.org/10.1016/0031-9422(95)00829-2.

Bucher, P., Karplus, K., Moeri, N., Hofmann, K., 1996. A flexible motif search technique based on generalized profiles. Comput. Chem. 20, 3-23. https://doi.org/10.1016/S0097-8485(96)80003-9.

Büdel, B., Dulic, T., Darienko, T., Rybalka, N., Friedl, T., 2016. Cynobacteria and algae of biological soil crusts. In: Weber, B. (eds) Biological Soil Crusts: An Organising Principle in Drylands. Ecological Studies 226. Springer, Cham, pp. 55-80.

Burman, J. D., Stevenson, C. E. M., Sawers, R. G., Lawson, D. M., 2007. The crystal structure of *Escherichia coli* TdcF, a member of the highly conserved YjgF/YER057c/UK114 family. BMC Struct. Biol. 7, 1-14. https://doi.org/10.1186/1472-6807-7-30.

Bursy, J., Pierik, A. J., Pica, N., Bremer, E., 2007. Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. J. Biol. Chem. 282(43), 31147-31155. https://doi.org/10.1074/jbc.M704023200.

Callebaut, A., Lambein, F., 1977. Biosynthesis of isoxazolinone amino acids: incorporation studies on the synthesis of 2-(2-amino-3-carboxypropyl)-isoxazolin-5-one in *Lathyrus odoratus* seedlings [proceedings]. Arch. Int. Physiol. Biochim. 85(1), 157-158.

Carmichael, W. W., 1994. The Toxins of Cyanobacteria. Sci. Am. 270, 78-86. https://doi.org/10.1038/scientificamerican0194-78.

Carmichael, W. W., 2001. Assessment of blue-green algal toxins in raw and finished drinking water. AWWA Research Foundation, Denver, pp. 179.

Carmichael, W. W., Jones, C. L. A., Mahmood, N. A., Theiss, W. C., Krogh, P., 1985. Algal toxins and water-based diseases. Crit. Rev. Environ. Control 15, 275-313. https://doi.org/10.1080/10643388509381734.

Carrick, H. J., Steinman, A. D., 2001. Variation in perypython biomass and species composition in Lake Okeechobee, Florida (USA): distribution of algal guilds along environmental gradients. Arch Hydrobiol 152, 411-438.

Carter, J. H., Du Bus, R. H., Dyer, J. R., Floyd, J. C., Rice, K. C., Shaw, P. D., 1974. Biosynthesis of viomycin. I. Origin of αβ-diaminopropionic acid and serine. Biochemistry 13, 1221-1227. https://doi.org/10.1021/bi00703a026.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., Parkhill, J., 2005. ACT: the Artemis comparison tool. Bioinformatics 21(16), 3422-3423. https://doi.org/10.1093/bioinformatics/bti553.

Casamatta, D. A., Johansen, J. R., Vis, M. L., Broadwater, S. T., 2005. Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria) 1. J. Phycol. 41, 421-438. https://doi.org/10.1111/j.1529-8817.2005.04062.x.

Catch, J. R., Jones, T. S., 1948. The chemical nature of aerosporin; isolation of a new natural amino-acid. Biochem. J. 42(4), lii.

Challis, G. L., 2005. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. ChemBioChem 6(4), 601-611. https://doi.org/10.1002/cbic.200400283.

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., Schomburg, D., 2021. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res. 49(D1), D498-D508. https://doi.org/10.1093/nar/gkaa1025.

Chen, C.-H., Flory, W., Koeppe, R. E., 1972. Variation of neurotoxicity of L-and D-2,4-diaminobutyric acid with route of administration. Toxicol. Appl. Pharmacol. 23(2), 334-338. https://doi.org/10.1016/0041-008X(72)90194-9.

Chen, J. Y., Bodley, J. W., 1988. Biosynthesis of diphthamide in *Saccharomyces cerevisiae*. Partial purification and characterization of a specific S-adenosylmethionine: elongation factor 2 methyltransferase. J. Biol. Chem. 263(24), 11692-11696. https://doi.org/10.1016/S0021-9258(18)37839-6.

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., Song, L.-R., Shu, W.-S., 2021. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. ISME J. 15, 211-227. https://doi.org/10.1038/s41396-020-00775-z.

Cheng, Y.-I., Chou, L., Chiu, Y.-F., Hsueh, H.-T., Kuo, C.-H., Chu, H.-A., 2020. Comparative genomic analysis of a novel strain of Taiwan hot-spring cyanobacterium *Thermosynechococcus* sp. CL-1. Front. Microbiol. 11, 82. https://doi.org/10.3389/fmicb.2020.00082.

Chernoff, N., Hill, D. J., Diggs, D. L., Faison, B. D., Francis, B. M., Lang, J. R., Larue, M. M., Le, T. T., Loftin, K. A., Lugo, J. N., 2017. A critical review of the postulated role of the non-essential amino acid, β-*N*-methylamino-L-alanine, in neurodegenerative disease in humans. J. Toxicol. Environ. Health B 20, 183-229. https://doi.org/10.1080/10937404.2017.1297592.

Cheung, J., Beasley, F. C., Liu, S., Lajoie, G. A., Heinrichs, D. E., 2009. Molecular characterization of staphyloferrin B biosynthesis in *Staphylococcus aureus*. Mol. Microbiol. 74, 594-608. https://doi.org/10.1111/j.1365-2958.2009.06880.x.

Chorus, I. (2001) Introduction: Cyanotoxins – Research for environmental safety and human helath, in: Chorus, I., (Eds.), Cyanotoxins – Occurrence, Causes, Consequences. Springer-Verlag, Berlin, 1-4. http://dx.doi.org/10.1007/978-3-642-59514-1_1.

Chorus, I., Welker, M. (Eds.), 2021. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. CRC Press. https://doi.org/10.1201/9781003081449.

Cochrane, G., Karsch-Mizrachi, I., Takagi, T., Sequence Database Collaboration, I. N., 2016. The international nucleotide sequence database collaboration. Nucleic Acids Res. 44, D48-D50. https://doi.org/10.1093/nar/gkv1323.

Codd, G. A., Morrison, L. F., Metcalf, J. S., 2005. Cyanobacterial toxins: risk management for health protection. Toxicol. Appl. Pharmacol. 203, 264-272. https://doi.org/10.1016/j.taap.2004.02.016.

Cohen, G. N., 1987. Regulation of Amino Acids Biosynthesis in Prokaryotes, in: Heilmeyer, L. M. G. (Eds.), Signal Transduction and Protein Phosphorylation. Springer, Boston, MA, pp. 273-287. https://doi.org/10.1007/978-1-4757-0166-1_35.

Cohen, S. A., 2012. Analytical techniques for the detection of α-amino-β-methylaminopropionic acid. Analyst 137(9), 1991-2005. https://doi.org/10.1039/C2AN16250D.

Colas, S., Marie, B., Lance, E., Quiblier, C., Tricoire-Leignel, H., Mattei, C., 2021. Anatoxin-a: overview on a harmful cyanobacterial neurotoxin from the environmental scale to the molecular target. Environ. Res. 193, 110590. https://doi.org/10.1016/j.envres.2020.110590.

Coleman, M. U., White, M. A., 1993. The role of biological disturbances in the prodution of solar salt. In: Seventh Symposium on Salt, Vol. 1. Elsevier Science Publishers, Amsterdam, pp. 623-631.

Cordeiro, R., Azevedo, J., Luz, R., Vasconcelos, V., Gonçalves, V., Fonseca, A., 2021. Cyanotoxin screening in BACA culture collection: identification of new cylindrospermopsin producing cyanobacteria. Toxins 13, 258. https://doi.org/10.3390/toxins13040258.

Costilow, R. N., Laycock, L., 1971. Ornithine cyclase (deaminating): purification of a protein that converts ornithine to proline and definition of the optimal assay conditions. J. Biol. Chem. 246, 6655-6660. https://doi.org/10.1016/S0021-9258(19)34165-1.

Cox, P. A., Banack, S. A., Murch, S. J., 2003. Biomagnification of cyanobacterial neurotoxins and neurodegenerative disease among the Chamorro people of Guam. PNAS 100, 13380-13383. https://doi.org/10.1073/pnas.2235808100.

Cox, P. A., Banack, S. A., Murch, S. J., Rasmussen, U., Tien, G., Bidigare, R. R., Metcalf, J. S., Morrison, L. F., Codd, G. A., Bergman, B., 2005. Diverse taxa of cyanobacteria produce β-*N*-methylamino-L-alanine, a neurotoxic amino acid. Proc. Natl. Acad. Sci. Unit. States Am. 102, 5074-5078. https://doi.org/10.1073/pnas.0501526102.

Cox, P. A., Davis, D. A., Mash, D. C., Metcalf, J. S., Banack, S. A., 2016. Dietary exposure to an environmental toxin triggers neurofibrillary tangles and amyloid deposits in the brain. Proc. R. Soc. B: Biol. Sci. 283, 20152397. https://doi.org/10.1098/rspb.2015.2397.

Cox, P. A., Sacks, O. W., 2002. Cycad neurotoxins, consumption of flying foxes, and ALS-PDC disease in Guam. Neurology 58, 956-959. https://doi.org/10.1212/WNL.58.6.956.

Craig, J. W., Cherry, M. A., Brady, S. F., 2011. Long-chain *N*-acyl amino acid synthases are linked to the putative PEP-CTERM/exosortase protein-sorting system in Gram-negative bacteria. J. Bacteriol. 193, 5707-5715. https://doi.org/10.1128/JB.05426-11.

Csűös, M., 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26, 1910-1912. https://doi.org/10.1093/bioinformatics/btq315.

Cummings, T. F. M., Gori, K., Sanchez-Pulido, L., Gavriilidis, G., Moi, D., Wilson, A. R., Murchison, E., Dessimoz, C., Ponting, C. P., Christophorou, M. A., 2021. Citrullination was introduced into animals by horizontal gene transfer from cyanobacteria. Mol. Biol. Evol. 39, msab317. https://doi.org/10.1093/molbev/msab317.

Cunningham, B. R., John, S. G., 2017. The effect of iron limitation on cyanobacteria major nutrient and trace element stoichiometry. Limnol. Oceanogr. 62, 846-858. https://doi.org/10.1002/lno.10484.

Czech, L., Hermann, L., Stöveken, N., Richter, A. A., Höppner, A., Smits, S. H. J., Heider, J., Bremer, E., 2018. Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: genetics, phylogenomics, biochemistry, and structural analysis. Genes 9(4), 177. https://doi.org/10.3390/genes9040177.

Daane, L. L., Molina, J. A., Berry, E. C., Sadowsky, M. J., 1996. Influence of earthworm activity on gene transfer from *Pseudomonas fluorescens* to indigenous soil bacteria. Appl. Environ. Microbiol. 62, 515-521. https://doi.org/10.1128/aem.62.2.515-521.1996.

Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., Durand, D., 2017. Xenolog classification. Bioinformatics 33, 640-649. https://doi.org/10.1093/bioinformatics/btw686.

Darriba, D., Taboada, G. L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164-1165. https://doi.org/10.1093/bioinformatics/btr088.

Darriba, D., Taboada, G. L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772-772. https://doi.org/10.1038/nmeth.2109.

Datta, P., Bhadra, R., 1978. Biodegradative threonine dehydratase. Reduction of ferricyanide by an intermediate of the enzyme-catalyzed reaction. Eur. J. Biochem. 91, 527-532. https://doi.org/10.1111/j.1432-1033.1978.tb12706.x.

David, L. A., Alm, E. J., 2011. Rapid evolutionary innovation during an Archaean genetic expansion. Nature 469, 93-96. https://doi.org/10.1038/nature09649.

Dawlaty, J., Zhang, X., Fischbach, M. A., Clardy, J., 2010. Dapdiamides, tripeptide antibiotics formed by unconventional amide ligases. J. Nat. Prod. 73, 441-446. https://doi.org/10.1021/np900685z.

de Crécy-Lagard, V., Hanson, A., 2013. Comparative Genomics. In: Maloy, S., Hughes, K. (Eds.), Brenner's Encyclopedia of Genetics (Second Edition). Academic Press, San Diego, pp. 102-105.

de Jong, A., van Heel, A. J., Kok, J., Kuipers, O. P., 2010. BAGEL2: mining for bacteriocins in genomic data. Nucleic Acids Res. 38, W647-W651. https://doi.org/10.1093/nar/gkq365.

de Lorenzo, V., Bindereif, A., Paw, B. H., Neilands, J. B., 1986. Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. J. Bacteriol. 165(2), 570-578. https://doi.org/10.1128/jb.165.2.570-578.1986.
de Lorenzo, V., Neilands, J. B., 1986. Characterization of iucA and iucC genes of the aerobactin system of plasmid ColV-K30 in *Escherichia coli*. J. Bacteriol. 167(1), 350-355. https://doi.org/10.1128/jb.167.1.350-355.1986.

de los Ríos, A., Wierzchos, J., Sancho, L. G., Ascaso, C., 2004. Exploring the physiological state of continental Antarctic endolithic microorganisms by microscopy. FEMS Microbiol. Ecol. 50, 143-152. https://doi.org/10.1016/j.femsec.2004.06.010.

Degani, G., Barbiroli, A., Regazzoni, L., Popolo, L., Vanoni, M. A., 2018. Imine deaminase activity and conformational stability of UK114, the mammalian member of the Rid protein family active in amino acid metabolism. Int. J. Mol. Sci. 19, 945. https://doi.org/10.3390/ijms19040945.

Deka, G., Bisht, S., Savithri, H. S., Murthy, M. R. N., 2018. Comparative structural and enzymatic studies on *Salmonella typhimurium* diaminopropionate ammonia lyase reveal its unique features. J. Struct. Biol. 202, 118-128. https://doi.org/10.1016/j.jsb.2017.12.012.

Doamekpor, L. K., Onwona-Agyeman, R., Ameka, G. K., 2019. Bioenergy: biodiesel from freshwater green microalgae and a cyanobacterium occurring in ghana. West Afr. J. Appl. Ecol. 27, 51-60.

Dodds, W. K., Castenholz, R. W., 1987. Effects of grazing and light on the growth of *Nostoc pruniforme* (Cyanobacteria). Br. Phycol. J. 23, 219-227. https://doi.org/10.1080/00071618800650251.

Dortch, Q., Parsons, M. L., Rabalais, N. N., Turner, R. E., 1999. What is the threat of harmful algal blooms in Louisiana Coastal Waters? In: Rozas, L. P., Nyman, J. A., Proffitt, C. E., Rabalais, N. N., Reed, D., J., Truner, R. E. (eds) Recent research in Costal Louisiana: natural system function and response to human influences. Louisiana Sea Grant College Program, Baton Rouge.

Dossaji, S. F., Bell, E. A., 1973. Distribution of α-amino-β-methylaminopropionic acid in *Cycas*. Phytochemistry 12, 143-144. https://doi.org/10.1016/S0031-9422(00)84634-8.

Downing, S., Banack, S. A., Metcalf, J. S., Cox, P. A., Downing, T. G., 2011. Nitrogen starvation of cyanobacteria results in the production of β-*N*-methylamino-L-alanine. Toxicon 58, 187-194. https://doi.org/10.1016/j.toxicon.2011.05.017.

Downing, S., Downing, T. G., 2016. The metabolism of the non-proteinogenic amino acid β-*N*-methylamino-L-alanine (BMAA) in the cyanobacterium *Synechocystis* PCC6803. Toxicon 115, 41-48. https://doi.org/10.1016/j.toxicon.2016.03.005.

Downs, D. M., Ernst, D. C., 2015. From microbiology to cancer biology: the Rid protein family prevents cellular damage caused by endogenously generated reactive nitrogen species. Mol. Microbiol. 96, 211-219. https://doi.org/10.1111/mmi.12945.

Doyon, J.-P., Ranwez, V., Daubin, V., Berry, V., 2011. Models, algorithms and programs for phylogeny reconciliation. Brief. Bioinformatics 12, 392-400. https://doi.org/10.1093/bib/bbr045.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307-1320. https://doi.org/10.1093/genetics/161.3.1307.

Dulić, T., Svirčev, Z., Palanački Malešević, T., Faassen, E. J., Savela, H., Hao, Q., Meriluoto, J., 2022. Assessment of common cyanotoxins in cyanobacteria of biological loess crusts. Toxins 14, 215. https://doi.org/10.3390/toxins14030215.

Dunlop, R. A., Banack, S. A., Bishop, S. L., Metcalf, J. S., Murch, S. J., Davis, D. A., Stommel, E. W., Karlsson, O., Brittebo, E. B., Chatziefthimiou, A. D., 2021. Is exposure to BMAA a risk factor for neurodegenerative diseases? A response to a critical review of the BMAA hypothesis. Neurotox. Res. 39, 81-106. https://doi.org/10.1007/s12640-020-00302-0.

Dutta, D., De, D., Chaudhuri, S., Bhattacharya, S. K., 2005. Hydrogen production by Cyanobacteria. Microb. Cell Fact. 4, 1-11. https://doi.org/10.1186/1475-2859-4-36.

Eddy, S. R., 1998. Profile hidden Markov models. Bioinformatics 14, 755-763. https://doi.org/10.1093/bioinformatics/14.9.755.

Eddy, S. R., 2004. Where did the BLOSUM62 alignment score matrix come from? Nat. Biotechnol. 22, 1035-1036. https://doi.org/10.1038/nbt0804-1035.

Edwards, K. J., Gihring, T. M., Banfield, J. F., 1999. Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. Appl. Environ. Microbiol. 65, 3627-3632. https://doi.org/10.1128/AEM.65.8.3627-3632.1999.

Efron, B., 1992. Bootstrap methods: another look at the jackknife. In: Kotz, S., Johnson, N. L. (Eds.) Breakthroughs in statistics. Springer Series in Statistics. Springer, New York, pp. 569-593. https://doi.org/10.1007/978-1-4612-4380-9_41.

Eiger, I. Z, Greenstein, J. P, 1948. Addition products of dehydropeptides. Arch. Biochem. 19, 467-473.

El-Baz, F. K., El-Senousy, W. M., El-Sayed, A. B., Kamel, M. M., 2013. *In vitro* antiviral and antimicrobial activities of *Spirulina platensis* extract. J. Appl. Pharm. Sci 3, 52-56. DOI: https://doi.org/10.7324/JAPS.2013.31209.

Emmert, E. A. B., Klimowicz, A. K., Thomas, M. G., Handelsman, J., 2004. Genetics of zwittermicin A production by *Bacillus cereus*. Appl. Environ. Microbiol. 70, 104-113. https://doi.org/10.1128/AEM.70.1.104-113.2004.

Emms, D. M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. https://doi.org/10.1186/s13059-015-0721-2.

Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., Gerwick, W. H., 2012. *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Int. J. Syst. Evol. Microbiol. 62, 1171-1178. https://doi.org/10.1099/ijs.0.033761-0.

Eriksson, J., Jonasson, S., Papaefthimiou, D., Rasmussen, U., Bergman, B., 2009. Improving derivatization efficiency of BMAA utilizing AccQ-Tag® in a complex cyanobacterial matrix. Amino Acids 36, 43-48. https://doi.org/10.1007/s00726-007-0023-4.

Ernst, B., Neser, S., O'Brien, E., Hoeger, S. J., Dietrich, D. R., 2006. Determination of the filamentous cyanobacteria *Planktothrix rubescens* in environmental water samples using an image processing system. Harmful Algae 5, 281-289. https://doi.org/10.1016/j.hal.2005.08.003.

Ernst, D. C., Anderson, M. E., Downs, D. M., 2016. L-2,3-diaminopropionate generates diverse metabolic stresses in *Salmonella enterica*. Mol. Microbiol. 101, 210-223. https://doi.org/10.1111/mmi.13384.

Faassen, E. J., 2017. Extraction and LC-MS/MS Analysis of Underivatised BMAA, in: Meriluoto, J., Spoof, L., Codd, G.A. (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis. John Wiley & Sons, Chichester, UK, pp. 439-446. https://doi.org/10.3390/md14030045.

Faassen, E. J., Antoniou, M. G., Beekman-Lukassen, W., Blahova, L., Chernova, E., Christophoridis, C., Combes, A., Edwards, C., Fastner, J., Harmsen, J., 2016. A collaborative evaluation of LC-MS/MS based methods for BMAA analysis: soluble bound BMAA found to be an important fraction. Mar. Drugs 14(3), 45. https://doi.org/10.3390/md14030045.

Faassen, E. J., Gillissen, F., Lürling, M., 2012. A comparative study on three analytical methods for the determination of the neurotoxin BMAA in cyanobacteria. PLoS One 7, e36667. https://doi.org/10.1371/journal.pone.0036667.

Faassen, J. E., 2014. Presence of the neurotoxin BMAA in aquatic ecosystems: what do we really know? Toxins 6, 1109-1138. https://doi.org/10.3390/toxins6031109.

Falcón, L. I., Magallón, S., Castillo, A., 2010. Dating the cyanobacterial ancestor of the chloroplast. ISME J. 4, 777-783. https://doi.org/10.1038/ismej.2010.2.

Fauman, E. B., Blumenthal, R. M., Cheng, X., 1999. Structure and evolution of Adomet-dependent methyltransferases, in: Cheng, X., Blumenthal, R. M. (Eds.), S-Adenosylmethionine-Dependent Methyltransferases. World Scientific Publishing, New Jersey, pp. 1-38. https://doi.org/10.1142/9789812813077_0001.

Felnagle, E. A., Rondon, M. R., Berti, A. D., Crosby, H. A., Thomas, M. G., 2007. Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Appl. Environ. Microbiol. 73, 4162-4170. https://doi.org/10.1128/AEM.00485-07.

Felsenstein, J., 1978. The number of evolutionary trees. Syst. Zool. 27, 27-33. https://doi.org/10.2307/2412810.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783-791. https://doi.org/10.2307/2408678.

Fisch, K. M., 2013. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. RSC Adv. 3(40), 18228-18247. https://doi.org/10.1039/C3RA42661K.

Fitch, W. M., 1970. Distinguishing homologous from analogous proteins. Syst. Zool. 19, 99-113. https://doi.org/10.2307/2412448.

Fitch, W. M., 2000. Homology: a personal view on some of the problems. Trends Genet. 16, 227-231. https://doi.org/10.1016/S0168-9525(00)02005-9.

Flannery, D. T., Walter, M. R., 2012. Archean tufted microbial mats and the Great Oxidation Event: new insights into an ancient problem. Aust. J. Earth Sci. 59, 1-11. https://doi.org/10.1080/08120099.2011.607849.

Fleischmann R. D., Adams M. D., White, O., Clayton R. A., Kirkness E. F., Kerlavage A. R., Bult C. J., Tomb, J.-F., Dougherty B. A., Merrick J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne J. D., Scott, J., Shirley, R., Liu, L.-l., Glodek, A., Kelley J. M., Weidman J. F., Phillips C. A., Spriggs, T., Hedblom, E., Cotton M. D., Utterback T. R., Hanna M. C., Nguyen D. T., Saudek D. M., Brandon R. C., Fine L. D., Fritchman J. L., Fuhrmann J. L., Geoghagen, N. S. M., Gnehm C. L., McDonald L. A., Small K. V., Fraser C. M., Smith H. O., Venter, J. C., 1995. Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. Science 269, 496-512. https://doi.org/10.1126/science.7542800.

Flicek, P., Keibler, E., Hu, P., Korf, I., Brent, M. R., 2003. Leveraging the mouse genome for gene prediction in human: from whole-genome shotgun reads to a global synteny map. Genome Res. 13, 46-54. https://doi.org/10.1101/gr.830003.

Flissi, A., Ricart, E., Campart, C., Chevalier, M., Dufresne, Y., Michalik, J., Jacques, P., Flahaut, C., Lisacek, F., Leclère, V., 2020. Norine: update of the nonribosomal peptide resource. Nucleic Acids Res. 48(D1), D465-D469. https://doi.org/10.1093/nar/gkz1000.

Flores, E., Herrero, A., 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. Nat. Rev. Microbiol. 8, 39-50. https://doi.org/10.1038/nrmicro2242.

Flouri, T., Izquierdo-Carrasco, F., Darriba, D., Aberer, A. J., Nguyen, L. T., Minh, B. Q., Von Haeseler, A., Stamatakis, A., 2015. The phylogenetic likelihood library. Syst. Biol. 64, 356-362. https://doi.org/10.1093/sysbio/syu084.

Flynn, J. M., Downs, D. M., 2013. In the absence of RidA, endogenous 2-aminoacrylate inactivates alanine racemases by modifying the pyridoxal 5′-phosphate cofactor. J. Bacteriol. 195, 3603-3609. https://doi.org/10.1128/JB.00463-13.

Forterre, P., 2016. Microbes from Hell. Microbes from Hell. University of Chicago Press, pp. 1-288.

Fox, G. E., Wisotzkey, J. D., Jurtshuk Jr, P., 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Evol. Microbiol. 42, 166-170. https://doi.org/10.1099/00207713-42-1-166.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., 1995. The minimal gene complement of *Mycoplasma genitalium*. Science 270, 397-404. https://doi.org/10.1126/science.270.5235.397

Freckleton, R. P., 2002. On the misuse of residuals in ecology: regression of residuals vs. multiple regression. J. Anim. Ecol., 542-545. https://doi.org/10.1046/j.1365-2656.2002.00618.x.

Frishman, D., Mokrejs, M., Kosykh, D., Kastenmüller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., 2003. The PEDANT genome database. Nucleic Acids Res. 31, 207-211. https://doi.org/10.1093/nar/gkg005.

Fuchsman, C. A., Collins, R. E., Rocap, G., Brazelton, W. J., 2017. Effect of the environment on horizontal gene transfer between bacteria and archaea. PeerJ 5, e3865. https://doi.org/10.7717/peerj.3865.

Funari, E., Testai, E., 2008. Human health risk assessment related to cyanotoxins exposure. Crit. Rev. Toxicol. 38, 97-125. https://doi.org/10.1080/10408440701749454.

Furnes, H., Banerjee, N. R., Muehlenbachs, K., Staudigel, H., de Wit, M., 2004. Early life recorded in Archean pillow lavas. Science 304, 578-581. https://doi.org/10.1126/science.1095858.

Gadagkar, S. R., Rosenberg, M. S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B Mol. Dev. Evol. 304, 64-74. https://doi.org/10.1002/jez.b.21026.

Galtier, N., 2007. A model of horizontal gene transfer and the bacterial phylogeny problem. Syst. Biol. 56(4), 633-642. https://doi.org/10.1080/10635150701546231.

Garcia-Pichel, F., 2009. Cyanobacteria. In: Schaechter, M. (Ed.), Encyclopedia of Microbiology (Third Edition). Academic Press, Oxford, pp. 107-124.

Garcia-Pichel, F., Nubel, U., Muyzer, G., 1998. The phylogeny of unicellular, extremely halotolerant cyanobacteria. Arch. Microbiol. 169, 469-482. https://doi.org/10.1007/s002030050599.

Garcia-Gonzalez, E., Müller, S., Hertlein, G., Heid, N., Süssmuth, R. D., Genersch, E., 2014. Biological effects of paenilamicin, a secondary metabolite antibiotic produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. MicrobiologyOpen 3, 642-656. https://doi.org/10.1002/mbo3.195.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C., Ritchie, M. G., 2008. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. Mol. Ecol. 17, 1648-1657. https://doi.org/10.1111/j.1365-294X.2008.03713.x.

Gaut, B. S., Lewis, P. O., 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol. Biol. Evol. 12, 152-162. https://doi.org/10.1093/oxfordjournals.molbev.a040183.

Geiger, O., Röhrs, V., Weissenmayer, B., Finan, T. M., Thomas-Oates, J. E., 1999. The regulator gene phoB mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl-*N,N,N*-trimethylhomoserine in *Rhizobium* (*Sinorhizobium*) *meliloti*. Mol. Microbiol. 32(1) 63-73. https://doi.org/10.1046/j.1365-2958.1999.01325.x.

Gerwick, W. H., Roberts, M. A., Proteau, P. J., Chen, J.-L., 1994. Screening cultured marine microalgae for anticancer-type activity. J. Appl. Phycol. 6, 143-149. https://doi.org/10.1007/BF02186068.

Giménez, M., Azziz, G., Gill, P. R., Batista, S., 2019. Horizontal gene transfer elements: plasmids in Antarctic microorganisms. In: Castro-Sowinski, S. (eds) The Ecological Role of Micro-organisms in the Antarctic Environment. Springer: Cham, Switzerland, pp. 85-107.

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J., Pace, N. R., 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. 170, 3584-3592. https://doi.org/10.1128/jb.170.8.3584-3592.1988.

Gmelin, R., 1959. The free amino acids in the seeds of *Acacia willardiana* (Mimosaceae). Isolation of willardiin, a new plant amino acid which is probably L-beta-(3-uracil)-alpha-aminopropionic acid. Hoppe-Seylers Z. Physiol. Chem. 316, 164-169. https://doi.org/10.1515/bchm2.1959.316.1.164.

Gobler, C. J., 2020. Climate change and Harmful Algal Blooms: insights and perspective. Harmful Algae 91, 101731. https://doi.org/10.1016/j.hal.2019.101731.

Godde, J. S., Baichoo, S., Mungloo-Dilmohamud, Z., Jaufeerally-Fakim, Y., 2018. Comparison of genomic islands in cyanobacteria: evidence of bacteriophage-mediated horizontal gene transfer from eukaryotes. Microbiol. Res. 211, 31-46. https://doi.org/10.1016/j.micres.2018.03.005.

Goldman, S. J., Lammers, P. J., Berman, M. S., Sanders-Loehr, J., 1983. Siderophore-mediated iron uptake in different strains of *Anabaena* sp. J. Bacteriol. 156(3), 1144-1150. https://doi.org/10.1128/jb.156.3.1144-1150.1983.

Golubchik, T., Wise, M. J., Easteal, S., Jermiin, L. S., 2007. Mind the gaps: evidence of bias in estimates of multiple sequence alignments. Mol. Biol. Evol. 24, 2433-2442. https://doi.org/10.1093/molbev/msm176.

Golubic, S., 1980. Halophily and halotolerance in cyanophytes Orig. Life Evol. Biosph. 10: 169–183. https://doi.org/10.1007/BF00928667.

Gophna, U., Kristensen, D. M., Wolf, Y. I., Popa, O., Drevet, C., Koonin, E. V., 2015. No evidence of inhibition of horizontal gene transfer by CRISPR–Cas on evolutionary timescales. ISME J. 9, 2021-2027. https://doi.org/10.1038/ismej.2015.20.

Goupil-Feuillerat, N., Cocaign-Bousquet, M., Godon, J.-J., Ehrlich, S. D., Renault, P., 1997. Dual role of alpha-acetolactate decarboxylase in *Lactococcus lactis* subsp. lactis. J. Bacteriol. 179, 6285-6293. https://doi.org/10.1128/jb.179.20.6285-6293.1997.

Goyal, A., Gelbwaser-Klimovsky, D., Gore, J., 2020. Horizontal gene transfer becomes disadvantageous in rapidly fluctuating environments. bioRxiv. https://doi.org/10.1101/2020.08.07.241406.

Grant, W. D., 1986. The alkaline saline environment. In: Herbert, R. A. and Codd, G. A. (Eds.) Microbes in extreme environments, Acadmic Press, London, pp. 25-54.

Gray, G. S., Fitch, W. M., 1983. Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus*. Mol. Biol. Evol. 1, 57-66. https://doi.org/10.1093/oxfordjournals.molbev.a040298.

Griffith, F., 1928. The significance of Pneumococcal types. J. Hyg. 27, 113-159. https://doi.org/10.1017/S0022172400031879.

Griffiths, D. J., Saker, M. L., 2003. The Palm Island mystery disease 20 years on: a review of research on the cyanotoxin cylindrospermopsin. Environ. Toxicol. 18, 78-93. https://doi.org/10.1002/tox.10103.

Griffiths, G. L., Sigel, S. P., Payne, S. M., Neilands, J. B., 1984. Vibriobactin, a siderophore from *Vibrio cholerae*. J. Biol. Chem. 259(1), 383-385. https://doi.org/10.1016/S0021-9258(17)43671-4.

Grigorieva, G., Shestakov, S., 1982. Transformation in cyanobacterium *Synechocystis* sp. 6803. FEMS Microbiol. Lett. 13, 367-370. https://doi.org/10.1111/j.1574-6968.1982.tb08289.x.

Gugger, M. F., Hoffmann, L., 2004. Polyphyly of true branching cyanobacteria (Stigonematales). Int. J. Syst. Evol. Microbiol. 54, 349-357. https://doi.org/10.1099/ijs.0.02744-0.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307-321. https://doi.org/10.1093/sysbio/syq010.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696-704. https://doi.org/10.1080/10635150390235520.

Guljamow, A., Jenke-Kodama, H., Saumweber, H., Quillardet, P., Frangeul, L., Castets, A. M., Bouchier, C., Tandeau de Marsac, N., Dittmann, E., 2007. Horizontal gene transfer of two cytoskeletal elements fro a eukaryote to a cyanobacterium. Curr. Biol. 17, R757-R759. https://doi.org/10.1016/j.cub.2007.06.063.

Guo, J., Wang, Q., Wang, X., Wang, F., Yao, J., Zhu, H., 2015. Horizontal gene transfer in an acid mine drainage microbial community. BMC Genom. 16, 1-11. https://doi.org/10.1186/s12864-015-1720-0.

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res. 46, D851-D860. https://doi.org/10.1093/nar/gkx1068.

Haft, D. H., Selengut, J. D., White, O., 2003. The TIGRFAMs database of protein families. Nucleic Acids Res. 31, 371-373. https://doi.org/10.1093/nar/gkg128.

Hagemann, M., Henneberg, M., Felde, V. J., Berkowicz, S. M., Raanan, H., Pade, N., Felix-Henningsen, P., Kaplan, A., 2017. Cyanobacterial populations in biological soil crusts of the northwest Negev Desert, Israel-effects of local conditions and disturbance. FEMS Microbiol. Ecol. 93, fiw228. https://doi.org/10.1093/femsec/fiw228.

Hall, R. J., Whelan, F. J., McInerney, J. O., Ou, Y., Domingo-Sananes, M. R., 2020. Horizontal gene transfer as a source of conflict and cooperation in prokaryotes. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.01569.

Hanfrey, C. C., Pearson, B. M., Hazeldine, S., Lee, J., Gaskin, D. J., Woster, P. M., Phillips, M. A., Michael, A. J., 2011. Alternative spermidine biosynthetic route is critical for growth of *Campylobacter jejuni* and is the dominant polyamine pathway in human gut microbiota. J. Biol. Chem. 286(50), 43301-43312. https://doi.org/10.1074/jbc.M111.307835.

Hardiman, C. A., Weingarten, R. A., Conlan, S., Khil, P., Dekker, J. P., Mathers, A. J., Sheppard, A. E., Segre, J. A., Frank, K. M., 2016. Horizontal transfer of carbapenemase-encoding plasmids and comparison with hospital epidemiology data. Antimicrob. Agents Chemother. 60, 4910-4919. https://doi.org/10.1128/AAC.00014-16.

Hardison, R. C., 2003. Comparative genomics. PLoS Biol. 1, e58. https://doi.org/10.1371/journal.pbio.0000058.

Harland, F. M. J., Wood, S. A., Moltchanova, E., Williamson, W. M., Gaw, S., 2013. *Phormidium autumnale* growth and anatoxin-a production under iron and copper stress. Toxins 5. https://doi.org/10.3390/toxins5122504.

Hayes, P. K., El Semary, N. A., Sánchez-Baracaldo, P., 2007. The taxonomy of cyanobacteria: molecular insights into a difficult problem. In: Brodie, J. and Lewis, J. (Eds.) Unravlling the Algae. Natural History Museum, CRC Press, pp. 93-101.

Hehnly, C., Zhang, L., Paulson, J. N., Almeida, M., von Bredow, B., Wijetunge, D. S. S., Galperin, M. Y., Sheldon, K., Schiff, S. J., Broach, J. R., 2020. Complete genome sequences of the human pathogen *Paenibacillus thiaminolyticus* Mbale and type Strain *P. thiaminolyticus* NRRL B-4156. Microbiol. Resour. Announc. 9, e00181-00120. https://doi.org/10.1128/MRA.00181-20.

Heine, A., Canaves, J. M., Von Delft, F., Brinen, L. S., Dai, X., Deacon, A. M., Elsliger, M. A., Eshaghi, S., Floyd, R., Godzik, A., 2004. Crystal structure of *O*-acetylserine sulfhydrylase (TM0665) from *Thermotoga maritima* at 1.8 Å resolution. Proteins 56, 387-391. https://doi.org/10.1002/prot.20003.

Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., Barry, K., Tringe, S. G., Watson, D. B., He, Z., Hazen, T. C., Tiedje, J. M., Rubin, E. M., Zhou, J., 2010. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. ISME J. 4, 660-672. https://doi.org/10.1038/ismej.2009.154.

Hernandez, C. E., Rodríguez-Serrano, E., Avaria-Llautureo, J., Inostroza-Michael, O., Morales-Pallero, B., Boric-Bargetto, D., Canales-Aguirre, C. B., Marquet, P. A., Meade, A., 2013. Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. Methods Ecol. Evol. 4, 401-415. https://doi.org/10.1111/2041-210X.12033.

Higgins, S. N., Kling, H. J., Hecky, R. E., Taylor, W. D., Bootsma, H. A., 2003. The community composition, distribution, and nutrient status of epilithic peripython at five rocky littoral zones sites in Lake Malawi, Africa. J. Gt. Lakes Res. 29, 181-189. https://doi.org/10.1016/S0380-1330(03)70547-4.

Hillis, D. M., Bull, J. J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182-192. https://doi.org/10.1093/sysbio/42.2.182.

Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R., Molineux, I. J., 1992. Experimental phylogenetics: generation of a known phylogeny. Science 255, 589-592. https://doi.org/10.1126/science.1736360.

Hindák, F., 2012. Hormogonia in two nostocalean cyanophytes (cyanobacteria) from the genera *Hapalosiphon* and *Fischerella*. Biologia 67, 1075-1079. https://doi.org/10.2478/s11756-012-0100-3.

Hitzfeld, B. C., Höger, S. J., Dietrich, D. R., 2000. Cyanobacterial toxins: removal during drinking water treatment, and human risk assessment. Environ. Health Perspect. 108, 113-122. https://doi.org/10.1289/ehp.00108s1113.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Vinh, L. S., 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35(2), 518-522. https://doi.org/10.1093/molbev/msx281.

Holland, H. D., 1978. The chemistry of the atmosphere and oceans. Wileu New York, 341.

Hollenhorst, M. A., Bumpus, S. B., Matthews, M. L., Bollinger Jr, J. M., Kelleher, N. L., Walsh, C. T., 2010. The NRPS enzyme DdaD tethers Nβ-fumaramoyl-DAP for Fe (II)/α-ketoglutarate-dependent epoxidation by DdaC during dapdiamide antibiotic biosynthesis. J. Am. Chem. Soc. 132, 15773-15781. https://doi.org/10.1021/ja1072367.

Holm, L., Sander, C., 1997. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins 28, 72-82. https://doi.org/10.1002/(SICI)1097-0134(199705)28:1<72::AID-PROT7>3.0.CO;2-L.

Holtcamp, W., 2012. The Emerging Science of BMAA: Do cyanobacteria contribute to neurodegenerative disease? Environ. Health Perspect. 120, a110-a116. https://doi.org/10.1289/ehp.120-a110.

Homann, M., Heubeck, C., Airo, A., Tice, M. M., 2015. Morphological adaptations of 3.22 Ga-old tufted microbial mats to Archean coastal habitats (Moodies Group, Barberton Greenstone Belt, South Africa). Precambrian Res. 266, 47-64. https://doi.org/10.1016/j.precamres.2015.04.018.

Honda, D., Yokota, A., Sugiyama, J., 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. J. Mol. Evol. 48, 723-739. https://doi.org/10.1007/PL00006517.

Horikoshi, K., Antranikian, G., Bull, A. T., Robb, F. T., Stetter, K. O., 2010. Extremophiles handbook. Springer Science & Business Media, pp. 19-26.

Hosoya, R., Hamana, K., Isobe, M., Yokota, A., 2005. Polyamine distribution profiles within cyanobacteria. Microbiol. Cult. Coll. 21, 3-8.

Hsu, C.-C., Lu, L.-Y., Yang, Y.-S., 2010. From sequence and structure of sulfotransferases and dihydropyrimidinases to an understanding of their mechanisms of action and function. Expert Opin. Drug Metab. Toxicol. 6, 591-601. https://doi.org/10.1517/17425251003601987.

Hsu, S.-H., Zhang, S., Huang, S.-C., Wu, T.-K., Xu, Z., Chang, C.-Y., 2020. Characterization of enzymes catalyzing the formation of the nonproteinogenic amino acid l-Dap in capreomycin biosynthesis. Biochemistry 60, 77-84. https://doi.org/10.1021/acs.biochem.0c00808.

Huang, C.-Y., 2015. Inhibition of a putative dihydropyrimidinase from *Pseudomonas aeruginosa* PAO1 by flavonoids and substrates of cyclic amidohydrolases. PLoS One 10, e0127634. https://doi.org/10.1371/journal.pone.0127634.

Huber, A. L., 1986. Nitrogen fixation by *Nodularia spumigena* Martens (Cyanobacteria). I. Field studies on the contribution of blooms to the nitrogen budget of the Peel-Harvey Estuary, Western Australia. Hydrobiologia 131, 193-203.

Huelsenbeck, J. P., 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44, 17-48. https://doi.org/10.1093/sysbio/44.1.17.

Huertas, M. J., López-Maury, L., Giner-Lamia, J., Sánchez-Riego, A. M., Florencio, F. J., 2014. Metals in cyanobacteria: analysis of the copper, nickel, cobalt and arsenic homeostasis mechanisms. Life 4, 865-886. https://doi.org/10.3390/life4040865.

Hughey, R., Krogh, A., 1996. Hidden Markov models for sequence analysis: extension and analysis of the basic method. Bioinformatics 12, 95-107. https://doi.org/10.1093/bioinformatics/12.2.95.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., Sigrist, C. J. A., 2006. The PROSITE database. Nucleic Acids Res. 34, D227-D230. https://doi.org/10.1093/nar/gkj063.

Humbert, J. F., 2017. Molecular tools for the detection of toxigenic Cyanobacteria in natural ecosystems, in: Meriluoto, J., Spoof, L., Codd, G.A. (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis. John Wiley & Sons, Chichester, UK, pp. 280-283. https://doi.org/10.1002/9781119068761.ch28.

Humphrey, J. M., Chamberlin, A. R., 1997. Chemical synthesis of natural product peptides: coupling methods for the incorporation of noncoded amino acids into peptides. Chem. Rev. 97, 2243-2266. https://doi.org/10.1021/cr950005s.

Husnik, F., McCutcheon, J. P., 2018. Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. 16, 67-79. https://doi.org/10.1038/nrmicro.2017.137.

Huson, D. H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23(2), 254-267. https://doi.org/10.1093/molbev/msj030.

Igarashi, K., Kashiwagi, K., 2000. Polyamines: mysterious modulators of cellular functions. Biochem. Biophys. Res. Commun. 271(3), 559-564. https://doi.org/10.1006/bbrc.2000.2601.

Ikai, H., Yamamoto, S., 1997. Identification and analysis of a gene encoding L-2,4-diaminobutyrate: 2-ketoglutarate 4-aminotransferase involved in the 1,3-diaminopropane production pathway in *Acinetobacter baumannii*. J. Bacteriol. 179(16), 5118-5125. https://doi.org/10.1128/jb.179.16.5118-5125.1997.

Ikegami, F., Itakagi, S., Ishikawa, T., Ongena, G., Kuo, Y.-H., Lambein, F., Murakoshi, I., 1991. Biosynthesis of β-(isoxazolin-5-on-2-yl) alanine, the precursor of the neurotoxic amino acid β-*N*-oxalyl-L-α, β-diaminopropionic acid. Chem. Pharm. Bull. 39(12), 3376-3377. https://doi.org/10.1248/cpb.39.3376.

Ikegami, F., Murakoshi, I., 1994. Enzymic synthesis of non-protein β-substituted alanines and some higher homologues in plants. Phytochemistry 35, 1089-1104. https://doi.org/10.1016/S0031-9422(00)94805-2.

Ikegami, F., Sakai, R., Ishikawa, T., Kuo, Y.-H., Lambein, F., Murakoshi, I., 1993. Biosynthesis in vitro of 2-(3-amino-3-carboxypropyl)-isoxazolin-5-one, the neurotoxic amino acid in *Lathyrus odoratus*. Biol. Pharm. Bull. 16(7), 732-734. https://doi.org/10.1248/bpb.16.732.

Ikegami, F., Yamamoto, A., Kuo, Y.-H., Lambein, F., 1999. Enzymatic formation of 2,3-diaminopropionic acid, the direct precursor of the neurotoxin β-ODAP, in *Lathyrus sativus*. Biol. Pharm. Bulletin 22, 770-771. https://doi.org/10.1248/bpb.22.770.

Imhoff, J. F., Sahl, H. G., Soliman, G. S. H., Trüper, H. G., 1979. The Wadi Natrun: chemical composition and microbial mass developments in alkaline brines of eutrophic desert lakes. Geomicrobiol. J. 1, 219-234. https://doi.org/10.1080/01490457909377733.

Irons, J. L., Hodge-Hanson, K., Downs, D. M., 2020. RidA proteins protect against metabolic damage by reactive intermediates. Microbiol. Mol. Biol. Rev. 84, e00024-00020. https://doi.org/10.1128/MMBR.00024-20.

Ishida, T., Watanabe, M. M., Sugiyama, J., Yokota, A., 2001. Evidence for polyphyletic origin of the members of the orders of Oscillatoriales and Pleurocapsales as determined by 16S rDNA analysis. FEMS Microbiol. Lett. 201, 79-82. https://doi.org/10.1111/j.1574-6968.2001.tb10736.x.

Iteman, I., Rippka, R., de Marsac, N. T., Herdman, M., 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. The GenBank accession numbers for the sequences reported in this paper are AF180968 and AF180969 for ITS-L and ITS-S, respectively. Microbiology 146, 1275-1286. https://doi.org/10.1099/00221287-146-6-1275.

Ito, Y., Butler, A., 2005. Structure of synechobactins, new siderophores of the marine cyanobacterium *Synechococcus* sp. PCC 7002. Limnol. Oceanogr. 50, 1918-1923. https://doi.org/10.4319/lo.2005.50.6.1918.

Jain, R., Rivera, M. C., Moore, J. E., Lake, J. A., 2002. Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. 61, 489-495. https://doi.org/10.1006/tpbi.2002.1596.

Jaki, B., Heilmann, J., Sticher, O., 2000. New antibacterial metabolites from the cyanobacterium *Nostoc commune* (EAWAG 122b). J. Nat Prod. 63, 1283-1285. https://doi.org/10.1021/np000033s.

Jander, G., Joshi, V., 2010. Recent progress in deciphering the biosynthesis of aspartate-derived amino acids in plants. Mol. Plant 3(1), 54-65. https://doi.org/10.1093/mp/ssp104.

Jansonius, J. N., Vincent, M.G, 1987. Structural basis for catalysis by aspartate aminotransferase, in: Jurnak, F.A., McPherson, A. (Eds.), Biological Macromolecules and Assemblies. Wiley, New York, pp. 187-285.

Javor, B., 1989. Hypersaline environments. Microbiology and Biogeochemistry. Springer, Berlin, pp. 328.

Jeanjean, R., Talla, E., Latifi, A., Havaux, M., Janicki, A., Zhang, C. C., 2008. A large gene cluster encoding peptide synthetases and polyketide synthases is involved in production of siderophores and oxidative stress response in the cyanobacterium *Anabaena* sp. strain PCC 7120. Environ. Microbiol. 10(10), 2574-2585. https://doi.org/10.1111/j.1462-2920.2008.01680.x.

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., Bork, P., 2007. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 36, D250-D254. https://doi.org/10.1093/nar/gkm796.

Jiang, L., Eriksson, J., Lage, S., Jonasson, S., Shams, S., Mehine, M., Ilag, L. L., Rasmussen, U., 2014. Diatoms: a novel source for the neurotoxin BMAA in aquatic environments. PloS One 9, e84578. https://doi.org/10.1371/journal.pone.0084578.

Jiang, L., Johnston, E., Åberg, K. M., Nilsson, U., Ilag, L. L., 2013. Strategy for quantifying trace levels of BMAA in cyanobacteria by LC/MS/MS. Anal. Bioanal. Chem. 405(4), 1283-1292. https://doi.org/10.1007/s00216-012-6550-1.

Johnsborg, O., Eldholm, V., Håvarstein, L. S., 2007. Natural genetic transformation: prevalence, mechanisms and function. Res. Microbiol. 158, 767-778. https://doi.org/10.1016/j.resmic.2007.09.004.

Jones, A. C., Monroe, E. A., Eisman, E. B., Gerwick, L., Sherman, D. H., Gerwick, W. H., 2010. The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. Nat. Prod. Rep. 27, 1048-1065. https://doi.org/10.1039/C000535E.

Jørgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., Bak, S., Møller, B. L., 2005. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr. Opin. Plant Biol. 8, 280-291. https://doi.org/10.1016/j.pbi.2005.03.014.

Kajiyama, S.-i., Kanzaki, H., Kawazu, K., Kobayashi, A., 1998. Nostofungicidine, an antifungal lipopeptide from the field-grown terrestrial blue-green alga *Nostoc commune*. Tetrahedron Lett. 39, 3737-3740. https://doi.org/10.1016/S0040-4039(98)00573-5.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., Jermiin, L. S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14(6), 587-589. https://doi.org/10.1038/nmeth.4285.

Kalyani, J. N., Ramachandra, N., Kachroo, A. H., Mahadevan, S., Savithri, H. S., 2012. Functional analysis of the genes encoding diaminopropionate ammonia lyase in *Escherichia coli* and *Salmonella enterica* serovar *typhimurium*. J. Bacteriol. 194, 5604-5612. https://doi.org/10.1128/JB.01362-12.

Kaplan-Levy, R. N., Hadas, O., Summers, M. L., Rücker, J., Sukenik, A., 2010. Akinetes: dormant cells of cyanobacteria. In: Lubzens, E., Cerda, J., Clarkm M. (Eds.) Dormancy and resistance in harsh environments. Topics in Current Genetics, vol. 21. Springer, Berlin, Heidelberg, pp. 5-27. https://doi.org/10.1007/978-3-642-12422-8_2.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D., Stal, L., 2002. Dinitrogen fixation in the world's oceans. Biogeochemistry 57, 47-98. https://doi.org/10.1023/A:1015798105851.

Kashyap, A. K., Dubey, S. K., Jain, B. P., 2022. Cyanobacterial diversity concerning the extreme environment and their bioprospecting. In: Cyanobacterial Lifestyle and its Applications in Biotechnology. Academic Press, pp. 1-22.

Katoh, K., Asimenos, G., Toh, H., 2009. Multiple alignment of DNA sequences with MAFFT. In: Posada, D. (Eds.) Bioinformatics for DNA Sequence Analysis. Methods in Molecular Biology, vol. 537. Humana Press. https://doi.org/10.1007/978-1-59745-251-9_3.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518. https://doi.org/10.1093/nar/gki198.

Katoh, K., Misawa, K., Kuma, K. i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066. https://doi.org/10.1093/nar/gkf436.

Katoh, K., Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4), 772-780. https://doi.org/10.1093/molbev/mst010.

Kauff, F., Büdel, B., 2010. Phylogeny of cyanobacteria: an overview. In: Lüttge, U., Beyschlag, W., Büdel, B., Francis, D. (Eds.) Progress in Botany vol. 72. Springer, Berlin, Heidelberg, pp. 209-224. https://doi.org/10.1007/978-3-642-13145-5_8.

Kaushik, B. D., Venkataraman, G. S., 1979. Effect of algal inoculation on the yield and vitamin C content of two varieties of tomato. Plant Soil 52, 135-137. https://doi.org/10.1007/BF02197740.

Keating, T. A., Marshall, C. G., Walsh, C. T., Keating, A. E., 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat. Struct. Biol. 9(7), 522-526. https://doi.org/10.1038/nsb810.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., Lander, E. S., 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423, 241-254. https://doi.org/10.1038/nature01644.

Kellmann, R., Michali, T. K., Neilan, B. A., 2008. Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers. J. Mol. Evol. 67(5), 526-538. https://doi.org/10.1007/s00239-008-9169-2.

Khalaf, E. M., Raizada, M. N., 2020. Draft genome sequences of *Bacillus* and *Paenibacillus* species isolated from seeds of *Citrullus lanata* (watermelon), *Cucurbita moschata* (butternut squash), and *Cucurbita pepo* L. var. pepo L.(pumpkin). Microbiol. Resour. Announc. 9, e00727-00720. https://doi.org/10.1128/MRA.00727-20.

Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., Fedorova, N. D., 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet. Biol. 47, 736-741. https://doi.org/10.1016/j.fgb.2010.06.003.

Khan, F., Jala, V. R., Rao, N. A., Savithri, H. S., 2003. Characterization of recombinant diaminopropionate ammonia-lyase from *Escherichia coli* and *Salmonella typhimurium*. Biochem. Biophys. Res. Commun. 306, 1083-1088. https://doi.org/10.1016/S0006-291X(03)01100-8.

Klähn, S., Hagemann, M., 2011. Compatible solute biosynthesis in cyanobacteria. Environ. Microbiol.13(3), 551-562. https://doi.org/10.1111/j.1462-2920.2010.02366.x.

Kloesges, T., Popa, O., Martin, W., Dagan, T., 2011. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. Mol. Biol. Evol. 28, 1057-1074. https://doi.org/10.1093/molbev/msq297.

Klümper, U., Droumpali, A., Dechesne, A., Smets, B. F., 2014. Novel assay to measure the plasmid mobilizing potential of mixed microbial communities. Front. Microbiol. 5, 730. https://doi.org/10.3389/fmicb.2014.00730.

Kobylarz, M. J., Grigg, J. C., Liu, Y., Lee, M. S. F., Heinrichs, D. E., Murphy, M. E. P., 2016. Deciphering the substrate specificity of SbnA, the enzyme catalyzing the first step in staphyloferrin B biosynthesis. Biochemistry 55, 927-939. https://doi.org/10.1021/acs.biochem.5b01045.

Kobylarz, M. J., Grigg, J. C., Shin-ichi, J. T., Rai, D. K., Heinrichs, D. E., Murphy, M. E. P., 2014. Synthesis of L-2,3-diaminopropionic acid, a siderophore and antibiotic precursor. Chem. Biol. 21, 379-388. https://doi.org/10.1016/j.chembiol.2013.12.011.

Komarék, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. Eur. J. Phycol. 51(3), 346-353. https://doi.org/10.1080/09670262.2016.1163738.

Komarék, J., Kaštovsky, J., Mareš, J., Johnansen, J. R., 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86, 295-335.

Komárek, J., Kling, H., Komárková, J., 2003. 4 - Filamentous Cyanobacteria. In: Wehr, J. D., Sheath, R. G. (Eds.) Freshwater Algae of North America. Academic Press, Burlington, pp. 117-196. https://doi.org/10.1016/B978-0-12-385876-4.00004-9.

Konetschny-Rapp, S., Jung, G., Meiwes, J., Zähner, H., 1990. Staphyloferrin A: a structurally new siderophore from *Staphylococci*. Eur. J. Biochem. 191, 65-74. https://doi.org/10.1111/j.1432-1033.1990.tb19094.x.

Kononen, K., Kuparinen, J., Mäkelä, J., Laanemets, J., Pavelson, J., Nõmmann, S., 1996. Initiation of cyanobacterial blooms in a frontal regions at the entrance to the Gulf of Finland, Baltic Sea. Limnol. Oceanogr. 41, 98-112. https://doi.org/10.4319/lo.1996.41.1.0098.

Koonin, E. V., Makarova, K. S., Aravind, L., 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. 55, 709-742. https://doi.org/10.1146/annurev.micro.55.1.709.

Koonin, E. V., Mushegian, A. R., Galperin, M. Y., Walker, D. R., 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. 25, 619-637. https://doi.org/10.1046/j.1365-2958.1997.4821861.x.

Kosten, S., Huszar, V. L. M., Bécares, E., Costa, L. S., van Donk, E., Hansson, L.-A., Jeppesen, E., Kruk, C., Lacerot, G., Mazzeo, N., De Meester, L., Moss, B., Lürling, M., Nõges, T., Romo, S., Scheffer, M., 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Glob. Change Biol. 18, 118-126. https://doi.org/10.1111/j.1365-2486.2011.02488.x.

Kristjansson, J. K., Hreggvidsson, G. O., 1995. Ecology and habitats of extremophiles. World J. Microbiol. Biotechnol. 11, 17-25. https://doi.org/10.1007/BF00339134.

Krüger, T., Mönch, B., Oppenhäuser, S., Luckas, B., 2010. LC–MS/MS determination of the isomeric neurotoxins BMAA (β-*N*-methylamino-L-alanine) and DAB (2,4-diaminobutyric acid) in cyanobacteria and seeds of *Cycas revoluta* and *Lathyrus latifolius*. Toxicon 55(2-3), 547-557. https://doi.org/10.1016/j.toxicon.2009.10.009.

Kuhner, M. K., Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11, 459-468. https://doi.org/10.1093/oxfordjournals.molbev.a040126.

Kumar, K., Mella-Herrera, R. A., Golden, J. W., 2010. Cyanobacterial heterocysts. Cold Spring Harbor Perspect. Biol. 2, a000315. https://doi.org/10.1101/cshperspect.a000315.

Kunin, V., Goldovsky, L., Darzentas, N., Ouzounis, C. A., 2005. The net of life: reconstructing the microbial phylogenetic network. Genome Res. 15, 954-959. https://doi.org/10.1101/gr.3666505.

Künzler, K., Eichenberger, W., 1997. Betaine lipids and zwitterionic phospholipids in plants and fungi. Phytochemistry 46(5), 883-892. https://doi.org/10.1016/S0031-9422(97)81274-5.

Kuo, Y.-H., Ikegami, F., Lambein, F., 1998. Metabolic routes of β-(isoxazolin-5-on-2-yl)-L-alanine (BIA), the precursor of the neurotoxin ODAP (β-*N*-oxalyl-L-α, β,-diaminopropionic acid), in different legume seedlings. Phytochemistry 49, 43-48. https://doi.org/10.1016/S0031-9422(97)01001-7.

Kuo, Y.-H., Lambein, F., Ikegami, F., Van Parijs, R., 1982. Isoxazolin-5-ones and amino acids in root exudates of pea and sweet pea seedlings. Plant Physiol. 70(5), 1283-1289. https://doi.org/10.1104/pp.70.5.1283.

Kurland, L. K., Mulder, D. W., 1954. Epidemiologic investigations of amyotrophic lateral sclerosis: 1. Preliminary report on geographic distribution, with special reference to the Mariana Islands, including clinical and pathologic observations. Neurology 4, 355-355. https://doi.org/10.1212/WNL.4.5.355.

Kurmayer, R., Sivonen, K., Wilmotte, A., Salmaso, N., Wiley, J., 2017. Molecular Tools for the Detection and Quantification of Toxigenic Cyanobacteria. Wiley Online Library, New York, pp. 440. https://doi.org/10.1002/9781119332169.

Kurz, M., Burch, A. Y., Seip, B., Lindow, S. E., Gross, H., 2010. Genome-driven investigation of compatible solute biosynthesis pathways of *Pseudomonas syringae* pv. syringae and their contribution to water stress tolerance. Appl. Environ. Microbiol. 76(16), 5452-5462. https://doi.org/10.1128/AEM.00686-10.

Labine, M. A., Minuk, G. Y., 2009. Cyanobacterial toxins and liver disease. Can. J. Physiol. Pharmacol. 87, 773-788. https://doi.org/10.1139/Y09-081.

Lacap, D. C., Warren-Rhodes, K. A., McKay, C. P., Pointing, S. B., 2011. Cyanobacteria and chloroflexi-dominated hypolithic colonisation of quartz the hyper-arid core of the Atacama Desert, Chile. Extremophiles 15, 31-38. https://doi.org/10.1007/s00792-010-0334-3.

Lage, S., Burian, A., Rasmussen, U., Costa, P. R., Annadotter, H., Godhe, A., Rydberg, S., 2016. BMAA extraction of cyanobacteria samples: which method to choose? Environ. Sci. Pollut. Res. 23(1), 338-350. https://doi.org/10.1007/s11356-015-5266-0.

Lambein, F., 1981. Lathyrism in young chicks, induced by isoxazolin-5-ones from *Lathyrus odoratus* seedlings [sweet-pea]. Arch. Int. Physiol. Biochim..

Lambein, F., De Bruyn, A., Ikegami, F., Kuo, Y. H., 1986. Distribution and biosynthesis of isoxazolinone derivatives in the genus *Lathyrus*. In: Kaul, A. K., Comb, D. (Eds.) *Lathyrus* and Lathyrism. Third World Medical Research Foundation, New York, pp. 246-256.

Lambrecht, J. A., Flynn, J. M., Downs, D. M., 2012. Conserved YjgF protein family deaminates reactive enamine/imine intermediates of pyridoxal 5′-phosphate (PLP)-dependent enzyme reactions. J. Biol. Chem. 287, 3454-3461. https://doi.org/10.1074/jbc.M111.304477.

Lambrecht, J. A., Schmitz, G. E., Downs, D. M., 2013. RidA proteins prevent metabolic damage inflicted by PLP-dependent dehydratases in all domains of life. MBio 4, e00033-00013. https://doi.org/10.1128/mBio.00033-13.

Lanfear, R., Calcott, B., Ho, S. Y. W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695-1701. https://doi.org/10.1093/molbev/mss020.

Latysheva, N., Junker, V. L., Palmer, W. J., Codd, G. A., Barker, D., 2012. The evolution of nitrogen fixation in cyanobacteria. Bioinformatics 28, 603-606. https://doi.org/10.1093/bioinformatics/bts008.

Lawrence, J. G., 2002. Gene transfer in bacteria: speciation without species? Theor. Popul. Biol. 61, 449-460. https://doi.org/10.1006/tpbi.2002.1587.

Lawrence, J. G., Ochman, H., 2002. Reconciling the many faces of lateral gene transfer. Trends Microbiol. 10, 1-4. https://doi.org/10.1016/S0966-842X(01)02282-X.

Lederberg, J., Tatum, E. L., 1946. Gene Recombination in *Escherichia coli*. Nature 158, 558-558. https://doi.org/10.1038/158558a0.

Lee, J., Sperandio, V., Frantz, D. E., Longgood, J., Camilli, A., Phillips, M. A., Michael, A. J., 2009. An alternative polyamine biosynthetic pathway is widespread in bacteria and essential for biofilm formation in *Vibrio cholerae*. J. Biol. Chem. 284(15), 9899-9907. https://doi.org/10.1074/jbc.M900110200.

Leikoski, N., Fewer, D. P., Sivonen, K., 2009. Widespread occurrence and lateral transfer of the cyanobactin biosynthetic gene cluster in cyanobacteria. Appl. Env. Microbiol. 75, 853-857. https://doi.org/10.1128/AEM.02134-08.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242-W245. https://doi.org/10.1093/nar/gkw290.

Levy, M. S., Balbinder, E., Nagel, R., 1993. Effect of mutations in SOS genes on UV-induced precise excision of Tn10 in *Escherichia coli*. Mutat. Res. DNA Repair 293, 241-247. https://doi.org/10.1016/0921-8777(93)90075-R.

Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J.-K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. 34, D572-D580. https://doi.org/10.1093/nar/gkj118.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, L., Dechesne, A., He, Z., Madsen, J. S., Nesme, J., Sørensen, S. J., Smets, B. F., 2018. Estimating the transfer range of plasmids encoding antimicrobial resistance in a wastewater treatment plant microbial community. Environ. Sci. Technol. Lett. 5, 260-265. https://doi.org/10.1021/acs.estlett.8b00105.

Li, L., Stoeckert, C. J., Roos, D. S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178-2189. https://doi.org/10.1101/gr.1224503.

Li, M. H. T., Ung, P. M. U., Zajkowski, J., Garneau-Tsodikova, S., Sherman, D. H., 2009. Automated genome mining for natural products. BMC Bioinformatics 10, 1-10. https://doi.org/10.1186/1471-2105-10-185.

Li, R., Oliver, R. A., Townsend, C. A., 2017. Identification and characterization of the sulfazecin monobactam biosynthetic gene cluster. Cell Chem. Biol. 24, 24-34. https://doi.org/10.1016/j.chembiol.2016.11.010.

Li, S.-J., Hua, Z.-S., Huang, L.-N., Li, J., Shi, S.-H., Chen, L.-X., Kuang, J.-L., Liu, J., Hu, M., Shu, W.-S., 2014. Microbial communities evolve faster in extreme environments. Sci. Rep. 4, 6205. https://doi.org/10.1038/srep06205.

Li, T., Liu, D., Yang, Y., Guo, J., Feng, Y., Zhang, X., Cheng, S., Feng, J., 2020. Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. Sci. Rep. 10, 1-9. https://doi.org/10.1038/s41598-020-79484-8.

Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M. K., Durkin, A. S., 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res. 49, D1020-D1028. https://doi.org/10.1093/nar/gkaa1105.

Ligrone, R., 2019. The great oxygenation event. In: Biological Innovations that Built the World. Springer, Cham, pp. 129-154. https://doi.org/10.1007/978-3-030-16057-9_5.

Lilley, A. K., Bailey, M. J., 1997. The acquisition of indigenous plasmids by a genetically marked pseudomonad population colonizing the sugar beet phytosphere is related to local environmental conditions. Appl. Environ. Microbiol. 63, 1577-1583. https://doi.org/10.1128/aem.63.4.1577-1583.1997.

Litvaitis, M. K., 2002. A molecular test of cyanobacterial phylogeny: inferences from constraint analyses. Hydrobiologia 468, 135-145. https://doi.org/10.1023/A:1015262621848.

Liu, X., Zeng, J., Chen, X., Xie, W., 2016. Crystal structures of RidA, an important enzyme for the prevention of toxic side products. Sci. Rep. 6, 1-9. https://doi.org/10.1038/srep30494.

Llewellyn, L. E., 2006. Saxitoxin, a toxic marine natural product that targets a multitude of receptors. Nat. Prod. Rep. 23, 200-222. https://doi.org/10.1039/B501296C.

Loeb, S. L., Reuter, J. E., 1981. The epilithic periphytion community: a five-lake comparative study of community productiviy, nitrogen metabolism and depth-distribution of standing crop. Verh. Int. Ver. Limnol. 21, 346-352. https://doi.org/10.1080/03680770.1980.11897005.

Loftie-Eaton, W., Tucker, A., Norton, A., Top Eva, M., Liu, S. J., 2014. Flow cytometry and real-time quantitative PCR as tools for assessing plasmid persistence. Appl. Environ. Microbiol. 80, 5439-5446. https://doi.org/10.1128/AEM.00793-14.

Lohkamp, B., Voevodskaya, N., Lindqvist, Y., Dobritzsch, D., 2010. Insights into the mechanism of dihydropyrimidine dehydrogenase from site-directed mutagenesis targeting the active site loop and redox cofactor coordination. Biochim. Biophys. Acta Proteins Proteom. 1804, 2198-2206. https://doi.org/10.1016/j.bbapap.2010.08.014.

Lopatkin, A. J., Huang, S., Smith, R. P., Srimani, J. K., Sysoeva, T. A., Bewick, S., Karig, D. K., You, L., 2016. Antibiotics as a selective driver for conjugation dynamics. Nat. Microbiol. 1, 16044. https://doi.org/10.1038/nmicrobiol.2016.44.

Louis, P., Galinski, E. A., 1997. Characterization of genes for the biosynthesis of the compatible solute ectoine from *Marinococcus halophilus* and osmoregulated expression in *Escherichia coli*. Microbiology 143, 1141-1149. https://doi.org/10.1099/00221287-143-4-1141.

Low-Décarie, E., Fussmann, G. F., Dumbrell, A. J., Bell, G., 2016. Communities that thrive in extreme conditions captured from a freshwater lake. Biol. Lett. 12, 20160562. https://doi.org/10.1098/rsbl.2016.0562.

Lyck, S., Gjølme, N., Utkilen, H., 1996. Iron starvation increases toxicity of *Microcystis aeruginosa* CYA 228/1 (Chroococcales, Cyanophyceae). Phycologia 35, 120-124. https://doi.org/10.2216/i0031-8884-35-6S-120.1.

Lynch, D., O'Brien, J., Welch, T., Clarke, P., ÓCuív, P., Crosa, J. H., O'Connell, M., 2001. Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. J. Bacteriol. 183(8), 2576-2585. https://doi.org/10.1128/JB.183.8.2576-2585.2001.

Lyons, T. W., Reinhard, C. T., Planavsky, N. J., 2014. The rise of oxygen in Earth's early ocean and atmosphere. Nature 506, 307-315. https://doi.org/10.1038/nature13068.

Maglott, D. R., Katz, K. S., Sicotte, H., Pruitt, K. D., 2000. NCBI's LocusLink and RefSeq. Nucleic Acids Res. 28, 126-128. https://doi.org/10.1093/nar/28.1.126.

Makarova, K. S., Wolf, Y. I., Snir, S., Koonin, E. V., 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J. Bacteriol. 193, 6039-6056. https://doi.org/10.1128/JB.05535-11.

Malathi, K., Padmanaban, G., Sarma, P. S., 1970. Biosynthesis of β-*N*-oxalyl-L-α, β-diaminopropionic acid, the *Lathyrus sativus* neurotoxin. Phytochemistry 9, 1603-1610. https://doi.org/10.1016/S0031-9422(00)85283-8.

Malhotra-Kumar, S., Xavier, B. B., Das, A. J., Lammens, C., Butaye, P., Goossens, H., 2016. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. The Lancet Infect. Dis. 16, 283-284. https://doi.org/10.1016/S1473-3099(16)00012-8.

Mantas, M. J. Q., Nunn, P. B., Codd, G. A., Barker, D., 2022. Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA). Phytochemistry 200, 113198. https://doi.org/10.1016/j.phytochem.2022.113198.

Mantas, M. J. Q., Nunn, P. B., Ke, Z., Codd, G. A., Barker, D., 2021. Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 2,4-diaminobutanoic acid (2,4-DAB). Phytochemistry 192, 112953. https://doi.org/10.1016/j.phytochem.2021.112953.

Mapstone, L. J., Leite, M. N., Purton, S., Crawford, I .A., Dartnell, L., 2022. Cyanobacteria and microalgae in supporting human habitation on Mars. Biotechnol. Adv. 59, 107946. https://doi.org/10.1016/j.biotechadv.2022.107946.

Marks, C. R., Stevenson, B. S., Rudd, S., Lawson, P. A., 2012. *Nitrospira*-dominated biofilm within a thermal artesian spring: a case for nitrification-driven primary production in a geothermal setting. Geobiology 10, 457-466. https://doi.org/10.1111/j.1472-4669.2012.00335.x.

Martin, J. L., McMillan, F. M., 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struct. Biol. 12(6), 783-793. https://doi.org/10.1016/S0959-440X(02)00391-3.

Martins, J., Peixe, L., Vasconcelos, V. M., 2011. Unraveling cyanobacteria ecology in wastewater treatment plants (WWTP). Microb. Ecol. 62, 241-256. https://doi.org/10.1007/s00248-011-9806-y.

Mathers, A. J., Peirano, G., Pitout, J. D. D., 2015. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clinical microbiology reviews 28, 565-591. https://doi.org/10.1128/CMR.00116-14.

Matz, C. J., Christensen, M. R., Bone, A. D., Gress, C. D., Widenmaier, S. B., Weger, H. G., 2004. Only iron-limited cells of the cyanobacterium *Anabaena flos-aquae* inhibit growth of the green alga *Chlamydomonas reinhardtii*. Can. J. Bot. 82, 436-442. https://doi.org/10.1139/b04-022.

Mayer, P. M., Galatowitsch, S. M., 2001. Assessing integrity of restored prairie wetlands from species production-diversity relationships. Hydrobiologia 443, 177-185. https://doi.org/10.1023/A:1017516724965.

Mbedi, S., Welker, M., Fastner, J., Wiedner, C., 2005. Variability of the microcystin synthetase gene cluster in the genus *Planktothrix* (Oscillatoriales, Cyanobacteria). FEMS Microbiol. Lett. 245(2), 299-306. https://doi.org/10.1016/j.femsle.2005.03.020.

McCormick, P. V., O'Deall, M. B., 1996. Quantifying peripython responses to phosphorus in the Florida Everglades: a synoptic experimental approach. J. N. Am. Benthol. Soc. 15, 450-468. https://doi.org/10.2307/1467798.

McGinnis, M. W., Parker, Z. M., Walter, N. E., Rutkovsky, A. C., Cartaya-Marin, C., Karatan, E., 2009. Spermidine regulates *Vibrio cholerae* biofilm formation via transport and signaling pathways. FEMS Microbiol. Lett. 299(2), 166-174. https://doi.org/10.1111/j.1574-6968.2009.01744.x.

McKnight, D. M., Morel, F. M. M., 1980. Copper complexation by siderophores from filamentous blue-green algae. Limnol. Oceanogr. 25, 62-71. https://doi.org/10.4319/lo.1980.25.1.0062.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339-W346. https://doi.org/10.1093/nar/gkr466.

Medema, M. H., Paalvast, Y., Nguyen, D. D., Melnik, A., Dorrestein, P. C., Takano, E., Breitling, R., 2014. Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. PLoS Comput. Biol. 10, e1003822. https://doi.org/10.1371/journal.pcbi.1003822.

Meriluoto, J., Spoof, L., Codd, G. A., 2017. Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis. John Wiley & Sons, Chichester, UK, pp. 548.

Mestichelli, L. J., Gupta, R. N., Spenser, I. D., 1979. The biosynthetic route from ornithine to proline. J. Biol. Chem. 254, 640-647. https://doi.org/10.1016/S0021-9258(17)37853-5.

Metcalf, J. S., Banack, S. A., Powell, J. T., Tymm, F. J. M., Murch, S. J., Brand, L. E., Cox, P. A., 2018. Public health responses to toxic cyanobacterial blooms: perspectives from the 2016 Florida event. Water Policy 20, 919-932. https://doi.org/10.2166/wp.2018.012.

Metcalf, J. S., Codd, G. A., 2012. Cyanotoxins. In: Whitton, B. A. (Ed.) Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands, Dordrecht, pp. 651-675.

Metcalf, J. S., Glover, W. B., Banack, S. A., Cox, P. A., 2017. Analysis of β-N-Methylamino-l-Alanine by UHPLC-MS/MS, in: Meriluoto, J., Spoof, L., Codd, G.A. (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis. John Wiley & Sons, Chichester, UK, pp. 435-438. https://doi.org/10.1002/9781119068761.ch54.

Meyer, B., Immer, C., Kaiser, S., Sharma, S., Yang, J., Watzinger, P., Weiß, L., Kotter, A., Helm, M., Seitz, H.-M., 2020. Identification of the 3-amino-3-carboxypropyl (acp) transferase enzyme responsible for acp3U formation at position 47 in *Escherichia coli* tRNAs. Nucleic Acids Res. 48(3), 1435-1450. https://doi.org/10.1093/nar/gkz1191.

Mez, K., Beattie, K. A., Codd, G. A., Hanselmann, K., Hauser, B., Naegeli, H., Preisig, H. R. 1997. Identification of a microcystin in benthic cyanobacteria linked to cattle deaths on alpine pastures in Switzerland. Eur. J. Pychol. 32, 111-117. https://doi.org/10.1080/09670269710001737029.

Mihali, T. K., Kellmann, R., Muenchhoff, J., Barrow, K. D., Neilan, B. A., 2008. Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. Appl. Environ. Microbiol.74(3), 716-722. https://doi.org/10.1128/AEM.01988-07.

Miller, B. R., Gulick, A. M., 2016. Structural biology of nonribosomal peptide synthetases, in: Evans, B. (Eds.), Nonribosomal Peptide and Polyketide Biosynthesis. Methods in Molecular Biology. Springer, 1401, Clifton, N.J., pp. 3-29. https://doi.org/10.1007/978-1-4939-3375-4_1.

Miller, S. R., Purugganan, M. D., Curtis, S. E., 2006. Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium *Mastigocladus laminosus*. Appl. Environ. Microbiol. 72, 2793-2800. https://doi.org/10.1128/AEM.72.4.2793-2800.2006.

Misner, I., Blouin, N., Leonard, G., Richards, T. A., Lane, C. E., 2015. The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. Genome Biol. Evol. 7, 120-135. https://doi.org/10.1093/gbe/evu276.

Mistiniene, E., Luksa, V., Sereikaite, J., Naktinis, V., 2003. Oligomeric assembly and ligand binding of the members of protein family YER057c/YIL051c/YJGF. Bioconjug. Chem. 14, 1243-1252. https://doi.org/10.1021/bc0341066.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Res. 49, D412-D419. https://doi.org/10.1093/nar/gkaa913.

Moffitt, M. C., Neilan, B. A., 2004. Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. Appl. Environ. Microbiol. 70(11), 6353-6362. https://doi.org/10.1128/AEM.70.11.6353-6362.2004.

Mølbak, L., Kroer, N., 2007. Root growth and exudate production define the frequency of horizontal plasmid transfer in the rhizosphere. FEMS Microbiol. Ecol. 59, 167-176. https://doi.org/10.1111/j.1574-6941.2006.00229.x.

Molin, S., Tolker-Nielsen, T., 2003. Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr. Opin. Biotechnol. 14, 255-261. https://doi.org/10.1016/S0958-1669(03)00036-3.

Monteil, C. L., Grouzdev, D. S., Perriere, G., Alonso, B., Rouy, Z., Cruveiller, S., Ginet, N., Pignol, D., Lefevre, C. T., 2020. Repeated horizontal gene transfers triggered parallel evolution of magnetotaxis in two evolutionary divergent lineages of magnetotactic bacteria. ISME J. 14, 1783-1794. https://doi.org/10.1038/s41396-020-0647-x.

Moore, K. R., Magnabosco, C., Momper, L., Gold, D. A., Bosak, T., Fournier, G. P., 2019. An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. Front. Microbiol., 1612. https://doi.org/10.3389/fmicb.2019.01612.

Morrison, S. S., Mullineaux, C. W., Ashby, M. K., 2005. The influence of acetyl phosphate on DspA signalling in the cyanobacterium Synechocystis sp. PCC6803. BMC Microbiol. 5, 47. https://doi.org/10.1186/1471-2180-5-47.

Müller, S., Garcia-Gonzalez, E., Mainz, A., Hertlein, G., Heid, N. C., Mösker, E., van den Elst, H., Overkleeft, H. S., Genersch, E., Süssmuth, R. D., 2014. Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. Angew. Chem. Int. Ed. 53, 10821-10825. https://doi.org/10.1002/anie.201404572.

Mungan, M. D., Alanjary, M., Blin, K., Weber, T., Medema, M. H., Ziemert, N., 2020. ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Res. 48, W546-W552. https://doi.org/10.1093/nar/gkaa374.

Murch, S. J., Cox, P. A., Banack, S. A., 2004. A mechanism for slow release of biomagnified cyanobacterial neurotoxins and neurodegenerative disease in Guam. Proc. Natl. Acad. Sci. U.S.A. 101, 12228-12231. https://doi.org/10.1073/pnas.0404926101.

Muriana, F. J. G., Alvarez-Ossorio, M. C., Relimpio, A. M., 1991. Purification and characterization of aspartate aminotransferase from the halophile archaebacterium *Haloferax mediterranei*. Biochem. J. 278, 149-154. https://doi.org/10.1042/bj2780149.

Muro-Pastor, A. M., Kuritz, T., Flores, E., Herrero, A., Wolk, C. P., 1994. Transfer of a genetic marker from a megaplasmid of Anabaena sp. strain PCC 7120 to a megaplasmid of a different Anabaena strain. J. Bacteriol. 176, 1093-1098. https://doi.org/10.1128/jb.176.4.1093-1098.1994.

Muth, W. L., Costilow, R. N., 1974. Ornithine cyclase (deaminating): II. Properties of the homogeneous enzyme. J. Biol. Chem. 249, 7457-7462. https://doi.org/10.1016/S0021-9258(19)81260-7.

Nagasawa, T., Tanizawa, K., Satoda, T., Yamada, H., 1988. Diaminopropionate ammonia-lyase from *Salmonella typhimurium*. Purification and characterization of the crystalline enzyme, and sequence determination of the pyridoxal 5'-phosphate binding peptide. J. Biol. Chem. 263, 958-964. https://doi.org/10.1016/S0021-9258(19)35446-8.

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M., 2000. Crystal structure of *N*-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. Structure 8, 729-738. https://doi.org/10.1016/S0969-2126(00)00160-X.

Nakamura, Y., Itoh, T., Matsuda, H., Gojobori, T., 2004. Biased biological functions of horriontally transferred genes in prokaryotic genomes. Nat. Genet. 36, 760-766. https://doi.org/10.1038/ng1381.

Nakasugi, K., Svenson, C. J., Neilan, B. A., 2006. The competence gene, comF, from *Synechocystis* sp. strain PCC 6803 is involved in natural transformation, phototactic motility and piliation. Microbiology 152, 3623-3631. https://doi.org/10.1099/mic.0.29189-0.

Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., De Los Santos, E. L. C., Yeong, M., Cruz-Morales, P., Abubucker, S., 2020. A computational framework to explore large-scale biosynthetic diversity. Nat. Chem. Biol. 16, 60-68. https://doi.org/10.1038/s41589-019-0400-9.

Neilan, B. A., Stuart, J. L., Goodman, A. E., Cox, P. T., Hawkins, P. R., 1997. Specific amplification and restriction polymorphisms of the cyanobacterial rRNA operon spacer region. Syst. Appl. Microbiol. 20, 612-621. https://doi.org/10.1016/S0723-2020(97)80033-1.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., Minh, B. Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32(1), 268-274. https://doi.org/10.1093/molbev/msu300.

Nicolaisen, K., Moslavac, S., Samborski, A., Valdebenito, M., Hantke, K., Maldener, I., Muro-Pastor, A. M., Flores, E., Schleiff, E., 2008. Alr0397 is an outer membrane transporter for the siderophore schizokinen in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 190(22), 7500-7507. https://doi.org/10.1128/JB.01062-08.

Niehaus, T. D., Gerdes, S., Hodge-Hanson, K., Zhukov, A., Cooper, A. J. L., ElBadawi-Sidhu, M., Fiehn, O., Downs, D. M., Hanson, A. D., 2015. Genomic and experimental evidence for multiple metabolic functions in the RidA/YjgF/YER057c/UK114 (Rid) protein family. BMC Genom. 16, 1-14. https://doi.org/10.1186/s12864-015-1584-3.

Nigam, S. N., Ressler, C., 1966. Biosynthesis of 2,4-diaminobutyric acid from L-[3H] homoserine and DL-[1-14C] aspartic acid in *Lathyrus sylvestris* W. Biochemistry 5(11), 3426-3431. https://doi.org/10.1021/bi00875a006.

Nishimura, S., Taya, Y., Kuchino, Y., Ohashi, Z., 1974. Enzymatic synthesis of 3-(3-amino-3-carboxypropyl) uridine in *Escherichia coli* phenylalanine transfer RNA: transfer of the 3-amino-3-carboxypropyl group from S-adenosylmethionine. Biochem. Biophys. Res. Commun. 57, 702-708. https://doi.org/10.1016/0006-291X(74)90603-2.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., 2015. Promoting an open research culture. Science 348, 1422-1425. https://doi.org/10.1126/science.aab2374.

Novakovskaya, I. V., Patova, E. N., Dubrovskiy, Y. A., Novakovskiy, A. B., Kulyugina, E. E., 2022. Distribition of algae and cyanobaceria of biological soil crusts along the elevation gradient in mountain plant communities at the Northern Urals (Russian European Northeast). J. Mt. Sci. 19, 637-646. https://doi.org/10.1007/s11629-021-6952-7.

Novo, C., Tata, R., Clemente, A., Brown, P. R., 1995. *Pseudomonas aeruginosa* aliphatic amidase is related to the nitrilase/cyanide hydratase enzyme family and Cys166 is predicted to be the active site nucleophile of the catalytic mechanism. FEBS Lett. 367, 275-279. https://doi.org/10.1016/0014-5793(95)00585-W.

Nowicka-Krawczyk, P., Mühlsteinová, R., Hauer, T., 2019. Detailed characterization of the *Arthrospira* type species separating commercially grown taxa into the new genus *Limnospira* (Cyanobacteria). Sci. Rep. 9, 694. https://doi.org/10.1038/s41598-018-36831-0.

Nuin, P. A. S., Wang, Z., Tillier, E. R. M., 2006. The accuracy of several multiple sequence alignment programs for proteins. BMC Bioinformatics 7, 1-18. https://doi.org/10.1186/1471-2105-7-471.

Nunn, P. B., 2017. 50 years of research on α-amino-β-methylaminopropionic acid (β-methylaminoalanine). Phytochemistry 144, 271-281. https://doi.org/10.1016/j.phytochem.2017.10.002.

Nunn, P. B., Codd, G. A., 2017. Metabolic solutions to the biosynthesis of some diaminomonocarboxylic acids in nature: formation in cyanobacteria of the neurotoxins 3-*N*-methyl-2,3-diaminopropanoic acid (BMAA) and 2,4-diaminobutanoic acid (2,4-DAB). Phytochemistry 144, 253-270. https://doi.org/10.1016/j.phytochem.2017.09.015.

Nunn, P. B., Codd, G. A., 2019. Environmental distribution of the neurotoxin L-BMAA in *Paenibacillus* species. Toxicol. Res. 8, 781-783. https://doi.org/10.1039/C9TX00203K.

Ochman, H., Lawrence, J. G., Groisman, E. A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299-304. https://doi.org/10.1038/35012500.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733-D745. https://doi.org/10.1093/nar/gkv1189.

Oka, T., Tsuji, H., Noda, C., Sakai, K., Hong, Y.-m., Suzuki, I., Muñoz, S., Natori, Y., 1995. Isolation and characterization of a novel perchloric acid-soluble protein inhibiting cell-free protein synthesis. J. Biol. Chem. 270, 30060-30067. https://doi.org/10.1074/jbc.270.50.30060.

Okeyoshi, K., Okajima, M. K., Kaneko, T., 2021. The cyanobacterial polysaccharide sacran: characteristics, structures, and preparation of LC gels. Polym. J. 53, 81-91. https://doi.org/10.1038/s41428-020-00426-2.

Oliver, A., Podell, S., Pinowska, A., Traller, J. C., Smith, S. R., McClure, R., Beliaev, A., Bohutskyi, P., Hill, E. A., Rabines, A., 2021. Diploid genomic architecture of *Nitzschia inconspicua*, an elite biomass production diatom. Sci. Rep. 11, 1-14. https://doi.org/10.1038/s41598-021-95106-3.

Oren, A., 2006. Halophilic microorganisms and their environments. Springer, Dordrecht, pp. 575.

Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M., Edwards, S. V., 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature 446, 180-184. https://doi.org/10.1038/nature05621.

Ovando, C. A., Carvalho, J. C. d., Vinícius de Melo Pereira, G., Jacques, P., Soccol, V. T., Soccol, C. R., 2018. Functional properties and health benefits of bioactive peptides derived from *Spirulina*: a review. Food Rev. Int. 34, 34-51. https://doi.org/10.1080/87559129.2016.1210632.

Paerl, H. W., 2012. Marine plankton. In: Whitton, B. (eds) The Ecology of Cyanobacteria II: their diversity in time and space. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-3855-3_5.

Paerl, H. W., Fulton, R. S., 2006. Ecology of harmful cyanobacteria. In: Graneli, E., Turner, J. (eds) Ecology of harmful marine algae. Springer, Berlin, pp. 95-107.

Paerl, H. W., Otten, T. G., 2013. Blooms bite the hand that feeds them. Science 342, 433-434. https://doi.org/10.1126/science.1245276.

Paerl, H. W., Pinckney, J. L., Steppe, T. F., 2000. Cyanobacterial–bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. Environ. Microbiol. 2, 11-26. https://doi.org/10.1046/j.1462-2920.2000.00071.x.

Pagel, M., 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B: Biol. Sci. 255, 37-45. https://doi.org/10.1098/rspb.1994.0006.

Pagel, M., 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26, 331-348. https://doi.org/10.1111/j.1463-6409.1997.tb00423.x.

Pagel, M., 1999a. Inferring the historical patterns of biological evolution. Nature 401, 877-884. https://doi.org/10.1038/44766.

Pagel, M., 1999b. The maximum likelihood approach to reconstructing ancetral character states of discrete characters on phylogenies. Syst. Biol. 48, 612-622. https://doi.org/10.1080/106351599260184.

Pagel, M., Meade, A., Barker, D., 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53, 673-684. https://doi.org/10.10635150490522232.

Pan, H., Song, L., Liu, Y., Börner, T., 2002. Detection of hepatotoxic *Microcystis* strains by PCR with intact cells from both culture and environmental samples. Arch. Microbiol. 178, 421-427. https://doi.org/10.1007/s00203-002-0464-9.

Papendorf, O., König, G. M., Wright, A. D., 1998. Hierridin B and 2,4-dimethoxy-6-heptadecyl-phenol, secondary metabolites from the cyanobacterium *Phormidium ectocarpi* with antiplasmodial activity. Phytochemistry 49, 2383-2386. https://doi.org/10.1016/S0031-9422(98)00440-3.

Parmeggiani, F., Weise, N. J., Ahmed, S. T., Turner, N. J., 2018. Synthetic and therapeutic applications of ammonia-lyases and aminomutases. Chem. Rev. 118, 73-118. https://doi.org/10.1021/acs.chemrev.6b00824.

Parsons, L., Bonander, N., Eisenstein, E., Gilson, M., Kairys, V., Orban, J., 2003. Solution structure and functional ligand screening of HI0719, a highly conserved protein from bacteria to humans in the YjgF/YER057c/UK114 family. Biochemistry 42, 80-89. https://doi.org/10.1021/bi020541w.

Parter, M., Kashtan, N., Alon, U., 2007. Environmental variability and modularity of bacterial metabolic networks. BMC Evol. Biol. 7, 1-8. https://doi.org/10.1186/1471-2148-7-169.

Patel, A., Matsakas, L., Rova, U., Christakopoulos, P., 2019. A perspective on biotechnological applications of thermophilic microalgae and cyanobacteria. Bioresour. Technol. 278, 424-434. https://doi.org/10.1016/j.biortech.2019.01.063.

Peng, W.-F., Huang, C.-Y., 2014. Allantoinase and dihydroorotase binding and inhibition by flavonols and the substrates of cyclic amidohydrolases. Biochimie 101, 113-122. https://doi.org/10.1016/j.biochi.2014.01.001.

Percival, S. L., Williams, D. W., 2014. Cyanobacteria. In: Percival, S. L., Yates, M. V., Williams, D. W., Chalmers, R. M., Gray, N. F. (Eds.) Microbiology of waterborne diseases. Academic Press, Boston, Massachusetts, pp. 79-88. https://doi.org/10.1016/C2010-0-67101-X.

Perkins, H. R., Cummins, C. S., 1964. Chemical structure of bacterial cell walls: ornithine and 2,4-diaminobutyric acid as components of the cell walls of plant pathogenic corynebacteria. Nature 201, 1105-1107. https://doi.org/10.1038/2011105a0.

Persmark, M., Pittman, P., Buyer, J. S., Schwyn, B., Gill Jr, P. R., Neilands, J. B., 1993. Isolation and structure of rhizobactin 1021, a siderophore from the alfalfa symbiont *Rhizobium meliloti* 1021. J. Am. Chem. Soc. 115(10), 3950-3956. https://doi.org/10.1021/ja00063a014.

Peters, P., Galinski, E. A., Trüper, H. G., 1990. The biosynthesis of ectoine. FEMS Microbiol. Lett. 71, 157-162. https://doi.org/10.1016/0378-1097(90)90049-V.

Pfleger, B. F., Lee, J. Y., Somu, R. V., Aldrich, C. C., Hanna, P. C., Sherman, D. H., 2007. Characterization and analysis of early enzymes for petrobactin biosynthesis in *Bacillus anthracis*. Biochemistry 46, 4147-4157. https://doi.org/10.1021/bi6023995.

Polsky, F. I., Nunn, P. B., Bell, E. A. 1972. Distribution and toxicity of alpha-amino-beta-methylaminopropionic acid. Fed. Proc. 31, 1473-1475.

Postgate, J., 1998. Nitrogen fixation. Cambridge University Press, Cambridge, UK.

Potts, M., 1980. Blue-green algae (Cyanophyta) in marine coastal environments of the Sinai Peninsula; distribution, zonation, stratification and taxonomic diversity. Phycologia 19, 60-73. https://doi.org/10.2216/i0031-8884-19-1-60.1.

Pouria, S., Andrade, A., Barbosa, J., Cavalcanti, R. L., Barreto, V. T., Ward, C. J., Preiser, W., Poon, G. K., Neild, G. H., Codd, G. A. 1998. Fatal microcystun intoxication in hemodialysis unit in Caruaru, Brazil. Lancet 351, 21-26. https://doi.org/10.1016/S0140-6736(97)12285-1.

Pruitt, K. D., Tatusova, T., Maglott, D. R., 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 33, D501-D504. https://doi.org/10.1093/nar/gkl842.

Quang, L. S., Gascuel, O., Lartillot, N., 2008. Empirical Profile Mixture Model for Phylogenetic Reconstruction. Bioinformatics 24(20), 2317-2323. https://doi.org/10.1093/bioinformatics/btn445.

Rambaut, A., Grassly, N. C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13(3), 235-238. https://doi.org/10.1093/bioinformatics/13.3.235.

Ramette, A., Tiedje, J. M., 2007. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. Microb. Ecol. 53, 197-207. https://doi.org/10.1007/s00248-005-5010-2.

Rampelotto, P. H., 2013. Extremophiles and extreme environments. Life 3, 482-485. https://doi.org/10.3390/life3030482.

Rao, D. R., Hariharan, K., Vijayalakshmi, K. R., 1969. A study of the metabolism of l-αγ-diaminobutyric acid in a *Xanthomonas* species. Biochem. J. 114(1), 107-115. https://doi.org/10.1042/bj1140107.

Rao, S. L. N., Adiga, P. R., Sarma, P. S., 1964. The isolation and characterization of β-*N*-oxalyl-L-α, β-diaminopropionic acid: a neurotoxin from the seeds of *Lathyrus sativus*. Biochemistry 3, 432-436. https://doi.org/10.1021/bi00891a022.

Rashmi, V., ShylajaNaciyar, M., Rajalakshmi, R., D'Souza, S. F., Prabaharan, D., Uma, L., 2013. Siderophore mediated uranium sequestration by marine cyanobacterium *Synechococcus elongatus* BDU 130911. Bioresour. Technol. 130, 204-210. https://doi.org/10.1016/j.biortech.2012.12.016.

Rastogi, R. P., Madamwar, D., Incharoensakdi, A., 2015. Bloom dynamics of cyanobacteria and their toxins: environmental health impacts and mitigation strategies. Front. Microbiol. 6, 1254. https://doi.org/10.3389/fmicb.2015.01254.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., Huson, D. H., 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. 33, 5799-5808. https://doi.org/10.1093/nar/gki885.

Raven, J. A., Allen, J. F., 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biol. 4, 209. https://doi.org/10.1186/gb-2003-4-3-209.

Ravenhall, M., Škunca, N., Lassalle, F., Dessimoz, C., 2015. Inferring Horizontal Gene Transfer. PLoS Comput. Biol. 11, e1004095. https://doi.org/10.1371/journal.pcbi.1004095.

Reeve, A. M., Breazeale, S. D., Townsend, C. A., 1998. Purification, characterization, and cloning of an S-adenosylmethionine-dependent 3-amino-3-carboxypropyltransferase in nocardicin biosynthesis. J. Biol. Chem. 273(46), 30695-30703. https://doi.org/10.1074/jbc.273.46.30695.

Ressler, C., 1956. Formation of α,γ-diaminobutyric acid from asparagine-containing peptides. J. Am. Chem. Soc. 78(22), 5956-5957. https://doi.org/10.1021/ja01603a068.

Reuter, K., Pittelkow, M., Bursy, J., Heine, A., Craan, T., Bremer, E., 2010. Synthesis of 5-hydroxyectoine from ectoine: crystal structure of the non-heme iron (II) and 2-oxoglutarate-dependent dioxygenase EctD. PloS One 5(5), e10647. https://doi.org/10.1371/journal.pone.0010647.

Réveillon, D., Séchet, V., Hess, P., Amzil, Z., 2016. Production of BMAA and DAB by diatoms (*Phaeodactylum tricornutum*, *Chaetoceros* sp., *Chaetoceros calcitrans* and, *Thalassiosira pseudonana*) and bacteria isolated from a diatom culture. Harmful Algae 58, 45-50. https://doi.org/10.1016/j.hal.2016.07.008.

Reyes, A., Alves, J. M. P., Durham, A. M., Gruber, A., 2017. Use of profile hidden Markov models in viral discovery: Current insights. Adv. Genet. 7, 29-45. https://doi.org/10.2147/AGG.S136574.

Richaume, A., Angle, J. S., Sadowsky, M. J., 1989. Influence of soil variables on *in situ* plasmid transfer from *Escherichia coli* to *Rhizobium fredii*. Appl. Environ. Microbiol. 55, 1730-1734. https://doi.org/10.1128/aem.55.7.1730-1734.1989.

Richaume, A., Smit, E., Faurie, G., van Elsas, J. D., 1992. Influence of soil type on the transfer of plasmid RP4p from *Pseudomonas fluorescens* to introduced recipient and to indigenous bacteria. FEMS Microbiol. Lett. 101, 281-291. https://doi.org/10.1111/j.1574-6968.1992.tb05785.x.

Richter, A. A., Kobus, S., Czech, L., Hoeppner, A., Zarzycki, J., Erb, T. J., Lauterbach, L., Dickschat, J. S., Bremer, E., Smits, S. H. J., 2020. The architecture of the diaminobutyrate acetyltransferase active site provides mechanistic insight into the biosynthesis of the chemical chaperone ectoine. J. Biol. Chem. 295(9), 2822-2838. https://doi.org/10.1074/jbc.RA119.011277.

Richter, A. A., Mais, C.-N., Czech, L., Geyer, K., Hoeppner, A., Smits, S. H. J., Erb, T. J., Bange, G., Bremer, E., 2019. Biosynthesis of the stress-protectant and chemical chaperon ectoine: biochemistry of the transaminase EctB. Front. Microbiol. 10, 2811. https://doi.org/10.3389/fmicb.2019.02811.

Riding, R., 1999. The term stromatolite: towards an essential definition. Lethaia 32, 321-330. https://doi.org/10.1111/j.1502-3931.1999.tb00550.x.

Riekhof, W. R., Andre, C., Benning, C., 2005. Two enzymes, BtaA and BtaB, are sufficient for betaine lipid biosynthesis in bacteria. Arch. Biochem. Biophys. 441(1), 96-105. https://doi.org/10.1016/j.abb.2005.07.001.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., Stanier, R. Y., 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology 111, 1-61. https://doi.org/10.1099/00221287-111-1-1.

Rivera-Aguilar, V., Montjano, G., Rodriguez-Zaragoza, S., Duran-Dias, A., 2006. Distribution and composition of cyanobacteria, mosses and lichens of the biological soil crusts of the Tehuacan Valley, Puebla, Mexico. J. Arid Environ. 67, 208-225. https://doi.org/10.1016/j.jaridenv.2006.02.013.

Robbins, T., Liu, Y.-C., Cane, D. E., Khosla, C., 2016. Structure and mechanism of assembly line polyketide synthases. Curr. Opin. Struct. Biol. 41, 10-18. https://doi.org/10.1016/j.sbi.2016.05.009.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53(1-2), 131-147. https://doi.org/10.1016/0025-5564(81)90043-2.

Roger, P. A., Ladha, J. K., 1992. Biological N_2 fixation in wetland rice fields: estimation and contribution to nitrogen balance. Plant Soil 141, 41-55. https://doi.org/10.1007/978-94-017-0910-1_3.

Ropars, J., de la Vega, R. C. R., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., 2015. Adaptive horizontal gene transfers between multiple cheese-associated fungi. Curr. Biol. 25, 2562-2569. https://doi.org/10.1016/j.cub.2015.08.025.

Rosén, J., Hellenäs, K.-E., 2008. Determination of the neurotoxin BMAA (β-*N*-methylamino-l-alanine) in cycad seed and cyanobacteria by LC-MS/MS (liquid chromatography tandem mass spectrometry). Analyst 133, 1785-1789. https://doi.org/10.1039/B809231A.

Rosén, J., Westerberg, E., Schmiedt, S., Hellenäs, K.-E., 2016. BMAA detected as neither free nor protein bound amino acid in blue mussels. Toxicon 109, 45-50. https://doi.org/10.1016/j.toxicon.2015.11.008.

Rossi, F., Micheletti, E., Bruno, L., Adhikary, S. P., Albertano, P., De Philippis, R., 2012. Characteristics and role of the exocellular polysaccharides produced by five cyanobacteria isolated from phototrophic biofilms growing on stone monuments. Biofouling 28, 215-224. https://doi.org/10.1080/08927014.2012.663751.

Saadi, O. E., Esterman, A. J., Cameron, S., Roder, D. M., 1995. Murray River water, raised cyanobacterial cell counts, and gastrointestinal and dermatological symptoms. Med. J. Aust. 162, 122-125. https://doi.org/10.5694/j.1326-5377.1995.tb138473.x.

Sagan, L., 1967. On the origin of mitosing cells. J. Theor. Biol. 14, 225-274. https://doi.org/10.1016/0022-5193(67)90079-3.

Sakai, N., Ohfune, Y., 1990. Total synthesis of galantin I. Revision of the original structure. Tetrahedron Lett. 31, 3183-3186. https://doi.org/10.1016/s0040-4039(00)94727-0.

Sakai, N., Ohfune, Y., 1992. Total synthesis of galantin I. Acid-catalyzed cyclization of galantinic acid. J. Am. Chem. Soc. 114, 998-1010. https://doi.org/10.1021/ja00029a031.

Saker, M. L., Thomas, A. D., Norton, J. H., 1999. Cattle mortality attributed to the toxic cyanobacterium *Cylindrospermopsis raciborskii* in an outbreak region of north Queensland. Environ. Toxicol. 14, 179-182. https://doi.org/10.1002/(SICI)1522-7278(199902)14:1<179::AID-TOX23>3.0.CO;2-G.

Sánchez-Baracaldo, P., Cardona, T., 2020. On the origin of oxygenic photosynthesis and Cyanobacteria. New Phytol. 225, 1440-1446. https://doi.org/10.1111/nph.16249.

Santos-Aberturas, J., Chandra, G., Frattaruolo, L., Lacret, R., Pham, T. H., Vior, N. M., Eyles, T. H., Truman, A. W., 2019. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. Nucleic Acids Res. 47, 4624-4637. https://doi.org/10.1093/nar/gkz192.

Schaefer, A. M., Yrastorza, L., Stockley, N., Harvey, K., Harris, N., Grady, R., Sullivan, J., McFarland, M., Reif, J. S., 2020. Exposure to microcystin among coastal residents during a cyanobacteria bloom in Florida. Harmful Algae 92, 101769. https://doi.org/10.1016/j.hal.2020.101769.

Schirmacher, A. M., Hanamghar, S. S., Zedler, J. A. Z., 2020. Function and benefits of natural competence in cyanobacteria: from ecology to targeted manipulation. Life 10, 249. https://doi.org/10.3390/life10110249.

Schirrmeister, B. E., Gugger, M., Donoghue, P. C. J., 2015. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. Palaeontology 58, 769-785. https://doi.org/10.1111/pala.12178.

Schirrmeister, B. E., Sanchez-Baracaldo, P., Wacey, D., 2016. Cyanobacterial evolution during the Precambrian. Int. J. Astrobiology 15, 187-204. https://doi.org/10.1017/S1473550415000579.

Schliep, K. P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schönknecht, G., Chen, W.-H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Bräutigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science 339, 1207-1210. https://doi.org/10.1126/science.1231707.

Schopf, J. W., 2012. The fossil record of cyanobacteria. In: Whitton, B. (Ed.) Ecology of Cyanobacteria II. Springer, Dordrecht, pp. 15-36. https://doi.org/10.1007/978-94-007-3855-3_2.

Schopf, J. W., Packer, B. M., 1987. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. Science 237, 70-73. https://doi.org/10.1126/science.11539686.

Seckbach, J., 2007. Algae and cyanobacteria in extreme environments. Springer Science & Business Media, Dordrecht, Netherlands, pp. 101-123.

Seckbach, J., Oren, A., 2007. Oxygenic photosynthetic microorganisms in extreme environments. In: Seckbach, J. (Ed.) Algae and Cyanobacteria in Extreme Environments. Springer Netherlands, Dordrecht, pp. 3-25.

Seneviratne, A. S., Fowden, L., 1968. The amino acids of the genus *Acacia*. Phytochemistry 7, 1039-1045. https://doi.org/10.1016/S0031-9422(00)88249-7.

Shah, P., Swiatlo, E., 2008. A multifaceted role for polyamines in bacterial pathogens. Mol. Microbiol. 68(1), 4-16. https://doi.org/10.1111/j.1365-2958.2008.06126.x.

Shcolnick, S., Keren, N., 2006. Metal homeostasis in cyanobacteria and chloroplasts. Balancing benefits and risks to the photosynthetic apparatus. Plant Physiol. 141(3), 805-810. https://doi.org/10.1104/pp.106.079251.

Sheath, R. G., Cole, K. M., 1992. Biogeography of stream macroalgae in North America. J. Phycol. 28, 448-460. https://doi.org/10.1111/j.0022-3646.1992.00448.x.

Sheppard, R. J., Beddis, A. E., Barraclough, T. G., 2020. The role of hosts, plasmids and environment in determining plasmid transfer rates: a meta-analysis. Plasmid 108, 102489. https://doi.org/10.1016/j.plasmid.2020.102489.

Shoji, J. I., Sakazaki, R., Wakisaka, Y., Koizumi, K., Mayama, M., Matsuura, S., 1975. Isolation of galantins I and II, water-soluble basic peptides studies on antibiotics from the genus *Bacillus*. III. J. Antibiot. 28, 122-125. https://doi.org/10.7164/antibiotics.28.122.

Shojima, S., Nishizawa, N.-K., Fushiya, S., Nozoe, S., Irifune, T., Mori, S., 1990. Biosynthesis of phytosiderophores: in vitro biosynthesis of 2′-deoxymugineic acid from L-methionine and nicotianamine. Plant Physiol. 93(4), 1497-1503. https://doi.org/10.1104/pp.93.4.1497.

Shojima, S., Nishizawa, N.-K., Mori, S., 1989. Establishment of a cell-free system for the biosynthesis of nicotianamine. Plant. Cell. Physiol. 30, 673-677. https://doi.org/10.1093/oxfordjournals.pcp.a077792.

Siegesmund, M. A., Johansen, J. R., Karsten, U., Friedl, T., 2008. *Coleofasciculus* gen. nov.(Cyanobacteria): morphological and molecular criteria for revision of the genus *Microcoleus gomont* 1. J. Phycol. 44, 1572-1585. https://doi.org/10.1111/j.1529-8817.2008.00604.x.

Simpson, F. B., Neilands, J. B., 1976. Siderochromes in cyanophyceae: isolation and characterization of schizokinen from *Anabaena* sp. 1. J. Phycol. 12, 44-48. https://doi.org/10.1111/j.1529-8817.1976.tb02824.x.

Singh, R., Parihar, P., Singh, M., Bajguz, A., Kumar, J., Singh, S., Singh, V. P., Prasad, S. M., 2017. Uncovering potential applications of cyanobacteria and algal metabolites in biology, agriculture and medicine: current status and future prospects. Front. Microbiol. 8, 515. https://doi.org/10.3389/fmicb.2017.00515.

Singh, S. P., Montgomery, B. L., 2011. Determining cell shape: adaptive regulation of cyanobacterial cellular differentiation and morphology. Trends Microbiol. 19, 278-285. https://doi.org/10.1016/j.tim.2011.03.001.

Siodłak, D., 2015. α, β-dehydroamino acids in naturally occurring peptides. Amino Acids 47, 1-17. https://doi.org/10.1007/s00726-014-1846-4.

Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., Li, H., Ranieri, M. R. M., Webster, A. L. H., Cao, M., 2020. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat. Commun. 11, 1-9. https://doi.org/10.1038/s41467-020-19986-1.

Sleator, R. D., 2011. Phylogenetics. Arch. Microbiol. 193, 235-239. https://doi.org/10.1007/s00203-011-0677-x.

Smiles, D. E., 1988. Aspects of the physical environment of soil organisms. Biol. Fertil. Soils 6, 204-215.

Sonnhammer, E. L. L., Eddy, S. R., Birney, E., Bateman, A., Durbin, R., 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 26(1), 320-322. https://doi.org/10.1093/nar/26.1.320.

Sotiroudis, T. G., Sotiroudis, G. T., 2013. Health aspects of Spirulina (Arthrospira) microalga food supplement. J. Serbian Chem. Soc. 78, 395-405. https://doi.org/10.2298/JSC121020152S.

Soule, T., Palmer, K., Gao, Q., Potrafka, R. M., Stout, V., Garcia-Pichel, F., 2009. A comparative genomics approach to understanding the biosynthesis of the sunscreen scytonemin in cyanobacteria. BMC Genom. 10, 1-10. https://doi.org/10.1186/1471-2164-10-336.

Spencer, P. S., 2019. Hypothesis: etiologic and molecular mechanistic leads for sporadic neurodegenerative diseases based on experience with Western Pacific ALS/PDC. Front. Neurol. 10, 754. https://doi.org/10.3389/fneur.2019.00754.

Spencer, P., Fry, R., Palmer, V. S., Kisby, G. E., 2012. Western Pacific ALS-PDC: a prototypical neurodegenerative disorder linked to DNA damage and aberrant proteogenesis? Front. Neurol. 3, 180. https://doi.org/10.3389/fneur.2012.00180.

Spencer, P., Nunn, P., Hugon, J., Ludolph, A., Roy, D., 1986. Motorneurone disease on Guam: possible role of a food neurotoxin. Lancet 327, 965. https://doi.org/10.1016/S0140-6736(86)91059-7.

Stal, L. J., Krumbein, W. E., 1987. Temporal separation of nitrogen fixation and photosynthesis in the filamentous, non-heterocystous cyanobacterium *Oscillatoria* sp. Arch. Microbiol. 149, 76-80. https://doi.org/10.1007/BF00423140.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690. https://doi.org/10.1093/bioinformatics/btl446.

Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J., Hranueli, D., 2008. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. Nucleic Acids Res. 36, 6882-6892. https://doi.org/10.1093/nar/gkn685.

Steffensen, D. A., 2008. Economic cost of cyanobacterial blooms. In: Hudnell, H. K. (Ed.), Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. Springer New York, New York, NY, pp. 855-865.

Steinbüchel, A., Füchtenbusch, B., Gorenflo, V., Hein, S., Jossek, R., Langenbach, S., Rehm, B. H. A., 1998. Biosynthesis of polyesters in bacteria and recombinant organisms. Polym. Degrad. Stab. 59, 177-182. https://doi.org/10.1016/S0141-3910(97)00173-0.

Stevens-Jr, S. E., Porter, R. D., 1986. Heterospecific transformation among cyanobacteria. J. Bacteriol. 167, 1074-1076. https://doi.org/10.1128/jb.167.3.1074-1076.1986.

Stewart, I., Schluter, P. J., Shaw, G. R., 2006. Cyanobacterial lipopolysaccharides and human health – a review. Environ. Health 5, 7. https://doi.org/10.1186/1476-069X-5-7.

Stibal, M., Sabacká, M., Kastovská, K., 2006. Microbial communities on glacier surfaces in Svalbard: impact of physical and chemical properties on abundance and structure of cyanobacteria and algae. Microb. Ecol. 52, 644-654. https://doi.org/10.1007/s00248-006-9083-3.

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., Durand, D., 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics 28, i409-i415. https://doi.org/10.1093/bioinformatics/bts386.

Sturgill, G., Rather, P. N., 2004. Evidence that putrescine acts as an extracellular signal required for swarming in *Proteus mirabilis*. Mol. Microbiol. 51(2), 437-446. https://doi.org/10.1046/j.1365-2958.2003.03835.x.

Sukenik, A., Zohary, T., Padisák, J., 2009. Cyanoprokaryota and other prokaryotic algae. In: Likens, G. E. (Ed.) Encyclopedia of Inland Waters, vol. 1. Elsevier, Oxford, UK, pp. 138-148. https://doi.org/10.1016/B978-012370626-3.00133-2.

Sun, D., Jeannot, K., Xiao, Y., Knapp, C. W., 2019. Editorial: horizontal gene transfer mediated bacterial antibiotic resistance. Front. Microbiol. 10, 1933-1933. https://doi.org/10.3389/fmicb.2019.01933.

Sunda, W. G., Huntsman, S. A., 2015. High iron requirement for growth, photosynthesis, and low-light acclimation in the coastal cyanobacterium *Synechococcus bacillaris*. Front. Microbiol. 6, 561. https://doi.org/10.3389/fmicb.2015.00561.

Susko, E., Lincker, L., Roger, A. J., 2018. Accelerated estimation of frequency classes in site-heterogeneous profile mixture models. Mol. Biol. Evol. 35(5), 1266–1283. https://doi.org/10.1093/molbev/msy026.

Svircev, Z., Krstic, S., Miladinov-Mikov, M., Baltic, V., Vidovic, M. 2009. Freshwater cyanobacterial blooms and primary liver cancer epidemiological studies in Serbia. J. Environ. Sci. Health C. 27, 36-55. https://doi.org/10.1080/10590500802668016.

Swofford, D. L., Olsen, G. J., Waddell, P. J., Hillis, D. M. 1996. Phylogenic inference. In: Hillis, D. M., Moritz, C., Mable, B. K. (Eds.) Molecular Systematics, 2nd Edition. Sinauer Associated Sunderland, MA, pp. 407-514.

Synge, R. L., 1939. The Hofmann degradation of glutamine residues in gliadin. Biochem. J. 33(5), 671–678. https://doi.org/10.1042/bj0330671.

Syvanen, M., 1994. Horizontal gene transfer: evidence and possible consequences. Annu. Rev. Genet. 28, 237-261. https://doi.org/10.1146/annurev.ge.28.120194.001321.

Takeuchi, N., 2001. The altitudinal distribution of snow alge on an Alaska glacier (Gulkana Glacier in the Alaska Range). Hydrol. Process 15, 3447-3459. https://doi.org/10.1002/hyp.1040.

Tan, V. X., Lassus, B., Lim, C. K., Tixador, P., Courte, J., Bessede, A., Guillemin, G. J., Peyrin, J.-M., 2018. Neurotoxicity of the cyanotoxin BMAA through axonal degeneration and intercellular spreading. Neurotox. Res. 33, 62-75. https://doi.org/10.1007/s12640-017-9790-1.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., Koonin, E. V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33-36. https://doi.org/10.1093/nar/28.1.33.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., Tolstoy, I., 2014. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 42, D553-D559. https://doi.org/10.1093/nar/gkt1274.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44, 6614-6624. https://doi.org/10.1093/nar/gkw569.

Taya, Y., Tanaka, Y., Nishimura, S., 1978. Cell-free biosynthesis of discadenine, a spore germination inhibitor of *Dictyostelium discoideum*. FEBS Lett. 89(2), 326-328. https://doi.org/10.1016/0014-5793(78)80247-6.

Taylor, E. L., Taylor, T. N., Krings, M., 2009. Paleobotany: The Biology and Evolution of Fossil Plants. Academic Press, Inc., San Diego, CA.

Teixeira, M. G., Costa, M. C., de Carvalho, V. L., Pereira, M. S., Hage, E. 1993. Gastroenteritis epidemic in the area of Itaparica Dam, Bahia, Brazil. Bull. Pan. Am. Health Organ. 27, 244-253.

Thomas, M. G., Chan, Y. A., Ozanick, S. G., 2003. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. Antimicrob. Agents Chemother. 47, 2823-2830. https://doi.org/10.1128/AAC.47.9.2823-2830.2003.

Tomitani, A., Knoll, A. H., Cavanaugh, C. M., Ohno, T., 2006. The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. Proc. Natl. Acad. Sci. 103, 5442-5447. https://doi.org/10.1073/pnas.0600999103.

Tooming-Klunderud, A., Sogge, H., Rounge, T. B., Nederbragt, A. J., Lagesen, K., Glöckner, G., Hayes, P. K., Rohrlack, T., Jakobsen, K. S., 2013. From green to red: horizontal gene transfer of the phycoerythrin gene cluster between *Planktothrix* strains. Appl. Environ. Microbiol. 79, 6803-6812. https://doi.org/10.1128/AEM.01455-13.

Trehan, K., Sinha, U., 1981. Genetic transfer in a nitrogen-fixing filamentous cyanobacterium. J. Gen. Microbiol. 124, 349-352. https://doi.org/10.1099/00221287-124-2-349.

Tripathi, A., Gottesman, S., 2016. Phosphate on, rubbish out. Nature 539, 38-39. https://doi.org/10.1038/539038a.

Turner, S., Pryer, K. M., Miao, V. P. W., Palmer, J. D., 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J. Eukaryot. Microbiol. 46, 327-338. https://doi.org/10.1111/j.1550-7408.1999.tb04612.x.

Tuner, P., Gammie, A., Hollinrake, K., Codd, G. A. 1990. Pneumonia associated with contact with cyanobacteria. B. J. M., 300, 1440-1441. https://doi.org/10.1136/bmj.300.6737.1440.

Uo, T., Yoshimura, T., Nishiyama, T., Esaki, N., 2002. Gene cloning, purification, and characterization of 2,3-diaminopropionate ammonia-lyase from *Escherichia coli*. Biosci., Biotechnol., Biochem. 66, 2639-2644. https://doi.org/10.1271/bbb.66.2639.

Utkilen, H., Gjølme, N., 1995. Iron-stimulated toxin production in *Microcystis aeruginosa*. Appl. Environ. Microbiol. 61, 797-800. https://doi.org/10.1128/aem.61.2.797-800.1995.

van Dongen, S., 2000. A cluster algorithm for graphs. Technical Report INSR0010. CWI.

van Heel, A. J., de Jong, A., Montalban-Lopez, M., Kok, J., Kuipers, O. P., 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides. Nucleic Acids Res. 41, W448-W453. https://doi.org/10.1093/nar/gkt391.

van Kranendonk, M. J., 2006. Volcanic degassing, hydrothermal circulation and the flourishing of early life on Earth: A review of the evidence from c. 3490-3240 Ma rocks of the Pilbara Supergroup, Pilbara Craton, Western Australia. Earth-Sci. Rev. 74, 197-240. https://doi.org/10.1016/j.earscirev.2005.09.005.

Vega, A., Bell, E. A., 1967. α-Amino-β-methylaminopropionic acid, a new amino acid from seeds of *Cycas circinalis*. Phytochemistry 6, 759-762. https://doi.org/10.1016/S0031-9422(00)86018-5.

Vincent, W. F., 2007. Cold Tolerance in Cyanobacteria and Life in the Cryosphere. In: Seckbach, J. (Ed.), Algae and Cyanobacteria in Extreme Environments. Springer Netherlands, Dordrecht, pp. 287-301.

Violi, J. P., Facey, J. A., Mitrovic, S. M., Colville, A., Rodgers, K. J., 2019. Production of β-methylamino-L-alanine (BMAA) and its isomers by freshwater diatoms. Toxins 11(9), 512. https://doi.org/10.3390/toxins11090512.

Volcani, B. E., 1944. The microorganisms of the Dead Sea. In: Papers collected to commemorate the 70th anniversary of Dr Chaim Weizmann. Collective volume. Daniel Sieff Research Institute, Rehovoth, pp. 71-85.

Walsh, T. A., Green, S. B., Larrinua, I. M., Schmitzer, P. R., 2001. Characterization of plant β-ureidopropionase and functional overexpression in *Escherichia coli*. Plant Physiol. 125, 1001-1011. https://doi.org/10.1104/pp.125.2.1001.

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., Thompson, C. C., 2017. Ecogenomics and taxonomy of cyanobacteria phylym. Front. Microbiol. 8, 2132. https://doi.org/10.3389/fmicb.2017.02132.

Wan, Z., Varshavsky, J., Teegala, S., McLawrence, J., Goddard, Noel L., 2011. Measuring the rate of conjugal plasmid transfer in a bacterial population using quantitative PCR. Biophys. J. 101, 237-244. https://doi.org/10.1016/j.bpj.2011.04.054.

Wang, H. C., Minh, B. Q., Susko, E., Roger, A. J., 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67(2), 216-235. https://doi.org/10.1093/sysbio/syx068.

Wang, M., Gould, S. J., 1993. Biosynthesis of capreomycin. 2. Incorporation of L-serine, L-alanine, and L-2,3-diaminopropionic acid. J. Org. Chem. 58, 5176-5180. https://doi.org/10.1021/jo00071a029.

Ward, A. K., 1985. Factors affecting distribution of Nostoc in Cascade Mountain streams of Western Oregon, USA. Verh. Int. Ver. Limnol. 22, 2799-2804. https://doi.org/10.1080/03680770.1983.11897777.

Ward, D. M., Bateson, M. M., Ferris, M. J., Kühl, M., Wieland, A., Koeppel, A., Cohan, F. M., 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. Philos Trans. R. Soc. Lond. B. 361, 1997-2008. https://doi.org/10.1098/rstb.2006.1919.

Ward, D. M., Castenholz, R. W., Miller, S. R., 2012. Cyanobacteria in Geothermal habitats. In: Whitton, B. A. (Ed.) The Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer, Dordrecht, pp. 39-63. https://doi.org/10.1007/978-94-007-3855-3_3.

Waskom, M. L., 2021. Seaborn: statistical data visualization. J. Open Source Softw. 6, 3021. https://doi.org/10.21105/joss.03021.

Watanabe, I., Rockwood, W. G., 1979. Nitrogen and rice. IRRI, Los Banos, Phillippines.

Watanabe, T., Horiike, T., 2021. The evolution of molybdenum dependent nitrogenase in cyanobacteria. Biology 10, 329. https://doi.org/10.3390/biology10040329.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., Barton, G. J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9), 1189-1191. https://doi.org/10.1093/bioinformatics/btp033.

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D. H., Wohlleben, W., 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J. Biotechnol. 140, 13-17. https://doi.org/10.1016/j.jbiotec.2009.01.007.

Webster, A. J., Payne, R. J. H., Pagel, M., 2003. Molecular phylogenies link rates of evolution and speciation. Science 301, 478-478. https://doi.org/10.1016/j.ijbiomac.2021.06.186.

Whelan, S., Liò, P., Goldman, N., 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. TRENDS Genet. 17, 262-272. https://doi.org/10.1016/S0168-9525(01)02272-7.

Whitton, B. A. (Ed.), 2012. Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer, Dordrecht, 760 pp.

Whitton, B. A., Potts, M., 2012. Introduction to the cyanobacteria. In: Whitton, B. A., Potts, M. (Eds.) The Ecology of Cyanobacteria II. Springer, Dordrecht, pp. 1-13. https://doi.org/10.1007/0-306-46855-7_1.
WHO (2014). Antimicrobial Resistance Global Report on Surveillance. World Health Organization. Available online at: https://www.who.int/antimicrobial-resistance/publications/en/

Widderich, N., Hoeppner, A., Pittelkow, M., Heider, J., Smits, S. H. J., Bremer, E., 2014a. Biochemical properties of ectoine hydroxylases from extremophiles and their wider taxonomic distribution among microorganisms. PloS One 9(4), e93809. https://doi.org/10.1371/journal.pone.0093809.

Widderich, N., Kobus, S., Höppner, A., Riclea, R., Seubert, A., Dickschat, J. S., Heider, J., Smits, S. H. J., Bremer, E., 2016. Biochemistry and crystal structure of ectoine synthase: a metal-containing member of the cupin superfamily. PLoS One 11(3), e0151285. https://doi.org/10.1371/journal.pone.0151285.

Widderich, N., Pittelkow, M., Höppner, A., Mulnaes, D., Buckel, W., Gohlke, H., Smits, S. H. J., Bremer, E., 2014b. Molecular dynamics simulations and structure-guided mutagenesis provide insight into the architecture of the catalytic core of the ectoine hydroxylase. J. Mol. Biol. 426(3), 586-600. https://doi.org/10.1016/j.jmb.2013.10.028.

Wiedenbeck, J., Cohan, F. M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35, 957-976. https://doi.org/10.1111/j.1574-6976.2011.00292.x.

Wilks, S. S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9, 60-62. https://doi.org/10.1214/aoms/1177732360.

Woese, C. R., Fox, G. E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. 74, 5088-5090. https://doi.org/10.1073/pnas.74.11.5088.

Wolk, C. P., Ernst, A., Elhai, J., 1994. Heterocyst metabolism and development. In: Bryant, D. A. (Ed.) The Molecular Biology of Cyanobacteria. Springer, Dordrecht, pp. 769-823.

Wolk, C. P., Vonshak, A., Kehoe, P., Elhai, J., 1984. Construction of shuttle vectors capable of conjugative transfer from Escherichia coli to nitrogen-fixing filamentous cyanobacteria. Proc. Natl. Acad. Sci. USA 81, 1561-1565. https://doi.org/10.1073/pnas.81.5.1561.

Wong, F. C. Y., Meeks, J. C., 2002. Establishment of a functional symbiosis between the cyanobacterium *Nostoc punctiforme* and the bryophyte *Anthoceros punctatus* requires genes involved in nitrogen control and initiation of heterocyst differentiation. The GenBank accession numbers for the coding regions of ntcA, hetR and hetF are AY038370, AF318069 and AF288130, respectively. Microbiology 148, 315-323. https://doi.org/10.1099/00221287-148-1-315.

Wu, M., Eisen, J. A., 2008. A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 9, R151. https://doi.org/10.1186/gb-2008-9-10-r151.

Wynn-Williams, D. D., 2000. Cyanobacteria in deserts—life at the limit? In: Whitton, B. A., Potts, M. (Eds.), The ecology of cyanobacteria. Springer, Kluwer, The Netherlands, pp. 341-366.

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., Huang, G., Li, Y., Yan, Q., Wu, S., Wang, X., Chen, S., He, G., Xiao, X., Xu, A., 2011. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J. 5, 414-426. https://doi.org/10.1038/ismej.2010.144.

Xu, Z., Sun, Z., Li, S., Xu, Z., Cao, C., Xu, Z., Feng, X., Xu, H., 2015. Systematic unravelling of the biosynthesis of poly (L-diaminopropionic acid) in *Streptomyces albulus* PD-1. Scientific Rep. 5, 1-10. https://doi.org/10.1038/srep17400.

Yamamoto, S., Hamanaka, K., Suemoto, Y., Ono, B.-i., Shinoda, S., 1986. Evidence for the presence of a novel biosynthetic pathway for norspermidine in *Vibrio*. Can. J. Microbiol. 32(2), 99-103. https://doi.org/10.1139/m86-021.

Yamamoto, S., Mutoh, N., Tsuzuki, D., IkAi, H., Nakao, H., Shinoda, S., NariMatsu, S., Miyoshi, S.-i., 2000. Cloning and characterization of the ddc homolog encoding L-2,4-diaminobutyrate decarboxylase in *Enterobacter aerogenes*. Biol. Pharm. Bull. 23(5), 649-653. https://doi.org/10.1248/bpb.23.649.

Yamamoto, S., Tsuzaki, Y., Tougou, K., Shinoda, S., 1992. Purification and characterization of L-2,4-diaminobutyrate decarboxylase from *Acinetobacter calcoaceticus*. Microbiol. 138(7), 1461-1465. https://doi.org/10.1099/00221287-138-7-1461.

Yamanaka, K., Fukumoto, H., Takehara, M., Hamano, Y., Oikawa, T., 2020. The stereocontrolled biosynthesis of mirror-symmetric 2,4-diaminobutyric acid homopolymers is critically governed by adenylation activations. ACS Chem. Biol. 15(7), 1964-1973. https://doi.org/10.1021/acschembio.0c00321.

Yang, Z. 2006. Computational molecular evolution. Oxford: Oxford University Press.

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13, 303-314. https://doi.org/10.1038/nrg3186.

Yerrapragada, S., Siefert, J. L., Fox, G. E., 2009. Horizontal gene transfer in cyanobacterial signature genes. In: Gogarten, M. B., Gogarten, J. P., Olendzenski, L. C. (Eds.) Horizontal Gene Transfer. Methods in Molecular Biology, vol. 532. Humana Press, pp. 339-366. https://doi.org/10.1007/978-1-60327-853-9_20.

Yoon, B.-J., 2009. Hidden Markov models and their applications in biological sequence analysis. Curr. genomics 10, 402-415. https://doi.org/10.2174/138920209789177575.

Žegura, B., Štraser, A., Filipič, M., 2011. Genotoxicity and potential carcinogenicity of cyanobacterial toxins–a review. Mutation Res. - Rev. Mutation Res. 727, 16-41. https://doi.org/10.1016/j.mrrev.2011.01.002.

Zeidner, G., Bielawski, J. P., Shmoish, M., Scanlan, D. J., Sabehi, G., Béjà, O., 2005. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. Environ. Microbiol. 7, 1505-1513. https://doi.org/10.1111/j.1462-2920.2005.00833.x.

Zhang, P., Peng, C., Zhang, J., Zhang, J., Chen, J., Zhao, H., 2022. Long-term harmful algal blooms and nutrients patterns affected by climate change and anthropogenic pressures in the Zhanjiang Bay, China. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.849819.

Zhao, C., Song, C., Luo, Y., Yu, Z., Sun, M., 2008. L-2,3-diaminopropionate: one of the building blocks for the biosynthesis of Zwittermicin A in *Bacillus thuringiensis* subsp. *kurstaki* strain YBT-1520. FEBS Lett. 582, 3125-3131. https://doi.org/10.1016/j.febslet.2008.07.054.

Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., Papke, R. T., 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res. 16, 1099-1108. http://www.genome.org/cgi/doi/10.1101/gr.5322306.

Zhou, X., Shen, X.-X., Hittinger, C. T., Rokas, A., 2018. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Mol. Biol. Evol. 35(2), 486-503. https://doi.org/10.1093/molbev/msx302.

Zhu, Q., Kosoy, M., Dittmar, K., 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. BMC Genom. 15, 1-18. https://doi.org/10.1186/1471-2164-15-717.

Zhu, X., Li, Q., Yin, C., Fang, X., Xu, X., 2015. Role of spermidine in overwintering of cyanobacteria. J. Bacteriol. 197(14), 2325-2334. https://doi.org/10.1128/JB.00153-15.

Zinder N. D., Lederberg, J., 1952. Genetic exchange in *Salmonella*. J. Bacteriol. 64, 679-699. https://doi.org/10.1128/jb.64.5.679-699.1952.

Zoonomia, Consortium., 2020. A comparative genomics multitool for scientific discovery and conservation. Nature 587, 240. https://doi.org/10.1038/s41586-020-2876-6.

Appendix A

Supplementary material for Chapter 2

Supplementary data to this chapter can be found online at the University of Edinburgh's DataShare repository, https://doi.org/10.7488/ds/3057.

Ziying Ke performed the reconstruction of the cyanobacterial species phylogeny (Section S6), including genome selection (Section S4) and orthogroup prediction (Section S5).

Appendix B

Supplementary material for Chapter 3

Supplementary data to this chapter can be found online at the University of Edinburgh's DataShare repository, https://doi.org/10.7488/ds/3270.

Appendix C

Supplementary material for Chapter 4

Supplementary data to this chapter was submitted to DataSync as a separate online supplement along with the thesis (file name: "Supplementary_materials_HGT").

Ziying Ke performed the orthogroup prediction (Section S1).