Title: The evolutionary advantage of toxin production among cyanobacteria, the oldest known organisms on Earth

Short title: Benefits of toxin production by cyanobacteria

Authors: Ravi José Tristão Ramos,^{a,b} Crina-Maria Ionescu, PhD,^{a,b} Jaroslav Koča, Prof, DSc^{a,b}

Affiliations

^aCEITEC-Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

^bNational Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

Authors' email addresses: RJTR, 455306@mail.muni.cz; CMI, crina-maria.ionescu@ceitec.muni.cz; JK, jkoca@ceitec.cz

Corresponding author: Jaroslav Koča, Professor, D.Sc. CEITEC-Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic. Telephone and fax: +420 54949 4947, +420 54949 2685, +420 54949 2552. Email address: jkoca@ceitec.cz

The evolutionary advantage of toxin production among cyanobacteria, the oldest known organisms on Earth

Cyanobacteria produce toxic secondary metabolites for reasons hitherto unclear. Using a phylogenetic approach that accounts for the high complexity of biosynthetic gene clusters (full or partial inversions, variable length, different number of genes, non-orthologues), we analyzed the sequences of 76 biosynthetic gene clusters covering 19 cyanotoxins. The phylogenetic tree of biosynthetic gene clusters branches first according to the bioactivity of the toxic metabolite (molecular target in another organism), then according to the chemical class and chemical structure of the secondary metabolite, and finally according to the organism and area of origin. The bioactivity of a toxic metabolite can be deduced directly from the nucleotide sequence of the biosynthetic gene cluster, without needing to examine the enzymes themselves or to measure expression levels. Bioactivity may have been the primary driving force behind the diversity of secondary metabolism in cyanobacteria. This genetic machinery evolved to facilitate three specific survival strategies acting separately or in tandem, with dominant cyanobacteria possessing the genetic machinery to support all three strategies. Transmembrane (direct) toxicity targeting ion channels, intracellular (indirect) toxicity targeting cell-cycle regulation, and digestion inhibition targeting proteases may have provided the survival advantage underpinning the evolutionary success of both cyanobacteria and their early symbiotic hosts.

Keywords: cyanotoxin, evolutionary benefit, bioactivity, phylogenetic analysis

Introduction

Bioactive secondary metabolites, which are products of pathways not directly involved in growth, development, or reproduction, often serve complex functions that aid survival and increase dominance of an organism in a given ecological niche. Snake venoms are probably the most famous examples of bioactive secondary metabolites, though not nearly the most versatile. Bacteria, fungi, and plants all exhibit an impressive array of secondary metabolites, with cyanobacteria being the earliest organisms to develop such capabilities. Cyanobacteria are unicellular photosynthetic prokaryotes populating marine, fresh-water, and land ecosystems, which drove the rapid oxygenation of Earth's atmosphere and are recognized as the ancestors of plastids¹. Cyanobacteria continue to exert a crucial environmental impact because they produce about 25% of all carbohydrates on Earth and perform massive nitrogen fixation together with sequestering trace metals and phosphorous from the environment, thus fueling most contemporary food chains. In addition, many cyanobacteria produce unique secondary metabolites (typically, peptides and alkaloids) via non-ribosomal peptide synthesis (NRPS) and/or polyketide synthesis (PKS), with complex environmental, economic, and health impact².

The exact function of cyanobacterial secondary metabolites remains largely unknown² despite the fact that a substantial portion of the genome can be dedicated to this secondary metabolism³. These bioactive compounds are often referred to generically as cyanotoxins because toxic metabolites are found in great quantities during cyanobacterial blooms that kill local aquatic fauna. However, not all secondary metabolites of cyanobacterial NRPS/PKS are especially toxic, and some even have strong potential for use as pharmaceuticals, pesticides, or biofuels^{4,5}. It is strongly suspected that production of toxins and other bioactive peptides is

linked to the evolution of certain cyanobacterial species or niche populations^{2,4,6}. To date, the following potential roles have been suggested for bioactive NRPS/PKS products: (i) photoprotection due to absorption of harmful ultra-violet radiation; (ii) facilitating the acquisition or more efficient use of available but limiting resources; (iii) impeding growth or resource use of competitors such as green algae or other cyanobacteria; and (iv) limiting predation by inducing toxicity in organisms predating on cyanobacteria^{2,4,6}. While the role of cyanobacterial metabolites in photoprotection, resource monopolization, and habitat dominance has been demonstrated experimentally, hypotheses regarding the anti-predation role are not entirely consistent^{2,4,6}. Specifically, some studies have provided evidence of co-evolutionary adaptation between toxin-producing cyanobacteria and its natural predator, zooplankton⁷, while others have suggested that the oldest cyanotoxin genes appeared before zooplankton⁸. Thus, to clarify the evolutionary role of bioactive secondary metabolites in cyanobacteria, it is necessary to take a second look at the genetic machinery responsible for the production of such compounds.

It is currently possible to study secondary metabolism in great detail using dedicated high-throughput omics technologies that enable to quantify the levels of entire pools of transcripts, proteins, and metabolites in a given cell, in a particular tissue, and even in environments populated by multiple species. However, substantial information regarding the nature of the final metabolite can be gained simply from examining the genes involved⁹. Moreover, it is currently possible to predict the chemical class of a metabolite using only the translated sequence of the genes coding for the set of enzymes responsible for metabolite production, together with information from dedicated databases covering known genes, enzymes, pathways, and products of secondary metabolism¹⁰. Going one step further, one may wonder whether the nucleotide sequences biosynthetic gene clusters themselves already contain the key

information regarding the impact of the secondary metabolite in the environment (i.e., bioactivity). Thus, we set out to examine the relationship between nucleotide sequence similarity and bioactivity of secondary metabolites produced by complex gene clusters performing NRPS/PKS in various species of cyanobacteria. Understanding these relationships may provide insight into the potential contribution of toxin-producing gene clusters in the overwhelming evolutionary success of cyanobacteria and in their role as promoters of life diversification on Earth.

However, analyzing cyanobacterial genomes is challenging because cyanobacteria, which are the earliest known organisms on our planet, have accumulated a huge amount of lateral gene transfer and gene shuffling events over geological time^{11,12}. In fact, the phylogenetic relationships among cyanobacterial species remain an important topic of debate¹³⁻¹⁵. Regarding the genetic basis of cyanotoxin production, the challenge lies in the high variability of the biosynthetic gene clusters. Tens of secondary metabolites with different structure and bioactivity may be produced via NRPS/PKS within the same genus, yet different strains of the same species might not produce the same metabolites^{6,12,13}. Corresponding NRPS/PKS gene clusters in different species or even in different strains of the same species differ regarding the presence, order, direction, and length of genes, which may or may not be orthologous^{16,17}. It remains unclear just how much horizontal gene transfer occurs from different strains and different species of cyanobacteria, as well as from non-cyanobacterial species. For example, upon phylogenetic analysis of cylindrospermopsin-producing genes, *Cylindrospermopsis raciborskii* from Spain grouped together with Tunisian and American strains and not with other European strains¹⁸ even though *C. raciborskii* is reported to form a tight monophyletic group with specific continental distribution in terms of normal phylogenetic markers such as the 16s rRNA gene^{19,20}. In the case

of the saxitoxin-producing gene cluster, mainly vertical transfer congruent with 16s rRNA evolution was suggested for some components (e.g., genes encoding for aminotransferase or histidine kinase), whereas horizontal gene transfer has been suggested for other components (e.g., genes encoding for polyketide synthase or amidinotransferase)^{21,22}. A recent comparative genomics analysis highlighted evidence for massive gene shuffling within biosynthetic gene clusters across the *Cyanobacteria* phylum, with potential crossover to and from non-cyanobacterial species, and without coherence with species phylogeny¹². It is clear that a multi-gene phylogenetic approach is required here, but such analyses involving highly divergent non-orthologues have been very challenging^{12,23}. Therefore, it is desirable to apply a unified measure of sequence similarity that can be used to compare complex stretches of DNA along entire biosynthetic gene clusters even in the absence of further information regarding evolutionary mechanisms.

To tackle this complex problem, we employed a phylogenetic approach that bridges alignment-based and alignment-free phylogenomics, facilitating the analysis of long stretches of DNA containing orthologous and non-orthologous sequences with high or uneven divergence, gene shuffling, and inversions. We found that the sequence of complex biosynthetic gene clusters involved in the secondary metabolism of cyanobacteria contains key information regarding the bioactivity of the final product. These relationships were apparent directly from the nucleotide sequence of the biosynthetic gene clusters, without the need to examine the enzymes encoded by the genes in each cluster or to measure protein or RNA expression levels, suggesting that bioactivity may have been the primary driving force behind the diversity of the genetic machinery for secondary metabolism in cyanobacteria. Finally, we explored the implications of our findings in the context of cyanobacterial evolution. Our findings suggest that the genetic

machinery responsible for secondary metabolism in cyanobacteria may have evolved to facilitate specific survival strategies in addition to a potential role in resource management. Transmembrane toxicity targeting ion channels, intracellular toxicity targeting cell-cycle regulation, and digestion inhibition targeting proteases may have provided the survival advantage underpinning the evolutionary success of both cyanobacteria and their early symbiotic hosts.

Results

Key information regarding the bioactivity of a secondary metabolite can be found directly within the nucleotide sequence of its biosynthetic gene cluster

We surveyed the NCBI Nucleotide database and obtained the sequences of 76 gene clusters responsible for NRPS/PKS in cyanobacteria (approximately 3000–64000 bp each), covering 19 different products (cyanotoxins). The complexity of these gene clusters is illustrated in Supplementary Table S1, which summarizes the organism of origin, the number and order of genes in each cluster, and the nature of the secondary metabolite produced by the enzymes encoded by the gene cluster.

To account for the complex nature of these NRPS/PKS gene clusters, we employed a phylogenetic approach built around local pairwise sequence alignment (LPSA), which is commonly used to identify, categorize, annotate, and filter out gene or protein sequences obtained from one or more species. LPSA has been shown to perform well for phylogenetic analyses involving multiple genes and even whole genomes of viruses, eukaryotes, and bacteria, including cyanobacteria²⁴⁻²⁸. The LPSA-based approach used in this study (<https://crocodile.ncbr.muni.cz/>) involves three steps, namely: (i) break each sequence into

subsequences; (ii) using the Basic Local Alignment Search Tool (BLAST)^{29,30}, perform local alignment of each pair of subsequences generated for the entire dataset; and (iii) using the BLAST report for subsequences (percentage of identities on matches, subsequence and alignment lengths), build a matrix that reflects the distance between each pair of sequences in the original dataset (full methodological details are provided in the Methods). The information contained in the distance matrix is used to generate a phylogenetic tree in which branch lengths are derived from the average pairwise similarity of subsequence alignments.

Upon analyzing the biosynthetic gene clusters using this LPSA-based approach, we obtained a phylogenetic tree describing the sequence similarity relationships among the 76 gene clusters (Figure 1). The phylogenetic tree of biosynthetic gene clusters branches first according to the bioactivity of the secondary metabolite, which is given by the molecular target of the cyanobacterial metabolite in another organism. Specifically, gene clusters producing ion channel modulators group separately from those producing phosphatase inhibitors and from those producing protease inhibitors. Furthermore, gene clusters producing neurotoxins targeting voltage-gated ion channels (saxitoxins) group separately from gene clusters producing neurotoxins targeting ligand-gated ion channels (anatoxins). These results confirm that the biosynthetic genes themselves already contain key information regarding the impact of the secondary metabolites in the environment (i.e., bioactivity). Furthermore, these findings suggest that the bioactivity of the secondary metabolites may be the primary driving force behind the diversity of the genetic machinery for secondary metabolism in cyanobacteria.

Bioactivity, rather than similarity of chemical structure, is the main explanation behind genetic diversity in the secondary metabolism of cyanobacteria

The next levels of branching of the phylogenetic tree correspond to the chemical class and chemical structure of the secondary metabolite, which are dictated by the biosynthetic enzymes and the order of biosynthetic reactions. Specifically, gene clusters producing cyclic peptides generally group separately from those producing alkaloids. Furthermore, among gene clusters producing indole-based alkaloids, there is distinct grouping according to whether the bioactive metabolite contains a nitrile group or an isonitrile group (NitFischerindole vs. Fischerindole in Figure 1). However, gene clusters producing anacyclamide and microcyclamide group with clusters producing indoles and not with clusters producing other modified cyclic peptides. Furthermore, it should be noted that most secondary metabolites produced via NRPS/PKS are modular compounds, and it is currently recognized that variability in one or more substituents leads to different target specificity. For example, the affinity for genetic variants of the same ion channel differs across saxitoxin analogs³¹, whereas the specificity for serine proteases with trypsin- and thrombin-like activity differs among aeruginosins³². Thus, the grouping of biosynthetic gene clusters according to the chemical structure of the secondary metabolite likely reflects the bioactivity of the metabolite rather than its specific chemical structure.

Indeed, gene clusters producing nodularins grouped closely together with those producing microcystins. While both nodularins and microcystins are cyclic peptides, nodularins typically contain five amino acid residues, whereas microcystins contain seven; however, the biological activity of nodularins and microcystins is equivalent, as these metabolites have the same molecular targets, the same binding site on their molecular target, and almost the same

interactions involved in binding with the molecular target^{16,33}. Furthermore, despite the fact that aeruginosins are linear and contain an indole part in addition to the oligopeptide chain, gene clusters producing aeruginosin grouped together with those producing cyclic peptides that share the same molecular target in other organisms (i.e., serine proteases)³². These findings suggest that genetic diversity in the secondary metabolism of cyanobacteria is explained mainly in terms of bioactivity rather than in terms of chemical structure similarity.

Non-uniform horizontal gene transfer of biosynthetic gene clusters of the secondary metabolism

The next level of branching of the phylogenetic tree corresponds to the genus/species and geographical area of origin. For example, saxitoxin-producing gene clusters originating from *Aphanizomenon* species group separately from clusters originating from other genera; similar findings are noted for anabaenopeptin-producing clusters originating from *Plankthotrix* species, as well as for cyanopeptolin-producing clusters originating from *Plankthotrix* species. Other relationships are less well defined. Cylindrospermopsin-producing cluster 7, which comes from a Chinese *C. raciborskii* strain, groups closer to clusters 4 and 5, which come from Chinese *Raphidiopsis curvata* strains, than with cluster 2, which comes from a different Chinese *C. raciborskii* strain and groups closer to cluster 6, which comes from an Australian *C. raciborskii* strain. The anabaenopeptin-producing cluster 8, which originates from a Finnish *Anabaena* species, groups separately from clusters 1, 3, and 12, which come from Finnish *Plankthotrix* species; on the other hand, cluster 3 groups close to cluster 12, as they both come from the same Finnish *Plankthotrix aghardii* strain, while cluster 1 from another Finnish *P. aghardii* strain groups closer to clusters from non-Finnish *Plankthotrix* species. These findings confirm that some

biosynthetic gene clusters may transfer more readily than others across species or strains, likely aided by geographical proximity or contamination events, while other signs of recombination and gene duplication are likely related to vertical inheritance¹².

Discussion

Survival strategies based on secondary metabolism

Based on the organization of the phylogenetic tree of biosynthetic gene clusters, three main strategies with potential survival benefit can be distinguished, namely transmembrane (direct) toxicity due to ion channel blockade, intracellular (indirect) toxicity due to protein phosphatase blockade, and digestion inhibition due to serine protease blockade (Figure 2). Various branches of secondary metabolism contributing to each strategy likely provide back-up or some form of specialization to facilitate the strategy. Direct and indirect toxicity may act separately or in tandem, with the effects of indirect toxicity taking longer to manifest, requiring facilitation by a membrane permeabilization agent, and being mainly aimed at organisms of the animal kingdom. Digestion inhibition would serve as a bail-out strategy in case the cell is ingested, not particularly aiming to kill the culprit organism; in other words, serine protease inhibition by cyanobacterial toxins likely arose as a purely protective mechanism, which is fundamentally different from the role of serine protease inhibitors produced by animals (e.g., ticks)³⁴. Dominant cyanobacteria such as *Planktothrix*, *Anabaena*, and *Microcystis* have the genetic machinery to support all three strategies.

Figure 2. Evolutionary advantage of toxin production by cyanobacteria. Three survival strategies can be distinguished. Dominant cyanobacteria such as *Planktothrix*, *Anabaena*, and *Microcystis* have the genetic machinery to support all three strategies.

Direct toxicity can be achieved by modulating the activity of voltage- and ligand-gated ion channels, which are ubiquitous across all life domains. Such mechanisms, which often involve alkaloid metabolites with relatively broad specificity but differential affinity, can be used effectively against competitors and against phagocytosis, infection, or predation by other organisms³⁵⁻³⁷. Our findings support the role of the neurotoxin saxitoxin as a modulator of voltage-gated ion channels² and further suggest that indole-based alkaloids and perhaps cyclamides may share such targets. The indoles welwitindolinone and fischerindole have rarely been studied. Nevertheless, related indole-based alkaloids produced by the genus *Fischerella* are suspected to act as ion channel modulators similar to saxitoxin³⁸. Furthermore, the indole moiety itself is a known modulator of voltage-dependent gating of potassium channels³⁹. Cyclamides are cyclic oligopeptides produced via the ribosomal pathway and heavily functionalized via post-translational modifications⁴⁰. While anacyclamide and microcyclamide have not been studied for their interaction with ion channels, the thiazoline moiety commonly encountered in such

cyclamides has been reported to be critical for the toxicity of kalkitoxin⁴¹, another cyanotoxin that targets voltage-gated ion channels⁴². At the same time, our findings support the role of the neurotoxin anatoxin as a modulator of ligand-gated ion channels², and further suggest that cylindrospermopsin, curacyn, nocuolin, and perhaps barbamide may share such targets. While cylindrospermopsin is typically referred to as a hepatotoxin, its neurotoxicity has been confirmed in fish, with the nicotinic acetylcholine receptor as the suspected target (i.e., the same target as that of anatoxin), though binding seems reversible⁴³. The interaction of curacin A with ligand-gated ion channels has not been tested to date. Curacin A-like compounds produced by the same species (tropical *Lyngbia majuscula*) were shown to induce neurotoxicity in rat neurons via interaction with ligand-gated ion channels, though curacin A itself did not have such effect in rat neurons⁴⁴. As curacin A is toxic to shrimp³⁶, the difference in toxicity (shrimp vs. rat neurons) might be related to differences in specificity towards the target molecule. Finally, nocuolin A and barbamide have not been studied for potential neurotoxicity or interaction with ion channels. Nocuolin A is a natural oxadiazine, a class of compounds shown to block sodium channels and exert neurotoxic effects in cockroaches⁴⁵. Barbamide is a strong molluscicide, though its mechanism of action has yet to be determined. Given that barbamide contains thiazole and trichloromethyl, two pharmacophores known to interact with ion channels^{45,46}, it is likely that barbamide also targets ion channels.

Indirect toxicity can be achieved by modulating intracellular signaling. Signaling processes involved in growth and reproduction rely on recognition of specific protein phosphorylation patterns created by the interplay of protein kinases and protein phosphatases. Microcystins/nodularins, which block protein phosphatases, interfere with the cell cycle leading to abnormal mitosis and even cell death. In addition to actively deterring predation^{35,47},

microcystins indirectly support dominance of the producing cyanobacterium by promoting preferential predation on competing organisms⁴⁸. It has been suggested that nodularin synthetase genes are derived from microcystin synthetase genes⁸. While the pentapeptide nodularin and the heptapeptide microcystin are similar in structure and bioactivity, neither has a strong effect on its own. Instead, anabaenolysin may potentiate the action of microcystins/nodularins by permeating cholesterol-based membranes⁴⁹. In microcystin/nodularin producers, biosynthetic clusters for anabaenolysin may have evolved as an adaptation to the evolution of animal organisms feeding on the symbiote. A single cyanobacterial strain may produce many anabaenolysin variants⁵⁰, probably as an adaptation to the diversification of marine life.

Digestion inhibition can be achieved by modulating the activity of proteases, which protects ingested cells from being degraded. As such digestive enzymes have highly efficient catalytic sites, inhibition requires high specificity, which is provided by peptide-based metabolites. Aeruginosin, anabaenopeptin, and cyanopeptolin are serine protease inhibitors, typically targeting the digestive enzymes trypsin and/or chymotrypsin, as well as similar proteases involved in the coagulation pathway. The same cyanobacterium may possess the genetic machinery to synthesize more than one type of serine protease inhibitors and there is substantial though not complete overlap in the specificities of aeruginosin, anabaenopeptin, and cyanopeptolin^{6,32}, suggesting that these biosynthetic gene clusters may provide mutual back-up for inhibiting digestion in high-level organisms predating/grazing on cyanobacterial symbiotic hosts. In this context, interference with the coagulation pathway may not be evolutionarily motivated in cyanobacteria, but rather a consequence of the similarity between digestive proteins and coagulation pathway proteins. At the same time, the evolution of genetic machineries to

produce serine-protease inhibitors in the form of cyanobacterial metabolites may have contributed to the diversification of tryptic enzymes in the animal kingdom.

Implications for the study of cyanobacterial evolution

While it is agreed that microcystin- and saxitoxin-producing gene clusters are likely the oldest toxin-producing gene clusters in cyanobacteria, it is believed that the ability to produce microcystin and saxitoxin has been lost repeatedly across multiple lineages of cyanobacteria and even among strains of the same species, suggesting that there may have been neither negative nor positive pressure for such biosynthetic genes, or that such pressure was circumstantial to some clades^{6,8,22}. However, others have pointed out that the genomes of South American saxitoxin-producing *C. raciborski* strains contain vestiges of cylindrospermopsin-producing genes, whereas no vestiges of saxitoxin-producing genes are seen in the genomes of the cylindrospermopsin-producing strains⁵¹. Our findings suggest that saxitoxin/cylindrospermopsin and microcystin/nodularin serve parallel survival strategies (transmembrane vs. intracellular toxicity, respectively; Figure 1). Due to its simplicity, accessibility, and fast acting mechanism, transmembrane toxicity by modulation of ion channels likely represents an older mechanism⁵².

Limitations

The phylogenetic analysis approach applied in the present study is conceptually similar to previous efforts employing average sequence similarity²⁴⁻²⁸, in that it relies on obtaining LPSA similarity scores for certain pieces of the genome, averaging the score for each full sequence analyzed, and creating a distance matrix based on the average similarity scores. However, unlike other LPSA-based phylogenetic approaches²⁸, the method used here does not employ information regarding evolutionary timelines. For this reason, the branch lengths of the

phylogenetic tree obtained in this study are not directly representative of evolutionary time but of sequence similarity, for which the evolutionary rate may vary throughout geological time and environment. On the other hand, cyanobacteria evolution would have been heavily influenced by their symbiotic relationships, which may preclude the application of current assumptions regarding evolutionary rate^{53,54}. Furthermore, we could not infer ancestral sequences because this LPSA-based approach obtains distance matrices based solely on average similarity scores, which inevitably results in partial loss of alignment information. Finally, while many cyanobacterial metabolites have been characterized² and many sequences potentially corresponding to NRPS/PKS enzymes have been detected⁵⁵, the coding sequences currently available for other such metabolites did not fit the inclusion criteria for this study. As more data accumulate, it will become possible to perform an all-encompassing phylogenetic analysis of cyanobacterial biosynthetic gene clusters and shed further light into the potential biological, ecological, and evolutionary role of cyanotoxins.

Scope for further research

The high level of organization of the phylogenetic tree of biosynthetic gene clusters illustrates the depth of information that can be obtained using LPSA combined with a simple method of overall sequence similarity. It is notable that we were able to obtain such information without having to evaluate non-coding regions, RNA/protein expression levels, or metabolomics profiles. This suggests that, when attempting to understand the relationships among complex phenotypes in prokaryotes, it may not be necessary to build one tree per gene or even one tree per gene cluster. We propose that the “sequence-structure-function” paradigm, typically discussed in the context of proteins, can be expanded from the nucleotide sequence of the gene cluster to the bioactivity of the metabolite synthesized by the enzyme ensemble encoded by the gene cluster.

Thus, our findings can also be interpreted in the context of the “It’s the song not the singer” (ITSNTS) theory, which proposes that the true units of evolutionary selection are not the organisms themselves (i.e., the singers in the ITSNTS theory) but the processes implemented by genes, cells, species, or communities (i.e., the song in the ITSNTS theory)⁵⁶. In this context, we may speak about the bioactivity of the biosynthetic gene cluster (i.e., the song in the ITSNTS theory) as an independent entity of a biosystem, driving the evolution of organisms or communities carrying such biosynthetic genes or interacting with the metabolites synthesized by the genes (i.e., the singers in the ITSNTS theory). Nevertheless, looking only at the gene cluster sequence is not sufficient to clarify whether and how much of a given toxin will be produced under different environmental conditions. While it is generally accepted that recent cyanobacterial invasion, proliferation, and ecological impact are strongly affected by climate change and environmental changes caused by human activities^{57,58}, it remains unclear what triggers abundant production of toxic metabolites by otherwise silent gene clusters, though it is quite evident that such triggers are multi-faceted². Further study is warranted to clarify whether or how secondary metabolites could enhance resource utilization or temperature and chemical resistance in cyanobacteria. Alternatively, since the feeding behavior of grazers and predators is sensitive to environmental changes, the fact that toxin production in cyanobacteria is triggered by environmental changes may simply reflect the association between such changes and feeding behavior of predators. Finally, the survival strategies proposed in this study may not have been the original ecological roles of cyanotoxins, as these early roles would have been lost or replaced with new functions over evolutionary time, reflecting the change in local and planet-wide ecological context. It is important to keep in mind that cyanobacteria are often encountered as endosymbionts of corals, diatoms, dinoflagellates, seagrass, and sponges. Thus, survival benefits

conferred by bioactive secondary metabolites may apply to the cyanobacterium itself or to its symbiotic host. For example, while there may be negative selective pressure for microcystin-producing gene clusters within the Cyanobacteria phylum⁸, such genes may provide benefit to the symbiotic host⁵⁹. Production of toxic metabolites by cyanobacteria may have promoted the adaptation of some animals to consuming a diet contaminated with toxin-producing cyanobacteria and even carrying the cyanobacteria as a symbiont; for example, all eight sodium channel variants in the pufferfish genome harbor mutations rendering them resistant to saxitoxin⁶⁰. The impact of the survival strategies discussed here should therefore be considered in the context of these symbiotic relationships. Furthermore, future genomic studies on cyanobacteria should take into account the symbiotic host, as gene transfer potential may be substantially altered for cyanobacteria living as symbionts⁵⁴.

Conclusion

The bioactivity of a toxic metabolite can be deduced directly from the nucleotide sequence of the biosynthetic gene cluster, without needing to examine the enzymes themselves or to measure expression levels. Bioactivity may have been the primary driving force behind the diversity of secondary metabolism in cyanobacteria. This genetic machinery evolved to facilitate three specific survival strategies acting separately or in tandem, with dominant cyanobacteria possessing the genetic machinery to support all three strategies. Transmembrane (direct) toxicity targeting ion channels, intracellular (indirect) toxicity targeting cell-cycle regulation, and digestion inhibition targeting proteases may have provided the survival advantage underpinning the evolutionary success of both cyanobacteria and their early symbiotic hosts.

Methods

Description of the LPSA-based approach for phylogenetic analysis

a. Definitions

S : A sequence to align

L : The length of a given sequence

s : A subsequence of S

σ_i : The subsequence of S_i with the highest similarity to a certain subsequence s

l : Length of a subsequence s

f : Frame shift from position 1 of a sequence S

n : Number of frames f within a subsequence s

$$n = \lfloor l/f \rfloor$$

N : Number of subsequences in S , including the subsequences from different frame shifts f

$$N = \lfloor L/l \rfloor + \lfloor (L-f)/l \rfloor + \lfloor (L-2f)/l \rfloor + \dots + \lfloor (L-(n-1)*f)/l \rfloor$$

N_0 : Number of subsequences in S_i for which no subsequence in S_j was found during the alignment

M : The distance matrix for all sequences S

b. Description

The approach involves three steps described schematically below for a dataset of three sequences.

- (i) Break each sequence S_i (for 3 subsequences: $i, j \in \{a, b, c\}$) into n_i subsequences of a chosen length l , equal for all i
- (ii) Perform pairwise local alignment (in this case, using BLAST) of all subsequences s_i vs all subsequences s_j , $i \neq j$
- (iii) Calculate a distance matrix M (for 3 subsequences, M is a 3×3 matrix) based on the BLAST report for all subsequence alignments. Each element M_{ij} of M is calculated as follows:

for $i = j$

$$\text{Distance}(S_i, S_j) = 0$$

for $i \neq j$

$\text{Distance}(S_i, S_j)$

$$= 1 - \frac{\sum_{k=1}^{N_i} \left(\max(\%identity(s_{i_k}, \sigma_j), \%identity(\sigma_j, s_{i_k})) \times \text{alignment_length}(s_{i_k}, \sigma_j) \right)}{\sum_{m=1}^{N_i - N_0} (\text{alignment_length}(s_{i_m}, \sigma_j))}$$

- c. The final result is a distance matrix containing the metric described in item b.

Various methodologies can then be used to derive a phylogenetic tree based on this matrix (e.g., neighbor joining, weighted least squares, etc.). The distance matrix can be represented as follows:

$$\begin{pmatrix} 0 & \text{Distance}(S_a, S_b) & \text{Distance}(S_a, S_c) \\ \text{Distance}(S_b, S_a) & 0 & \text{Distance}(S_b, S_c) \\ \text{Distance}(S_c, S_a) & \text{Distance}(S_c, S_b) & 0 \end{pmatrix}$$

445

446 *Dataset of cyanobacterial NRPS/PKS gene clusters*

447 The sequences of toxin-producing gene clusters were obtained from the Nucleotide database of
 448 NCBI (queried in February 2018) using the search terms ((NRPS) OR (PKS) OR (gene_cluster)
 449 OR (polyketide_synthase) OR (nonribosomal_peptide_synthase)) AND
 450 “cyanobacteria”[porgn: __txid1117]. Subsequently, we filtered out all sequences shorter than
 451 1,500 bp, and manually removed all entries whose annotation did not refer to NRPS or PKS
 452 proteins (contig, scaffold, whole genome, plasmid, etc.), entries containing patent sequences, and
 453 entries that did not specify the final product of the gene cluster. To ensure that all relevant entries
 454 were identified, we performed additional searches using the identified gene cluster products as
 455 keywords (e.g., microcystin, cylindrospermopsin), and applied the same filtering protocol. For
 456 each hit, we downloaded a FASTA file containing the coding sequences (CDS), and we labeled
 457 each file with its corresponding GenBank accession code. Each file was downloaded and
 458 inspected separately. If a file contained also sequences annotated to code for proteins unrelated
 459 to the gene cluster product, we removed these particular CDS. If a file contained several regions
 460 annotated as different clusters (i.e., resulting in different products), we moved the CDS for each
 461 different cluster to a separate file. Finally, we retained only files that contained at least three
 462 genes. In these files, we concatenated the remaining sequences by removing non-sequence
 463 information (FASTA headers), without changing the order of genes in the original file. The final
 464 analysis included only toxins for which at least two biosynthetic gene clusters satisfied the
 465 above-listed conditions (19 toxins, 76 biosynthetic gene clusters). Thus, the phylogenetic
 466 analysis was performed on a dataset of 76 files, each pertaining to a single biosynthetic cluster
 467 and being referred to using a single identifier (e.g., “Microcystin_1”) (Figure 3). Prior to

sequence alignment, ambiguous nucleotides were replaced as follows: *n* and *v* were replaced with *a*; *s* and *b* were replaced with *c*; *r*, *m*, and *d* were replaced with *g*; *k*, *w*, *h*, and *u* were replaced with *t*. The dataset (nucleotide sequences in FASTA format) is available as Supplementary Data S1, whereas the phylogenetic tree (in Newick format) is available as Supplementary Data S2.

Software and hardware used

The datasets were obtained and processed manually. The phylogenetic analysis was conducted using CrocoPhylogeny version 0.0.18.6.30 with default settings⁶¹, running NCBI BLAST version 2.6³⁰. The phylogenetic tree was built using neighbor joining as implemented in SplitsTree4⁶², with default parameters, starting from the distance matrix. All calculations were performed on a computer with the following specifications: Intel Core i7-5820K CPU (6 cores, 12 threads); 32.0 GB RAM (DDR4); ext4 file system on SSD; Ubuntu 16.4. The phylogenetic tree was visualized using iTOL v3⁶³.

Figure 3. Overview of methods. Full information about the 76 gene clusters included in the analysis is provided in Table S1. The dataset (nucleotide sequences in FASTA format) is available as Supplementary Data S1, whereas the phylogenetic tree (in Newick format) is available as Supplementary Data S2. CDS, coding sequences; PKS, polyketide synthesis; NJ, neighbor joining; NRPS, non-ribosomal peptide synthesis.

488

489 **Data availability**

490 All data generated or analyzed during this study are included in this published article and its
491 supplementary information files.

492

493 **Acknowledgements**

494 The authors thank Dr. Luísa Hoffmann, Mr. Taraka Ramji Moturu, and Mrs. Michelle Almeida
495 da Paz for helpful comments and useful discussions.

496

497 **Author contributions**

498 RJTR, CMI, and JK designed the study. RJTR and CMI obtained the source data and analyzed
499 the results. RJTR wrote the code and ran the calculations. CMI wrote the manuscript. All authors
500 contributed substantially to discussion of the content and reviewed and edited the manuscript
501 before submission.

502

503 **Additional information**

504 ***Competing interests***

505 The authors declare no competing interests.

506 ***Funding***

This research was carried out under the project CEITEC 2020 (LQ1601) with financial support from the Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme II. The funder had no role in the design of the study, in data collection, analysis, or interpretation, in writing of the manuscript, or in the decision to publish.

Ethics approval and consent to participate

Not applicable, as no human or animal studies were conducted. All data for analysis were obtained from publicly available databases.

Consent for publication

Not applicable.

Supplementary information legends

Supplementary Table S1. Overview of biosynthetic gene clusters used in this study. Table containing cyanotoxin name, GenBank ID, organism, length, and list of genes in each cluster.

Supplementary Data S1. Nucleotide sequences of the biosynthetic gene clusters. The FASTA sequences of all biosynthetic gene clusters are provided in a single file for easy download.

Supplementary Data S2. Phylogenetic tree of biosynthetic gene clusters. Phylogenetic tree (Newick format) in which branch lengths are derived from the average pairwise similarity of subsequence alignments.

527

528 **References**

529 1. Ponce-Toledo, R. I. *et al.* An early-branching freshwater cyanobacterium at the origin of plastids.

530 *Curr. Biol.* **27**, 386-391 (2017).

531 2. Huisman, J. *et al.* Cyanobacterial blooms. *Nat. Rev. Microbiol.* **16**, 471-483 (2018).

532 3. Leao, T. *et al.* Comparative genomics uncovers the prolific and distinctive metabolic potential of
533 the cyanobacterial genus *Moorea*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3198-3203 (2017).

534 4. Leao, P. N., Engene, N., Antunes, A., Gerwick, W. H. & Vasconcelos, V. The chemical ecology
535 of cyanobacteria. *Nat. Prod. Rep.* **29**, 372-391 (2012).

536 5. Mazard, S., Penesyan, A., Ostrowski, M., Paulsen, I. T. & Egan, S. Tiny microbes with a big
537 impact: the role of cyanobacteria and their metabolites in shaping our future. *Mar. Drugs* **14**, 97;
538 10.3390/md14050097 (2016).

539 6. Kurmayer, R., Deng, L. & Entfellner, E. Role of toxic and bioactive secondary metabolites in
540 colonization and bloom formation by filamentous cyanobacteria *Planktothrix*. *Harmful Algae* **54**,
541 69-86 (2016).

542 7. Lemaire, V. *et al.* Genotype $\times$ genotype interactions between the toxic cyanobacterium
543 *Microcystis* and its grazer, the waterflea *Daphnia*. *Evol. Appl.* **5**, 168-182 (2012).

544 8. Rantala, A. *et al.* Phylogenetic evidence for the early evolution of microcystin synthesis. *Proc.*
545 *Natl. Acad. Sci. U.S.A.* **101**, 568-573 (2004).

- 546 9. Weber, T. & Kim, H. U. The secondary metabolite bioinformatics portal: computational tools to
547 facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.* **1**, 69-79
548 (2016).
- 549 10. Khater, S., Anand, S. & Mohanty, D. In silico methods for linking genes and secondary
550 metabolites: the way forward. *Synth. Syst. Biotechnol.* **1**, 80-88 (2016).
- 551 11. Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T.
552 Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer
553 events. *Genome Res.* **16**, 1099-1108 (2006).
- 554 12. Calteau, A. *et al.* Phylum-wide comparative genomics unravel the diversity of secondary
555 metabolism in *Cyanobacteria*. *BMC Genomics* **15**, 977; 10.1186/1471-2164-15-977 (2014).
- 556 13. Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven
557 genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1053-1058 (2013).
- 558 14. Walter, J. M. *et al.* Ecogenomics and taxonomy of *Cyanobacteria* phylum. *Front. Microbiol.* **8**,
559 2132; 10.3389/fmicb.2017.02132 (2017).
- 560 15. Beck, C., Knoop, H. & Steuer, R. Modules of co-occurrence in the cyanobacterial pan-genome
561 reveal functional associations between groups of ortholog genes. *PLoS Genet.* **14**, e1007239;
562 10.1371/journal.pgen.1007239 (2018).
- 563 16. Pearson, L., Mihali, T., Moffitt, M., Kellmann, R. & Neilan, B. On the chemistry, toxicology
564 and genetics of the cyanobacterial toxins, microcystin, nodularin, saxitoxin and
565 cylindrospermopsin. *Mar. Drugs* **8**, 1650-1680 (2010).

- 566 17. Dittmann, E., Fewer, D. P. & Neilan, B. A. Cyanobacterial toxins: biosynthetic routes and
567 evolutionary roots. *FEMS Microbiol. Rev.* **37**, 23-43 (2013).
- 568 18. Cirés, S. *et al.* Phylogeography of cylindrospermopsin and paralytic shellfish toxin-producing
569 *Nostocales* cyanobacteria from Mediterranean Europe (Spain). *Appl. Environ. Microbiol.* **80**,
570 1359-1370 (2014).
- 571 19. Gugger, M. *et al.* Genetic diversity of *Cylindrospermopsis* strains (cyanobacteria) isolated from
572 four continents. *Appl. Environ. Microbiol.* **71**, 1097-1100 (2005).
- 573 20. Stucken, K. *et al.* Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis*
574 *raciborskii* strains. *Syst. Appl. Microbiol.* **32**, 37-48 (2009).
- 575 21. Jungblut, A. D. & Neilan, B. A. Molecular identification and evolution of the cyclic peptide
576 hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. *Arch.*
577 *Microbiol.* **185**, 107-114 (2006).
- 578 22. Moustafa, A. *et al.* Origin of saxitoxin biosynthetic genes in cyanobacteria. *PLoS ONE* **4**,
579 e5758; 10.1371/journal.pone.0005758 (2009).
- 580 23. Moreira, C., Vasconcelos, V. & Antunes, A. Phylogeny and biogeography of cyanobacteria and
581 their produced toxins. *Mar. Drugs* **11**, 4350-4369 (2013).
- 582 24. Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K. & Schuster, S. C. Whole-genome
583 prokaryotic phylogeny. *Bioinformatics* **21**, 2329-2335 (2005).
- 584 25. Satoh, S., Mimuro, M. & Tanaka, A. Construction of a phylogenetic tree of photosynthetic
585 prokaryotes based on average similarities of whole genome sequences. *PLoS ONE* **8**, e70290;
586 10.1371/journal.pone.0070290 (2013).

- 587 26. Xia, X. PhyPA: phylogenetic method with pairwise sequence alignment outperforms likelihood
588 methods in phylogenetics involving highly diverged sequences. *Mol. Phylogenet. Evol.* **102**, 331-
589 343 (2016).
- 590 27. Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of
591 prokaryotic viruses. *Bioinformatics* **33**, 3396-3404 (2017).
- 592 28. Yokono, M., Satoh, S. & Tanaka, A. Comparative analyses of whole-genome protein sequences
593 from multiple organisms. *Sci. Rep.* **8**, 6800; 10.1038/s41598-018-25090-8 (2018).
- 594 29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
595 search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
- 596 30. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421;
597 10.1186/1471-2105-10-421 (2009).
- 598 31. Thomas-Tran, R. & Du Bois, J. Mutant cycle analysis with modified saxitoxins reveals specific
599 interactions critical to attaining high-affinity inhibition of hNaV1.7. *Proc. Natl. Acad. Sci. U.S.A.*
600 **113**, 5856-5861 (2016).
- 601 32. Ersmark, K., Del Valle, J. & Hanessian, S. Chemistry and biology of the aeruginosin family of
602 serine protease inhibitors. *Angew. Chem. Int. Ed. Engl.* **47**, 1202-1223 (2008).
- 603 33. Kelker, M. S., Page, R. & Peti, W. Crystal structures of protein phosphatase-1 bound to
604 nodularin-R and tautomycin: a novel scaffold for structure-based drug design of serine/threonine
605 phosphatase inhibitors. *J. Mol. Biol.* **385**, 11-21 (2009).

- 606 34. Blisnick, A. A., Foulon, T. & Bonnet, S. I. Serine protease inhibitors in ticks: an overview of
607 their role in tick biology and tick-borne pathogen transmission. *Front. Cell. Infect. Microbiol.* **7**,
608 199; 10.3389/fcimb.2017.00199 (2017).
- 609 35. Zanchett, G. & Oliveira-Filho, E. C. Cyanobacteria and cyanotoxins: from impacts on aquatic
610 ecosystems and human health to anticarcinogenic effects. *Toxins* **5**, 1896-1917 (2013).
- 611 36. Essack, M., Alzubaidy, H. S., Bajic, V. B. & Archer, J. A. C. Chemical compounds toxic to
612 invertebrates isolated from marine cyanobacteria of potential relevance to the agricultural
613 industry. *Toxins* **6**, 3058-3076 (2014).
- 614 37. Rzymiski, P. & Poniedziałek, B. In search of environmental role of cylindrospermopsin: a
615 review on global distribution and ecology of its producers. *Water Res.* **66**, 320-337 (2014).
- 616 38. Cagide, E. *et al.* Hapalindoles from the cyanobacterium *Fischerella*: potential sodium channel
617 modulators. *Chem. Res. Toxicol.* **27**, 1696-1706 (2014).
- 618 39. Avdonin, V. & Hoshi, T. Modification of voltage-dependent gating of potassium channels by
619 free form of tryptophan side chain. *Biophys. J.* **81**, 97-106 (2001).
- 620 40. Martins, J. & Vasconcelos, V. Cyanobactins from cyanobacteria: current genetic and chemical
621 state of knowledge. *Mar. Drugs* **13**, 6910-6946 (2015).
- 622 41. White, J. D., Xu, Q., Lee, C. S. & Valeriote, F. A. Total synthesis and biological evaluation of
623 +-kalkitoxin, a cytotoxic metabolite of the cyanobacterium *Lyngbya majuscula*. *Org. Biomol.*
624 *Chem.* **2**, 2092-2102 (2004).
- 625 42. LePage, K. T. *et al.* The neurotoxic lipopeptide kalkitoxin interacts with voltage-sensitive
626 sodium channels in cerebellar granule neurons. *Toxicol. Lett.* **158**, 133-139 (2005).

- 627 43. Guzmán-Guillén, R. *et al.* Cylindrospermopsin induces neurotoxicity in tilapia fish
628 (*Oreochromis niloticus*) exposed to *Aphanizomenon ovalisporum*. *Aquat. Toxicol.* **161**, 17-24
629 (2015).
- 630 44. Berman, F. W., Gerwick, W. H. & Murray, T. F. Antillatoxin and kalkitoxin, ichthyotoxins from
631 the tropical cyanobacterium *Lyngbya majuscula*, induce distinct temporal patterns of NMDA
632 receptor-mediated neurotoxicity. *Toxicon* **37**, 1645-1648 (1999).
- 633 45. Djemaoun, A., Habes, D. & Soltani, N. Effects of ingested indoxacarb (oxadiazine) on
634 biochemical composition of ovaries in *Blattella germanica* (Dictyoptera, Blattellidae). *J.*
635 *Entomol. Zool. Stud.* **3**, 362-366 (2015).
- 636 46. Zhorov, B. S. & Dong, K. Elucidation of pyrethroid and DDT receptor sites in the voltage-gated
637 sodium channel. *Neurotoxicology* **60**, 171-177 (2017).
- 638 47. Jang, M. H., Jung, J. M. & Takamura, N. Changes in microcystin production in cyanobacteria
639 exposed to zooplankton at different population densities and infochemical concentrations.
640 *Limnol. Oceanogr.* **52**, 1454-1466 (2007).
- 641 48. Leitão, E., Ger, K. A. & Panosso, R. Selective grazing by a tropical copepod (*Notodiaptomus*
642 *iheringi*) facilitates *Microcystis* dominance. *Front. Microbiol.* **9**, 301; 10.3389/fmicb.2018.00301
643 (2018).
- 644 49. Oftedal, L. *et al.* The lipopeptide toxins anabaenolysin A and B target biological membranes in
645 a cholesterol-dependent manner. *Biochim. Biophys. Acta* **1818**, 3000-3009 (2012).
- 646 50. Jokela, J. *et al.* Anabaenolysins, novel cytolytic lipopeptides from benthic *Anabaena*
647 cyanobacteria. *PLoS ONE* **7**, e41222; 10.1371/journal.pone.0041222 (2012).

- 648 51. Hoff-Risseti, C. *et al.* Cyldindropermopsin and saxitoxin synthetase genes in
649 *Cylindropermopsis raciborskii* strains from Brazilian freshwater. *PLoS ONE* **8**, e74238;
650 10.1371/journal.pone.0074238 (2013).
- 651 52. Murray, S. A., Mihali, T. K. & Neilan, B. A. Extraordinary conservation, gene loss, and positive
652 selection in the evolution of an ancient neurotoxin. *Mol. Biol. Evol.* **28**, 1173-1182 (2011).
- 653 53. Itoh, T., Martin, W. & Nei, M. Acceleration of genomic evolution caused by enhanced mutation
654 rate in endocellular symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12944-12948 (2002).
- 655 54. Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev.*
656 *Microbiol.* **8**, 218-230 (2010).
- 657 55. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide
658 and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes.
659 *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9259-9264 (2014).
- 660 56. Doolittle, W. F. & Inkpen, S. A. Processes and patterns of interaction as units of selection: An
661 introduction to ITSNTS thinking. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4006-4014 (2018).
- 662 57. Hansson, L.-A. *et al.* Food-chain length alters community responses to global change in aquatic
663 systems. *Nat. Clim. Change* **3**, 228-233 (2013).
- 664 58. Wells, M. L. *et al.* Harmful algal blooms and climate change: learning from the past and present
665 to forecast the future. *Harmful Algae* **49**, 68-93 (2015).
- 666 59. Kaasalainen, U. *et al.* Cyanobacteria produce a high variety of hepatotoxic peptides in lichen
667 symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5886-5891 (2012).

- 668 60. Jost, M. C. *et al.* Toxin-resistant sodium channels: parallel adaptive evolution across a complete
669 gene family. *Mol. Biol. Evol.* **25**, 1016-1024 (2008).
- 670 61. Tristão Ramos, R. J., Almeida da Paz, M., Ionescu, C. M. & Koča, J. CrocoPhylogeny: web-
671 based phylogenetic analysis of whole genomes and complex sequences in just three clicks.
672 <https://crocodile.ncbr.muni.cz/> (2018).
- 673 62. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol.*
674 *Biol. Evol.* **23**, 254-267 (2006).
- 675 63. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
676 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-W245 (2016).
- 677
- 678
- 679