



## OPEN ACCESS

## EDITED BY

Richard Allen White III,  
University of North Carolina at Charlotte,  
United States

## REVIEWED BY

Paraskevi Mara,  
Woods Hole Oceanographic Institution,  
United States  
Mercedes Nieves Morión,  
Spanish National Research Council (CSIC),  
Spain

## \*CORRESPONDENCE

Ruth Pérez Gallego  
✉ ruth.perez.gallego@nioz.nl

RECEIVED 11 July 2023

ACCEPTED 05 September 2023

PUBLISHED 27 September 2023

## CITATION

Pérez Gallego R, Bale NJ, Sinninghe  
Damste JS and Villanueva L (2023) Developing  
a genetic approach to target cyanobacterial  
producers of heterocyte glycolipids in the  
environment.

*Front. Microbiol.* 14:1257040.  
doi: 10.3389/fmicb.2023.1257040

## COPYRIGHT

© 2023 Pérez Gallego, Bale, Sinninghe Damste  
and Villanueva. This is an open-access article  
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Developing a genetic approach to target cyanobacterial producers of heterocyte glycolipids in the environment

Ruth Pérez Gallego<sup>4\*</sup>, Nicole J. Bale<sup>1</sup>, Jaap S. Sinninghe Damste<sup>1,2</sup>  
and Laura Villanueva<sup>1,2</sup>

<sup>1</sup>Department of Marine Microbiology and Biogeochemistry (MMB), NIOZ Royal Netherlands Institute for Sea Research, Den Burg, Netherlands, <sup>2</sup>Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands

Heterocytous cyanobacteria are important players in the carbon and nitrogen cycle. They can fix dinitrogen by using heterocytes, specialized cells containing the oxygen-sensitive nitrogenase enzyme surrounded by a thick polysaccharide and glycolipid layer which prevents oxygen diffusion and nitrogenase inactivation. Heterocyte glycolipids can be used to detect the presence of heterocytous cyanobacteria in present-day and past environments, providing insight into the functioning of the studied ecosystems. However, due to their good preservation throughout time, heterocyte glycolipids are not ideal to detect and study living communities, instead methods based on DNA are preferred. Currently cyanobacteria can be detected using untargeted genomic approaches such as metagenomics, or they can be specifically targeted by, for example, the use of primers that preferentially amplify their 16S rRNA gene or their *nifH* gene in the case of nitrogen fixing cyanobacteria. However, since not all cyanobacterial nitrogen fixers are heterocytous, there is currently no fast gene-based method to specifically detect and distinguish heterocytous cyanobacteria. Here, we developed a PCR-based method to specifically detect heterocytous cyanobacteria by designing primers targeting the gene (*hglT*) encoding the enzyme responsible for the last step in the biosynthesis of heterocyte glycolipid (i.e., a glycosyltransferase). We designed several primer sets using the publicly available sequences of 23 heterocytous cyanobacteria, after testing them on DNA extracts of 21 heterocyte-forming and 7 non-heterocyte forming freshwater cyanobacteria. The best primer set was chosen and successfully used to confirm the presence of heterocytous cyanobacteria in a marine environmental sample.

## KEYWORDS

heterocytous cyanobacteria, N<sub>2</sub> fixation, heterocyte glycolipids, glycosyltransferase, lipid biosynthesis, polymerase chain reaction, primer design, environmental detection

## Introduction

Cyanobacteria are cosmopolitan photosynthetic organisms with a central role in earth's biogeochemical cycles. Besides their role as primary producers, fixing carbon through photosynthesis using only sunlight and CO<sub>2</sub>, a variety of cyanobacterial species are also capable of fixing atmospheric nitrogen (N<sub>2</sub>), which also confers them an important role in the nitrogen cycle (Zehr, 2011; Mishra et al., 2021).

Nitrogen fixation is catalyzed by the nitrogenase enzyme which is easily inactivated in presence of oxygen. Hence, cyanobacteria have developed two main strategies to protect the nitrogenase from the oxygen generated during photosynthesis. One strategy is to separate both processes in time; in most nitrogen-fixing unicellular cyanobacteria  $N_2$  fixation only occurs at night, when the photosynthetic machinery is not in use and oxygen is not being produced (Millineaux et al., 1981). Filamentous cyanobacteria developed an alternative strategy, i.e., the spatial separation of both processes. Heterocytous cyanobacteria can develop heterocytes, a type of specialized cell where, when in absence of combined nitrogen, nitrogen fixation takes place (Klugkist and Haaker, 1984; Knutson et al., 2018). Heterocytes protect the nitrogenase enzyme from oxygen using several strategies such as (1) providing spatial separation of nitrogen and carbon fixation by confining nitrogen fixation to specialized cells surrounded by a thick-walled outer membrane composed of polysaccharide (heterocyte envelope polysaccharide, HEP) and an inner layer composed of heterocyte glycolipids (HGs) (Haselkorn, 1978), (2) inactivation of photosystem II in heterocytes to avoid production of oxygen (Tel-Or et al., 1997), and (3) enhanced respiration inside the heterocytes to consume internal oxygen (Wolk et al., 1994).

Currently, the presence of cyanobacteria in the environment can be confirmed by using cyanobacteria-specific 16S rRNA gene primers (Nübel et al., 1997), and the presence of dinitrogen-fixing cyanobacteria by applying primers targeting *nifH* (Olson et al., 1998), a key gene involved in dinitrogen fixation. However, there are no screening techniques based on genetic methods allowing the detection of heterocytous cyanobacteria in the environment, which could be used to evaluate the potential for cyanobacterial  $N_2$  fixation. Thus far, we can only determine the presence of heterocytous cyanobacteria using time and resource intensive methods such as microscopy and lipid analysis, or by using untargeted metagenomic methods. Taxonomic classification of cyanobacteria via microscopic analysis requires expertise and relies solely in morphological traits (Castenholz et al., 2001).

Cyanobacterial taxonomy is complex and has undergone extensive restructuring and revisions in recent years. It has transformed from being a hierarchical system that simply placed morphologically similar taxa into a classification system based on evolutionary relationships. Most of the studies mentioned in this work used the classification system proposed by Rippka et al. (1979), later adopted as the classification standard in the Bergey's Manual of Systematic Bacteriology (Castenholz et al., 2001). This classification method recognized five subsections or orders, which in turn contained "families," divided mainly according to their morphology. Heterocytous cyanobacteria occur only in subsections IV (*Nostocales*) and V (*Stigonematales*). Although this nomenclature did not reflect evolutionary relationships, it was adopted as convenient and temporary method to classify cyanobacterial strains (Castenholz et al., 2001). However, a recent revision of cyanobacterial taxonomy by Komarek et al. (2014) proposed a system that left aside classification based solely on morphology and focused on a polyphasic approach focusing on gene phylogeny (e.g., 16S rRNA) to achieve monophyletic clusters representing "orders."

The chemical structure of the HGs and their distribution (Figure 1) has chemotaxonomical value and can potentially also be used to distinguish between cyanobacteria of different orders and families (Table 1; Supplementary Table S1; Gambacorta et al., 1998; Bauersachs et al., 2009a, 2013). For example, hexose-sugar HGs composed of  $C_{26}$  ( $HG_{26}$ ) diol and keto-ol are mostly present in species of the *Nostocaceae*

family (*Nostocaceae* and *Aphanizomenonaceae* orders according to the new 16S rRNA-based classification; Gambacorta et al., 1998; Bauersachs et al., 2009a; Komarek et al., 2014; Table 1; Supplementary Table S1), while hexose  $C_{28}$  HGs ( $HG_{28}$ ) containing three functional groups (keto-diol and triol) are predominant in members of the *Rivulariaceae* (*Calotrichaceae*) family. Likewise, hexose  $C_{30}$  HG ( $HG_{30}$ ) triols and keto-diols are characteristic of members of the *Scytonemataceae* (*Scytonemataceae* and *Tolypothrichaceae*) family, while members of the *Stigonemataceae* (*Stigonemataceae*, *Chlorogloeopsidaceae* and *Hapalosiphonaceae*) mostly synthesize hexose  $C_{32}$  HGs ( $HG_{32}$ ) with three functional groups (Gambacorta et al., 1998). However, there are some exceptions to these general trends. For example, the *Nostocaceae* (*Nostocaceae* and *Aphanizomenonaceae*) family mostly produces hexose  $HG_{26}$  diol and keto-ol, but there are members of this family that also produce hexose  $HG_{28}$  triol and keto-diol (Supplementary Table S1). Although they are usually produced in small amounts, in some cases they are synthesized in larger amounts, such as in *Aphanizomenon* sp. TR83 (12% of the total HGs) and in *Dolichospermum* sp. BIR169 (33%) (Bauersachs et al., 2017; Supplementary Table S1). Moreover, *Stigonema ocellatum* SAG 48.90, a strain belonging to the *Stigonematales* order (*Stigonemataceae*), was recently shown to produce mostly hexose  $HG_{28}$  keto-diol (67%) and triol (27%), HGs thought to be characteristic of the *Rivulariaceae* (*Calotrichaceae*) family (Supplementary Table S1; Bauersachs et al., 2019).

Additionally, the type of headgroup of HGs (Figure 1) can also be an indicator of the lifestyle and habitat of the producing heterocytous cyanobacterial strain. Most HG-producing cyanobacteria biosynthesize HGs with a hexose ( $C_6$ ) as headgroup, however some cyanobacterial species also produce HGs with alternative headgroups such as pentoses ( $C_5$ ), deoxyhexoses ( $C_6$ ) or methyl-hexoses ( $C_6$ ) (Schouten et al., 2013; Bale et al., 2018a; Figure 1; Supplementary Table S1). Most of the heterocytous cyanobacterial strains studied to date that synthesize HGs with a hexose headgroup are free-living and inhabit freshwater and terrestrial environments, with a few exceptions, i.e., *Nostoc muscorum* UTEX 1933 and *Calothrix* sp. UTEX 2589 synthesize  $HG_{26}$  and  $HG_{28}$  with hexose as headgroup but were enriched from marine environments (Schouten et al., 2013). In contrast, heterocytous cyanobacteria that are symbionts of marine diatoms, synthesize  $HG_{30}$  and  $HG_{32}$  HGs containing a pentose headgroup, and were, consequently, proposed as biomarkers for tracing endosymbiotic heterocytous cyanobacteria in present and past marine environments (Schouten et al., 2013; Bale et al., 2018b).

Due to their potential preservation over geological time scales, HGs have also been used as biomarkers to study the occurrence of nitrogen fixation in ancient sediments (Bauersachs et al., 2010). In addition, changes in the abundances of  $HG_{26}$  diols and keto-ols can be used to reconstruct surface water temperatures in freshwater environments using the  $HDI_{26}$  index (Bauersachs et al., 2015). However, our current understanding of HG production among heterocytous cyanobacteria is restricted, because only a limited amount of HG-producing cyanobacterial cultures has been examined. A solution would be to study the heterocytous cyanobacterial community occurring in the environment using both approaches, i.e., HG and genetic analyses. One could use untargeted genetic approaches such as 16S rRNA and *nifH* gene analysis using primers that specially target cyanobacteria. However, these approaches might not be specific enough, since they will target all cyanobacteria and  $N_2$  fixing cyanobacteria (among other  $N_2$  fixing organisms; Olson et al., 1998). Hence, if the species of interest are sparse, there is the risk of overlooking them and detecting only the



TABLE 1 Summary of HGs produced by different cyanobacterial families according to the classification described in Komarek et al. (2014) and to the nomenclature in the original studies. \*,\*\*

Family in original study	Family as per Komarek et al. (2014)	HG <sub>26</sub> diol keto-ol and diol	HG <sub>28</sub> keto-ol and diol	HG <sub>28</sub> keto-diol and triol	HG <sub>30</sub> keto-ol and diol	HG <sub>30</sub> keto-diol and triol	HG <sub>32</sub> keto-diol and triol
<i>Nostocaceae</i> <sup>abc</sup>	<i>Aphanizomenonaceae</i>	++	(++ – tr.)	(+ – tr.)			
<i>Nostocaceae</i> <sup>abdefg</sup>	<i>Nostocaceae</i>	++	(+ – tr.)				
<i>Rivulariaceae</i> <sup>abehi</sup>	<i>Calotrichaceae</i>		(+ – tr.)	++			
<i>Rivulariaceae</i> <sup>h</sup>	<i>Fortieaceae</i>		++		tr.	tr.	
<i>Scytonemataceae</i> <sup>h</sup>	<i>Tolypothrichaceae</i>		++			+	
<i>Scytonemataceae</i> <sup>hi</sup>	<i>Scytonemataceae</i>					++	
<i>Stigonemataceae</i> <sup>i</sup>	<i>Stigonemataceae</i>		+	++		tr.	
<i>Stigonemataceae</i> <sup>h</sup>	<i>Chlorogloeopsidaceae</i>						++
<i>Stigonemataceae</i> <sup>hij</sup>	<i>Hapalosiphonaceae</i>						++
<i>Fischerellaceae</i> <sup>i</sup>	<i>Hapalosiphonaceae</i>						++
<i>Nostochopsidaceae</i> <sup>i</sup>	<i>Hapalosiphonaceae</i>						++

++ dominant, + present, tr. trace amounts.

<sup>a</sup>Bauersachs et al. (2009b), <sup>b</sup>Wörmer et al. (2012), <sup>c</sup>Bauersachs et al. (2014), <sup>d</sup>Bauersachs et al. (2010), <sup>e</sup>Bauersachs et al. (2009a), <sup>f</sup>Gambacorta et al. (1996), <sup>g</sup>Schouten et al. (2013), <sup>h</sup>Gambacorta et al. (1998), <sup>i</sup>Bauersachs et al. (2019), <sup>j</sup>Bauersachs et al. (2014).

\*In this work when discussing strains analyzed in previous studies in the main text, we will use the taxonomic classification as described in the original study (first column of this table), as well as their family-level classification as per described in Komarek et al. (2014) (second column of this table), the latter will be shown in between brackets.

\*\*Values without brackets indicate HGs present in all strains of the indicated family, while values in brackets indicate HGs present only in some strains. Hyphen symbol (–) indicates a range of abundances.

the Central Atlantic Ocean (station 24: 8° 26' 6.1548", –49° 44' 27.2142") during cruise 64PE455 on board of the *R/V Pelagia* during July–August 2019 as described in Theirlinck et al. (2023). DNA extraction of *S. fluitans* sample number 38 is described in Theirlinck et al. (2023).

## Lipid extraction and analysis

Lipids present in the microbial mat sample were extracted from freeze dried aliquots using a modified Bligh-Dyer procedure as described by Bauersachs et al. (2011 and references within). The extracted intact polar lipids were analyzed by high performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry (HPLC/ESI-MS<sup>2</sup>) as described in Bauersachs et al. (2009b).

## Polymerase chain reaction (PCR) and sequencing

### Validation of the DNA extracted from cyanobacterial cultures

To ensure that the DNA extracted from cyanobacterial cultures was of good quality and suitable for our study, we performed a PCR using universal primers 515F-Y and 806RB targeting the 16S rRNA gene (Supplementary Table S2; Walters et al., 2015), resulting in the expected ~300 bp band. All PCR reaction mixes and cycles used in this study were as described in Besseling et al. (2018) with minor modifications, i.e., PCR annealing temperature was 50°C, extension time was 30" and it was repeated for 30 cycles (hereafter these specifications will be shown within parenthesis in the main text).

### Amplification and sequencing of *hglT*

Upon optimisation of the PCR conditions (see Results), amplification of *hglT* using the primers designed in this study was carried out using an annealing temperature 56.2°C, 1'20" extension time, for 30 cycles, and an equimolar mixture of primers, for example Rv1 refers to an equimolar mixture of Rv1\_G1, Rv1\_G2, Rv1\_G3, Rv1\_G4 and Rv1\_G5.

Upon amplification of *hglT* with primer mixes Fw1 mixB and Rv1, the presence of a single or multiple PCR products was analyzed via electrophoresis on an agarose gel (1%) stained with ethidium bromide or GelRed® (Biotium, Hayward, CA, USA), using SmartLadder (Eurogentec, Seraing, Belgium) as molecular weight marker. If only one band was observed the PCR product was purified using a PCR purification kit (QIAquick PCR Purification kit, Qiagen, Hilden, Germany). If more than one band was observed, then the PCR product was run on 1% or 1.5% agarose gels stained with SYBR Safe® (ThermoFisher Scientific, USA), bands were then excised and purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany). To confirm the purity of the purified bands a small aliquot was run again on an agarose gel (1%) as described above. If the amount of the PCR product was not sufficient for direct sequencing, the amount was increased via concentration of the product of several PCR reactions using a PCR purification kit (QIAquick PCR Purification kit, Qiagen, Hilden, Germany). Purified PCR products were sequenced via Sanger sequencing at MacroGen Europe (Amsterdam, The Netherlands) using primer mixes Fw1 mixB and Rv1.

If after concentration the amount of PCR product was still not sufficient for direct sequencing, it was then cloned into the pCR™4-TOPO™ vector (Invitrogen, Carlsbad, CA, USA) using the TOPO-TA cloning kit for sequencing (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The ligation product was subsequently transformed into TOP10 competent cells (Invitrogen, Carlsbad, CA, USA) and plated in LB plates containing ampicillin

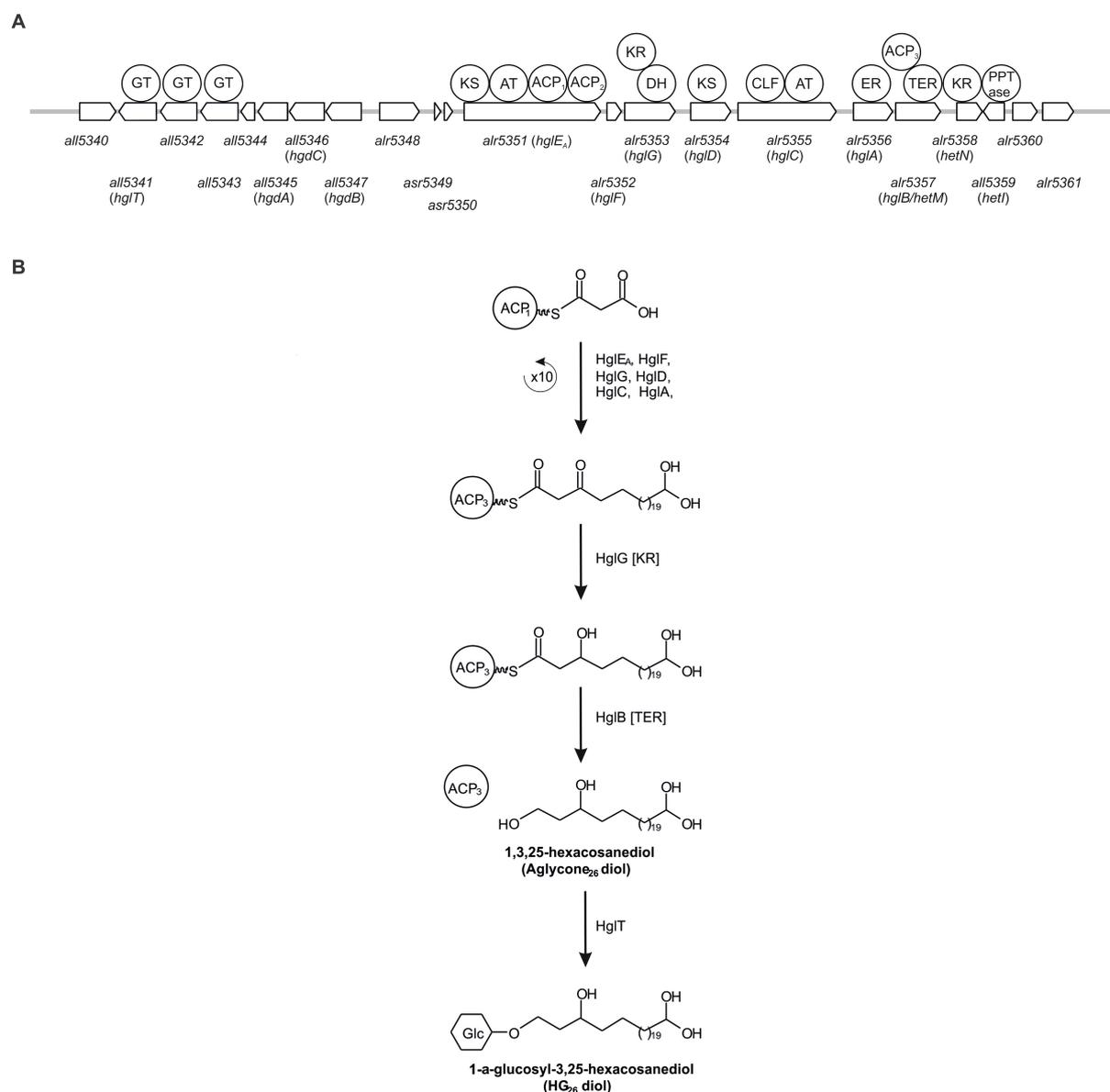


FIGURE 2

Schematic representation of the genes involved in HG biosynthesis in *Anabaena* sp. PCC 7120 (A) and simplified representation of the biosynthesis pathway of HG26 diols (B). ACP, acyl carrier protein; AT, acyl transferase; CLF, chain length factor; DH, dehydrase; ER, enoyl reductase; GT, Glycosyltransferase; KR, ketoacyl reductase; KS,  $\beta$ -ketoacyl synthase; TER, thioester reductase.

(100  $\mu$ g/mL), kanamycin (20  $\mu$ g/mL), IPTG (0.1 mM) (UltraPure™ IPTG, Invitrogen Europe, manufactured in Italy) and X-Gal (0.2 mg/mL) (Thermo Scientific, Vilnius, Lithuania). The resulting white colonies were collected using a sterile toothpick, dissolved in 50  $\mu$ L TE 1x buffer and lysed by incubation at 95°C for 10 min. Upon centrifugation at 2000 rpm for 1 min, 1  $\mu$ L supernatant was used as template in 25  $\mu$ L PCR reactions using M13 primers (Supplementary Table S2) and the PCR reaction mix and amplification program described above. PCR products were sent to MacroGen Europe for PCR purification and Sanger sequencing using primers T3 and T7 (Supplementary Table S2).

### Amplification and sequencing of 16S rRNA

16S rRNA V4-V5 regions from DNA present in the microbial mat sample was amplified using barcoded primers 515F-Y and 926Rbc

(Parada et al., 2016) via PCR reaction using Phusion Taq Polymerase (Thermo Fischer Scientific). And the following PCR program: 30" at 98°C, 30 cycles of 10" at 98°C, 20" at 50°C, 30" at 72°C; final extension at 72°C for 7" and storage at 4°C. Samples were amplified in triplicate and pooled together after positive confirmation via electrophoresis on an agarose gel (2%). Positive controls (*E. coli* lysate) and negative controls (PCR water without DNA) were included. Concentration of each PCR product was quantified on an agarose gel (2%) using a quantification standard. PCR products from all samples were pooled in equimolar concentrations, ran on an agarose gel (2%) and the resulting ~400 bp band was cut off the gel and purified using QIAquick gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Sequencing was performed by MacroGen Europe (Amsterdam, The Netherlands) using an Illumina MiSeq 2  $\times$  300 bp V3 kit (Illumina, CA,

TABLE 2 Summary of the strains used in this study to test the designed primers.

Heterocytous	
From Culture Collection Yerseke (CCY)	
CCY 0012	<i>Nostoc</i>
CCY 0017	<i>Anabaena</i>
CCY 0018	<i>Calothrix</i>
CCY 0103	<i>Nodularia chucula</i>
CCY 0202	<i>Calothrix</i>
CCY 9402-a	<i>Anabaena</i>
CCY 9613	<i>Anabaena</i>
CCY 9614	<i>Anabaena</i>
CCY 9626*	<i>Nostoc</i>
CCY 9910	<i>Anabaena</i>
CCY 9921	<i>Anabaena cylindrica</i>
CCY 9922	<i>Anabaena variabilis</i>
CCY 9923	<i>Calothrix</i>
CCY 9924	<i>Chlorogloeopsis</i> sp.
CCY 9925	<i>Nostoc</i>
CCY 9926	<i>Nostoc</i>
From Pasteur Culture Collection (PCC)	
PCC 7101	<i>Tolypothrix tenuis</i>
PCC 10023	<i>Scytonema</i> sp.
Other sources	
PCC 7120*	<i>Anabaena</i> sp.
UAM 290	<i>Chrysochlorium ovalisporum</i>
UAM 292	<i>Chrysochlorium ovalisporum</i>
Non heterocytous	
From Culture Collection Yerseke (CCY)	
CCY 9627	<i>Leptolyngbya</i>
CCY 9409	<i>Lyngbya</i> sp.
CCY 9703	<i>Pseudanabaena</i>
CCY 0011	<i>Synechococcus</i>
CCY 9201	<i>Synechococcus</i>
CCY 9503	<i>Synechococcus</i>
CCY 9506	<i>Synechococcus</i>

\*Indicates the same strain kept in two different culture collections under different identification numbers.

USA). Amplification and analysis of 16S rRNA gene of the *S. fluitans* sample is described in [Theirlinck et al. \(2023\)](#).

## Bioinformatic analysis

### Search of HglT homologs and primer design

Homologs of the protein HglT of *Anabaena* sp. PCC 7120, encoded by gene *all5351* (IMG gene ID 637235753 hereafter, *hglT*) were found via BLASTx (translated nucleotide sequence searched against protein sequences, accessed on May 2020) ([Altschul et al.,](#)

1990; [Camacho et al., 2009](#)) against 23 genomes ([Supplementary Table S3](#) and [Supplementary Figure S1](#)). The resulting gene sequences ([Supplementary Table S4](#)) were aligned in MEGA (v11.0.8) using the in-built MUSCLE ([Edgar, 2004](#)) sequence aligner with default parameters [gap open penalties: -400, gap extend penalties: 0, max iterations: 16, Cluster Method: UPGMA, Min Diag Length (Lambda): 24]. To select suitable regions for the design of our primers we used the following guidelines: (1) Binding region should be 17–22 base pairs (bp) long; (2) Binding region Tm's should be 50–65°C; (3) Binding region should have moderate GC content (<50%); (4) Repetitive poly-N regions should be avoided; (5) Regions prone to secondary structure formation (hairpins, self-dimerization or primer dimer), such as regions containing complementary sequences (palindromes) within themselves should be avoided; (6) Primer pairs should have similar Tm's ( $\pm 4^\circ\text{C}$ ); and (7) Distance between the two binding regions (Fw and Rv) should be as large as possible. The regions comprised between the base pairs 112..131 and 299..320, were selected as suitable regions for the design of forward primers, and regions 949..968 and 973..989 were selected as targets for the design of reverse primers.

Primers were designed according to the general primer design guidelines described by [Abd-El Salam \(2003\)](#), which we applied by setting the following rules: (1) length of the primer should be between 17 and 22 bp (2) degenerate nucleotides should not be consecutive (3) degenerate nucleotides should be as far apart from each other as possible (4) each primer sequence should have the least number of degenerate nucleotides possible but no more than four. Primer properties such as melting temperature (Tm) and percentage of GC content were also taken into consideration, aiming to design primers with similar Tm and moderate GC content (<50%). The potential of each primer to create hairpins and primer dimers due to potential self-annealing was also checked by using OligoCalc ([Kibbe, 2007](#); see [Supplementary Table S5](#)). All primers were synthesized by Integrated DNA Technologies (IDT, Leuven, Belgium).

### Analysis of *hglT* amplicons

Nucleotide sequences obtained via Sanger sequencing were processed using Geneious Prime (v 2023.0.4). Sequences were trimmed using an error probability limit of 0.01. Potential heterozygous bases in single reads were identified using a 50% peak similarity cutoff, peak detection height was set at 10%. When available, the consensus sequence was obtained by aligning forward and reverse reads with Geneious assembler using the highest sensitivity settings. When appropriate, sequences belonging to the pCR<sup>TM</sup>4-TOPO<sup>TM</sup> vector (Invitrogen, Carlsbad, CA, USA) were identified and removed.

To determine their origin, the resulting sequences were queried using Geneious built-in blastN (BLAST+ v 2.12.0, accessed in March 2023) ([Altschul et al., 1990; Camacho et al., 2009](#)) with default parameters (low complexity filter, mask for lookup table, max e-value: 0.05, word size: 11, open gap cost: 5, extend gap cost: 2, scoring match: 2, scoring mismatch: 2–3).

To generate the phylogenetic trees sequences were aligned using MAFFT (v7.407) with L-INS-i iterative refinement method ([Katoh and Standley, 2013](#)) and poorly aligned regions were removed using trimAl ([Capella-Gutiérrez et al., 2009](#)). Phylogenetic trees were built using IQ-tree (v1.6.7) and its in-built nucleotide substitution model finder ([Kalyaanamoorthy et al., 2017](#)), using 1,000 replicates to perform SH-like approximate likelihood ratio test (SH-aLRT)

(Guindon et al., 2010) and 1,000 bootstrap replicates. Phylogenetic trees were visualized using iTOL (v6.7.4) (Letunic and Bork, 2021).

### Analysis of 16S rRNA amplicons

The resulting amplicon sequencing data was processed using the Cascabel pipeline (Abdala Asbun et al., 2020). Within the Cascabel pipeline raw reads were filtered, trimmed and ASVs identified using the dada2 R package (v 1.19.1) (Callahan et al., 2016). Taxonomy was assigned using RDP (Wang et al., 2007) classifier within dada2, using SILVA SSU database 138 Ref NR 99 (Quast et al., 2013) and setting the minimum bootstrapping support required to return a taxonomic classification to 45% (minBoot=45). Resulting taxa with a minimum count of two observations were summarized using QIIME (using summarize\_taxa.py v 1.9.1 and filter\_otus\_from\_otu\_table.py v 1.9.1) (Caporaso et al., 2010) and BIOM tables were converted using BIOM tool (v 2.1.6) (McDonald et al., 2012). Processing and analysis of 16S rRNA amplicon sequencing data of the *S. fluitans* sample is described in Theirlinck et al. (2023).

## Results and discussion

### Assessing the potential of *hglT* as phylogenetic marker gene

To assess the phylogenetic potential of *hglT*, we generated a phylogenetic tree using complete *hglT* sequences from genomes available on the Integrated Microbial Genomes & Microbiomes system (IMG/M<sup>1</sup>) (Chen et al., 2021). We searched for homologs of the protein HglT of *Anabaena* sp. PCC 7120, encoded by gene *all5351* (IMG gene ID 637235753 hereafter, *hglT*) against 23 genomes (Supplementary Table S3) using BLASTx (translated nucleotide sequence searched against protein sequences), accessed on May 2020 (Altschul et al., 1990; Camacho et al., 2009). These included the genomes of 11 strains known to produce HGs, and 12 genomes of strains closely related to strains known to produce HGs. For example, *Chrysochloris ovalisporum* UAM 290 has been described to synthesize HGs (Wörmer et al., 2012) but there is no genome available for this strain, so we chose the genome of *Chrysochloris ovalisporum* UAM-MAO. In the BLASTx search, we selected for each genome, the protein sequence with the highest percentage of homology (average 75%;  $n=23$ ) and highest bit-score (average 421,  $n=23$ ) (Supplementary Table S4). These sequences were used to build the *hglT* phylogenetic tree (Figure 3A).

To ensure that the *hglT* gene tree provides enough resolution, we compared it to a 16S rRNA tree generated using sequences obtained from the same (or closely related) genomes as *hglT* (Figure 3B; Supplementary Figure S2; Supplementary Table S6). Cyanobacteria can possess multiple copies of the 16S rRNA gene, oftentimes also of different sequence length. Thus, to avoid potential misalignments caused by large differences in sequence length we only included sequences which were at least 1,000 bp long. Figure 3B reveals that, in most cases, the 16S rRNA sequences of the examined cyanobacteria cluster together, with a few exceptions, such as the two sequences of *Chrysochloris ovalisporum* UAM-MAO.

The results from the *hglT* and 16S rRNA tree do not always show the same evolutionary history for a given genome. However, both trees agree with observations of previous studies, such as the polyphyletic origin of nearly all cyanobacterial families present in the trees, confirming previous reports in literature about the polyphyletic origin of *Nostoc*, *Anabaena*, *Calothrix*, *Tolypothrix* and *Scytonema* (Komárek et al., 2013; Komárek et al., 2014; Berrendero Gómez et al., 2016 and references cited therein), and the monophyletic origin of the *Stigonemataceae* family (*Chlorogloeopsidaceae* and *Hapalosiphonaceae*) (Castenholz et al., 2001 and references within). In general, it can be concluded that the *hglT* tree provides enough resolution to use this gene as a marker for HG-producing cyanobacteria.

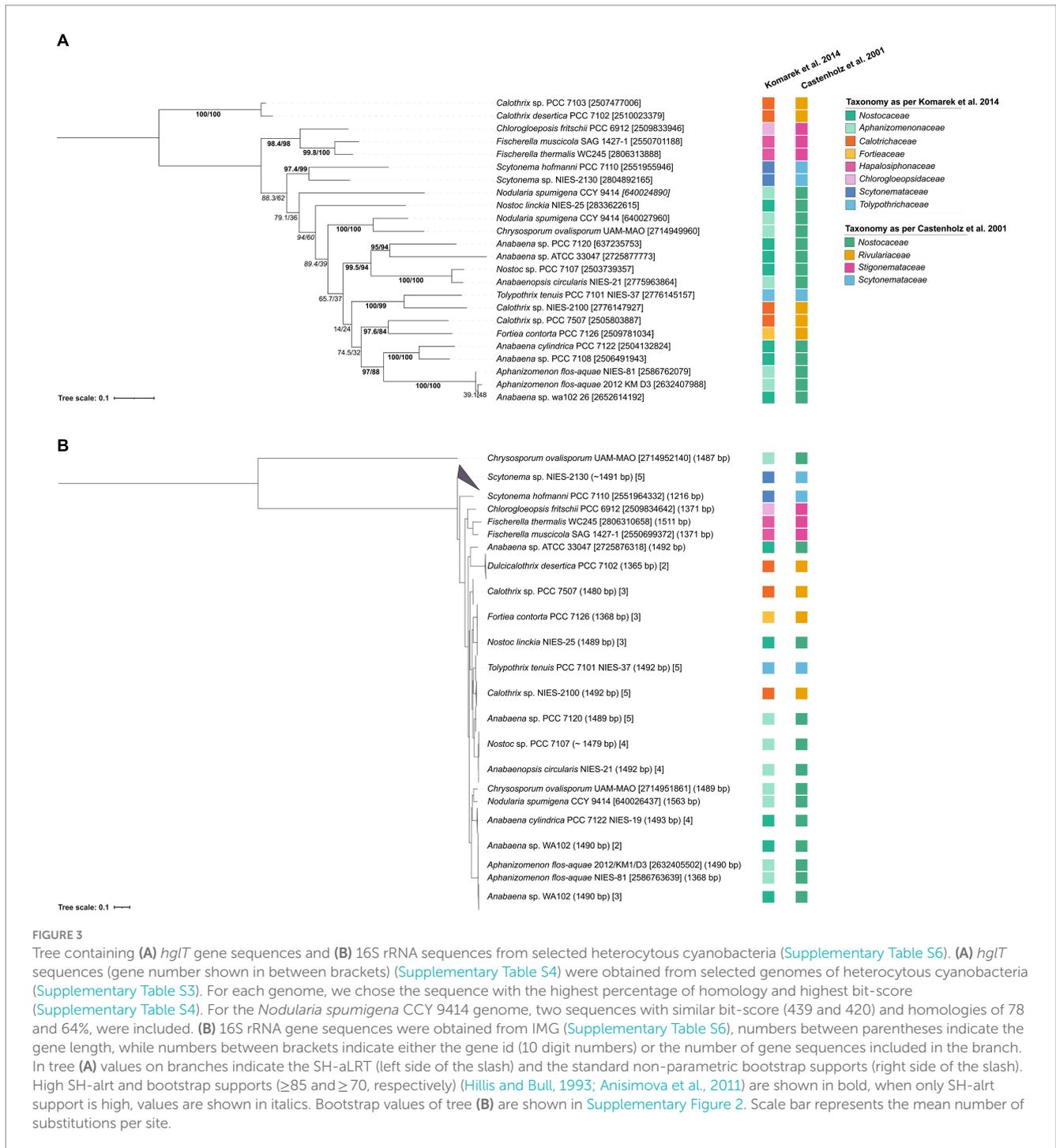
### Primer design and validation in pure cultures

To design primers targeting *hglT*, we used all sequences used to generate the *hglT* tree (Figure 4; Supplementary Table S4), aligned them in MEGA (v 11.0.8) using the in-built MUSCLE (Edgar, 2004) sequence aligner with default parameters, and manually identified regions suitable as primer targets (see Supplementary Figure S1 for details).

To be able to target the largest number of cyanobacterial species, while still following the rules considered for the design of the degenerate primers (i.e., only a maximum of four degenerate nucleotides should be included; see Material and Methods for details), we decided to design several primers targeting the same binding region but containing different combinations of degenerate bases (Figure 4; Supplementary Tables S2, S5), and use them in equimolar amounts in the PCR reaction. For some primers (Fw1\_G1, Fw1\_G2, Fw2\_G3 and Fw2\_G4), we also designed an alternative variant (Fw1\_G1n, Fw1\_G2af, Fw2\_G3p and Fw2\_G4c, respectively) (Supplementary Table S5) containing one additional degenerate nucleotide that would allow to increase the number of targeted strains (Supplementary Table S7). For example, Fw1\_G1n contains 4 degenerate bases and targets 8 of the 24 strains, while Fw1\_G1 contains 3 degenerate bases and targets only 6 strains. Both primer variants were tested in parallel and in combination with primers targeting the same region for which only one variant was designed (i.e., Fw1\_G3 to Fw1\_G7). For example, primer variants with the least number of degenerations (i.e., Fw1\_G1 and Fw1\_G2) were tested in combination with Fw1\_G3 to Fw1\_G7 (hereafter referred to as “mix A”), and primer variants with the greatest number of degenerations (i.e., Fw1\_G1n and Fw1\_G2af) were also tested in combination with Fw1\_G3 to Fw1\_G7 (hereafter referred to as “mix B”) (Figure 4).

The designed primers were tested using DNA extracted from 21 heterocyte-forming cyanobacterial strains (Table 2; Supplementary Table S8). Next, in order to develop the best suited primer set and its optimal annealing temperature, we performed a PCR with an annealing temperature gradient (52–63°C annealing temperature, 1'20" extension time, 30 cycles) on two selected strains (*Calothrix* CCY 9923 and *Chlorogloeopsis* sp. CCY 9924) using all possible primer combinations (Supplementary Figure S3). All primer combinations successfully targeted the selected strains, and primer combinations containing a larger number of degenerate bases (mix B) did not show more unspecific bands than those containing the more conservative primers (mix A)

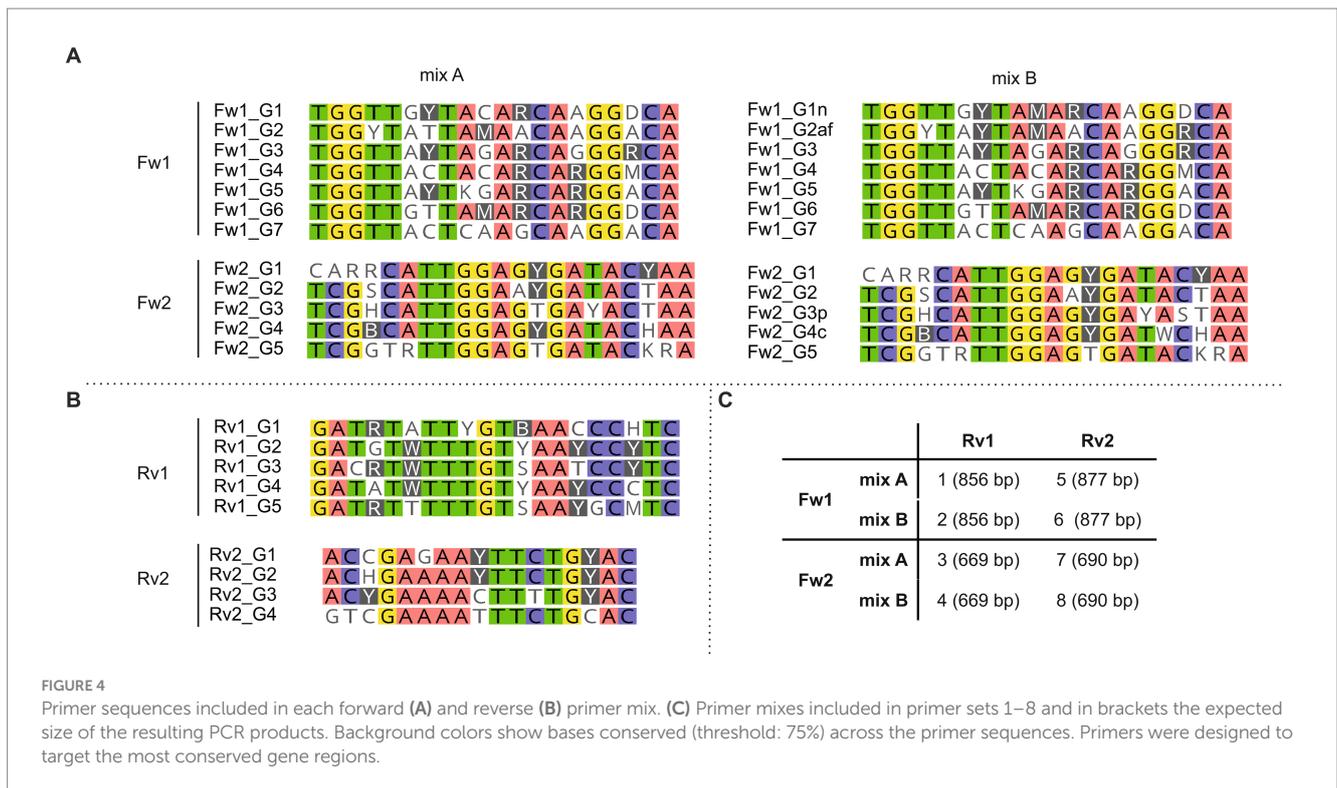
1 <https://img.jgi.doe.gov/>



**FIGURE 3**  
 Tree containing (A) *hglT* gene sequences and (B) 16S rRNA sequences from selected heterocytous cyanobacteria (Supplementary Table S6). (A) *hglT* sequences (gene number shown in between brackets) (Supplementary Table S4) were obtained from selected genomes of heterocytous cyanobacteria (Supplementary Table S3). For each genome, we chose the sequence with the highest percentage of homology and highest bit-score (Supplementary Table S4). For the *Nodularia spumigena* CCY 9414 genome, two sequences with similar bit-score (439 and 420) and homologies of 78 and 64%, were included. (B) 16S rRNA gene sequences were obtained from IMG (Supplementary Table S6), numbers between parentheses indicate the gene length, while numbers between brackets indicate either the gene id (10 digit numbers) or the number of gene sequences included in the branch. In tree (A) values on branches indicate the SH-aLRT (left side of the slash) and the standard non-parametric bootstrap supports (right side of the slash). High SH-aLRT and bootstrap supports ( $\geq 85$  and  $\geq 70$ , respectively) (Hillis and Bull, 1993; Anisimova et al., 2011) are shown in bold, when only SH-aLRT support is high, values are shown in italics. Bootstrap values of tree (B) are shown in Supplementary Figure 2. Scale bar represents the mean number of substitutions per site.

(Supplementary Figure S3). Consequently, we decided to only use primer mix B to be able to target the largest number of cyanobacterial strains, and thus reduce the total number of primer combinations from eight to four. For strain *Calothrix* CCY 9923 primer combinations containing Fw1 primers resulted in an additional smaller (~400 bp) and unexpected band in the agarose gel when using an annealing temperature of 56.2°C or lower (Supplementary Figure S3). However, in some cases using an annealing temperature higher than 56.2°C resulted in much lower concentrations of PCR product, as was observed, for example, when primers Fw1 mixB and Rv2 and an annealing temperature of 58.2°C

were used on *Chlorogloeopsis* sp. CCY 9924 (Supplementary Figure S3). Hence, despite the appearance of the unspecific band, we chose in favor of higher PCR product concentrations and chose to use an annealing temperature of 56.2°C in all subsequent PCR reactions. Next, we checked which of the four remaining primer pairs could amplify the *hglT* gene from the largest number of strains (Supplementary Figure S4). Primer pair 2 (Fw1 mix B + Rv1) successfully generated a PCR product of the expected size (~850 bp) for most of the tested strains (Supplementary Figure S4) and was selected as the best primer pair to potentially detect C<sub>6</sub> HG producers in the environment.



To check the specificity of this primer pair (Fw1 mix B + Rv1) we also tested it in non-heterocyte forming cyanobacterial strains (Table 2; Supplementary Table S8). In four out of seven of these strains (*Synechococcus* strains CCY 0011, CCY 9201, CCY 9503 and CCY 9506) PCR resulted in the production of two very faint bands (~1,000 bp and ~600 bp) (Supplementary Figure S5), which upon cloning and sequencing showed some homology to sequences present in *Synechococcus* and *Cyanobium*, indicating that the primers are not fully specific for heterocytous cyanobacteria. Nonetheless, this unspecific and inefficient amplification using the designed primers can be discerned by the size of the PCR product and ultimately by sequencing the amplified products (Supplementary Tables S9, S10).

To check if the selected primer pair (Fw1 mix B + Rv1) actually amplified *hglT*, the most abundant PCR fragments obtained from the tested cyanobacterial cultures (shown as ++ in Supplementary Figure S4) were purified and sequenced. Initially, to ensure that all the PCR fragments obtained corresponded to heterocytous cyanobacteria, we carried out an NCBI blastN search and confirmed that all analyzed PCR fragments corresponded to HG-producing cyanobacterial strains (Supplementary Table S9). However, the genera (or species) of the hit with the highest bit-score did not always correspond to genera assigned to the tested (query) strain.

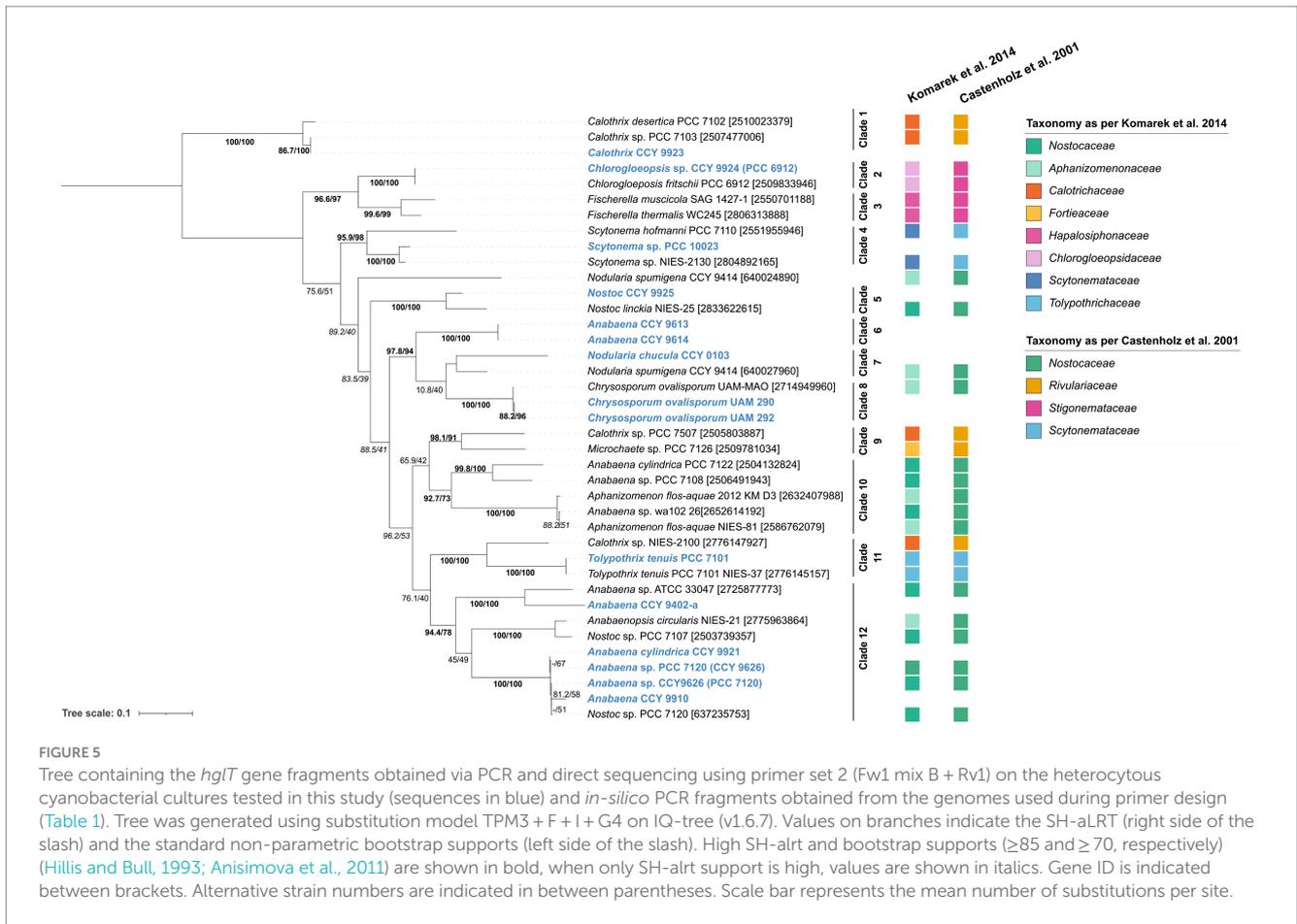
Next, we wanted to see if the amplified fragments corresponding to *hglT* could be used as phylogenetic marker and assist during strain identification and naming. To do so we generated a phylogenetic tree containing the PCR fragments (700–870 bp) obtained from the cultures analyzed in this study and PCR fragments generated *in silico* (Figure 5). *In silico* PCR fragments were obtained from the *hglT* sequences initially used to design the primers (Supplementary Table S4) by selecting the DNA fragment of *hglT* contained between primers Fw1 mixB and Rv1. This tree (Figure 5) shows that the targeted *hglT* fragment has potential as a phylogenetic marker up to the genus level.

We observe that the sequences obtained using our primers, in most cases, cluster with *hglT* fragments of strains of the same genera (Figure 5). For example, *Calothrix* CCY 9923 clusters with *Calothrix* sp. PCC 7103 and *Calothrix desertica* PCC 7102 (Clade 1), confirming its affiliation to the genus *Calothrix*, while *Scytonema* sp. PCC 10023 clusters with *Scytonema* sp. NIES-2130 and *Scytonema hofmanmi* PCC 7110 (Clade 4), also confirming its affiliation to the genera *Scytonema*. Clade 12 contains closely related sequences that are identified with different names and strain identifiers, including *Anabaena* CCY 9910, *Anabaena cylindrica* CCY 9921 and three *Anabaena/Nostoc* sp. PCC 7120. The latter is also shown in the tree as *Anabaena* sp. CCY 9626, which refers to strain PCC 7120 after being kept in the former Culture Collection Yerseke (CCY). Strain PCC 7120 has been renamed several times, underscoring the difficulties faced by taxonomists during cyanobacterial classification. It was known first as *Nostoc muscorum*, it was then classified as *Anabaena* (Rippka et al., 1979) and later renamed as *Nostoc* although this reassignment has been questioned by some studies (Svenning et al., 2005 and references within).

Overall, these results show that *hglT*-primer set 2 is sensitive enough to detect the gene of the glycosyltransferase involved in HG biosynthesis in heterocyte-forming cyanobacteria and that upon sequencing the resulting *hglT* fragments can be used as taxonomic marker.

## Application of the developed genetic approach to the environment

After confirmation that the primer set 2 (Fw mix B and Rv1) successfully targeted *hglT* in heterocytous cyanobacterial cultures, it was tested in the environment using a microbial mat collected from the Dutch Wadden island of Schiermonnikoog. The lipid and microbial composition of microbial mats from this location had been



**TABLE 3** Heterocyte glycolipids present in the microbial mat collected from Schiermonnikoog (53°29.445'N, 6°8.342'E) in April 2010.

	Relative abundance (%)
1-(O-hexose)-3,25-hexacosanediol (HG <sub>26</sub> diol)	38%
1-(O-hexose)-3-keto-25-hexacosanol (HG <sub>26</sub> keto-ol)	17%
1-(O-hexose)-27-keto-3,25-octacosanediol (HG <sub>28</sub> keto-diol)	40%
1-(O-hexose)-3,27-octacosanediol (HG <sub>28</sub> diol)	1%
1-(O-hexose)-3,25,27-octacosanetriol (HG <sub>28</sub> triol)	5%

subject of several studies in the past and were shown to contain cyanobacterial communities and heterocyte glycolipids, suggesting the presence of heterocytous cyanobacteria (Severin and Stal, 2008; Severin et al., 2010; Bauersachs et al., 2011; Fan et al., 2015).

Lipid analysis of the microbial mat confirmed the presence of heterocyte glycolipids, albeit at very low concentrations. We detected presence of 1-(O-hexose)-3,25-hexacosanediol (HG<sub>26</sub> diol), 1-(O-hexose)-3-keto-25-hexacosanol (HG<sub>26</sub> keto-ol), and 1-(O-hexose)-27-keto-3,25-octacosanediol (HG<sub>28</sub> keto-diol), and also small amounts of 1-(O-hexose)-3,27-octacosanediol (HG<sub>28</sub> diol) and 1-(O-hexose)-3,25,27-octacosanetriol (HG<sub>28</sub> triol) (Table 3). Detection of HG<sub>26</sub> keto-ol and diol suggest the presence of cyanobacterial members of the *Nostocaceae* family (*Nostocaceae* and *Aphanizomenonaceae*), and presence of HG<sub>28</sub> keto-diol and triol

suggested the presence of members of the *Rivulariaceae* (*Calotrichaceae*) family (Table 1; Supplementary Table S1). While traces of HG<sub>28</sub> diol could also indicate the presence of organisms belonging to other cyanobacterial families such as *Fortiea contorta* PCC 7126 (formerly known as *Microchaete* sp.), *Stigonema ocellatum* SAG 48.90 or *Tolypothrix tenuis* PCC 7101. (Supplementary Table S1).

These HG results agree with the study of Bauersachs et al. (2011). These authors reported the presence of 1-(O-hexose)3,25-hexacosanediol (HG<sub>26</sub> diol), suggesting the presence of members of the *Nostocaceae* family (*Nostocaceae* and *Aphanizomenonaceae*), in supratidal microbial mats collected near our sampling site. Additionally, Bauersachs et al. (2011) also detected the presence of *Calothrix* sp. and *Nodularia* sp. (*Aphanizomenaceae*) in nearby sampling sites using 16S rRNA gene libraries, although HGs generally associated with *Calothrix* were not detected.

Next, we wanted to detect potential HG-producing heterocytous cyanobacteria in the microbial mat using our newly developed *hglT* primers. DNA from the freeze-dried microbial mat was extracted as described previously for the cyanobacterial cultures (Table 2). We carried out a PCR reaction on the microbial mat DNA extract using primer pair 2 (Fw1 mixB + Rv2) and the same annealing temperature that was selected for PCR on DNA extracts of cyanobacterial cultures (56°C annealing temperature, 1' 30" extension time, 30 cycles). Unfortunately, however, no PCR products were obtained from the microbial mat (Supplementary Figure S6). We carried out additional tests to rule out the presence of inhibitors that could seriously hinder the reaction, and to ensure that the selected sample contained DNA

belonging to cyanobacteria and nitrogen fixing bacteria (See [Supplementary Results](#) and [Supplementary Figures S7–S9](#)).

We performed a gradient PCR (48–58°C annealing temperature, 1'10" extension time, 37 cycles) on our microbial mat sample using the *hglT*-primer mix 2 (Fw1 mixB and Rv1) and increased the number of amplification cycles to try to obtain a PCR product. However, no clear band was obtained at any of the annealing temperatures tested ([Supplementary Figure S10](#)). Previous genetic studies showed that microbial mats collected from this location contained diazotrophic cyanobacteria, and lipid analysis indicated that heterocytous cyanobacteria were present in the microbial mat ([Severin et al., 2010](#); [Bauersachs et al., 2011](#)). However, our primers designed for the amplification of *hglT* gene fragments of heterocyte cyanobacteria did not yield any band when applied on the DNA extract of the microbial mat. This can be explained in two ways: (1) no heterocytous cyanobacteria were present in the microbial mat or they were present in very low abundances (2) our primers cannot target the heterocytous cyanobacterial species present in this microbial mat.

To answer this question, we carried out 16S rRNA gene amplicon sequencing analysis. Sequencing results showed that nearly half of the reads (54%) corresponded to Amplicon Sequence Variants (ASVs) ([Table 4](#)) of cyanobacteria. *Coleofasciculus chthonoplastes* PCC 7420, a non-heterocytous diazotroph, was the most relatively abundant (ra) cyanobacteria (43.7% ra). Other cyanobacterial species such as the nitrogen fixers *Lyngbya aestuarii* PCC-7419 (2.8% ra) and *Trichodesmium* IMS101 (0.28% ra) were also detected but no heterocytous cyanobacteria were detected ([Table 4](#)). Hence, we conclude that the amplification of *hglT* using our primers failed simply because heterocytous cyanobacteria were absent or present in very low abundances. Since the use of PCR-based approaches (including the use of universal 16S rRNA primers) typically favors amplification of more abundant microbial species, potentially masking the presence of those present in lower proportion, it is possible that the heterocytous cyanobacteria were present in the microbial mat but in too low abundances to be detected by our PCR methods. Alternatively, given the excellent preservation of heterocyte glycolipids over time ([Schaefer et al., 2020](#)), and the relatively short life of DNA when released in the environment after cell death, it is also plausible that the heterocyte glycolipids detected in the microbial mat were preserved remnants of dead heterocytous cyanobacteria whose DNA had been degraded. Therefore, both explanations would be consistent with the very low levels of HG lipids detected in the microbial mat and the absence of DNA from heterocytous cyanobacteria.

Next, we used a specimen of the floating macroalgae *S. fluitans* collected in the Central Atlantic Ocean. Its microbiome consists of 35.9% (ra) cyanobacteria, including several heterocytous cyanobacteria such as *Mastigocoleus* (11.7% ra) and *Rivularia* PCC 7116 (4.74% ra) ([Table 4](#); [Theirlynyck et al., 2023](#)), making it a good candidate to test our primers. Accordingly, we performed a gradient PCR (48–58°C annealing temperature, 1'10" extension time, 37 cycles) on this specimen using the *hglT*-primer mix 2 (Fw1 mixB and Rv1). This led to successful amplification of the partial *hglT* gene when using annealing temperatures above 54.2°C. The most intense bands were observed at 54.2 and 56°C ([Supplementary Figure S11](#)), these results are in line with previous results obtained from testing the primer mix on cyanobacterial cultures and support our choice to use 56°C

**TABLE 4** Relative abundances of the cyanobacterial ASVs\* present in the microbial mat and *S. fluitans* sample, obtained after running the Cascabel pipeline ([Abdala Asbun et al., 2020](#)).\*\*

ASV ID	Microbial Mat	<i>Sargassum fluitans</i> <sup>1</sup>
<i>Acrophormium</i> PCC 7375	–	3.69%
Chloroplast	3.96%	–
<i>Chroococcidiopsis</i> PCC 6712	–	1.34%
<i>Coleofasciculaceae</i>	0.11%	–
<i>Coleofasciculus</i> PCC 7420	0.47%	–
<i>Coleofasciculus</i> PCC 7420 <i>chthonoplastes</i>	43.69%	–
<i>Cyanobacteriales</i>	–	0.38%
<i>Cyanobacteriia</i>	–	4.46%
<i>Halomicronema</i> TFEP1	–	0.22%
<i>Leptolyngbya</i> PCC 6406	–	0.16%
<i>Limnothrix</i>	0.26%	<0.01%
<i>Lyngbya</i> PCC 7419	0.16%	–
<i>Lyngbya</i> PCC 7419 <i>aestuarii</i>	2.77%	–
<b><i>Mastigocoleus</i> BC008</b>	–	<b>0.47%</b>
<b><i>Mastigocoleus</i> BC008 <i>testarum</i></b>	–	<b>11.26%</b>
MBIC10086	4.04%	0.03%
<i>Nodosilinea</i> PCC 7104	0.12%	–
<i>Nodosilineaceae</i>	–	0.22%
<b><i>Nostocaceae</i></b>	–	<b>1.05%</b>
<i>Oscillatoria</i> PCC 6304 <i>acumiata</i>	0.21%	–
<i>Oscillatoriaceae</i>	0.06%	–
<i>Oxyphotobacteria</i> <i>Incertae Sedis</i>	1.65%	0.09%
<i>Phormidesmiaceae</i>	–	4.71%
<i>Phormidesmiales</i>	–	0.09%
<i>Phormidesmis</i> ANT.LACV5.1	–	0.50%
<i>Phormidium</i> MBIC10003 <i>lucidum</i>	0.11%	0.04%
<i>Pleurocapsa</i> PCC 7319	–	0.57%
<b><i>Rivularia</i> PCC 7116</b>	–	<b>4.74%</b>
<i>Spirulina</i> P7	0.06%	–
<i>Symphotheca</i> PCC 7002	0.09%	0.01%
<i>Synechococcus</i> CC9902	–	0.07%
<i>Synechococcus</i> PCC 7336	–	0.47%
<i>Trichodesmium</i> IMS101	0.28%	<0.01%
<i>Vampirovibrionaceae</i>	–	0.08%
Total % of cyanobacteria and chloroplasts	58.08%	34.87%

\*Only ASVs with relative abundances  $\geq 0.05$  in at least one of the samples are included.

\*\*Cyanobacteria known to synthesize heterocyte glycolipids are shown in bold.

<sup>1</sup>[Theirlynyck et al. \(2023\)](#).

(instead of 58.2°C) as the preferred annealing temperature for all cyanobacterial cultures.

Thus, we have shown that the primers designed in this study can target and amplify the *hglT* gene of a wide variety of cyanobacterial species in the environment. This includes species not considered

during primer design, such as *Mastigocoleous* and *Rivularia* PCC 7116, present in the microbiome of *S. fluitans*.

## Conclusion

We conclude that the primer set 2 designed in this study and consisting of an equimolar mix of seven forward primers (Fw1 mixB) and five reverse primers (Rv1) can successfully target and partially amplify the *hglT* gene of heterocytous cyanobacteria in cultures and the environment. However, it may also target non-heterocytous cyanobacteria such as *Synechococcus* albeit to a lesser extent. We conclude that our primers can be used qualitatively to assess the presence of heterocytous cyanobacteria. However, to ensure accurate results we suggest that (1) a temperature gradient PCR should be performed to ensure that the best annealing temperature for the sample analyzed is used, and (2) the PCR product should be sequenced to confirm the presence of the suspected heterocytous cyanobacteria in the sample.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#). The 16S rRNA amplicon sequencing data presented in the study was deposited in the European Nucleotide Archive (ENA) repository, accession number PRJEB65747. Partial *hglT* sequences and phylogenetic trees presented in the study were deposited in the Zenodo repository, accession <https://doi.org/10.5281/zenodo.8309792>. Further inquiries can be directed to the corresponding author.

## Author contributions

RP: Conceptualization, Formal analysis, Investigation, Resources, Visualization, Writing – original draft, Writing – review & editing. NB: Formal analysis, Resources, Writing – review & editing. JS: Funding acquisition, Writing – review & editing. LV: Resources, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no. 694569-MICROLIPIDS) to JS. LV and JS

## References

- Abdala Asbun, A., Besseling, M. A., Balzano, S., van Bleijswijk, J. D. L., Witte, H. J., Villanueva, L., et al. (2020). Cascabel: a scalable and versatile amplicon sequence data analysis pipeline delivering reproducible and documented results. *Front. Genet.* 11:489357. doi: 10.3389/fgene.2020.489357
- Abd-El Salam, K. A. (2003). Bioinformatic tools and guideline for PCR primer design. *Afr. J. Biotechnol.* 2, 91–95. doi: 10.4314/ajb.v2i5.14794
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., and Gascuel, O. (2011). Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast

received funding from the Soehngen Institute for Anaerobic Microbiology (SIAM) through a Gravitation Grant (024.002.002) from the Dutch Ministry of Education, Culture, and Science (OCW).

## Acknowledgments

We would like to thank Samuel Cirés for providing biomass of the *C. ovalisporum* strains UAM290 and UAM292, Koichiro Awai for providing the strain *Anabaena* sp. PCC 7120, Hoaxin Fan and Lucas Stal for providing the microbial mat samples, Tom Theirlynck and Linda Amaral-Zettler for providing the *S. fluitans* sample, Michele Grego for assistance handling the cyanobacterial cultures, Dina Castillo Boukhchtaber for her assistance preparing the microbial mat samples for 16S rRNA gene amplicon sequencing, and for carrying out the PCR on the *S. fluitans* sample, Lora Strack van Schijndel for her assistance extracting DNA from the microbial mat sample, and Bastiaan von Meijenfeldt for helpful discussions. We would also like to thank two referees for comments on an earlier version of this paper.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1257040/full#supplementary-material>

Likelihood-based Approximation Schemes. *Syst Biol* 60, 685–699. doi: 10.1093/sysbio/syr041

Awai, K., and Wolk, C. P. (2007). Identification of the glycosyl transferase required for synthesis of the principal glycolipid characteristic of heterocysts of *Anabaena* sp. strain PCC 7120. *FEMS Microbiol. Lett.* 266, 98–102. doi: 10.1111/j.1574-6968.2006.00512.x

Bale, N. J., Hopmans, E. C., Dorhout, D., Stal, L. J., Grego, M., van Bleijswijk, J., et al. (2018a). A novel heterocyst glycolipid detected in a pelagic N<sub>2</sub>-fixing cyanobacterium of the genus *Calothrix*. *Org. Geochem.* 123, 44–47. doi: 10.1016/j.orggeochem.2018.06.009

Bale, N. J., Villareal, T. A., Hopmans, E. C., Brussaard, C. P. D., Besseling, M., Dorhout, D., et al. (2018b). C<sub>5</sub> glycolipids of heterocytous cyanobacteria track symbiont

- abundance in the diatom *Hemiaulus hauckii* across the tropical North Atlantic. *Biogeosciences* 15, 1229–1241. doi: 10.5194/bg-15-1229-2018
- Bauersachs, T., Compaoré, J., Hopmans, E. C., Stal, L. J., Schouten, S., and Sinninghe Damsté, J. S. (2009a). Distribution of heterocyst glycolipids in cyanobacteria. *Phytochemistry* 70, 2034–2039. doi: 10.1016/j.phytochem.2009.08.014
- Bauersachs, T., Compaoré, J., Severin, I., Hopmans, E. C., Schouten, S., Stal, L. J., et al. (2011). Diazotrophic microbial community of coastal microbial mats of the southern North Sea. *Geobiology* 9, 349–359. doi: 10.1111/j.1472-4669.2011.00280.x
- Bauersachs, T., Hopmans, E. C., Compaoré, J., Stal, L. J., Schouten, S., and Damsté, J. S. S. (2009b). Rapid analysis of long-chain glycolipids in heterocystous cyanobacteria using high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 23, 1387–1394. doi: 10.1002/rcm.4009
- Bauersachs, T., Miller, S. R., Gugger, M., Mudimu, O., Friedl, T., and Schwark, L. (2019). Heterocyste glycolipids indicate polyphyly of stigonematalean cyanobacteria. *Phytochemistry* 166:112059. doi: 10.1016/j.phytochem.2019.112059
- Bauersachs, T., Miller, S. R., van der Meer, M. T. J., Hopmans, E. C., Schouten, S., and Sinninghe Damsté, J. S. (2013). Distribution of long chain heterocyst glycolipids in cultures of the thermophilic cyanobacterium *Mastigocladus laminosus* and a hot spring microbial mat. *Org. Geochem.* 56, 19–24. doi: 10.1016/j.orggeochem.2012.11.013
- Bauersachs, T., Mudimu, O., Schulz, R., and Schwark, L. (2014). Distribution of long chain heterocyst glycolipids in N<sub>2</sub>-fixing cyanobacteria of the order Stigonematales. *Phytochemistry* 98, 145–150. doi: 10.1016/j.phytochem.2013.11.007
- Bauersachs, T., Rochelmeier, J., and Schwark, L. (2015). Seasonal lake surface water temperature trends reflected by heterocyst glycolipid-based molecular thermometers. *Biogeosciences* 12, 3741–3751. doi: 10.5194/bg-12-3741-2015
- Bauersachs, T., Speelman, E. N., Hopmans, E. C., Reichart, G.-J., Schouten, S., and Damsté, J. S. S. (2010). Fossilized glycolipids reveal past oceanic N<sub>2</sub> fixation by heterocystous cyanobacteria. *Proc. Natl. Acad. Sci.* 107, 19190–19194. doi: 10.1073/pnas.1007526107
- Bauersachs, T., Talbot, H. M., Sidgwick, F., Sivonen, K., and Schwark, L. (2017). Lipid biomarker signatures as tracers for harmful cyanobacterial blooms in the Baltic Sea. *PLoS One* 12:e0186360. doi: 10.1371/journal.pone.0186360
- Berrendero Gómez, E., Johansen, J. R., Kaštokský, J., Bohunická, M., and Čapková, K. (2016). Macrochaete gen. Nov. (Nostocales, cyanobacteria), a taxon morphologically and molecularly distinct from *Calothrix*. *J. Phycol.* 52, 638–655. doi: 10.1111/jpy.12425
- Besseling, M. A., Hopmans, E. C., Boschman, R. C., Sinninghe Damsté, J. S., and Villanueva, L. (2018). Benthic archaea as potential sources of tetraether membrane lipids in sediments across an oxygen minimum zone. *Biogeosciences* 15, 4047–4064. doi: 10.5194/bg-15-4047-2018
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Castenholz, R. W., Wilmotte, A., Herdman, M., Rippka, R., Waterbury, J. B., Iteman, I., et al. (2001). “Phylum BX. Cyanobacteria” in *Bergey’s manual® of systematic bacteriology: Volume one: the archaea and the deeply branching and phototrophic bacteria*. eds. D. R. Boone, R. W. Castenholz and G. M. Garrity (New York, NY: Springer), 473–599.
- Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi: 10.1093/nar/gkaa939
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fan, H., Bolhuis, H., and Stal, L. J. (2015). Nitrification and nitrifying bacteria in a coastal microbial mat. *Front. Microbiol.* 6:1367. doi: 10.3389/fmicb.2015.01367
- Gambacorta, A., Pagnotta, E., Romano, I., Sodano, G., and Trincone, A. (1998). Heterocyst glycolipids from nitrogen-fixing cyanobacteria other than nostocaceae. *Phytochemistry* 48, 801–805. doi: 10.1016/S0031-9422(97)00954-0
- Gambacorta, A., Romano, I., Trincone, A., Soriente, A., Giordano, M., and Sodano, G. (1996). Heterocyst glycolipids from five nitrogen-fixing cyanobacteria. *Gazz. Chim. Ital.* 126, 653–656.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Halimatul, H. S. M., Ehira, S., and Awai, K. (2014). Fatty alcohols can complement functions of heterocyst specific glycolipids in *Anabaena* sp. PCC 7120. *Biochem. Biophys. Res. Commun.* 450, 178–183. doi: 10.1016/j.bbrc.2014.05.093
- Haselkorn, R. (1978). Heterocysts. *Annu. Rev. Plant Physiol.* 29, 319–344. doi: 10.1146/annurev.pp.29.060178.001535
- Hillis, D. M., and Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Syst. Biol.* 42, 182–192. doi: 10.1093/sysbio/42.2.182
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35, W43–46. doi: 10.1093/nar/gkm234
- Klugkist, J., and Haaker, H. (1984). Inhibition of nitrogenase activity by ammonium chloride in *Azotobacter vinelandii*. *J. Bacteriol.* 157, 148–151.
- Knutson, C. M., Plunkett, M. H., Liming, R. A., and Barney, B. M. (2018). Efforts toward optimization of aerobic biohydrogen reveal details of secondary regulation of biological nitrogen fixation by nitrogenous compounds in *Azotobacter vinelandii*. *Appl. Microbiol. Biotechnol.* 102, 10315–10325. doi: 10.1007/s00253-018-9363-0
- Komarek, J., Kastovsky, J., Mares, J., and Johansen, J. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86, 295–335.
- Komárek, J., Santanna, C. L., Bohunická, M., Mareš, J., Hentschke, G. S., Rigonato, J., et al. (2013). Phenotype diversity and phylogeny of selected Scytonema-species (Cyanoprokaryota) from SE Brazil. *Fottea* 13, 173–200. doi: 10.5507/fot.2013.015
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., et al. (2012). The biological observation matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 1:7. doi: 10.1186/2047-217X-1-7
- Millineaux, P. M., Gallon, J. R., and Chaplin, A. E. (1981). Acetylene reduction (nitrogen fixation) by cyanobacteria grown under alternating light-dark cycles. *FEMS Microbiol. Lett.* 10, 245–247. doi: 10.1111/j.1574-6968.1981.tb06249.x
- Mishra, A., Rajput, S., Gupta, P. S., Goyal, V., Singh, S., Sharma, S., et al. (2021). “Role of cyanobacteria in Rhizospheric nitrogen fixation” in *Soil nitrogen ecology soil biology*. eds. C. Cruz, K. Vishwakarma, D. K. Choudhary and A. Varma (Cham: Springer International Publishing), 497–519.
- Nübel, U., Garcia-Pichel, F., and Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl. Environ. Microbiol.* 63, 3327–3332.
- Olson, J. B., Steppe, T. F., Litaker, R. W., and Paerl, H. W. (1998). N<sub>2</sub>-fixing microbial consortia associated with the ice cover of Lake Bonney, Antarctica. *Microb. Ecol.* 36, 231–238. doi: 10.1007/s002489900110
- Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111, 1–61. doi: 10.1099/00221287-111-1-1
- Schaefer, B., Grice, K., Coolen, M. J. L., Summons, R. E., Cui, X., Bauersachs, T., et al. (2020). Microbial life in the nascent Chicxulub crater. *Geology* 48, 328–332. doi: 10.1130/G46799.1
- Schouten, S., Villareal, T. A., Hopmans, E. C., Mets, A., Swanson, K. M., and Sinninghe Damsté, J. S. (2013). Endosymbiotic heterocystous cyanobacteria synthesize different heterocyst glycolipids than free-living heterocystous cyanobacteria. *Phytochemistry* 85, 115–121. doi: 10.1016/j.phytochem.2012.09.002
- Severin, I., Acinas, S. G., and Stal, L. J. (2010). Diversity of nitrogen-fixing bacteria in cyanobacterial mats. *FEMS Microbiol. Ecol.* 73, 514–525. doi: 10.1111/j.1574-6941.2010.00925.x
- Severin, I., and Stal, L. J. (2008). Light dependency of nitrogen fixation in a coastal cyanobacterial mat. *ISME J.* 2, 1077–1088. doi: 10.1038/ismej.2008.63
- Svenning, M. M., Eriksson, T., and Rasmussen, U. (2005). Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. *Arch. Microbiol.* 183, 19–26. doi: 10.1007/s00203-004-0740-y

- Tel-Or, E., Stewart, W. D. P., and Fogg, G. E. (1997). Photosynthetic components and activities of nitrogen-fixing isolated heterocysts of *Anabaena cylindrica*. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 198, 61–86. doi: 10.1098/rspb.1977.0086
- Theirlynck, T., Mendonça, I. R. W., Engelen, A. H., Bolhuis, H., Collado-Vides, L., van Tussenbroek, B. I., et al. (2023). Diversity of the holopelagic Sargassum microbiome from the great Atlantic Sargassum Belt to coastal stranding locations. *Harmful Algae* 122:102369. doi: 10.1016/j.hal.2022.102369
- Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* 1:e00009–15. doi: 10.1128/mSystems.00009-15
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Wolk, C. P., Ernst, A., and Elhai, J. (1994). “Heterocyst metabolism and development” in *The molecular biology of cyanobacteria advances in photosynthesis*. ed. D. A. Bryant (Dordrecht: Springer Netherlands), 769–823.
- Wörmer, L., CirĚs, S., Velázquez, D., Quesada, A., and Hinrichs, K.-U. (2012). Cyanobacterial heterocyst glycolipids in cultures and environmental samples: diversity and biomarker potential. *Limnol. Oceanogr.* 57, 1775–1788. doi: 10.4319/lo.2012.57.6.1775
- Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* 19, 162–173. doi: 10.1016/j.tim.2010.12.004