

PhD

3rd  
CYCLE

FCUP  
2017

U. PORTO

Environmental and applied aspects of cyanobacteria  
diversity and the importance of its preservation

Vitor Manuel Capela  
Ramos

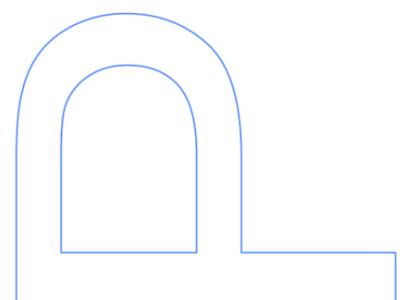
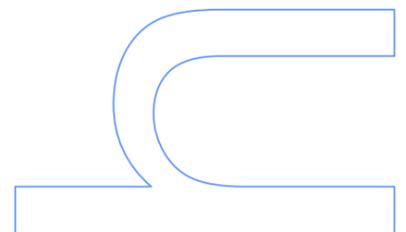
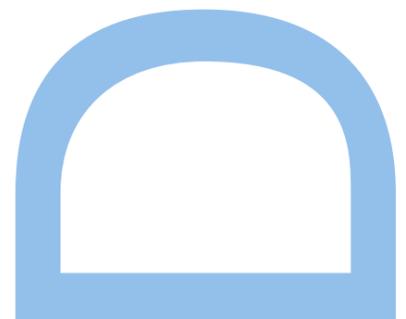
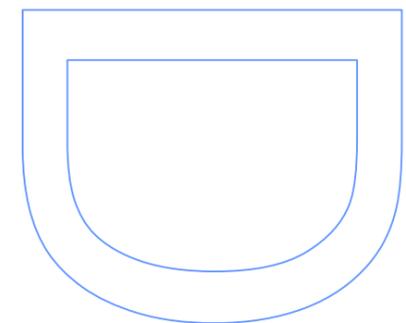
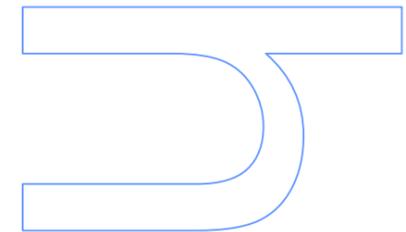
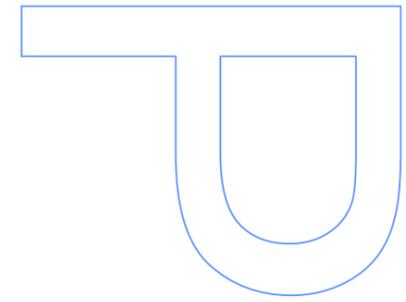
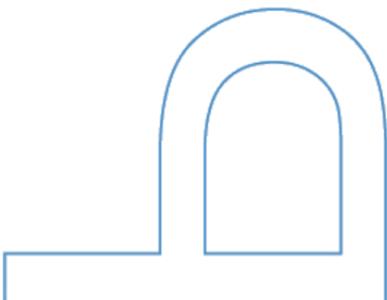
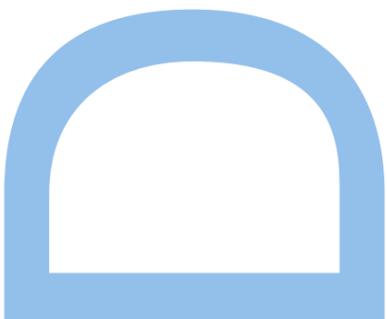
FC

U. PORTO  
FC FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO

# Environmental and applied aspects of cyanobacteria diversity and the importance of its preservation

Vitor Manuel Capela Ramos  
PhD thesis presented to the  
Faculty of Sciences of the University of Porto  
Environmental Sciences and Technology  
2017

U. PORTO  
FC FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO



Vitor Manuel Capela Ramos

# Environmental and applied aspects of cyanobacteria diversity and the importance of its preservation

Supervisor: Vitor Vasconcelos



Dissertation presented to the Faculty of Sciences from the University of Porto to obtain the PhD in Environmental Sciences and Technology

Dissertação apresentada à Faculdade de Ciências da Universidade do Porto para obtenção do grau de Doutor em Ciências e Tecnologia do Ambiente

Porto, 2017

*To my wife and children*

## Acknowledgements

First I would like to thank my advisor, Professor Vitor Vasconcelos, for his endless support, encouragement and empowerment, and for providing his thoughtful guidance and expertise over the years as a mentor, teacher and friend.

I would like to thank Blue Biotechnology and Ecotoxicology (BBE) laboratory at Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) for the infrastructures and equipment facilities.

This work was supported by the Structured Program of R&D&I INNOVMAR - Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER- 000035, Research Line NOVELMAR), funded by the Northern Regional Operational Program (NORTE2020) through the European Regional Development Fund (ERDF). It was also supported by NoMorFilm project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 634588.

I acknowledge the support given by Fundação para a Ciência e Tecnologia (FCT), through my PhD fellowship (SFRH/BD/80153/2011), which was supported by POPH/FSE funding.

To my colleagues in BBE, thank you for your support and friendship, and above all for keeping an excellent working environment. I owe debts of gratitude to João Morais, Raquel Castelo Branco, Pedro Leão and Micaela Vale who, in different ways, have helped me in this work.

A special thank to Ângela Brito from IBMC/i3S and to Frederico Sobrinho and Sinda Carvalhal-Gomes from UFRJ.

Finally, I especially thank my family for your love, patience and inspiration, and for being my biggest supporters.



## Previous note

The thesis is presented as a series of linked chapters, each including one or more sections. Each section is a manuscript. Given that such work was performed in collaboration with other authors, the candidate clarifies that, in all of them actively participated in its design, data gathering and analysis, discussion of results, as well as in preparing the manuscripts. A general discussion integrates and synthesizes the work, along with main conclusions. In the end of the thesis, appendices give access to supplementary information to the main document.

### List of manuscripts meant for publication

#### In peer-reviewed scientific journals:

- Ramos V., Morais J., Vasconcelos V.M. A curated database of cyanobacterial strains relevant for modern taxonomy and phylogenetic studies (submitted)

- Ramos V.\*, Castelo-Branco R.\*, Leão P.N., Martins J., Carvalhal-Gomes S., da Silva F.S., Filho J.G.M., Vasconcelos V.M. Cyanobacterial diversity in microbial mats from the hypersaline lagoon system of Araruama, Brazil: an in-depth polyphasic study (submitted)

\* these authors contributed equally to this work

- Brito Â., Ramos V., Mota R., Lima S., Santos A., Vieira J., Vieira C.P., Kaštovský J., Vasconcelos V.M., Tamagnini P. Description of new genera and species of marine cyanobacteria isolated from the Portuguese Atlantic coast (submitted)

- Ramos V.\*, Morais J.\*, Castelo-Branco R.\*, Pinheiro Â., Martins J., Regueiras A., Pereira A.L., Lopes V., Frazão B., Gomes D., Moreira C., Costa M.S., Brûle S., Faustino S., Martins R., Saker M., Osswald J., Leão P.N., Vasconcelos V.M. Cyanobacterial diversity held in mBRCs as a biotechnological asset: the case study of the newly established LEGE Culture Collection (submitted)

\* these authors contributed equally to this work

#### In peer-reviewed book chapter:

- Ramos V., Moreira C., Mankiewicz-Boczek J., Vasconcelos V.M. Application of molecular tools in monitoring cyanobacteria and their potential toxin production (submitted)

## Sumário

As cianobactérias são um grupo de procariotas fotossintéticos conhecidos pela sua grande biodiversidade e quimiodiversidade, ainda que a sua verdadeira dimensão esteja ainda longe de conhecida. De facto, a sistemática e metodologias recentes para a identificação destes microorganismos foram capazes de expôr todo um mundo novo relativamente a estas matérias. O valor intrínseco da diversidade cianobacteriana advém não só da sua importância enquanto fornecedores de bens e serviços ambientais, mas também da sua utilização, direta ou indireta, enquanto recurso natural. Por outro lado, algumas cianobactérias geram custos económicos, uma vez que causam efeitos nocivos ao ambiente, incluindo o ser humano. Assim sendo, continua a ser importante explorar a sua diversidade biológica e química, algo que deverá ser realizado não só a partir de amostras colhidas diretamente na natureza, mas também a partir de isolados em Coleções de Cultura mundiais, uma vez que podem conter diversidade escondida. Isto constitui o racional para esta tese, cujas questões de investigação se focam na implicação das abordagens metodológicas e das classificações taxonómicas para um melhor conhecimento da diversidade real das cianobactérias, e na relevância da existência de isolados e sua preservação para fins taxonómicos, ambientais ou aplicados.

Uma vez que as estirpes de referência em Coleções de Cultura são particularmente importantes para fins comparativos em estudos taxonómicos e filogenéticos, e em função do atual estado da sistemática das cianobactérias, foi criada uma base de dados *online* onde facilmente se podem encontrar dados e metadados de estirpes relevantes, do ponto de vista taxonómico, genómico ou filogenómico, de cianobactérias.

O estudo de isolados cianobacterianos marinhos, intertidais e de tapetes microbianos hipersalinos, revelaram uma rica e inexplorada diversidade cianobacteriana. A reexaminação e estudo taxonómico realizado às estirpes intertidais levaram à proposta de 5 novas taxa de cianobactérias marinhas, representando 3 novos géneros e 2 novas espécies. As diferentes abordagens metodológicas e atribuições taxonómicas usadas para estudar os tapetes microbianos mostraram diferir bastante na sua capacidade de expôr a diversidade cianobacteriana presente nas amostras. Para se conseguir um panorama mais completo da diversidade presente, a complementaridade entre essas abordagens demonstrou ser essencial. Mostrou-se ainda que a maioria dos isolados presentes nesses dois ambientes possuem o potencial de produzir metabolitos secundários, algo que é adicionalmente relevante no caso dos isolados hipersalinos, já que as cianobactérias oriundas destes ambientes extremos continuam pouco estudadas enquanto possível fonte de novos compostos bioactivos.

A descoberta e anotação de agrupamentos de genes que codificam para cianotoxinas permitiram o desenvolvimento e utilização de diferentes abordagens moleculares para monitorizar cianobactérias potencialmente nefastas, as quais são revistas nesta tese. Devido à sua especificidade, é dada especial atenção à vigilância de suplementos alimentares, incluindo

o desenvolvimento e disponibilização de Procedimentos Operacionais Padrão (POPs) para serem utilizados por profissionais que trabalhem no sector da segurança alimentar e controlo de qualidade.

Finalmente, e tomando a Coleção de Cultura LEGE como caso de estudo, é desmacarado o processo de conversão de um grupo de isolados avulsos numa Coleção de Cultura. Outros resultados importantes que resultaram desse processo foram as criações de uma base de dados *online* para pesquisa da coleção, bem como o catálogo de estirpes LEGE. Além do mais, através da combinação e integração de dados biológicos e químicos e do uso de árvores filogenéticas, foi demonstrada a sumptuosa bio- e quimiodiversidade presente na coleção. A partir das árvores mostra-se que, apesar dos dados sobre a diversidade filogenética possam ser boas ferramentas para guiar o processo de seleção de estirpes para despistagem de produtos naturais, os programas de descoberta de novos compostos não devem negligenciar uma avaliação estirpe-a-estirpe.

No seu conjunto, esta tese reforça a pertinência da utilização de sistemas de classificação robustas e filogeneticamente inferidas, e de métodos apropriados e complementares (p.e. abordagem polifásica), para que se consiga uma maior e melhor compreensão sobre a diversidade cianobacteriana. Destaca também a relevância do isolamento e do cultivo de cianobactérias, evidenciando o papel central desempenhado pelas Coleções de Culturas na caracterização e preservação desse recurso natural, para melhor responderem às necessidades das várias áreas de investigação.

## Summary

Cyanobacteria are a group of photosynthetic prokaryotes recognized for their wide biodiversity and chemodiversity, whose true extents are far from known. Indeed, the modern systematics and methodologies for the identification of these microorganisms exposed a whole new world related to these issues. The intrinsic value of cyanobacterial diversity arises not only from the importance of cyanobacteria as providers of ecological goods and services, but also from their use as direct or indirect natural resource. On the other hand, some cyanobacteria generate economic costs to society, since they may cause deleterious effects to the environment, including humans. Therefore, it continues to be important to explore the bio- and chemodiversity of these microorganisms, something that should be performed not only from natural samples, but also from isolates present in Culture Collections worldwide, which might contain hidden diversity. This constituted the rationale for this thesis, whose main research questions focus on the implications of methodological approaches and taxonomic classifications to achieve a greater knowledge of the real diversity of cyanobacteria, and on the relevance of the existence and preservation of isolates for taxonomic, environmental and applied purposes.

Since reference strains held in Culture Collections are particularly important for comparative purposes in taxonomic or phylogenetic studies, and due to the current problematic status of cyanobacterial systematics, I have created a curated, online database where data and metadata for relevant cyanobacterial strains, from the taxonomic, genomic or phylogenetic point of view, can be easily found.

The study of marine, intertidal isolates and hypersaline microbial mats, respectively from the Portuguese coast and from Brazilian lagoons, exposed a rich, still unexplored cyanobacterial biodiversity. A reexamination and taxonomic study made for the intertidal isolates led to the proposal of 5 new cyanobacterial taxa, referring to 3 new genera and 2 new species. The different methodological approaches and taxonomic assignments used to study the hypersaline mats varied radically in their ability to capture the cyanobacterial diversity present in the samples. The complementarity between those approaches proved to be necessary to attain the most complete picture of the cyanobacterial diversity. It was demonstrated that the majority of the cyanobacterial isolates obtained from these two environments have the potential to produce secondary metabolites, which is additionally relevant for the hypersaline isolates, since cyanobacteria from these extreme environments are still understudied as a source for new bioactive compounds.

The discovery and annotation of modular biosynthetic gene clusters that encode for cyanotoxins allowed the development and implementation of different molecular-based approaches for the monitoring of potentially hazardous cyanobacteria, which are reviewed on this thesis. Due to its specificity, a special focus is given to food supplement surveillance, including the development

and provision of Standard Operation Procedures (SOPs) to be used by professionals working in food safety/quality assurance.

Finally, taking the LEGE Culture Collection as a case study, the process of turning in-house cyanobacterial isolates into a Culture Collection is unmasked. Other important outcomes from this process were the establishment of its associated searchable, online database, and the creation of the catalog of strains. Also, by integrating biological and chemical data, the rich and diverse bio- and chemodiversity present in the collection is exposed through phylogenetic trees. From these trees it is shown that, even though the use of phylogenetic diversity data is a good way for directing the strains selection for natural products screening, strain-by-strain assessments should not be neglected by discovery programs.

Overall, this thesis reinforces the pertinence of using accurate, phylogenetically derived classifications, and of appropriate and complementary methods (e.g. polyphasic approach) for the full understanding of the cyanobacterial diversity. It also highlights the relevance of isolation and cultivation of cyanobacteria, and underlines the pivotal role played by Cultures Collections in characterizing and preserving this bioresource, in order to better supply the needs of taxonomic, ecological, environmental or biotechnological researches.

# Index

## Contents

|  |             |
|--|-------------|
| <b>Acknowledgements .....</b>  | <b>i</b>    |
| <b>Previous note.....</b>  | <b>ii</b>   |
| <b>Sumário .....</b>   | <b>iii</b>  |
| <b>Summary .....</b>   | <b>v</b>    |
| <b>Index.....</b>  | <b>vii</b>  |
| <b>Contents.....</b>   | <b>vii</b>  |
| <b>Table Index.....</b>  | <b>x</b>    |
| <b>Figure Index .....</b>  | <b>xi</b>   |
| <b>Abbreviations.....</b>  | <b>xiii</b> |
| <b>1. General Introduction.....</b>  | <b>1</b>    |
| <b>1.1 Cyanobacterial diversity .....</b>  | <b>1</b>    |
| <b>1.2 Ecological, environmental, and economic issues related to cyanobacteria .....</b>                           | <b>2</b>    |
| <b>1.3 Preservation of microbial diversity and the need for strains .....</b>                                      | <b>4</b>    |
| 1.3.1 <i>Ex situ</i> preservation of cyanobacterial diversity .....  | 5           |
| <b>1.4 Objectives and Thesis Structure.....</b>  | <b>6</b>    |
| <b>1.5 References.....</b>   | <b>7</b>    |
| <b>2. Classification matters: Its key role in understanding diversity .....</b>                                    | <b>10</b>   |
| <b>2.1 A curated database of cyanobacterial strains relevant for modern taxonomy and phylogenetic studies.....</b> | <b>10</b>   |
| 2.1.1 Abstract.....  | 10          |
| 2.1.2 Background & Summary .....   | 10          |
| 2.1.3 Methods .....  | 12          |
| 2.1.4 Data Records .....   | 17          |
| 2.1.5 Technical Validation .....   | 20          |
| 2.1.6 Acknowledgements .....   | 20          |
| 2.1.7 References.....  | 20          |

|   |            |
|---|------------|
| <b>2.2 Description of new genera and species of marine cyanobacteria isolated from the Portuguese Atlantic coast .....</b>                      | <b>22</b>  |
| 2.2.1 Abstract.....   | 22         |
| 2.2.2 Introduction .....  | 23         |
| 2.2.3 Materials and methods .....   | 24         |
| 2.2.4 Results and discussion.....   | 27         |
| 2.2.5 Conclusions.....  | 45         |
| 2.2.6 Acknowledgments .....   | 46         |
| 2.2.7 Supplementary Data .....  | 47         |
| 2.2.8 References.....   | 51         |
| <b>2.3 Cyanobacterial diversity in microbial mats from the hypersaline lagoon system of Araruama, Brazil: an in-depth polyphasic study.....</b> | <b>56</b>  |
| 2.3.1 Abstract.....   | 56         |
| 2.3.2 Introduction .....  | 57         |
| 2.3.3 Materials and Methods .....   | 59         |
| 2.3.4 Results.....  | 66         |
| 2.3.5 Discussion.....   | 77         |
| 2.3.6 Acknowledgments .....   | 83         |
| 2.3.7 Supplementary Data .....  | 84         |
| 2.3.8 References.....   | 98         |
| <b>3. Monitoring cyanobacteria and their environmental disservices .....</b>  | <b>102</b> |
| <b>3.1 Application of molecular tools in monitoring cyanobacteria and their potential toxin production .....</b>                                | <b>102</b> |
| 3.1.1 Abstract.....   | 102        |
| 3.1.2 Introduction .....  | 102        |
| 3.1.3 Possible Applications .....   | 104        |
| 3.1.4 Checklist of Publications, Applications and Lessons from Practice.....  | 112        |
| 3.1.5 General conclusions.....  | 121        |
| 3.1.6 Acknowledgements .....  | 121        |
| 3.1.7 References.....   | 121        |
| <b>4. Preserving cyanobacterial diversity as a natural resource .....</b>   | <b>130</b> |

|  |            |
|--|------------|
| <b>4.1 Cyanobacterial diversity held in mBRCs as a biotechnological asset: the case study of the newly established LEGE Culture Collection .....</b> | <b>130</b> |
| 4.1.1 Abstract.....  | 130        |
| 4.1.2 Introduction .....   | 131        |
| 4.1.3 Material and Methods.....  | 132        |
| 4.1.4 Results and Discussion .....   | 136        |
| 4.1.5 Conclusions .....  | 146        |
| 4.1.6 Acknowledgements .....   | 147        |
| 4.1.7 Supplementary Data .....   | 148        |
| 4.1.8 References.....  | 149        |
| <b>5. General Discussion .....</b>   | <b>153</b> |
| <b>5.1 Final remarks.....</b>  | <b>157</b> |
| <b>5.2 References.....</b>   | <b>158</b> |
| <b>Appendices .....</b>  | <b>161</b> |
| Appendix A.....  | 161        |
| Appendix B.....  | 162        |
| Appendix C .....   | 163        |
| Appendix D .....   | 164        |

## Table Index

|                   |  |     |
|-------------------|--|-----|
| <b>Table 2.1</b>  | Number of cyanobacterial strains included in the dataset of CyanoType (version 1.0), by category   | 13  |
| <b>Table 2.2</b>  | Predicted secondary structure of the 16S-23S rRNA internal transcribed spacer (ITS) of <i>Geminobacterium atlanticum</i> LEGE 07459  | 29  |
| <b>Table 2.3</b>  | Comparison of the predicted secondary structures of the 16S-23S rRNA internal transcribed spacers (ITS) from our isolate <i>Jaaginema litorale</i> LEGE 07176 and from other strains | 41  |
| <b>Table 2.4</b>  | Cyanobacterial strains, culture collection code and sampling sites   | 47  |
| <b>Table 2.5</b>  | Oligonucleotide primers used in this study   | 47  |
| <b>Table 2.6</b>  | Sequences of the conserved and non-conserved regions of the 16S-23S rRNA internal transcribed spacer (ITS) of the seven LEGE isolates  | 48  |
| <b>Table 2.7</b>  | Physicochemical parameters of the three studied sites located in the Araruama lagoon complex   | 61  |
| <b>Table 2.8</b>  | Number of cyanobacterial 16S rRNA gene sequences used in or discarded from phylogeny, by sample  | 65  |
| <b>Table 2.9</b>  | Species composition and morphological-based characterization of cyanobacteria observed in the upper layer of the microbial mats collected at the Araruama's lagoons                  | 67  |
| <b>Table 2.10</b> | List of cyanobacterial strains isolated from hypersaline microbial mats collected at the Araruama lagoon system (RJ, Brazil)   | 71  |
| <b>Table 2.11</b> | Taxon richness comparison, by taxonomic order, of morphospecies and phylotypes identified in this study and morphospecies previously reported  | 76  |
| <b>Table 2.12</b> | Diversity estimates, considering different categorizations of taxa and/or molecular data processing  | 77  |
| <b>Table 2.13</b> | Checklist of cyanobacteria species reported for different water bodies from the Araruama lagoon system, as depicted from this study or as retrieved from the literature survey       | 84  |
| <b>Table 2.14</b> | Percentage of pyrosequencing reads, by sample, included in each phylotype identified in the phylogenetic tree  | 90  |
| <b>Table 3.1</b>  | List of studies showing the main different applications in which the detection/identification of tox(igen)ic cyanobacteria have been pursued, by molecular-based method used         | 107 |
| <b>Table 3.2</b>  | Examples of other applications/studies where molecular methods are being applied for the detection/identification of cyanobacteria and of microbial communities related to them      | 120 |
| <b>Table 4.1</b>  | Number of cyanobacterial strains, by genera, in LEGE CC  | 137 |

## Figure Index

|                    |   |    |
|--------------------|---|----|
| <b>Figure 2.1</b>  | Diagram illustrating the workflow followed during the construction and release of the dataset   | 12 |
| <b>Figure 2.2</b>  | Example of the use of the proposed subset of strains representing the cyanobacterial “tree of life” to evaluate the phylogenetic placement of strains not included in the main tree due to having short 16S rRNA gene sequences | 16 |
| <b>Figure 2.3</b>  | Micrographs of <i>Geminobacterium atlanticum</i> LEGE 07459   | 28 |
| <b>Figure 2.4</b>  | Maximum likelihood phylogenetic tree based on partial 16S rRNA gene sequences   | 31 |
| <b>Figure 2.5</b>  | Similarity between cyanobacterial strains based on 16S rRNA gene sequences  | 32 |
| <b>Figure 2.6</b>  | Micrographs of <i>Hyella patelloides</i> LEGE 07179   | 33 |
| <b>Figure 2.7</b>  | Micrographs of three strains of <i>Lusitaniella coriacea</i> LEGE 07167, LEGE 07157, and LEGE 06111   | 35 |
| <b>Figure 2.8</b>  | Micrographs of <i>Jaaginema litorale</i> LEGE 07176   | 38 |
| <b>Figure 2.9</b>  | Unrooted maximum-likelihood phylogram for 16S-23S rRNA ITS sequences of <i>Jaaginema litorale</i> LEGE 07176 and other strains within the same sub-clade of the 16S rRNA-based phylogenetic tree                                | 42 |
| <b>Figure 2.10</b> | Micrographs of <i>Calenema singularis</i> LEGE 06188  | 43 |
| <b>Figure 2.11</b> | Absorption spectra of cell-free extracts of the seven cyanobacterial strains described in this study  | 49 |
| <b>Figure 2.12</b> | Maximum likelihood phylogenetic tree based on the 16S-ITS-23S rRNA gene sequences   | 50 |
| <b>Figure 2.13</b> | Maximum likelihood phylogenetic tree based on partial 16S rRNA gene sequences including all <i>Jaaginema</i> sequences available  | 51 |
| <b>Figure 2.14</b> | Location map of the Araruama’s complex and sampling sites   | 59 |
| <b>Figure 2.15</b> | Sampling sites and sample mats  | 60 |
| <b>Figure 2.16</b> | Schematic overview of the experimental design   | 61 |
| <b>Figure 2.17</b> | Epifluorescence and bright field micrographs showing ubiquitous, abundant, or dominant cyanobacteria in the environmental samples   | 69 |
| <b>Figure 2.18</b> | Cyanobacterial isolates obtained in this study  | 70 |
| <b>Figure 2.19</b> | Circular phylogenetic ML trees based on cyanobacterial 16S rDNA sequences   | 73 |
| <b>Figure 2.20</b> | Venn diagrams showing the number of distinct cyanobacterial taxa distinguished in each mat sample, by different approaches  | 74 |
| <b>Figure 2.21</b> | Polyacrylamide gel image showing the DGGE bands excised for cloning and sequencing  | 91 |
| <b>Figure 2.22</b> | Light micrographs showing unicellular and colonial morphospecies from the cyanobacterial mat layers   | 92 |
| <b>Figure 2.23</b> | Light micrographs showing filamentous non-heterocystous morphospecies   | 93 |

|                    |   |     |
|--------------------|---|-----|
| <b>Figure 2.24</b> | Light micrographs showing filamentous non-heterocystous Cyanobacteria and Proteobacteria  | 94  |
| <b>Figure 2.25</b> | Number and relative abundance of OTUs (clustered at 98% similarity) generated by the SILVAngs pipeline, at the phylum level   | 94  |
| <b>Figure 2.26</b> | Krona charts screenshots illustrating the metagenomics results for the mat samples collected at EB1, EB2 and EB3.   | 95  |
| <b>Figure 2.27</b> | Complementary ML tree based on a selection of cyanobacterial 16S rDNA sequences from this study, from the study of Clementino et al. (2008) and from <i>Oxynema</i> spp.  | 96  |
| <b>Figure 2.28</b> | Agarose gel electrophoresis image of PCR products for PKS genes, and BLAST results for the obtained sequences   | 97  |
| <b>Figure 3.1</b>  | Evolution of the number of new sequences, deposited in the GenBank nucleotide database, for genes involved in the biosynthesis of cyanotoxins   | 103 |
| <b>Figure 3.2</b>  | Percentage of publications on toxic cyanobacteria that used molecular methods, until mid-2016; by cyanotoxin  | 104 |
| <b>Figure 3.3</b>  | Overview of the main different applications of molecular-based methods for cyanobacteria (and/or their toxins) detection and identification   | 105 |
| <b>Figure 4.1</b>  | Diagram showing the workflow followed during the data gathering on the LEGE CC strains, the completed and expected outputs of the process and the planned updates   | 133 |
| <b>Figure 4.2</b>  | Example of morphological diversity among cyanobacterial isolates from LEGE CC   | 138 |
| <b>Figure 4.3</b>  | Circular ML tree of 16S rRNA gene sequences illustrating the phylogenetic diversity of LEGE CC strains, their placement at the order level, and some traits or information relevant for biotechnological purposes | 139 |
| <b>Figure 4.4</b>  | General statistics for the LEGE CC strains concerning geographical origin, source/type of habitat, and taxonomic classification at the order level  | 142 |
| <b>Figure 4.5</b>  | Number of LEGE CC strains included in journal articles  | 142 |
| <b>Figure 4.6</b>  | ML cladogram for 165 LEGE CC strains having available data related to natural products  | 143 |
| <b>Figure 4.7</b>  | Examples of strains evaluated for EPSs production   | 148 |

## Abbreviations

|           |   |
|-----------|---|
| ANA       | Anatoxin-a  |
| BGAS      | Blue-Green Algae Supplements                                    |
| BMAA      | $\beta$ -Methylamino-L-Alanine                                  |
| BRCs      | Biological Research Centres                                     |
| CC        | Culture Collection  |
| CYN       | Cylindrosdpermopsin   |
| DGGE      | Denaturing Gradient Gel Electrophoresis                         |
| ECCO      | European Culture Collections' Organisation                      |
| eDNA      | Environmental DNA   |
| ELISA     | Enzyme-Linked Immunosorbent Assay                               |
| EMBRC     | European Marine Biological Resource Centre                      |
| EPS       | Extracellular Polymeric Substances                              |
| FISH      | Fluorescent <i>in situ</i> Hybridization                        |
| gDNA      | Genomic DNA   |
| HABs      | Harmful Algal Blooms  |
| ICN       | International Code of Nomenclature for algae, fungi, and plants |
| ICNP      | International Code of Nomenclature of Prokaryotes               |
| ITS       | Internal Transcribed Spacer                                     |
| LC-MS/MS  | Liquid Chromatography Mass Spectrometry                         |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization Time Of Flight      |
| MC        | Microcystin   |
| MIRRI     | Microbial Resource Research Infrastructure                      |
| NGS       | Next-Generation Sequencing                                      |
| NRPS      | Nonribosomal Peptide-Synthetase                                 |
| OTU       | Operational Taxonomic Unit                                      |
| PKS       | Polyketide Synthase   |
| SEM       | Scanning Electron Microscopy                                    |
| SOP       | Standard Operating Procedure                                    |
| SXT       | Saxitoxin   |
| TEM       | Transmission Electron Microscopy                                |
| WHO       | World Health Organization                                       |
| WFCC      | World Federation for Culture Collections                        |

# 1. General Introduction

Cyanobacteria are a large, diverse group (phylum) of phototrophic prokaryotes, inhabiting Earth since nearly 3 billion years ago (Dvorák et al. 2013; Schirrmeyer et al. 2015). They are the most ancient organisms capable of oxygenic photosynthesis and were responsible for the first long-term oxygenation of the Earth's atmosphere which occurred ca. 2.5-2.3 billion years ago, known by the name of "Great Oxidation Event" (Cavalier-Smith 2006; Schirrmeyer et al. 2015). This event can be considered the first ecological service, at the global scale. It was characterized by an acute rise of atmospheric O<sub>2</sub> concentration, and marked the transition from a reducing to an oxidizing atmosphere, enabling the emergence of new and more complex forms of life (Sánchez-Baracaldo et al. 2014; Schirrmeyer et al. 2015). Ever since, in a continually changing environment, cyanobacteria have evolved under different pressure forces that drive evolution, allowing them to selectively colonize new ecosystems and to become morphologically and physiologically more and more diverse (Schirrmeyer et al. 2013). Indeed, currently these microorganisms exhibit a high diversity in terms of their habitats, morphology, physiology, and metabolic capabilities (Komárek 2010; Beck et al. 2012; Whitton 2012). This diversity has an intrinsic value not only for ecosystems, where cyanobacteria perform important ecological functions (e.g. contributions to biogeochemical cycling of carbon and nitrogen), but also for human activities, since some of these organisms (and/or their products) are being used as an increasingly important bioresource (Rastogi & Sinha 2009; Sohm et al. 2011; Paerl 2012; Sharma et al. 2014). Despite the ecological significance and the perceived value as a natural resource, cyanobacteria may also have deleterious effects on the environment and humans, namely because of their capability to produce toxins (Codd et al. 2005).

## 1.1 Cyanobacterial diversity

Cyanobacteria can be found in aquatic (marine, brackish or fresh waters) or terrestrial (e.g. in soil, or being endolithic or sub-aerial) environments (Whitton 2012; Hauer et al. 2015). Some are extremophiles, living in hypersaline, highly alkaline, thermophilic, or psychrophilic environments (Whitton 2012; Makhalanyane et al. 2015). Quantitatively, cyanobacteria are among the most important organisms on Earth. An estimate of their global biomass is  $600 \times 10^{12}$  g C (dry biomass) or  $3 \times 10^{15}$  g C (wet biomass), values that represent around 1/2000 of the global carbon biomass for primary producers (García-Pichel et al. 2003).

In terms of number of taxa, estimates of the current cyanobacterial biodiversity range from 3166-6280 (Nabout et al. 2013) to 8000 (Guiry 2012) species. However, many more species are believed to exist, in particular in understudied ecosystems and geographic regions (Nabout et al. 2013; Dvorák et al. 2014; Sherwood et al. 2015).

Traditionally, the classification and identification of cyanobacteria has been based on the analysis of morphological characteristics using light-microscopy observations. Nevertheless, owing to the morphological simplicity of the organisms, the variability of phenotypic traits within one species, and the dependence of selected characters from environmental factors (strains cultured and maintained for long times under artificial conditions in the laboratory, may exhibit uncommon or "distorted" features), make the taxonomic identification of cyanobacteria based only in microscopy quite problematic (Gugger et al. 2002; Dvorák et al. 2014; Komárek 2016). For this reason, the adoption of a combination of approaches for the identification of cyanobacteria was widely advocated and put into practice in the last decade (Palinska & Surosz 2014; Komárek 2016), culminating in the most recent classification proposal for cyanobacterial genera (Komárek et al. 2014). This proposal followed a polyphasic approach, in which species (or genus) are defined both by genotypic (notably, the use of the 16S rRNA gene) and phenotypic characterizations, including ultrastructural, physiological and ecological characters (see, for example, Wilmotte & Herdman 2001, Dvorák et al. 2014; Komárek 2016). In the wake of the "molecular revolution" in systematics and taxonomy of cyanobacteria, the number of newly erected or revised taxa has increased dramatically in the last few years (e.g. see Komárek 2010 for a review of older works). For instance, 16 new genera were presented at the 19th Symposium of the International Society for Cyanophyte Research, in 2013 (Dvorák et al. 2014; Komárek et al. 2014). Accordingly, the number of Type strains deposited in Culture Collections (CCs) has increased as well. Apart their biological diversity, cyanobacteria are also well-known for their chemodiversity, producing a wide variety of secondary metabolites (Wiegand & Pflugmacher 2005; Rastogi & Sinha 2009; Leão et al. 2012; Brito et al. 2015).

## **1.2 Ecological, environmental, and economic issues related to cyanobacteria**

Cyanobacteria were pioneers in respect to two major metabolic innovations: oxygenic photosynthesis and aerobic N<sub>2</sub> fixation (Cavalier-Smith 2006; Sánchez-Baracaldo et al. 2014). In the past, they laid the foundations of Earth's atmosphere and biosphere, as we know them nowadays (Cavalier-Smith 2006; Schirrmeister et al. 2015). Currently, these metabolically-diverse microorganisms are still governing important geochemical and biotic changes, at different spatial scales, and in different environments (Sohm et al. 2011; Paerl 2012; Whitton 2012; Yang et al 2016). Other examples of ecological services provided by cyanobacteria include soil formation and stabilization (Garcia-Pichel & Wojciechowski 2009; Hauer et al. 2015), calcification (Merz-Preiß & Riding 1999; Jansson & Northen 2010), production of growth regulators on other organisms through allelopathy (Leão et al. 2015), etc.

As abovementioned, cyanobacteria can survive in almost every habitat such as from oceans to fresh water, soil to bare rocks, deserts to ice shelves, hot springs to polar lakes, existing either as free living organisms or in symbioses with plants, lichens and several protists (Whitton 2012; Hauer et al. 2015; Makhalanyane et al. 2015; Oren 2015; Rikkinen 2015). In some of these habitats they represent dominant biota in terms of total biomass and productivity (e.g. Partensky et al. 1999; Scanlan et al. 2009; Sánchez-Baracaldo et al. 2014; Yang et al 2016). As a result of evolution and survival in such assorted habitats, cyanobacteria exhibit a broad spectrum of secondary metabolites, which are likely to give them ecological advantages to compete successfully on the planet (Sivonen & Börner 2008; Leão et al. 2012). Due to this, cyanobacteria are considered a rich and emerging source of novel metabolites from a biotechnological and industrial point of view (Rastogi & Sinha 2009; Singh et al. 2011). For marine cyanobacteria, these secondary metabolites include 40% lipopeptides, 9% amides, 5.6% amino acids, 4.2% fatty acids, and 4.2% macrolides (Burja et al. 2001). In terms of their bioactivities, such compounds tend to be lipopeptides (66%) with different activities: cytotoxicity (41%), antitumor (13%), antiviral (4%), and antibiotic (12%), with the remaining 18% being antimalarials, antimycotics, multi-drug resistance reversers, and immunosuppressive agents (Burja et al., 2001).

#### *Environmental risks and ecological disservices*

Ecological or environmental disservices are typically understood as functions of ecosystems, or of their components, that are perceived as negative for human well-being (Lyytimäki et al. 2008). One such example are Harmful Algal Blooms (HABs), which are a rapid and uncontrolled growth of algae, or cyanobacteria, in aquatic environments. In the case of cyanobacteria, negative impacts came through the release of odors (i.e. production of odiferous metabolites, such as 2-methylisoborneol or geosmin, which confer earthy/musty taste and odor to water; see Giglio et al. 2010), or more dangerously through the release of toxic compounds (Codd et al. 2005; Moreira et al. 2014). Indeed, a number of cyanobacterial secondary metabolites are recognized as being cyanotoxins, which led the World Health Organization (WHO) to state that some cyanobacterial species may act as causative agents of poisoning in animals and humans (for a review of the different classes of toxins, and the taxa producing them, see Codd et al. 2005). People may be exposed to cyanotoxins by drinking or bathing in contaminated water, but human health may be affected by different, less obvious routes (e.g. Ibelings & Chorus 2007). Other examples of cyanobacteria-related disservices include aesthetic impacts such as water discoloration due to HABs or discolouring of calcite formations in buildings and monuments (Albertano 2012). Biodeterioration of touristic attractions (e.g. caves and monuments) caused by cyanobacterial mats formation is also a matter of concern (Albertano 2012).

### *Economic-related issues*

The work of Sharma and collaborators (2014) presents cyanobacteria from an economic perspective, where the plethora of present and possible biotechnological applications of cyanobacteria is scrutinized. Uses varies from pharmaceuticals (Singh et al. 2011), cosmetics (Rastogi & Sinha 2009), biofertilizers (Vaishampayan et al. 2001), soil conditioners (Hu et al. 2012), feed/food or additives (Gantar & Svirčev 2008; Ohmori & Ehira 2014), to biofuels (Machado & Atsumi 2012), among many others applications (Sharma et al. 2014).

For oncology drugs only, the pharmaceutical value of the estimated marine cyanobacteria diversity was evaluated in \$37.5-181.5 billion dollars (Erwin et al. 2010), although very few cyanobacterial compounds have yet entered clinical trials (Singh et al. 2011). The use of cyanobacterial bioactive molecules has been a subject of very active patenting activity in the past (Sekar & Pauraj 2007), but in the last few years has somewhat decreased (Borowitzka 2014). Economic costs due to cyanobacterial blooms directly affects water supply and management agencies, and may also represent indirect costs to tourism and agriculture sectors (Steffensen 2008; Albertano 2012). There is a good number of patent applications that are concerned with the detection or controlling of toxic cyanobacteria and their toxins, but the largest number of recent patents relates to the production of biofuels (for a review of patenting activity for these and other cyanobacterial products, see Borowitzka 2014). From ancient times, edible cyanobacteria have been included in the informal market of food, in different parts of the world (Ohmori & Ehira 2014). The increasing recognition of the nutritional value of such cyanobacteria by consumers worldwide, led to a real, global market for food or dietary supplements containing cyanobacteria (Ohmori & Ehira 2014).

Overall, there are clear commercial and scientific drivers for an increasing demand for (new) cyanobacterial isolates, to associated data and expertise (e.g. consultation or training services) (see also Antunes et al. 2016).

## **1.3 Preservation of microbial diversity and the need for strains**

As the bioindustry is evolving, the provision of microbial resources for research is now recognised as being an essential component in the advancement of biotechnology and life sciences, thus contributing to discovery and helping provide solutions to societal and economic challenges (Smith et al. 2014; Stackebrandt et al. 2015). This trend has resulted in a “renaissance” of (the importance of) microbial culture collections, and to the recognition of the need to conserve and to increase as much as possible the wealth of microorganisms included therein, for future applications (Stackebrandt et al. 2015). The Organisation for Economic Cooperation and Development (OECD) established the concept of Biological Resource Centre

(Smith 2003), being the variant devoted to microbes known as microbial Biological Resource Centres (mBRCs). The general concept of a BRC presented at that time includes service providers and repositories of the living cells, genomes of organisms, and information relating to heredity and the functions of biological systems (Smith et al. 2014; Antunes et al. 2016).

The information resources endorsed by the World Federation of Culture Collections (WFCC), and available from its website (<http://www.wfcc.info/wdcmdb/>), provide a comprehensive directory of culture collections, and a gateway to microbial biodiversity, molecular biology and genome projects. At present, 709 collections are registered in the World Data Centre for Microorganisms (WDCM; [www.wfcc.info/ccinfo/](http://www.wfcc.info/ccinfo/)), 220 of them just in Europe. European collections house about 795,000 microbial strains, including algae, bacteria, yeasts and fungi (Stackebrandt et al. 2015).

### 1.3.1 *Ex situ* preservation of cyanobacterial diversity

Cyanobacterial cultures are held in formal or informal collections located in academic, public-service, private, government, and commercial organizations worldwide (Day 2014). These collections perform a key role in delivering documented, characterized cultures as “seed stocks” for use in different sectors such as health and education, agriculture, or bioindustry. They are also providers of reference strains for biological assays (e.g. for toxicity testing) or of strains cited in scientific papers, including Type strains from taxonomic studies. Lastly they are also centres for conservation (i.e. preservation) of cyanobacterial biodiversity (Day 2014).

According to data in the WDCM, in the world, more than 100 algal- and cyanobacterial-based CCs maintain cyanobacteria (> 11000 strains), based on data for the year 2016. Some of the world’s most important repositories of living cyanobacteria are the Pasteur Culture Collection of Cyanobacteria (PCC), in France; the Culture Collection of Algae at Goettingen University (SAG), in Germany; the Culture Collection of Autotrophic Organisms (CCALA), in Czech Republic; the Culture Centre of Algae and Protozoa (CCAP), in Scotland; the American Type Culture Collection (ATCC) and the Culture Collection of Algae at the University of Texas (UTEX), in the USA; and, the Microbial Culture Collection at the National Institute for Environmental Studies (NIES-Collection), in Japan. Some of those international CCs maintain cyanobacterial strains recognized as being toxic.

In Portugal, the Coimbra Collection of Algae (ACOI) – a CC initiated in 1972 (Santos & Santos 2004) - has 445 strains from 48 cyanobacterial genera, according to data retrieved from its online database (<http://acoi.ci.uc.pt>; accessed December 2016). Cyanobacterial strains at ACOI were collected mostly in Portugal, and primarily from fresh water and some terrestrial environments (Santos & Santos 2004). Another CC in Portugal that comprises cyanobacterial strains is the Estela Sousa e Silva Algal Culture Collection (ESSACC), at the National Institute of Health Dr. Ricardo Jorge (Paulino et al. 2009). The first strains at ESSACC were collected

more than 40 years ago. This CC has 127 cyanobacterial strains, several of them are recognizably toxigenic. They were collected mostly from mainland Portugal and almost exclusively from fresh water environments (Paulino et al. 2009). While the former CC is a member of WFCC, the second is not.

## 1.4 Objectives and Thesis Structure

The general objective of this research is to boost awareness of the richness of cyanobacterial bio- and chemo-diversity, whose extents are not yet fully known, of their associated risks, and of the continuing relevance of preserving such diversity through *ex situ* conservation. Specific goals pursued included (1) to contribute to the still ongoing improvement of the systematics of cyanobacteria, (2) to test and to evaluate the complementarity of different approaches to assess their biodiversity, (3) to recognize and discuss the molecular-based tools being used for the monitoring of (potentially) toxic cyanobacteria, in risk assessment systems, and (4) to empirically evidence the worth of obtaining and preserving isolates for taxonomic, environmental and applied purposes.

The present compilation of studies can be conceptually divided into three main sections. The first section reinforces the relevance of an accurate classification for the understanding of the cyanobacterial diversity, and explores its role for either the study of cyanobacteria in cultures or in natural settings. Since reference strains held in culture collections are particularly important for comparative purposes in current taxonomic or phylogenetic studies, it is thus believed that make available a general listing on such isolates is serviceable to microbiologists in general, and cyanobacteriologists in particular (subsection 2.1). A reexamination and taxonomic study of marine cyanobacteria previously isolated from a narrow area of the Atlantic coast, in Portugal, led to the proposal of 3 new genera and 2 new species (subsection 2.2). The high cyanobacterial diversity present in hypersaline microbial mats is extensively characterized, while the implication of different methods and classification approaches for its study is discussed (subsection 2.3). The second section is dedicated to review the application of molecular tools on the detection of cyanobacteria and their toxins in the environment, with a special emphasis to food supplements surveillance (section 3). The third and last section describes the process that was followed to organize in-house cyanobacterial isolates into a Culture Collection (section 4). The case study exposed in this section also stresses the (potential) value of cyanobacterial isolates as a resource for biotechnology applications. Finally, the last chapter discusses the main findings of the thesis and summarizes significant outcomes derived from it.

## 1.5 References

- ALBERTANO, P. 2012. Cyanobacterial biofilms in monuments and caves. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, B.A. Whitton (ed.), Springer, pp. 317-343.
- ANTUNES, A., STACKEBRANDT, E. & LIMA, N. 2016. Fueling the bio-economy: European culture collections and microbiology education and training. *Trends in Microbiology*, 24, 77-79.
- BECK, C., KNOOP, H., AXMANN, I. M. & STEUER, R. 2012. The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics*, 13, 56.
- BOROWITZKA, M. A. 2014. Patents on cyanobacteria and cyanobacterial products and uses. In *Cyanobacteria: An Economic Perspective, 1st edition*, N.K. Sharma, A.K. Rai, L.J. Stal (eds.), John Wiley & Sons, pp. 329-338.
- BRITO, Á., GAIFEM, J., RAMOS, V., GLUKHOV, E., DORRESTEIN, P. C., GERWICK, W. H., VASCONCELOS, V. M., MENDES, M. V. & TAMAGNINI, P. 2015. Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*, 9, 218-226.
- CAVALIER-SMITH, T. 2006. Cell evolution and Earth history: stasis and revolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361, 969-1006.
- CODD, G. A., MORRISON, L. F. & METCALF, J. S. 2005. Cyanobacterial toxins: risk management for health protection. *Toxicology and Applied Pharmacology*, 203, 264-272.
- CROWE, S. A., DØSSING, L. N., BEUKES, N. J., BAU, M., KRUGER, S. J., FREI, R. & CANFIELD, D. E. 2013. Atmospheric oxygenation three billion years ago. *Nature*, 501, 535-538.
- DAY, J.G. 2014. Cryopreservation of cyanobacteria. In *Cyanobacteria: An Economic Perspective, 1st edition*, N.K. Sharma, A.K. Rai, L.J. Stal (eds.), John Wiley & Sons, pp. 319-327.
- DVOŘÁK, P., CASAMATTA, D. A., POULÍČKOVÁ, A., HAŠLER, P., ONDŘEJ, V. & SANGES, R. 2014. Synechococcus: 3 billion years of global dominance. *Molecular Ecology*, 23, 5538-5551.
- GANTAR, M. & SVIRČEV, Z. 2008. Microalgae and cyanobacteria: food for thought. *Journal of Phycology*, 44, 260-268.
- GARCIA-PICHEL, F., BELNAP, J., NEUER, S. & SCHANZ, F. 2003. Estimates of global cyanobacterial biomass and its distribution. *Algological Studies*, 109, 213-217.
- GARCIA-PICHEL, F. & WOJCIECHOWSKI, M. F. 2009. The evolution of a capacity to build supra-cellular ropes enabled filamentous cyanobacteria to colonize highly erodible substrates. *PLoS One*, 4, e7801.
- GIGLIO, S., CHOU, W. K. W., IKEDA, H., CANE, D. E. & MONIS, P. T. 2010. Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environmental Science & Technology*, 45, 992-998.
- GUGGER, M., LYRA, C., HENRIKSEN, P., COUTE, A., HUMBERT, J. F. & SIVONEN, K. 2002. Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. *International Journal of Systematic and Evolutionary Microbiology*, 52, 1867-1880.
- GUIRY, M. D. 2012. How many species of algae are there? *Journal of Phycology*, 48, 1057-1063.
- HAUER, T., MÜHLSTEINOVÁ, R., BOHUNICKÁ, M., KAŠTOVSKÝ, J. & MAREŠ, J. 2015. Diversity of cyanobacteria on rock surfaces. *Biodiversity and Conservation*, 24, 759-779.
- HU, C., GAO, K. & WHITTON, B. A. 2012. Semi-arid regions and deserts. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, B.A. Whitton (ed.), Springer, pp. 345-369.
- IBELINGS, B. W. & CHORUS, I. 2007. Accumulation of cyanobacterial toxins in freshwater "seafood" and its consequences for public health: a review. *Environmental Pollution*, 150, 177-192.
- JANSSON, C. & NORTHEN, T. 2010. Calcifying cyanobacteria—the potential of biomineralization for carbon capture and storage. *Current Opinion in Biotechnology*, 21, 365-371.
- KOMÁREK, J. 2010. Recent changes (2008) in cyanobacteria taxonomy based on a combination of molecular background with phenotype and ecological consequences (genus and species concept). *Hydrobiologia*, 1, 245-259.
- KOMÁREK, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, 51, 346-353.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J. R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- LEÃO, P. N., ENGENE, N., ANTUNES, A., GERWICK, W. H. & VASCONCELOS, V. 2012. The chemical ecology of cyanobacteria. *Natural Product Reports*, 29, 372-391.
- LEE, E., RYAN, U. M., MONIS, P., MCGREGOR, G. B., BATH, A., GORDON, C. & PAPANINI, A. 2014. Polyphasic identification of cyanobacterial isolates from Australia. *Water Research*, 59, 248-261.
- LYYTIMÄKI, J., PETERSEN, L. K., NORMANDER, B. & BEZÁK, P. 2008. Nature as a nuisance? Ecosystem services and disservices to urban lifestyle. *Environmental Sciences*, 5, 161-172.

- MACHADO, I. M. & ATSUMI, S. 2012. Cyanobacterial biofuel production. *Journal of Biotechnology*, 162, 50-56.
- MAKHALANYANE, T. P., VALVERDE, A., VELÁZQUEZ, D., GUNNIGLE, E., VAN GOETHEM, M. W., QUESADA, A. & COWAN, D. A. 2015. Ecology and biogeochemistry of cyanobacteria in soils, permafrost, aquatic and cryptic polar habitats. *Biodiversity and Conservation*, 24, 819-840.
- MERZ-PREIB, M. & RIDING, R. 1999. Cyanobacterial tufa calcification in two freshwater streams: ambient environment, chemical thresholds and biological processes. *Sedimentary Geology*, 126, 103-124.
- MOREIRA, C., RAMOS, V., AZEVEDO, J. & VASCONCELOS, V. 2014. Methods to detect cyanobacteria and their toxins in the environment. *Applied Microbiology and Biotechnology*, 98, 8073-8082.
- NABOUT, J. C., DA SILVA ROCHA, B., CARNEIRO, F. M. & SANT'ANNA, C. L. 2013. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodiversity and Conservation*, 22, 2907-2918.
- OHMORI, M. & EHIRA, S. 2014. *Spirulina*: an example of cyanobacteria as nutraceuticals. In *Cyanobacteria: An Economic Perspective, 1st edition*, N.K. Sharma, A.K. Rai, L.J. Stal (eds.), John Wiley & Sons, pp., 103-118.
- OLSON, S. L., REINHARD, C. T. & LYONS, T. W. 2016. Cyanobacterial diazotrophy and Earth's delayed oxygenation. *Frontiers in Microbiology*, 7, 1526.
- OREN, A. 2015. Cyanobacteria in hypersaline environments: biodiversity and physiological properties. *Biodiversity and Conservation*, 24, 781-798.
- PAERL, H. W. 2012. Marine plankton. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, B.A. Whitton (ed.), Springer, pp. 127-153.
- PALINSKA, K. A. & SUROSZ, W. 2014. Taxonomy of cyanobacteria: a contribution to consensus approach. *Hydrobiologia*, 740, 1-11.
- PARTENSKY, F., HESS, W. R. & VAULOT, D. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews*, 63, 106-127.
- PAULINO, S., SAM-BENTO, F., CHURRO, C., ALVERCA, E., DIAS, E., VALÉRIO, E. & PEREIRA, P. 2009. The Estela Sousa e Silva Algal Culture Collection: a resource of biological and toxicological interest. *Hydrobiologia*, 636, 489-492.
- RASTOGI, R. P. & SINHA, R. P. 2009. Biotechnological and industrial significance of cyanobacterial secondary metabolites. *Biotechnology Advances*, 27, 521-539.
- RIKKINEN, J. 2015. Cyanolichens. *Biodiversity and Conservation*, 24, 973-993.
- SÁNCHEZ-BARACALDO, P., RIDGWELL, A. & RAVEN, J. A. 2014. A neoproterozoic transition in the marine nitrogen cycle. *Current Biology*, 24, 652-657.
- SANTOS, L. & SANTOS, F. M. 2004. The Coimbra Culture Collection of Algae (ACOI). *Nova Hedwigia*, 79, 39-47.
- SCANLAN, D. J., OSTROWSKI, M., MAZARD, S., DUFRESNE, A., GARCZAREK, L., HESS, W. R., POST, A. F., HAGEMANN, M., PAULSEN, I. & PARTENSKY, F. 2009. Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews*, 73, 249-299.
- SCHIRRMEISTER, B. E., DE VOS, J. M., ANTONELLI, A. & BAGHERI, H. C. 2013. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences*, 110, 1791-1796.
- SCHIRRMEISTER, B. E., GUGGER, M. & DONOGHUE, P. C. 2015. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology*, 58, 769-785.
- SEKAR, S. & PAULRAJ, P. 2007. Strategic mining of cyanobacterial patents from the USPTO patent database and analysis of their scope and implications. *Journal of Applied Phycology*, 19, 277-292.
- SHARMA, N.K., RAI, A.K. & STAL, L.J. (eds.). 2014. *Cyanobacteria: An Economic Perspective, 1st edition*. John Wiley & Sons, 376 p.
- SHERWOOD, A. R., CARLILE, A. L., VACCARINO, M. A. & JOHANSEN, J. R. 2015. Characterization of Hawaiian freshwater and terrestrial cyanobacteria reveals high diversity and numerous putative endemics. *Phycological Research*, 63, 85-92.
- SINGH, R. K., TIWARI, S. P., RAI, A. K. & MOHAPATRA, T. M. 2011. Cyanobacteria: an emerging source for drug discovery. *The Journal of Antibiotics*, 64, 401-412.
- SIVONEN, K. & BÖRNER, T. 2008. Bioactive compounds produced by cyanobacteria. In *The Cyanobacteria: Molecular Biology, Genomics and Evolution*, A. Herrero, E. Flores (eds.), Caister Academic Press, pp. 159-197.
- SMITH, D. 2003. Culture collections over the world. *International Microbiology*, 6, 95-100.
- SMITH, D., MCCLUSKEY, K. & STACKEBRANDT, E. 2014. Investment into the future of microbial resources: culture collection funding models and BRC business plans for biological resource centres. *SpringerPlus*, 3, 81.
- SOHM, J. A., WEBB, E. A. & CAPONE, D. G. 2011. Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, 9, 499-508.

- STACKEBRANDT, E., SCHÜNGEL, M., MARTIN, D. & SMITH, D. 2015. The microbial resource research infrastructure MIRRI: strength through coordination. *Microorganisms*, 3, 890-902.
- STEFFENSEN, D. A. (2008). Economic cost of cyanobacterial blooms. In *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs*. Springer New York, pp. 855-865.
- VAISHAMPAYAN, A., SINHA, R. P., HADER, D. P., DEY, T., GUPTA, A. K., BHAN, U. & RAO, A. L. 2001. Cyanobacterial biofertilizers in rice agriculture. *The Botanical Review*, 67, 453-516.
- WHITTON, B. A. (ed.). 2012. *Ecology of Cyanobacteria II - Their Diversity in Space and Time*. Springer Netherlands. 760 p.
- WIEGAND, C. & PFLUGMACHER, S. 2005. Ecotoxicological effects of selected cyanobacterial secondary metabolites a short review. *Toxicology and Applied Pharmacology*, 203, 201-218.
- WILMOTTE, A. & HERDMAN, M. 2001. Phylogenetic relationships among the cyanobacteria based on 16S rRNA sequences. In *Bergey's Manual of Systematics of Archaea and Bacteria: Volume one*. D.R. Boone, R.W. Castenholz & G.M. Garrity, eds. Springer: New York, USA, pp. 487-493.
- YANG, J., LV, H., LIU, L., YU, X. & CHEN, H. 2016. Decline in water level boosts cyanobacteria dominance in subtropical reservoirs. *Science of the Total Environment*, 557, 445-452.

## 2. Classification matters: Its key role in understanding diversity

### 2.1 A curated database of cyanobacterial strains relevant for modern taxonomy and phylogenetic studies

Vitor Ramos<sup>1,2</sup>, João Morais<sup>1</sup>, Vitor M. Vasconcelos<sup>1,2,\*</sup>

<sup>1</sup> Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal

<sup>2</sup> Faculty of Sciences, University of Porto, Rua do Campo Alegre, Porto 4169-007, Portugal

\* Corresponding author. Email address: vmascon@fc.up.pt

#### 2.1.1 Abstract

The dataset herein described lays the groundwork for an online database of relevant cyanobacterial strains, named CyanoType (<http://lege.ciimar.up.pt/cyanotype>). It is a database that includes categorized cyanobacterial strains useful for taxonomic, phylogenetic or genomic purposes, with associated information obtained by means of a literature-based curation. The dataset lists 371 strains and represents the first version of the database (CyanoType v.1). Information for each strain includes strain synonymy and/or co-identity, strain categorization, habitat, accession numbers for molecular data, taxonomy and nomenclature notes according to three different classification schemes, hierarchical automatic classification, phylogenetic placement according to a selection of relevant studies (including this), and important bibliographic references. The database will be updated periodically, namely by adding new strains meeting the criteria for inclusion and by revising and adding up-to-date metadata for strains already listed. A global 16S rDNA-based phylogeny is provided in order to assist users when choosing the appropriate strains for their studies.

#### 2.1.2 Background & Summary

Strains held in culture collections are pivotal for comparative purposes in current taxonomic or phylogenetic studies of prokaryotes in general, and cyanobacteria in particular. In recent years, the number of new cyanobacterial genera established by a modern polyphasic taxonomy (i.e., following a combination of different techniques) has greatly increased (Komárek et al. 2014), resulting in the designation of several new Type strains (i.e., an isolate based on which the author describes a new species or genus; it is often the holotype specimen itself).

Concerning nomenclature, cyanobacteria (formerly known as blue-green algae) are a special case among the prokaryotes, since they are ruled either by the International Code of Nomenclature for algae, fungi, and plants (ICN; formerly the International Code of Botanical Nomenclature, ICBN) or by the International Code of Nomenclature of Prokaryotes (ICNP, formerly the International Code of Nomenclature of Bacteria, ICNB). Nomenclature rules governed by these two entities are converging (Oren 2004, 2011; Komárek et al. 2014; Palinska et al. 2014) but due to this duality two general types of systematics still exist: the more ancient botanical/phycological classification scheme and the bacteriological scheme (Palinska et al. 2014). Available keys for the identification of cyanobacteria are mostly based on the botanical system proposed by Geitler in 1932 (Pinevich 2015), including the key present in the pioneering bacteriological system of Stanier and colleagues (Rippka et al. 1979; Oren 2004). One important classification system followed by microbiologists is the Bergey's Manual of Systematic Bacteriology, often confused as an "official classification", which is not the case (Brenner et al. 2005). The manual classifies the cyanobacteria in "form genera" which, in turn, are divided into clusters or subclusters (Boone & Castenholz 2001). For each (sub)cluster at least one Reference strain is assigned. This strain category, as presented in the manual, should not be confused with being a Type strain (though some of them effectively are). Moreover, the term "form genus" has no standing under the Bacteriological or under the Botanical Codes of Nomenclature (Oren 2004) and the authors of the cyanobacterial section of the manual early admitted that the proposed classification is a temporary one (Oren 2004; Boone & Castenholz 2001). Despite these taxonomic issues, the Bergey's Manual is an important body of work, since it systematizes, lists and characterizes a good number of cyanobacterial strains, most of which are widely used as reference in phylogenetic studies.

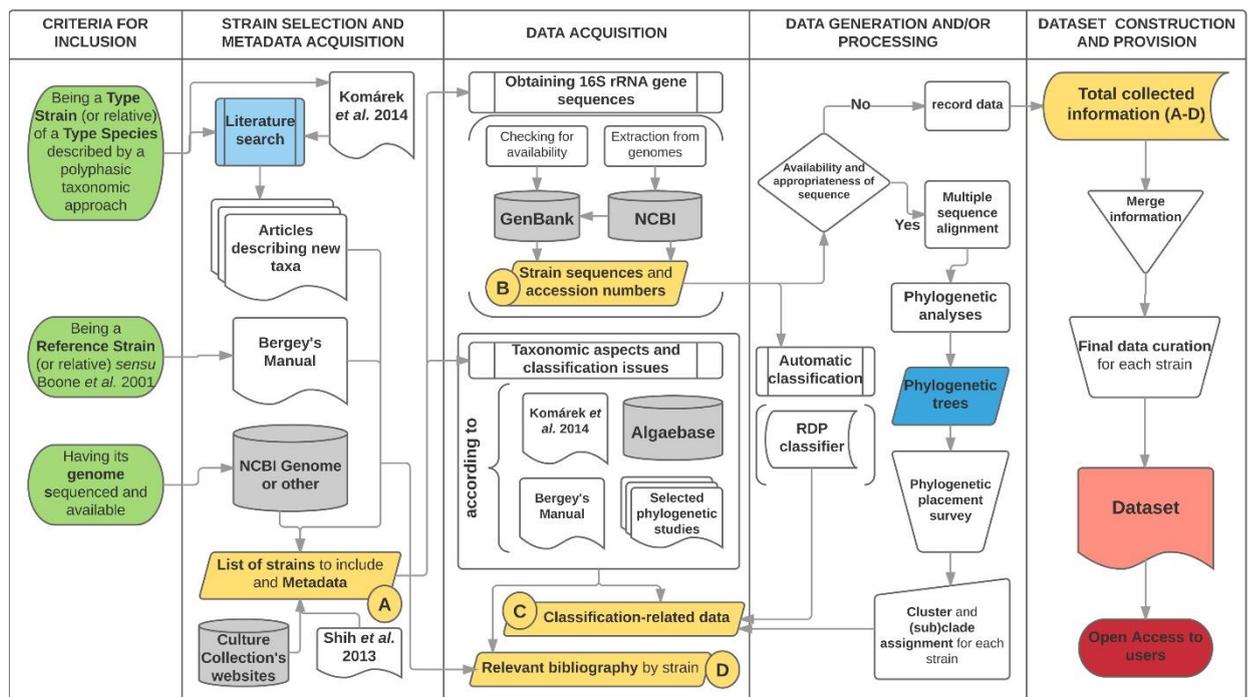
Several new taxa that have been recently established emerge from taxonomic revisions of "classical" botanical genera, which have been described primarily by their morphological features. Most of them are in fact polyphyletic, as depicted from 16S rRNA gene-based phylogenies using strains assigned to different species of such genera (Oren 2004; Boone & Castenholz 2001; Komárek et al. 2014). Since the pioneering work of Carl Woese, George Fox and colleagues (Balch et al. 1977; Fox et al. 1977; Woese & Fox 1977), the 16S rRNA gene became, and still is, the most important and widely used molecular marker for the identification of prokaryotes. However, its resolving power at species level is low (Rosselló-Mora & Amann 2001) and should therefore be employed to obtain identifications at the genus level. Nonetheless, its appropriateness for phylogenetic-based classifications was again demonstrated more recently. By following a phylogenomic approach involving 54 cyanobacterial genomes, Shih et al. (2013) have demonstrated that the 16S rRNA gene phylogeny is highly congruent with that obtained from a concatenation of 31 conserved proteins. Thus, it is likely that the 16S rRNA gene will continue to be the standard molecular marker for proposing new cyanobacterial genera (Komárek et al. 2014). The emergence of genome-based taxonomy

(Konstantinidis & Tiedje 2005) approaches, however, renders genome-sequenced strains increasingly important to the field.

Due to the above-mentioned issues, choosing the proper strains to include in taxonomic, phylogenetic or comparative genomic studies on cyanobacteria is very often a challenging task. In order to overcome this difficulty, we introduce the curated dataset of CyanoType v.1, a database with an extensive list of relevant cyanobacterial strains classified by importance category, i.e. with the indication about being a Type strain, a Reference strain *sensu* Bergey's Manual, and/or a strain having its genome sequenced. The dataset encompasses different types of metadata (e.g. strain synonymy and/or co-identity), including a reference list for each strain. In order to help users in their process of selecting strains, we provide two 16S rDNA-based phylogenetic trees for guidance. The main phylogenetic tree and the information for each strain included in the dataset is available in the searchable, online database at <http://lege.ciimar.up.pt/cyanotype> (see also Appendix C).

### 2.1.3 Methods

The workflow performed in this study for the literature and database searches, for the data compilation and processing, and for the dataset construction is illustrated in Fig. 2.1.



**Figure 2.1.** Diagram illustrating the workflow followed during the construction and release of the dataset (standard flowchart symbols were used).

### Data acquisition

We initially established the criteria for inclusion of cyanobacterial strains in the dataset. We have considered three main groups of strains to be included, representing different levels of

importance from the taxonomic point of view: (1) strains that were used as Type strains for the proposal or establishment of a new taxon by mean of a modern, polyphasic taxonomic approach (Komárek et al. 2014), (2) strains that are included as Reference strains in Bergey's Manual of Systematic Bacteriology (Boone & Castenholz 2001), and (3) strains that have their genome sequenced and publicly available. Following these criteria, and through literature and online database searches, we have obtained a list of relevant strains categorized by taxonomic importance.

Literature searches for Type strains were firstly guided by information included in the work of Komárek et al. (2014), which lists the cyanobacterial genera for which the holotype (i.e. Type species) was described using a modern, polyphasic taxonomic approach. To avoid missing any Type strain (e.g. strains from new genera arising from later studies than those included in Komárek et al. 2014) we have performed complementary searches in literature databases such as ISI Web of Science, Scopus, PubMed and Google Scholar using the following Boolean search string: ((cyanobact\* OR cyanophy\*) AND ((gen. nov. OR gen. et sp. nov.) OR "new genus" OR "novel genus" OR "new genera" OR "novel genera")) for the fields [Title, Abstract, Keywords]. Duplicate articles were eliminated. We then fully examined the search results to evaluate the suitability of the articles for our research.

**Table 2.1** Number of cyanobacterial strains included in the dataset of CyanoType (version 1.0), by category.

| Strain category *       | Numbers of strains | Number of strains included in the phylogenetic trees |
|-------------------------|--------------------|--|
| T or t, only            | 73                 | 63   |
| T or t and R or r       | 5                  | 4  |
| T or t and G            | 10                 | 10   |
| T or t and R or r and G | 9                  | 9  |
| R or r and G            | 60                 | 60   |
| R or r, only            | 41                 | 30   |
| G, only                 | 172                | 155  |
| E                       | 1                  | 1  |
| <b>TOTAL</b>            | <b>371</b>         | <b>332</b>   |

\* T - Type strain of the Type species; t - not the type strain, but phylogenetically close-related; R - Reference strain in Bergey's Manual of Systematic Bacteriology (Boone & Castenholz 2001); r - not the reference strain, but phylogenetically close-related; G - strain with its genome sequenced and publicly available; E - strain studied from exsiccata; see also categories descriptions in the Data Records section

The dataset also includes Reference strains from all clusters or subclusters defined in the Bergey's Manual (Boone & Castenholz 2001), and known relatives (as indicated in the Manual) if the 16S rRNA gene sequences for the Reference strains were not available. The number of strains fitting in each category are summarized in Table 2.1. A full description of each category type can be found in the Data Records section (see Strain\_Category). We have also – as was done for strains – performed literature and online database searches for data and metadata

acquisition. For instance, we manually performed data mining in public molecular (e.g. NCBI) and taxonomic (e.g. AlgaeBase; Guiry & Guiry 2015) databases and searches on websites of Culture Collections.

Finally, strains with available cyanobacterial genomes were also included in the dataset. The work of Shih et al. (2013) was used as a reference first list. To obtain the full list, we have then used the Assembly database and other NCBI resources (e.g. Genome, Genome BLAST, BLASTn) in order to search for genomes and to obtain accession numbers and 16S rRNA gene nucleotide sequences from the strains. The search term Cyanobacteria (Taxonomy ID: 1117) was used to obtain the list of available cyanobacterial genomes. For our study, we have considered 251 out of 372 strains having their genomes available in NCBI (until the end of 2015). Missing strains are *Prochlorococcus* spp. which were not included in the dataset due to overrepresentation and phylogenetic redundancy. Even so, the dataset comprises 28 representatives, by far the most represented genus (Appendix B).

All strains with available 16S rRNA gene nucleotide sequences were then subjected to a phylogenetic study (see Subsection “Phylogenetic analyses” below). First, in order to obtain the sequences, we have performed Boolean searches in the NCBI Nucleotide database (which includes GenBank). For some strains it was necessary to extract the sequences by mining their genome. Accession numbers were recorded in the dataset (Appendix B). Adequacy of the sequences length for multiple alignment and further analyses was then checked (see Subsection “Phylogenetic analyses”). Additionally, the 16S rRNA gene sequences were submitted to the automatic RDP Naive Bayesian rRNA Classifier v2.6 pipeline (Wang et al. 2007). Strains were ranked and the hierarchical classification result recorded (Appendix B).

Moreover, we have classified the strains at higher taxonomic levels (Order and Family) and verified the nomenclatural status of taxon names according to taxonomic concepts followed in Komarek et al. (2014) (at the Genus level), and in AlgaeBase (Species level) (Guiry & Guiry 2015), and recorded it in the dataset (Appendix B). The same was made to other names by which the strain may be known or to conflicting identifications of co-identical strains. Whenever relevant, we have also added additional taxonomy or nomenclature notes/clarifications to the dataset (e.g., indication of whether it is the Type strain of the holotype).

### **Phylogenetic analyses**

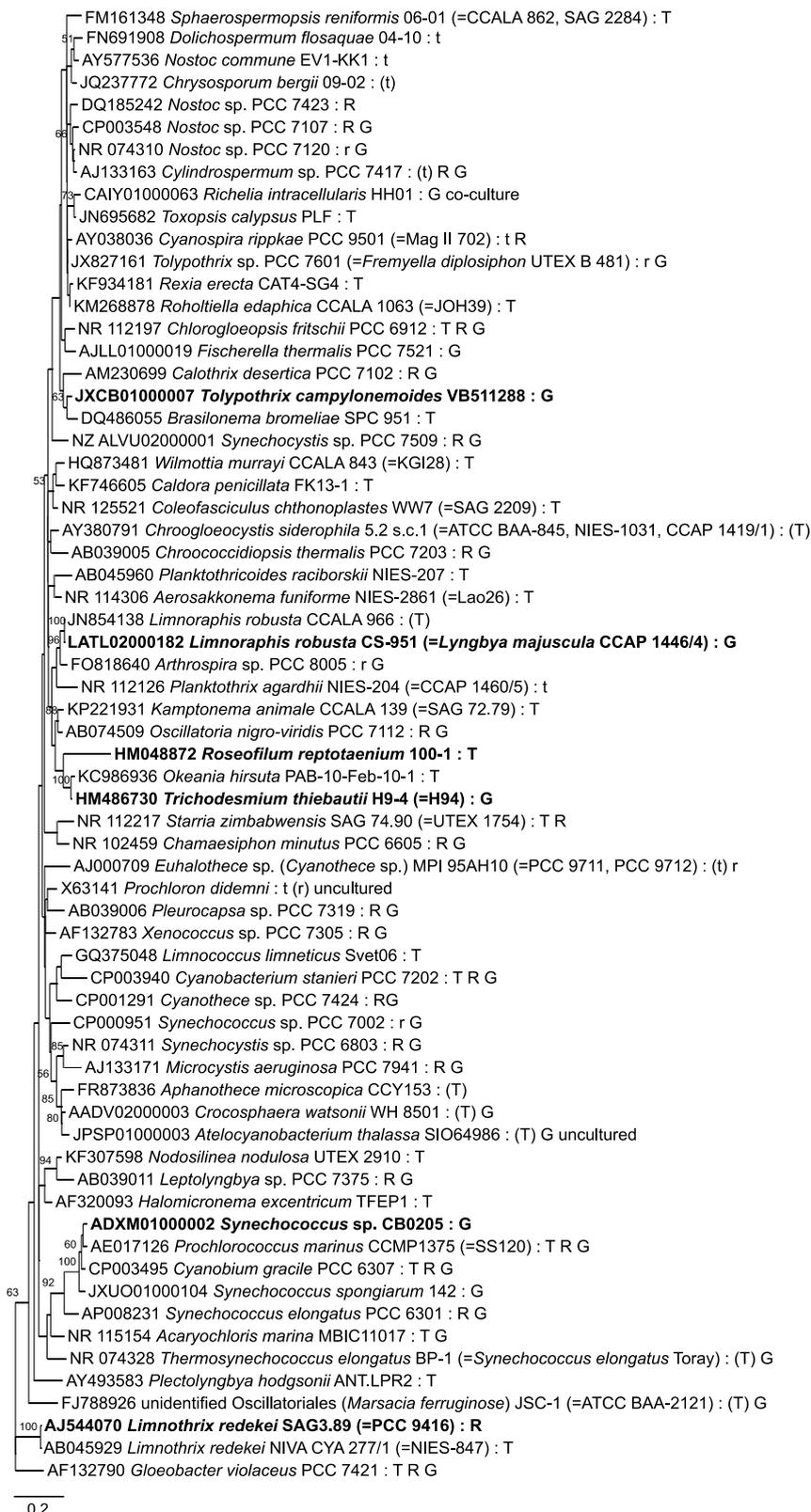
All bioinformatics procedures and analyses were conducted using the MEGA7 software package (Kumar et al. 2016). Sequences were aligned using the ClustalW algorithm. Strains with small-sized sequences (<1000 nt) were treated separately, to avoid reducing the number of unambiguously aligned nucleotide positions, and thus preventing distortion of the main

phylogeny. Molecular phylogenetic analyses were inferred by using the Maximum Likelihood (ML) method, based on the nucleotide substitution model that best fit the alignment data. By applying the corrected Akaike's Information Criterion (AICc), the chosen nucleotide substitution model was General Time Reversible (GTR) for both analyses. A discrete Gamma distribution ([+G]) was used to model evolutionary rate differences among sites, while the rate variation model allowed for some sites to be evolutionarily invariable ([+I]). The trees with the highest log likelihood (-27524.3318 and -7982.8912, respectively) are shown for the main (Appendix A1) and complementary (Fig. 2.2) phylogenies. Both trees were rooted with the outgroup *Chloroflexus aurantiacus* J-10-fl (NR\_074263).

The phylogenetic analysis for the main tree (Appendix A1) involved 333 nucleotide sequences. All positions containing gaps and missing data were eliminated. The final alignment dataset consisted of 863 positions. In order to systematize the phylogenetic placement of the cyanobacterial strains, we have grouped the strains into clusters (broader groups; for sequences placed together but lacking bootstrap support) and in clades (for groups of sequences with a ML bootstrap support), as depicted in Appendix A1. This primary data was included in the dataset (see also *Phylog\_This\_Work*, in the Data Records section). We have also described the phylogenetic placement of the strains according to a selection of important studies (Ahlgren et al. 2006; Kettler et al. 2007; Scanlan et al. 2009; Schirmer et al. 2011; Ahlgren & Rocop, 2012; Howard-Azzeh et al. 2014) (Appendix B).

In turn, the analysis performed for the complementary phylogenetic tree (Fig. 2.2) involved 67 nucleotide sequences. This tree is meant to show the placement of those shorter sequences that were not included in the main tree (six sequences, ranging from 381 to 898 nt; three strains with sequences <315 nt were discarded from the analysis). To do so, we have also included 60 cyanobacterial sequences used in the main tree. The selection of strains involved representatives from all the clades identified in Appendix A1 (for larger clades, we have selected a number of divergent strains), intending to cover the cyanobacterial diversity contained in CyanoType v.1. Due to the inclusion of short sequences, less than 5% gaps, missing data, and ambiguous bases were allowed at any position of the alignment. This resulted in a total of 533 positions in the final alignment.

The survey and collection of the different data or metadata (including important bibliographic references for each strain) was finished by the end of 2015, for this version of the database (Appendix B).



**Figure 2.2.** Example of the use of the proposed subset of strains representing the cyanobacterial “tree of life” (see Subset\_Condens\_Tree in the Data Records section and Phylogenetic analyses in Methods) to evaluate the phylogenetic placement of strains not included in the main tree (Appendix A1) due to having short 16S rRNA gene sequences (in bold). The evolutionary history was inferred by using the Maximum Likelihood method based on the GTR+G+I model. Bootstrap values indicated near internal branches; values below 50% were omitted. Information for each cyanobacterial strain include accession number of the nucleotide sequence, strain ID, eventual taxonomic synonyms or other strain names (in parentheses), and co-identical strains or other strain codes (in parentheses). Letters after colon indicate the categorization of strains as follows (see Strain\_Category in Data Record section): The outgroup was pruned from the tree for clarity. The scale bar represents nucleotide substitutions per site.

#### 2.1.4 Data Records

The dataset, the sequence alignments and the tree files (see Appendix B) obtained in this work are deposited at FigShare.

The dataset (CyanoType\_data\_v1.0.csv) is a semi-comma separated values file containing taxonomic- and phylogenetic-related data and other useful information for each cyanobacterial strain considered in this work, including important strain-related references when available (e.g. literature for strain origin, identification/characterization, taxonomy, phylogeny and/or genome sequencing). Rows represent single strains, for which data were integrated. Columns are for useful information and metadata, as follow:

##### *Entry\_number*

It is the entry number of the cyanobacterial strain in the dataset.

##### *Strain\_ID*

Taxon name and strain code.

##### *Strain\_Other\_ID*

Other taxon name(s) previously assigned to the strain, synonym(s) of the taxon name, or other putative taxonomic designation(s).

##### *Strain\_Co-Ident*

Older code(s) for the strain, misspellings or code(s) from co-identical strains (e.g. same strain deposited in other(s) culture collection(s); not an exhaustive list).

##### *Strain\_Category*

Categorization of strains by relevance, as defined in this work, and additional important strain characterization. T - Type strain of the Type species (taxon established by modern polyphasic taxonomy); t - not the type strain but known to have the same phylogenetic placement as the Type species, after taxonomic revision; R - Reference strain in Bergey's Manual of Systematic Bacteriology (Boone & Castenholz 2001); r - strain known to be included in the same phylogenetic cluster as the reference strain, as mentioned in the Bergey's Manual (Boone & Castenholz 2001); G - strain with its Genome sequenced and publicly available; E - strain studied from Exsiccata (dried herbarium specimens of cyanobacteria). A letter in parentheses means that there is a taxonomic-related uncertainty with the taxon name (see taxonomic comments) or the assigned strain's category could not be satisfactorily confirmed (e.g. for unpublished, provisional species names).

*Strain\_Addition*

Additional characterization of the strain concerning its isolation status. “Co-culture” is for strains in culture but associated with other organism (i.e. not free-living isolates).

*Environment*

Type of environment from which the strain was obtained.

*Habitat\_notes*

Additional details on the source/origin or lifestyle of the strain.

*16S\_Acc\_Nbr*

GenBank accession number for the 16S rRNA gene sequence.

*NCBI\_ID*

NCBI Assembly or BioProject numbers for available cyanobacterial genomes.

*Tax\_Komarek\_Ord\_Fam*

Order and family assignments for the strain identification(s), according to the recent classification scheme proposed by Komárek et al. (2014).

*Tax\_Status\_Genus*

Status of the genus as depicted from the Appendix 1 in Komárek et al. (2014), as follows: 1 – genera supported by a molecular phylogeny, including a 16S rRNA gene sequence of the type species; 2 – genera, from which only one or a few species were studied using molecular methods and for which there is no 16S rRNA gene data for the type species; 3 – genera studied using molecular methods and found to be poly/paraphyletic or with no clear relationship with other genera; 4 – genera not yet studied using molecular methods; 5 – genera not yet validly described; *16S-Type* – genera for which there is a 16S rRNA sequence for the type material publicly available (in parentheses, when this availability is not indicated in Komárek et al. 2014); [?] problematic genera from the taxonomic point of view.

*Tax\_AlgaeBase\_Ord\_Fam*

Order and family assignments for the strain identification(s), according to the online database AlgaeBase (Guiry & Guiry 2015).

*Tax\_Status\_AlgaeBase\_&\_Tax\_Notes*

Status of the strain's taxon name as present in AlgaeBase (Guiry & Guiry 2015). When applicable, we indicate whether it is a type strain (i.e. holotype or epitype). It might also include other primary data, such as taxonomic relevant comments or notes.

*Tax\_AlgaeBase\_Holotype*

Type species of the genus (holotype) and authority as indicated in AlgaeBase (Guiry & Guiry 2015). It may include some additional taxonomic relevant comments or notes.

#### *Tax\_Bergey's*

Classification according to the Bergey's Manual scheme (Boone & Castenholz 2001), in condensed form. The first roman numerals refer to subsections, while the second refer to form-genus within that subsection.

#### *Phylog\_This\_Work*

Position of the strain within the phylogenetic tree illustrated in Appendix A1 (capital letters and numbers refer to clusters and clades, respectively).

#### *Subset\_Condens\_Tree*

Subset of 60 strains for a proposal of a condensed phylogenetic tree covering the cyanobacterial diversity included in CyanoType (see also Fig. 2.2 and the Subsection “Phylogenetic analyses” in Methods). The goal of this suggested subset is to aid users in preliminary phylogenetic analyses, namely to discern the placement of their sequences in relation to relevant strains.

#### *Phylog\_RDP\_Classifier*

Classification according to the automatic RDP Naive Bayesian rRNA Classifier (Wang et al. 2007).

#### *Phylog\_Shih*

Phylogenetic placement of the strain (clade or sub-clade) as established in Shih et al. (2013).

#### *Phylog\_Howard-Azzeh*

Phylogenetic placement of the strain (clade or sub-clade) as established in Howard-Azzeh et al. (2014).

#### *Phylog\_Schirrmeister*

Phylogenetic placement of the strain (clade or sub-clade) as established in Schirrmeister et al. (2011).

#### *Phylog\_Picocyano*

Ecotypes as established or present in Ahlgren & Rocop (2012), Ahlgren et al. (2006), Kettler et al. (2007) or Scanlan et al. (2009). For *Prochlorococcus* and *Synechococcus* spp. strains only.

#### *Metadata\_Shih*

Information for additional metadata present in Shih et al. (2013).

#### *References*

Important literature related with the strain (e.g. with information on isolation/source origin, identification/taxonomy, phylogeny, genome sequencing, etc.).

### 2.1.5 Technical Validation

The dataset (Appendix B) was extensively checked for double entries, errors or inconsistencies (all fields), while data or metadata concerning each entry (i.e. strain) was further revised, very particularly decisions about category attribution (see Fig. 2.1). Whenever available, bibliographic references are provided for each entry, enabling any user to get access to the original data. Researchers making use of the dataset (Appendix B) or the database are encouraged to assess the validity and accuracy of the data and send us feedback through the website database, at <http://leqe.ciimar.up.pt/cyanotype>. The information will be updated after curation by our team.

In the future, it is intended that the information for any given entry (i.e. strain) in the database may be curated on a voluntary basis. To this end, administrative and managerial procedures for quality control of data will be implemented. For instance, users will need to request permission to become a contributor and will have a user account. Any observation made by a contributor will be flagged and simultaneously an automatic message will be sent to the administrator. The 'pending' flag will be removed only after administrator approval. The observation made by the contributor for a particular strain will be then recorded and become accessible to other users, as updated information for that strain.

### 2.1.6 Acknowledgements

This work was supported in part by FCT – Foundation for Science and Technology under the project UID/Multi/04423/2013 and by the Structured Program of R&D&I INNOVMAR - Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER-000035, Research Line NOVELMAR), funded by the Northern Regional Operational Program (NORTE2020) through the European Regional Development Fund (ERDF). VR was supported by the FCT fellowship SFRH/BD/80153/2011.

### 2.1.7 References

- AHLGREN, N. A. & ROCAP, G. 2012. Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Frontiers in Microbiology*, 3, 213.
- AHLGREN, N. A., ROCAP, G. & CHISHOLM, S. W. 2006. Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology*, 8, 441-454.
- BALCH, W. E., MAGRUM, L. J., FOX, G. E., WOLFE, R. S. & WOESE, C. R. 1977. An ancient divergence among the bacteria. *Journal of Molecular Evolution*, 9, 305-311.

- BOONE, D.R. & CASTENHOLZ, R.W. (eds) 2001. *Bergey's Manual of Systematic Bacteriology 2nd edn. Volume 1. The Archaea and the Deeply Branching and Phototrophic Bacteria*, Springer-Verlag.
- BRENNER, D. J., STALEY, J. T. & KRIEG, N. R. 2005. Classification of procaryotic organisms and the concept of bacterial speciation. In *Bergey's Manual of Systematic Bacteriology, 2nd edn. Volume 2. The Proteobacteria*, Springer Nature. pp. 27–32
- FOX, G. E., MAGRUM, L. J., BALCH, W. E., WOLFE, R. S. & WOESE, C. R. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences*, 74, 4537-4541.
- GEITLER, L. 1932. Cyanophyceae. In *Dr L. Rabenhorst's Kryptogamen-Flora von Deutschland Österreich und der Schweiz. Volume XIV*, Akademische Verlag.
- GUIRY, M. D. & GUIRY, G. M. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. Available at: <http://www.algaebase.org.>, accessed between Nov. 2015 and Dec. 2015.
- HOWARD-AZZEH, M., SHAMSEER, L., SCHELLHORN, H. E. & GUPTA, R. S. 2014. Phylogenetic analysis and molecular signatures defining a monophyletic clade of heterocystous cyanobacteria and identifying its closest relatives. *Photosynthesis Research*, 122, 171-185.
- KETTLER, G. C., MARTINY, A. C., HUANG, K., ZUCKER, J., COLEMAN, M. L., RODRIGUE, S., CHEN, F., LAPIDUS, A., FERRIERA, S., JOHNSON, J., STEGLICH, C., CHURCH, G. M., RICHARDSON, P. & CHISHOLM, S. W. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genetics*, 3, e231.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J. R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005. Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology*, 187, 6258-6264.
- KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870-1874.
- OREN, A. 2004. A proposal for further integration of the cyanobacteria under the Bacteriological Code. *International Journal of Systematic and Evolutionary Microbiology*, 54, 1895-1902.
- OREN, A. 2010. Cyanobacterial systematics and nomenclature as featured in the International Bulletin of Bacteriological Nomenclature and Taxonomy / International Journal of Systematic Bacteriology / International Journal of Systematic and Evolutionary Microbiology. *International Journal of Systematic and Evolutionary Microbiology*, 61, 10-15.
- PALINSKA, K. A. & SUROSZ, W. 2014. Taxonomy of cyanobacteria: a contribution to consensus approach. *Hydrobiologia*, 740, 1-11.
- PINEVICH, A. V. 2014. Proposal to consistently apply the International Code of Nomenclature of Prokaryotes (ICNP) to names of the oxygenic photosynthetic bacteria (cyanobacteria), including those validly published under the International Code of Botanical Nomenclature (ICBN)/International Code of Nomenclature for algae, fungi and plants (ICN), and proposal to change Principle 2 of the ICNP. *International Journal of Systematic and Evolutionary Microbiology*, 65, 1070-1074.
- RIPPKA, R., STANIER, R. Y., DERUELLES, J., HERDMAN, M. & WATERBURY, J. B. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology*, 111, 1-61.
- ROSSELLÓ-MORA, R. & AMANN, R. 2001. The species concept for prokaryotes. *FEMS Microbiology Reviews*, 25, 39-67.
- SCANLAN, D. J., OSTROWSKI, M., MAZARD, S., DUFRESNE, A., GARCZAREK, L., HESS, W. R., POST, A. F., HAGEMANN, M., PAULSEN, I. & PARTENSKY, F. 2009. Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews*, 73, 249-299.
- SCHIRRMESTER, B. E., ANTONELLI, A. & BAGHERI, H. C. 2011. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology*, 11, 45.
- SHIH, P. M., WU, D., LATIFI, A., AXEN, S. D., FEWER, D. P., TALLA, E., CALTEAU, A., CAI, F., TANDEAU DE MARSAC, N., RIPPKA, R., HERDMAN, M., SIVONEN, K., COURSIIN, T., LAURENT, T., GOODWIN, L., NOLAN, M., DAVENPORT, K. W., HAN, C. S., RUBIN, E. M., EISEN, J. A., WOYKE, T., GUGGER, M. & KERFELD, C. A. 2012. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, 110, 1053-1058.
- WANG, Q., GARRITY, G. M., TIEDJE, J. M. & COLE, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261-5267.
- WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74, 5088-5090.

## 2.2 Description of new genera and species of marine cyanobacteria isolated from the Portuguese Atlantic coast

Ângela Brito<sup>1,2</sup>, Vitor Ramos<sup>3,4</sup>, Rita Mota<sup>1,2,3</sup>, Steeve Lima<sup>1,2</sup>, Arlete Santos<sup>1,2,3</sup>, Jorge Vieira<sup>1,2</sup>, Cristina P. Vieira<sup>1,2</sup>, Jan Kaštovský<sup>5</sup>, Vitor M. Vasconcelos<sup>3,4</sup> and Paula Tamagnini<sup>1,2,3</sup> \*

<sup>1</sup> i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal

<sup>2</sup> IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal

<sup>3</sup> Faculdade de Ciências, Departamento de Biologia, Universidade do Porto, Rua do Campo Alegre, Edifício FC4, 4169-007 Porto, Portugal

<sup>4</sup> Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Terminal de Cruzeiros do Porto de Leixões, 4450-208 Matosinhos, Portugal

<sup>5</sup> Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, CZ 37005 České Budějovice, Czech Republic

\* Corresponding author. Email address: pmtamagn@ibmc.up.pt

### 2.2.1 Abstract

Aiming at increasing the knowledge on marine cyanobacteria from temperate regions, we previously isolated and characterized sixty strains from the Portuguese foreshore and evaluate their potential to produce secondary metabolites. About 15% of the obtained 16S rRNA gene sequences showed less than 97% similarity to sequences in the databases revealing novel biodiversity. Herein, seven of these strains were extensively characterized and their classification was re-evaluated. Our study led to the proposal of five new taxa 3 genera (*Geminobacterium*, *Lusitaniella*, *Calenema*) and 2 species (*Hyella patelloides* and *Jaaginema litorale*).

*Geminobacterium atlanticum* LEGE 07459 is a Chroococcalean that shares morphological characteristics with other unicellular cyanobacterial genera but has a distinct phylogenetic position and particular ultrastructural features. The description of the Pleurocapsales *Hyella patelloides* LEGE 07179 includes novel molecular data for members of this genus. The filamentous isolates of *Lusitaniella coriacea* - LEGE 07167, 07157 and 06111 - constitute a very distinct lineage, and seem to be ubiquitous on the Portuguese coast. *Jaaginema litorale* has distinct characteristics compared to their marine counterparts, and our analysis indicates that this genus is polyphyletic. The Synechococcales *Calenema singularis* possess wider trichomes than other *Leptolyngbya*, and its phylogenetic position reinforces the establishment of this new genus.

### 2.2.2 Introduction

Due to the long evolutionary history and metabolic plasticity, cyanobacteria are photosynthetic prokaryotes that can be found in a wide range of habitats including the extreme ones (Whitton & Potts 2000). Marine cyanobacteria play an important role in the global carbon cycle as primary producers and the ability to fix nitrogen confers them an advantage in the often nutrient-poor/nitrogen-depleted marine environments (Knoll 2008; Zehr 2011). In the intertidal zones cyanobacteria are exposed to extremely stressful conditions such as desiccation, temperature and salinity fluctuations. In these areas, cyanobacteria often form cohesive mats held together by extracellular polymeric substances (EPS) that, besides contributing to a structurally-stable and hydrated microenvironment, are believed to confer protection against UV-radiation and promote the concentration of nutrients and metal ions (Rossi & De Philippis 2015; Stuart et al. 2016). Continental Portugal has an extensive coastline with climatic influences from the North Atlantic Ocean and the Mediterranean Sea and varied topographies, beach morphologies and wave regimens (for details see Pontes et al. (2005), Lima et al. (2006), Kottek et al. (2006), Seabra et al., (2011) and Brito et al. (2012)). In the last years, although several reports on the diversity of cyanobacteria from Portuguese fresh and estuarine waters have been published (De Figueiredo et al. 2006; Galhano et al. 2011; Lopes et al. 2012; Valério et al. 2009), only a few studies have focused on the diversity of marine strains. Araújo et al. (2009) produced a morphologically-based checklist of benthic marine algae and cyanobacteria from the north of Portugal based on new records, literature references and herbarium data. Subsequently, Brito et al. (2012) isolated and characterized, using morphological characters and molecular-based methods, 60 cyanobacterial strains from the intertidal zones along the Portuguese coast, that are maintained at LEGE Culture Collection (CIIMAR, Matosinhos, Portugal) (see also Ramos et al. 2010). The morphological analysis revealed 35 morphotypes (15 genera and 16 species) belonging to four out of the five Orders usually recognized in the traditional classification systems of Cyanobacteria (e.g. Castenholz (2001) and Komárek et al. (2014)). The molecular characterization of these isolates, based on partial 16S rRNA gene sequences, showed that 15% of these sequences had less than 97% similarity compared to the sequences available in the databases emphasizing the presence of novel cyanobacterial diversity (Bruto et al. 2012). Despite all these strains have been previously identified at the genus or species level (Bruto et al. 2012), it is well known that traditional cyanobacterial taxa may encompass several cryptic species or even genera i.e. morphologically (almost) identical organisms that fit the same current taxon but that belong to phylogenetically distinct clades/lineages (Casamatta et al. 2003; Komárek et al. 2014; Komárek 2016). Therefore, the main goal of this study was to further characterize seven of those isolates (two unicellular and five filamentous) and to re-evaluate their classification according to the recent scheme proposed by Komárek et al. (2014), and the subsequent taxonomic recommendations and principles (Komárek 2016). Our polyphasic study

led to the proposal of five new taxa, contributing to increase the knowledge on cyanobacteria from temperate marine environments.

### 2.2.3 Materials and methods

#### Cyanobacterial strains and culture conditions

The seven cyanobacterial strains studied in this work (Supplementary Table 2.4) were previously isolated from the intertidal zones along the Portuguese Atlantic coast and are deposited at the Blue Biotechnology and Ecotoxicology Culture Collection (acronym LEGE) located at CIIMAR, Matosinhos, Portugal (Brito et al. 2012). The uni-cyanobacterial strain *Calenema singularis* LEGE 06188 was grown in Z8 medium (Kotai 1972) supplemented with 25 g L<sup>-1</sup> NaCl and the remaining uni-cyanobacterial cultures were maintained in MN medium (Rippka 1988). The media were supplemented with 10 µg mL<sup>-1</sup> of vitamin B12. The cultures were kept at 25 °C, under a 16 h light (15 - 25 µmol photons m<sup>-2</sup> s<sup>-1</sup>) / 8 h dark regimen. Voucher specimens were deposited at the Herbarium for Nonvascular Cryptogams (CBFS) at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic with accession numbers CBFS A-071 to 076 and at the Herbarium of University of Porto, Porto, Portugal, with the accession numbers PO-T4749 to T4755.

#### Light microscopy

Cells were observed directly or stained with Alcian Blue (0.5% (w/v) in 50% ethanol) and observed using an Olympus X31 light microscope (Olympus, Japan). Light micrographs were acquired with an Olympus DP25 camera and the Cell<sup>^</sup>B image software (Olympus, Japan).

#### Transmission Electron Microscopy (TEM)

For negative staining, 10 µl of culture was placed on a Formvar carbon-coated nickel grids (200 mesh) for 2 min, the liquid in excess was removed with a filter paper, stained with 1% uranyl acetate for 5 sec with the excess of liquid removed.

For TEM, cells were collected, centrifuged and processed as previously described (Seabra et al., 2009) with the exception of the resin used that was EMBed-812 resin (Electron Microscopy Sciences, Hatfield, PA, USA). For the isolates LEGE 07157, LEGE 07176 and LEGE 07459 the cells were fixed with 2.5% glutaraldehyde and 2% paraformaldehyde in 50 mM sodium cacodylate buffer (pH 7.2) for 72 h and postfixed overnight with 2% osmium tetroxide in 50 mM sodium cacodylate buffer (pH 7.2). Subsequently, the cells were washed three times in sodium cacodylate buffer and stained en bloc overnight with 2% uranyl acetate.

The samples or ultrathin sections were examined using a JEM-1400Plus (Jeol, MA, USA) electron microscope operating at 80 kV.

### **Scanning Electron Microscopy (SEM)**

For SEM analysis, cells were collected by centrifugation and fixed in 2% glutaraldehyde in 50 mM sodium cacodylate buffer (pH 7.2) and washed one time in double-strength buffer. The dehydration was performed in a graded series of ethanol (25 - 100%; v/v) and dried by critical-point drying. Samples were coated with a gold/palladium thin film, for 100 sec and a 15 mA current by sputtering, using the SPI Module Sputter Coater equipment (Structure Probe, Inc., PA, USA). The SEM analysis were performed using a High Resolution Scanning Electron Microscope (JSM 6301F, Jeol Ltd., Tokyo, Japan), operating at 15 kV.

### **DNA extraction, PCR amplification, cloning and sequencing**

Genomic DNA was extracted using the phenol/chloroform method previously described (Tamagnini et al. 1997) and the regions encoding part of 16S rRNA, the internal transcribed spacer (ITS) and most of the 23S rRNA gene were amplified using the oligonucleotide primers listed on Supplementary Table 2.5. PCR reactions were performed using a thermal cycler (MyCycler™, Bio-Rad laboratories Inc., Hercules, CA, USA) following procedures previously described (Tamagnini et al. 1997). The PCR profiles included an initial denaturation at 94 °C for 5 min, followed by 35 or 30 cycles (for ITS) at 94 °C for 1 min, 52 °C or 60 °C (for ITS) for 1 min, 72 °C for 1 min and a final extension at 72 °C for 7 min. The amplification of *nifH* was performed following the methodology previously described (Omoregie et al. 2004). The PCR products were separated by agarose gel electrophoresis (Sambrook & Russell 2001) and DNA fragments were isolated from gels using the NZYGelpure Kit (NZYTech, Lisbon, Portugal), according to the manufacturer's instructions. Purified products were cloned into pGEM®-T Easy vector (Promega, Madison, WI, USA) following the methodology earlier described (Ramos et al. 2010).

### **Nucleotide sequence accession numbers**

Sequence data were deposited in the GenBank database under the accession numbers KR676346 to KR676352 (rRNA genes), and KC256766 and KC256774 (*nifH*).

### **Secondary structure of ITS**

16S-23S ITS sequences were folded using CLC Genomics Workbench software 8 (CLC Bio-Qiagen, Aarhus, Denmark). The different conserved and variable regions were identified according to Iteaman et al. (2000) and each sequence fragment was folded individually. Default parameter settings were used in this analysis.

### **Phylogenetic analysis**

The 16S-ITS-23S rRNA sequences obtained in this study (Acc. numbers KR676346 to KR676352 deposited in GenBank) were used as query in a megaBLAST, at NCBI (database limited to cyanobacteria, December 2015) and only those above 90% sequence query coverage,

were retrieved. Sequences were aligned using both Clustal Omega (Sievers et al. 2011) and MUSCLE (Edgar 2004) to determine the impact of using a given alignment algorithm. Phylogenetic relationships were inferred using both Maximum-Likelihood (ML) [Fasttree; (Price et al. 2009, 2010)] using 500 bootstrap replicates, and Bayesian approaches [MrBayes; (Ronquist et al. 2012)]. When using either the Maximum-Likelihood or the Bayesian approaches, the model used was the General Time Reversible model with a proportion of invariant sites and a gamma distribution (GTR+I+G), since this is the selected model when using the Akaike information criterion (AIC), as implemented in Modeltest (Darriba et al., 2012; Guindon and Gascuel, 2003). It should be noted that the size of the hypervariable region is highly variable in the set of the sequences used, and since gapped positions are not used in phylogenetic analyses, this region does not contribute much to the phylogenetic inferences. There were a total of 2767 (clustal alignment) and 2829 (MUSCLE alignment) ungapped positions in the final dataset.

Since the information available in databases for the 16S-ITS-23S region could not cover the overall diversity of the group, a phylogenetic tree using only the 16S rRNA gene sequences was also performed. For this purpose, the 16S rRNA gene region of the seven sequences obtained in this study was searched against the NCBI BLASTn database (April 2016) and their best hits (using the expect value as the criterion) were retrieved. In addition, in order to have a reliable and robust backbone representation of the cyanobacterial diversity, selected sequences from relevant cyanobacteria taxa, including reference strains, were also included in this analysis. The final dataset included 97 sequences. The 16S rRNA gene phylogenetic tree was constructed using the procedures described above. There were a total of 602 (clustal alignment) and 609 (MUSCLE alignment) ungapped positions in the final dataset. *Chloroflexus aurantiacus* strain J-10-fl was used as outgroup in all analyses.

The strain assigned to the Type species of the genus *Jaaginema* was not included in the above datasets due to its short length (see Results). Therefore, we also performed phylogenetic analyses using all available 16S rRNA gene sequences of *Jaaginema* independently of their length. Moreover, as noted above, when using the set of divergent 16S-ITS-23S rRNA sequences, the hypervariable region is almost ignored in the phylogenetic analyses since gapped positions are excluded. Nevertheless, this region may be very informative at the species level or below when using not so divergent sequence sets (Iteman et al. 2002). Therefore, an additional phylogenetic tree was constructed using only the ITS region from *Jaaginema litorale* LEGE 07176 and its four closest related isolates, as depicted by the 16S rDNA analysis. For both cases the ML method and the General Time Reversible (GTR) model were used, considering 500 pseudo-replicates for bootstrap analysis, as implemented in MEGA7 (Kumar et al. 2016).

The 16S rRNA gene similarity matrix for sequences from our strains and from relevant strains was conducted in MEGA7 (Kumar et al. 2016), and calculated from the number of base

differences per sequence. This analysis involved 42 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 918 positions in the final dataset.

### **Phycobiliproteins extraction**

Phycobiliproteins were extracted by disrupting the cells (2 - 4 mg dry weight) in 1 mL 0.01 M sodium phosphate buffer pH 7.0, plus 0.15 M NaCl containing 0.4 g of 0.2 mm-diameter acid washed glass beads (Sigma-Aldrich, Inc., MO, USA) using a FastPrep®-24 instrument (MP Biomedicals, CA, USA). Afterwards, samples were centrifuged at 4 °C and 16000 g for 90 min (Bennett and Bogorad, 1973). All absorbance spectra were recorded using an UV-2401PC spectrophotometer (Shimadzu, Corp., Japan) in the range 350-750 nm.

### **Nitrogenase activity**

For the *in vivo* nitrogenase activity measured by the acetylene reduction assay (ARA), cultures were grown in ASNIII<sub>0</sub> medium (ASNIII without nitrate) (Rippka et al. 1979) under 12 h light (25 μmol photons m<sup>-2</sup> s<sup>-1</sup>) / 12 h dark cycles at 30 °C for 5 days. 10 ml of the culture were transferred to Erlenmeyer flasks with a total volume of 130 ml and incubated with 13% C<sub>2</sub>H<sub>2</sub> for 12 h in sealed flasks. One millimeter of gas samples were withdrawn to determine the amount of acetylene reduced to ethylene. The measurements were performed using the gas chromatograph (GC) Clarus 480® (Perkin Elmer, MA, USA), equipped with a Thermal Conductivity Detector (TDC) and a capillary Porapak N 80/100 column, under the control of the TotalChrom® software (Perkin Elmer, MA, USA). Nitrogen was used as a carrier gas and the following temperatures were used; injector - 120 °C, detector - 120 °C and column - 60 °C. Calibration was performed using different ethylene concentrations. The nitrogenase activity was expressed per chlorophyll *a* and time. Chlorophyll *a* content was determined as previously described (Meeks & Castenholz 1971). For positive and negative controls, cells of *Anabaena* sp. PCC 7120 grown in BG11<sub>0</sub> (Stanier et al. 1971) or BG11<sub>0</sub> supplemented with ammonium chloride were used, respectively.

## **2.2.4 Results and discussion**

In a previous study, 60 cyanobacterial strains from the Portuguese foreshore were characterized and among these isolates there was a high degree of novel diversity, with nine 16S rRNA gene sequences exhibiting 93 to 97% similarity compared to sequences available in the databases (Brito et al., 2012), which are values within the recommended threshold values for bacterial species (98 to 99%) or genera (94.5 to 95%) demarcation [see for e.g. (Kim et al. 2014; Yarza et al. 2014)]. These sequences belong to two unicellular and seven filamentous strains, but only seven strains could be maintained for an extended period of time (> 1 year) in the LEGE Culture

Collection, and therefore are the ones characterized in the present study. One of the cultures, LEGE 07161, contained in fact two cyanobacterial species (Fig. 2.3, A): one previously identified as *Chroococcopsis* sp., and a smaller unicellular strain that was initially mistaken for nanocytes/baeocytes and to which the 16S rRNA gene sequence corresponds (details corrected at GenBank: KR676352). To this new isolate was attributed the code LEGE 07459.

The detailed study presented here comprises the description of new genera/species of marine cyanobacteria: The unicellular *Geminobacterium atlanticum*, the colonial *Hyella patelloides*, and the filamentous non-heterocystous *Lusitaniella coriacea*, *Jaaginema litorale*, and *Calenema singularis*.

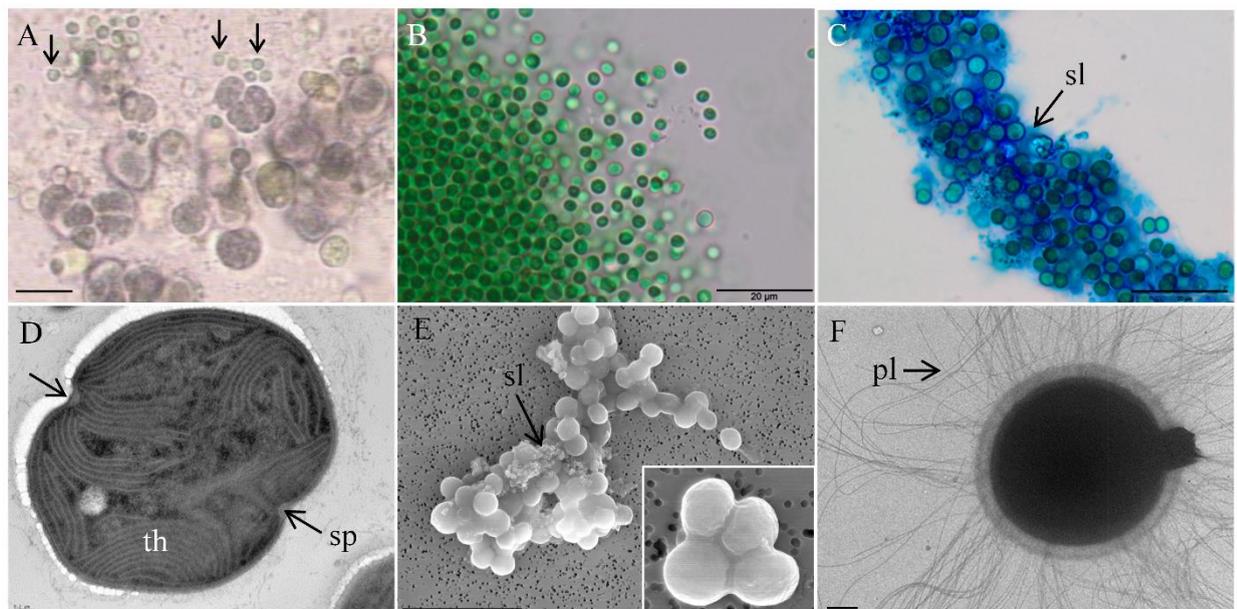
### Taxonomic descriptions

#### ***Geminobacterium* Ramos, Brito et Kaštovský, gen. nov. (Fig. 2.3)**

Spherical cells, solitary or forming aggregates. Division by binary fission in one plane. Irregular, longitudinal or semi radial arrangement of thylakoids. Presence of long and abundant pilus-like structures.

*Type species: Geminobacterium atlanticum* Ramos, Brito et Kaštovský (see below).

*Etymology:* “Gemino” and “bacterium” due to having a mixture of the diacritical features of the relative genera *Geminocystis* and *Cyanobacterium*.

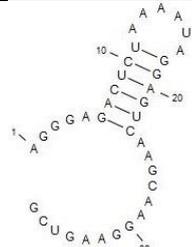
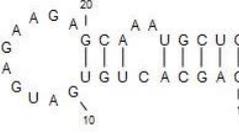


**Figure 2.3.** Micrographs of *Geminobacterium atlanticum* LEGE 07459. (A) Light microscopy of the initial culture containing two cyanobacterial strains, one morphologically identified as *Chroococcopsis* sp. LEGE 07161 (larger cells, Brito et al. 2012) and the other identified here as *G. atlanticum* LEGE 07459 (cells indicated by arrows); (B) Uni-cyanobacterial culture of *G. atlanticum* LEGE 07459 and (C) Culture stained with Alcian blue, slime (sl); (D) TEM ultrathin section showing the irregular thylakoids (th) arrangement and a forming septum (sp; arrows); (E) SEM image where it is possible to observe slime surrounding cell aggregates; (F) Negative staining showing the presence of pilus-like structures (pl). Scale bars: (A) = 10  $\mu$ m; (B, C) = 20  $\mu$ m; (D) = 0.2  $\mu$ m; (E) = 10  $\mu$ m; (E - insert) = 3  $\mu$ m; (F) = 0.5  $\mu$ m.

***Geminobacterium atlanticum* Ramos, Brito et Kaštovský sp. nov. (Fig. 2.3)**

Spherical cells with 2.0 to 3.0 µm in diameter. Blue- or olive-green cells solitary or more often forming aggregates (Fig. 2.3, E) surrounded by slime (Fig. 2.3, C and E). Irregular arrangement of the thylakoids (Fig. 2.3, D), and presence of long and abundant pilus-like structures (Fig. 2.3, F). Contains phycocyanin (PC) but not phycoerythrin (PE) (Supplementary Fig. 2.11). 16S-23S rRNA ITS region 367 nucleotides long: D1-D1' helix with 37 nucleotides; Box-B 29 nucleotides long with 10 nucleotides in the terminal loop; tRNA<sup>Ile</sup> and tRNA<sup>Ala</sup> with 74 and 73 nucleotides, respectively (Table 2.2 and Supplementary Table 2.6).

**Table 2.2** Predicted secondary structure of the 16S-23S rRNA internal transcribed spacer (ITS) of *Geminobacterium atlanticum* LEGE 07459.

| D1-D1'   | Box-B  | V3             | tRNA     |
|--|--|----------------|----------|
|  |  | Not determined | Ala, Ile |

*Type:* Portugal, Martinhal beach, N 37° 01' 07.30"/W 8° 55' 36.17", collected by Rui Seabra and Vitor Ramos, 2/May/2007, Holotype CBFS A-073-1, Herbarium for Nonvascular Cryptogams at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic. Isotype PO-T4750, Herbarium of University of Porto, Porto, Portugal. Type strain LEGE 07459.

*Etymology:* atlanticum - referring to the type locality, Atlantic coast of Portugal.

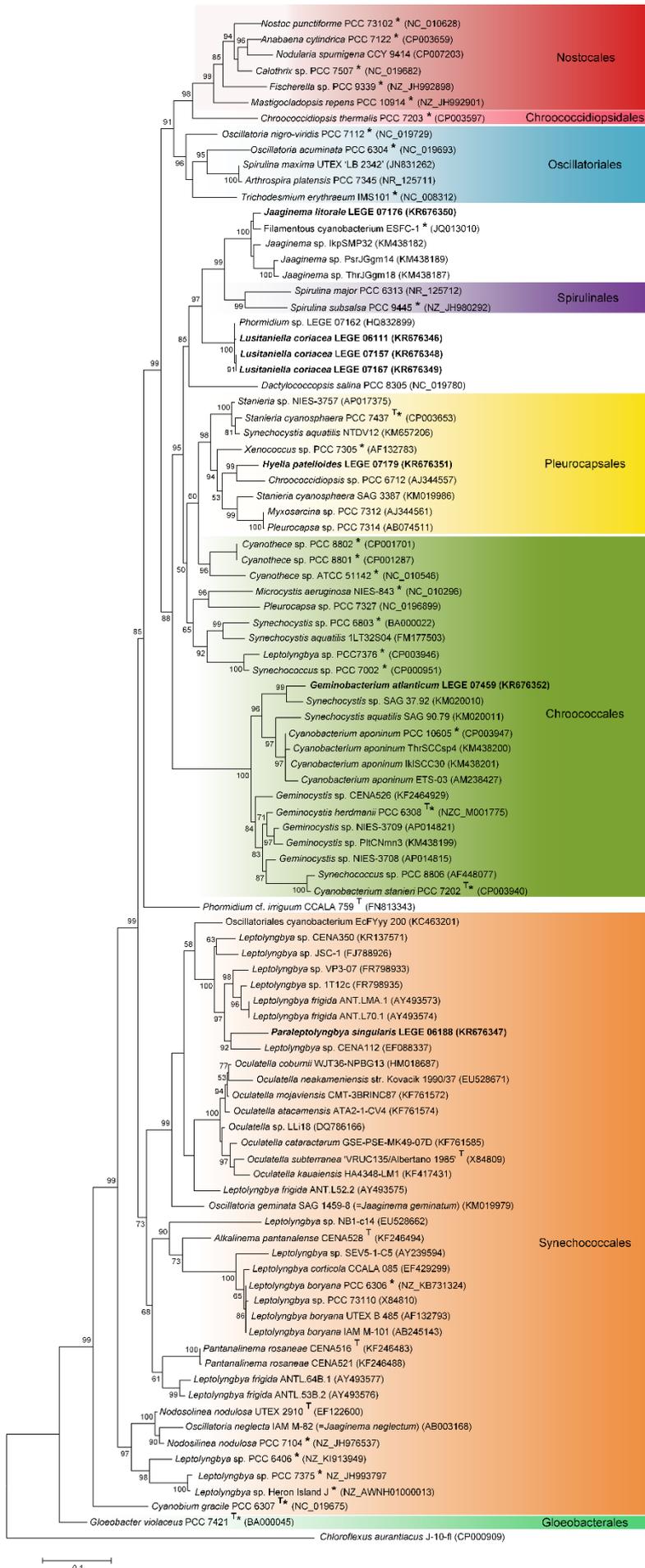
*Ecology and distribution:* This taxon occurred rarely as epizoid on shells of the marine gastropod mollusc *Patella* sp..

*Comparison with similar taxa:* LEGE 07459 has morphological characteristics similar to those described for the genera *Geminocystis* Korelusová, Kaštovský et Komárek, *Cyanobacterium* Rippka et Cohen-Bazire and *Synechocystis* Sauvageau. The cell size is smaller than most of *Geminocystis* members and similar to members of *Synechocystis*. However, the arrangement of the thylakoids is irregular over almost the whole cell (Fig. 2.3, D) as in *Geminocystis* and *Cyanobacterium*, contrasting with the genus *Synechocystis* in which the arrangement is generally parietal (Korelusová et al. 2009). Moreover, long and abundant pilus-like structure were observed (Fig. 2.3, F) resembling *Synechocystis* sp. PCC 6803 type IV pili (Bhaya et al. 2000). It was previously described that *Geminocystis* differs from *Synechocystis* by its position

in the phylogenetic tree, arrangement of thylakoids and ITS secondary structure, sharing these characteristics with *Cyanobacterium*. *Geminocystis* differs from *Cyanobacterium* in cell shape and number of planes of fission (Korelusová et al. 2009). The spherical cell shape of LEGE 07459 is similar to *Geminocystis* but divides in one plane as *Cyanobacterium*. The pigment composition of LEGE 07459 indicates the presence of PC but not PE (Supplementary Fig. 2.11). The 16S-23S rRNA ITS region of LEGE 07459 (Table 2.2 and Supplementary Table 2.6) is 367 nucleotides long and the secondary structure of the region D1-D1' helix is shorter than the one reported for *Synechocystis* but similar to those reported for *Geminocystis* and *Cyanobacterium* (Korelusová et al. 2009).

The Box-B of our isolate is slightly different and has a longer terminal loop (10 nucleotides) than the ones reported for *Geminocystis*, *Cyanobacterium* and *Synechocystis* (4-6 nucleotides). The V3 region could not be determined (as for *Geminocystis papuanica*; Korelusová et al. 2009). The ITS region of LEGE 07459 contains also the tRNA<sup>Ala</sup> and tRNA<sup>Ala</sup> as *Geminocystis* and *Cyanobacterium*, contrasting with *Synechocystis* that only contains tRNA<sup>Ala</sup> (Korelusová et al. 2009).

From the phylogenetic analysis based on 16S-ITS-23S rRNA gene sequence, it can be shown that LEGE 07459 is in a cluster comprising members of the genera *Geminocystis* and *Cyanobacterium* with a strong bootstrap support, while members of *Synechocystis* are in a different position in the tree (Supplementary Fig. 2.12). In the phylogenetic reconstruction based on 16S rRNA gene sequences only (Fig. 2.4), the isolate LEGE 07459 is within a robust cluster (100% bootstrap value) composed mainly by members of *Geminocystis* and *Cyanobacterium*, but together with the strain SAG 37.92 (a marine strain assigned to *Synechocystis* sp.). In order to clarify the classification of SAG 37.92, available sequences from isolates identified as belonging to the holotype species of *Synechocystis*, *S. aquatilis* Sauvageau, were included in this analysis. These strains appear scattered along the tree (Fig. 2.4), but SAG 90.79 is in a subclade with *Cyanobacterium aponinum* isolates (98% of sequence similarity, see Fig. 2.5) and within a larger clade that includes our strain and SAG 37.92. This clade is included in a broader clade with the Type strains of the genera *Geminocystis* and *Cyanobacterium* (*G. herdmanii* PCC 6308 and *C. stanieri* PCC 7202, respectively) (Korelusová et al. 2009; Rippka & Cohen-Bazire 1983). It is also worth to notice that the holotype strain of the recently described genus *Geminocystis* - *G. herdmanii* PCC 6308 - was previously assigned to the genus *Synechocystis* (cluster 1) according to the Bergey's Manual of Systematic Bacteriology (Castenholz 2001). At that time, it was already recognized that none of the *Synechocystis* strains included in the Bergey's Manual corresponded to the Type species, *Synechocystis aquatilis* Sauvageau, as stated in Castenholz (2001, p. 514).



**Figure 2.4.** Maximum likelihood (ML) phylogenetic tree based on partial 16S rRNA gene sequences. The studied strains are indicated in bold and by the respective GenBank accession number (brackets). *Chloroflexus aurantiacus* strain J-10-fl was used as outgroup. Numbers along branches indicate bootstrap values considering 500 pseudo-replicates: only those equal or above 50% are indicated. Taxonomic levels higher than genus are according to Komárek et al. (2014). The orders are highlighted in different colours along the tree. \* strains included in the most recent taxonomic classification of cyanobacteria (Komárek et al. 2014). T Type species.

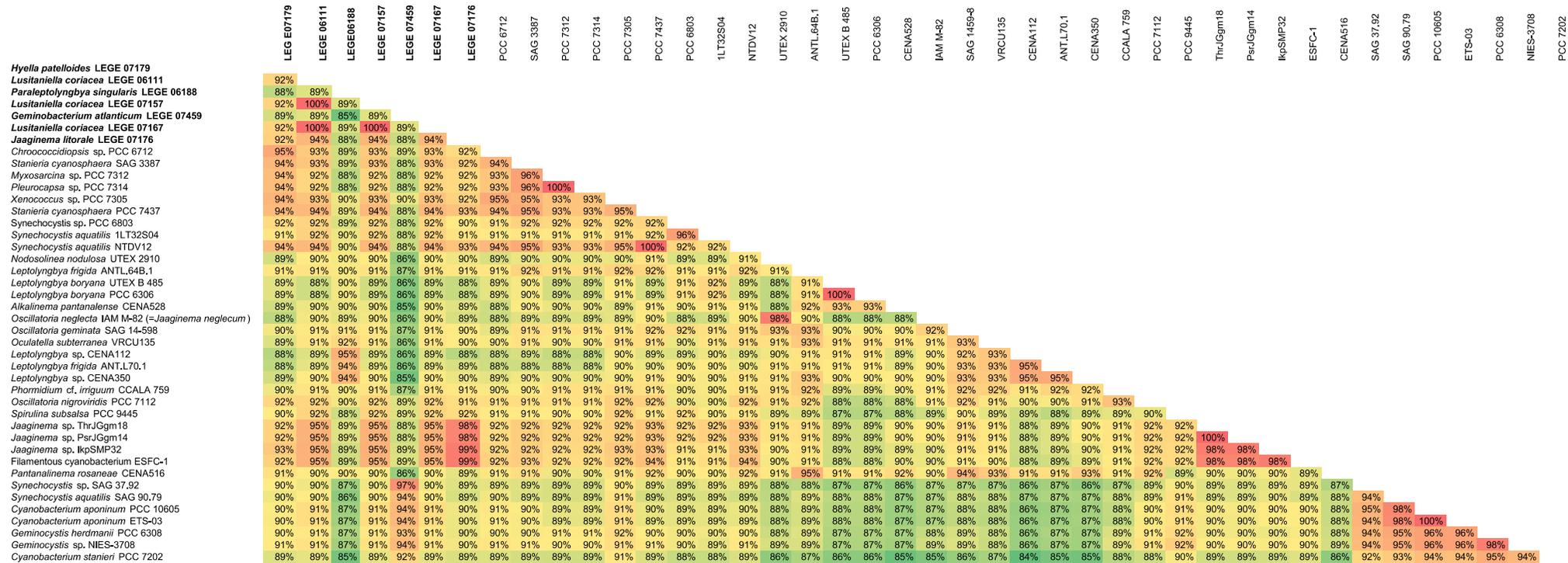
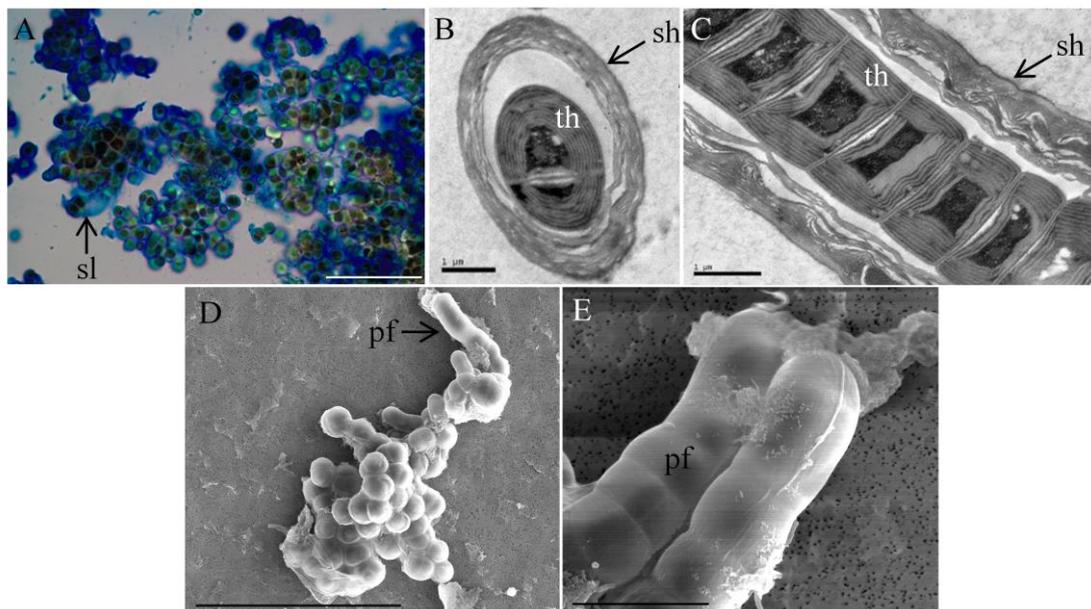


Figure 2.5. Similarity between cyanobacterial strains based on 16S rRNA gene sequences. The seven LEGE isolates are indicated in bold.

LEGE 07459 displays relatively low 16S rRNA gene sequence similarities (Fig. 2.5) to the Type strains *Cyanobacterium stanieri* PCC 7202 (92%), *Geminocystis herdmanii* PCC 6308 (93%), and to *Synechocystis* sp. PCC 6803 (88%), the Reference strain for cluster 2.1 in the Bergey's Manual (Castenholz 2001). LEGE 07459 shares 97% similarity with *Synechocystis* sp. SAG 37.92, while both strains share only 94% of similarity with the 16S rRNA gene sequences from *Synechocystis aquatilis* SAG 90.79 and *Cyanobacterium aponinum* isolates (Fig. 2.5).

The morphological and ultrastructural characteristics, low similarities at 16S rRNA gene sequences level, phylogenetic position and ITS structure revealed the distinctive features of this isolate and enable the proposal of LEGE 07459 as a new genus (*Geminobacterium* gen. nov.) as well as the description of its Type species (*Geminobacterium atlanticum* sp. nov.). Since the freshwater strain *Synechocystis* sp. SAG 37.92 clusters together with our isolate and shows 97% similarity (16S rRNA gene sequence), it is possible that SAG 37.92 is a new species within the recently established genus *Geminobacterium*, however this hypothesis requires further studies.



**Figure 2.6.** Micrographs of *Hyella patelloides* LEGE 07179. (A) Light microscopy showing the presence of slime (sl) surrounding the cells stained with Alcian blue; (B) and (C) TEM ultrathin sections of cells and a pseudofilament, respectively. The arrangement of the thylakoids (th) and the thick stratified sheath (sh) are indicated; (D) and (E) SEM images where it is possible to observe the colonial morphology, an emerging pseudofilament (pf), and detail of pseudofilaments (pf), respectively. Scale bars: (A) = 50 µm; (B, C) = 1 µm; (D) = 60 µm; (E) = 10 µm.

***Hyella patelloides* Ramos, Brito et Kaštovský sp. nov. (Fig. 2.6)**

Thallus brown, rarely red- or violet brown. Proximal part of pseudofilament uni-, bi- or multiseriate, with more or less spherical, polygonal rounded or oval cells 2.5 - 5.5 µm in diameter. Distal part of thallus usually uniseriate, cells club-shaped or cylindrical, 3 - 5 µm wide and up to 10 µm length (terminal cells rarely up to 20). Colourless multi-layered sheath, easily observed by TEM (Fig. 2.6, B and C). Colonies held together by extracellular polymeric

substances (EPS) (Fig. 2.6, A) from which the pseudofilaments emerge (Fig. 2.6, C to E). Cells divide by binary fission, reproduction by baeocytes. The thylakoids arrangement is parietal (Fig. 2.6, B and C). Contains PE and PC but high PE:PC ratio (Supplementary Fig. 2.11). 16S-23S rRNA ITS region 657 nucleotides long: D1-D1' helix with 69 nucleotides; Box-B 31 nucleotides long with 6 nucleotides in the terminal loop; tRNA<sup>Ala</sup> and tRNA<sup>Ala</sup> with 74 and 73 nucleotides, respectively (Supplementary Table 2.6).

*Type:* Portugal, Moledo do Minho beach, N 41°50'58.68''/W 8°52'0.18'' collected by Rui Seabra and Vitor Ramos, 21/May/2007, Holotype CBFS A-075-1, Herbarium for Nonvascular Cryptogams at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic. Isotype PO-T4754, Herbarium of University of Porto, Porto, Portugal. Type strain LEGE 07179.

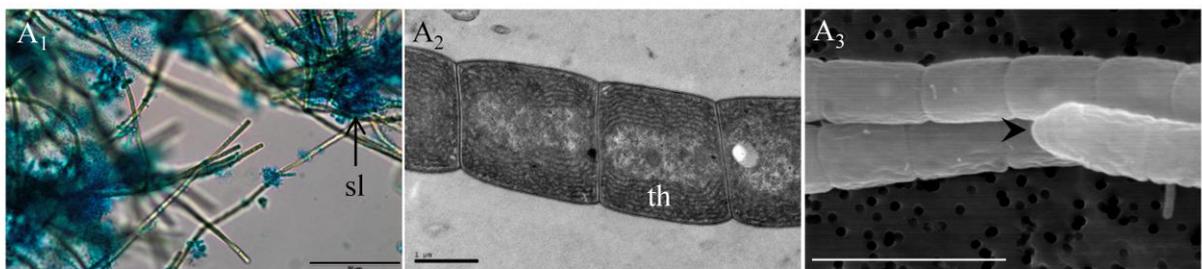
*Etymology:* patelloides - referring to the ecology.

*Ecology and distribution:* This taxon occurred rarely as epizoic or euendolithic on shells of the marine gastropod mollusc *Patella* sp..

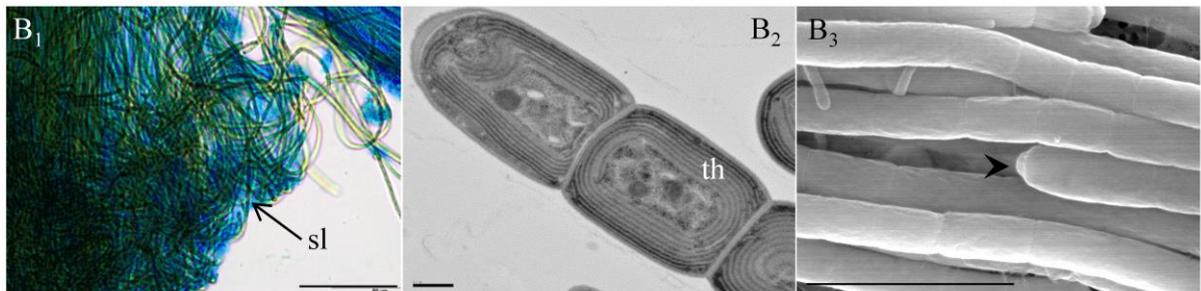
*Comparison with similar taxa:* Some members of the cyanobacterial genus *Hyella* Bornet & Flahault have been described as endoliths penetrating calcareous substrates (e.g. shells) (Al-Thukair 2011; Al-Thukair & Golubic 1991; Castenholz 2001; Ramírez-Reinat & Garcia-Pichel 2012). *H. patelloides* differs significantly from other known species. The morphology of thallus resembles slightly the one of the freshwater *H. fontana* Huber et Jardin, but the cells of *H. patelloides* are distinctly smaller and display a different colour (and ecology). The arrangement of thylakoids (parietal) differs from the one from the Type species *Hyella caespitosa* Bornet et Flahault which is irregular (Komárek & Anagnostidis 1998). This observation suggests that the genus *Hyella* is not monophyletic, although this requires to be clarified. The pigment composition analyses indicates the presence of both PE and PC but with a high PE:PC ration, consistent with the colour exhibited by our strain (Supplementary Fig. 2.11). In the phylogenetic analysis based on 16S-ITS-23S rRNA sequences, *Hyella patelloides* clusters with *Stanieria cyanosphaera* PCC 7437, a member of Pleurocapsales but it is a very divergent sequence (Supplementary Fig. 2.12). In the 16S rRNA phylogenetic tree (Fig. 2.4) this isolate is within a coherent cluster composed by members of Pleurocapsales together with *Chroococidiopsis* sp. PCC 6712. LEGE 07179 only shares 95% similarity (16S rRNA gene sequence) with its closest related *Chroococidiopsis* sp. PCC 6712 (Fig. 2.5), and 94% with the others Pleurocapsales strains. *Hyella* is a highly understudied genus (Komárek et al. 2014), and our data are the first molecular data available for the *Hyella* genus (Brito et al. 2012; Komárek et al. 2014). The different regions of ITS of LEGE 07179 were identified (Supplementary Table 2.6) but due to

the lack of data regarding 16S-23S rRNA ITS for others *Hyella* specimens, no comparisons were possible. Despite the lack of molecular data, the ecological, morphological and ultrastructural data demonstrated that LEGE 07179 possess distinctive characteristics comparing with other *Hyella* species, and allow us to propose LEGE 07179 as a new species of *Hyella* (*Hyella patteloides* sp. nov.). Nevertheless, the phylogenetic position of our *Hyella* within the Pleurocapsales (Fig. 2.4) supports the monophyly of this cyanobacterial Order, as redefined by Komárek et al. (2014).

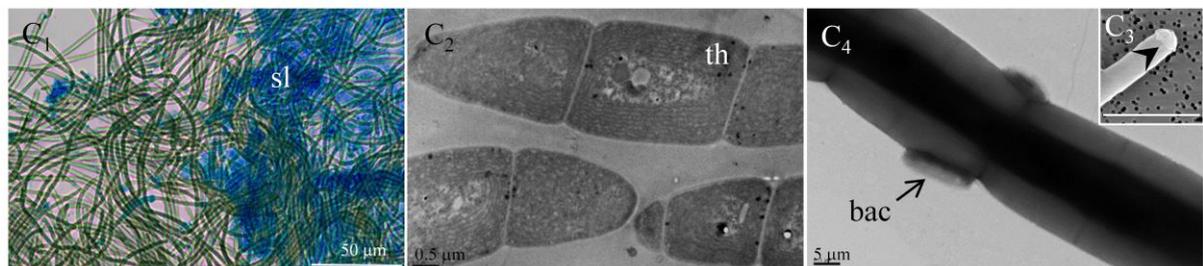
A: LEGE 07167



B: LEGE 07157



C: LEGE 06111



**Figure 2.7.** Micrographs of three strains of *Lusitaniella coriacea* LEGE 07167 (A), LEGE 07157 (B), and LEGE 06111 (C). (A1, B1, C1) Light microscopy of the filaments stained with Alcian blue, slime (sl); (A2, B2, C2) TEM longitudinal section of the filaments in which are evident the parietal arrangement of thylakoids (th); (A3, B3, C3) SEM images showing the shape of the cells, terminal cells, and calyptra (arrow head). (C4) Negative staining showing bacteria (bac) embed in the thin sheath of LEGE 06111. Scale bars: (A1, B1, C1) = 50  $\mu\text{m}$ ; (A2) = 1  $\mu\text{m}$ ; (B2, C2) = 0.5  $\mu\text{m}$ ; (A3) = 4  $\mu\text{m}$ ; (B3) = 6  $\mu\text{m}$ ; (C3) = 5  $\mu\text{m}$ , (C4) = 0.5  $\mu\text{m}$ .

***Lusitaniella* Ramos, Brito et Kaštovský, gen. nov. (Fig. 2.7)**

Thin trichomes (less than 3  $\mu\text{m}$  width) composed by cylindrical, isodiametric or elongated cells and more or less conical terminal cells. Large amount of slime can be observed surrounding the

filaments (Fig. 2.7, A1, B1 and C1). Parietal arrangement of the thylakoids (Fig. 2.7, A2, B2 and C2). Reproduction by hormogonia, necridic cells not observed. Motile.

*Type species: Lusitaniella coriacea* Ramos, Brito et Kaštovský (see below).

*Etymology:* Lusitania = the Latin name of Portugal in Roman times, refers to locality of occurrence.

***Lusitaniella coriacea* Ramos, Brito et Kaštovský sp. nov. (Fig. 2.7)**

Trichomes straight or curved, irregularly clustered or in bundles, forming "leathery" mats, motile. Cylindrical cells 1.5-2.8 µm wide, 2-6 µm long, isodiametric or elongated (Fig. 2.7, A2, B2 and C2). The terminal cells are more or less conical (Fig. 2.7, C2), sometimes rounded with calyptra (Fig. 2.7, A3, B3 and C3). Slime surrounding the filaments can be observed (Fig. 2.7 A1, B1, and C1), structured sheaths not visible. Contains PE and PC (Supplementary Fig. 2.11). 16S-23S rRNA ITS region 444 nucleotides long: D1-D1' helix with 98 nucleotides; Box-B 35 nucleotides long with 4 nucleotides in the terminal loop; tRNA<sup>Ile</sup> with 74 nucleotides (Supplementary Table 2.6).

*Type:* Portugal, Lavadores beach (Gaia), N 41° 07' 45.07"/W 8° 40' 6.88", collected by Rui Seabra and Vitor Ramos, 15/May/2007, Holotype CBFS A074-2, Herbarium for Nonvascular Cryptogams at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic. Isotype PO-T4752, Herbarium of University of Porto, Porto, Portugal. Paratypes: CBFS A074-1 and PO-T4751 (LEGE 06111), CBFS A-074-3 and PO-T4753 (LEGE 07157). Type strain LEGE 07167.

*Etymology:* coriacea – leather like in Latin, referring to the shape and consistency of the mats.

*Ecology and distribution:* This taxon occurred abundant as epilithic mats, on the surface of (immersed) rocks or in tide puddles, not only found in type locality.

*Comparison with similar taxa:* The isolates LEGE 07167, 07157 and 06111 were collected from two different sampling sites in Portugal (Northwest - LEGE 07167, 07157 and South - LEGE 06111) and previously assigned to *Leptolyngbya fragilis*, *L. mycoidea*, and *Phormidium/Leptolyngbya*, respectively (Brito et al. 2012). However, in the phylogenetic analysis based on 16S-ITS-23S rRNA gene sequences they are in a distinct sub-cluster, supported by a strong bootstrap value (Supplementary Fig. 2.12). This tight sub-cluster is within a clade composed by somewhat divergent sequences from *Spirulina subsalsa* PCC 9445 and from another LEGE isolate - LEGE 07176 here studied, see below. In agreement, in the tree

based on 16S rRNA gene sequences only (Fig. 2.4), the three isolates appear in a robust and tight sub-cluster together with LEGE 07162 (Brito et al. 2012). LEGE 07162 is no longer available in LEGE Culture Collection and therefore could not be included in the present study. In any case, LEGE 07167, LEGE 07157 and LEGE 06111 form a well-supported distinct sub-cluster, and are phylogenetically apart from any other cyanobacteria, despite being in a large sub-clade with *Jaaginema* and *Spirulina* strains. The three LEGE strains are 100% similar concerning the 918 bp 16S rRNA sequence analysed (Fig. 2.5), and their closest relatives are the strains from the *Jaaginema* and *Spirulina* sub-clades but with only 95% and 92% of sequence similarities, respectively. The molecular data highlighted these three isolates and led to the description of the new genus (*Lusitaniella* gen. nov.) as well as the description of the Type species (*Lusitaniella coriacea* sp. nov.). However, there are no distinctive morphological characteristics, being *Lusitaniella* a typical cryptogenus (Komárek 2016; Komárek et al. 2014) – i.e. high molecular but no morphological support. As a consequence, the strains belonging to this genus were previously identified as *Leptolyngbya* or *Phormidium* (Brito et al. 2012), two genera known to be polyphyletic and most probably encompassing several cryptic species or even genera (Casamatta et al. 2003; Komárek et al. 2014). In fact, these two taxa are highlighted by Komárek (2016) as examples of traditional genera in which several cryptogenera are prone to exist. Not surprisingly, several new genera emerging from these two taxa have been recently proposed e.g. *Pantalinema* Vaz et al.; *Alkalinema* Vaz et al.; *Oculatella* Zammit, Billi et Albertano; *Kamptonema* Strunecký, Komárek et Smarda; and *Ammassolinea* Hašler et al. (Komárek et al. 2009; Strunecký et al. 2010; Johansen et al. 2011; Zammit et al. 2012; Hašler et al. 2014; Osorio-Santos et al. 2014; Strunecký et al. 2014; Vaz et al. 2015). The three *Lusitaniella coriacea* strains have 16S-23S rRNA ITS sequences with the same length (444 nucleotides) and they contain only tRNA<sup>Ala</sup> (Supplementary Table 2.6). The D1-D1' helix has 98 nucleotides, longer than those of other strains, e.g. *Leptolyngbya* strains (51-65 nucleotides) and *Phormidium* strains (57-58 nucleotides). The Box-B is 35 nucleotides long, within the range of those of *Leptolyngbya* strains (11 - 41 nucleotides) and *Phormidium* (36-38 nucleotides) (Johansen et al. 2011; Hašler et al. 2012). The pigment composition analysis of the three *Lusitaniella coriacea* strains indicated the presence of both PE and PC in slightly different ratios (Supplementary Fig. 2.11).

### ***Jaaginema litorale* Ramos, Brito et Kaštovský sp. nov. (Fig. 2.8)**

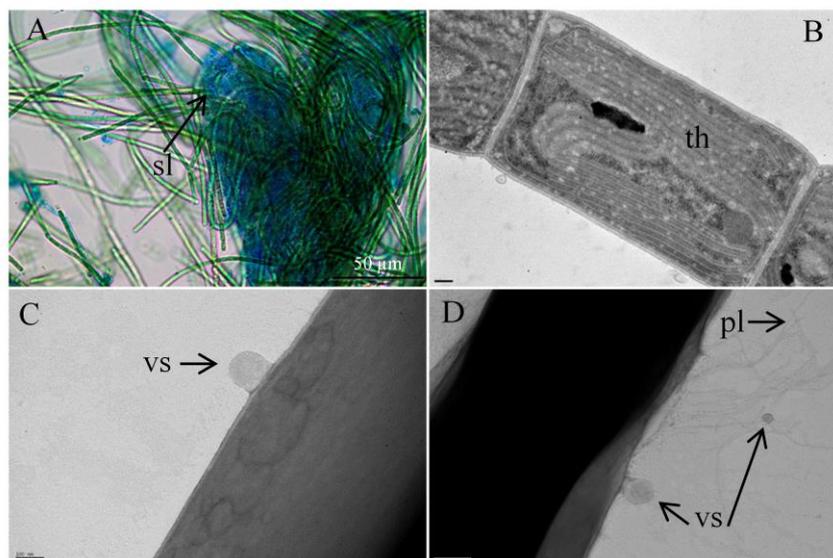
Thallus filamentous, blue-green (Fig. 2.8, A), without sheath or with very fine mucilaginous layer around the trichome, immotile, forming a firm biofilm. Trichomes slightly constricted at the cross walls. Cylindrical cells 1.2 - 2.1 µm wide and 2 - 4 µm long (Fig. 2.8, B), apical cells shortly narrowed or rounded, without calyptra. Parietal arrangement of the thylakoids (Fig. 2.8, B). Contains PC but not PE (Supplementary Fig. 2.11). 16S-23S rRNA ITS region 447 nucleotides

long: D1-D1' helix with 67 nucleotides; Box-B 30 nucleotides long with 5 nucleotides in the terminal loop; tRNA<sup>Ala</sup> with 74 nucleotides (Supplementary Table 2.6).

*Type:* Portugal, Aguda beach (Gaia), N 41° 02' 58.35"/W 8° 39' 19.22" collected by Rui Seabra and Vitor Ramos, 15/May/2007, Holotype CBFS A-076-1, Herbarium for Nonvascular Cryptogams at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic. Isotype PO-T4755, Herbarium of University of Porto, Porto, Portugal. Type strain LEGE 07176.

*Etymology:* litorale – referring to ecology.

*Ecology and distribution:* This taxon occurred rarely as epilithic on wave-exposed rocks in intertidal zones.



**Figure 2.8.** Micrographs of *Jaaginema litorale* LEGE 07176. (A) Aggregated filaments stained with Alcian blue highlighting the presence of slime (sl); (B) TEM ultrathin section showing the parietal arrangement of the thylakoids (th); (C) and (D) Negative staining showing the presence of vesicles (vs) emerging from the outer membrane and the presence of thin pilus-like structures (pl). Scale bars: (A) = 50 µm; (B) = 0.2 µm; (C) = 0.1 µm; (D) = 0.2 µm.

*Comparison with similar taxa:* The isolate LEGE 07176 was previously assigned to *Leptolyngbya fragilis* (Brito et al. 2012). However, this study showed that this isolate could fit by phylogeny and morphology to the genus *Jaaginema* Anagnostidis et Komárek, exhibiting however some differences from the already known marine species: *J. cavanillesianum* (González Guerrero) Anagnostidis, isolated from the Cadiz isthmus in Iberian Peninsula has narrow trichome (1-1.5 µm wide) and acute conical apical cells. *J. longarticulatum* Anagnostidis & Komárek, epiphytic, isolated from the Adriatic Sea is also narrower (1-1.5 µm wide), *J. crassum* (Woronichin) Anagnostidis is larger (2.5-3 µm wide), and *J. woronichinii* (Anissimova) Anagnostidis & Komárek has longer cells (2.8-7.5 µm long) and is unconstricted at cross-walls (Komárek &

Anagnostidis 1998). The pigment composition of LEGE 07176 indicates the presence of PC but not PE (Supplementary Fig. 2.11). In negatively stained LEGE 07176 cells, vesicles emerging from the outer membrane were observed (Fig. 2.8, C and D). It is known that many bacteria can release vesicles that play a role in different processes such as quorum signalling (Mashburn & Whiteley 2005) and horizontal gene transfer (Kolling & Matthews 1999; Rumbo et al. 2011). Recently, the formation of extracellular vesicles was reported in the unicellular marine cyanobacteria *Prochlorococcus* (Biller et al. 2014) and in the filamentous freshwater *Anabaena* sp. PCC 7120 strain (Oliveira et al. 2015), indicating that a wide range of cyanobacteria have the ability to produce and release vesicles that, most probably, play an important role in the interactions between the microorganisms and the environment (Biller et al. 2014).

In the phylogenetic analysis based on 16S-ITS-23S rRNA gene sequences, and as mentioned before, LEGE 07176 is within a cluster (100% bootstrap value) composed by *Spirulina subsalsa* PCC 9445 and the *Lusitaniella coriacea* isolates LEGE 07167, LEGE 07157 and LEGE 06111, but forming a distinct branch appearing as a “loner sequence” (Supplementary Fig. 2.12). In the tree based on 16S rRNA gene sequences only (Fig. 2.4), this isolate is within a cluster (99% bootstrap) composed by *Spirulina* strains and together with the filamentous cyanobacterium ESFC-1 and three *Jaaginema* strains in a well-supported cluster (100% bootstrap value). The filamentous cyanobacterium ESFC-1 was collected from microbial mats, Elkhorn Slough estuary, CA, USA (Woebken et al. 2012; Everroad et al. 2016) and the three *Jaaginema* strains were isolated from thermal springs (Greece) (Bravakos et al. 2016). LEGE 07176 shares 98-99% similarity (16S rRNA gene sequence) with the *Jaaginema* strains and 99% with the filamentous cyanobacterium ESFC-1 (Fig. 2.5). Other sequences of *Jaaginema* strains, some originally named by their basionyms (*Oscillatoria neglecta* = *Jaaginema neglectum* (Lemmermann) Anagnostidis et Komárek and *Oscillatoria geminata* = *Jaaginema geminatum* (Schwabe ex Gomont) Anagnostidis et Komárek), are also included in the tree but are placed distantly from these isolates. In this analysis, we did not include the available sequence of a strain assigned to the Type species, *Jaaginema subtilissimum* (Kützing ex Forti) Anagnostidis et Komárek, due to its short length sequence (635 bp). However, an additional phylogenetic analysis was performed with all *Jaaginema* sequences available in GenBank (including the one assigned to the Type species) (Fig. 2.13). In this tree, it is possible to observe that *Jaaginema* strains are distributed along the tree and that *J. subtilissimum* BDU 92183 is placed distantly from the clade composed by our strain, the filamentous cyanobacterium ESFC-1, and the five strains from hot springs: the three *Jaaginema* strains from Greece, reported by Bravakos et al. (2016), and two *J. minimum* isolates from Iran. All these data suggest that the traditional, morphological-based *Jaaginema* genus is polyphyletic, deserving further taxonomic revision. It is important to notice that the three *Jaaginema* strains of Bravakos et al. (2016) were morphologically assigned to this genus but this was not supported by molecular data. The best BLAST hit for the 16S rRNA gene sequences of these strains is the estuarine filamentous

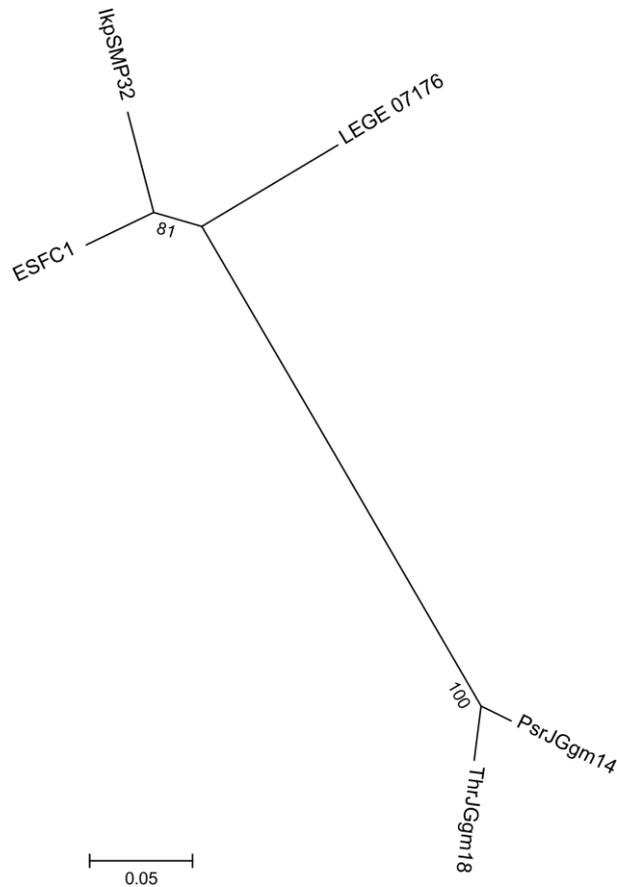
cyanobacterium ESFC-1 that appears in the same cluster (Bravakos et al. 2016), in accordance with our data. Furthermore, according with the authors, the three Greek strains are morphologically similar to *Oscillatoria geminata* (= *Jaaginema geminatum* (Schwabe ex Gomont) Anagnostidis et Komárek). However, the only sequence available in GenBank for this species (strain SAG 1459-8) has only 90% similarity with the Greek strains and, in accordance, our isolate shares the same percentage of identity (Fig. 2.13). Moreover, the recently described filamentous cyanobacterium ESFC-1 (Everroad et al. 2016), was included in Spirulinales, since *Spirulina subsalsa* PCC 9445 was its closest relative (Komárek et al. 2014). This order is characterized by typical screw-like coiled trichomes and its members have a particular phylogenetic position. However, ESFC-1 and although it clusters together with *Spirulina* strains, exhibits straight to slightly curved trichomes with rounded to slightly conical ends. According to Everroad et al. (2016), ESFC-1 is morphological similar to *Geitlerinema* with a cell size typical of *Leptolyngbya*. It also important to notice that the overlapping between the genera *Geitlerinema* and *Jaaginema* has been previously reported (Anagnostidis 1989). Although our isolate LEGE 07176, ESFC-1, and the three *Jaaginema* strains cluster together with *Spirulina* strains, they are in a clearly different sub-cluster. In addition, neither LEGE 07176, ESFC-1, nor the three *Jaaginema* strains display the typical screw-like coiled trichomes. Altogether, the data suggest that ESFC-1 strain does not belongs to Spirulinales and that together with our strain and the other three *Jaaginema* strains might represent a new order. This certainly requires further investigation and revision of the Order Spirulinales.

Since phylogenies based on the 16S-23S rRNA ITS region are commonly used to establish cyanobacterial ecotypes or species (Rocap et al. 2002; Sciuto & Moro 2016), an additional phylogenetic analysis based on this hypervariable region, including all strains within the sub-clade of LEGE 07176, was performed. The obtained unrooted tree (Fig. 2.9) clearly shows three main lineages, one composed by the Greek strains PsrJGgm14 and ThrJGm18, another with the other Greek isolate lkpSMP32 and the estuarine strain ESFC-1, and the third composed only by our strain. Furthermore, taking into account the 16S-23S rRNA ITS secondary structures it was possible to observe D1-D1' helixes with three different patterns. One shared by PsrJGgm14 and ThrJGm18 (66 nucleotides), another by lkpSMP32 and the filamentous cyanobacterium ESFC-1 (68 nucleotides), and a third corresponding to our strain (67 nucleotides). The Box-B is similar in all strains and all contain only the tRNA<sup>Ala</sup> (Table 2.3).

**Table 2.3** Comparison of the predicted secondary structures of the 16S-23S rRNA internal transcribed spacers (ITS) from our isolate *Jaaginema litorale* LEGE 07176 and other strains within the same sub-clade of the 16S rRNA-based phylogenetic tree (see Fig. 2.4)

| Strain                                  | D1-D1' | Box-B | tRNA |
|---|--------|-------|------|
| <i>Jaaginema litorale</i><br>LEGE 07176 |        |       | Ile  |
| Filamentous<br>cyanobacterium<br>ESFC-1 |        |       | Ile  |
| <i>Jaaginema</i> sp.<br>IkpSM32         |        |       | Ile  |
| <i>Jaaginema</i> sp.<br>PsrJGm14        |        |       | Ile  |
| <i>Jaaginema</i> sp.<br>ThrJGm18        |        |       | Ile  |

The morphological and phylogenetic data demonstrated that LEGE 07176 could fit into the genus *Jaaginema* Anagnostidis et Komárek, however this genus is most probably polyphyletic deserving further taxonomic revision. The ecology and the inferred phylogeny based on the ITS, as well as the ITS secondary structures, highlight the differences between our isolate and its closest relatives. Therefore, in this work we propose LEGE 07176 as a new species of *Jaaginema* (*Jaaginema litorale* sp. nov.).



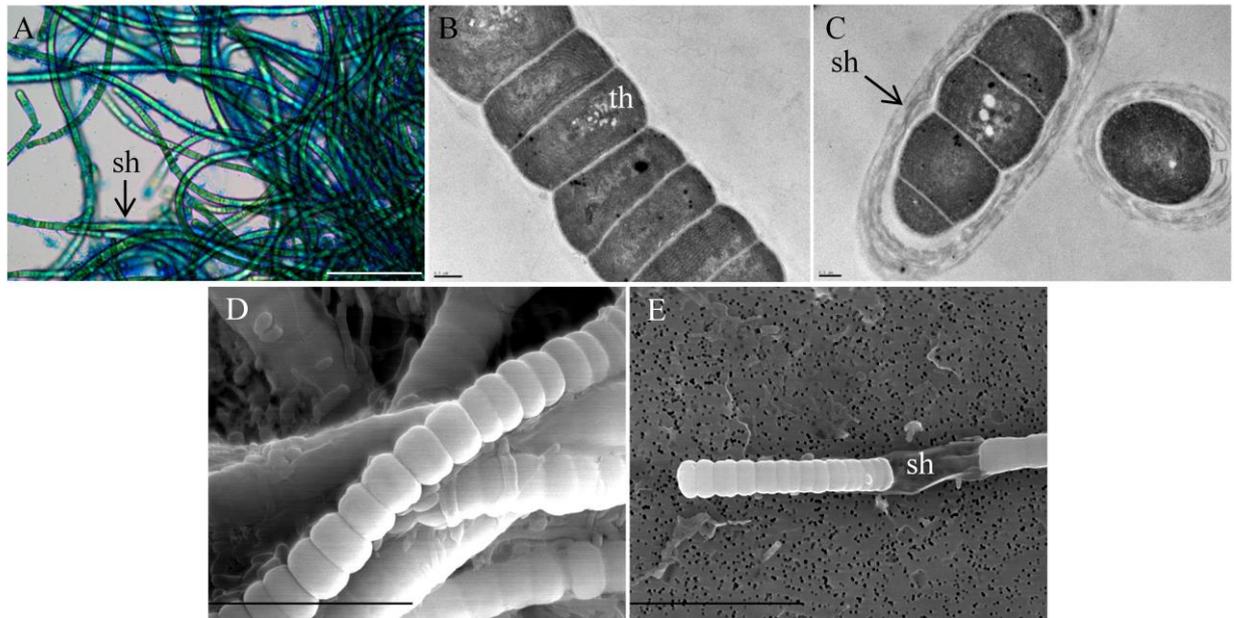
**Figure 2.9.** Unrooted maximum-likelihood phylogram for 16S-23S rRNA internal transcribed spacer (ITS) sequences of *Jaaginema littorale* LEGE 07176 and other strains within the same sub-clade of the 16S rRNA-based phylogenetic tree (see Fig. 2). ESFC-1: filamentous cyanobacterium; PsrJGgm14, ThrJGgm18 and IkpSMP32: other *Jaaginema* strains.

***Calenema* Ramos, Brito et Kaštovský, gen. nov. (Fig. 2.10)**

Trichomes with parietal arrangement of thylakoids (distinguishable centro- and chromatoplasm), constricted at the cross walls, cells shorter than wide or almost isodiametric. End of filament is distinctly wider and cells are more constricted than the normal vegetative cells. Colourless fine sheath, sometimes wider and layered. Reproduction by hormogonia, necridic cells present.

*Type species: Calenema singularis* Ramos, Brito et Kaštovský (see below).

*Etymology:* Cale – the Latin name of the city where it was collected, in Roman times; nema – referring to thread, filament.



**Figure 2.10.** Micrographs of *Calenema singularis* LEGE 06188. (A) Light micrograph of the filaments stained with Alcian blue highlighting the sheath (sh); (B) and (C) TEM ultrathin sections showing the parietal arrangement of the thylakoids (th) and the stratified sheath (sh); (D) and (E) SEM images where it is possible to observe the barrel shaped cells and the breakage of a trichome with the two parts still within the sheath (sh). Scale bars: (A) = 50  $\mu\text{m}$ ; (B, C) = 0.5  $\mu\text{m}$ ; (D) = 6  $\mu\text{m}$ ; (E) = 10  $\mu\text{m}$ .

***Calenema singularis* Ramos, Brito et Kaštovský sp. nov. (Fig. 2.10)**

Filaments forming tuft-like aggregates (Fig. 2.10, A), trichome olive green to dark brown, (1.5 - 1.9 - 4 (5.5)  $\mu\text{m}$  wide, cells length irregular (0.5 - 5  $\mu\text{m}$ ), mostly shorter than wide or isodiametric (Fig. 2.10, B and D). Cells barrel shaped with sharp constrictions at the cells junctions (Fig. 2.10, B to E), and at the end of filaments are deeply constricted, more spherical and slightly wider than normal cells, later separated by necridic cells as hormogonia (Fig. 2.10, E). Colourless sheath, well defined sometimes wider and layered (more visible with TEM; Fig. 2.10, C). Contains the phycobilin pigments phycoerythrin (PE) and phycocyanin (PC) (Supplementary Fig. 2.11). 16S-23S rRNA ITS region 553 nucleotides long: D1-D1' helix with 67 nucleotides; Box-B 56 nucleotides long with 3 nucleotides in the terminal loop; tRNA<sup>Ile</sup> and tRNA<sup>Ala</sup> with 74 and 73 nucleotides, respectively (Supplementary Table 2.6).

*Type:* Portugal, Lavadores beach (Gaia), N 41° 07' 45.07"/W 8° 40' 6.88", collected by Rui Seabra and Vitor Ramos, 22/November/2006, Holotype CBFS A-072-1, Herbarium for Nonvascular Cryptogams at the Department of Botany, Faculty of Science, University of South Bohemia, Czech Republic. Isotype PO-T4749, Herbarium of University of Porto, Porto, Portugal. Type strain LEGE 06188.

*Etymology:* singularis, Lat. unique

*Ecology and distribution:* This taxon occurred rarely in coastal sea water in surf zone (tychoplankton?).

*Comparison with similar taxa:* The isolate LEGE 06188 is morphologically very similar to *Leptolyngbya* Anagnostidis et Komárek. The distinct diacritical features are the wide of the trichomes (older trichomes of *Calenema* are sometimes wider than 5 µm in contrast with typical *Leptolyngbya* trichomes that reaches cca. 3 µm) and the slightly deeper constriction at the cells walls. The pigment composition analysis indicated the presence of both PE and PC (Supplementary Fig. 2.11).

In the phylogenetic analysis based on 16S-ITS-23S rRNA gene sequences (Supplementary Fig. 2.12), *Calenema singularis* LEGE 06188 is placed within a clade along with the *Leptolyngbya* sub-clade (*sensu stricto*; contains representative strains of the Type species of the genus, *L. boryana* Anagnostidis et Komárek; see Vaz et al. (2015)), despite a noticeable sequence divergence. However, in the tree based on the 16S rRNA (Fig. 2.4) LEGE 06188 is still placed within a clade with *Leptolyngbya*-like cyanobacteria (including some of the newly established genera derived from it, e.g. *Oculatella*) but it is clearly not closely related with the *Leptolyngbya sensu stricto* subclade. In fact, the 16S rRNA gene sequence similarity between our strain and *Leptolyngbya boryana* PCC 6306 or *L. boryana* UTEX B 485 is only 90% (Fig. 2.4). In addition, LEGE 06188 has low similarity with other cyanobacterial sequence, sharing 95% similarity with its closest relative, the Brazilian freshwater strain *Leptolyngbya* sp. CENA 112 (Furtado et al. 2009), and 94% with the strains *Leptolyngbya* sp. CENA 350 (subaerial, from Brazil) and *Leptolyngbya frigida* ANT.L70.1 from an Antarctic lake (Taton et al. 2006). Since these values are within the boundaries to distinguishing genera (Kim et al. 2014; Yarza et al. 2014), these strains could be placed within the genus *Calenema*, as different species. However, this requires further investigation.

As already mentioned before, *Leptolyngbya* is a polyphyletic genus, and recently new genera have been emerging e.g. *Pantanalinema*, *Alkalinema*, *Oculatella*, *Nodosilinea*, allowing a clearer view on the systematics of the family Leptolyngbyaceae (Komárek et al. 2014). Therefore, sequences from members of these genera were included and coherent groups were observed, showing that our isolate does not fit in any of the previously established genus. Furthermore, taking into account the 16S-23S rRNA ITS secondary structures, it was possible to observe that while the D1-D1' helix is 66 nucleotides long, with a similar length of those reported to *Leptolyngbya* (51-65 nucleotides), the Box-B has 56 nucleotides, longer than the majority of those reported for *Leptolyngbya* strains (11-41 nucleotides) (Supplementary Table 2.6). Both tRNA<sup>Ile</sup> and tRNA<sup>Ala</sup> are present as reported for *Leptolyngbya* strains (Hašler et al. 2012; Johansen et al. 2011).

The morphological data demonstrated that LEGE 06188 possess distinctive characteristics comparing with other *Leptolyngbya* spp., namely wider trichomes. The phylogenetic position

and the low 16S rRNA sequence similarity with other *Leptolyngbya*, allow us to propose LEGE 06188 as a new genus (*Calenema* gen. nov.) as well as the description of its Type species (*Calenema singularis* sp. nov.).

All phylogenetic analyses were also performed using a different alignments and/or methods (for details see material and methods), and the results obtained supported the majority of the findings/conclusions presented here.

### ***nif* genes/capacity to fix atmospheric nitrogen and to produce bioactive compounds**

In this work, and confirming the PCR screening performed previously (Brito et al., 2012), partial sequences were obtained for *nifH* (dinitrogenase reductase) for the isolates *Hyella patteloides* LEGE 07179 and *Calenema singularis* LEGE 06188 (GenBank accession numbers: KC256774, KC256766). Interestingly, the *nifH* sequence of *Calenema singularis* LEGE 06188 has 87% similarity with the *nifH* of *Leptolyngbya* sp. CENA 350 (KR137611), a strain that was isolated from a bamboo phyllosphere and that is one of the phylogenetically closest relatives in the 16S rRNA gene sequence-based tree. In addition, the ability of these cyanobacteria to fix nitrogen was further evaluated by growing the strains in ASNIII medium with decreasing concentrations of combined nitrogen until depletion and the nitrogenase activity was evaluated by the acetylene reduction assay (ARA). Although the strains could grow (*Calenema singularis* LEGE 06188) or survive (*Hyella patteloides* LEGE 07179) in medium without combined nitrogen the nmoles of ethylene produced/chl *a*/h were residual compared to our control strain *Anabaena/Nostoc* sp. PCC 7120 (data not shown), indicating that the capacity to fix nitrogen might have been suppressed during the prolonged period of cultivation in media with combined nitrogen.

Recently, the capacity of these strains to produce secondary metabolites was also evaluated showing that *Hyella patteloides* LEGE 07179 probably produces a compound related to naopeptin or antanapeptin, while *Lusitaniella coriacea* LEGE 07167 and *Jaaginema litorale* LEGE 07176 might produce compounds related to hepatotoxins (Brito et al. 2012, 2015).

### **2.2.5 Conclusions**

The extensive polyphasic approach employed on isolates from a restricted area of the Portuguese Atlantic coast, led to the proposal of 3 new genera (*Geminobacterium*, *Lusitaniella*, *Calenema*) and 2 new species (*Hyella patelloides* and *Jaaginema litorale*) of marine cyanobacteria belonging to at least 3 different orders according to Komárek et al. (2014).

The Chroococcalean *Geminobacterium atlanticum* LEGE 07459 shares morphological characteristics with other unicellular cyanobacterial genera but its phylogenetic position and particular ultrastructural features definitively support the establishment of a new genus. The description of a new species of *Hyella*, *H. patelloides* LEGE 07179, includes the disclosure of

the first molecular data for members of this genus, contributing to increase the knowledge on the relatively understudied group of marine Pleurocapsales. Three of our filamentous isolates - LEGE 07167, LEGE 07157 and LEGE 06111 - are phylogenetically apart from any other cyanobacteria constituting a very distinct lineage, the incertis sedis genus *Lusitaniella*, that seems to be ubiquitous on the Portuguese coast. The newly proposed species *Jaaginema litorale* displays some distinct characteristics comparing with other *Jaaginema* species, in particular with their marine counterparts, and our data analysis indicate that this genus might be polyphyletic deserving further taxonomic revision. The Synechococcales species *Calenema singularis* possess wider trichomes than other Leptolyngbyaceae, namely than *Leptolyngbya*, and its phylogenetic position reinforced the establishment of this new genus.

The new taxa described in this study are also certainly ecologically relevant as N<sub>2</sub>-fixers and/or as promising sources of natural compounds. These findings reinforce the perception of the relevance of “culturomics” (Dubourg et al. 2013; Lagier et al. 2015) to study unexplored microbial diversity, allowing efficient and individual functional analyses and downstream investigations envisaging biotechnological applications.

### 2.2.6 Acknowledgments

This work was funded by National Funds through FCT - Fundação para a Ciência e a Tecnologia, grants SFRH/BD/70284/2010 (to AB), SFRH/BD/80153/2011 (to VR), SFRH/BD/84914/2012 (to RM), and GACR 15-11912S (to JK). To this work also contributed the project MARBIOTECH (reference NORTE-07-0124-FEDER-000047) within the SR&TD Integrated Program MARVALOR - Building research and innovation capacity for improved management and valorisation of marine resources, supported by ON.2 Program and by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Programme, through the project NOVOMAR and national funds through FCT – Foundation for Science and Technology, through the project “PEst-C/MAR/LA0015/2013”. We are also grateful to Daniela Silva from Centro de Materiais da Universidade do Porto (CEMUP), Porto, Portugal, for technical assistance with SEM and to Paulo Oliveira, IBMC, for the help with the GC measurements.

## 2.2.7 Supplementary Data

**Table 2.4** Cyanobacterial strains, culture collection code and sampling sites.

| Cyanobacterium                    | LEGE Code | Sampling Site * | Latitude/Longitude               |
|-----------------------------------|-----------|-----------------|----------------------------------|
| <i>Geminobacterium atlanticum</i> | 07459     | Martinhal       | N 37° 01' 07.30"/W 8° 55' 36.17" |
| <i>Hyella patelloides</i>         | 07179     | Moledo do Minho | N 41° 50' 58.68"/W 8° 52' 0.18"  |
| <i>Lusitaniella coriacea</i>      | 07167     | Lavadores       | N 41° 07' 45.07"/W 8° 40' 6.88"  |
| <i>Lusitaniella coriacea</i>      | 07157     | Lavadores       | N 41° 07' 45.07"/W 8° 40' 6.88"  |
| <i>Lusitaniella coriacea</i>      | 06111     | Martinhal       | N 37° 01' 07.30"/W 8° 55' 36.17" |
| <i>Jaaginema litorale</i>         | 07176     | Aguda           | N 41° 02' 58.35"/W 8° 39' 19.22" |
| <i>Calenema singularis</i>        | 06188     | Lavadores       | N 41° 07' 45.07"/W 8° 40' 6.88"  |

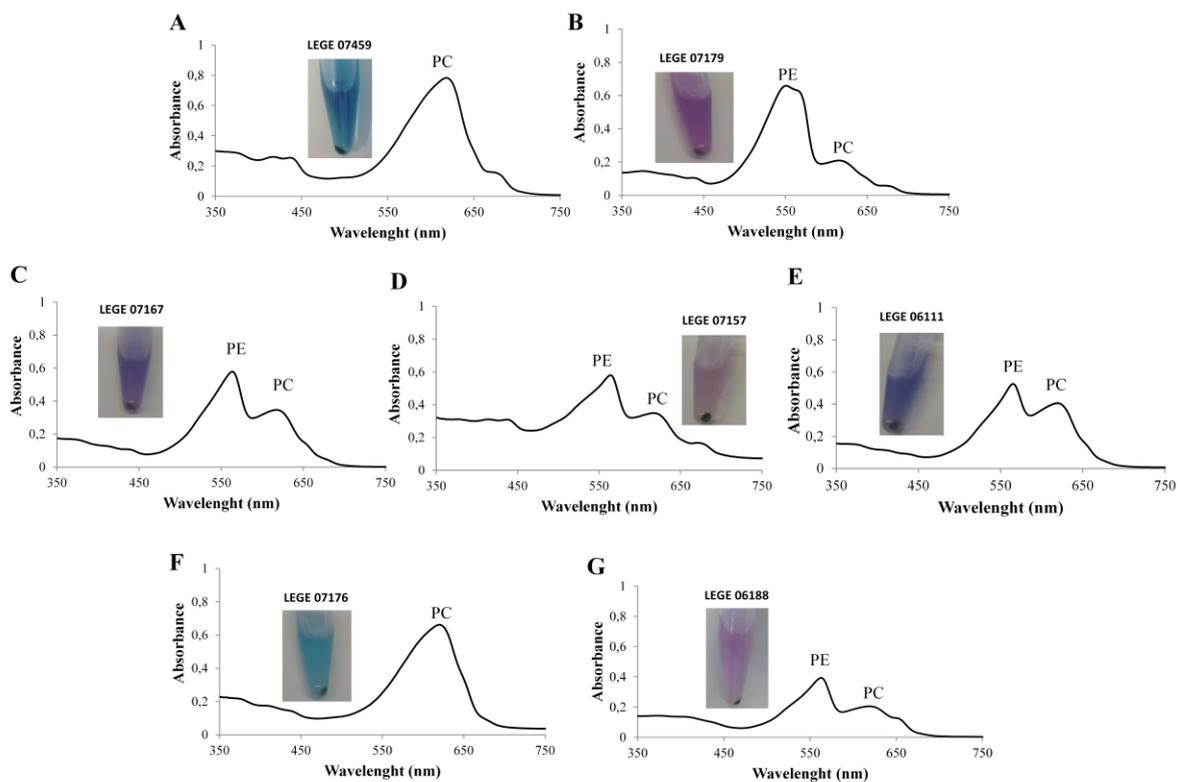
\* for details see Brito et al. (2012)

**Table 2.5** Oligonucleotide primers used in this study.

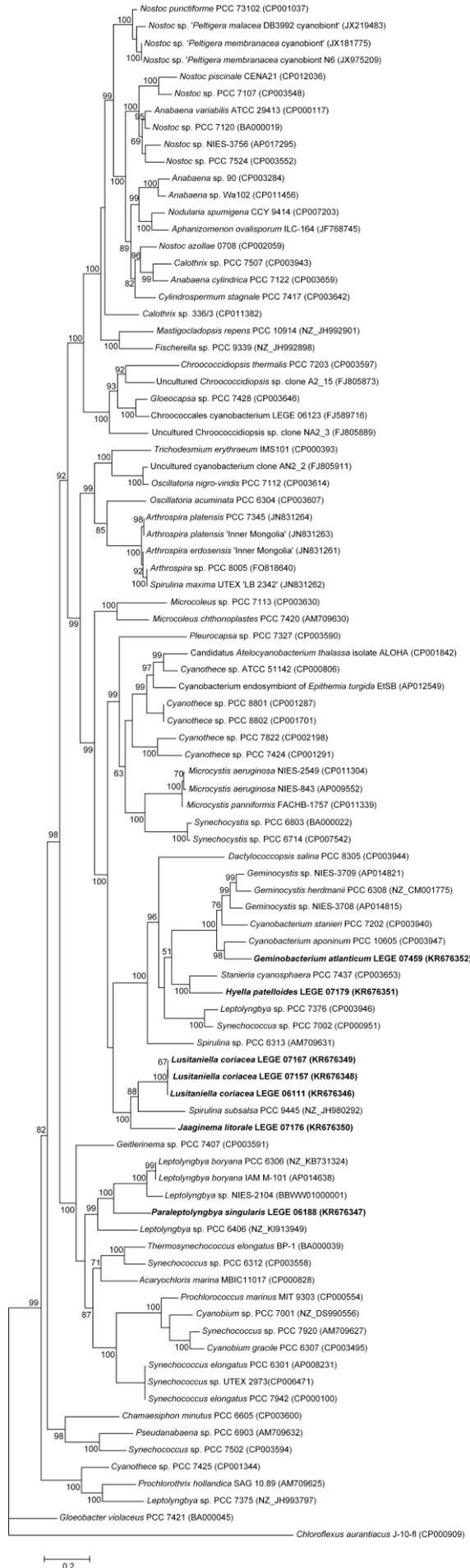
| Target gene sequence | Primer    | Sequence 5' → 3'       | Reference                |
|----------------------|-----------|------------------------|--------------------------|
| 16S rRNA             | 16SF      | CGCTAGTAATCGCAGGTCA    | This work                |
|                      | ITSR      | CGGCTACCTTGTTACGACTT   | This work                |
| rRNA ITS             | CSIF      | GYCACGCCCGAAGTCRRTAC   | Janse et al. 2004        |
|                      | ULR       | CCTCTGTGTGCCTAGGTATC   | Janse et al. 2004        |
| 23S rRNA             | ITSF      | GGAGAACGCCAATACCCA     | This work                |
|                      | ULF       | GATACCTAGGCACACAGAGG   | Ramos et al. 2010        |
|                      | 23S800F   | GGTTAGGGGTGAAATGCCAA   | Ramos et al. 2010        |
|                      | 23S867F   | GTTGGCGGATGAGGTGTGGTTA | This work                |
|                      | 23S893F   | GCGGTACCAAATCGAGACAAAC | This work                |
|                      | 23S1682F  | CTCTCTAAGGAACCTCGGCAAA | Ramos et al. 2010        |
|                      | 23S819R   | TTGGCATTTCACCCCTAACC   | This work                |
|                      | 23S1027R  | GCTGGTGGTCTGGGCTGTTT   | Ramos et al. 2010        |
|                      | 23S1932R  | CCTTAGGACCGTTATAGTTA   | Ramos et al. 2010        |
|                      | p23SrV_r1 | TCAGCCTGTTATCCCTAGAG   | Sherwood & Presting 2007 |
| <i>nifH</i>          | nifH4     | TTYTAYGGNAARGGNGG      | Omeregíe et al. 2004     |
|                      | nifH3     | ATRTTRTTNGCNGCRTA      | Omeregíe et al. 2004     |
|                      | nifH2     | ADNGCCATCATYTCNCC      | Omeregíe et al. 2004     |
|                      | nifH1     | TGYGAYCCNAARGCNGA      | Omeregíe et al. 2004     |

**Table 2.6** Sequences of the conserved and non-conserved regions of the 16S-23S rRNA internal transcribed spacer (ITS) of the seven LEGE isolates.

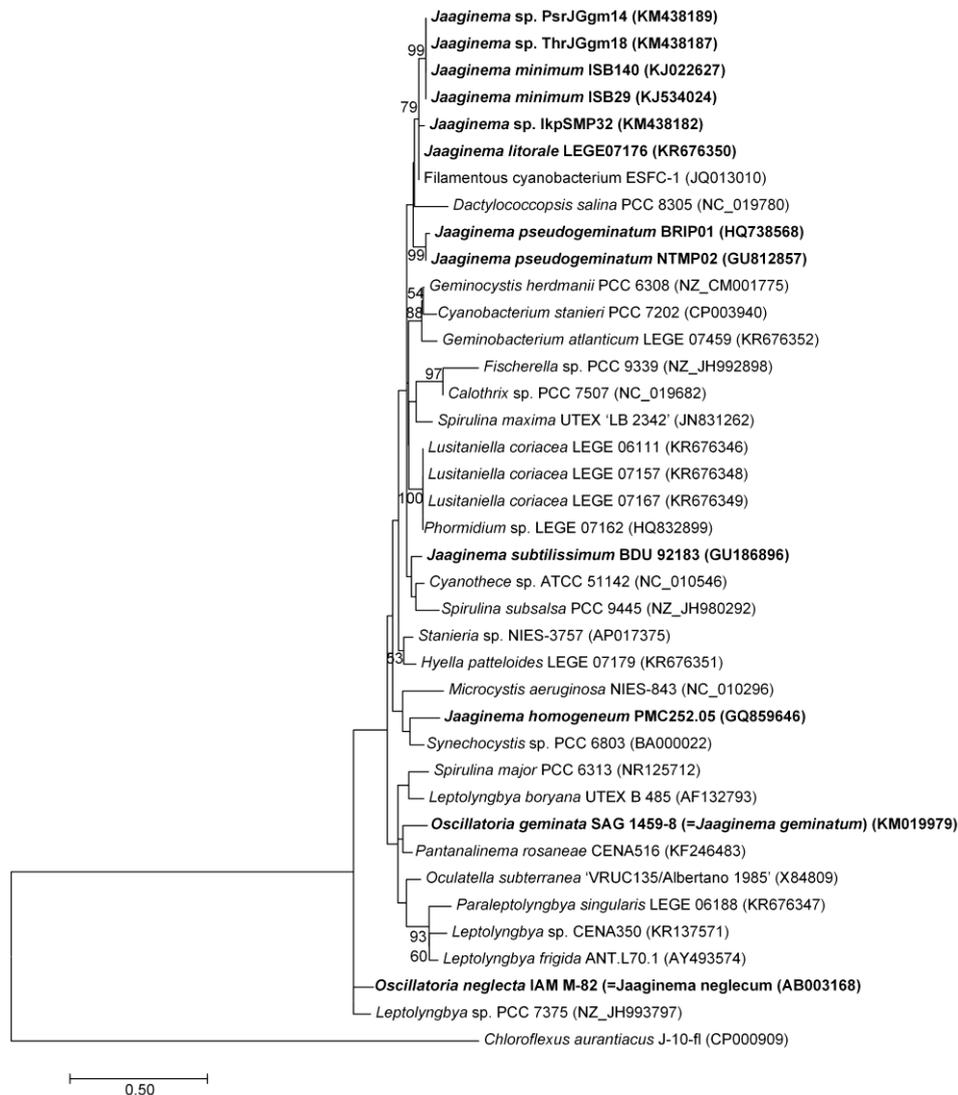
| Strain  | D1-D1'  | D2          | V2 | Box-B  | Box-A        | D4      | V3 | tRNA       |
|---|---|-------------|----|--|--------------|---------|----|------------|
| <i>Calenema singularis</i><br>LEGE 06188        | AGGGAGACUCUAAAAUAGGA<br>GUCAAGCAAGGAAAGUCG  | .           | -- | CAGCACUGUGAUGAGAAAGAGC<br>AAAUGCUG                             | GAACCUUGAAAA | CUGCAUA | -- | Ala<br>Ile |
| <i>Jaaginema litorale</i><br>LEGE 07176         | AGGGAGACCUUUAACCCU<br>UCAAGACAGUCAAAAAGUAGC<br>UUGAAGAGUCUAGACAAUCC<br>CAAGGUCG                                 | CUUCAAACUA  | -- | CAGAAGCCCCAAAGCCAGACU<br>GAGAAUGCUG                            | GAACCUUGAAAA | CUGCAUA | -- | Ala<br>Ile |
| <i>Lusitaniella coriacea</i><br>LEGE 06111      | AGGGAGACCUAACUGAACGA<br>AUAAAAAAGGCGAAGAUU<br>GACCUAGUGUUAAUCAAAAU<br>CAUAAAAGUCUUAUCGAUU<br>CAGAGAAUCCCAAGGUCG | CUUUCAAACUA | -- | CAGCACCGACUCUCAUAAAA<br>GAGAGAGACUGCUG                         | GAACCUUGAAAA | CUGCAUA | -- | Ile        |
| <i>Lusitaniella coriacea</i><br>LEGE 07157      | AGGGAGACCUAACUGAACGA<br>AUAAAAAAGGCGAAGAUUG<br>ACCUAGUGUUAAUCAAAAU<br>AUAAAAGUCUUAUCGAUU<br>GAGAAUCCCAAGGUCG    | CUUUAACUA   | -- | CAGCACCGACUCUCAUAAAA<br>GAGAGAGACUGCUG                         | GAACCUUGAAAA | CUGCAUA | -- | Ile        |
| <i>Lusitaniella coriacea</i><br>LEGE 07167      | AGGGAGACCUAACUGAACGA<br>AUAAAAAAGGCGAAGAUUG<br>ACCUAGUGUUAAUCAAAAU<br>AUAAAAGUCUUAUCGAUU<br>GAGAAUCCCAAGGUCG    | CUUUAACUA   | -- | CAGCACCGACUCUCAUAAAA<br>GAGAGAGACUGCUG                         | GAACCUUGAAAA | CUGCAUA | -- | Ile        |
| <i>Hyella patteloides</i><br>LEGE 07179         | AGGGAGACCAACCCAUUAG<br>ACUCCGAAACAUUUUGGAU<br>AAGAGCUAAUGAGGUCAUCC<br>UAGGUCG                                   | CUUUAACUA   | -- | UAGCACCGUUGCCUCACGAG<br>CGACUGCUG                              | GAACCUUGAAAA | CUGCAUA | -- | Ile        |
| <i>Geminobacterium atlanticum</i><br>LEGE 07459 | AGGGAGACCUACUUCUGAUA<br>UACCGACCAUUUUUAUUG<br>GUAUUCAGUAAGUCAUCCCA<br>AGGUCG                                    | CUUUAAGCUA  | -- | CAGCACCUAGUUUCACUAUUG<br>CUUUAUAGGUAAGGGAUAGG<br>GAGGGAGGUCGUG | GAACCUUGAAAA | CUGCAUA | -- | Ala<br>Ile |



**Figure 2.11.** Absorption spectra of cell-free extracts of the seven cyanobacterial strains described in this study. (A) *Geminobacterium atlanticum* LEGE 07459; (B) *Hyella patelloides* LEGE 07179; (C, D, E) *Lusitaniella coriacea* strains LEGE 07167, LEGE 07157, and LEGE 06111, respectively; (F) *Jaaginema litorale* LEGE 07176, and (G) *Paraleptolyngbya singularis* LEGE 06188. PE - phycoerythrin; PC - phycocyanin.



**Figure 2.12.** Maximum likelihood (ML) phylogenetic tree based on the 16S-ITS-23S rRNA gene sequences. The studied strains are indicated in bold and by the respective GenBank accession number (brackets). *Chloroflexus aurantiacus* strain J-10-fl was used as outgroup. Numbers along branches indicate bootstrap values considering 500 pseudo-replicates: only those equal or above 50% are indicated.



**Figure 2.13.** Maximum likelihood (ML) phylogenetic tree based on partial 16S rRNA gene sequences including all *Jaaginema* sequences available, the sequence of our LEGE strain classified here as *Jaaginema litorale*, and other cyanobacteria (including the other strains described in this study). The *Jaaginema* strains are indicated in bold. *Chloroflexus aurantiacus* strain J-10-fl was used as outgroup. Numbers along branches indicate bootstrap values considering 500 pseudo-replicates: only those equal or above 50% are indicated.

## 2.2.8 References

- AL-THUKAIR, A. A. 2011. Calculating boring rate of endolithic cyanobacteria *Hyella immanis* under laboratory conditions. *International Biodeterioration & Biodegradation*, 65, 664-667.
- AL-THUKAIR, A. A. & GOLUBIC, S. 1991. New endolithic cyanobacteria from the Arabian Gulf. I. *Hyella immanis* sp. nov.. *Journal of Phycology*, 27, 766-780.
- ANAGNOSTIDIS, K. 1989. *Geitlerinema*, a new genus of oscillatoriacean cyanophytes. *Plant Systematics and Evolution*, 164, 33-46.
- ARAÚJO, R., BÁRBARA, I., TIBALDO, M., BERECIBAR, E., TAPIA, P. D., PEREIRA, R., SANTOS, R. & PINTO, I. S. 2009. Checklist of benthic marine algae and cyanobacteria of northern Portugal. *Botanica Marina*, 52.
- BENNETT, A., BOGORAD, L. 1973. Complementary chromatic adaptation in a filamentous blue-green alga. *The Journal of Cell Biology*, 58, 419-435.
- BHAYA, D., BIANCO, N. R., BRYANT, D. & GROSSMAN, A. 2000. Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC 6803. *Molecular Microbiology*, 37, 941-951.
- BILLER, S. J., SCHUBOTZ, F., ROGGENSACK, S. E., THOMPSON, A. W., SUMMONS, R. E. & CHISHOLM, S. W. 2014. Bacterial vesicles in marine ecosystems. *Science*, 343, 183-186.

- BORNET, E., FLAHAULT, C. 1888. Note sur deux nouveaux genres d'algues perforantes. *Journal de Botanique (Morot)*, 2, 161-165.
- BRAVAKOS, P., KOTOULAS, G., SKARAKI, K., PANTAZIDOU, A. & ECONOMOU-AMILLI, A. 2016. A polyphasic taxonomic approach in isolated strains of Cyanobacteria from thermal springs of Greece. *Molecular Phylogenetics and Evolution*, 98, 147-160.
- BRITO, Â., GAIFEM, J., RAMOS, V., GLUKHOV, E., DORRESTEIN, P. C., GERWICK, W. H., VASCONCELOS, V. M., MENDES, M. V. & TAMAGNINI, P. 2015. Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*, 9, 218-226.
- BRITO, Â., RAMOS, V., SEABRA, R., SANTOS, A., SANTOS, C. L., LOPO, M., FERREIRA, S., MARTINS, A., MOTA, R., FRAZÃO, B., MARTINS, R., VASCONCELOS, V. & TAMAGNINI, P. 2012. Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Systematic and Applied Microbiology*, 35, 110-119.
- CASAMATTA, D. A., VIS, M. L. & SHEATH, R. G. 2003. Cryptic species in cyanobacterial systematics: a case study of *Phormidium retzii* (Oscillatoriales) using RAPD molecular markers and 16S rDNA sequence data. *Aquatic Botany*, 77, 295-309.
- CASTENHOLZ, R. W., WILMOTTE, A., HERDMAN, M., RIPPKA, R., WATERBURY, J. B., ITEMAN, I. & HOFFMANN, L. 2001. Phylum BX. Cyanobacteria, In *Bergey's Manual of Systematic Bacteriology, Volume One: The Archaea and the Deeply Branching and Phototrophic Bacteria*, D. R. Boone, R. W. Castenholz, G. M. Garrity (eds.), Springer, pp. 473-599.
- DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, 772-772.
- DE FIGUEIREDO, D. R., REBOLEIRA, A. S. S. P., ANTUNES, S. C., ABRANTES, N., AZEITEIRO, U., GONÇALVES, F. & PEREIRA, M. J. 2006. The effect of environmental parameters and cyanobacterial blooms on phytoplankton dynamics of a Portuguese temperate Lake. *Hydrobiologia*, 568, 145-157.
- DUBOURG, G., LAGIER, J. C., ARMOUGOM, F., ROBERT, C., HAMAD, I., BROUQUI, P. & RAOULT, D. 2013. The proof of concept that culturomics can be superior to metagenomics to study atypical stool samples. *European Journal of Clinical Microbiology & Infectious Diseases*, 32, 1099-1099.
- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-1797.
- EVERROAD, R. C., STUART, R. K., BEBOUT, B. M., DETWEILER, A. M., LEE, J. Z., WOEBKEN, D., PRUFERT-BEBOUT, L. & PETT-RIDGE, J. 2016. Permanent draft genome of strain ESFC-1: ecological genomics of a newly discovered lineage of filamentous diazotrophic cyanobacteria. *Standards in Genomic Sciences*, 11.
- FURTADO, A. L. F. F., CALIJURI, M. D. C., LORENZI, A. S., HONDA, R. Y., GENUÁRIO, D. B. & FIORE, M. F. 2009. Morphological and molecular characterization of cyanobacteria from a Brazilian facultative wastewater stabilization pond and evaluation of microcystin production. *Hydrobiologia*, 627, 195-209.
- GALHANO, V., DE FIGUEIREDO, D. R., ALVES, A., CORREIA, A., PEREIRA, M. J., GOMES-LARANJO, J. & PEIXOTO, F. 2011. Morphological, biochemical and molecular characterization of *Anabaena*, *Aphanizomenon* and *Nostoc* strains (Cyanobacteria, Nostocales) isolated from Portuguese freshwater habitats. *Hydrobiologia*, 663, 187-203.
- GUINDON, S., X000E, PHANE & GASCUEL, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Systematic Biology*, 52, 696-704.
- HASLER, P., DVORAK, P., JOHANSEN, J. R., KITNER, M., ONDREJ, V. & POULICKOVA, A. 2012. Morphological and molecular study of epipellic filamentous genera *Phormidium*, *Microcoleus* and *Geitlerinema* (Oscillatoriales, Cyanophyta/Cyanobacteria). *Fottea*, 341-356.
- HASLER, P., DVORAK, P., POULICKOVA, A. & CASAMATTA, D. A. 2014. A novel genus *Ammassolinea* gen. nov. (Cyanobacteria) isolated from sub-tropical epipellic habitats. *Fottea*, 241-248.
- ITEMAN, I., RIPPKA, R., TANDEAU DE MARSAC, N. & HERDMAN, M. 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. *Microbiology*, 146, 1275-1286.
- ITEMAN, I., TANDEAU DE MARSAC, N., HERDMAN, M. & RIPPKA, R. 2002. rDNA analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira*. *Microbiology*, 148, 481-496.
- JANSE, I., KARDINAAL, W. E. A., MEIMA, M., FASTNER, J., VISSER, P. M. & ZWART, G. 2004. Toxic and nontoxic *Microcystis* colonies in natural populations can be differentiated on the basis of rRNA gene internal transcribed spacer diversity. *Applied and Environmental Microbiology*, 70(7), 3979-3987.
- JOHANSEN, J. R., KOVACIK, L., CASAMATTA, D. A., IKOVÁ, K. F. & KAŠTOVSKÝ, J. 2011. Utility of 16S-23S ITS sequence and secondary structure for recognition of intrageneric and intergeneric

- limits within cyanobacterial taxa: *Leptolyngbya corticola* sp. nov. (Pseudanabaenaceae, Cyanobacteria). *Nova Hedwigia*, 92, 283-302.
- KIM, M., OH, H. S., PARK, S. C. & CHUN, J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 64, 346-351.
- KNOLL, A.H. 2008. Cyanobacteria and earth history. In *The Cyanobacteria: Molecular Biology, Genomics, and Evolution*, Herrero A., Flores E. (eds.), Caister Academic Press. pp. 1-20.
- KOLLING, G. L., & MATTHEWS, K. R. 1999. Export of virulence genes and Shiga toxin by membrane vesicles of *Escherichia coli* O157: H7. *Applied and Environmental Microbiology*, 65(5), 1843-1848.
- KOMÁREK, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, 51, 346-353.
- KOMÁREK, J., & ANAGNOSTIDIS, K. 1998. Cyanoprokaryota 1. Teil: Chroococcales. In *Süßwasserflora von Mitteleuropa 19/1*, Ettl, H., Gärtner, G., Heynig, H. & Mollenhauer, D. (eds.), Gustav Fischer Verlag. p. 548.
- KOMÁREK, J., KAŠTOVSKÝ, J., VENTURA, S., TURICCHIA, S. & ŠMARDÁ, J. 2009. The cyanobacterial genus *Phormidesmis*. *Algological Studies*, 129, 41-59.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- KORELUSOVÁ, J., KAŠTOVSKÝ, J. & KOMÁREK, J. 2009. Heterogeneity of the cyanobacterial genus *Synechocystis* and description of a new genus, *Geminocystis*. *Journal of Phycology*, 45, 928-937.
- KOTAI, J. (1972). Instructions for preparation of modified nutrient solution Z8 for algae. Norwegian Institute for Water Research, Oslo, 11, 5.
- KOTTEK, M., GRIESER, J., BECK, C., RUDOLF, B. & RUBEL, F. 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15, 259-263.
- KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870-1874.
- LAGIER, J.-C., HUGON, P., KHELALFIA, S., FOURNIER, P.-E., LA SCOLA, B. & RAOULT, D. 2015. The Rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews*, 28, 237-264.
- LIMA, F. P., QUEIROZ, N., RIBEIRO, P. A., HAWKINS, S. J. & SANTOS, A. M. 2006. Recent changes in the distribution of a marine gastropod, *Patella rustica* Linnaeus, 1758, and their relationship to unusual climatic events. *Journal of Biogeography*, 33, 812-822.
- LOPES, V. R., RAMOS, V., MARTINS, A., SOUSA, M., WELKER, M., ANTUNES, A. & VASCONCELOS, V. M. 2012. Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries. *Marine Environmental Research*, 73, 7-16.
- MASHBURN, L. M. & WHITELEY, M. 2005. Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature*, 437, 422-425.
- MEEKS, J. C. & CASTENHOLZ, R. W. 1971. Growth and photosynthesis in an extreme thermophile, *Synechococcus lividus* (Cyanophyta). *Archiv für Mikrobiologie*, 78, 25-41.
- OLIVEIRA, P., MARTINS, N., SANTOS, M., COUTO, N., WRIGHT, P. & TAMAGNINI, P. 2015. The *Anabaena* sp. PCC 7120 exoproteome: Taking a peek outside the box. *Life*, 5, 130-163.
- OMOREGIE, E. O., CRUMBLISS, L. L., BEBOUT, B. M. & ZEHR, J. P. 2004. Comparison of diazotroph community structure in *Lyngbya* sp. and *Microcoleus chthonoplastes* dominated microbial mats from Guerrero Negro, Baja, Mexico. *FEMS Microbiology Ecology*, 47, 305-308.
- OSORIO-SANTOS, K., PIETRASIAK, N., BOHUNICKÁ, M., MISCOE, L. H., KOVÁČIK, L., MARTIN, M. P. & JOHANSEN, J. R. 2014. Seven new species of *Oculatella* (Pseudanabaenales, Cyanobacteria): taxonomically recognizing cryptic diversification. *European Journal of Phycology*, 49, 450-470.
- PONTES, M. T., AGUIAR, R. & OLIVEIRA PIRES, H. 2005. A nearshore wave energy atlas for Portugal. *Journal of Offshore Mechanics and Arctic Engineering*, 127, 249.
- PRICE, M. N., DEHAL, P. S. & ARKIN, A. P. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26, 1641-1650.
- PRICE, M. N., DEHAL, P. S. & ARKIN, A. P. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5, e9490.
- RAMIREZ-REINAT, E. L. & GARCIA-PICHEL, F. 2012. Prevalence of Ca<sup>2+</sup>-ATPase-mediated carbonate dissolution among cyanobacterial euendoliths. *Applied and Environmental Microbiology*, 78, 7-13.
- RAMOS, V., SEABRA, R., BRITO, Á., SANTOS, A., SANTOS, C. L., LOPO, M., MORADAS-FERREIRA, P., VASCONCELOS, V. M. & TAMAGNINI, P. 2010. Characterization of an intertidal cyanobacterium that constitutes a separate clade together with thermophilic strains. *European Journal of Phycology*, 45, 394-403.
- RIPPKA, R. 1988. Isolation and purification of cyanobacteria. *Methods in Enzymology*, 167, 3-27.

- RIPPKA, R. & COHEN-BAZIRE, G. 1983. The cyanobacteriales: A legitimate order based on the type strain *Cyanobacterium stanieri*? *Annales de l'Institut Pasteur / Microbiologie*, 134, 21-36.
- RIPPKA, R., STANIER, R. Y., DERUELLES, J., HERDMAN, M. & WATERBURY, J. B. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology*, 111, 1-61.
- ROCAP, G., DISTEL, D. L., WATERBURY, J. B. & CHISHOLM, S. W. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16s-23s ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, 68, 1180-1191.
- RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HOHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, 539-542.
- ROSSI, F. & DE PHILIPPIS, R. 2015. Role of cyanobacterial exopolysaccharides in phototrophic biofilms and in complex microbial mats. *Life*, 5, 1218-1238.
- RUMBO, C., FERNANDEZ-MOREIRA, E., MERINO, M., POZA, M., MENDEZ, J. A., SOARES, N. C., MOSQUERA, A., CHAVES, F. & BOU, G. 2011. Horizontal transfer of the OXA-24 carbapenemase gene via outer membrane vesicles: a new mechanism of dissemination of carbapenem resistance genes in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 55, 3084-3090.
- SAMBROOK, J., RUSSELL, D.W. 2001. Molecular cloning: a laboratory manual. 3rd ed., Cold Spring Harbor Laboratory Press. p. 2110.
- SCIUTO, K. & MORO, I. 2016. Detection of the new cosmopolitan genus *Thermoleptolyngbya* (Cyanobacteria, Leptolyngbyaceae) using the 16S rRNA gene and 16S–23S ITS region. *Molecular Phylogenetics and Evolution*, 105, 15-35.
- SEABRA, R., SANTOS, A., PEREIRA, S., MORADAS-FERREIRA, P. & TAMAGNINI, P. 2009. Immunolocalization of the uptake hydrogenase in the marine cyanobacterium *Lyngbya majuscula* CCAP 1446/4 and two *Nostoc* strains. *FEMS Microbiology Letters*, 292, 57-62.
- SEABRA, R., WETHEY, D. S., SANTOS, A. M. & LIMA, F. P. 2011. Side matters: microhabitat influence on intertidal heat stress over a large geographical scale. *Journal of Experimental Marine Biology and Ecology*, 400, 200-208.
- SHERWOOD, A. R. & PRESTING G. G. 2007. Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. *Journal of Phycology*, 43(3), 605-608.
- SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539-539.
- STANIER, R., KUNISAWA, R., MANDEL, M., COHEN-BAZIRE, G. 1971. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriological Reviews*, 35, 171.
- STRUNECKÝ, O., ELSTER, J. & KOMÁREK, J. 2010. Phylogenetic relationships between geographically separate *Phormidium* cyanobacteria: is there a link between north and south Polar Regions? *Polar Biology*, 33, 1419-1428.
- STRUNECKÝ, O., KOMÁREK, J., ŠMARDA, J., 2014. *Kamptonema* (Microcoleaceae, Cyanobacteria), a new genus derived from the polyphyletic *Phormidium* on the basis of combined molecular and cytomorphological markers. *Preslia*, 86, 193-207.
- STUART, R. K., MAYALI, X., LEE, J. Z., CRAIG EVERROAD, R., HWANG, M., BEBOUT, B. M., WEBER, P. K., PETT-RIDGE, J. & THELEN, M. P. 2016. Cyanobacterial reuse of extracellular organic carbon in microbial mats. *The ISME Journal*, 10, 1240-1251.
- TAMAGNINI, P., TROSHINA, O., OXELFELT, F., SALEMA, R. & LINDBLAD, P. 1997. Hydrogenases in *Nostoc* sp. strain PCC 73102, a strain lacking a bidirectional enzyme. *Applied and Environmental Microbiology*, 63, 1801-1807.
- TATON, A., GRUBISIC, S., ERTZ, D., HODGSON, D. A., PICCARDI, R., BIONDI, N., TREDICI, M. R., MAININI, M., LOSI, D., MARINELLI, F. & WILMOTTE, A. 2006. Polyphasic study of antarctic cyanobacterial strains. *Journal of Phycology*, 42, 1257-1270.
- VALERIO, E., CHAMBEL, L., PAULINO, S., FARIA, N., PEREIRA, P. & TENREIRO, R. 2009. Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. *Microbiology*, 155, 642-656.
- VAZ, M. G. M. V., GENUARIO, D. B., ANDREOTE, A. P. D., MALONE, C. F. S., SANT'ANNA, C. L., BARBIERO, L. & FIORE, M. F. 2014. *Pantanalinema* gen. nov. and *Alkalinema* gen. nov.: novel pseudanabaenacean genera (Cyanobacteria) isolated from saline-alkaline lakes. *International Journal of Systematic and Evolutionary Microbiology*, 65, 298-308.
- WHITTON, B. A., POTTS, M. 2000. Introduction to the cyanobacteria. In *The ecology of cyanobacteria: their diversity in time and space*. Whitton, B. A. & Potts M. (eds.), Springer. pp. 1-11.

- WOEBKEN, D., BUROW, L. C., PRUFERT-BEBOUT, L., BEBOUT, B. M., HOEHLER, T. M., PETT-RIDGE, J., SPORMANN, A. M., WEBER, P. K. & SINGER, S. W. 2012. Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *The ISME Journal*, 6, 1427-1439.
- YARZA, P., YILMAZ, P., PRUESSE, E., GLÖCKNER, F. O., LUDWIG, W., SCHLEIFER, K.-H., WHITMAN, W. B., EUZÉBY, J., AMANN, R. & ROSSELLÓ-MÓRA, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, 12, 635-645.
- ZAMMIT, G., BILLI, D. & ALBERTANO, P. 2012. The subaerophytic cyanobacterium *Oculatella subterranea* (Oscillatoriales, Cyanophyceae) gen. et sp. nov.: a cytomorphological and molecular description. *European Journal of Phycology*, 47, 341-354.
- ZEHR, J. P. 2011. Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology*, 19, 162-173.

## 2.3 Cyanobacterial diversity in microbial mats from the hypersaline lagoon system of Araruama, Brazil: an in-depth polyphasic study

Vitor Ramos<sup>1,2,#</sup>, Raquel Castelo-Branco<sup>2,#</sup>, Pedro Nuno Leão<sup>1</sup>, Joana Martins<sup>1,2</sup>, Sinda Carvalhal-Gomes<sup>3</sup>, Frederico Sobrinho da Silva<sup>3</sup>, João Graciano Mendonça Filho<sup>3</sup>, Vitor Manuel Vasconcelos<sup>1,2\*</sup>

<sup>1</sup> Faculty of Sciences, University of Porto, Porto, Portugal

<sup>2</sup> Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Matosinhos, Portugal

<sup>3</sup> Palynofacies & Organic Facies Laboratory, Department of Geology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

\* Corresponding author. Email address: Email: vmvascon@fc.up.pt

# these authors contributed equally to this work

### 2.3.1 Abstract

Microbial mats are complex micro-scale ecosystems composed by different microorganisms, and can be found in a wide range of environments. In photosynthetic mats from hypersaline environments, cyanobacteria typically predominate in the top layer and play crucial ecological roles within the microbial consortium. Studying their diversity is thus relevant to provide new insights into the functioning and dynamics of these microbial communities. In this work, we have extensively characterized the cyanobacterial diversity present in three hypersaline microbial mat samples from the coastal lagoon system of Araruama, state of Rio de Janeiro, Brazil. For this purpose, we have scrutinized the samples by combining morphological and molecular approaches, as well as culture-dependent and independent methods, and discussing the different outcomes. We have also compared diversity indices and taxonomic assignments obtained by different classification methods. In addition, a PCR-based screening was performed to evaluate the potential production of secondary metabolites by *Geitlerinema*, *Leptolyngbya*, *Nodosilinea*, and *Synechococcus* strains isolated in this study. Thirty-six unique morphospecies could be differentiated in the three original samples. The results were then compared with previous taxonomic surveys, showing that this study increased by more than 15% the number of morphospecies and genera reported for the entire Araruama system. After quality filtering of the 16S rRNA gene sequences, molecular independent techniques allowed to obtain 38 DGGE band sequences and 105 cyanobacterial operational taxonomic units (OTUs) consensus sequences from >9000 reads, respectively. The phylogeny inferred with these molecular data was then compared with morphological data. Most of the 48 phylotypes classified from phylogenetic analysis could be associated with the morphospecies observed at the order level,

while a good number of sequences demonstrated to be closely affiliated with cyanobacteria from ecologically similar habitats. Other interesting finding from the phylogenetic analysis was that some sequences had no close relatives in the public databases, including one from an isolate, being placed as 'loner' sequences within different orders. This seems indicate the presence of novel cyanobacterial diversity in the mats of the Araruama system.

### 2.3.2 Introduction

Photosynthetic microbial mats are complex micro-scale ecosystems that can be found globally, in a wide range of environments, and are a major driving force in the formation of some modern microbialites, i.e. rock formation driven by microbial growth (Dupraz et al. 2009; Decho 2010; Stal 2012). Very often, mats are vertically structured, composed by distinct layers containing different pigments. Such colored layers are characterized by having phylogenetically different and metabolically complementary groups of prokaryotes whose composition can vary greatly depending on light quality and intensity, oxygen levels and nutrients sources (Ward et al. 2006; Harris et al. 2013). Photosynthetic mats usually have an upper green layer where cyanobacteria predominate and play a crucial role for the functioning of these consortia (Ley et al. 2006; Ward et al. 2006; Dupraz et al. 2009; Stal 2012). Typically, cyanobacteria constitute a large fraction of the biomass in the upper few millimeters (Ley et al. 2006; Harris et al. 2013) and are followed by aerobic or facultative heterotrophic bacteria, chemolithotrophic bacteria (among them colorless sulfur bacteria), anoxygenic phototrophs (purple and green bacteria) and sulfate-reducing bacteria (Visscher et al. 1992). In the top layer of mats, the most exposed to environmental changes and disturbances, cyanobacteria act as primary producers (Ley et al. 2006; Harris et al. 2013; Stal 2012) and atmospheric nitrogen fixers (Díez et al. 2007; Bauersachs et al. 2011; Stal 2012), while being responsible for the production of a matrix of extracellular polymeric substances (EPSs) that provide physical protection and resistance to desiccation for the microbial mat community (Dupraz et al. 2009; Franks & Stolz 2009; Stal 2012).

In saline aquatic systems, these photosynthetic mats can be observed in flat, undisturbed, sheltered marine or estuarine coasts, in salterns or salt evaporation ponds, or in hypersaline lagoons (Oren 2012; Stal 2012). Most of these ecosystems are considered harsh environments, with extreme environmental fluctuations are likely to occur on different timescales, and have been shown to be highly biodiverse (Ley et al. 2006; Harris et al. 2013). In hypersaline lagoon margins, where the photosynthetic mats develop, high salinity, seasonal desiccation and high light irradiance are the main environmental stressors influencing microbial mat community adaptation processes (Stal 2012). For instance, such factors determine that only halophilic or halotolerant cyanobacteria are able to inhabit and appropriately develop. Although there are some studies that list the taxa that occur at high salt concentrations (see Oren 2012), it is difficult

to obtain a clear picture of the cyanobacterial diversity occurring in hypersaline environments due to the unsatisfactory state of the taxonomy and nomenclature of the cyanobacteria (Oren 2012; Dvořák et al. 2015). Traditional systems of classification of cyanobacteria were based mainly on morphological criteria (for a review, see Komárek et al. 2014), something that is now recognized as having a lack of resolution at the species level, while completely ignoring cryptic species (Dvořák et al. 2015). Indeed, there are numerous morphologically indistinguishable cyanobacteria, the so-called cryptic species, which do not share a common evolutionary history (Komárek et al. 2014; Dvořák et al. 2015; Komárek 2016). Therefore, a polyphasic approach including molecular, morphological and ecophysiological traits is now mandatory for the taxonomy and identification of cyanobacteria (Komárek et al. 2014; Komárek 2016).

The east coastline region of the State of Rio de Janeiro in Brazil harbors a series of shallow coastal lagoons, forming one of the major hypersaline systems of the world (Clementino et al. 2008). The main waterbody is the Araruama lagoon, which gives the name to the system, and with an average of 5.2% total salts its water has a remarkable high salt content (Clementino et al. 2008). This region has a typical tropical climate with wet and dry seasons, where low levels of annual rainfall and high evaporation rates favored the development of several salty ponds around the lagoons (Kjerfve et al. 1996; Clementino et al. 2008). The cyanobacterial species present in microbial mats and/or water samples from several lagoons of the Araruama system have been extensively studied through culture-independent, morphological-based identifications, resulting in several species inventory lists (e.g. Iespa & Silva 2005; Silva et al. 2006, 2007a, 2007b, 2011). By contrast, a single molecular-based study of the total microbial diversity has been performed by Clementino et al. (2008), using water samples from the Araruama lagoon. In this work, the authors have detected three different cyanobacterial phylotypes.

With the aim of strengthening the knowledge on cyanobacteria living in hypersaline environments from tropical regions and (1) by considering the existing morphological-based species inventories, which show that the Araruama's lagoons harbor a high diversity of cyanobacteria, (2) by taking into account the findings from Mobberley et al. (2012) and Harris et al. (2013), who demonstrate the power of 454 sequencing technology for the study of the microbial diversity in very complex samples as are photosynthetic hypersaline mats, (3) by realizing that an accurate identification may be hampered by low resolution of classification methods (Dvořák et al. 2015; Nguyen et al. 2016), and (4) by following the more recent principles and recommendations for studying cyanobacterial taxa (Dvořák et al. 2015; Komárek 2016), we have exhaustively and comprehensively characterized the cyanobacterial diversity present in three mats from three lagoons of the Araruama system. For this purpose, we have combined morphological and molecular approaches, as well as culture-dependent and independent methods. In order to understand how different definitions of "units of diversity" shape the cyanobacterial community structure (composition, richness and diversity) we have also

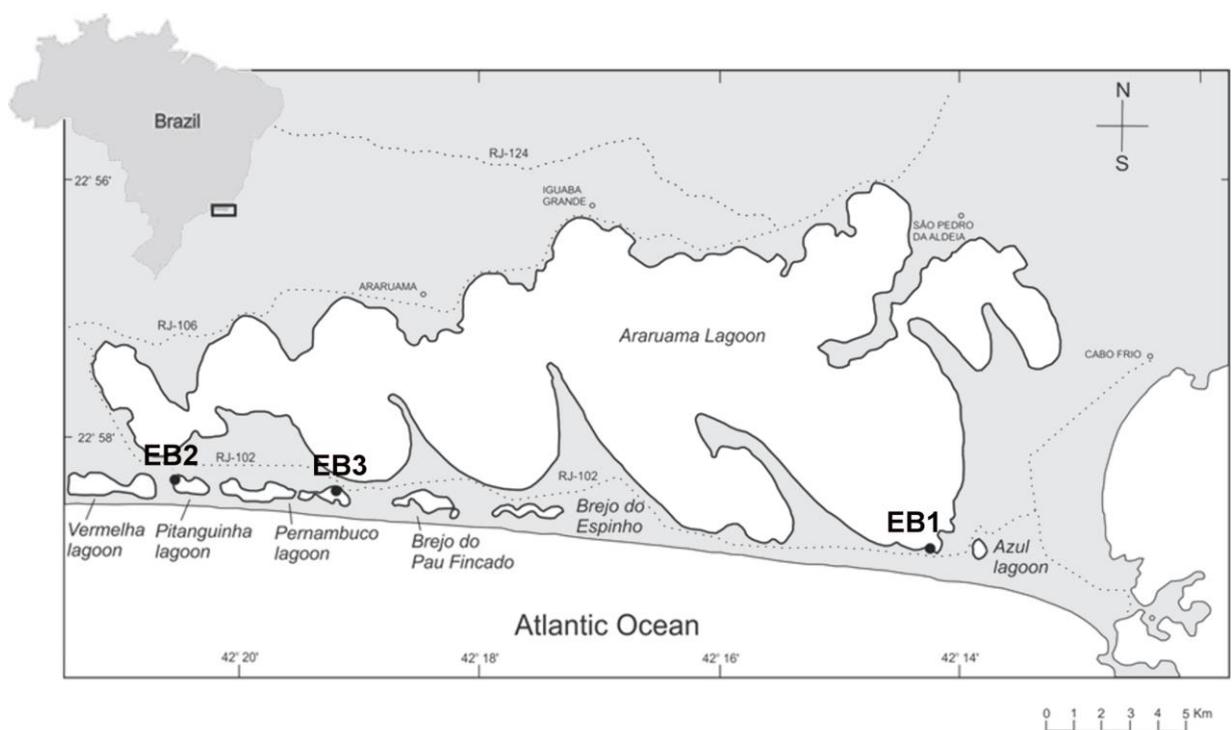
compared diversity indices and taxonomic assignments obtained by different classification methods. In addition, a PCR-based screening was performed to evaluate the potential production of secondary metabolites by the cyanobacterial strains isolated in this study.

### 2.3.3 Materials and Methods

#### Sampling sites and methodological approach

##### Sites

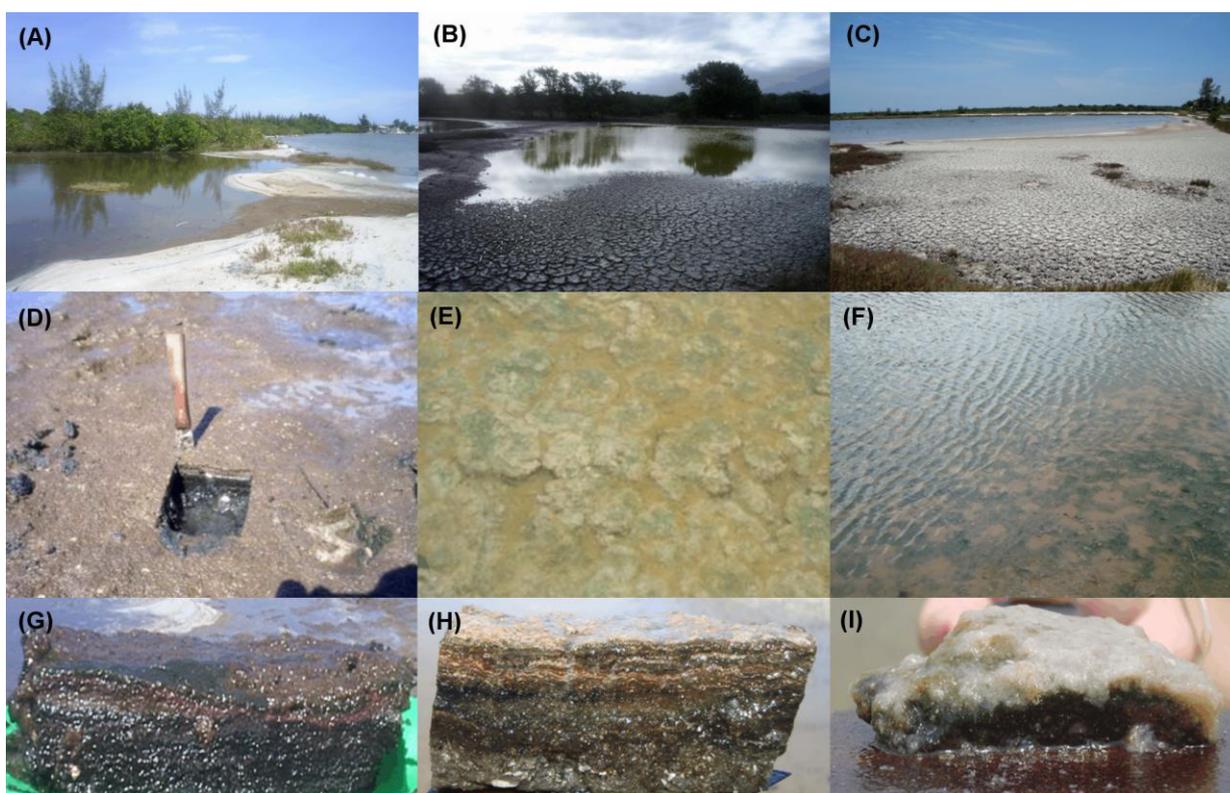
Samples were collected from three lagoons of the Araruama complex (Fig. 2.14): Araruama (the main lagoon; 22°56'36.0"S 42°06'02.0"W), Pitanguinha (22°55'39.0"S 42°21'20.0"W) and Pernambuco (22°55'50.0"S 42°18'86.0"W). Sampling sites (EB1, EB2 and EB3, respectively) were selected based on the occurrence of cyanobacterial-dominated mats, as previously indicated in Damazio and Silva (2006), Iespa and Silva (2005), and Silva et al. (2005, 2006). EB1 is placed in the eastern part of Araruama, at the entry of a confined, temporary pond connected to the main lagoon by a small channel (Fig. 2.15). The site is near the Channel Itajuru in Cabo Frio (Fig. 2.14), which connects the lagoon Araruama with the Atlantic Ocean. EB2 is surrounded by typical restinga vegetation (Fig. 2.15) and is located near to a salt pan. EB3 is located in an artificial pond surrounded by grass-like vegetation, adjacent to the lagoon, and is also near a salt pan.



**Figure 2.14.** Location map of the Araruama’s complex and sampling sites. EB1 is located in the main lagoon (Araruama), EB2 in lagoon Pitanguinha, and EB3 in lagoon Pernambuco.

### *Field sampling and samples processing*

Mat samples from each site were collected from an area of 1 m<sup>2</sup>, in February during the rainy season (Fig. 2.15). Physicochemical parameters of the water above or near the sampled mats were determined and are presented in Table 2.7. The shape of the mats was recorded during sampling, while their structural characteristics were examined at the laboratory. For this purpose, subsamples consisting of mat sections of about 10 cm<sup>2</sup> (Fig. 2.15) were collected, stored into polypropylene bags and transported to the lab. Mats were then characterized by color, cohesion and carbonate lamination under a light stereoscopic microscope. Subsamples used for isolation and morphological and molecular characterizations of cyanobacteria present in the mats were separated just after sample collection. In each sampling site, subsamples of 2 cm<sup>3</sup> were randomly collected from the top layers of the mats and distributed into sterile tubes.



**Figure 2.15.** Sampling sites and sample mats. First column refers to EB1 (Araruama), the second to EB2 (Pitanguinha), and the third to EB3 (Pernambuco). (A-C) Images of the sampling sites during the dry season; (D-F) Cyanobacterial mats at the moment of sampling. In (D) is shown the local where the 1x1 m sample was collected, at site EB1. (G-H) Cross section view of the samples, showing the multi-layered profile of the mats.

All subsamples were then preserved in the dark at 4 °C until processed for further analysis. At the laboratory, subsamples were processed aseptically and carefully restricted to their top photosynthetic layer, using sterile scalpel blades. All subsamples were screened for the presence of cyanobacteria by observing a piece of the mat under a light microscope (Leica DMLB, Bensheim, Germany).

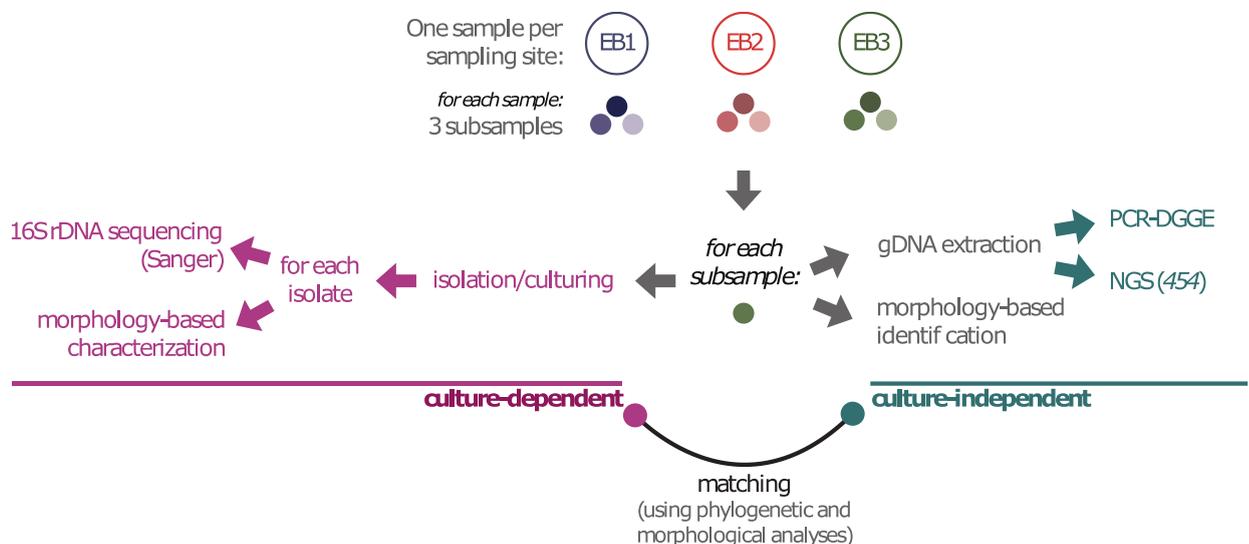
**Table 2.7** Physicochemical parameters\* of the three studied sites located in the Araruama lagoon complex.

| Sampling site | Lagoon      | pH           | Conductivity (μS/cm) | Salinity (psu) | Water temperature (°C) | Air temperature (°C) |
|---------------|-------------|--------------|----------------------|----------------|------------------------|----------------------|
| EB1           | Araruama    | 8,3<br>(8,5) | 24,2                 | (35,3)         | 24 (30,1)              | 27 (27,5)            |
| EB2           | Pitanguinha | 8,4<br>(8,6) | 9,2                  | (56,9)         | 34 (32,1)              | 31 (28,9)            |
| EB3           | Pernambuco  | 7,7<br>(8,8) | 11,2                 | (59,8)         | 28 (31,5)              | 27 (28,8)            |

\* at the time of sampling. Within brackets are average values for the 13 months prior to sampling.

### Study design

A workflow diagram illustrating the experimental procedures used in this polyphasic study is shown in Fig. 2.16. For each mat sample, three subsamples were independently used in each methodological approach. For instance, three independent slide preparations were observed for the microscope-based characterization of each environmental sample. The same applies for the isolation of cyanobacteria and of environmental DNA.



**Figure 2.16.** Schematic overview of the experimental design. See text for methodological details.

### Isolation, culturing and morphological-based characterization of cyanobacteria

For the isolation of cyanobacteria, subsamples were subjected to liquid culture enrichment, streaking in agar plates or micromanipulation (Rippka 1988; Waterbury 2006; see also Brito et al. 2012), or to a combination thereof, using different cultures media and salinities. Whenever feasible (i.e. for dominant species) single cells, colonies or filaments were isolated under the microscope with the help of a stretched Pasteur pipette, and transferred directly from raw biological material to different liquid or solid media (Ramos et al. 2010). When growth was evident, aliquots from the enriched cultures or agar plates were transferred and streaked again into fresh agar plates, or isolated by micromanipulation. The process was repeated until clonal

cyanobacterial isolates were obtained. The isolates were then transferred and grown in the correspondent liquid medium. The different media used during isolation were MN, BG11<sub>0</sub>, and Z8, at NaCl concentrations of 25‰, 40‰ or 55‰ (Rippka 1988; Waterbury 2006) and were supplemented with B12 vitamin and cycloheximide (Rippka 1988). During the isolation process, cultures were kept under a light/dark regime of 14:10h, irradiance of 10–30  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ , and temperature of 25 °C.

Isolates are deposited at the Blue Biotechnology and Ecotoxicology Culture Collection (acronym LEGE), at CIIMAR, Matosinhos, Portugal.

Microphotographs of environmental samples and isolates (either bright field or fluorescence) were obtained using a microscope (Model BX41, Olympus, Hamburg, Germany) coupled to an image analysis system (Model DP72 microscope digital camera, Olympus). Filament and/or cell dimensions were measured using the software Cell B (Olympus), with the same equipment. Dominant or abundant species (qualitative measure) present in each mat sample were recorded.

### **Survey of cyanobacterial taxa from inventory lists**

A primary literature search was performed to assess the cyanobacterial species richness in the Araruama's complex. A list of taxa was compiled along with the findings from this study (Supplementary Table 2.13).

### **Molecular-based methods**

#### *DNA isolation and 16S rRNA gene amplification*

Microbial mat subsamples were homogenized using sterile mortars and pestles. Approximately 400 mg (wet weight) of material was used for each DNA extraction. Total environmental DNA (eDNA) was extracted from samples using the Zymo Research Soil Microbe DNA kit (Zymo Research Corp, Irvine, CA, USA), according to the manufacturer's instructions. With respect to isolates, genomic DNA (gDNA) was extracted from fresh biomass samples, harvested from log-phase cultures, using the commercial kit PureLink Genomic DNA Mini Kit (Invitrogen, Carlsbad, USA). DNA integrity was checked by agarose gel electrophoresis with ethidium bromide staining and the triplicates pooled together.

In the case of gDNA from isolates, PCRs were performed using the conditions and the primer sets previously described in Brito et al. (2012). Regarding eDNA samples, a fragment of 422 bp length was amplified using the cyanobacteria-specific primer pair CYA-359F/CYA-781R (Nübel et al. 1997). In PCRs for denaturing gradient gel electrophoresis (DGGE) analysis, the forward primer (CYA-359F-GC) had a 40-nucleotide GC-rich sequence (GC clamp) attached to its 5'-end. The PCR reactions for DGGE were prepared in a volume of 20  $\mu\text{l}$  containing 1 $\times$  Reaction Buffer, 2.5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  of each deoxynucleotide triphosphate, 20.0 pmol of each primer, 0.5 U of GoTaq® Flexi DNA Polymerase (Promega, Madison, WI, USA), 20 mg  $\text{ml}^{-1}$  of bovine

serum albumin (BSA) and 5-10 ng of DNA template. Thermal cycling was carried out in a T-Professional Standard thermocycler (Biometra, Goettingen, Germany) under the following conditions: initial denaturation at 94°C for 2 min, followed by 11 cycles at 94 °C for 1 min, 65 °C for 1 min, and 72 °C for 1 min. This first step was followed by 32 cycles at 94 °C for 1 min, 55 °C for 1 min and 72 °C for 4 min and a final extension step at 72 °C for 4 min. PCR products were separated by 1.5% (w/v) agarose gel in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). Gels were stained with ethidium bromide and photographed under UV transillumination.

#### *Denaturing gradient gel electrophoresis, cloning and sequencing*

After gel visualization, PCR products from the same mat sample were pooled together. Twenty microliters of each pooled sample were loaded onto 6% polyacrylamide 1 mm gels, using a 40–60% denaturing gradient (100% denaturing conditions correspond to 7 M urea and 40% formamide). The electrophoresis was performed using a DCode system (Bio-Rad, CA, USA) at 60 V for 16 h, in 1× TAE buffer. The gel was stained with 1× SYBR Gold nucleic acid stain (Invitrogen, San Diego, CA). Small pieces of visible DGGE bands (Supplementary Fig. 2.21) were punched from the gel with sterile pipette tips. Each piece was then transferred into PCR tubes containing 30 µl of sterile water and incubated at 37 °C for 30 minutes to allow diffusion of the DNA. Two microliters of the eluted DNA were used as template for the re-amplification of the 16S rRNA gene, as described above. In this case, CYA-359F (i.e. without the GC clamp) was the forward primer used, as described by Nübel and colleagues (1997). PCR products were then extracted from the agarose gel and purified by using the spin columns Cut & Spin Gel Extraction (GRiSP, Porto, Portugal). Purified PCR products from each DGGE band were cloned using a pGEM® - T Easy Vector System Kit (Promega, Madison, WI, USA), and transformed into *Escherichia coli* ONE SHOT® TOP10 chemically competent cells (Invitrogen, San Diego, CA), following the instructions of the manufacturers. Colonies were selected by blue-white screening, and the presence of the appropriate insert was evaluated by colony PCR, using the primers pUCF/pUCR. Colonies with the insert were grown overnight at 37 °C, in liquid LB medium supplemented with 100 µg ml<sup>-1</sup> of ampicillin, with shaking at 200 rpm and plasmidic DNA was isolated from the overnight cultures using the GenElute Plasmid Miniprep Kit (Sigma, USA).

#### *Sanger Sequencing*

Purified plasmids and PCR products obtained from isolates (purified with the same spin columns mentioned above) were sent to sequencing at MacroGen (Amsterdam, Netherlands). All sequences were checked for chimera formation using the software DECIPHER (Wright et al. 2012).

### *High-throughput amplicons sequencing*

PCR amplifications from eDNA were obtained using the same primers used for PCR-DGGE, but without the GC clamp in the forward primer. They were originally designed (Nübel et al. 1997) to target the V3-V4 region of the 16S rRNA gene for cyanobacteria (including plastids). The amplification of PCR products was carried out using a barcode-tagged PCR primers approach (bcPCR), following the same conditions, adaptors and reagents as described in Pinto et al. (2014). Pre-sequencing processing such as amplicon library generation, barcoding and emulsification are described elsewhere (Pinto et al. 2014). Massive parallel sequencing was performed using the Genome Sequencer FLX System Instrument (454 Life Sciences, Roche) at Biocant, Portugal. Raw sequence reads were then analyzed and processed using an in-house, automatic pipeline from Biocant, Portugal. Processing steps performed include sorting of sequences by sample, dereplication, filtering of low-quality sequences, detection and removing of DNA chimeras, Operational Taxonomic Units (OTUs) clustering (sequence similarity cut-off value of 97%) and generation of OTUs consensus sequences, as described in Pinto et al. (2014).

After removing primers and barcode tags, the raw reads obtained were additionally analyzed in the SILVAngs pipeline (Quast et al. 2013). Steps of quality control included the exclusion from further processing of putative contaminations and artefacts, of reads < 100 aligned nucleotides or with a low alignment quality, and of reads with more than 2% of ambiguities, or 2% of homopolymers. Reads were then dereplicated and the unique reads were clustered into OTUs. The reference read of each OTU (i.e., the longest read in each cluster) was classified by a local BLASTn search against the non-redundant version of the SILVA SSU Ref dataset (release 123; <http://www.arb-silva.de>) with standard settings (Camacho et al. 2009). In this case, the sequence similarity value for defining distinct OTUs was 98%.

### *Nucleotide sequence accession numbers*

Novel PCR-based sequences associated with this study are available in GenBank under the accession numbers KT730170-KT730215. Sequence reads obtained in this study were deposited in NCBI's Sequence Read Archive (SRA) with the project number PRJNA294527 (SRA identifier: SRP063335); for corresponding accession numbers and further details on sequences see Appendix A4.

### *Phylogenetic analysis*

The cyanobacterial 16S rRNA gene sequences from isolates (9), DGGE bands (38) and consensus sequences of pyrosequencing derived "97% cutoff" OTUs (105) were joined together in a single phylogenetic analysis (Table 2.8 and Appendix A4). Singletons (i.e, a read with a sequence that is present exactly once) and OTUs consensus sequences with less than 300

nucleotides length were removed from phylogenetic and downstream analyses. A second round of identification and removal of chimeras was performed for pyrosequencing sequences using DECIPHER (Wright et al. 2012). In order to include the most similar sequences and to attain a reliable and robust backbone representation of the cyanobacterial diversity, the best BLAST hits for our sequences (and the closest known relative, if the best hit was an unidentified organism) were also included in the phylogeny (see Appendix A4), together with all the available sequences from Reference strains included in the Bergey's Manual of Systematic Bacteriology (Castenholz et al. 2001). The sequences from the rumen *Melainabacterium* strain YS2 and *Chloroflexus auranticus* J-10-fl were used as outgroup.

Multiple sequence alignment, evolutionary analyses and phylogenetic tree reconstructions were carried out using the software package MEGA6 (Tamura et al., 2013). Kimura 2-parameter was the model of nucleotide substitution used to infer the Maximum Likelihood (ML) tree (1000 replicates), as chosen by the corrected Akaike's Information Criterion (AICc). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3556)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 31.8460% sites). The final analysis involved 402 nucleotide sequences with a total of 345 positions in the dataset.

**Table 2.8** Number of cyanobacterial 16S rRNA gene sequences used in or discarded from phylogeny, by sample.

| Sequences from        | EB1 | EB2 | EB3 | Total        |
|-----------------------|-----|-----|-----|--------------|
| <b>Included</b>       |     |     |     | <b>(145)</b> |
| Isolates              | 4   | 3   | 2   | <b>9</b>     |
| DGGE bands            | 8   | 13  | 17  | <b>38</b>    |
| Pyrosequencing OTUs # | 64  | 24  | 10  | <b>98</b>    |
| <b>Not included</b>   |     |     |     | <b>(67)</b>  |
| Singlets              | 27  | 18  | 9   | <b>54</b>    |
| Sequences < 300 bp §  | 4   | 1   | 2   | <b>7</b>     |
| Chimeras *            | 0   | 2   | 4   | <b>6</b>     |

# Sequence similarity threshold of 97%; generated by the Biocant pipeline

§ But included in further analyses (e.g. see Table 2.12)

\* Detected in a second round of screening, using DECIPHER

### *PCR screening for secondary metabolites*

An additional PCR-based screening was performed to evaluate the potential production of toxins and bioactive compounds by the cyanobacterial strains isolated in this study. Primer sets used were: (1) HEPF/HEPR (Jungblut & Neilan 2006), that targets the aminotransferase region of *mcyE* and *ndaF*, two genes that are involved in the biosynthesis of the hepatotoxins microcystin and nodularin, respectively; (2) *sxtI682F/sxtI877R* (Lopes et al., 2012), targeting a gene involved in the biosynthesis of saxitoxin, *sxtI*; (3) *anaC-genF/anaC-genR* (Rantala-Ylinen et al. 2011), which targets *anaC*, a gene from the anatoxin-a synthetase gene cluster; and (4) the degenerate

primers DKF/DKR (Moffitt & Neilan 2001), to generically amplify polyketide synthase (PKS) genes. PCR conditions used were the same as originally described. Genomic DNA (gDNA) from the microcystin producer *Microcystis aeruginosa* LEGE 00063 was used as positive control in the PCRs with the primer sets HEPF/HEPR and DFK/DKR. Other gDNA used as positive controls were from *Anabaena* sp. LEGE X-002 (anatoxin-a producer) and *Cuspidothrix issatschenkoi* LMECYA 031 (saxitoxin producer).

### **Taxonomic assignments and cyanobacterial diversity comparison**

Cyanobacteria were identified based on morphology (hereafter called morphospecies) following standard taxonomic references (Komárek & Anagnostidis 1998, 2005). Nomenclature of all taxa was then brought to its latest system of classification (Komárek et al. 2014; Guiry & Guiry 2016). At the end, this list of morphospecies was compared with the list of morphospecies acquired from the survey (Supplementary Table 2.13).

Cyanobacterial 16S rRNA gene sequences obtained in this study were classified by means of different automatic, hierarchical taxonomies such as Greengenes v13.8 (McDonald et al. 2012), RDP II classifier v11.4 (Wang et al. 2007), NCBI Taxonomy (Federhen 2012), and SILVA Taxonomy v123 (Quast et al. 2013) using standard settings. Furthermore, using a phylogeny-guided clustering approach as recommended by Nguyen et al. (2016), we have manually curated and categorized the sequences into phylotypes according to their phylogenetic placement and bootstrap support of clades (Appendix A2). Thus, for purposes of this study, a phylotype should be seen as a taxon *sensu lato*, which may embrace diversity corresponding to more than one traditional, taxonomic rank.

The number of taxa (“species” richness, S) determined by morphological- (i.e., morphospecies) and DNA-based approaches (i.e., phylotypes and OTUs defined by a 97% or a 98% identity threshold) were compared between samples and among methods. Furthermore, using the number of reads encompassed in each OTU (97%) or phylotype, i.e. their relative abundance (Supplementary Table 2.14 and Appendix A4), other diversity indices were calculated (according to Morris et al. 2014) and compared between samples: the Shannon's diversity ( $H'$ ), Simpson's diversity ( $1/D$ ), and Shannon's evenness ( $E_H$ ) indices.

### **2.3.4 Results**

#### **Characterization of microbial mats**

Morphologically, the cyanobacterial mats found at the sampling sites belonged to the smooth (EB1) or polygonal (EB2 and EB3) types, while structurally they were layered (Fig. 2.15 and Table 2.13). At the time of sampling, smooth mats from EB1 presented a carpet-like form, covering a large area of the pond bottom. The sampled mat had a thin green layer on top (about 3.5 mm), followed by a purple (5.5 mm) and dark (5.48 mm) layers (Fig. 2.15, G). The mats

found in EB2 and EB3 were characterized by their large polygonal plates. Sampled mats (Fig. 2.15, H and I) showed a thin yellow-greenish layer on top (2.4 and 2.7 mm in EB2 and EB3, respectively), followed by a purple-brown layer (2.9 and 3.4 mm) and then a dark layer (24.6 and 27.9 mm). Some thin and discontinuous calcium carbonate layers were found.

### Morphological characterization of cyanobacterial diversity

Thirty-six species belonging to 22 genera were distinguished in the culture-independent study of the three mat samples (Table 2.9, Figs. 2.17 and Supplementary Figs. 2.22, 2.23 and 2.24).

**Table 2.9** Species composition and morphological-based characterization of cyanobacteria observed in the upper layer of the microbial mats collected at Araruama (EB1), Pitanguinha (EB2) and Pernambuco (EB3) lagoons.

| Taxa<br>(Order/Morphospecies)   | Sample <sup>a</sup> |     |     | Cell size<br>(Ø or length × width; in µm) | Figures/panels<br>#            |
|---|---------------------|-----|-----|---|--------------------------------|
|   | EB1                 | EB2 | EB3 |   |                                |
| <b>Chroococcales</b>  |                     |     |     |   |                                |
| <i>Aphanothece</i> cf. <i>conglomerata</i> Rich   | ●                   |     |     | 1.3 ± 0.1 × 0.9 ± 0.1                     | Suppl. 2.22 (C)                |
| <i>Aphanothece</i> aff. <i>salina</i> Elenkin & Danilov   | ●                   |     |     | 6.0 ± 1.2 × 3.5 ± 1.0                     | Suppl. 2.22 (D)                |
| <i>Aphanothece</i> cf. <i>stagnina</i> (Sprengel) A.Braun   | ●                   | ○   | ○   | 9.5 ± 1.9 × 4.6 ± 0.6                     | 2.17 (H) and Suppl. 2.22 (E)   |
| <i>Chroococcus</i> aff. <i>turgidus</i> (Kützing) Nägeli  | ●                   |     |     | 22.9 ± 3.2 × 33.2 ± 5.1                   | Suppl. 2.22 (F)                |
| <i>Cyanosarcina</i> aff. <i>thalassia</i> Anagnostidis & Pantazidou                                   |                     | ○   |     | 1.3 ± 0.2                                 | Suppl. 2.22 (H)                |
| <i>Geminocystis</i> sp. Korelusová, Kastovský & Komárek   | ●                   | ●   | ●   | 4.1 ± 0.8                                 | 2.17 (I) and Suppl. 2.22 (I)   |
| <i>Gloeocapsopsis</i> cf. <i>crepidinum</i> (Thuret) Geitler ex Komárek                               | ●                   | ○   |     | 4.3 ± 1.3                                 | Suppl. 2.22 (J)                |
| <i>Gloeothece</i> cf. <i>subtilis</i> Skuja   | ●                   |     |     | 1.5 ± 0.1 × 0.8 ± 0.1                     | Suppl. 2.22 (K)                |
| <b>Oscillatoriales</b>  |                     |     |     |   |                                |
| <i>Coleofasciculus chthonoplastes</i> (Thuret ex Gomont) Siegesmund, Johansen & Friedl                |                     | □   | □   | 8.0 ± 2.0 × 6.5 ± 0.7                     | 2.17 (D) and Suppl. 2.23 (A)   |
| <i>Geitlerinema</i> aff. <i>amphibium</i> (Agardh ex Gomont) Anagnostidis                             |                     | ●   |     | 5.2 ± 0.7 × 2.8 ± 0.5                     | Suppl. 2.23 (B)                |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> (Woloszynska) Anagnostidis                                | ■                   | ●   | ■   | 5.2 ± 1.3 × 2.1 ± 0.3                     | 2.17 (B) and Suppl. 2.23 (C)   |
| <i>Microcoleus</i> aff. <i>steenstrupii</i> Petersen  | ■                   | ●   |     | 6.3 ± 1.3 × 3.9 ± 0.6                     | 2.17 (C) and Suppl. 2.23 (I)   |
| <i>Oscillatoria limosa</i> Agardh ex Gomont   |                     | ●   | ○   | 4.4 ± 0.7 × 9.7 ± 0.3                     | Suppl. 2.23 (L)                |
| <i>Oscillatoria margaritifera</i> Kützing ex Gomont   |                     | ●   |     | 6.2 ± 0.7 × 27.6 ± 1.9                    | Suppl. 2.23 (M)                |
| <i>Oscillatoria subbrevis</i> Schmidle  |                     | ●   |     | 2.5 ± 0.4 × 7.3 ± 0.3                     | Suppl. 2.23 (N)                |
| <i>Oxynema</i> cf. <i>lloydianum</i> (Gomont) Chatchawan, Komárek, Strunecky, Smarda & Peerapornpisal |                     | ●   | ■   | 3.4 ± 0.7 × 8.5 ± 0.8                     | 2.17 (F-G) and Suppl. 2.23 (O) |
| <i>Phormidium nigroviride</i> (Thwaites ex Gomont) Anagnostidis & Komárek                             |                     | ●   |     | 4.5 ± 0.8 × 14.3 ± 0.7                    | Suppl. 2.23 (P)                |
| <b>Spirulinales</b>   |                     |     |     |   |                                |

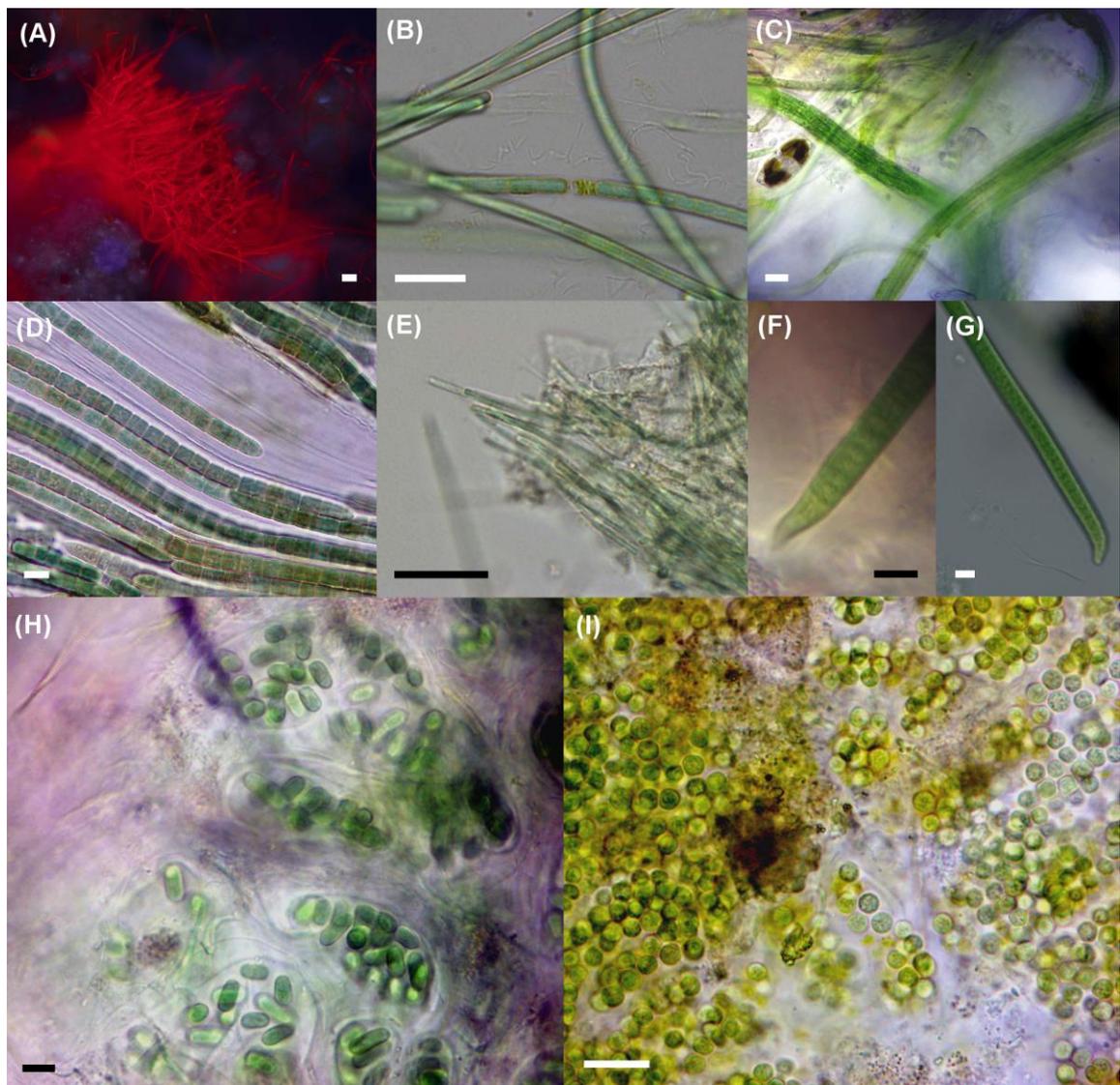
|  |       |   |                                 |
|--|-------|---|---------------------------------|
| <i>Spirulina labyrinthiformis</i> Gomont                                 | ●     | 0.7 ± 0.1 wide,<br>spirals width: 1.9 ± 0.5 | Suppl. 2.24 (B-C)               |
| <i>Spirulina subsalsa</i> Oerstedt ex Gomont                             | ○     | 1.3 ± 0.3 wide,<br>spirals width: 3.4 ± 0.3 | Suppl. 2.24 (D)                 |
| <i>Spirulina tenerrima</i> Kützing ex Gomont                             | ●     | 0.4 ± 0.1 wide,<br>spirals width: 1.4 ± 0.1 | Suppl. 2.24 (E)                 |
| <b>Synechococcales</b>   |       |   |                                 |
| <i>Aphanocapsa litoralis</i> (Hansgirg) Komárek & Anagnostidis           | ●     | 3.0 ± 0.5                                   | Suppl. 2.22 (A)                 |
| <i>Aphanocapsa</i> cf. <i>salina</i> Woronichin                          | ● ●   | 1.2 ± 0.2                                   | Suppl. 2.22 (B)                 |
| <i>Coelosphaeriopsis</i> cf. <i>halophila</i> Lemmermann                 | ●     | 2.8 ± 0.2                                   | Suppl. 2.22 (G)                 |
| <i>Halomicronema excentricum</i> Abed, Garcia-Pichel & Hernández-Mariné  | ● ■ ● | 2.8 ± 0.6 × 0.8 ± 0.1                       | 2.17 (E) and<br>Suppl. 2.23 (H) |
| <i>Komvophoron breve</i> (Carter) Anagnostidis                           | ●     | 1.1 ± 0.1 × 2.0 ± 0.2                       | Suppl. 2.23 (D)                 |
| <i>Komvophoron</i> cf. <i>minutum</i> (Skuja) Anagnostidis & Komárek     | ●     | 2.0 ± 0.3 × 2.2 ± 0.3                       | Suppl. 2.23 (E)                 |
| <i>Lemmermanniella</i> sp. Geitler in Engler & Prantl                    | ●     | 1.0 ± 0.1 × 0.6 ± 0.1                       | Suppl. 2.22 (L)                 |
| <i>Leptolyngbya crosbyana</i> (Tilden) Anagnostidis & Komárek            | ● ●   | 1.9 ± 0.6 × 1.3 ± 0.1                       | Suppl. 2.23 (F)                 |
| <i>Leptolyngbya</i> cf. <i>ectocarpi</i> (Gomont) Anagnostidis & Komárek | ●     | 3.0 ± 1.0 × 1.5 ± 0.4                       | Suppl. 2.23 (G)                 |
| <i>Nodosilinea</i> sp. Perkerson & Casamatta                             | ●     | 1.7 ± 0.5 × 2.5 ± 0.5                       | Suppl. 2.23 (J)                 |
| <i>Nodosilinea nodulosa</i> (Li & Brand) Perkerson & Casamatta           | ●     | 1.0 ± 0.2 × 1.1 ± 0.2                       | Suppl. 2.23 (K)                 |
| <i>Pseudanabaena</i> aff. <i>amphigranulata</i> (Goor) Anagnostidis      | ●     | 3.8 ± 0.9 × 1.3 ± 0.1                       | Suppl. 2.23 (Q)                 |
| <i>Pseudanabaena</i> cf. <i>limnetica</i> (Lemmermann) Komárek           | ●     | 3.5 ± 1.1 × 0.8 ± 0.3                       | Suppl. 2.23 (R)                 |
| <i>Pseudanabaena</i> cf. <i>minima</i> (G.S.An) Anagnostidis             | ●     | 1.5 ± 0.5 × 1.4 ± 0.2                       | Suppl. 2.24 (A)                 |
| <i>Synechococcus</i> sp. Nägeli  | ●     | 2.1 ± 0.6 × 0.8 ± 0.1                       | Suppl. 2.22 (M)                 |
| <i>Synechocystis salina</i> Wislouch                                     | ●     | 2.1 ± 0.2                                   | Suppl. 2.22 (N)                 |

<sup>a</sup> Underlined symbols refer to first reported observations for the entire lagoons complex. Black circles or squares denote first reported observations of a species for that water body. White circles or squares indicate that the species was also observed in previous studies (see also the full checklist available in Supplementary Table 2.13). Squares mean that the taxon was found to dominate or was abundant in samples from this study.

<sup>#</sup> Suppl. refer to figures in Supplementary Data section

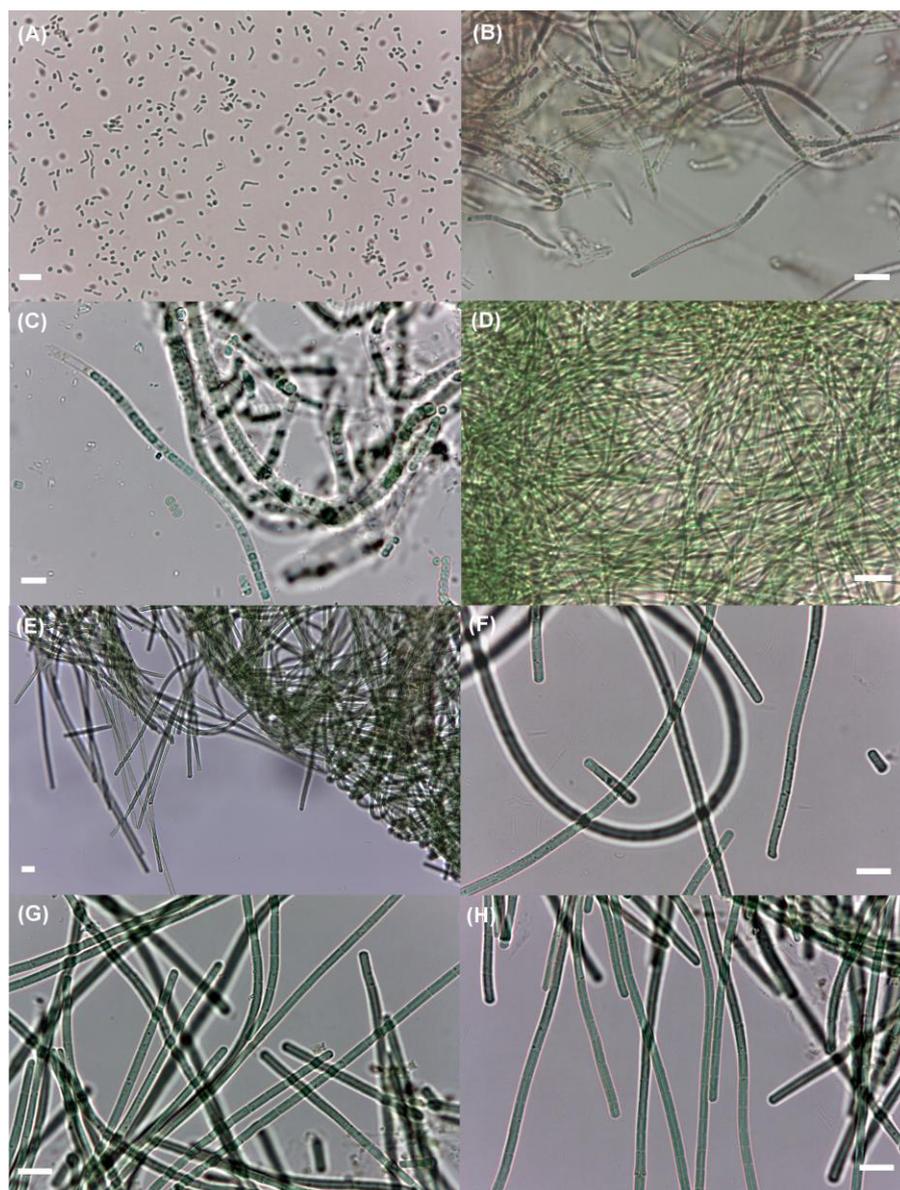
In the mat collected at EB1 7 species of Chroococcales, 2 of Oscillatoriales and 12 of Synechococcales (21 species in total) were observed. The mat from EB2 showed the presence of 4 Chroococcales, 9 Oscillatoriales and 5 Synechococcales species (18 in total). In the sample from EB3, in addition to species of Chroococcales (2), Oscillatoriales (4) and Synechococcales (3), species from the order Spirulinales (3) were also observed (12 in total). Filamentous heterocystous cyanobacteria from the order Nostocales, or colonial species from the orders Pleurocapsales or Chroococciopsidales were not observed in any of the samples. With three species detected, the most represented genera were *Aphanothece*, *Oscillatoria*, *Spirulina* (only observed in the mat sample collected at EB3) and *Pseudanabaena*.

Four species were common to all three samples (Table 2.9): the colonial, rod-shaped *Aphanothece* cf. *stagnina* (Fig. 2.17, H), the spherical, unicellular *Geminocystis* sp. (Fig. 2.17, I), the very thin, filamentous *Halomicronema excentricum* (Fig. 2.17, A and E), and the highly motile, filamentous *Geitlerinema* cf. *lemmermannii* (Fig. 2.17, B). This latter species dominated the mat sample from EB1 and was abundant at EB3. *Microcoleus* aff. *steenstrupii* (Fig. 2.17, C), with its trichomes densely packed in fascicles, was also abundant at EB1, being present at EB2. The wide sheathed, bundle-forming species *Coleofasciculus chthonoplastes* (Fig. 2.17, D) dominated the mats collected at EB2 and EB3, but was not observed in the mat from EB1. Other abundant taxa were *Leptolyngbya minuta*, in the mat from EB2, and the filamentous with bent ends *Oxynema* cf. *lloydianum* (Fig. 2.17, F-G) at EB3. *O. lloydianum* was also detected at EB2.



**Figure 2.17.** Epifluorescence (A) and bright field micrographs (B-I) showing ubiquitous, abundant, or dominant cyanobacteria in the environmental samples. (A) Tuft of filaments from *Halomicronema excentricum*, a thin cyanobacterium common to the three samples and abundant in the mat from EB2; (B) *Geitlerinema* cf. *lemmermannii*, present in the three samples and being dominant at EB1 and abundant at EB3; (C) *Microcoleus* aff. *steenstrupii*, abundant in EB1; (D) *Coleofasciculus chthonoplastes*, a dominant species in mats collected at EB2 and EB3; (E) *Halomicronema excentricum*; (F-G) *Oxynema* cf. *lloydianum* abundant at EB3; (H) *Aphanothece* cf. *stagnina* and (I) *Geminocystis* sp., both common to all three samples. Scale bar: 10 µm.

Although clearly dominated by cyanobacteria, other groups of microorganisms were present in the top layer of the mats, including some diatoms and Proteobacteria (Supplementary Fig. 2.24). The culture-dependent approach allowed the isolation and identification of nine cyanobacterial strains, belonging to five different taxa (Table 2.10 and Fig. 2.18), one from the order Oscillatoriales and the other four from three genera of Synechococcales. Five of them are strains from the Oscillatoriales species *Geitlerinema* cf. *lemmermannii*. Four isolates from two different species, *Leptolyngbya* aff. *ectocarpus* and *Geitlerinema* cf. *lemmermannii*, were obtained from the sample collected at EB1. Three strains from two species, *Nodosilinea* sp. and *Geitlerinema* cf. *lemmermannii*, were isolated from the mat sampled at EB2. Finally, one *Leptolyngbya* sp. and one *Synechococcus* sp. strains were obtained from the EB3 sample.



**Figure 2.18.** Cyanobacterial isolates obtained in this study. (A) The small, unicellular *Synechococcus* sp. LEGE 11394; (B) the brownish, filamentous *Leptolyngbya* aff. *ectocarpus* LEGE 11389; (C) the sheathed filamentous *Nodosilinea* sp. LEGE 11395; (D) the thin, filamentous *Leptolyngbya* sp. LEGE 11392; (E) *Geitlerinema* cf. *lemmermannii* LEGE 11390, at 400x magnification; (F-H) the same non-sheathed filamentous species as in (E), at 1000x magnification; strains LEGE 11393, 11391 and LEGE 11396, respectively. The short filaments in (F) are hormogonia. Scale bar: 10 µm.

**Table 2.10** List of cyanobacterial strains isolated from hypersaline microbial mats collected at the Araruama lagoon system (RJ, Brazil).

| Strain   | Order           | Sampling site | Figure/panel |
|--|-----------------|---------------|--------------|
| <i>Leptolyngbya</i> aff. <i>ectocarpi</i> LEGE 11389   | Synechococcales | EB1           | 2.18 (B)     |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> LEGE 11390 | Oscillatoriales | EB1           | 2.18 (E)     |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> LEGE 11391 | Oscillatoriales | EB1           | 2.18 (G)     |
| <i>Leptolyngbya</i> sp. LEGE 11392                     | Synechococcales | EB3           | 2.18 (D)     |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> LEGE 11393 | Oscillatoriales | EB2           | 2.18 (F)     |
| <i>Synechococcus</i> sp. LEGE 11394                    | Synechococcales | EB3           | 2.18 (A)     |
| <i>Nodosilinea</i> sp. LEGE 11395                      | Synechococcales | EB2           | 2.18 (C)     |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> LEGE 11396 | Oscillatoriales | EB1           | 2.18 (H)     |
| <i>Geitlerinema</i> cf. <i>lemmermannii</i> LEGE 11401 | Oscillatoriales | EB2           | #            |

# this strain was lost

### Molecular and phylogenetic characterization

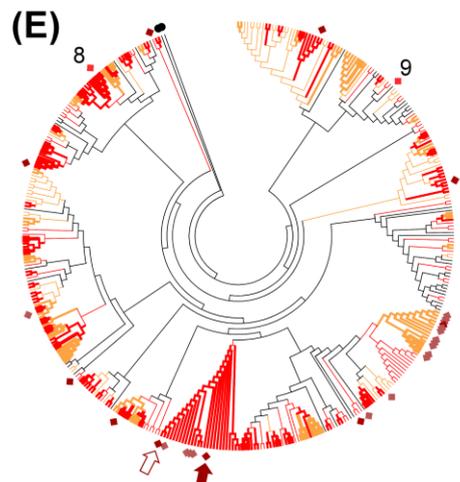
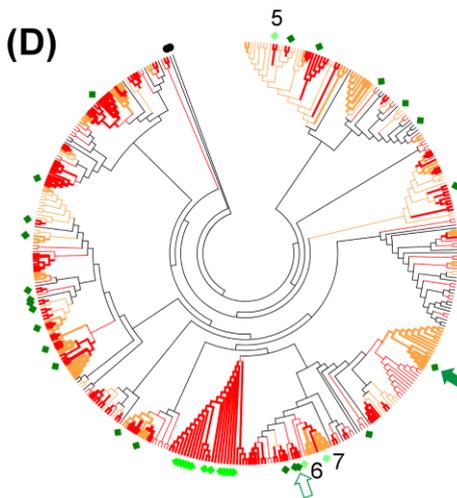
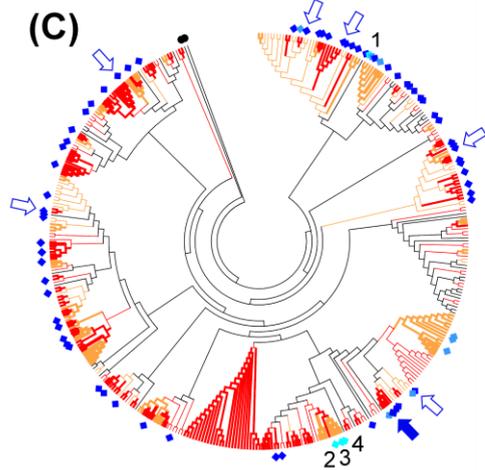
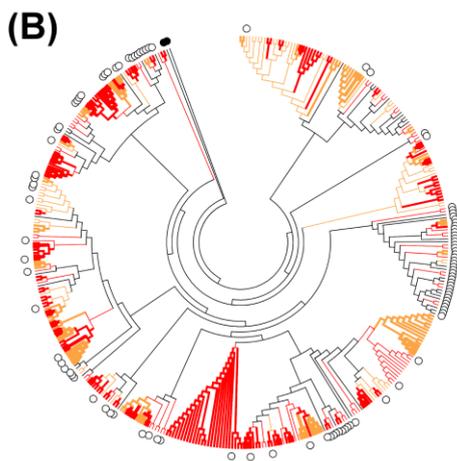
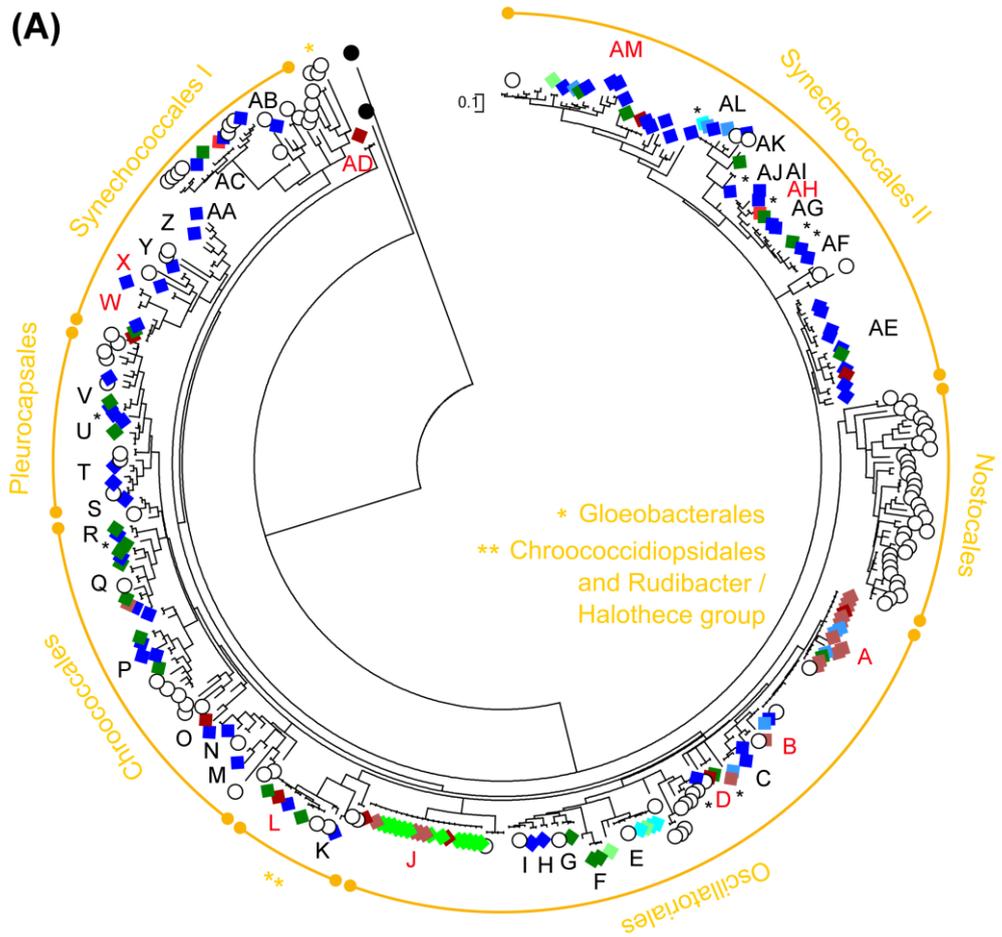
The 454-pyrosequencing generated 10836 high quality reads in total, for the three mat samples. Reads were reduced to 10487 after removing singletons (54 were from cyanobacteria; see Table 2.8). From those, cyanobacterial 16S rDNA sequences accounted for > 85% of total reads in any of the samples, while plastids sequences only accounted for  $\leq 0.8\%$  in any of the mats (see also Supplementary Figs. 2.25 and 2.26). In Fig. 2.19 the circular ML tree with 145 16S rDNA sequences obtained in this study (Table 2.8) is shown, along with sequences from Reference strains (Castenholz et al. 2001) and from BLAST search results. The same tree but with sequence labels is available from Appendix A3; additionally, the tree in a rectangular form can be found through Appendix A2, enabling the detailed visualization of the phylotypes assigned to sequences. The sequences obtained from mats of the Araruama lagoons are widespread across virtually the entire tree (Fig. 2.19, A). The same holds true when looking at sequences from each mat (Fig. 2.19, C-E). Exceptions are the clade of the order Nostocales and the early-branching Gloeobacterales, without sequences obtained in this study. Fifty-seven sequences (39% of total) were included in the Synechococcales, corresponding to 25 phylotypes (see also Table 2.11), and 55 sequences (38%) in the clade of Oscillatoriales, corresponding to 12 phylotypes (Fig. 2.19, A). The remaining sequences were distributed by the orders Chroococcales (13%; eight phylotypes), Pleurocapsales (6%; four phylotypes), Chroococciopsidales and *Rudibacter-Halotheca* group (3%, one phylotype in each clade). Looking at the 76 sequences obtained from EB1 (Fig. 2.19, C), the most abundant 97% OTU, encompassing 20.6% of the total number of 454-reads of this sample, is placed in the clade of phylotype C in the Oscillatoriales (see also Appendix A2). Six other OTUs sequences, placed in phylotypes from different lineages of Synechococcales, Pleurocapsales or Oscillatoriales, had over 4% relative abundance. These observations contrast with the pyrosequencing data for mats collected at the other two sites. The mat from EB2 (Fig. 2.19, D) was clearly dominated by a single OTU sequence which

represented 84.7% of the total reads. This sequence is placed in phylotype A, which includes the Reference strains *Coleofasciculus* (ex-*Microcoleus*) *chthonoplastes* CCY9606 and PCC 7420 (Siegesmund et al. 2008; Appendix A2). The mat from EB3 (Fig. 2.19, E) was also largely dominated by a single sequence, representing 85.9% of the reads from this sample. It was included in phylotype J, which also encompasses the second most abundant sequence in the sample (the only other above 4% of total reads). Thus, this unidentified Oscillatoriales phylotype represented 90% of the 454-reads obtained from the mat collected at EB3 (Supplementary Fig. 2.22).

Regarding the DGGE band sequences, those from the EB1 and EB3 mats were placed among different lineages of the tree (7 sequences in 6 phylotypes and 17 sequences in 5 phylotypes, respectively; see also Appendix A2 and Appendix A4). The DGGE band sequences from EB2 were all placed in the clade of phylotype J. Concerning the isolate-derived sequences, these were found to belong to different lineages of the order Synechococcales (4 isolates from the three mats) or to the same lineage within the Oscillatoriales (phylotype E), as shown Fig. 2.19. The clade of this latter phylotype contains the Reference strain *Geitlerinema* sp. PCC 7105. These findings are in accordance with the morphological-based identification (Table 2.10).

Interestingly, of the 75 sequences (51.7% of total) obtained in this study with best Blast hit results of  $\geq 99\%$ , 51 had a sequence obtained from a saline environment as best hit (Appendix A4) and only four produced a best Blast hit for sequences obtained from non-saline habitats. Even more remarkable is that 42 of the 51 matching sequences from saline environments were obtained from hypersaline microbial mats collected at a single location – Guerrero Negro, Baja California Sur, Mexico (Harris et al. 2013). The Araruama's and the highly similar Guerrero Negro's hypersaline cyanobacterial sequences grouped into nine distinct lineages (phylotypes A, B, D, J, L, W, X, AD, AH and AM) (Fig. 2.19; see also Appendix A2 and Appendix A4).

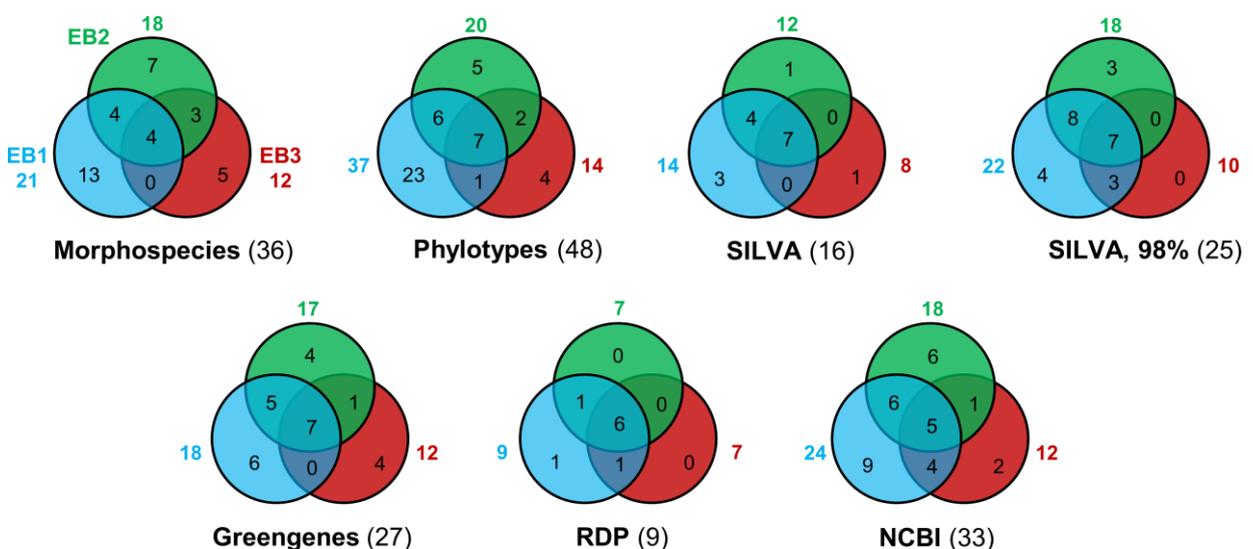
The metagenomic results obtained using the SILVAngs pipeline, where the similarity cutoff chosen for clustering OTUs was 98%, can be visualized in interactive, hierarchical Krona charts (Ondov et al., 2011) in a permalink that was archived by WebCite at <http://www.webcitation.org/6kiUALfVA> (see also Supplementary Fig. 2.26)



**Figure 2.19.** Circular phylogenetic ML trees (log-likelihood of -15279.7163) based on cyanobacterial 16S rDNA sequences. Sequences obtained in this study are marked with colored diamond squares (see below), while white circles denote sequences from Reference strains. The phylogram tree in (A) shows the taxonomic classification for the sequences, at the order level, and their assigned phylotypes (capital letters; highlighted in red are phylotypes that also include sequences obtained from hypersaline microbial mats from Guerrero Negro (see Harris et al. (2013) and text for details). Black asterisks indicate phylotypes consisting in single sequences that have no close relatives (i.e. loner sequences sensu Wilmotte and Herdman, 2001). The trees in (B-E) are the cladogram version of the tree in (A); tree branches in orange represent values of bootstrap support >50%, and in red >75% (1000 replicates). In (B) are highlighted the Reference strains sequences; in (C) the sequences from the mat collected at EB1 (Araruama lagoon); in (D) those from EB2 (Pitanguinha); and in (E) the sequences from EB3 (Pernambuco). Bluish diamonds indicate sequences from EB1, greenish are from EB2, and brownish are from EB3. Darker colors refer to 454-OTUs, lighter to isolates and normal colors are for DGGE-derived sequences. Numbers in (C-E) highlight the isolates obtained from each mat: 1. *Leptolyngbya* aff. *ectocarpus* LEGE 11389; 2., 3. and 4. *Geitlerinema* cf. *lemmermannii* strains LEGE 11390, 11391 and 11396; 5. *Nodosilinea* sp. LEGE 11395; 6. and 7. *Geitlerinema* cf. *lemmermannii* strains LEGE 11393 and 11401; 8. *Synechococcus* sp. LEGE 11394; 9. *Leptolyngbya* sp. LEGE 11392. Arrows point out all OTUs encompassing more than 4% of the total pyrosequencing reads from a sample. In addition, filled arrows indicate the most abundant OTU of each sample. Tree was rooted with the rumen *Melainabacterium* strain YS2 (AF544207) and *Chloroflexus aurantiacus* J-10-f (CP000909) as outgroup.

### Richness, diversity and species composition comparisons

The congruence or divergence between taxon richness ( $S$ ) values obtained by the different approaches is illustrated by Venn diagrams, in Fig. 2.20. These depict “species” richness values obtained following diverse identification methodologies, either manually curated classifications (i.e. morphologically- and phylogenetically-based) or automatic hierarchical classifications (SILVA, Greengenes, RDP, NCBI Taxonomy; these classifiers assign query sequences into hierarchical categories or paths which often are not taxonomic entities. However, for simplicity, from now on they are referred to as taxa).



**Figure 2.20.** Venn diagrams showing the number of distinct cyanobacterial taxa distinguished in each mat sample, by different approaches (including a morphological-based identification, a phylogenetic-guided categorization or an automatic taxonomic classification using different classifiers). If not explicitly specified, OTU consensus sequences were defined as a cluster of reads with 97% similarity. In parentheses are the number of unique taxa identified in all samples.

In addition, the Venn diagrams allow for comparison between mat sample data. Regardless of the method used, EB1 was invariably the mat that showed a higher number of taxa (Fig. 2.20).

By contrast, EB3 was the mat with the lowest number of taxa (the only exception was with the RDP classifier, when the 16S rDNA sequences were distinctly classified into seven taxa, and the same value as in EB2). The different methods show that the number of unique taxa in each mat (that is, that only was present in that sample) was also higher in EB1 than in the other two mats, and, in general, higher in EB2 than in EB3. The number of common taxa present in all three mats varied from four, a value obtained by the morphological-based identification, to seven, as demonstrated by several other classification approaches. There were more taxa shared by EB1 and EB2 than by EB1 and EB3, or by EB2 and EB3. With regard to the methods, the number of unique taxa recognized in all samples was higher when looking at phylotypes (48), morphospecies (36) or at sequences classified using the NCBI Taxonomy database (33). The RDP classifier (9) had the lowest performance in differentiating the cyanobacterial diversity present on the mats from this system. Comparing the two groups of sequences classified using the SILVA database, one with the 454-OTUs defined by a similarity cutoff of 97% and the other by a cutoff of 98%, the most stringent definition of OTU (i.e. 98%) increased the number of distinct taxa obtained (25 vs 16 when using the threshold of 97%); however, falling short of the number of taxa classified by using the 97% OTUs and the Greengenes database, which was 27. It is worth to mention that the number of unclassified sequences obtained when using the automatic classifiers was high, with NCBI Taxonomy being the exception (see Appendix A4). For instance, there were 42 unclassified sequences using RDP and 55 using the SILVA classifier. In these cases, dissimilar (i.e. phylogenetically divergent) sequences are indistinctly classified as belonging to a same entity, generically termed “unclassified” or “unclassified cyanobacteria”. With Greengenes, however, it was possible to differentiate unclassified sequences into distinct OTUs. In brief, the mat sample from EB1 was the more diverse and rich in terms of cyanobacterial taxa present, while EB3 was the less diverse and rich. The categorization of phylotypes by distinguishing the higher number of taxa, was the method which attributed greater diversity to the samples, followed by the morphological-based identification of taxa and the hierarchical classification of sequences following the NCBI Taxonomy.

Table 2.11 shows the species richness by cyanobacterial order and compares values obtained in this study, for morphospecies and phylotypes, with the number of morphospecies previously reported for the Araruama lagoon system, as recovered from the survey (Supplementary Table 2.13). With 21 new cyanobacterial species records and eight new genera records, this study has increased by 16.3% and 20% the number of (morpho-) species and genera reported for the Araruama system, respectively. The Synechococcales was the order with more species (12) and genera (6) newly reported, followed by Oscillatoriales, with six species and one genus newly reported. In Table 2.9 are shown which are the morphospecies reported for the first time for the Araruama's system. Comparable results were obtained for taxonomic assignments of morphospecies and phylotypes at the order level (Table 2.11). The main differences were the identification of Spirulinales species by the morphological-based approach, an order not

detected in the 16S rDNA-based phylogeny of Araruama's sequences, and the detection of phylotypes within the Pleurocapsales, Chroococciopsidales and *Halothece*-related lineages, 3 groups of taxa that could not be identified morphologically.

**Table 2.11** Taxon richness comparison, by taxonomic order, of morphospecies and phylotypes identified in this study and morphospecies previously reported for lagoons from the Araruama's entire complex, as retrieved from the literature survey (see Supplementary Table 2.13 for the full checklist).

| Order                                       | Number of morphospecies (this study) # | Number of unique phylotypes (this study) | Number of morphospecies (other studies) # | Total number of morphospecies, including this study # |
|---|--|--|---|---|
| Chroococcales                               | 8 (6), 1 new record &                  | 8 [including 1 loner sequence]           | 46 (14)                                   | 47 (15)   |
| Chroococciopsidales                         | 0                                      | 1  | 1 (1)                                     | 1 (1)   |
| Gloeobacterales                             | 0                                      | 0  | 0   | 0   |
| Nostocales                                  | 0                                      | 0  | 4 (3)                                     | 4 (3)   |
| Oscillatoriales                             | 9 (6), 6 new                           | 12 [2 loner sequences]                   | 27 (9)                                    | 33 (10)   |
| Pleurocapsales                              | 0                                      | 4 [1 loner sequence]                     | 1 (1)                                     | 1 (1)   |
| <i>Rudibacter</i> / <i>Halothece</i> group* | 0                                      | 1  | 0   | 0   |
| Spirulinales                                | 3 (1), 2 new                           | 0  | 5 (1)                                     | 7 (1)   |
| Synechococcales                             | 16 (10), 12 new                        | 22 [5 loner sequences]                   | 24 (11)                                   | 36 (17)   |
| Total                                       | 36 (22), 21 new records                | 48 [9 loner sequences]                   | 108 (40)                                  | 129 (48)  |

# Number of genera shown in parentheses

& New species records for Araruama's entire system

\* This is a cyanobacterial lineage that will probably give origin to a separate order, as stated in Komárek et al. (2014). For now, these genera are placed within the order Chroococcales (Komárek et al., 2014). Some OTUs belonging to this lineage were previously detected by Clementino et al. (2008), in water samples from Araruama's main lagoon. Moreover, the genus *Halothece* may actually contain some species of *Aphanothece* (order Synechococcales), including *A. halophytica*, a halophilic species common in hypersaline environments (see Oren, 2012 for taxonomic details)

The cyanobacterial species richness estimates for the samples, obtained after applying morphological- or phylogenetically-based, manually curated classifications, or just after clustering of OTUs directly derived from metagenomic pipelines are depicted in Table 2.12. This table also shows other diversity measures for the cyanobacteria present in the mat samples but, in this case, considering only phylotypes or 97% OTUs (i.e., the cutoff level of sequences used in the phylogenetic analysis; see Table 2.8). The value of *S* was higher when looking at unclassified OTUs than when looking at morphospecies or phylotypes, and among those, and expectedly, was higher for the more stringent 98% OTUs, ranging from 68 in the EB3 mat to 258 in EB1. Irrespectively of the type of taxa categorization, *S* was consistently higher for the mat collected at EB1 (ranged between 21 and 258, depending on the type of categorization) and consistently lower for that from EB3 (12-68). In general, *H'* and *1/D* values varied accordingly among samples and between the two types of taxa categorization. They were higher in the mat from EB1, with values of 2.57 and 3.12 for *H'* estimates and 8.87 and 12.41 for *1/D* estimates, concerning phylotypes or 97% OTUs, respectively. The values for mats from EB2 and EB3 were similar between them, although slightly lower for EB3. For EB2, *H'* was 0.65 and

$1/D$  was 1.39 if looking at phylotypes, and 0.72 and 1.39 if looking at 97% OTUs. For EB3, the estimates of these two diversity indices were 0.42 and 1.23, for phylotypes, and 0.59 and 1.35, for 97% OTUs. Values of  $E_H$  estimates were higher for the EB1 sample (0.71 and 0.74 respectively for phylotypes and 97% OTUs) and lower for EB3 if looking at phylotypes (0.16 vs 0.22 for EB2) or for EB2 if looking at 97% OTUs (0.22 vs 0.24 for EB3).

**Table 2.12** Diversity estimates, considering different categorizations of taxa and/or molecular data processing

| Mat<br>sample | "Species" richness (S)     |                         |                          |                          | Shannon's<br>diversity ( $H'$ ) |                         | Simpson's<br>diversity ( $1/D$ ) |                         | Shannon's<br>evenness ( $E_H$ ) |                         |
|---------------|----------------------------|-------------------------|--------------------------|--------------------------|---------------------------------|-------------------------|----------------------------------|-------------------------|---------------------------------|-------------------------|
|               | Morphospecies <sup>§</sup> | Phylotypes <sup>§</sup> | 97%<br>OTUs <sup>#</sup> | 98%<br>OTUs <sup>#</sup> | Phylotypes <sup>§</sup>         | 97%<br>OTU <sup>§</sup> | Phylotypes <sup>§</sup>          | 97%<br>OTU <sup>§</sup> | Phylotypes <sup>§</sup>         | 97%<br>OTU <sup>§</sup> |
| EB1           | 21                         | 37                      | 68                       | 258                      | 2.57                            | 3.12                    | 8.87                             | 12.41                   | 0.71                            | 0.74                    |
| EB2           | 18                         | 20                      | 25                       | 128                      | 0.65                            | 0.72                    | 1.39                             | 1.39                    | 0.22                            | 0.22                    |
| EB3           | 12                         | 14                      | 12                       | 68                       | 0.42                            | 0.59                    | 1.23                             | 1.35                    | 0.16                            | 0.24                    |

<sup>§</sup> see also Fig. 2.20

<sup>#</sup> unclassified pyrosequencing OTUs; directly derived from Biocant or SILVAngs metagenomic pipelines according to a 97% or 98% identity threshold for clustering, respectively.

<sup>§</sup> abundances from 454 pyrosequencing-derived data (see Appendix A7); based on the total number of reads that gave rise to each OTU (i.e. consensus) sequence.

### Screening for secondary metabolites

Isolates were screened for the presence of genes involved in the biosynthesis of cyanotoxins and polyketides. With regard to putative toxin producers, no PCR amplification was obtained for any of the nine isolates. On the other hand, PKS genes were detected by PCR in 6 strains (see Supplementary Fig. 2.28): *Leptolyngbya* sp. LEGE 11389, *Geitlerinema* sp. LEGE 11390, LEGE 11391, LEGE 11396, and LEGE 11401, and *Nodosilinea* sp. LEGE 11395.

### 2.3.5 Discussion

In this work, we attempted to study and uncover the cyanobacterial diversity present in hypersaline mats from three lagoons of the Araruama system, while exploring the impact of different classification methods or procedures to evaluate such diversity. The thorough polyphasic approach used confirms and extends the high cyanobacterial diversity reported previously in morphological-based studies, for the entire Araruama system (see Supplementary Table 2.13). The cyanobacterial diversity found in the mat collected at Araruama's main lagoon (EB1) was more rich and more even (i.e. balanced) than that from the other two mats, either by the morphological-based perspective, or by the molecular one (Table 2.12). On the opposite, the mat collected at the Pernambuco lagoon (EB3) was the less diverse and more uneven, although the mat from the Pitanguinha lagoon (EB2) was comparable in terms of diversity indices/estimates (Table 2.12). These two mats shown to be largely dominated by one phylotype (Supplementary Table 2.14), the phylotype A that corresponds to *Coleofasciculus*

*chthonoplastes* (Appendix A2), in the mat from EB2, and the phylotype J, probably *Oxynema* spp. (see below), in the mat from EB3.

Expectedly, the 454-pyrosequencing technology allowed a deep coverage of the diversity present in the samples, particularly when compared to the other culture-independent, molecular technique employed, the PCR-DGGE (Table 2.8). In fact, PCR-DGGE was less capable in unravelling diversity than anticipated, in part due to the multitude of faint bands not able to be recovered for sequencing (Supplementary Fig. 2.21). This issue, that eludes detection, is well known in DGGE (Sánchez et al. 2009). Another recognized artifact that can arise from this technique is the preferential amplification bias (Neilson et al. 2013). It consists of a strong preferential 16S rDNA amplification of one species, and due to the high-resolution discriminatory power of DGGE this can produce a high microdiversity pattern among strains from a same species (Nübel et al. 1997; Neilson et al. 2013). This could explain the DGGE band sequences from EB2 present in the clade of phylotype J, and those from EB3 in phylotype A (Fig. 2.19 and Appendix A2). Curiously, however, phylotype J could not be detected in the EB2 sample by pyrosequencing and the same was also true for phylotype A, which was not detected by pyrosequencing in EB1, despite the same pool of DNA has been used in both techniques (Fig. 2.16). The possibility of primer mismatch seems limited, since the same primer pair was used in the PCRs for both analyses; the only difference is that, for the pyrosequencing approach, CYA-359F had not the GC-clamp. The primers have been designed to specifically target the V3–V4 region of the 16S rRNA gene of cyanobacteria and chloroplasts (Nübel et al. 1997), but may also target other bacterial groups to a lesser extent (Mühling et al. 2008; see also Supplementary Figs. 2.25 and 2.26). They are widely used for the study of cyanobacteria in environmental samples (Mühling et al. 2008) and their suitability for studying complex microbial communities using PCR-DGGE has been demonstrated by Nübel et al. (1997). The 16S rDNA region targeted by these primers is also highly adequate for studying cyanobacterial diversity by NGS methodologies (Mizrahi-Man et al. 2013; Nguyen et al. 2016). Like most PCR-based molecular approaches (Wintzingerode et al. 1997; Speksnijder et al. 2001), these two techniques are prone to bias, artifacts, pitfalls and have limitations, whose discussion and explanation falls beyond the scope of this study (for details on these issues see Mühling et al. 2008; Green et al. 2009; Berry et al. 2011; Scholz et al. 2012; Bragg & Tyson 2014). However, the overarching methodological strategy of this study envisioned complementarity between the techniques. Moreover, we have used a conservative approach to the use of 454-read sequences, for example by avoiding singletons and small sequences (Table 2.8), that may have caused that some sequences phylogenetically close to DGGE band sequences were not included in the phylogeny.

Given the cyanobacterial diversity exposed by the culture-independent methods used, the number of isolates obtained following the laborious and time consuming culture-dependent approaches was scarce. Moreover, five of the nine isolates belonged to a single species,

*Geitlerinema cf. lemmermannii* (Table 2.10 and Fig. 2.19). But quite surprisingly, the clade of this phylotype does not include any sequence obtained by the culture-independent approaches (Appendix A2), even though this same morphospecies was detected in all samples and shown to be abundant in EB1 and EB3 after using the morphological, culture-independent approach (Table 2.9 and Fig. 2.17, B). One possible explanation is bias in the DNA extraction from this cyanobacterium from environmental samples (since the gDNA extraction was not problematic when using the cultured isolates). Once again, this finding reinforces the relevance of complementary methods. Nonetheless, and although we are in the “NGS era”, this inconsistency is in agreement with findings from other fields of microbial ecology that points out the usefulness of culturing due to its capacity to unveil novel microbial diversity, undetected by metagenomics (Lagier et al. 2015). In accordance, the phylogenetic position of *Leptolyngbya* sp. LEGE 11392 (Fig. 2.19 and Appendix A2) revealed that this isolate is a loner sequence (*sensu* Wilmotte & Herdman 2001). Several other sequences showed no close relatives in the public databases, being also placed as loner sequences in the tree (Fig. 2.19). This indicates the presence of novel cyanobacterial diversity, not yet studied at the molecular level. However, the isolation of *Leptolyngbya* sp. LEGE 11392 allowed us to identify this phylogenetically loner sequence at the genus level, hinting that it is very likely a cryptic species within the recognized polyphyletic genus *Leptolyngbya* (Komárek 2016).

The difficulty in isolating and cultivating the different species observed, as portrayed by the low number of isolates achieved and the apparent species selectivity discussed, suggests that improvements will have to be made in isolation strategies (e.g. circumscribe rapidly growing and mobile species, such as *Geitlerinema cf. lemmermannii*, by phototaxis), and cultivation (e.g. change culture media and/or make adjustments to the recipes in order to reach near optimal nutritional requirements for most of the hypersaline species, or use culture medium specifically developed for some species, as for *Aphanothece halophytica* in Yopp et al. 1978). In fact, the most similar GenBank sequences ( $\geq 99\%$ ) for the sequences that we obtained were predominantly from saline environments (91%), evidencing a likely ecological specificity (e.g. salts or other nutritional requirements) of the cyanobacteria living in this ecosystems, an issue that would deserve further investigation. Besides the advantage for taxonomy, characterization and identification, the isolation of strains allows for further analyses, for example those envisaging biotechnological applications or physiological studies. The preliminary results on the production of secondary metabolites obtained for the isolates indicate that they have a potential to produce polyketides, a family of compounds that include many bioactive metabolites, known to be produced by marine cyanobacteria (Brito et al. 2015). None of the isolates shown to possess genes involved in the production of cyanotoxins commonly found in fresh or brackish waters. This finding is in accordance with that from Brito et al. (2012), for marine isolates.

For the first time, a thorough molecular study was performed in order to help characterize and classify the cyanobacterial diversity present in the microbial mats from the Araruama's lagoons.

The only available molecular sequences for cyanobacteria from this lagoon complex were from water samples, obtained from 16S rDNA and *nifH* clone libraries targeting the whole prokaryotic diversity (Clementino et al. 2008). In that study, the water was sampled at two sites, both in the western part of the main waterbody of the complex, the Araruama lagoon (Clementino et al. 2008). EB1 is located in this same lagoon, but on the opposite side, closer to the channel that connects the lagoon with the Atlantic Ocean. The 16S rDNA cyanobacterial sequences in Clementino et al. (2008) were phylogenetically placed within three different clades, one including *Coleofasciculus chthonoplastes* PCC 7420 (X70770), other with *Halotheca* sp. PCC 7418 (AJ000708), and the third containing *Synechococcus* sp. WH8101 (AF001480). The two first strains are well-known halophilic or extremely halotolerant species (Garcia-Pichel et al. 1998; Oren 2012). The strain PCC 7420 was previously known as *Microcoleus chthonoplastes* but the taxonomy of this species was later revised to *Coleofasciculus chthonoplastes* (Siegmond et al. 2008; see also nomenclatural comments in Oren 2012). *Halotheca* sp. PCC 7418 was firstly identified as *Aphanothece halophytica*, and is also known as *Cyanothece* sp. (Garcia-Pichel et al. 1998) due to confusing nomenclatural issues regarding related forms of *A. halophytica*, and which are better explained in Oren (2012). Strains PCC 7420, PCC 7418 and *Synechococcus* sp. WH8101 are included, respectively, in the clades of phylotypes A, L and AC in our study (Fig. 2.19 and Appendix A2) which encompass sequences from the three studied mats. Sequences from Clementino et al. (2008) are compared with sequences from these three phylotypes in a complementary phylogenetic tree presented in Supplementary Fig. 2.27; the sequences were not included in the main phylogeny since these had only a small overlap with sequences obtained in this study (i.e. they would generate a low number of phylogenetic informative nucleotide positions in the alignment). Clades of phylotypes A and L also harbor Araruama's closely-related sequences (>99% similarity) from Guerrero Negro (Harris et al. 2013; Fig. 2.19). This location in Baja California, Mexico, contains one of the most well-studied hypersaline microbial mats, dominated by *Coleofasciculus chthonoplastes* (Garcia-Pichel et al. 1996; Stal 2012). These findings seem to indicate that these cyanobacterial lineages are ubiquitous in hypersaline environments, namely in Araruama, and to some extent compare favorably with the results from morphological-based identifications. For instance, *Coleofasciculus chthonoplastes* was observed to dominate the samples from EB2 and EB3, but was not observed in EB1 (Table 2.9). Three *Aphanothece* spp. were detected, although none could be assigned to *Aphanothece halophytica*. The only *Aphanothece* species observed to be present in the three mats was *A. cf. stagnina* (Table 2.9), which could correspond to sequences in phylotype L (clade of *Halotheca*), also observed in all studied mats. The picocyanobacterium *Synechococcus* sp. was detected by microscopy only in EB3 (Table 2.9), from the same sample from which *Synechococcus* sp. LEGE 11394 was successfully isolated (Table 2.10). However, globally, the correspondence between morphospecies and phylotypes at the species/genus level was not as straightforward. Nevertheless, other categorical examples

include the phylotype E, a clade with the Reference strain *Geitlerinema* sp. PCC 7105 and all our *Geitlerinema* isolates (Table 2.10, Fig. 2.19 and Appendix A2), and the phylotype AE, a clade with the Type strain *Halomicronema excentricum* TFEP1 (Fig. 2.19, Supplementary Fig. 2.27, Appendix A2), a very thin, filamentous Synechococcales that was also detected in all three samples by morphology and shown to be abundant in EB2 (Table 2.10 and Fig. 2.17). Accordingly, the clade of phylotype AE includes 454-pyrosequencing sequences obtained from the three mats (Fig. 2.19). *Halomicronema excentricum* was firstly described from microbial mats, in man-made solar ponds at Eilat, Israel (Abed et al. 2002). There are some other likely positive examples, but that deserve further investigation, such as phylotype J, which includes several DGGE bands from EB2 and EB3 and 454-derived sequences that represent more than 90% of the total reads for EB3 (Fig. 2.19 and Supplementary Table 2.14). This clade, that also contain a sub-clade with *Oscillatoria acuminata* PCC 6304 (see Appendix A2), very likely include *Oxynema lloydianum* CCALA 960 (Chatchawan et al. 2012) since this Type strain is phylogenetically placed in a sister sub-clade of *Oscillatoria acuminata* PCC 6304 (Chatchawan et al. 2012). *Oxynema* cf. *lloydianum* is a morphospecies that was observed to be abundant in EB3 (Table 2.9), and is characterized by having cylindrical filaments, narrowed and bent at their ends (Fig. 2.17, F-G), as described in Chatchawan et al. (2012). Indeed, in a complementary tree performed (Supplementary Fig. 2.27) it is possible to observe the conjoint position of the three sub-clades, although the node supports and robustness of the phylogenetic inference may be compromised due to the short length of sequences. At the order level, the taxonomic correlation between the morphological identification and phylogeny was fairly satisfactory, with a good number of phylotypes being associated with the observed morphospecies (Table 2.11, see also Table 2.9 and Fig. 2.19). This was achieved after following the new classification of cyanobacteria based on phylogeny and morphology, proposed by Komárek et al. (2014), which includes the following phylogenetically delimited new or redefined orders: Chroococcales, Chroococciopsidales, Gloeobacterales, Nostocales, Oscillatoriales, Pleurocapsales, Spirulinales and Synechococcales, as highlighted in Fig. 2.19.

The characterization of microbial diversity is essential to understand the relationship between environmental conditions and ecosystem functioning. From the taxonomic point of view the fundamental units of bacterial diversity are species, while from the ecological point of view these biological units are frequently seen as ecotypes, functional groups or phylogenetic groups (Cohan & Perry 2007; Koepfel et al. 2008; Krause et al. 2014; Schmidt et al. 2015). The recommended sequence similarity threshold values for demarcation of bacterial species or genera using the 16S rRNA gene are 98-99% and 94.5-95%, respectively (Kim et al. 2014; Yarza et al. 2014), that should be complemented with other type of data (i.e. using a polyphasic approach) for an accurate taxonomic classification (e.g. Dvořák et al. 2015; Komárek 2016). To define units of diversity, microbial ecologists rely on clustering of 16S rRNA sequences into OTUs, most commonly using a single definition, where they share >97% or >98-99% sequence

identity for a more flexible or stringent delineation, respectively (Youngblut et al. 2013; Schmidt et al. 2015). However, when trying to identify cyanobacteria using molecular data (e.g. using GenBank or other databases such as classifiers for NGS metagenome data) it often happens that a taxonomic assignment cannot be satisfactorily determined, at least at lower taxonomic ranks. It might be either because of nonexistence of sequences closely enough related in databases (i.e. low-identity between query and match sequences; that generates the already mentioned loner sequences in phylogeny), or due to the lack of reliable and curated sequences in closely related matches (e.g. lack of match sequences from reference strains). This is especially problematic for complex environmental samples, since it is virtually impossible to know from which organisms the sequences correspond to, even if we characterize the sample composition using other criteria (e.g. identification of species by morphological traits). Again, “culturomics” (Lagier et al. 2015) are a good complementary option to overcome the depth bias, or other problems, inherent to molecular/metagenomic approaches. This was also demonstrated in this study, in the above mentioned cases of *Geitlerinema* strains and *Leptolyngbya* sp. LEGE 11392.

Despite being a critical step, achieving an assignment of taxa to sequences is often a challenge and a major issue in molecular-based methods, chiefly in the analysis of metagenomic data directly retrieved from environmental samples (Mobberley et al. 2012; Garcia-Etxebarria et al. 2014; Tuzhikov et al. 2014). Furthermore, coupled with the problem of cyanobacterial taxonomy and the existence of cryptic species (Dvořák et al. 2015; Komárek 2016), it is difficult to obtain a reliable identification at the species or genus level for this group of organisms. For instance, one of the reasons why the number of taxa generated by the automatic classifiers was smaller than that obtained by the two manually curated, i.e. morphospecies and phylotypes, classifications (Fig. 2.20) is linked to the size of the classifier databases, namely the number and diversity of sequences covering the cyanobacteria phylum. Classifiers seek a content curation (Garcia-Etxebarria et al. 2014), and therefore might only incorporate 16S rDNA sequences from Reference strains (Tuzhikov et al. 2014), leading to an underrepresentation of cyanobacterial genetic diversity within the database. Thus, due to the absence of proper reference sequences covering the cyanobacterial diversity, these databases may be unable to classify a significant part of a given data set (Garcia-Etxebarria et al. 2014; Tuzhikov et al. 2014). This results in a classification of sequences into higher taxonomic ranks or paths, or may cause the categorization of several phylogenetically divergent sequences into a same category termed, for example, “unclassified cyanobacteria”. Both of these issues lead to an underestimation of the number of unique taxa by comparison with other classification methods (Fig. 2.20). In fact, Nguyen et al. (2016) concluded that we need to move beyond 16S rDNA sequence similarity-based clustering techniques for 16S rDNA metagenomic analysis, and suggest a phylogeny-based OTU clustering with the corresponding development of algorithms and bioinformatic tools for the future. Naturally, our manual curation process for classifying

phylotypes, although quite laborious, was possible because we were focusing on a specific group of microorganisms, and thus a relatively small number of 454-read sequences was obtained (Table 2.8). The classification was based on a simple criterion, the bootstrap support of clades (Appendix A2). It still is a broad, inexact demarcation of “taxa”, since clades may include lineages more or less divergent (i.e. sequences more or less similar), but ensures that phylogenetically close related sequences are grouped together.

In conclusion, it was shown that the mats from coastal lagoons of Araruama harbor a high cyanobacterial diversity. This assumes greater relevance because an exhaustive examination of single samples, collected at each mat, was followed instead of studying diverse samples from each mat. The different approaches and taxonomic assignment methods varied radically in their ability to capture the diversity present in the samples. Nevertheless, the different approaches need to be seen as complementary, in order to attain the most complete picture of the cyanobacterial diversity. The phylogenetic-guided classification of sequences generated the higher number of unique taxa, although several of them could not have been identified. In fact, only with the morphological-based approach it was possible to identify almost all the recognized cyanobacterial components of the mat samples, at least at the genus level. However, at the order level, the taxonomic inferences were generally congruent between the two classification approaches. Finally, we have demonstrated that some of the cyanobacterial isolates obtained in this study have the potential to produce bioactive compounds, although the microbial diversity from the extreme environment of these lagoon’s ecosystems is still understudied as a source for new secondary metabolites.

### **2.3.6 Acknowledgments**

*Funding:* This work was funded by Portuguese National Funds through FCT - Fundação para a Ciência e a Tecnologia, grants SFRH/BD/80153/2011 (to VR).

### 2.3.7 Supplementary Data

**Table 2.13** Checklist of cyanobacteria species reported for different water bodies from the Araruama lagoon system, as depicted from this study (highlighted in bold) or retrieved from the literature survey.

| Taxa <sup>#</sup> +   | Sample description         |                                  |  | Lagoon <sup>£</sup>  | References <sup>@</sup>                 |
|---|----------------------------|----------------------------------|--|--|---|
|   | Type                       | Microbial structure <sup>*</sup> | Shape <sup>§</sup>                     |  |   |
| <b>Chroococcales</b>  |                            |                                  |  |  |   |
| <i>Aphanothece castagnei</i> (Brébisson) Rabenhorst                                     | mat, sediment, rock        | l, lay, s                        | b, colu, d, f, po, pu, sm, t           | ARA (man-made saltern), BES, BPF, PER, PIT, SAJ                      | 1-4, 6, 8, 17                           |
| <i>Aphanothece clathrata</i> West & G.S.West <sup>(17)</sup>                            | mat, rock                  | l, lay, non                      | b, f, po, pu, sm, t                    | ARA (man-made saltern), BES, PER                                     | 2, 4, 17                                |
| <b><i>Aphanothece conglomerata</i> Rich</b>   | mat, sediment, rock, water | l, lay, non                      | coll, f, sm                            | <b>ARA</b> , ARA (man-made saltern), BES, BPF, PER, PIT              | 1, 4, 6, 7, 10, 17, <b>this study</b>   |
| <i>Aphanothece halophytica</i> Frémy <sup>(6)(7)</sup>                                  | mat, sediment, rock, water | l, lay, s, t                     | b, coll, colu, d, f, po, pu, sm, t     | ARA (man-made saltern), BES, BPF, PER, PIT, SAJ                      | 1-8, 10-12, 15, 17                      |
| <i>Aphanothece marina</i> (Ercegovic) Komárek & Anagnostidis                            | mat, sediment, rock        | l, lay                           | b, coll, f, po, pu, sm, t              | ARA (man-made saltern), BES, PER, PIT                                | 1-4, 10, 11, 15, 17                     |
| <i>Aphanothece microscopica</i> Nägeli  | mat, water                 | lay                              | f, sm                                  | ARA (man-made saltern), BES, PIT                                     | 7, 17                                   |
| <i>Aphanothece pallida</i> (Kützing) Rabenhorst   | mat, rock, water           | l, non                           | b, coll, f, po, pu, t                  | PIT  | 1-4, 7, 10                              |
| <b><i>Aphanothece salina</i> Elenkin &amp; Danilov</b>                                  | mat, sediment, rock, water | l, lay, non                      | coll, f, po, pu, sm                    | <b>ARA</b> , ARA (man-made saltern), BPF, PER, PIT, SAJ              | 1, 3, 4, 6-8, 11, 17, <b>this study</b> |
| <i>Aphanothece saxicola</i> Nägeli  | mat, sediment, rock        | l, lay, s, t                     | b, colu, d, f, po, pu, sm, st, t       | ARA (man-made saltern), BES, BPF, PER, PIT                           | 1-6, 10-12, 15, 17                      |
| <b><i>Aphanothece stagnina</i> (Sprengel) A.Braum <sup>(6)</sup></b><br><sup>(17)</sup> | mat, sediment, rock, water | l, lay, s                        | b, coll, f, po, pu, sm, t              | <b>ARA</b> , ARA (man-made saltern), BES, BPF, <b>PER, PIT</b> , SAJ | 1-8, 10, 14, 17, <b>this study</b>      |
| <i>Chlorogloea tuberculosa</i> (Hansgirg) Wille   | mat                        | lay                              | f, sm                                  | BES  | 17                                      |
| <i>Chroococcus</i> sp. <sup>(17)</sup>  | mat                        | lay                              | f, sm                                  | BES  | 17                                      |
| <i>Chroococcus dispersus</i> (Keissler) Lemmermann                                      | mat, sediment, rock        | l, lay                           | coll, f, po, sm                        | PER, PIT   | 1, 3, 4, 11, 15                         |
| <i>Chroococcus giganteus</i> West   | mat, rock                  | l, lay                           | f, po, sm                              | PER, PIT   | 1, 3, 4, 11                             |
| <i>Chroococcus membraninus</i> (Meneghini) Nägeli                                       | mat, rock, water           | l, lay, s, t                     | b, coll, f, po, sm                     | AZU, BES, BPF, PER, PIT  | 1, 3-5, 7, 9-12, 14, 15, 17             |
| <i>Chroococcus microscopicus</i> J.Komárková-Legnerová & G.Cronberg <sup>(8)</sup>      | mat, sediment, rock        | l, lay, t                        | b, coll, f, po, pu, sm                 | ARA (man-made saltern), BES, BPF, PER, PIT, SAJ                      | 1, 3-6, 8, 10, 12, 14, 15, 17           |
| <i>Chroococcus minimus</i> (Keissler) Lemmermann <sup>(8)</sup><br><sup>(17)</sup>      | mat, sediment, rock, water | l, lay, t                        | b, coll, f, po, pu, sm                 | ARA (man-made saltern), BES, BPF, PER, PIT, SAJ                      | 1, 3, 4, 6-8, 10-12, 14, 15, 17         |
| <i>Chroococcus minor</i> (Kützing) Nägeli   | mat, sediment, rock        | l, lay, s, t                     | b, coll, colu, d, f, po, pu, sm, st, t | ARA (man-made saltern), AZU, BES, BPF, PER, PIT                      | 1-4, 6, 9-12, 14, 15, 17                |
| <i>Chroococcus minutus</i> (Kützing) Nägeli <sup>(4)</sup><br><sup>(17)</sup>           | mat, sediment, rock, water | l, lay, t                        | b, coll, f, po, pu, sm                 | ARA (man-made saltern), BES, BPF, PER, PIT, SAJ                      | 1, 3-8, 10-12, 14, 15, 17               |
| <i>Chroococcus obliteratus</i> Richter  | mat, rock                  | l, non                           | coll, po, sm                           | PER, PIT   | 1, 3, 4, 10                             |
| <i>Chroococcus prescottii</i> Drouet & Daily  | mat                        | non                              | b, coll                                | BPF, PIT   | 1, 5, 10                                |

|  |                               |                             |   |  |  |
|--|-------------------------------|-----------------------------|---|--|--|
| <i>Chroococcus quaternarius</i> Zalesky  | mat, sediment,<br>rock        | l, lay                      | coll, f, po, sm                           | AZU, PER, PIT  | 1, 3, 4, 9, 11                             |
| <i>Chroococcus submarinus</i> (Hansgirg) Kováčik   | sediment                      | -                           | -   | AZU  | 9  |
| <i>Chroococcus tenax</i> (Kirchner) Hieronymus   | mat                           | lay                         | coll, f, sm                               | BES, PIT   | 10, 11, 14                                 |
| <b><i>Chroococcus turgidus</i> (Kützing) Nägeli</b> <sup>(4)(8)</sup>  | mat, sediment,<br>rock, water | l, lay, s                   | b, coll, colu, d, f, po, pu,<br>sm, st    | <u>ARA</u> , ARA (man-made saltern),<br>AZU, BES, BPF, PER, PIT, SAJ | 2-6, 8-12, 14, 15, 17, <b>this study</b>   |
| <i>Chroococcus turicensis</i> (Nägeli) Hansgirg  | water                         | -                           | -   | PIT  | 7  |
| <b><i>Cyanosarcina thalassia</i> Anagnostidis &amp; Pantazidou</b>   | mat, rock                     | l, lay, t                   | coll, f, po, pu, sm                       | BES, PER, <b>PIT</b>   | 1, 3, 4, 11, 12, 14, 17, <b>this study</b> |
| <i>Entophysalis conferta</i> (Kützing) Drouet & Daily <sup>(1)</sup>   | mat                           | lay                         | b, f, po, pu, sm, t                       | BES, PIT   | 1, 11, 14, 17                              |
| <i>Entophysalis granulosa</i> Kützing <sup>(1)(4)(11)</sup>  | mat, rock                     | l, lay                      | coll, f, po, pu, sm                       | BES, PER, PIT  | 1, 3, 4, 10, 11, 14, 17                    |
| <b><i>Geminocystis</i> sp.</b>   | mat                           | lay                         | po, sm                                    | <u>ARA</u> , <u>PER</u> , <u>PIT</u>                                 | <b>this study</b>                          |
| <i>Gloeocapsa</i> sp.  | mat, rock                     | stromatolitic-<br>like, lay | -   | VER  | 16   |
| <i>Gloeocapsa punctata</i> Nägeli  | mat                           | lay                         | b, f, po, pu, t                           | PIT  | 1, 11                                      |
| <b><i>Gloeocapsopsis</i> cf. <i>crepidinum</i> (Thuret) Geitler ex Komárek</b>   | mat                           | lay                         | b, f, po, pu, sm, t                       | <u>ARA</u> , BES, <b>PIT</b>   | 1, 17, <b>this study</b>                   |
| <i>Gloeocapsopsis magma</i> (Brébisson) Komárek & Anagnostidis   | mat                           | lay                         | b, f, po, pu, sm, t                       | ARA (man-made saltern), PIT  | 1, 11, 17                                  |
| <i>Gloeothece confluens</i> Nägeli   | mat, rock                     | l, non                      | b, coll, f, po, pu, sm, t                 | PER, PIT, SAJ  | 1, 3, 4, 8, 10                             |
| <i>Gloeothece linearis</i> Nägeli  | mat, rock                     | l, lay                      | b, coll, f, po, pu, sm, t                 | BES, PER, PIT  | 1, 3, 4, 11, 17                            |
| <b><i>Gloeothece subtilis</i> Skuja</b>  | mat, sediment                 | lay                         | b, f, po, pu, sm, t                       | <u>ARA</u> , PIT   | 1, 11, 15, <b>this study</b>               |
| <i>Gloeothece vibrio</i> N.Carter <sup>(9)</sup>   | mat                           | lay, non                    | b, f, po, pu, sm, t                       | ARA (man-made saltern), PIT, SAJ                                     | 1, 8, 17                                   |
| <i>Gomphosphaeria aponina</i> Kützing  | mat, rock                     | l, non                      | coll, po, sm                              | PER  | 3, 4                                       |
| <i>Halothece</i> sp. (including " <i>Euhalothece</i> " sensu Garcia-Pichel <i>et al.</i> (1998); 11 OTUs) <sup>&amp;</sup>         | water                         | -                           | -   | ARA  | 13   |
| <i>Johannesbaptistia pellucida</i> (Dickie) W.R. Taylor & Drouet   | mat, sediment,<br>rock        | l, lay                      | b, coll, f, po, pu, sm, t                 | AZU, BES, PER, PIT   | 1-4, 9, 14, 15, 17                         |
| <i>Lithocapsa fasciculata</i> Ercegovic  | mat                           | non                         | b, f, po, pu, t                           | PIT  | 1  |
| <i>Pseudocapsa dubia</i> Ercegovic   | mat                           | lay                         | f, sm                                     | BES, PIT   | 1, 11, 14                                  |
| <i>Pseudocapsa maritima</i> Komárek  | mat                           | lay                         | f, sm                                     | BES  | 17   |
| <i>Pseudocapsa sphaerica</i> (Proskina-Lavrenko) Kováčik   | mat                           | lay                         | b, f, po, pu, sm, t                       | ARA (man-made saltern), BES, PIT, SAJ                                | 1, 8, 11, 14, 17                           |
| <i>Xenococcus schousboei</i> Thuret  | mat, sediment                 | lay                         | b, f, po, pu, sm, t                       | PIT  | 1, 11, 15                                  |
| <i>Xenotholos kernerii</i> (Hansgirg) M.Gold-Morgan <i>et al.</i>  | mat                           | lay                         | b, f, po, pu, sm, t                       | ARA (man-made saltern), PIT, SAJ                                     | 1, 8, 11, 17                               |
| <b>Chroococciopsidales</b>   |                               |                             |   |  |  |
| <i>Chroococciopsis fissuratum</i> (Ercegovic) Komárek & Anagnostidis   | water                         | -                           | -   | PIT  | 7  |
| <b>Oscillatoriales</b>   |                               |                             |   |  |  |
| <b><i>Coleofasciculus</i> (=Microcoleus) <i>chthonoplastes</i> (Gomont) M.Siegesmund <i>et al.</i></b> <sup>(1)(2)(3)(4)(10)</sup> | mat, sediment,<br>rock, water | l, lay, s                   | b, coll, colu, d, f, po, pu,<br>sm, st, t | ARA (man-made saltern), AZU,<br>BES, <b>PER</b> , <b>PIT</b> , SAJ   | 1-4, 7-10, 14, 15, 17, <b>this study</b>   |

|  |                     |                         |  |  |                            |
|--|---------------------|-------------------------|--|--|----------------------------|
| <b>Geitlerinema aff. amphibium (Agardh ex Gomont) Anagnostidis</b>                                 | mat                 | lay                     | po                                     | <b>PIT</b>                                 | <b>this study</b>          |
| <b>Geitlerinema cf. lemmermannii (Woloszynska) Anagnostidis</b>                                    | mat                 | lay                     | po, sm                                 | <b>ARA, PER, PIT</b>                       | <b>this study</b>          |
| <i>Lyngbya aestuarii</i> Liebman ex Gomont <sup>(1)(4)(10)(14)</sup>                               | mat, sediment, rock | l, lay, s               | b, coll, colu, d, f, po, pu, sm, st, t | BES, BPF, PER, PIT                         | 1-4, 6, 10, 14             |
| <i>Lyngbya confervoides</i> Agardh ex Gomont   | mat                 | non                     | b, f, po, pu, t                        | PER  | 2                          |
| <i>Lyngbya (=Porphyrosiphon) martensianus</i> Meneghini ex Gomont                                  | Mat                 | lay                     | sm                                     | BES  | 14                         |
| <b>Microcoleus aff. steenstrupii Petersen</b>  | Mat                 | lay                     | po, sm                                 | <b>ARA, PER</b>                            | <b>this study</b>          |
| <i>Microcoleus vaginatus</i> Gomont ex Gomont  | mat, rock, water    | l, non, s               | b, coll, colu, d, f, po, pu, sm, st, t | PER, PIT, SAJ                              | 1-4, 7, 8, 11              |
| <i>Microcoleus</i> sp.   | mat, rock           | stromatolitic-like, lay | -                                      | VER  | 16                         |
| <b>Oscillatoria limosa Agardh ex Gomont</b>  | Mat                 | lay, non                | b, f, po, pu, t                        | <b>PER, PIT</b>                            | <b>2, this study</b>       |
| <b>Oscillatoria margaritifera Kützing ex Gomont</b>  | Mat                 | lay                     | po                                     | <b>PIT</b>                                 | <b>this study</b>          |
| <i>Oscillatoria (=Spirulina) meneghiniana</i> Zanardini ex Gomont                                  | mat, water          | non                     | b                                      | BPF, PIT                                   | 5, 7                       |
| <b>Oscillatoria subbrevis Schmidle</b>   | mat, sediment, rock | l, non                  | b, coll, f, po, pu, sm, t              | AZU, PER, <b>PIT</b>                       | <b>2-4, 9, this study</b>  |
| <i>Oscillatoria terebriformis</i> f. <i>amphigranulata</i> Elenkin & Kossinskaja                   | mat                 | non                     | b, f, po, pu, t                        | PER  | 2                          |
| <i>Oscillatoria</i> sp.  | mat                 | lay                     | f, sm                                  | ARA (man-made saltern)                     | 17                         |
| <i>Oxynema (=Phormidium) acuminatum</i> (Gomont) Chatchawan <i>et al.</i>                          | mat, rock           | l, lay                  | b, coll, f, po, pu, sm, t              | BES, PER, PIT                              | 1, 3, 4, 11, 14, 17        |
| <b>Oxynema lloydianum (Gomont) Chatchawan, Komárek, Strunecky, Smarda &amp; Peerapornpisal</b>     | mat                 | lay                     | po, sm                                 | <b>PER, PIT</b>                            | <b>this study</b>          |
| <i>Phormidium acutum</i> (Brühl & Biswas) Anagnostidis & Komárek                                   | mat, rock           | l, non                  | b, coll, f, po, pu, sm, t              | PER, PIT                                   | 1, 3, 4                    |
| <i>Phormidium articulatum</i> (= <i>Oscillatoria articulata</i> ) (Gardner) Anagnostidis & Komárek | mat                 | Non                     | b, f, po, pu, t                        | PER  | 2                          |
| <i>Phormidium breve</i> (Kützing ex Gomont) Anagnostidis & Komárek                                 | mat, sediment, rock | l, lay, t               | b, coll, f, po, pu, sm, t              | AZU, BPF; PER, PIT                         | 1, 3, 4, 6, 9, 11, 12      |
| <i>Phormidium</i> (= <i>Oscillatoria</i> ) <i>foreau</i> (Frémy) Umezaki & Watanabe                | water               | -                       | -                                      | PIT  | 7                          |
| <i>Phormidium formosum</i> (Bory ex Gomont) Anagnostidis & Komárek                                 | mat, sediment       | Non                     | b, f, po, pu, t                        | AZU, PER, PIT                              | 1, 3, 4, 9                 |
| <i>Phormidium hamelii</i> (Frémy) Anagnostidis & Komárek   | mat, rock           | l, lay                  | b, coll, f, po, pu, sm, t              | PER, PIT                                   | 1, 3, 4, 11                |
| <i>Phormidium hormoides</i> Setchell & Gardner   | mat, rock           | l, non                  | b, coll, f, po, pu, sm, t              | PER  | 3, 4                       |
| <i>Phormidium minnesotense</i> (Tilden) Drouet   | mat                 | non                     | po                                     | PER  | 3, 4                       |
| <b>Phormidium nigroviride (Thwaites ex Gomont) Anagnostidis &amp; Komárek</b>                      | mat                 | lay                     | po                                     | <b>PIT</b>                                 | <b>this study</b>          |
| <i>Phormidium okenii</i> (Agardh) Anagnostidis & Komárek <sup>(4)</sup>                            | mat, sediment, rock | l, lay, t               | b, f, po, pu, sm, t                    | ARA (man-made saltern), AZU, BES, PER, PIT | 1, 3, 4, 9, 11, 12, 14, 17 |

|  |                     |                         |                              |   |  |
|--|---------------------|-------------------------|------------------------------|---|--|
| <i>Phormidium terebriforme</i> (Agardh ex Gomont) Anagnostidis & Komárek   | mat, rock           | l, non                  | b, f, po, pu, sm, t          | PER   | 1, 3, 4                                  |
| <i>Phormidium willei</i> (Gardner) Anagnostidis & Komárek <sup>(1)</sup>   | mat, rock           | l, non, t               | b, coll, f, po, pu, sm, t    | PIT   | 1, 3, 4, 12                              |
| <i>Planktothrix rubescens</i> (De Candolle ex Gomont) Anagnostidis & Komárek                                     | rock                | L                       | sm                           | PER   | 4  |
| <i>Symplocastrum</i> (= <i>Schizothrix</i> ) <i>friesii</i> (Gomont ex Gomont) Kirchner <sup>(1)(2)(4)(14)</sup> | mat, sediment, rock | l, lay, s, t            | b, coll, f, po, pu, sm, t    | ARA (man-made saltern), BES, PER, PIT, SAJ      | 1-4, 8, 10-12, 14, 15, 17                |
| <i>Trichodesmium lacustre</i> (= <i>Oscillatoria lacustris</i> ) (Klebahn) Geitler                               | mat                 | lay                     | b, f, po, pu, t              | PIT   | 1, 11                                    |
| unidentified Oscillatoriales ( <i>Coleofasciculus</i> -related, 87-93% similarity; 8 OTUs) <sup>a</sup>          | water               | -                       | -                            | ARA   | 13                                       |
| <b>Nostocales</b>  |                     |                         |                              |   |  |
| <i>Brachytrichia quoyi</i> Bornet & Flahault   | mat, rock           | non, s                  | b, colu, d, f, po, pu, st, t | PER   | 2  |
| <i>Calothrix crustacea</i> Schousboe ex Thuret   | mat, rock           | non, s                  | b, colu, d, f, po, pu, st, t | PER   | 2  |
| <i>Calothrix scopulorum</i> (Weber & Mohr) Agardh  | rock                | l                       | sm                           | PER   | 4  |
| <i>Kyrtuthrix maculans</i> (Gomont) Umezaki  | mat, sediment, rock | l, lay                  | b, coll, f, po, pu, sm, t    | AZU, PER, PIT                                   | 1, 3, 4, 9, 11, 15                       |
| <b>Pleurocapsales</b>  |                     |                         |                              |   |  |
| <i>Pleurocapsa fuliginosa</i> Hauck  | mat, rock           | lay, t                  | f                            | PER, PIT  | 1, 11, 12                                |
| <b>Spirulinales</b>  |                     |                         |                              |   |  |
| <b><i>Spirulina labyrinthiformis</i> Gomont</b>  | mat                 | lay                     | po                           | <b>PER</b>                                      | <b>this study</b>                        |
| <i>Spirulina laxissima</i> G.S.West  | sediment            | -                       | -                            | BPF   | 6  |
| <i>Spirulina meneghiniana</i> Zanardini ex Gomont  | mat                 | lay                     | f, sm                        | BES   | 17                                       |
| <b><i>Spirulina subsalsa</i> Oerstedt ex Gomont <sup>(6)(8)</sup></b>  | mat, sediment       | l, lay, non             | b, coll, f, po, pu, sm, t    | ARA (man-made saltern), BES, BPF; PER, PIT, SAJ | 1, 2, 4, 6, 8, 10, 17, <b>this study</b> |
| <i>Spirulina subtilissima</i> Kützing ex Gomont  | mat, sediment, rock | l, lay, s               | b, coll, f, po, pu, sm, t    | BES, PER, PIT, SAJ                              | 1-4, 8, 10, 11, 14, 15, 17               |
| <b><i>Spirulina tenerrima</i> Kützing ex Gomont</b>  | mat                 | lay                     | po                           | <b>PER</b>                                      | <b>this study</b>                        |
| <i>Spirulina</i> sp.   | mat, rock           | stromatolitic-like, lay | -                            | VER   | 16                                       |
| <b>Synechococcales</b>   |                     |                         |                              |   |  |
| <b><i>Aphanocapsa litoralis</i> (Hansgirg) Komárek &amp; Anagnostidis</b>  | mat                 | l, lay                  | pu, sm                       | <b>ARA, BES, PER, PIT, SAJ</b>                  | 1, 4, 8, 17, <b>this study</b>           |
| <i>Aphanocapsa rivularis</i> (Carmichael) Rabenhorst   | mat                 | l, lay                  | f, sm                        | BES   | 17                                       |
| <b><i>Aphanocapsa salina</i> Woronichin</b>  | mat                 | lay                     | po, pu, sm                   | <b>ARA, PIT, SAJ</b>                            | 8, <b>this study</b>                     |
| <i>Aphanocapsa salinarum</i> Hansgirg  | -                   | -                       | -                            | PIT   | 1  |
| <i>Bacularia caerulescens</i> Borzi  | mat, rock, water    | lay                     | f                            | PIT   | 1, 7, 11, 15                             |
| <b><i>Coelosphaeriopsis</i> cf. <i>halophila</i> Lemmermann</b>  | mat                 | lay                     | sm                           | <b>ARA</b>                                      | <b>this study</b>                        |
| <b><i>Halomicronema excentricum</i> Abed, Garcia-Pichel &amp; Hernández-Maríné <sup>&amp;</sup></b>              | mat                 | lay                     | po, sm                       | <b>ARA, PER, PIT</b>                            | <b>this study</b>                        |

|   |                     |             |  |  |                                   |
|---|---------------------|-------------|--|--|-----------------------------------|
| <i>Jaaginema subtilissimum</i> (=Oscillatoria subtilissima) (Kützing ex De Toni) Anagnostidis & Komárek           | mat, water          | lay         | f, sm                                  | BES, PIT                                   | 7, 17                             |
| <b><i>Komvophoron breve</i> (Carter) Anagnostidis (subgenus <i>Alyssophoron</i> Anagnostidis et Komárek 1988)</b> | mat                 | lay         | sm                                     | <u>ARA</u>                                 | this study                        |
| <b><i>Komvophoron</i> cf. <i>minutum</i> (Skuja) Anagnostidis &amp; Komárek</b>                                   | mat                 | lay         | po                                     | <u>PER</u>                                 | this study                        |
| <b><i>Lemmermanniella</i> sp. Geitler in Engler &amp; Prantl</b>  | mat                 | lay         | sm                                     | <u>ARA</u>                                 | this study                        |
| <b><i>Leptolyngbya crosbyana</i> (Tilden) Anagnostidis &amp; Komárek</b>  | mat                 | lay         | po, sm                                 | <u>ARA, PIT</u>                            | this study                        |
| <b><i>Leptolyngbya</i> cf. <i>ectocarpus</i> (Gomont) Anagnostidis &amp; Komárek<sup>% &amp;</sup></b>            | mat                 | lay         | sm                                     | <u>ARA</u>                                 | this study                        |
| <i>Leptolyngbya</i> (=Lyngbya) <i>fragilis</i> (Gomont) Anagnostidis & Komárek <sup>(1)</sup>                     | mat, sediment, rock | lay, t      | coll, po, sm                           | BES, BPF, PER, PIT                         | 1, 3, 4, 6, 10, 12, 14            |
| <i>Leptolyngbya hypolimnetica</i> (=Phormidium hypolimneticum) (Campbell) Anagnostidis <sup>(17)</sup>            | mat, rock           | l, lay, non | coll, f, po, pu, sm                    | ARA (man-made saltern), BES, PER           | 3, 4, 17                          |
| <i>Leptolyngbya komarovii</i> (Anissimova) Anagnostidis & Komárek   | mat, rock           | l, lay, s   | b, coll, colu, d, f, po, pu, sm, st, t | BES, PER, PIT                              | 1-4, 14                           |
| <i>Leptolyngbya</i> (=Lyngbya) <i>submonilifera</i> (Frémy) Senna & Compère                                       | Mat                 | lay         | f, sm                                  | BES  | 17                                |
| <i>Leptolyngbya tenuis</i> (Gomont) Anagnostidis & Komárek <sup>(17)</sup>  | mat, sediment, rock | l, lay      | coll, f, po, sm                        | ARA (man-made saltern), AZU, BES, PER, PIT | 1, 3, 4, 9-11, 14, 15, 17         |
| <i>Leptolyngbya</i> sp. <sup>(17)</sup>   | mat                 | lay         | f, sm                                  | BES  | 17                                |
| <i>Limnococcus</i> (=Chroococcus) <i>limneticus</i> (Lemmermann) Komárková <i>et al.</i>                          | mat                 | non         | b, f, po, pu, t                        | PER  | 2                                 |
| <i>Merismopedia warmingiana</i> Lagerheim   | mat                 | non         | b                                      | BPF  | 5                                 |
| <b><i>Nodosilinea nodulosa</i> (Li &amp; Brand) Perkerson &amp; Casamatta</b>                                     | Mat                 | lay         | sm                                     | <u>ARA</u>                                 | this study                        |
| <b><i>Nodosilinea</i> sp.<sup>% &amp;</sup></b>   | Mat                 | lay         | po                                     | <u>PIT</u>                                 | this study                        |
| <i>Planktolyngbya limnetica</i> (=subtilis) (Lemmermann) J.Komárková-Legnerová & G.Cronberg                       | mat, sediment, rock | l, lay      | b, f, po, pu, sm, t                    | BES, BPF; PER, PIT                         | 1, 4, 6, 14, 17                   |
| <b><i>Pseudanabaena</i> aff. <i>amphigranulata</i> (Goor) Anagnostidis</b>  | mat                 | lay         | po                                     | <u>PIT</u>                                 | this study                        |
| <b><i>Pseudanabaena</i> (=Oscillatoria) <i>limnetica</i> (Lemmermann) Komárek</b>                                 | mat, rock           | l, non      | b, coll, f, po, pu, sm, t              | <u>ARA</u> , PER, PIT                      | 1, 3, 4, this study               |
| <b><i>Pseudanabaena minima</i> (G.S.An) Anagnostidis</b>  | mat                 | lay         | sm                                     | <u>ARA</u>                                 | this study                        |
| <i>Schizothrix arenaria</i> Gomont  | mat, rock           | l, lay      | b, coll, f, po, pu, sm, t              | BES, PER, PIT                              | 1, 3, 4, 14                       |
| <i>Synechococcus elongatus</i> (Nägeli) Nägeli  | mat, sediment, rock | l, lay, t   | b, coll, f, po, pu, sm, t              | AZU, BES, BPF; PER, PIT                    | 1, 3, 4, 6, 9, 11, 12, 14, 15, 17 |
| <i>Synechococcus salinarum</i> Komárek  | mat, rock           | l, non      | b, f, po, pu, sm, t                    | PER, PIT, SAJ                              | 1, 4, 8                           |

|   |                  |           |  |                  |                      |
|---|------------------|-----------|--|------------------|----------------------|
| <i>Synechococcus subsalsus</i> Skuja  | water            | -         | -                                      | PIT              | 7                    |
| <i>Synechococcus</i> sp.  | mat              | lay       | f, sm                                  | BES              | 17                   |
| <b><i>Synechococcus</i> sp.</b> % &   | mat              | lay       | po                                     | <u>PER</u>       | <b>this study</b>    |
| <i>Synechococcus</i> sp. (6 OTUs) <sup>&amp;</sup>  | water            | -         | -                                      | ARA              | 13                   |
| <b><i>Synechocystis salina</i> Wislouch</b>   | mat, sediment    | lay       | sm                                     | <u>ARA</u> , BPF | <b>6, this study</b> |
| <i>Trichocoleus</i> (=Microcoleus) <i>tenerrimus</i> (Gomont) Anagnostidis <sup>(4)</sup> | mat, rock, water | l, non, s | b, coll, colu, d, f, po, pu, sm, st, t | PER, PIT         | 2-4, 7               |

# Numbers in brackets are bibliographic references of studies indicating that the species was dominating/was in high abundance in the sample (qualitative estimates).

+ Species names are currently accepted taxonomic entities according to Guiry and Guiry [18] and higher taxonomic ranks are according to the most recent system of classification for cyanobacteria, proposed by Komárek *et al.* [19] (see also Material and Methods section; main text). Older synonyms used in the consulted references are also shown (in brackets)

& Phylotypes or operational taxonomic units (OTUs) identified/characterized by 16S rRNA gene sequencing.

% Organism (also) detected by culture-dependent methodology. See Table 2.10 for a list of isolates obtained in this study and Table 2.9 and Fig. 2.18 for their characterization.

\* l, laminites; lay, layered organic (i.e. non-lithifying) mat; non, organic (i.e. non-lithifying) mat; s, stromatolites; t, thrombolites

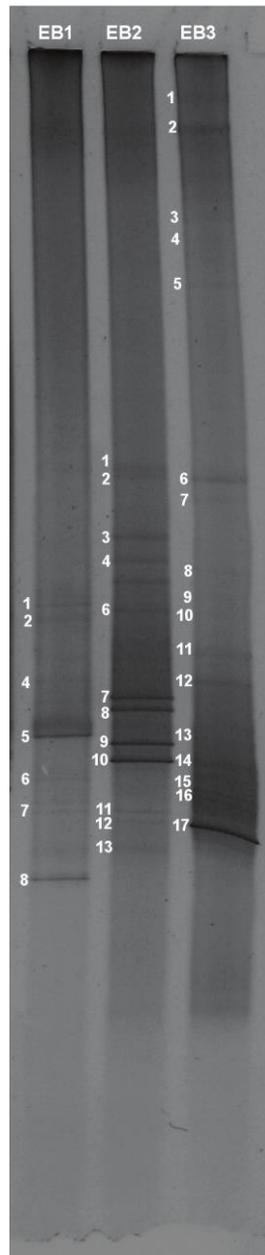
§ b, blistered; coll, colloform; colu, columnar; d, domal; f, flat; po, polygonal; pu, pustular; sm, smooth; st, stratiform; t, tufted

£ Abbreviations for the lagoons: **ARA**, Araruama; AZU, Azul; BES: Brejo do Espinho; BPF, Brejo do Pau Fincado; **PER**, Pernambuco; **PIT**, Pitanguinha; SAJ, Salina Julieta; VER, Vermelha. Studied water bodies (EB1 is located in **ARA**; EB2 → **PIT**; EB3 → **PER**) are highlighted in bold if the organism was observed in samples from this work and are underlined if the observation corresponds to a first report (see Table 2.9).

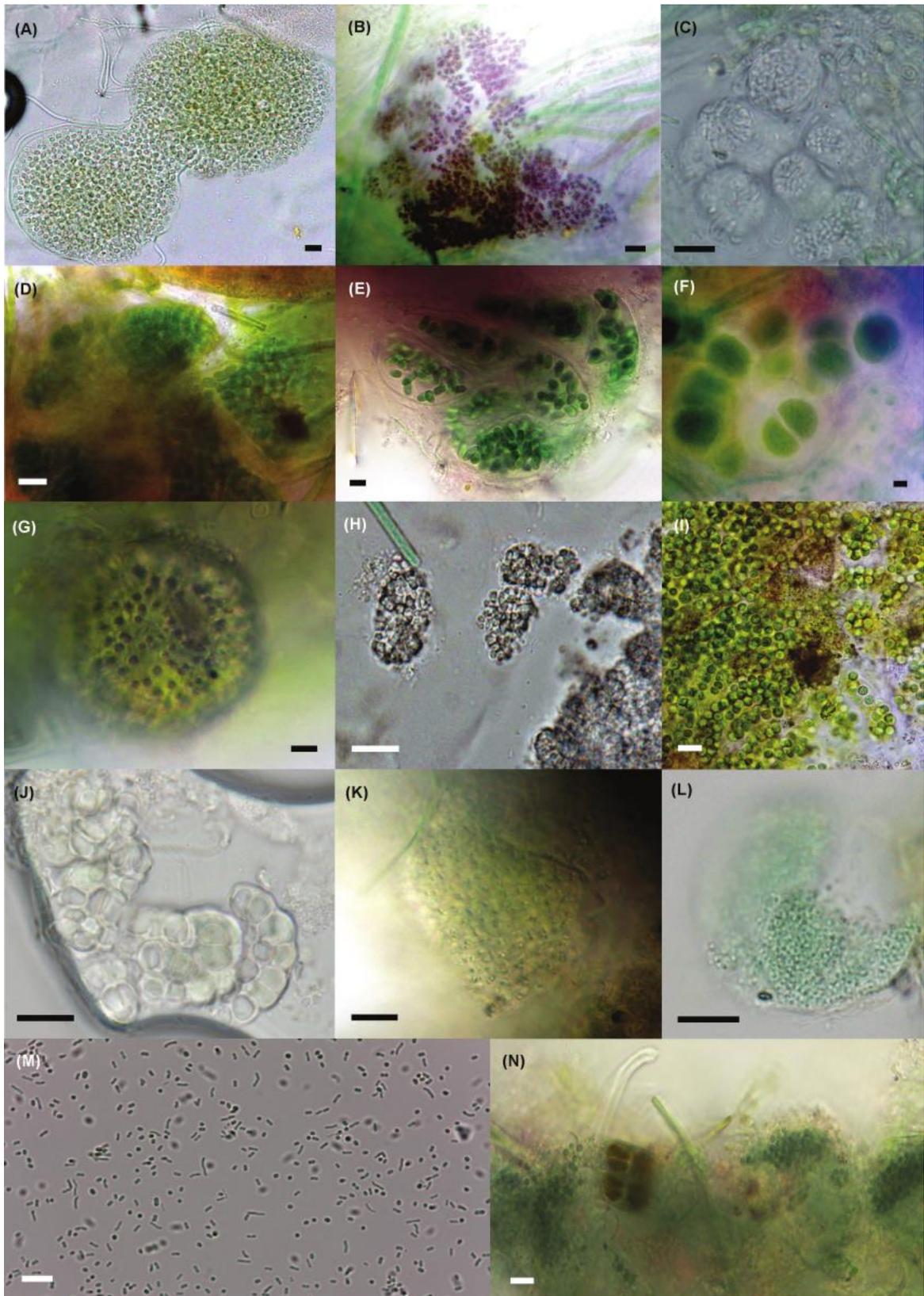
@ References: **1.** Damazio, C.M. *et al.*, 2005. *Geociências* 10 (6): 11-16. [in Portuguese], **2.** Silva, L.H.S. *et al.*, 2004. *Revista Brasileira de Paleontologia* 7(2): 189-192. [in Portuguese], **3.** Iespa, A.A.C. *et al.*, 2009. *Journal of Geoscience* 5(1): 35-41., **4.** Iespa, A.A.C. *et al.*, 2005. *Geociências* 10 (6): 5-10. [in Portuguese], **5.** Silva, L.H.S. *et al.*, 2007. *Anuário do Instituto de Geociências* 30(1): 188-193. [in Portuguese], **6.** Silva, L.H.S. *et al.*, 2011. *Anuário do Instituto de Geociências* 34(2) 14-23. [in Portuguese], **7.** Silva, L.H.S. *et al.*, 2007. *Revista Biociências* 13(1-2): 63-70. [in Portuguese], **8.** Silva, L.H.S. *et al.*, 2007. *Anuário do Instituto de Geociências* 30(1): 175-180. [in Portuguese], **9.** Silva, L.H.S. *et al.*, 2008. *Anuário do Instituto de Geociências* 31(1): 24-29. [in Portuguese], **10.** Damazio *et al.*, 2006. *Revista Brasileira de Paleontologia* 9(1):165-170. [in Portuguese], **11.** Silva, L.H.S. *et al.*, 2007 *Anuário do Instituto de Geociências* 30(1): 67-72. [in Portuguese], **12.** Silva, L.H.S. *et al.*, 2006 *Revista de Biologia e Ciências da Terra* 6(2): 243-250. [in Portuguese], **13.** Clementino, M.M. *et al.*, 2008 *Extremophiles* 12: 595-604., **14.** Silva, L.H.S. *et al.*, 2007 *Anuário do Instituto de Geociências* 30(1): 181-187. [in Portuguese], **15.** Silva, L.H.S. *et al.*, 2005 *Anuário do Instituto de Geociências* 28(1): 92-100. [in Portuguese], **16.** Vasconcelos, C. *et al.*, 2006 *Sedimentary Geology* 185: 175-183., **17.** Feder, F. *et al.*, 2013. *Advances in Microbiology* 3: 47-54., **18.** Guiry, M.D. and Guiry, G.M. 2016. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; searched on 18 March 2016. **19.** Komárek, J. *et al.*, 2014. *Preslia*, 86: 295-335.

**Table 2.14** Percentage of pyrosequencing reads, by sample, included in each phylotype identified in the phylogenetic tree. Only clusters that include sequences derived from this study (i.e. obtained from 454-pyrosequencing, isolates or DGGE bands) were considered to distinguish phylotypes (see also Appendix A2). Phylotypes with cells highlighted in blue indicate presence or absence of 454-read sequences in all three samples. Dark gray cells highlight the higher value for each sample. Light gray cells highlight values >4%.

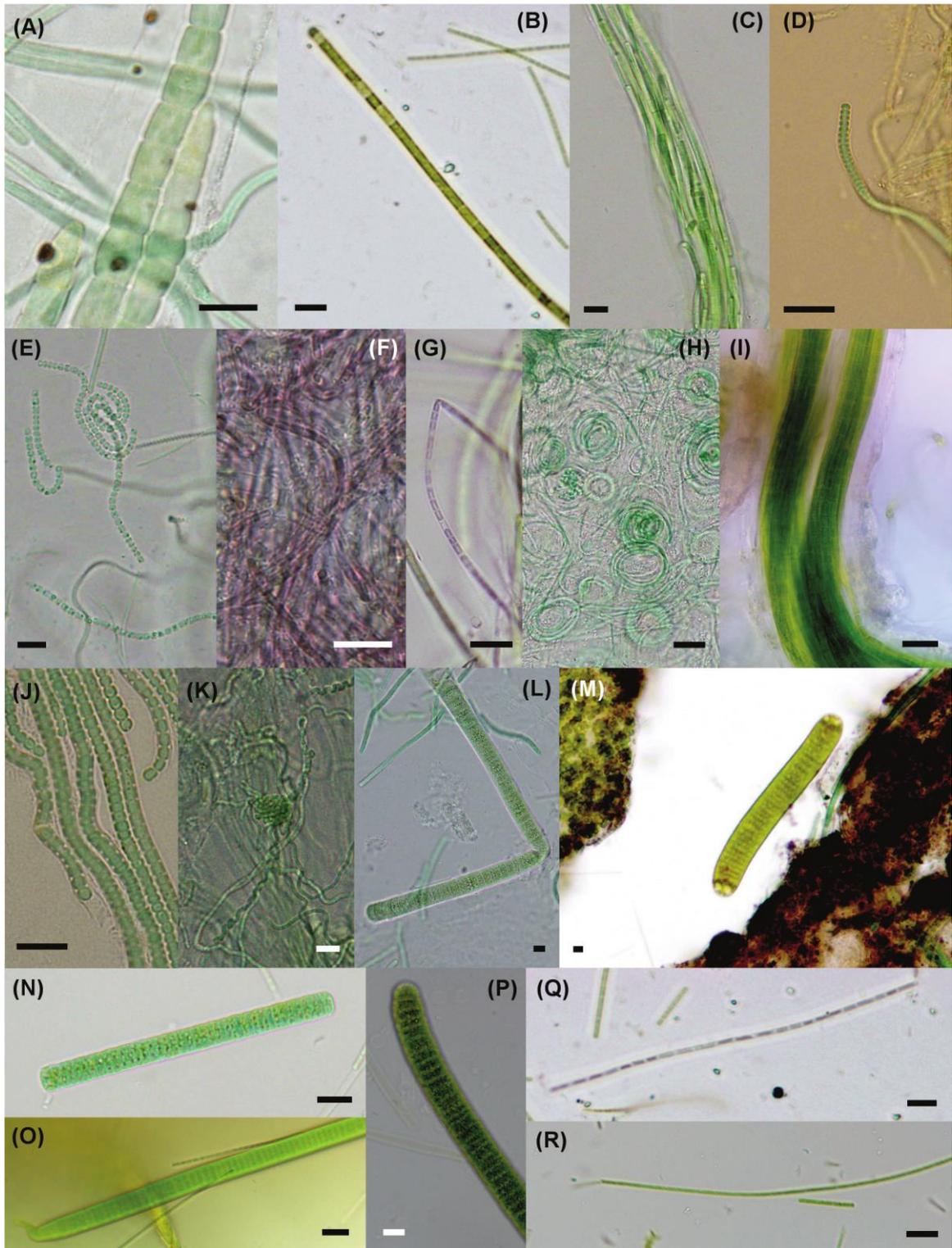
| Phylotypes        | EB1    | EB2    | EB3    |
|-------------------|--------|--------|--------|
| A                 | -      | 84,65% | 3,08%  |
| B                 | 13,18% | -      | -      |
| C                 | 22,86% | -      | -      |
| D                 | -      | 0,09%  | 0,07%  |
| E                 | -      | -      | -      |
| F                 | -      | 5,27%  | -      |
| G                 | -      | 2,18%  | -      |
| H                 | 0,12%  | -      | -      |
| I                 | 0,31%  | -      | -      |
| J                 | -      | -      | 90,03% |
| K                 | 0,25%  | -      | -      |
| L                 | 1,59%  | 0,34%  | 1,43%  |
| M                 | 3,25%  | -      | -      |
| N                 | 1,78%  | -      | -      |
| O                 | -      | -      | 0,14%  |
| P                 | 1,00%  | 0,25%  | -      |
| Q                 | 0,56%  | 1,53%  | -      |
| R                 | 0,47%  | 0,12%  | -      |
| S                 | 0,31%  | -      | -      |
| T                 | 1,09%  | -      | -      |
| U                 | 1,56%  | 0,06%  | -      |
| V                 | 5,03%  | 0,18%  | -      |
| W                 | 1,25%  | 0,06%  | 1,26%  |
| X                 | 3,03%  | -      | -      |
| Y                 | 0,94%  | -      | -      |
| Z                 | 0,12%  | -      | -      |
| AA                | 0,12%  | -      | -      |
| AB                | 0,09%  | -      | -      |
| AC                | 7,93%  | 0,18%  |        |
| AD                | -      | -      | 0,25%  |
| AE                | 10,87% | 0,12%  | 0,07%  |
| AF                | 0,09%  | -      | -      |
| AG                | 1,59%  | -      | -      |
| AH                | 0,19%  | 0,06%  | -      |
| AI                | 0,19%  | -      | -      |
| AJ                | 0,09%  | -      | -      |
| AL                | 3,72%  | -      | -      |
| AM                | 12,99% | 2,14%  | 1,69%  |
| <b>loner OTUs</b> |        |        |        |
| EB1_33_V3Z02DFMNB | 2,34%  | -      | -      |
| EB1_39_V3Z02ERQP0 | 0,22%  | -      | -      |
| EB1_77_V3Z02D1AW0 | 0,19%  | -      | -      |
| EB1_75_V3Z02DA6N6 | 0,16%  | -      | -      |
| EB1_70_V3Z02DPWRC | 0,12%  | -      | -      |
| EB2_2_V3Z04J04BU  | -      | 0,09%  | -      |
| EB2_15_V3Z04I8TO7 | -      | 0,09%  | -      |



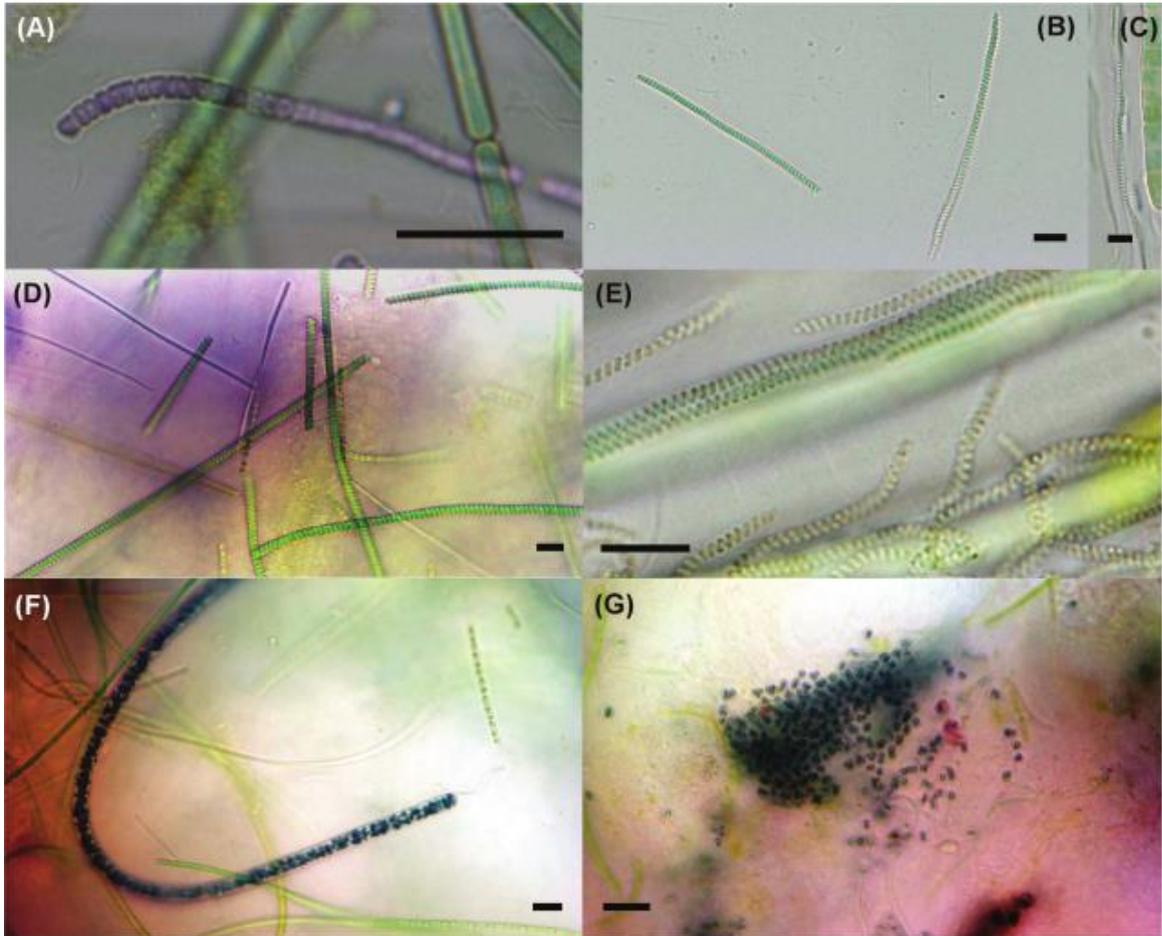
**Figure 2.21.** Polyacrylamide gel image showing the DGGE bands excised for cloning and sequencing.



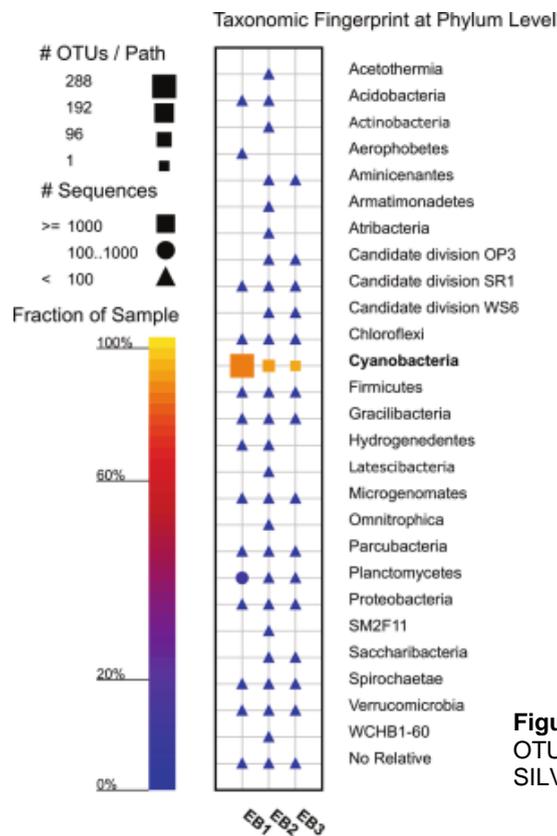
**Figure 2.22.** Light micrographs showing unicellular and colonial morphospecies from the cyanobacterial mat layers. (A) *Aphanocapsa littoralis*, (B) *Aphanocapsa* cf. *salina*, (C) *Aphanothece* cf. *conglomerata*, (D) *Aphanothece* aff. *salina*, (E) *Aphanothece* cf. *stagnina*, (F) *Chroococcus* aff. *turgidus*, (G) *Coelosphaeriopsis* cf. *halophila*, (H) *Cyanosarcina* aff. *thalassia*, (I) *Geminocystis* sp., (J) *Gloeocapsopsis crepidinum*, (K) *Gloeothece* cf. *subtilis*, (L) *Lemmermanniella* sp., (M) *Synechococcus* sp., (N) *Synechocystis salina*. Scale bar: 10  $\mu$ m.



**Figure 2.23.** Light micrographs showing filamentous non-heterocystous morphospecies. (A) *Coleofasciculus chthonoplastes*, (B) *Geitlerinema* aff. *amphibium*, (C) *Geitlerinema* cf. *lemmermannii*, (D) *Komvophoron breve*, (E) *Komvophoron* cf. *minutum*, (F) *Leptolyngbya crosbyana*, (G) *Leptolyngbya* cf. *ectocarpi*, (H) *Halomiconema excentricum*, (I) *Microcoleus* aff. *steenstrupii*, (J) *Nodosilinea* sp., (K) *Nodosilinea nodulosa*, (L) *Oscillatoria limosa*, (M) *Oscillatoria margaritifera*, (N) *Oscillatoria subbrevis*, (O) *Oxynema lloydianum*, (P) *Phormidium nigroviride*, (Q) *Pseudanabaena* aff. *amphigranulata*, (R) *Pseudanabaena* cf. *limnetica*, (S) *Pseudanabaena* cf. *minima*, (T) *Spirulina labyrinthiformis*. Scale bar: 10  $\mu$ m.

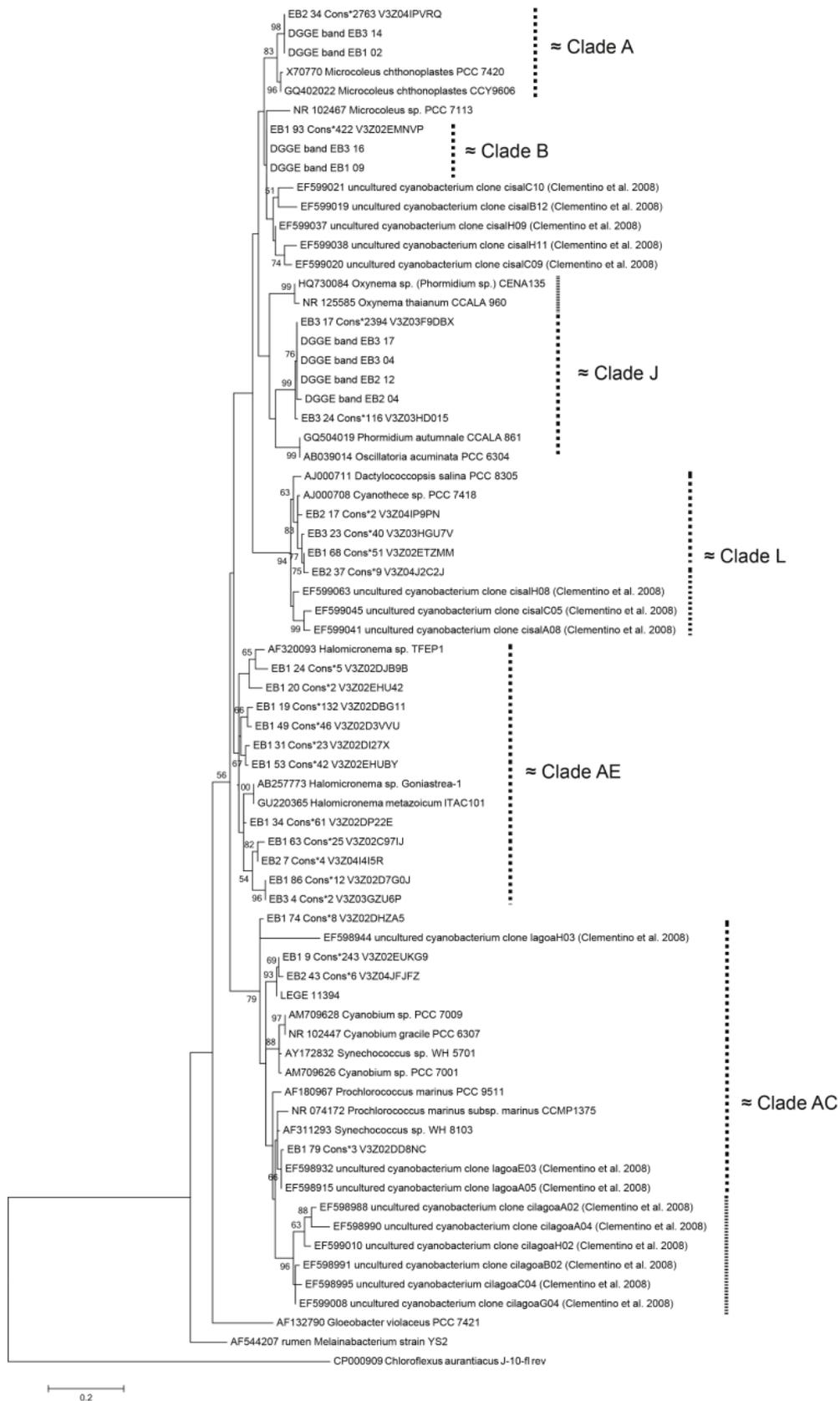


**Figure 2.24.** Light micrographs showing filamentous non-heterocystous Cyanobacteria (A-E) and Proteobacteria (F-G). (A) *Pseudanabaena* cf. *minima*, (B) *Spirulina labyrinthiformis*, (C) *S. labyrinthiformis* embedded within the sheath of *Coleofasciculus chthonoplastes*, (D) *Spirulina subsalsa*, (E) *Spirulina tenerrima*, (F) the motile, colorless, filamentous sulfur bacteria *Beggiatoa* sp., (G) a species of purple sulfur bacteria from the family Chromatiaceae. Scale bar: 10 µm.

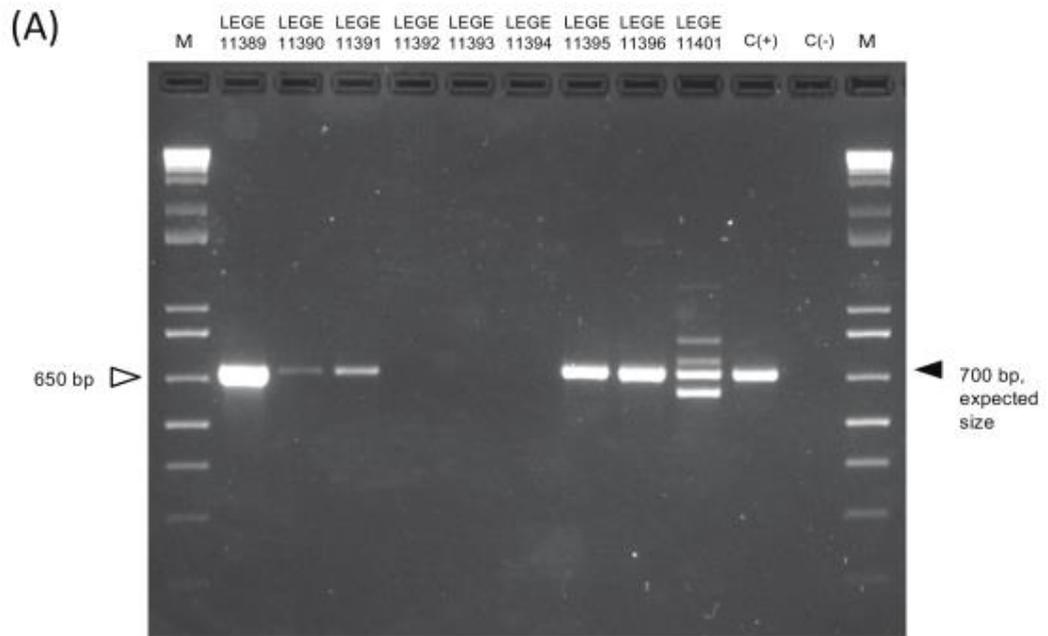


**Figure 2.25.** Number and relative abundance of OTUs (clustered at 98% similarity) generated by the SILVAngs pipeline, at the phylum level.





**Figure 2.27.** Complementary ML tree based on a selection of cyanobacterial 16S rDNA sequences from this study (with indication of the respective clade; see supplementary image via Appendix A2), from the study of Clementino et al. (2008) and from *Oxytnema* spp. (Chatchawan et al. 2012). The analysis involved 71 nucleotide sequences. Values of bootstrap support >50% are shown near the nodes. Due to sequence shortness and/or low overlapping, fewer than 6% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 203 positions in the final dataset.



**Legend:**  
 M – size marker DNA (1-kb Plus DNA Ladder; Invitrogen)  
 C(+) – positive control (gDNA from *Microcystis aeruginosa* LEGE 00063)  
 C(-) – negative control

(B)

| Query sequence | Top BLASTn hit   |          |                  |
|----------------|--|----------|------------------|
|                | Sequence   | Identity | Accession number |
| LEGE 11389     | <i>Nostoc</i> sp. ATCC 53789 isolate 5 polyketide synthase ketosynthase domain gene, partial cds                               | 72%      | AY695332.1       |
| LEGE 11390     | <i>Calothrix</i> sp. LEGE 07177 clone 1 polyketide synthase gene, partial cds  | 69%      | KC842345.1       |
| LEGE 11391     | <i>Calothrix</i> sp. LEGE 07177 clone 1 polyketide synthase gene, partial cds  | 69%      | KC842345.1       |
| LEGE 11395     | Unidentified microorganism clone 3 type I polyketide synthase-like gene, partial cds   | 78%      | AY210792.1       |
| LEGE 11396     | <i>Calothrix</i> sp. LEGE 07177 clone 1 polyketide synthase gene, partial cds  | 70%      | KC842345.1       |
| LEGE 11401     | <i>Pseudanabaena</i> cf. <i>curta</i> LEGE 07160 clone 1 polyketide synthase gene, partial cds                                 | 76%      | KC842368.1       |
| LEGE 00063     | <i>Microcystis aeruginosa</i> mcyD, mcyE, mcyF, ..., polyketide synthase and peptide synthetase, ..., complete and partial cds | 94%      | AB032549.2       |

**Figure 2.28.** Agarose gel electrophoresis image of PCR products for PKS genes (A), and BLAST results for the obtained sequences (B).

### 2.3.8 References

- ABED, R., GARCIA-PICHEL, F. & HERNÁNDEZ-MARINÉ, M. 2002. Polyphasic characterization of benthic, moderately halophilic, moderately thermophilic cyanobacteria with very thin trichomes and the proposal of *Halomicronema excentricum* gen. nov., sp. nov. *Archives of Microbiology*, 177, 361-370.
- BAUERSACHS, T., COMPAORÉ, J., SEVERIN, I., HOPMANS, E. C., SCHOUTEN, S., STAL, L. J. & SINNINGHE DAMSTÉ, J. S. 2011. Diazotrophic microbial community of coastal microbial mats of the southern North Sea. *Geobiology*, 9, 349-359.
- BERRY, D., BEN MAHFOUDH, K., WAGNER, M. & LOY, A. 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, 77, 7846-7849.
- BRAGG, L. & TYSON, G. W. 2014. Metagenomics using next-generation sequencing. *Methods in Molecular Biology*, 1096, 183-201.
- BRITO, Â., GAIFEM, J., RAMOS, V., GLUKHOV, E., DORRESTEIN, P. C., GERWICK, W. H., VASCONCELOS, V. M., MENDES, M. V. & TAMAGNINI, P. 2015. Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*, 9, 218-226.
- BRITO, Â., RAMOS, V., SEABRA, R., SANTOS, A., SANTOS, C. L., LOPO, M., FERREIRA, S., MARTINS, A., MOTA, R., FRAZÃO, B., MARTINS, R., VASCONCELOS, V. & TAMAGNINI, P. 2012. Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Systematic and Applied Microbiology*, 35, 110-119.
- CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421.
- CASTENHOLZ, R. W., WILMOTTE, A., HERDMAN, M., RIPPKA, R., WATERBURY, J. B., ITEMAN, I. & HOFFMANN, L. 2001. Phylum BX. Cyanobacteria. In *Bergey's Manual of Systematic Bacteriology: Volume One*. D.R. Boone, R.W. Castenholz & G.M. Garrity, eds. Springer: New York, USA, pp. 473-599.
- CHATCHAWAN, T., KOMÁREK, J., STRUNECKÝ, O., ŠMARDA, J. & PEERAPORNPIHAL, Y. 2012. *Oxynema*, a new genus separated from the genus *Phormidium* (Cyanophyta). *Cryptogamie, Algologie*, 33, 41-59.
- CLEMENTINO, M. M., VIEIRA, R. P., CARDOSO, A. M., NASCIMENTO, A. P. A., SILVEIRA, C. B., RIVA, T. C., GONZALEZ, A. S. M., PARANHOS, R., ALBANO, R. M., VENTOSA, A. & MARTINS, O. B. 2008. Prokaryotic diversity in one of the largest hypersaline coastal lagoons in the world. *Extremophiles*, 12, 595-604.
- COHAN, F. M. & PERRY, E. B. 2007. A systematics for discovering the fundamental units of bacterial diversity. *Current Biology*, 17, R373-R386.
- DAMAZIO, C. & SILVA, L. H. 2006. Cianobactérias em esteiras microbianas coliformes da lagoa Pitanguinha, Rio de Janeiro, Brasil. *Revista Brasileira de Paleontologia*, 9, 165-170.
- DECHO, A. W. 2010. Overview of biopolymer-induced mineralization: What goes on in biofilms? *Ecological Engineering*, 36, 137-144.
- DIEZ, B., BAUER, K. & BERGMAN, B. 2007. Epilithic cyanobacterial communities of a marine tropical beach rock (Heron Island, Great Barrier Reef): diversity and diazotrophy. *Applied and Environmental Microbiology*, 73, 3656-3668.
- DUPRAZ, C., REID, R. P., BRAISSANT, O., DECHO, A. W., NORMAN, R. S. & VISSCHER, P. T. 2009. Processes of carbonate precipitation in modern microbial mats. *Earth-Science Reviews*, 96, 141-162.
- DVOŘÁK, P., POULÍČKOVÁ, A., HAŠLER, P., BELLI, M., CASAMATTA, D. A. & PAPINI, A. 2015. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. *Biodiversity and Conservation*, 24, 739-757.
- FEDERHEN, S. 2011. The NCBI Taxonomy database. *Nucleic Acids Research*, 40, D136-D143.
- FRANKS, J. & STOLZ, J. F. 2009. Flat laminated microbial mat communities. *Earth-Science Reviews*, 96, 163-172.
- GARCIA-ETXEARRIA, K., GARCIA-GARCERÀ, M. & CALAFELL, F. 2014. Consistency of metagenomic assignment programs in simulated and real data. *BMC Bioinformatics*, 15, 90.
- GARCIA-PICHEL, F., PRUFERT-BEBOUT, L. & MUYZER, G. 1996. Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium. *Applied and Environmental Microbiology*, 62, 3284-3291.
- GREEN, S. J., LEIGH, M. B. & NEUFELD, J. D. 2010. Denaturing gradient gel electrophoresis (DGGE) for microbial community analysis. In *Handbook of Hydrocarbon and Lipid Microbiology*, K.N. Timmis, ed. Springer-Verlag: Berlin, Germany, pp.4137-4158.

- GUIRY, M. D. & GUIRY, G. M. 2016. *AlgaeBase*. World-wide Electronic Publication, National University of Ireland, Galway. Available online at: <http://www.algaebase.org> (Accessed up to 2 July 2016).
- HARRIS, J. K., CAPORASO, J. G., WALKER, J. J., SPEAR, J. R., GOLD, N. J., ROBERTSON, C. E., HUGENHOLTZ, P., GOODRICH, J., MCDONALD, D., KNIGHTS, D., MARSHALL, P., TUFO, H., KNIGHT, R. & PACE, N. R. 2013. Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *The ISME Journal*, 7, 50-60.
- IESPA, A.A.C. & SILVA, L.H. (2005). Laminitos microbianos e cianobactérias na lagoa Pernambuco, Neogeno do Rio de Janeiro, Brasil. *Geociências* 10, 5-10.
- JUNGBLUT, A.-D. & NEILAN, B. A. 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. *Archives of Microbiology*, 185, 107-114.
- KIM, M., OH, H. S., PARK, S. C. & CHUN, J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 64, 346-351.
- KJERFVE, B., SCHETTINI, C. A. F., KNOPPERS, B., LESSA, G. & FERREIRA, H. O. 1996. Hydrology and salt balance in a large, hypersaline coastal lagoon: Lagoa de Araruama, Brazil. *Estuarine, Coastal and Shelf Science*, 42, 701-725.
- KOEPPEL, A., PERRY, E. B., SIKORSKI, J., KRIZANC, D., WARNER, A., WARD, D. M., ROONEY, A. P., BRAMBILLA, E., CONNOR, N., RATCLIFF, R. M., NEVO, E. & COHAN, F. M. 2008. Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences*, 105, 2504-2509.
- KOEPPEL, A. F. & WU, M. 2014. Species matter: the role of competition in the assembly of congeneric bacteria. *The ISME Journal*, 8, 531-540.
- KOMÁREK, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, 51, 346-353.
- KOMÁREK, J., & ANAGNOSTIDIS, K. 1998. Cyanoprokaryota 1. Teil: Chroococcales. In *Süßwasserflora von Mitteleuropa 19/1*, Ettl, H., Gärtner, G., Heynig, H. & Mollenhauer, D. (eds.), Gustav Fischer Verlag. p. 548.
- KOMÁREK, J., & ANAGNOSTIDIS, K. 2005. Cyanoprokaryota 2. Teil: Oscillatoriales. In *Süßwasserflora von Mitteleuropa 19/2*, Büdel, B., Krienitz, L., Gärtner, G. & Schagerl, M. (eds.), Elsevier/Spektrum. p. 759.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- KRAUSE, S., LE ROUX, X., NIKLAUS, P. A., VAN BODEGOM, P. M., LENNON, J. T., BERTILSSON, S., GROSSART, H.-P., PHILIPPOT, L. & BODELIER, P. L. E. 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology*, 5.
- LAGIER, J.-C., HUGON, P., KHELAIPIA, S., FOURNIER, P.-E., LA SCOLA, B. & RAOULT, D. 2015. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews*, 28, 237-264.
- LEY, R. E., HARRIS, J. K., WILCOX, J., SPEAR, J. R., MILLER, S. R., BEBOUT, B. M., MARESCA, J. A., BRYANT, D. A., SOGIN, M. L. & PACE, N. R. 2006. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Applied and Environmental Microbiology*, 72, 3685-3695.
- LOPES, V. R., RAMOS, V., MARTINS, A., SOUSA, M., WELKER, M., ANTUNES, A. & VASCONCELOS, V. M. 2012. Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries. *Marine Environmental Research*, 73, 7-16.
- MCDONALD, D., PRICE, M. N., GOODRICH, J., NAWROCKI, E. P., DESANTIS, T. Z., PROBST, A., ANDERSEN, G. L., KNIGHT, R. & HUGENHOLTZ, P. 2011. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6, 610-618.
- MIZRAHI-MAN, O., DAVENPORT, E. R. & GILAD, Y. 2013. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE*, 8, e53608.
- MOBBERLEY, J. M., ORTEGA, M. C. & FOSTER, J. S. 2011. Comparative microbial diversity analyses of modern marine thrombolitic mats by barcoded pyrosequencing. *Environmental Microbiology*, 14, 82-100.
- MOFFITT, M. C. & NEILAN, B. A. 2001. On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus *Nodularia*. *FEMS Microbiology Letters*, 196, 207-214.
- MORRIS, E. K., CARUSO, T., BUSCOT, F., FISCHER, M., HANCOCK, C., MAIER, T. S., MEINERS, T., MÜLLER, C., OBERMAIER, E., PRATI, D., SOCHER, S. A., SONNEMANN, I., WÄSCHKE, N., WUBET, T., WURST, S. & RILLIG, M. C. 2014. Choosing and using diversity indices: insights for

- ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, 4, 3514-3524.
- MÜHLING, M., WOOLVEN-ALLEN, J., MURRELL, J. C. & JOINT, I. 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *The ISME Journal*, 2, 379-392.
- NEILSON, J. W., JORDAN, F. L. & MAIER, R. M. 2013. Analysis of artifacts suggests DGGE should not be used for quantitative diversity analysis. *Journal of Microbiological Methods*, 92, 256-263.
- NGUYEN, N.-P., WARNOW, T., POP, M. & WHITE, B. 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms and Microbiomes*, 2, 16004.
- NÜBEL, U., GARCIA-PICHEL, F. & MUYZER, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, 63, 3327-3332.
- ONDOV, B. D., BERGMAN, N. H. & PHILLIPPY, A. M. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12, 385.
- OREN, A. 2012. Salts and Brines. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, A.B. Whitton, ed. Springer: Dordrecht, The Netherlands, pp. 401-426.
- PINTO, C., PINHO, D., SOUSA, S., PINHEIRO, M., EGAS, C. & C. GOMES, A. 2014. Unravelling the diversity of grapevine microbiome. *PLoS ONE*, 9, e85622.
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLOCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590-D596.
- RAMOS, V., SEABRA, R., BRITO, Â., SANTOS, A., SANTOS, C. L., LOPO, M., MORADAS-FERREIRA, P., VASCONCELOS, V. M. & TAMAGNINI, P. 2010. Characterization of an intertidal cyanobacterium that constitutes a separate clade together with thermophilic strains. *European Journal of Phycology*, 45, 394-403.
- RANTALA-YLINEN, A., KANA, S., WANG, H., ROUHIAINEN, L., WAHLSTEN, M., RIZZI, E., BERG, K., GUGGER, M. & SIVONEN, K. 2011. Anatoxin-a synthetase gene cluster of the cyanobacterium *Anabaena* sp. strain 37 and molecular methods to detect potential producers. *Applied and Environmental Microbiology*, 77, 7271-7278.
- RIPPKA, R. 1988. Isolation and purification of cyanobacteria. *Methods in Enzymology*, 167, 3-27.
- SÁNCHEZ, O., GASOL, J. M., BALAGUÉ, V., MASSANA, R., MAS, J. & PEDRÓS-ALIÓ, C. 2009. Influence of primer mismatch and microdiversity on DGGE results: a case study with SAR11. *Aquatic Microbial Ecology*, 54, 211-216.
- SCHMIDT, T. S. B., MATIAS RODRIGUES, J. F. & VON MERING, C. 2015. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environmental Microbiology*, 17, 1689-1706.
- SCHOLZ, M. B., LO, C.-C. & CHAIN, P. S. G. 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology*, 23, 9-15.
- SIEGESMUND, M. A., JOHANSEN, J. R., KARSTEN, U. & FRIEDL, T. 2008. *Coleofasciculus* gen. nov. (Cyanobacteria): morphological and molecular criteria for revision of the genus *Microcoleus* Gomont. *Journal of Phycology*, 44, 1572-1585.
- SILVA, L. H., ALVES, S. A., MAGINA, F. C., GOMES, S. B., DELFINO, D. & LOPES, F. A. 2011. Determinação das cianobactérias de subsuperfície de sedimentos marginais do Brejo do Pau Fincado do complexo lagunar de Araruama, RJ, Brasil. *Anuário do Instituto de Geociências*, 34, 14-23.
- SILVA, L.H., FEDER, F., DELFINO, D.O. & LOPES, F.A.S. 2007a. Análise da composição cianobacteriana das esteiras pustulares em salina, Araruama, Rio de Janeiro. *Anuário do Instituto de Geociências*, 31, 169-174.
- SILVA, L.H., SANTOS, R.C., IESPA, A.A.C., DAMAZIO, C.M., DELFINO, D.O. & LOPES, F.A.S. 2007b. Cianobactérias planctônicas da lagoa Pitanguinha, RJ, Brasil. *Revista de Biociências*, 13, 63-70.
- SILVA, L.H., IESPA, A.A.C. & DAMAZIO, C.M. 2006. Trombólitos e cianobactérias da lagoa Pernambuco, Holoceno do Rio de Janeiro, Brasil. *Revista de Biologia e Ciências da Terra*, 6, 243-250.
- SILVA, L.H., DAMAZIO, C.M. & IESPA, A.A.C. 2005. Identificação de cianobactérias em sedimentos da lagoa Pitanguinha, Estado do Rio de Janeiro, Brasil. *Anuário do Instituto de Geociências*, 28, 92-100.
- SPEKSNIJDER, A. G. C. L., KOWALCHUK, G. A., DE JONG, S., KLINE, E., STEPHEN, J. R. & LAANBROEK, H. J. 2001. Microvariation artifacts introduced by pcr and cloning of closely related 16S rRNA gene sequences. *Applied and Environmental Microbiology*, 67, 469-472.
- STAL, L. J. 2012. Cyanobacterial mats and stromatolites. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, A.B. Whitton, ed. Springer: Dordrecht, The Netherlands, pp. 65-125.
- TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.

- TUZHNIKOV, A., PANCHIN, A. & SHESTOPALOV, V. I. 2014. TUIT, a BLAST-based tool for taxonomic classification of nucleotide sequences. *BioTechniques*, 56.
- VON WINTZINGERODE, F., GÖBEL, U. B. & STACKEBRANDT, E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, 21, 213-229.
- VISSCHER, P.T., PRINS, R.A., & VAN GEMERDEN, H. 1992. Rates of sulfate reduction and thiosulfate consumption in a marine microbial mat. *FEMS Microbiology Letters*, 86, 283-288.
- WANG, Q., GARRITY, G. M., TIEDJE, J. M. & COLE, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261-5267.
- WARD, D. M., BATESON, M. M., FERRIS, M. J., KUHL, M., WIELAND, A., KOEPEL, A. & COHAN, F. M. 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361, 1997-2008.
- WATERBURY, J. B. 2006. The Cyanobacteria - isolation, purification and identification. In *The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria*. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt, eds. Springer: New York, USA, pp. 1053-1073.
- WILMOTTE, A. & HERDMAN, M. 2001. Phylogenetic relationships among the cyanobacteria based on 16S rRNA sequences. In *Bergey's Manual of Systematics of Archaea and Bacteria: Volume one*. D.R. Boone, R.W. Castenholz & G.M. Garrity, eds. Springer: New York, USA, pp. 487-493.
- WRIGHT, E. S., YILMAZ, L. S. & NOGUERA, D. R. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Applied and Environmental Microbiology*, 78, 717-725.
- YARZA, P., YILMAZ, P., PRUESSE, E., GLÖCKNER, F. O., LUDWIG, W., SCHLEIFER, K.-H., WHITMAN, W. B., EUZÉBY, J., AMANN, R. & ROSSELLÓ-MÓRA, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, 12, 635-645.
- YOPP, J. H., TINDALL, D. R., MILLER, D. M. & SCHMID, W. E. 1978. Isolation, purification and evidence for a halophilic nature of the blue-green alga *Aphanothece halophytica* Frey (Chroococcales). *Phycologia*, 17, 172-178.
- YOUNGBLUT, N. D., SHADE, A., READ, J. S., MCMAHON, K. D. & WHITAKER, R. J. 2013. Lineage-specific responses of microbial communities to environmental change. *Applied and Environmental Microbiology*, 79, 39-47.

## 3. Monitoring cyanobacteria and their environmental disservices

### 3.1 Application of molecular tools in monitoring cyanobacteria and their potential toxin production

Vitor Ramos<sup>1,2</sup>, Cristiana Moreira<sup>2</sup>, Joanna Mankiewicz-Boczek<sup>3,4</sup>, Vitor Vasconcelos<sup>1,2\*</sup>

<sup>1</sup> Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

<sup>2</sup> Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Terminal de Cruzeiros do Porto de Leixões, 4450-208 Matosinhos, Portugal

<sup>3</sup> European Regional Centre for Ecohydrology of the Polish Academy of Sciences, 3 Tylna, 90-364 Łódź, Poland

<sup>4</sup> Department of Applied Ecology, Faculty of Biology and Environmental Protection, University of Lodz, Banacha Str. 12/16, 90-237, Łódź, Poland

\*Corresponding author: Vitor Vasconcelos, email: [vmvascon@fc.up.pt](mailto:vmvascon@fc.up.pt)

#### 3.1.1 Abstract

The discovery and annotation of biosynthetic gene clusters that encode for cyanotoxins allowed the development and implementation of different molecular-based approaches for the monitoring of potentially hazardous cyanobacteria. Here we present an overview on the usage of different molecular tools in the detection of these microorganisms and their toxins in the environment, while discussing practical and methodological issues related to it. Due to its specificity, a special focus is given to food supplement surveillance, including the development of Standard Operation Procedures (SOPs).

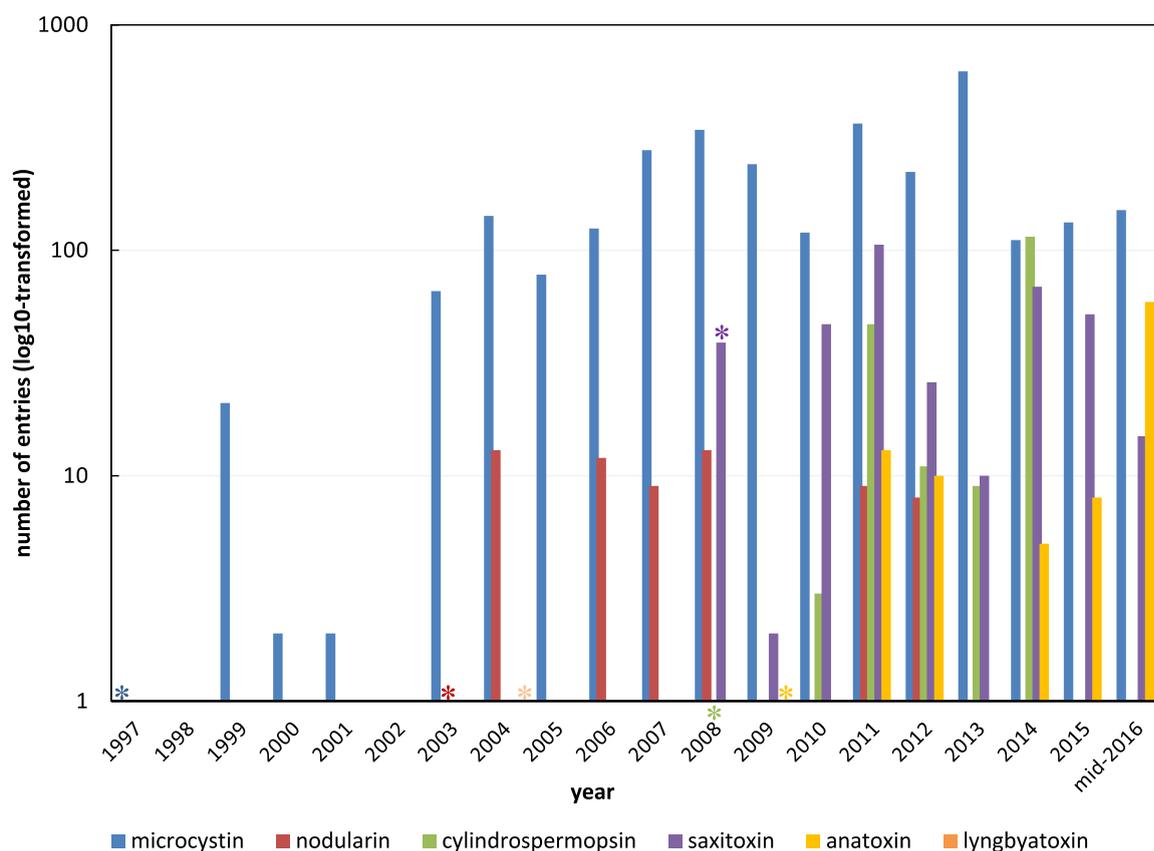
#### Keywords

cyanotoxins; toxigenicity assessment; review; molecular markers; environmental monitoring; water quality; food supplements; human health risk

#### 3.1.2 Introduction

After the advent of PCR-based methods, soon began the development of molecular tools for the study and monitoring toxigenic cyanobacteria in the environment. In fact, these methods have been applied for that purpose for almost two decades, and still increasingly being as it can be easily deduced (albeit indirectly) from Fig. 3.1. Before the genetics behind the biosynthesis of

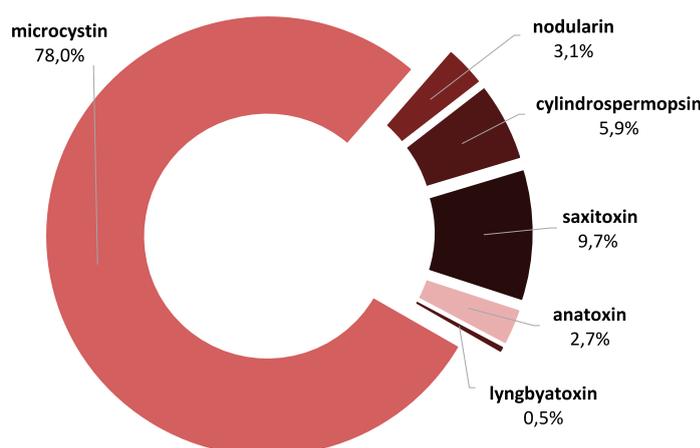
the different cyanotoxins started to be unravelled, a common molecular-based approach for the detection of putative toxic cyanobacteria in environmental samples consisted in performing DNA



**Figure 3.1.** Evolution of the number of new sequences, deposited in the GenBank nucleotide database, for genes involved in the biosynthesis of cyanotoxins. Example of a Boolean search string used: (((microcystin[Title]) OR mcy\*[Title]) AND cyanobacter\*[Organism]) AND ("1997"[Publication Date] : "1997"[Publication Date]). Asterisks (\*) refer to the release date of the first annotated gene sequence for each type of toxin (see text to an accession number and publication correspondence).

fingerprinting analyses to discriminate between species or strains/genotypes. The genetic markers used were not related (or correlated) with toxin synthesis, and thus obtain a clear and undeniable relationship between those fingerprint profiles and toxicity were difficult or not feasible at all (e.g. Nishihara et al. 1997; Neilan et al. 2003). Other approach is to use phylogenetically useful genes, also known as taxonomic markers, to detect and identify cyanobacterial taxa. These markers, very particularly the 16S rRNA gene, are still being extensively used. Not surprisingly, when cyanotoxin gene clusters are discovered and characterized, and their annotated sequences are made publicly available, the genes involved in the biosynthetic pathway of the different cyanotoxins rapidly become the preferred target region for the development of molecular tools to study and monitoring toxigenic cyanobacteria. As can be seen in Fig. 3.1, the microcystin gene cluster was the first to be identified and (partially) sequenced (Dittmann et al. 1997; GenBank sequence accession number, U97078.1); then, were the gene clusters of nodularin (Moffitt & Neilan 2004; AY210783.2), lyngbyatoxin (Edwards & Gerwick 2004; AY588942.1), cylindrospermopsin (Mihali et al. 2008; EU140798.1),

saxitoxin (Kellmann et al. 2008b; DQ787200.1), and finally the gene cluster for anatoxin biosynthesis (Méjean et al. 2009; FJ477836.1). As a result, the number of available gene sequences of the several cyanotoxins in the GenBank database has increased exponentially since the late 1990's, particularly for microcystin (Fig. 3.1). This is also in agreement with the higher number of publications on this cyanotoxin when compared with the others (Fig. 3.2), as retrieved by a literature search on the Scopus database conducted in early September 2016. Remarkably, the biosynthetic gene cluster of the dermatotoxin lyngbyatoxin – known to cause the so-called "swimmer's itch" – has earned little attention from molecular ecotoxicologists, so far (Figs. 3.1 and 3.2).



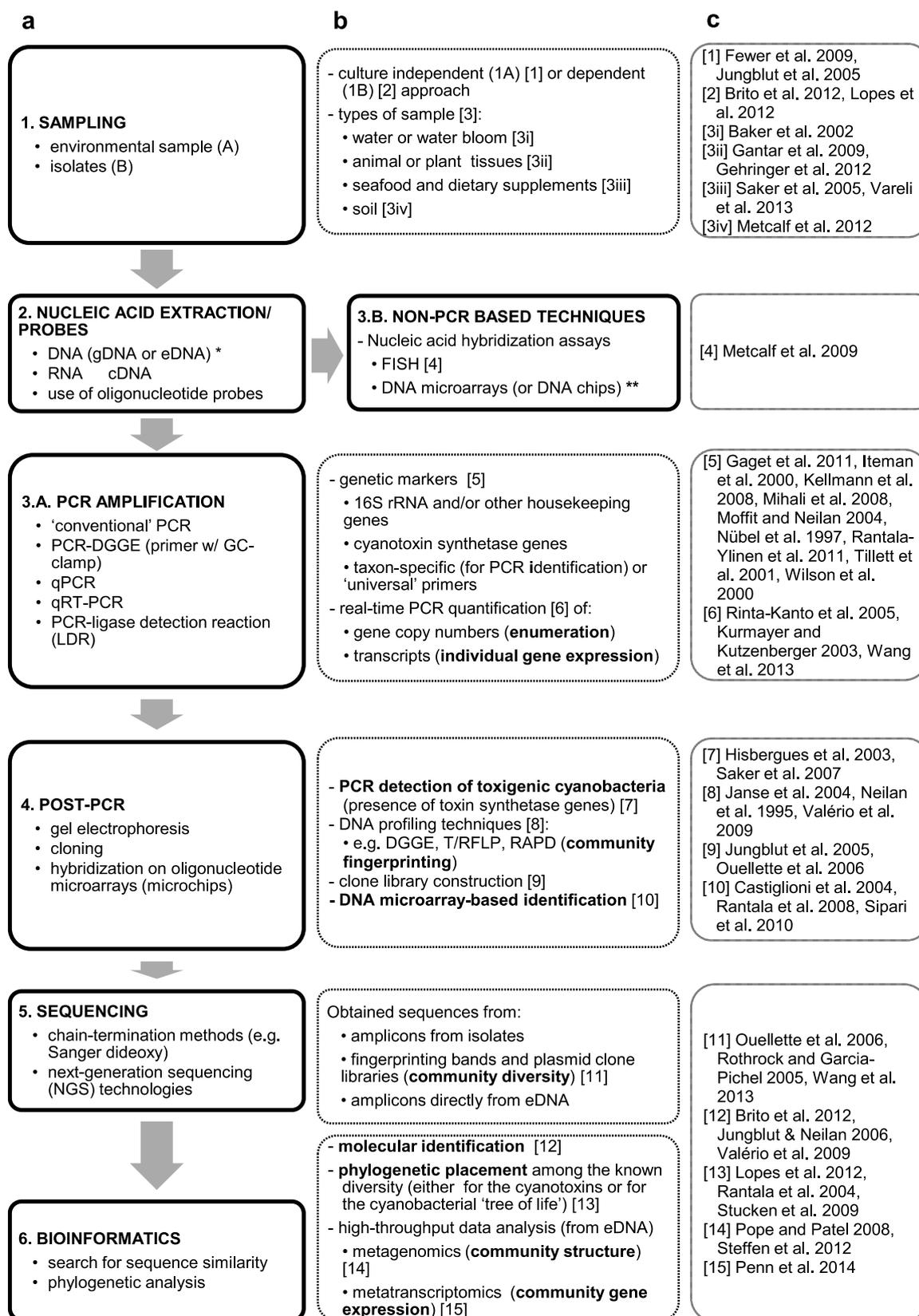
**Figure 3.2.** Percentage of publications on toxic cyanobacteria that used molecular methods, until mid-2016; by cyanotoxin.

### 3.1.3 Possible Applications

The availability of cyanotoxins gene cluster sequences led to the possibility of applying different molecular methods and approaches for monitoring toxigenic cyanobacteria directly in the environment. Fig. 3.3 summarizes the various applications and different molecular methods being used for the study of naturally occurring (toxic) cyanobacteria. It shows the main stages need to be taken during molecular-based approaches, pointing out the outcomes that may be achieved at the various stages, while providing some relevant bibliography for each specific issue. A more complete, although not exhaustive, list of publications is presented below in Table 3.1.

Despite the wide use of molecular tools for the detection of toxic cyanobacteria by the scientific community (which can be inferred from Table 3.1), very little is known about their use (especially on a regular basis) by other health- or environmental-related entities, namely by regulatory and monitoring agencies or by companies. For instance, Wood et al. (2013) stated that in New Zealand “despite research and validation demonstrating their potential, the application of these

tools by monitoring agencies has been limited". In fact, the attempt to develop molecular tools and to transfer this technology from academia to society has been a recurrent challenge during the last decade. Several transnational, regional or national projects, with different types of stakeholders, were or still are funded in Europe (e.g. CYANOCOST; B-BLOOMS2), North America (e.g. MERHAB-LGL; Water Research Foundation 2007a) and Oceania (e.g. Water Research Foundation 2007b; Wood et al. 2013). These projects have the purpose, at least as a specific goal, of developing and establish cost-effective molecular-based tools, with quick turnaround times able to assist monitoring decision processes. However, constraints to their adoption and implementation by potential stakeholders include (lack of) legislative requirements, perceived costs, and unwillingness to change procedures and practices (Water Research Foundation 2007b; Wood et al. 2013).



**Figure 3.3.** Overview of the main different applications of molecular-based methods for cyanobacteria (and/or their toxins) detection and identification. a – General flow diagram showing main stages; b – Methodological notes or applications description; in bold are possible outcomes or achievements, at this particular stage; c – Examples of relevant literature for topics identified in b), within square brackets. \* gDNA and eDNA regard to isolate genomic DNA and environmental DNA, respectively; \*\* in the scarce studies on cyanobacteria, hybridization is preceded by PCR (see stage 4) [adapted from Moreira et al. (2014) with permission from Springer.]

**Table 3.1** List of studies showing the main different applications (and some methodological features) in which the detection/identification of tox(igen)ic cyanobacteria have been pursued, by molecular-based method used. Note: the complete table is available in digital format (see Appendix A6)

| Method%          | Application        | Type of sample                 | Targeted molecular marker(s) <sup>&amp;</sup>                     | Cyanotoxin biosynthesis genes <sup>#</sup> | Complementary evaluation/validation <sup>§</sup> | References <sup>£</sup>        |
|------------------|--------------------|--------------------------------|---|--|--|--------------------------------|
| Conventional PCR | aquatic ecosystems | isolate(s)                     | 16S rRNA gene   | n.d.                                       | ELISA (+, for MC), LC-MS/MS (+, for MC)          | Carmichael & Li 2006           |
|                  | aquatic ecosystems | water                          | <i>mcyA; mcyB; mcyC; mcyD; mcyE; mcyG; ndaF; cyrB; cyrC; sxtI</i> | MC (+), NOD (+), CYL (+), SXT (+)          | ELISA (-), MALDI-TOF/MS (-)                      | Lopes et al. 2012              |
|                  | aquatic ecosystems | isolate(s)                     | <i>aoaA; aoaB; aoaC</i>   | CYL (-)                                    | ELISA (-), HPLC (-)                              | Fathalli et al. 2011           |
|                  | aquatic ecosystems | water                          | <i>mcyA; mcyB; mcyC; mcyD; mcyE</i>                               | MC (+)                                     | ELISA (+)  | Te & Gin 2011                  |
|                  | aquatic ecosystems | isolate(s)                     | <i>mcyB</i>   | MC (+)                                     | MALDI-TOF (+)                                    | Kurmayer et al. 2002           |
|                  | aquatic ecosystems | isolate(s)                     | <i>sxtA; sxtG; sxtH; sxtI; sxtX</i>                               | SXT (+)                                    | ELISA (+) and reported in previous studies       | Casero et al. 2014             |
|                  | aquatic ecosystems | water                          | <i>mcyA; mcyB; mcyD; mcyE</i>                                     | MC (+)                                     | HPLC (+), PPIA (+)                               | Mankiewicz-Boczek et al. 2006  |
|                  | aquatic ecosystems | water                          | <i>cyrJ</i>   | CYL (+)                                    | HPLC (+)   | Mankiewicz-Boczek et al. 2012a |
|                  | aquatic ecosystems | isolate(s) and water           | <i>cyrJ, pks</i>  | CYL (+)                                    | ELISA (+), HPLC (+), HPLC-MS/MS (+)              | Kokociński et al. 2013         |
|                  | aquatic ecosystems | isolate(s)                     | <i>aoaB; aoaC</i>   | CYL (-)                                    | HPLC (-)   | Valério et al. 2005            |
|                  | aquatic ecosystems | isolate(s) and water           | <i>mcyA; mcyB; mcyE; aoaA; aoaB; aoaC</i>                         | MC (+)                                     | ELISA (+)  | Martins et al. 2010            |
|                  | aquatic ecosystems | water                          | <i>mcyA</i>   | MC (+)                                     | ELISA (+)  | Saker et al. 2007b             |
|                  | aquatic ecosystems | water                          | <i>mcyA; mcyB</i>   | MC (+)                                     | PPIA (+), LC-MS (+)                              | Allender et al. 2009           |
|                  | aquatic ecosystems | water (from food fish farming) | <i>mcyB</i> ; 16S rRNA gene                                       | MC (+)                                     | HPLC (+)   | Nonneman & Zimba 2002          |
|                  | aquatic ecosystems | bloom                          | <i>mcyE; ndaF; sxtA; cyrA</i> ; 16S rRNA gene                     | MC and/or NOD (+), SXT (+), CYL (+)        | ELISA (-;+;+)                                    | Al-Tebrineh et al. 2012a       |
|                  | aquatic ecosystems | bloom                          | <i>mcyA; mcyC</i>   | MC (+)                                     | Mouse bioassay (+), PPIA (+)                     | Baker et al. 2001              |
|                  | aquatic ecosystems | bloom                          | <i>mcyA</i> ; 16S rRNA gene                                       | MC (+), SXT (+)                            | HPLC (+;+)                                       | Baker et al. 2002              |
|                  | aquatic ecosystems | water                          | <i>cyrC; mcyA; mcyC; sxtI</i>                                     | CYL (+), MC (-), SXT (-)                   | HPLC-MS/MS (+)                                   | McGregor et al. 2011           |

|               |                              |  |  |                                   |   |                                   |
|---------------|------------------------------|--|--|-----------------------------------|---|-----------------------------------|
|               | aquatic ecosystems           | water                                      | <i>mcyE</i>  | MC (+)                            | HPLC (+)  | Belykh et al. 2011                |
|               | aquatic ecosystems           | bloom                                      | <i>mcyB</i>  | MC (+)                            | ELISA (+)   | Bittencourt-Oliveira 2003         |
|               | aquatic ecosystems           | isolate(s)                                 | <i>cyrA; cyrB; cyrC; cyrJ; sxtA; sxtB; sxtI</i>                              | CYL (-), SXT (+)                  | LC-MS (-;+)   | Hoff-Rissetti et al. 2013         |
|               | aquatic ecosystems           | mat  | <i>mcyE</i>  | MC (+)                            | LC-MS (+)   | Wood et al. 2008                  |
|               | aquatic ecosystems           | isolate(s)                                 | <i>mcyA; mcyB; mcyC; ndaF;</i> CYL biosynthesis-related genes; 16S rRNA gene | MC (+), CYL (-)                   | <i>Artemia salina</i> bioassay (+), MALDI-TOF/MS (-)                    | Frazão et al. 2010                |
|               | aquatic ecosystems           | isolate(s)                                 | 16S rRNA gene  | n.d.                              | PPIA (+), ELISA (+)   | Gantar et al. 2009                |
|               | aquatic ecosystems           | mat (coral tissue)                         | <i>mcyA, mcyE; ndaF; cyrJ; sxtI</i>  | MC (+), NOD (-), CYL (-), SXT (-) | LC-MS (-;-;-)   | Glas et al. 2010                  |
|               | aquatic ecosystems           | herbarium specimens                        | <i>mcyD</i>  | MC (+)                            | HPLC (+), ELISA (+), MALDI-TOF (+)                                      | Metcalf et al. 2012b              |
|               | food (additive) surveillance | dietary supplements                        | <i>mcyE; ndaF; cpcBA</i> -IGS  | MC and/or NOD (+)                 | PPIA, ELISA and LC-MS (+, for MC), LC-MS (-, for CYL, ANA and SXT/PSPs) | Heussner et al. 2012              |
|               | food (additive) surveillance | dietary supplements                        | <i>mcyA; mcyE; ndaF; cpcBA</i> -IGS  | MC (+), MC and/or NOD (+)         | ELISA (+, for MC)   | Saker et al. 2005                 |
|               | food web bioaccumulation     | bloom and grazers (gut content)            | <i>mcyA; mcyB</i>  | MC (+)                            | PPIA (+), ELISA (+)   | Oberholster et al. 2006           |
|               | terrestrial ecosystems       | cryptogamic (soil) crust                   | <i>mcyD; aoaB; aoaC</i>  | MC (+), CYL (-)                   | HPLC (+; -), ELISA (+, for MC)  | Metcalf et al. 2012a              |
|               | terrestrial ecosystems       | cyanobiont isolate(s) and <i>in planta</i> | <i>mcyE; ndaA; ndaB; ndaF;</i> 16S rRNA gene                                 | NOD (+), MC and/or NOD (+)        | HPLC HESI-MS/MS (+, for NOD) (-, for MC)                                | Gehring et al. 2012               |
|               | terrestrial ecosystems       | cyanobiont isolate(s) and lichen tissue    | <i>mcyE</i>  | MC (+)                            | LC-MS (+, for MC and NOD)   | Kaasalainen et al. 2012           |
|               | terrestrial ecosystems       | cyanobiont isolate(s)                      | <i>mcyA; mcyD; mcyE; rpoC1;</i> 16S rRNA gene                                | MC (+)                            | PPIA (+), ELISA (+), HPLC-MS/MS (+)                                     | Oksanen et al. 2004               |
| Multiplex-PCR | aquatic ecosystems           | water                                      | <i>mcyA; mcyB; mcyC; mcyD; mcyE; mcyG; mcyJ; cyrC</i>                        | MC (+), CYL (+)                   | HPLC-MS/MS (+;+)  | Zhang et al. 2014                 |
|               | aquatic ecosystems           | bloom                                      | <i>mcyA; mcyB; mcyC; mcyD; mcyE; mcyG</i>                                    | MC (+)                            | MALDI-TOF (+), HPLC (+)   | Ouahid & Fernández del Campo 2009 |
|               | aquatic ecosystems           | water                                      | <i>mcyB; mcyD</i>  | MC (+)                            | n.d.  | Ouellette et al. 2006             |

|                        |                              |  |   |                                   |   |                              |
|------------------------|------------------------------|--|---|-----------------------------------|---|------------------------------|
|                        | food (additive) surveillance | dietary supplements                    | <i>mcyA; cpcBA</i> -IGS                               | MC (+)                            | MALDI-TOF/MS (+)                        | Saker et al. 2007a           |
| Multiplexed-tandem-PCR | aquatic ecosystems           | bloom                                  | <i>mcyE; ndaF; cyrC; sxtI</i>                         | MC (+), NOD (+), CYL (+), SXT (+) | ELISA (+;+;+;+)                         | Baker et al. 2013            |
| Quantitative PCR       | aquatic ecosystems           | water                                  | <i>mcyD; cpcBA</i> -IGS                               | MC (+)                            | n.d.                                    | Li et al. 2012               |
|                        | aquatic ecosystems           | water                                  | <i>mcyA</i>   | MC (+)                            | HPLC (+), ELISA (+)                     | Yoshida et al. 2005          |
|                        | aquatic ecosystems           | water                                  | <i>mcyA</i>   | MC (+)                            | n.d.                                    | Yoshida et al. 2007          |
|                        | aquatic ecosystems           | water                                  | <i>mcyA</i>   | MC (+)                            | ELISA (+)                               | Ha et al. 2011               |
|                        | aquatic ecosystems           | water                                  | <i>mcyA</i> , 16S rRNA gene                           | MC (+)                            | HPLC (+), PPIA (+)                      | Gagała et al. 2014           |
|                        | aquatic ecosystem            | water                                  | <i>cyrA</i>   | CYL (+)                           | ELISA (+), HPLC (+), HPLC-MS/MS (+)     | Kokociński et al. 2013       |
|                        | aquatic ecosystems           | water                                  | <i>mcyA; mcyB; mcyC; mcyD; mcyE; mcyG; mcyJ; cyrC</i> | MC (+), CYL (+)                   | HPLC-MS/MS (+;+)                        | Zhang et al. 2014            |
|                        | aquatic ecosystems           | water                                  | <i>mcyE</i>   | MC (+)                            | ELISA (+)                               | Te & Gin 2011                |
|                        | aquatic ecosystems           | water                                  | <i>mcyJ</i>   | MC (+)                            | HPLC (+)                                | Joung et al. 2011            |
|                        | aquatic ecosystems           | bloom                                  | <i>mcyE; ndaF</i>                                     | MC (+), NOD (+)                   | HPLC (+;+)                              | Al-Tebrineh et al. 2011      |
|                        | aquatic ecosystems           | water                                  | <i>mcyE</i>   | MC (+)                            | n.d.                                    | Vaitomaa et al. 2003         |
|                        | aquatic ecosystems           | sediment                               | <i>mcyB</i>   | MC (+)                            | PPIA (+)                                | Misson et al. 2012           |
|                        | aquatic ecosystems           | water                                  | <i>mcyA; cpc</i> operons                              | MC (+)                            | PPIA (+)                                | Briand et al. 2008           |
|                        | aquatic ecosystems           | water                                  | <i>mcyB; cpc</i> operons                              | MC (+)                            | n.d.                                    | Kurmayer & Kutzenberger 2003 |
|                        | aquatic ecosystems           | water                                  | <i>mcyA</i>   | MC (+)                            | HPLC (+)                                | Ostermaier & Kurmayer 2010   |
|                        | aquatic ecosystems           | water                                  | <i>aoaA; aoaB; aoaC</i>                               | CYL (+)                           | HPLC LC-MS (+)                          | Rasmussen et al. 2008        |
|                        | food (additive) surveillance | dietary supplements                    | <i>mcyB; cpcBA</i> -IGS; 16S rRNA gene                | MC (+)                            | ELISA (+), LC-MS/MS (+)                 | Vichi et al. 2012            |
|                        | food web bioaccumulation     | gut content of grazers and edible fish | 16S rRNA gene   | n.d.                              | ELISA (+, for MC), HPLC/PDA (+, for MC) | Sotton et al. 2014           |

|   |                          |   |   |                            |   |   |                          |
|---|--------------------------|---|---|----------------------------|---|---|--------------------------|
|   | food web bioaccumulation | bloom                                       | <i>mcyA; mcyB</i>                                   |                            | MC (+)                                    | PPIA (+), ELISA (+)   | Oberholster et al. 2006  |
| Multiplex real-time PCR                 | aquatic ecosystems       | isolate(s)                                  | <i>mcyE; ndaF; sxtA; cyrA</i>                       |                            | MC and/or NOD (+), SXT (+), CYL (+)       | n.d.  | Al-Tebrineh et al. 2012b |
|   | aquatic ecosystems       | bloom                                       | <i>mcyB; cpc</i> operons                            |                            | MC (+)                                    | n.d.  | Briand et al. 2009       |
|   | aquatic ecosystems       | bloom                                       | <i>mcyE; ndaF; sxtA; cyrA</i>                       |                            | MC and/or NOD (+), SXT (+), CYL (+)       | ELISA (-;+;+)   | Al-Tebrineh et al. 2012a |
| Microarrays                             | aquatic ecosystems       | isolate(s) and water                        | <i>mcyE</i> ; 16S rRNA gene                         |                            | MC (+)                                    | HPLC-MS (+), LC-MS (+)  | Sipari et al. 2010       |
|   | aquatic ecosystems       | isolate(s)                                  | HIP1; putative transporter gene                     | Na <sup>+</sup> -dependent | n.d.                                      | toxigenicity of strains (SXT-producer) reported in previous studies | Pomati et al. 2004       |
|   | aquatic ecosystems       | isolate(s) and water                        | <i>mcyE; ndaF</i>                                   |                            | MC and/or NOD (+)                         | n.d.  | Rantala et al. 2008      |
| DGGE/clone libraries/DNA typing methods | aquatic ecosystems       | isolate(s)                                  | <i>cpcBA</i> -IGS; rRNA ITS                         |                            | n.d.                                      | toxigenicity of strains reported in previous studies                | Neilan 2002              |
|   | aquatic ecosystems       | water                                       | <i>mcyE; ndaF</i> ; 16S rRNA gene                   |                            | MC and/or NOD (+)                         | ELISA (-;-)   | Amer et al. 2009         |
|   | aquatic ecosystems       | isolate(s)                                  | RAPD markers  |                            | n.d.                                      | toxigenicity of strains (MC-producer) reported in previous studies  | Nishihara et al. 1997    |
|   | aquatic ecosystems       | sediment                                    | <i>mcyB</i> ; 16S rRNA gene; rRNA ITS; <i>rpoC1</i> |                            | MC (-)?                                   | n.d.  | Innok et al. 2005        |
|   | aquatic ecosystems       | isolate(s)                                  | STR sequences; 16S rRNA gene                        |                            | n.d.                                      | HPLC-MS (+, for CYL)  | Chonudomkul et al. 2004  |
|   | aquatic ecosystems       | sediment                                    | <i>mcyB</i> ; rRNA ITS                              |                            | MC (+)                                    | PPIA (+)  | Misson et al. 2012       |
|   | aquatic ecosystems       | isolate(s) and water                        | <i>mcyA</i>   |                            | MC (+)                                    | ELISA (+), MALDI-TOF (+)  | Hisbergues et al. 2003   |
|   | aquatic ecosystems       | water                                       | rRNA ITS  |                            | n.d.                                      | HPLC (+, for MC)  | Kardinaal et al. 2007    |
|   | aquatic ecosystems       | isolate(s) and water                        | <i>mcyB; mcyD</i> ; 16S rRNA gene                   |                            | MC (+)                                    | toxigenicity of strains (MC-producer) reported in previous studies  | Ouellette et al. 2006    |
|   | aquatic ecosystems       | isolate(s)                                  | uncharacterized PS and PKS genes, 16S rRNA gene     |                            | PS/PKS (+)                                | HPLC (+, for CYL), (-, for SXT (i.e. PSP))                          | Valério et al. 2005      |
| aquatic ecosystems                      | mat                      | <i>mcyE; ndaF</i> ; rRNA ITS; 16S rRNA gene |   | MC and/or NOD (+)          | LC-MS (+, NOD), LC-MS (-, for MC and ANA) | Wood et al. 2012a   |                          |
| aquatic ecosystems                      | sediment                 | rRNA ITS                                    |   | n.d.                       | LC-MS (+, for ANA)                        | Wood et al. 2009  |                          |

|                                    |                          |  |  |  |  |                        |
|------------------------------------|--------------------------|--|--|--|--|------------------------|
|                                    | aquatic ecosystems       | isolate(s)                             | <i>mcyE</i> ; <i>ndaF</i> ; STRR sequences; 16S rRNA gene  | MC and/or NOD (+)                                    | PPIA (+); ELISA (+)  | Lyra et al. 2005       |
|                                    | aquatic ecosystems       | isolate(s)                             | HIP1; 16S rRNA gene  | n.d.   | Mouse bioassay (+), HPLC (+, for CYL)                              | Neilan et al. 2003     |
|                                    | aquatic ecosystems       | naturally occurring colonies and water | rRNA ITS   | n.d.   | MALDI-TOF/MS (+, for MC)   | Janse et al. 2004      |
|                                    | aquatic ecosystems       | mat                                    | <i>mcyE</i> ; rRNA ITS; 16S rRNA gene  | MC (+)   | PPIA (+), ELISA (+), LC-MS (+)                                     | Wood et al. 2008       |
|                                    | aquatic ecosystems       | isolate(s) and mat (coral tissue)      | <i>mcyA</i> ; 16S rRNA gene  | MC (+)   | PPIA (+), ELISA (+), HPLC-MS (+), HPLC-MS (-, for ANA and SXT)     | Richardson et al. 2007 |
|                                    | aquatic ecosystems       | mat                                    | <i>mcyE</i> ; <i>ndaF</i> ; ANA biosynthesis-related gene; <i>cyrJ</i> ; <i>sxtG</i> ; rRNA ITS; 16S rRNA gene | MC and/or NOD (-), ANA-related (-), CYL (-), SXT (+) | n.d.   | Villeneuve et al. 2012 |
|                                    | food web bioaccumulation | seawater and edible mussels            | <i>mcyA</i> ; <i>mcyE</i> ; uncharacterized genes; rRNA ITS  | MC (+) and PKS (+)                                   | ELISA (+, for MC), HPLC-MS: (+, for MC)                            | Vareli et al. 2012     |
| NGS and/or metagenomics techniques | aquatic ecosystems       | bloom                                  | <i>mcyA</i> ; 16S rRNA gene  | MC (+)   | PPIA (+), LC-MS (+), LC-MS (-, for CYL, SXT and BMAA)              | Steffen et al. 2012    |
|                                    | aquatic ecosystems       | bloom                                  | <i>Microcystis</i> metagenome  | n.d.   | toxigenicity of strains (MC-producer) reported in previous studies | Wilhelm et al. 2011    |
|                                    | aquatic ecosystems       | bloom                                  | 16S rRNA gene; BAC library   | n.d.   | n.d.   | Pope & Patel 2008      |
| FISH                               | aquatic ecosystems       | isolate(s) and water                   | <i>mcyA</i>  | MC (+)   | ELISA (+), HPLC (+)  | Gan et al. 2010        |
|                                    | aquatic ecosystems       | isolate(s) and water                   | <i>mcyA</i>  | MC (+)   | ELISA (+)  | Metcalf et al. 2009    |

% DGGE – Denaturing gradient gel electrophoresis, NGS – Next generation sequencing, FISH – Fluorescence *in situ* hybridization

& Biosynthetic gene cluster (cyanotoxin): *mcy* (microcystin), *nda* (nodularin), *cyr* ≡ *aoa* (cylindrospermopsin; gene name as it is in the original study), *sxt* (saxitoxin)

Other markers: rRNA ITS – 16S-23S rRNA internal transcribed spacer region; *cpc* operon – phycocyanin operon; *cpcBA*-IGS – phycocyanin intergenic spacer and flanking regions; *rpoC1* - RNA polymerase subunit beta; PS – nonribosomal peptide synthetase; PKS – polyketide synthase

Other abbreviations: HIP1 – highly iterated palindrome; RAPD – Random Amplified Polymorphic DNA; STRR – short tandemly repeated repetitive; BAC – bacterial artificial chromosome; CYL – cylindrospermopsin; ANA – anatoxin-a

# Molecular-based assessment of the presence (+) or absence (-) of biosynthesis genes for the production of the following toxins: MC – microcystin; NOD – nodularin; CYL – cylindrospermopsin; ANA – anatoxin-a; SXT – saxitoxin. n.d – not determined.

§ Toxicity/Toxin detection (+) or no detection (-) by additional complementary methods. Biological methods (non-specific): Mouse and brine shrimp (*Artemia salina*) bioassays; Biochemical methods: ELISA – enzyme-linked immunosorbent assay; PPIA – protein phosphatase inhibition assays; Analytical methods: LC-MS – liquid chromatography–mass spectrometry; HPLC – high-performance liquid chromatography; MALDI-TOF – matrix-assisted laser desorption/ionization time-of-flight (several variants of these techniques exist). n.d – not determined.

§ Refer to reference list in Chapter 3

### 3.1.4 Checklist of Publications, Applications and Lessons from Practice

#### ***Molecular-Based Studies on (Toxic) Cyanobacteria: Overview of Methods Being Used, and Generic Findings and Concerns***

In Table 3.1 is shown an extensive yet not full list of different molecular-based studies dealing with (possible) toxic cyanobacteria. The effective or potential toxicity of cyanobacteria determined in these studies was achieved either by molecular methods or other techniques (such as ELISA, HPLC, LC-MS, etc), or both. Since they are larger in number, studies based on conventional or quantitative PCR techniques included in Table 3.1 are intentionally underrepresented. Nonetheless, the selected examples intend to embrace as much as possible the different aspects of the overall studies/applications.

Thus, although incomplete, this table in conjunction with its full form (see Appendix A6; it has additional information on the sampling location, type of habitat and cyanobacterial identification), allows to gather some general conclusions:

- the large majority of the studies are being applied to aquatic environments, particularly surface freshwater (including planktonic bloom) (Appendix A6).
- other applications include food/additive surveillance (which will receive particular attention further ahead), as well as monitoring of terrestrial environments and of bioaccumulation through food web.
- in addition to water or bloom samples, other types of samples include tissues from different organisms (lichens, plants or animals), sediment or soil, as well as cyanobacterial isolates (either live cultures or dried material; for the latter, see Metcalf et al. 2012a, 2012b).
- DNA samples to be used in PCR-based techniques are obtained either by direct DNA isolation from environmental samples (eDNA) (culture-independent techniques; see Fig. 3.3) or through genomic DNA (gDNA) extracted from strains previously isolated (culture-dependent techniques; see Fig. 3.3).
- conventional PCR is the most employed technique, while qPCR is becoming more prevalent in recent years. With the exception of the methods usually used to profile the diversity of a community (especially DGGE), other molecular techniques (e.g. DNA chips, NGS, FISH, etc) are still scarcely used.
- the most preferred molecular markers are genes from the cyanotoxins' biosynthetic pathways. Nevertheless, indirect markers for identifying toxigenic cyanobacteria, such as the 16S rRNA gene, are still being used.
- most of the researches are designed to look for a specific cyanotoxin. In this particular, microcystins seem to be the preferred target, which is in accordance with Figs 3.1 and 3.2.
- related with it, the most studied taxon is *Microcystis* (Appendix A6).

- molecular-based studies are being preponderantly performed by, and in economically more developed countries (Appendix A6).

### ***The Need for Complementary Approaches***

Due to their high sensitivity, fairly rapid procedure time, and obtainment of reliable results, molecular-based methods are widely and routinely used in the analysis of natural samples from potentially toxic environments (Fig. 3.3 and Table 3.1), either through simple (i.e. conventional) PCR amplification or qPCR quantification or by using oligonucleotide-arrays. In fact, they allow a fast screening for the presence of toxin-related genes or other markers for toxic cyanobacteria, having the advantage of giving a prompt alarm if there is toxicity potential. But since they generate qualitative or indirect results, complementary approaches for assessing the presence of cyanotoxins in the samples are further required to corroborate or fully validate the molecular data. These include screening biochemical assays developed for detecting toxins (ELISA) or toxicity determination of specific cyanotoxins (PPIA), and stricter analysis obtained by chemical analytical techniques, such as HPLC, LC-MS, etc (Moreira et al. 2014). Notice that, in comparison to the molecular approach, these chemical methods are relatively expensive, and may require more specialized technicians and equipments. Preparation of the samples for chemical analyses is also more time consuming.

### ***Interpreting Results***

Possible false negative results may occur due to improper sample collection/nucleic acid extraction or from poor handling of nucleic acid samples before testing. It can also have origin in subsequent procedural steps, due to a wrong decision on which molecular tools to use (for instance, by choosing a PCR primer set or a probe not suitably designed for the genetic diversity present in the sample). False positive PCR results for the detection of cyanotoxin genes (indeed, as happens with other molecular markers) are likely to occur (Nonneman & Zimba 2002; Villeneuve et al. 2012). Thus, sequencing confirmation is highly recommended, or mandatory if a PCR-based technique is the only method being used. Even so, one should remember that the presence of these genes by itself does not confirm the existence of toxin production. Consequently, contrasting results are expected between molecular-derived data and those from complementary methods. One of these common conflicts is the evidence of a toxigenic potential inferred by molecular analysis, as deduced by the presence of known cyanotoxin genes (often confirmed by sequencing), which is not corroborated by data from other methods. Assuming that no procedural errors were made, this may be due to the fact that a) the gene cluster is present but inactive (Christiansen et al. 2008), or b) the toxin value present in the sample is below the detection level of the biochemical or analytical methods applied. As an illustration of

the latter, Saker et al. (2007a) only were able to detect microcystins in food supplements in a second round of MALDI-TOF analyses, after concentrating the samples.

Yet, looking at Table 3.1 it can be observed that the presence/absence of genes involved in the biosynthesis of cyanotoxins seems to broadly correlate positively with the results from biochemical and analytical methods. This general observation may derive from a positive-outcome bias (a type of publication bias) since congruent findings between molecular and biochemical or analytical data are more prone to be published than apparently incongruent results between them. Still, there are exceptions (see Table 3.1) to the aforementioned harmony between results from different types of data (Amer et al. 2009; Al-Tebrineh et al. 2012a), namely in those obtained from coastal environments (Frazão et al. 2010; Glas et al. 2010; Lopes et al. 2012; Vareli et al. 2012; see also Vareli et al. 2013 for a review on the presence of microcystin in marine ecosystems).

In brief, it is important to be aware of the possibility of false (positive or negative) results, to know how to interpret the data, and to recognize which are the limitations of the method(s).

### ***Choice of Molecular Tools for Toxigenicity Assessment***

Several group- and species-specific PCR primer sets have already been developed, using taxonomic markers such as the 16S rRNA gene, *gyrA* or *rpoB*, with the goal of distinguish toxic cyanobacteria. But actually, species identification is not conclusive in ascertaining if a cyanobacterium is a potential cyanotoxin producer (this is true either the identification has been achieved by molecular- or morphological-based methods). At most, only indirectly indicates such a capability, and only if it undoubtedly fit in a known phylo- or morphotype (i.e. a taxon defined by a gene marker or by morphological features, respectively) for which toxigenicity has been already proven. Though, even within these taxa there are some genotypes (i.e. strains) unable to produce toxins. For example, only 10 to 50% of *Microcystis* cells in natural systems have the genetic capability to produce microcystin (Joung et al. 2011; Rinta-Kanto et al. 2009). DNA profiling techniques using short tandem repeat (STR) analyses also fail to distinguish between toxic and non-toxic strains (Chonudomkul et al. 2004; Lyra et al. 2005).

Thus, the main tools available to assess toxigenicity by molecular methods are primers or probes designed after the sequencing and characterization of the several cyanotoxins gene clusters (for primers and probes development see other Chapters included in this book). However, given the foregoing (see previous sub-section), the choice of which is the best genetic marker (i.e. gene region within the cluster) that corroborates the existence of toxicity is an issue that is still being debated. For instance in microcystins some authors claim that the best candidate marker that confirms the presence of this cyanotoxin is the *mcyE* (Rantala et al. 2006; Mankiewicz-Boczek et al. 2006) while others refer to the *mcyG* (Tanabe et al. 2007). Both markers showed to have had the best correlation with the chemical data, as their detection by

PCR always returns a detection of microcystins by analytical methods. In the same way, if it is not detected by PCR, neither it is by analytical techniques. For cylindrospermopsin the *cyrJ* appears to be the best candidate marker (Mankiewicz-Boczek et al. 2012a; Hoff-Rissetti et al. 2013). For saxitoxin, *sxtI* seems to be useful to discriminate between SXT producers and non-producers species in the genus *Aphanizomenon* (Casero et al. 2014). However, this does not apply to other SXT producing genera (Casero et al. 2014). Regarding anatoxin-a (including its analog homoanatoxin-a), *anaC* shown to not be able to discriminate between all ANA producing and non-producing strains of *Anabaena* spp. evaluated by Rantala-Ylinen et al. (2011), giving also a false positive for *Aphanizomenon* sp.. In the other hand, it was successful to distinguish producers from non-producers among eleven tested strains of *Oscillatoria* spp. (Rantala-Ylinen et al. 2011). PCR results for *anaF*, another gene from the *ana* gene cluster, were generally congruent with the detection of ANA production in *Phormidium* spp. (Wood et al. 2012b) or *Tychonema* spp. (Shams et al. 2015; Salmaso et al. 2016). The only exception was obtained for *T. bornetii* NIVA-CYA 60, which has tested positive by PCR, but negative for ANA production by the receptor-binding assay (Shams et al. 2015).

### **Common and Possible Applications of Molecular Tools**

#### *Examples from Food (Additive) Surveillance*

The consumption of cyanobacteria, either as direct food or as an additive/supplement, represents a not so well known risk for human or animal health despite several evidences already reported (Heussner et al. 2012; Saker et al. 2005; Vichi et al. 2012), including some obtained by means of molecular techniques (Table 3.1). A special focus will be given to this issue, exploring how molecular techniques can be helpful to the control of food safety and quality. For that reason, a short summary on this topic will be given below. Moreover, three interlinked SOPs were conceived for “blue–green algae” food supplements (BGAS), where one can also find additional information and advices (Appendix A5). The first (SOP A) regards to sampling and processing of food supplement samples, the second (SOP B) concerns to methodological steps related with DNA extraction from those samples, while the third (SOP C) comprises a multiplex-PCR protocol previously developed by Saker et al. (2007a) for the detection of microcystin. All SOPs give methodological guidance, alternatives and suggestions (Appendix A5).

Common edible “microalgae” include cyanobacteria as *Arthrospira* spp. (which may also be known as *Spirulina* spp.), *Nostoc* spp. and *Aphanizomenon flos-aquae* or the green algae *Chlorella* spp.. Usually, they are commercially distributed worldwide as (blue-green) algae dietary supplements, but may also be used as a whole food (some *Nostoc* species; see But et al. 2002; Gao 1998; Johnson et al. 2008). The food supplements can be found in the marketplace in the form of tablets, capsules, flakes or in powder, and are commercialized as

being one hundred percent pure, as a varying mixture of the aforementioned organisms or as a mixture with other foodstuffs (Heussner et al. 2012). Besides the nutraceutical/food sector, edible microalgae are also used as feed supplement in the aquaculture, fishkeeping and poultry industries. A typical commercial system for algal biomass production involves two main stages (for a review see Takenaka & Yamaguchi 2014). The inoculum production stage (starter cultures) usually occurs in closed systems (e.g. small-scale photobioreactors). The second stage is the biomass production, which is usually conducted in large open ponds, normally in batch process. It is in this stage that contaminations with other organisms (Görs et al. 2010) are more prone to occur. Some of them can be toxic cyanobacteria, as detected in some BGAS supplements (Saker et al. 2005; Vichi et al. 2012). Moreover, the source of the edible microalgae (i.e. the origin of the strain used as starter culture) is very likely to be different between different manufacturers/producers (Habib et al. 2008). Therefore, even the species identification may often be disputable (Görs et al. 2010), and consequently it should be inspected. All these subjects are likely to be monitored by means of molecular tools (SOP B and SOP C, from Appendix A5).

Another topic that should represent a concern is the possible production of secondary metabolites other than cyanotoxins. This applies not only to potential contaminants, but to BGAS strains itself. This finding was exposed by Shih et al. (2013) who showed that some strains assigned to *Arthrospira* and *Spirulina* spp. possess gene clusters for the synthesis of bacteriocins, cyanobactins and terpenes. Remarkably, two of those strains are known for being used as BGAS (*Arthrospira* sp. PCC 8005 and *Arthrospira platensis* NIES-39; Habib et al. 2008 and Janssen et al. 2010, respectively). Their effects on humans can be either beneficial or adverse, depending furthermore on the substance itself, the dosage, its bioavailability, the frequency and duration of exposure, etc. Until these questions are investigated and satisfactorily clarified, the intake of such secondary metabolites may also raise concern for health, due to the chronic exposure that may incur frequent consumers (see also Gantar et al. 2008). A molecular-based monitoring may also be applied in this case.

#### *Examples from Aquatic Ecosystems: the Particular Cases of Poland and Portugal*

Unsurprisingly, molecular methods are above all applied to aquatic environments, especially in monitoring freshwater ecosystems, as it can be perceived by the number of related studies published (Table 3.1 and Appendix A6). Since it is not feasible to cover them all in detail, below are given some outcomes possible to be obtained through the application of molecular-based techniques, taking as examples studies performed in Polish and Portuguese fresh water aquatic systems.

In Poland, by routinely obtaining PCR products from water samples, it was possible to determine the occurrence of toxic genotypes responsible for synthesis of microcystins (target genes: *mcyA*,

*mcyB*, *mcyD* or *mcyE*) or cylindrospermopsin (target genes: *cyrJ* or *pks*) in different water bodies (Mankiewicz-Boczek et al. 2006, 2011a, 2012a; Kokociński et al. 2013). Molecular analysis indicated that *mcyE* gene – which takes part in the process of addition of the ADDA moiety, responsible for toxic properties to the microcystin molecule – was a very sensitive molecular marker for the determination of potential hepatotoxicity of cyanobacteria in different environmental samples, even if the cyanobacterial biomass in water was below 0.1 mg L<sup>-1</sup> (Mankiewicz-Boczek et al. 2006; Mankiewicz-Boczek et al. 2011b). Also through the utilization of PCR, and as stated before, *cyrJ* – the sulfotransferase gene required for tailoring reaction to complete the biosynthesis of cylindrospermopsin – showed to be an appropriate genetic marker for detection of cylindrospermopsin producers in Polish lakes (Mankiewicz-Boczek et al. 2012; Kokociński et al. 2013). Quantitative real-time PCR (qPCR) techniques are becoming widely used in the monitoring of toxic cyanobacterial blooms. They combine the qualitative confirmation of cyanobacterial presence along with their quantitative determination, allowing also a comparison with other measurable environmental variables. This, in turn, promotes better understanding of the mechanisms determining the dynamics of blooms, including the detection of toxic genotypes and the evaluation of their potential toxicity to other organisms, and the recognition of key environmental factors for blooms emergence and duration. Such knowledge is important to better understand the interaction between cyanobacteria and environmental factors, and to plan further remediation strategies for contaminated water body (Mankiewicz-Boczek 2012). The application of qPCR (based on 16S rRNA and *mcyA* genes copies, respectively for total *Microcystis* population and its toxic genotypes) allowed to observe that due to poor hydrological conditions in a Polish aquatic system (manifested by short water retention time) the total *Microcystis* gene copy number decreased, but the proportion of toxic *Microcystis* genotypes rised in comparison to values previously observed for the same location (Gaęała et al. 2014). This led to maintaining a similar level of microcystins concentration as that observed earlier, when the average *Microcystis* quantity was 10 time higher (Gaęała et al. 2014).

In Portugal, studies on the use of culture dependent or independent molecular methodologies in the screening of cyanotoxins has began little more than a decade ago. These studies used primers previously published in the literature, based on toxicity markers from the gene clusters of microcystins (*mcy*), cylindospermopsin (*cyr*), saxitoxin (*stx*) and anatoxin (*ana*) (for a revision, see Moreira et al. (2014)). Microcystins are the most widely screened and the most prevalent cyanotoxins in almost every freshwater system studied in Portugal. In two of the works published so far by Saker et al. (2007b) and Valério et al. (2010), single conventional PCR and multiplex-PCR techniques were routinely implemented in field samples. The molecular markers targeted were *mcyA*, in the two studies, and additionally *mcyB* in Valério et al. (2010). Both studies came to demonstrate a good correlation between the presence or absence of genes and the detection of the toxin by analytical methods (e.g. ELISA and HPLC), suggesting that these markers can be used in the early warning of cyanotoxins with a good certainty. Recently, through single PCR

reactions, Moreira et al. (2011a) establish the potential presence of the cyanotoxins microcystins (*mcyABE*), cylindrospermopsin (*cyrBC*) and saxitoxins (*stx1*) in the lakes and hotspots of the Azores islands. This study revealed that microcystin genes are the most commonly found in these aquatic systems, and that cylindrospermopsin genes are also present. Similarly, qPCR studies have also been applied in Portuguese freshwater samples, namely for the cyanotoxins microcystins and cylindrospermopsin. Martins et al. (2011) used specific marker genes for cyanobacteria, for *Microcystis* spp. and for MC-producing *Microcystis* (*mcyAB*). During this qPCR study it was found a negative significant correlation between the toxic (with *mcy* genes) and the non-toxic (without *mcy* genes) genotypes ratio and the overall *Microcystis* density observed. The findings from this study support that qPCR allows evaluating the dynamics of toxic cyanobacterial genotypes in a given freshwater system as well as to establish its seasonal variations in order to be used as a tool in a risk management decision. In other work using qPCR, Moreira et al. (2011b) have assessed the presence of *Cylindrospermopsis raciborskii* and the *cyrC* gene for cylindrospermopsin in a pond located in the centre of Portugal and revealed, for the first time, evidences of the presence of this cyanotoxin in continental waters of this country. Recently, the same authors reported the presence of the toxin itself, by analytical methodologies, in the same previously studied aquatic system (Moreira et al. unpublished data). See Table 3.1 to get to know the different molecular techniques being applied in studies on aquatic environments or Fig. 3.3 for a complete picture on the main outcomes that may arise from molecular-based studies.

#### *Other Examples*

Surveys concerning terrestrial environments and food web bioaccumulation are other important applications of molecular methods for monitoring/detect the presence of potential toxic cyanobacteria (Table 3.1). For instance, the usage of conventional PCR has exposed the hepatotoxin production capability of *Nostoc* cyanobionts (Appendix A6) living in symbioses with lichens (Kaasalainen et al. 2012; Oksanen et al. 2004) and plants (Gehring et al. 2012). Using a similar methodological approach, Metcalf et al. (2012a) revealed the risk that soil dust may represent to human health after discovering that cyanobacteria inhabiting soil crusts commonly found on arid environments are toxic. Conventional PCR (Oberholster et al. 2006), qPCR (Oberholster et al. 2006; Sotton et al. 2014) and PCR-DGGE (Vareli et al. 2012) methodologies have also allowed determining that toxic cyanobacteria may accumulate, sometimes extensively, in different organisms along the food chain, including in edible fishes and mussels (see Sukenik et al. 2015 for a review on bioaccumulation and effects of cyanotoxins on trophic webs).

*Other Than Toxigenicity-Related Examples*

In Table 3.2 are listed some studies that did not have as their main purpose the investigation of toxic cyanobacteria, although it may be related. Hence, the toxigenicity was not determined, neither by molecular techniques nor by other methods. Instead, these are studies that show other/new possibilities on the application of molecular techniques for the study of cyanobacteria, several of them also related to public health and environmental issues. For instance, Loza et al. (2013) have studied cyanobacterial communities by TGGE (a very similar technique to DGGE, but in this case a temperature gradient rather than a chemical gradient is used to denature the DNA) in order to develop a monitoring tool for water quality assessment. Likewise, Gomila et al. (2006) repeatedly detected the presence of cyanobacteria in hemodialysis water while screening clone libraries obtained from different points of a water treatment system. Other applications include, for example, attempts to DNA barcoding of potential toxigenic cyanobacteria (Kurobe et al. 2013) and development of a microarray for detecting genomic diversity associated with secondary metabolites in cyanobacteria (Pomati & Neilan 2004). Also worthy of record is the emergence of microbial community studies employing newer molecular methods (e.g. NGS technologies).

The full version of Table 3.2 is available from Appendix A7.

**Table 3.2** Examples of other applications/studies where molecular methods are being applied for the detection/identification of cyanobacteria (including potentially toxic or harmful taxa) and of microbial communities related to them. Note: the complete table is available in digital format (see Appendix A7)

| Application/<br>type of study                                       | Method  | Type of<br>sample      | Targeted<br>molecular<br>marker(s)  | Studied<br>organism(s)  | References <sup>§</sup> |
|---|---|------------------------|---|---|-------------------------|
| ecosystem health<br>and conservation<br>(coral reefs<br>monitoring) | DGGE/clone<br>libraries/DNA<br>typing methods | coral tissue           | 16S rRNA gene   | <i>Lyngbya</i> -like species<br>and unidentified<br>cyanobacterium                            | Sekar et al.<br>2006    |
| food web<br>bioaccumulation   | Conventional<br>PCR                           | water and<br>grazers   | 16S rRNA gene   | <i>Nodularia spumigena</i>  | Gorokhova<br>2009       |
| genetics of<br><i>secondary</i><br>metabolites                      | Microarrays                                   | isolates               | several, genes<br>encoding<br>unknown/putative<br>proteins  | <i>Anabaena circinalis</i>  | Pomati &<br>Neilan 2004 |
| microbial<br>community<br>structure                                 | DGGE/clone<br>libraries/DNA<br>typing methods | tree leaves            | 16S rRNA gene   | several epiphytic<br>cyanobacteria  | Rigonato et<br>al. 2012 |
| microbial<br>community<br>structure                                 | NGS and/or<br>metagenomics<br>techniques      | water                  | 16S rRNA gene   | Cyanobacteria (at<br>the phylum level)  | Ghai et al.<br>2011     |
| microbial<br>community<br>structure                                 | NGS and/or<br>metagenomics<br>techniques      | water                  | 16S rRNA gene   | <i>Cyanobium</i> ,<br><i>Merismopedia</i> ,<br><i>Synechococcus</i> ,<br><i>Pseudanabaena</i> | Ghai et al.<br>2012     |
| microbial<br>community<br>structure                                 | NGS and/or<br>metagenomics<br>techniques      | bloom                  | 16S rRNA gene   | <i>Microcystis</i> sp.  | Chen et al.<br>2011     |
| microbial<br>community<br>structure                                 | NGS and/or<br>metagenomics<br>techniques      | bloom                  | several, including<br>putative genes of<br>MC-LR cleavage<br>pathway ( <i>mlr</i> ) and<br>xenobiotic<br>metabolisms (GST<br>genes) * | transforming and<br>detoxifying<br>microcystins<br>microbiota                                 | Mou et al.<br>2013      |
| microbial<br>community<br>structure                                 | NGS and/or<br>metagenomics<br>techniques      | bloom                  | 23S rRNA gene   | Cyanobacteria (at<br>the order level) and<br>eukaryotic plastids<br>(at the phylum level)     | Steven et al.<br>2012   |
| taxonomy  | Conventional<br>PCR                           | herbarium<br>specimens | 16S rRNA gene   | several<br>cyanobacteria  | Palinska et al.<br>2006 |
| taxonomy (DNA<br>barcoding of<br>toxigenic<br>cyanobacteria)        | DGGE/clone<br>libraries/DNA<br>typing methods | water                  | 16S rRNA gene   | several potential<br>toxin-producing<br>cyanobacteria   | Kurobe et al.<br>2013   |
| water quality<br>(cyanobacterial<br>ecotypes as<br>bioindicators)   | DGGE/clone<br>libraries/DNA<br>typing methods | epilithic<br>biofilms  | 16S rRNA gene   | several epilithic<br>cyanobacteria  | Loza et al.<br>2013     |
| water quality (for<br>hemodialysis)                                 | DGGE/clone<br>libraries/DNA<br>typing methods | hemodialysis<br>water  | 16S rRNA gene   | several unidentified<br>cyanobacteria   | Gomila et al.<br>2006   |

\* MC-LR - microcystin-LR; GST - glutathione S-transferase

<sup>§</sup> Refer to reference list in Chapter 3

### 3.1.5 General conclusions

Since the late 1990's the number of available gene sequences of the several cyanotoxins has increased exponentially in the GenBank database, particularly for the microcystins which have become predominant until nowadays. This is in agreement with the higher number of publications on this cyanotoxin. The characterization of the biosynthetic gene cluster of this and the other cyanotoxins gave rise to the possibility of new applications and methodological improvements. In fact, since the detection of genes involved in cyanotoxins synthesis allows the early recognition of cyanobacterial threats, soon different molecular methods were adopted and developed to become a valuable tool for monitoring toxigenic cyanobacteria in the environment. Nevertheless, genetic information has to be seen as a primary assessment of toxicity, and so, the presence of toxins in a given sample should be validated analytically.

Due to its specificity, from all the applications discussed above food supplement surveillance has deserved special attention. Also, the lack of regulatory standards for the commercialization of BGAS (Gilroy et al. 2000; Habib et al. 2008) gives further importance to the monitoring of these products. As a result, three interlinked SOPs regarding the monitoring of BGAS were developed and made available (Appendix A5).

### 3.1.6 Acknowledgements

This work was partially supported by COST Action CYANOCOST, the Structured Program of R&D&I INNOVMAR - Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER-000035, Research Line NOVELMAR), funded by the Northern Regional Operational Program (NORTE2020) through the European Regional Development Fund (ERDF) and project FCT UID/Multi/04423/2013. VR would also like to acknowledge financial support from Fundação para a Ciência e a Tecnologia (FCT) fellowship SFRH/BD/80153/2011.

### 3.1.7 References

- ALLENDER, C. J., LECLEIR, G. R., RINTA-KANTO, J. M., SMALL, R. L., SATCHWELL, M. F., BOYER, G. L. & WILHELM, S. W. 2009. Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells. *Applied and Environmental Microbiology*, 75, 3598-3604.
- AL-TEBRINEH, J., GEHRINGER, M. M., AKCAALAN, R. & NEILAN, B. A. 2011. A new quantitative PCR assay for the detection of hepatotoxigenic cyanobacteria. *Toxicon*, 57, 546-554.
- AL-TEBRINEH, J., MERRICK, C., RYAN, D., HUMPAGE, A., BOWLING, L. & NEILAN, B. A. 2012a. Community composition, toxigenicity, and environmental conditions during a cyanobacterial bloom occurring along 1,100 kilometers of the Murray River. *Applied and Environmental Microbiology*, 78, 263-272.
- AL-TEBRINEH, J., PEARSON, L. A., YASAR, S. A. & NEILAN, B. A. 2012b. A multiplex qPCR targeting hepato- and neurotoxic cyanobacteria of global significance. *Harmful Algae*, 15, 19-25.

- AMER, R., DÍEZ, B. & EL-SHEHAWY, R. 2009. Diversity of hepatotoxic cyanobacteria in the Nile Delta, Egypt. *Journal of Environmental Monitoring*, 11, 126-133.
- BAKER, J. A., ENTSCH, B., NEILAN, B. A. & MCKAY, D. B. 2002. Monitoring changing toxigenicity of a cyanobacterial bloom by molecular methods. *Applied and Environmental Microbiology*, 68, 6070-6076.
- BAKER, J. A., NEILAN, B. A., ENTSCH, B. & MCKAY, D. B. 2001. Identification of cyanobacteria and their toxigenicity in environmental samples by rapid molecular analysis. *Environmental Toxicology*, 16, 472-482.
- BAKER, L., SENDALL, B. C., GASSER, R. B., MENJIVAR, T., NEILAN, B. A. & JEX, A. R. 2013. Rapid, multiplex-tandem PCR assay for automated detection and differentiation of toxigenic cyanobacterial blooms. *Molecular and Cellular Probes*, 27, 208-214.
- BELYKH, O. I., SOROKOVIKOVA, E. G., FEDOROVA, G. A., KALUZHNYAYA, O. V., KORNEVA, E. S., SAKIRKO, M. V. & SHERBAKOVA, T. A. 2011. Presence and genetic diversity of microcystin-producing cyanobacteria (*Anabaena* and *Microcystis*) in Lake Kotokel (Russia, Lake Baikal region). *Hydrobiologia*, 671, 241-252.
- BITTENCOURT-OLIVEIRA, M. D. C. 2003. Detection of potential microcystin-producing cyanobacteria in Brazilian reservoirs with a *mcyB* molecular marker. *Harmful Algae*, 2, 51-60.
- BRIAND, E., ESCOFFIER, N., STRAUB, C., SABART, M., QUIBLIER, C. & HUMBERT, J.-F. 2008. Spatiotemporal changes in the genetic diversity of a bloom-forming *Microcystis aeruginosa* (cyanobacteria) population. *The ISME Journal*, 3, 419-429.
- BRIAND, E., GUGGER, M., FRANCOIS, J. C., BERNARD, C., HUMBERT, J. F. & QUIBLIER, C. 2008. Temporal variations in the dynamics of potentially microcystin-producing strains in a bloom-forming *Planktothrix agardhii* (cyanobacterium) population. *Applied and Environmental Microbiology*, 74, 3839-3848.
- BRITO, Â., RAMOS, V., SEABRA, R., SANTOS, A., SANTOS, C. L., LOPO, M., FERREIRA, S., MARTINS, A., MOTA, R., FRAZÃO, B., MARTINS, R., VASCONCELOS, V. & TAMAGNINI, P. 2012. Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Systematic and Applied Microbiology*, 35, 110-119.
- BUT, P. P.-H., CHENG, L., CHAN, P. K., LAU, D. T.-W. & BUT, J. W.-H. 2002. *Nostoc flagelliforme* and faked items retailed in Hong Kong. *Journal of Applied Phycology*, 14, 143-145.
- CARMICHAEL, W. W. & LI, R. 2006. Cyanobacteria toxins in the Salton Sea. *Saline Systems*, 2, 5.
- CASERO, M. C., BALLOT, A., AGHA, R., QUESADA, A. & CIRÉS, S. 2014. Characterization of saxitoxin production and release and phylogeny of sxt genes in paralytic shellfish poisoning toxin-producing *Aphanizomenon gracile*. *Harmful Algae*, 37, 28-37.
- CASTIGLIONI, B., RIZZI, E., FROSINI, A., SIVONEN, K., RAJANIEMI, P., RANTALA, A., MUGNAI, M. A., VENTURA, S., WILMOTTE, A., BOUTTE, C., GRUBISIC, S., BALTHASART, P., CONSOLANDI, C., BORDONI, R., MEZZELANI, A., BATTAGLIA, C. & DE BELLIS, G. 2004. Development of a Universal Microarray Based on the Ligation Detection Reaction and 16S rRNA Gene Polymorphism To Target Diversity of Cyanobacteria. *Applied and Environmental Microbiology*, 70, 7161-7172.
- CHENG, C., ZAICHAO, Z., AIZHONG, D., JIAYAN, W., JINGFA, X. & YUJIAO, S. 2011. Bar-Coded Pyrosequencing Reveals the Bacterial Community during *Microcystis* water Bloom in Guanting Reservoir, Beijing. *Procedia Engineering*, 18, 341-346.
- CHONUDOMKUL, D., YONGMANITCHAI, W., THEERAGOOL, G., KAWACHI, M., KASAI, F., KAYA, K. & WATANABE, M. M. 2004. Morphology, genetic diversity, temperature tolerance and toxicity of *Cylindrospermopsis raciborskii* (Nostocales, Cyanobacteria) strains from Thailand and Japan. *FEMS Microbiology Ecology*, 48, 345-355.
- CHRISTIANSEN, G., MOLITOR, C., PHILMUS, B. & KURMAYER, R. 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Molecular Biology and Evolution*, 25, 1695-1704.
- DIETRICH, D. & HOEGER, S. 2005. Guidance values for microcystins in water and cyanobacterial supplement products (blue-green algal supplements): a reasonable or misguided approach? *Toxicology and Applied Pharmacology*, 203, 273-289.
- DITTMANN, E., NEILAN, B. A., ERHARD, M., VON DÖHREN, H. & BÖRNER, T. 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Molecular Microbiology*, 26, 779-787.
- EDWARDS, D. J. & GERWICK, W. H. 2004. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *Journal of the American Chemical Society*, 126, 11432-11433.
- FATHALLI, A., JENHANI, A. B. R., MOREIRA, C., AZEVEDO, J., WELKER, M., ROMDHANE, M., ANTUNES, A. & VASCONCELOS, V. 2011. Genetic variability of the invasive cyanobacteria

- Cylindrospermopsis raciborskii* from Bir M'cherga reservoir (Tunisia). *Archives of Microbiology*, 193, 595-604.
- FEWER, D. P., KÖYKKÄ, M., HALINEN, K., JOKELA, J., LYRA, C. & SIVONEN, K. 2009. Culture-independent evidence for the persistent presence and genetic diversity of microcystin-producing *Anabaena* (Cyanobacteria) in the Gulf of Finland. *Environmental Microbiology*, 11, 855-866.
- FRAZÃO, B., MARTINS, R. & VASCONCELOS, V. 2010. Are known cyanotoxins involved in the toxicity of picoplanktonic and filamentous North Atlantic marine cyanobacteria? *Marine Drugs*, 8, 1908-1919.
- GAĞAŁA, I., IZYDORCZYK, K., JURCZAK, T., PAWEŁCZYK, J., DZIADEK, J., WOJTAL-FRANKIEWICZ, A., JÓŻWIK, A., JASKULSKA, A. & MANKIEWICZ-BOCZEK, J. 2013. Role of environmental factors and toxic genotypes in the regulation of microcystins-producing cyanobacterial blooms. *Microbial Ecology*, 67, 465-479.
- GAGET, V., GRIBALDO, S. & TANDEAU DE MARSAC, N. 2010. An rpoB signature sequence provides unique resolution for the molecular typing of cyanobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 61, 170-183.
- GAN, N., HUANG, Q., ZHENG, L. & SONG, L. 2010. Quantitative assessment of toxic and nontoxic *Microcystis* colonies in natural environments using fluorescence in situ hybridization and flow cytometry. *Science China Life Sciences*, 53, 973-980.
- GANTAR, M., SEKAR, R. & RICHARDSON, L. L. 2009. Cyanotoxins from black band disease of corals and from other coral reef environments. *Microbial Ecology*, 58, 856-864.
- GANTAR, M. & SVIRČEV, Z. 2008. Microalgae and cyanobacteria: food for thought. *Journal of Phycology*, 44, 260-268.
- GAO, K. 1998. Chinese studies on the edible blue-green alga, *Nostoc flagelliforme*: a review. *Journal of Applied Phycology*, 10, 37-49.
- GEHRINGER, M. M., ADLER, L., ROBERTS, A. A., MOFFITT, M. C., MIHALI, T. K., MILLS, T. J. T., FIEKER, C. & NEILAN, B. A. 2012. Nodularin, a cyanobacterial toxin, is synthesized *in planta* by symbiotic *Nostoc* sp. *The ISME Journal*, 6, 1834-1847.
- GHAI, R., HERNANDEZ, C. M., PICAZO, A., MIZUNO, C. M., ININBERGS, K., DÍEZ, B., VALAS, R., DUPONT, C. L., MCMAHON, K. D., CAMACHO, A. & RODRIGUEZ-VALERA, F. 2012. Metagenomes of mediterranean coastal lagoons. *Scientific Reports*, 2.
- GHAI, R., RODRÍGUEZ-VALERA, F., MCMAHON, K. D., TOYAMA, D., RINKE, R., CRISTINA SOUZA DE OLIVEIRA, T., WAGNER GARCIA, J., PELLON DE MIRANDA, F. & HENRIQUE-SILVA, F. 2011. Metagenomics of the water column in the pristine upper course of the Amazon River. *PLoS ONE*, 6, e23785.
- GILROY, D. J., KAUFFMAN, K. W., HALL, R. A., HUANG, X. & CHU, F. S. 2000. Assessing potential health risks from microcystin toxins in blue-green algae dietary supplements. *Environmental Health Perspectives*, 108, 435-439.
- GLAS, M. S., MOTTI, C. A., NEGRI, A. P., SATO, Y., FROSCIO, S., HUMPAGE, A. R., KROCK, B., CEMBELLA, A. & BOURNE, D. G. 2010. Cyanotoxins are not implicated in the etiology of coral black band disease outbreaks on Pelorus Island, Great Barrier Reef. *FEMS Microbiology Ecology*, 73, 43-54.
- GOMILA, M., GASCÓ, J., GIL, J., BERNABEU, R., IÑIGO, V. & LALUCAT, J. 2006. A molecular microbial ecology approach to studying hemodialysis water and fluid. *Kidney International*, 70, 1567-1576.
- GOROKHOVA, E. 2009. Toxic cyanobacteria *Nodularia spumigena* in the diet of Baltic mysids: Evidence from molecular diet analysis. *Harmful Algae*, 8, 264-272.
- GÖRS, M., SCHUMANN, R., HEPPELE, D. & KARSTEN, U. 2010. Quality analysis of commercial *Chlorella* products used as dietary supplement in human nutrition. *Journal of Applied Phycology*, 22, 265-276.
- HA, J. H., HIDAKA, T. & TSUNO, H. 2011. Analysis of factors affecting the ratio of microcystin to chlorophyll-a in cyanobacterial blooms using real-time polymerase chain reaction. *Environmental Toxicology*, 26, 21-28.
- HABIB, M. A. B., PARIV, M., HUNTINGTON, T. C. & HASAN, M. R. 2008. *A review on culture, production and use of Spirulina as food for humans and feeds for domestic animals and fish*. FAO Fisheries and Aquaculture Circular no. 1034, Food and Agriculture Organisation (FAO) of the United Nations, Rome, Italy.
- HEUSSNER, A. H., MAZIJA, L., FASTNER, J. & DIETRICH, D. R. 2012. Toxin content and cytotoxicity of algal dietary supplements. *Toxicology and Applied Pharmacology*, 265, 263-271.
- HISBERGUES, M., CHRISTIANSEN, G., ROUHIAINEN, L., SIVONEN, K. & BÖRNER, T. 2003. PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. *Archives of Microbiology*, 180, 402-410.

- HOFF-RISSETI, C., DÖRR, F. A., SCHAKER, P. D. C., PINTO, E., WERNER, V. R. & FIORE, M. F. 2013. Cylindrospermopsin and saxitoxin synthetase genes in *Cylindrospermopsis raciborskii* strains from Brazilian freshwater. *PLoS ONE*, 8, e74238.
- IBELINGS, B. W. & CHORUS, I. 2007. Accumulation of cyanobacterial toxins in freshwater “seafood” and its consequences for public health: A review. *Environmental Pollution*, 150, 177-192.
- INNOK, S., MATSUMURA, M., BOONKERD, N. & TEAUMROONG, N. 2005. Detection of *Microcystis* in lake sediment using molecular genetic techniques. *World Journal of Microbiology and Biotechnology*, 21, 1559-1568.
- ITEMAN, I., RIPPKA, R., TANDEAU DE MARSAC, N. & HERDMAN, M. 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. *Microbiology*, 146, 1275-1286.
- JANSE, I., KARDINAAL, W. E. A., MEIMA, M., FASTNER, J., VISSER, P. M. & ZWART, G. 2004. Toxic and nontoxic *Microcystis* colonies in natural populations can be differentiated on the basis of rRNA gene internal transcribed spacer diversity. *Applied and Environmental Microbiology*, 70, 3979-3987.
- JANSSEN, P. J., MORIN, N., MERGEAY, M., LEROY, B., WATTIEZ, R., VALLAEYS, T., WALERON, K., WALERON, M., WILMOTTE, A., QUILLARDET, P., DE MARSAC, N. T., TALLA, E., ZHANG, C. C. & LEYS, N. 2010. Genome sequence of the edible cyanobacterium *Arthrospira* sp. PCC 8005. *Journal of Bacteriology*, 192, 2465-2466.
- JOHNSON, H. E., KING, S. R., BANACK, S. A., WEBSTER, C., CALLANAUPA, W. J. & COX, P. A. 2008. Cyanobacteria (*Nostoc commune*) used as a dietary item in the Peruvian highlands produce the neurotoxic amino acid BMAA. *Journal of Ethnopharmacology*, 118, 159-165.
- JOUNG, S.-H., OH, H.-M., KO, S.-R. & AHN, C.-Y. 2011. Correlations between environmental factors and toxic and non-toxic *Microcystis* dynamics during bloom in Daechung Reservoir, Korea. *Harmful Algae*, 10, 188-193.
- JUNGBLUT, A.-D., HAWES, I., MOUNTFORT, D., HITZFELD, B., DIETRICH, D. R., BURNS, B. P. & NEILAN, B. A. 2005. Diversity within cyanobacterial mat communities in variable salinity meltwater ponds of McMurdo Ice Shelf, Antarctica. *Environmental Microbiology*, 7, 519-529.
- JUNGBLUT, A.-D. & NEILAN, B. A. 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. *Archives of Microbiology*, 185, 107-114.
- KAASALAINEN, U., FEWER, D. P., JOKELA, J., WAHLSTEN, M., SIVONEN, K. & RIKKINEN, J. 2012. Cyanobacteria produce a high variety of hepatotoxic peptides in lichen symbiosis. *Proceedings of the National Academy of Sciences*, 109, 5886-5891.
- KARDINAAL, W. E. A., JANSE, I., KAMST-VAN AGTERVELD, M., MEIMA, M., SNOEK, J., MUR, L. R., HUISMAN, J., ZWART, G. & VISSER, P. 2007. *Microcystis* genotype succession in relation to microcystin concentrations in freshwater lakes. *Aquatic Microbial Ecology*, 48, 1-12.
- KELLMANN, R., MICHALI, T. K. & NEILAN, B. A. 2008a. Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers. *Journal of Molecular Evolution*, 67, 526-538.
- KELLMANN, R., MIHALI, T. K., JEON, Y. J., PICKFORD, R., POMATI, F. & NEILAN, B. A. 2008b. Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Applied and Environmental Microbiology*, 74, 4044-4053.
- KOKOCIŃSKI, M., MANKIEWICZ-BOCZEK, J., JURCZAK, T., SPOOF, L., MERILUOTO, J., REJMONCZYK, E., HAUTALA, H., VEHNÄÄINEN, M., PAWEŁCZYK, J. & SOININEN, J. 2013. *Aphanizomenon gracile* (Nostocales), a cylindrospermopsin-producing cyanobacterium in Polish lakes. *Environmental Science and Pollution Research*, 20, 5243-5264.
- KURMAYER, R., DITTMANN, E., FASTNER, J. & CHORUS, I. 2002. Diversity of microcystin genes within a population of the toxic cyanobacterium *Microcystis* spp. in Lake Wannsee (Berlin, Germany). *Microbial Ecology*, 43, 107-118.
- KURMAYER, R. & KUTZENBERGER, T. 2003. Application of Real-Time PCR for quantification of microcystin genotypes in a population of the toxic cyanobacterium *Microcystis* sp. *Applied and Environmental Microbiology*, 69, 6723-6730.
- KUROBE, T., BAXA, D. V., MIONI, C. E., KUDELA, R. M., SMYTHE, T. R., WALLER, S., CHAPMAN, A. D. & TEH, S. J. 2013. Identification of harmful cyanobacteria in the Sacramento-San Joaquin Delta and Clear Lake, California by DNA barcoding. *SpringerPlus*, 2, 491.
- LI, D., KONG, F., SHI, X., YE, L., YU, Y. & YANG, Z. 2012. Quantification of microcystin-producing and non-microcystin producing *Microcystis* populations during the 2009 and 2010 blooms in Lake Taihu using quantitative real-time PCR. *Journal of Environmental Sciences*, 24, 284-290.
- LOPES, V. R., RAMOS, V., MARTINS, A., SOUSA, M., WELKER, M., ANTUNES, A. & VASCONCELOS, V. M. 2012. Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries. *Marine Environmental Research*, 73, 7-16.
- LOZA, V., PERONA, E. & MATEO, P. 2012. Molecular fingerprinting of cyanobacteria from river biofilms as a water quality monitoring tool. *Applied and Environmental Microbiology*, 79, 1459-1472.

- LYRA, C. 2005. Benthic cyanobacteria of the genus *Nodularia* are non-toxic, without gas vacuoles, able to glide and genetically more diverse than planktonic *Nodularia*. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 55, 555-568.
- MANKIEWICZ-BOCZEK, J. 2012. Application of molecular tools in ecohydrology. *Ecohydrology & Hydrobiology*, 12, 165-170.
- MANKIEWICZ-BOCZEK, J., IZYDORCZYK, K., ROMANOWSKA-DUDA, Z., JURCZAK, T., STEFANIAK, K. & KOKOCIŃSKI, M. 2006. Detection and monitoring toxigenicity of cyanobacteria by application of molecular methods. *Environmental Toxicology*, 21, 380-387.
- MANKIEWICZ-BOCZEK, J., GAĞAŁA, I., KOKOCIŃSKI, M., JURCZAK, T. & STEFANIAK, K. 2011a. Perennial toxigenic *Planktothrix agardhii* bloom in selected lakes of Western Poland. *Environmental Toxicology*, 26, 10-20.
- MANKIEWICZ-BOCZEK, J., PALUS, J., GAĞAŁA, I., IZYDORCZYK, K., JURCZAK, T., DZIUBAŁTOWSKA, E., STĘPNIK, M., ARKUSZ, J., KOMOROWSKA, M., SKOWRON, A. & ZALEWSKI, M. 2011b. Effects of microcystins-containing cyanobacteria from a temperate ecosystem on human lymphocytes culture and their potential for adverse human health effects. *Harmful Algae*, 10, 356-365.
- MANKIEWICZ-BOCZEK, J., KOKOCIŃSKI, M., GAĞAŁA, I., PAWEŁCZYK, J., JURCZAK, T. & DZIADEK, J. 2012. Preliminary molecular identification of cylindrospermopsin-producing Cyanobacteria in two Polish lakes (Central Europe). *FEMS Microbiology Letters*, 326, 173-179.
- MARTINS, A., MOREIRA, C., VALE, M., FREITAS, M., REGUEIRAS, A., ANTUNES, A. & VASCONCELOS, V. 2011. Seasonal dynamics of *Microcystis* spp. and their toxigenicity as assessed by qPCR in a temperate reservoir. *Marine Drugs*, 9, 1715-1730.
- MARTINS, J., PEIXE, L. & VASCONCELOS, V. 2010. Cyanobacteria and bacteria co-occurrence in a wastewater treatment plant: absence of allelopathic effects. *Water Science & Technology*, 62, 1954.
- MASSERET, E. & SUKENIK, A. 2008. Amplified rDNA Restriction Analysis (ARDRA) for monitoring of potentially toxic cyanobacteria in water samples. *Israel Journal of Plant Sciences*, 56, 75-82.
- MCGREGOR, G. B., SENDALL, B. C., HUNT, L. T. & EAGLESHAM, G. K. 2011. Report of the cyanotoxins cylindrospermopsin and deoxy-cylindrospermopsin from *Raphidiopsis mediterranea* Skuja (Cyanobacteria/Nostocales). *Harmful Algae*, 10, 402-410.
- MÉJEAN, A., MANN, S. P., MALDINEY, T., VASSILIADIS, G. L., LEQUIN, O. & PLOUX, O. 2009. Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-a in the cyanobacterium *Oscillatoria* PCC 6506 occurs on a modular polyketide synthase initiated by L-proline. *Journal of the American Chemical Society*, 131, 7512-7513.
- METCALF, J. S., REILLY, M., YOUNG, F. M. & CODD, G. A. 2009. Localization of microcystin synthetase genes in colonies of the cyanobacterium *Microcystis* using fluorescence in situ hybridization. *Journal of Phycology*, 45, 1400-1404.
- METCALF, J. S., BEATTIE, K. A., PURDIE, E. L., BRYANT, J. A., IRVINE, L. M. & CODD, G. A. 2012b. Analysis of microcystins and microcystin genes in 60–170-year-old dried herbarium specimens of cyanobacteria. *Harmful Algae*, 15, 47-52.
- METCALF, J. S., RICHER, R., COX, P. A. & CODD, G. A. 2012a. Cyanotoxins in desert environments may present a risk to human health. *Science of the Total Environment*, 421-422, 118-123.
- MIHALI, T. K., KELLMANN, R., MUENCHHOFF, J., BARROW, K. D. & NEILAN, B. A. 2008. Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. *Applied and Environmental Microbiology*, 74, 716-722.
- MISSON, B., DONNADIEU-BERNARD, F., GODON, J.-J., AMBLARD, C. & LATOUR, D. 2012. Short- and long-term dynamics of the toxic potential and genotypic structure in benthic populations of *Microcystis*. *Water Research*, 46, 1438-1446.
- MOFFITT, M. C. & NEILAN, B. A. 2004. Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Applied and Environmental Microbiology*, 70, 6353-6362.
- MOREIRA, C., RAMOS, V., AZEVEDO, J. & VASCONCELOS, V. 2014. Methods to detect cyanobacteria and their toxins in the environment. *Applied Microbiology and Biotechnology*, 98, 8073-8082.
- MOREIRA, C., MARTINS, A., MOREIRA, C., & VASCONCELOS, V., 2011a. Toxigenic cyanobacteria in volcanic lakes and hot springs of a North Atlantic Island (S. Miguel, Azores, Portugal). *Fresenius Environmental Bulletin*, 20, 420-426.
- MOREIRA, C., MARTINS, A., AZEVEDO, J., FREITAS, M., REGUEIRAS, A., VALE, M., ANTUNES, A. & VASCONCELOS, V. 2011b. Application of real-time PCR in the assessment of the toxic cyanobacterium *Cylindrospermopsis raciborskii* abundance and toxicological potential. *Applied Microbiology and Biotechnology*, 92, 189-197.
- MOU, X., LU, X., JACOB, J., SUN, S. & HEATH, R. 2013. Metagenomic identification of bacterioplankton taxa and pathways involved in microcystin degradation in Lake Erie. *PLoS ONE*, 8, e61890.

- NEILAN, B. 2002. The molecular evolution and DNA profiling of toxic cyanobacteria. *Current Issues in Molecular Biology*, 4, 1-11.
- NEILAN, B. A., SAKER, M. L., FASTNER, J., TÖRÖKNÉ, A. & BURNS, B. P. 2003. Phylogeography of the invasive cyanobacterium *Cylindrospermopsis raciborskii*. *Molecular Ecology*, 12, 133-140.
- NEILAN, B. A., JACOBS, D. & GOODMAN, A. E. 1995. Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. *Applied and Environmental Microbiology*, 61, 3875-3883.
- NISHIHARA, H., MIWA, H., WATANABE, M., NAGASHIMA, M., YAGI, O. & TAKAMURA, Y. 1997. Random Amplified Polymorphic DNA (RAPD) analyses for discriminating genotypes of *Microcystis* cyanobacteria. *Bioscience, Biotechnology, and Biochemistry*, 61, 1067-1072.
- NONNEMAN, D. & ZIMBA, P. V. 2002. A PCR-based test to assess the potential for microcystin occurrence in channel catfish production ponds. *Journal of Phycology*, 38, 230-233.
- NÜBEL, U., GARCIA-PICHEL, F. & MUYZER, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, 63, 3327-3332.
- ÖBERHOLSTER, P. J., BOTHA, A. M. & CLOETE, T. E. 2006. Use of molecular markers as indicators for winter zooplankton grazing on toxic benthic cyanobacteria colonies in an urban Colorado lake. *Harmful Algae*, 5, 705-716.
- OXSANEN, I., JOKELA, J., FEWER, D. P., WAHLSTEN, M., RIKKINEN, J. & SIVONEN, K. 2004. Discovery of rare and highly toxic microcystins from lichen-associated cyanobacterium *Nostoc* sp. strain IO-102-I. *Applied and Environmental Microbiology*, 70, 5756-5763.
- OSTERMAIER, V. & KURMAYER, R. 2010. Application of Real-Time PCR to estimate toxin production by the cyanobacterium *Planktothrix* sp. *Applied and Environmental Microbiology*, 76, 3495-3502.
- OUAHID, Y. & DEL CAMPO, F. F. 2009. Typing of toxigenic *Microcystis* from environmental samples by multiplex PCR. *Applied Microbiology and Biotechnology*, 85, 405-412.
- OUELLETTE, A. J. A., HANDY, S. M. & WILHELM, S. W. 2006. Toxic *Microcystis* is widespread in Lake Erie: PCR detection of toxin genes and molecular characterization of associated cyanobacterial communities. *Microbial Ecology*, 51, 154-165.
- PALINSKA, K. A. 2006. Phylogenetic evaluation of cyanobacteria preserved as historic herbarium exsiccata. *International Journal of Systematic and Evolutionary Microbiology*, 56, 2253-2263.
- PENN, K., WANG, J., FERNANDO, S. C. & THOMPSON, J. R. 2014. Secondary metabolite gene expression and interplay of bacterial functions in a tropical freshwater cyanobacterial bloom. *The ISME Journal*, 8, 1866-1878.
- POMATI, F. & NEILAN, B. A. 2004. PCR-based positive hybridization to detect genomic diversity associated with bacterial secondary metabolism. *Nucleic Acids Research*, 32, e7.
- POMATI, F., BURNS, B. P. & NEILAN, B. A. 2004. Identification of an Na<sup>+</sup>-dependent transporter associated with saxitoxin-producing strains of the cyanobacterium *Anabaena circinalis*. *Applied and Environmental Microbiology*, 70, 4711-4719.
- POPE, P. B. & PATEL, B. K. C. 2008. Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS Microbiology Ecology*, 64, 9-27.
- RANTALA, A., FEWER, D. P., HISBERGUES, M., ROUHIAINEN, L., VAITOMAA, J., BORNER, T. & SIVONEN, K. 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. *Proceedings of the National Academy of Sciences*, 101, 568-573.
- RANTALA, A., RAJANIEMI-WACKLIN, P., LYRA, C., LEPISTO, L., RINTALA, J., MANKIEWICZ-BOCZEK, J. & SIVONEN, K. 2006. Detection of microcystin-producing cyanobacteria in Finnish lakes with genus-specific microcystin synthetase gene E (*mcyE*) PCR and associations with environmental factors. *Applied and Environmental Microbiology*, 72, 6101-6110.
- RANTALA, A., RIZZI, E., CASTIGLIONI, B., DE BELLIS, G. & SIVONEN, K. 2008. Identification of hepatotoxin-producing cyanobacteria by DNA-chip. *Environmental Microbiology*, 10, 653-664.
- RANTALA-YLINEN, A., KANA, S., WANG, H., ROUHIAINEN, L., WAHLSTEN, M., RIZZI, E., BERG, K., GUGGER, M. & SIVONEN, K. 2011. Anatoxin-a synthetase gene cluster of the cyanobacterium *Anabaena* sp. strain 37 and molecular methods to detect potential producers. *Applied and Environmental Microbiology*, 77, 7271-7278.
- RASMUSSEN, J. P., GIGLIO, S., MONIS, P. T., CAMPBELL, R. J. & SAINT, C. P. 2008. Development and field testing of a real-time PCR assay for cylindrospermopsin-producing cyanobacteria. *Journal of Applied Microbiology*, 104, 1503-1515.
- RICHARDSON, L. L., SEKAR, R., MYERS, J. L., GANTAR, M., VOSS, J. D., KACZMARSKY, L., REMILY, E. R., BOYER, G. L. & ZIMBA, P. V. 2007. The presence of the cyanobacterial toxin microcystin in black band disease of corals. *FEMS Microbiology Letters*, 272, 182-187.
- RIGONATO, J., ALVARENGA, D. O., ANDREOTE, F. D., DIAS, A. C. F., MELO, I. S., KENT, A. & FIORE, M. F. 2012. Cyanobacterial diversity in the phyllosphere of a mangrove forest. *FEMS Microbiology Ecology*, 80, 312-322.

- RINTA-KANTO, J. M., KONOPKO, E. A., DEBRUYN, J. M., BOURBONNIERE, R. A., BOYER, G. L. & WILHELM, S. W. 2009. Lake Erie *Microcystis*: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. *Harmful Algae*, 8, 665-673.
- RINTA-KANTO, J. M., OUELLETTE, A. J. A., BOYER, G. L., TWISS, M. R., BRIDGEMAN, T. B. & WILHELM, S. W. 2005. Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative Real-Time PCR. *Environmental Science & Technology*, 39, 4198-4205.
- ROTHROCK, M. J. & GARCIA-PICHEL, F. 2005. Microbial diversity of benthic mats along a tidal desiccation gradient. *Environmental Microbiology*, 7, 593-601.
- SAKER, M., MOREIRA, C., MARTINS, J., NEILAN, B. & VASCONCELOS, V. M. 2009. DNA profiling of complex bacterial populations: toxic cyanobacterial blooms. *Applied Microbiology and Biotechnology*, 85, 237-252.
- SAKER, M. L., WELKER, M. & VASCONCELOS, V. M. 2007a. Multiplex PCR for the detection of toxigenic cyanobacteria in dietary supplements produced for human consumption. *Applied Microbiology and Biotechnology*, 73, 1136-1142.
- SAKER, M. L., VALE, M., KRAMER, D. & VASCONCELOS, V. M. 2007b. Molecular techniques for the early warning of toxic cyanobacteria blooms in freshwater lakes and rivers. *Applied Microbiology and Biotechnology*, 75, 441-449.
- SAKER, M. L., JUNGBLUT, A. D., NEILAN, B. A., RAWN, D. F. K. & VASCONCELOS, V. M. 2005. Detection of microcystin synthetase genes in health food supplements containing the freshwater cyanobacterium *Aphanizomenon flos-aquae*. *Toxicon*, 46, 555-562.
- SALMASO, N., CERASINO, L., BOSCAINI, A. & CAPELLI, C. 2016. Planktic *Tychonema* (Cyanobacteria) in the large lakes south of the Alps: phylogenetic assessment and toxigenic potential. *FEMS Microbiology Ecology*, doi: 10.1093/femsec/fiw155.
- SEKAR, R., MILLS, D. K., REMILY, E. R., VOSS, J. D. & RICHARDSON, L. L. 2006. Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased *Siderastrea siderea*. *Applied and Environmental Microbiology*, 72, 5963-5973.
- SHAMS, S., CAPELLI, C., CERASINO, L., BALLOT, A., DIETRICH, D. R., SIVONEN, K. & SALMASO, N. 2015. Anatoxin-a producing *Tychonema* (Cyanobacteria) in European waterbodies. *Water Research*, 69, 68-79.
- SHIH, P. M., WU, D., LATIFI, A., AXEN, S. D., FEWER, D. P., TALLA, E., CALTEAU, A., CAI, F., TANDEAU DE MARSAC, N., RIPPKA, R., HERDMAN, M., SIVONEN, K., COURSIIN, T., LAURENT, T., GOODWIN, L., NOLAN, M., DAVENPORT, K. W., HAN, C. S., RUBIN, E. M., EISEN, J. A., WOYKE, T., GUGGER, M. & KERFELD, C. A. 2012. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, 110, 1053-1058.
- SIPARI, H., RANTALA-YLINEN, A., JOKELA, J., OKSANEN, I. & SIVONEN, K. 2010. Development of a chip assay and quantitative PCR for detecting microcystin synthetase E gene expression. *Applied and Environmental Microbiology*, 76, 3797-3805.
- SOTTON, B., GUILLARD, J., ANNEVILLE, O., MARÉCHAL, M., SAVICHTCHEVA, O. & DOMAIZON, I. 2014. Trophic transfer of microcystins through the lake pelagic food web: Evidence for the role of zooplankton as a vector in fish contamination. *Science of the Total Environment*, 466-467, 152-163.
- STEFFEN, M. M., LI, Z., EFFLER, T. C., HAUSER, L. J., BOYER, G. L. & WILHELM, S. W. 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS ONE*, 7, e44002.
- STEVEN, B., MCCANN, S. & WARD, N. L. 2012. Pyrosequencing of plastid 23S rRNA genes reveals diverse and dynamic cyanobacterial and algal populations in two eutrophic lakes. *FEMS Microbiology Ecology*, 82, 607-615.
- STUCKEN, K., MURILLO, A. A., SOTO-LIEBE, K., FUENTES-VALDÉS, J. J., MÉNDEZ, M. A. & VÁSQUEZ, M. 2009. Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis raciborskii* strains. *Systematic and Applied Microbiology*, 32, 37-48.
- SUKENIK, A., QUESADA, A. & SALMASO, N. 2015. Global expansion of toxic and non-toxic cyanobacteria: effect on ecosystem functioning. *Biodiversity and Conservation*, 24, 889-908.
- TAKENAKA, H. & YAMAGUCHI, Y. 2013. Commercial-scale culturing of cyanobacteria: an industrial experience. In N. Sharma, A. Rai & L. Stal (eds) *Cyanobacteria: An Economic Perspective*, John Wiley & Sons, Chichester, UK.
- TANABE, Y., KASAI, F. & WATANABE, M. M. 2007. Multilocus sequence typing (MLST) reveals high genetic diversity and clonal population structure of the toxic cyanobacterium *Microcystis aeruginosa*. *Microbiology*, 153, 3695-3703.
- TE, S. H. & GIN, K. Y.-H. 2011. The dynamics of cyanobacteria and microcystin production in a tropical reservoir of Singapore. *Harmful Algae*, 10, 319-329.

- TILLET, D., PARKER, D. L. & NEILAN, B. A. 2001. Detection of toxigenicity by a probe for the microcystin synthetase A gene (*mcyA*) of the cyanobacterial genus *Microcystis*: comparison of toxicities with 16S rRNA and phycocyanin operon (phycocyanin intergenic spacer) phylogenies. *Applied and Environmental Microbiology*, 67, 2810-2818.
- VAITOMAA, J., RANTALA, A., HALINEN, K., ROUHIAINEN, L., TALLBERG, P., MOKELKE, L. & SIVONEN, K. 2003. Quantitative Real-Time PCR for determination of microcystin synthetase E copy numbers for *Microcystis* and *Anabaena* in lakes. *Applied and Environmental Microbiology*, 69, 7289-7297.
- VALÉRIO, E., CHAMBEL, L., PAULINO, S., FARIA, N., PEREIRA, P. & TENREIRO, R. 2010. Multiplex PCR for detection of microcystins-producing cyanobacteria from freshwater samples. *Environmental Toxicology*, 25, 251-260.
- VALERIO, E., CHAMBEL, L., PAULINO, S., FARIA, N., PEREIRA, P. & TENREIRO, R. 2009. Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. *Microbiology*, 155, 642-656.
- VALÉRIO, E., PEREIRA, P., SAKER, M. L., FRANCA, S. & TENREIRO, R. 2005. Molecular characterization of *Cylindrospermopsis raciborskii* strains isolated from Portuguese freshwaters. *Harmful Algae*, 4, 1044-1052.
- VARELI, K., JAEGER, W., TOUKA, A., FRILLINGOS, S., BRIASOULIS, E. & SAINIS, I. 2013. Hepatotoxic seafood poisoning (HSP) due to microcystins: A threat from the ocean? *Marine Drugs*, 11, 2751-2768.
- VARELI, K., ZARALI, E., ZACHARIOUDAKIS, G. S. A., VAGENAS, G., VARELIS, V., PILIDIS, G., BRIASOULIS, E. & SAINIS, I. 2012. Microcystin producing cyanobacterial communities in Amvrakikos Gulf (Mediterranean Sea, NW Greece) and toxin accumulation in mussels (*Mytilus galloprovincialis*). *Harmful Algae*, 15, 109-118.
- VICHI, S., LAVORINI, P., FUNARI, E., SCARDALA, S. & TESTAI, E. 2012. Contamination by *Microcystis* and microcystins of blue-green algae food supplements (BGAS) on the Italian market and possible risk for the exposed population. *Food and Chemical Toxicology*, 50, 4493-4499.
- VILLENEUVE, A., LAURENT, D., CHINAIN, M., GUGGER, M. & HUMBERT, J.-F. 2012. Molecular characterization of the diversity and potential toxicity of cyanobacterial mats in two tropical lagoons in the South Pacific Ocean. *Journal of Phycology*, 48, 275-284.
- WANG, X., SUN, M., XIE, M., LIU, M., LUO, L., LI, P. & KONG, F. 2013. Differences in microcystin production and genotype composition among *Microcystis* colonies of different sizes in Lake Taihu. *Water Research*, 47, 5659-5669.
- WATER RESEARCH FOUNDATION. 2007a. Development of molecular reporters for microcystin activity and toxicity, Project #2818. AwwaRF, Denver, USA. Online, viewed 15 April 2016 at <http://www.waterrf.org/PublicReportLibrary/2818.pdf>.
- WATER RESEARCH FOUNDATION. 2007b. Early Detection of Cyanobacterial Toxins Using Genetic Methods, Project #2881. AwwaRF, Denver, USA. Online, viewed 15 April 2016 at [http://www.waterrf.org/ExecutiveSummaryLibrary/91198\\_2881\\_profile.pdf](http://www.waterrf.org/ExecutiveSummaryLibrary/91198_2881_profile.pdf).
- WILHELM, S. W., FARNSLEY, S. E., LECLEIR, G. R., LAYTON, A. C., SATCHWELL, M. F., DEBRUYN, J. M., BOYER, G. L., ZHU, G. & PAERL, H. W. 2011. The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae*, 10, 207-215.
- WILSON, K. M., SCHEMBRI, M. A., BAKER, P. D. & SAINT, C. P. 2000. Molecular characterization of the toxic cyanobacterium *Cylindrospermopsis raciborskii* and design of a species-specific PCR. *Applied and Environmental Microbiology*, 66, 332-338.
- WOOD, S. A., SMITH, K. F., BANKS, J. C., TREMBLAY, L. A., RHODES, L., MOUNTFORT, D., CARY, S. C. & POCHON, X. 2013. Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zealand Journal of Marine and Freshwater Research*, 47, 90-119.
- WOOD, S. A., KUHAJEK, J. M., WINTON, M. & PHILLIPS, N. R. 2012a. Species composition and cyanotoxin production in periphyton mats from three lakes of varying trophic status. *FEMS Microbiology Ecology*, 79, 312-326.
- WOOD, S. A., SMITH, F. M. J., HEATH, M. W., PALFROY, T., GAW, S., YOUNG, R. G. & RYAN, K. G. 2012b. Within-mat variability in anatoxin-a and homoanatoxin-a production among benthic *Phormidium* (Cyanobacteria) strains. *Toxins*, 4, 900-912.
- WOOD, S. A., JENTZSCH, K., RUECKERT, A., HAMILTON, D. P. & CARY, S. C. 2009. Hindcasting cyanobacterial communities in Lake Okaro with germination experiments and genetic analyses. *FEMS Microbiology Ecology*, 67, 252-260.
- WOOD, S. A., MOUNTFORT, D., SELWOOD, A. I., HOLLAND, P. T., PUDDICK, J. & CARY, S. C. 2008. Widespread distribution and identification of eight novel microcystins in antarctic cyanobacterial mats. *Applied and Environmental Microbiology*, 74, 7243-7251.

- YOSHIDA, M., YOSHIDA, T., TAKASHIMA, Y., HOSODA, N. & HIROISHI, S. 2007. Dynamics of microcystin-producing and non-microcystin-producing *Microcystis* populations is correlated with nitrate concentration in a Japanese lake. *FEMS Microbiology Letters*, 266, 49-53.
- YOSHIDA, M., YOSHIDA, T., TAKASHIMA, Y., KONDO, R. & HIROISHI, S. 2005. Genetic diversity of the toxic cyanobacterium *Microcystis* in Lake Mikata. *Environmental Toxicology*, 20, 229-234.
- ZHANG, W., LOU, I., UNG, W. K., KONG, Y. & MOK, K. M. 2014. Analysis of cylindrospermopsin- and microcystin-producing genotypes and cyanotoxin concentrations in the Macau storage reservoir. *Hydrobiologia*, 741, 51-68.

## 4. Preserving cyanobacterial diversity as a natural resource

### 4.1 Cyanobacterial diversity held in mBRCs as a biotechnological asset: the case study of the newly established LEGE Culture Collection

Vitor Ramos<sup>1,2,#</sup>, João Morais<sup>1,#</sup>, Raquel Castelo-Branco<sup>1,#</sup>, Ângela Pinheiro<sup>1</sup>, Joana Martins<sup>1</sup>, Ana Regueiras<sup>1,2</sup>, Ana L. Pereira<sup>1</sup>, Viviana R. Lopes<sup>1,3</sup>, Bárbara Frazão<sup>1,4</sup>, Dina Gomes<sup>1,5</sup>, Cristiana Moreira<sup>1</sup>, Maria Sofia Costa<sup>1</sup>, Sébastien Brûle<sup>6</sup>, Silvia Faustino<sup>7</sup>, Rosário Martins<sup>1,8</sup>, Martin Saker<sup>1,9</sup>, Joana Osswald<sup>1</sup>, Pedro N. Leão<sup>1</sup>, Vitor M. Vasconcelos<sup>1,2\*</sup>

<sup>1</sup> Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Matosinhos, Portugal

<sup>2</sup> Faculdade de Ciências, Universidade do Porto, Porto, Portugal

<sup>3</sup> Nanotechnology and Functional Materials, Department of Engineering Sciences, Uppsala University, Uppsala, Sweden

<sup>4</sup> IPMA - Portuguese Institute of Sea and Atmosphere, Lisbon, Portugal

<sup>5</sup> ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

<sup>6</sup> Master 2 Biotechnologie, Université de Bretagne-Sud, Lorient/Vannes, France

<sup>7</sup> Federal Amapá University (UNIFAP), Laboratory of Algae Cultivation and Bioprospection, Macapá, Amapá, Brazil

<sup>8</sup> Health and Environment Research Centre, School of Health, Polytechnic Institute of Porto, Porto, Portugal

<sup>9</sup> Alpha Environmental Solutions, Dubai, United Arab Emirates

\* Corresponding author. Email address: Email: vmvascon@fc.up.pt

# these authors contributed equally to this work

#### 4.1.1 Abstract

Cyanobacteria are a well-known source of bioproducts. This turns cyanobacterial isolates present in microbial culture collections valuable for biotechnology purposes. Acting as repositories of strains and of their genetic material, culture collections (CC) facilitate the access to cyanobacterial diversity and their associated natural compounds, being able to satisfy the needs of academia or the industry. With this in mind, we decided to organize our cyanobacterial strains into a publicly available culture collection, the LEGE CC. During its implementation, we

surveyed and summarized all relevant information available for the strains, with an emphasis on biotechnological related issues. In this study, we reveal the biodiversity and chemodiversity of LEGE isolates, provide the first version of the strain catalog, and discuss the main findings gained from the data survey. The strains were extensively characterized and taxonomically classified using a polyphasic approach. Currently, LEGE CC comprises 386 cyanobacterial strains belonging to six orders and at least 46 genera, several of them unique among the phylogenetic diversity of the group, and thus taxonomically challenging. The notable biodiversity of the Collection is accompanied by a potentially valuable chemodiversity. Indeed, based on less than half of the existing cyanobacterial LEGE CC strains, several studies have unveiled their (actual or potential) capability to produce a variety of biotechnologically interesting compounds, including toxins or newly discovered bioactive molecules. Additionally, LEGE CC strains also show to be a promising source of extracellular polymeric substances (EPS). Finally, by linking the existing information related to secondary metabolites for LEGE CC strains and their phylogenetic diversity, we put forward some considerations on the use of cyanobacterial strains as a source for natural products bioprospecting.

#### **4.1.2 Introduction**

Microbial Biological Resource Centers (mBRCs) are quality-managed culture collections that ensure the *ex situ* preservation of microorganisms, while providing public access to their microbial diversity (i.e. to live strains or to genomic DNA from these strains), to relevant data related to it (e.g. taxonomic identification, culture conditions, ecophysiological features, etc), and also to expertise services such as training or consulting (Antunes et al. 2016). mBRCs have made numerous and significant contributions to science and society, and still are pivotal in underpinning the bioeconomy derived from microbial resources (Smith et al. 2014). In this particular case, cyanobacteria have been pointed out in the past few decades as one of the most promising groups of microorganisms for the discovery of natural compounds with pharmacological and other biotechnological applications (Margesin et al. 2001; Abed et al. 2009; Singh et al. 2011; Wijffels et al. 2013). For oncology drugs only, the pharmaceutical value of the estimated marine cyanobacteria diversity was evaluated in \$37.5-181.5 billion, in 2010 U.S. dollars (Erwin et al. 2010). One other relevant property of cyanobacteria with biotechnological interest is the production of extracellular polymeric substances (EPSs) (Abed et al. 2009; Pereira et al. 2011).

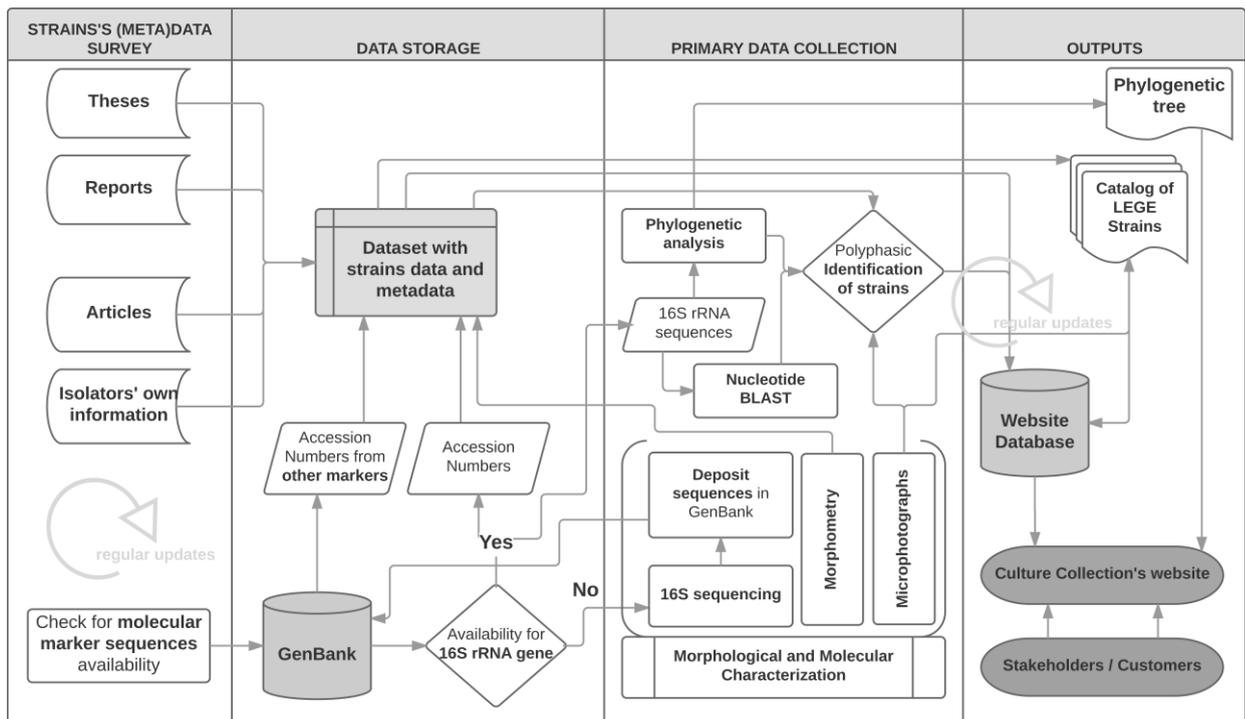
The Blue Biotechnology and Ecotoxicology group (BBE), at CIIMAR, Portugal, has recently undertaken a process of organizing its cyanobacterial isolates into a Culture Collection (acronym LEGE). It began as an in-house collection, in 1991, when a number of strains from the colonial,

toxic cyanobacterium *Microcystis aeruginosa* were isolated from freshwater water bodies, in Portugal (Vasconcelos et al. 1995). Since then, members from the now called BBE group have been isolating and maintaining diverse cyanobacterial isolates that were collected in different ecosystems and locations. Most of these isolates were used by BBE members for their own research. Not surprisingly, a good number of strains have been assessed in ecotoxicological studies or used for the discovery of biologically active compounds, the main research lines of the group. As a consequence, a considerable body of research emphasizes that the strains now deposited at the LEGE Culture Collection (CC) have the potential or actual capacity to produce a myriad of chemical compounds, including toxins or newly discovered bioactive molecules (e.g. Vasconcelos et al. 1995; Martins et al. 2005; Leão et al. 2013a; Martins et al. 2013; Brito et al. 2015). In contrast, some of the isolates at BBE were poorly characterized or even unidentified. Moreover, along these years the strains were kept independently by their isolators and named inconsistently. For that reason, and by recognizing the value of this biological resource and the relevance of making it publicly available, a decision was made to characterize and organize the isolates and establish them into a Culture Collection in accordance to the World Federation for Culture Collections (WFCC) guidelines (WFCC 2010).

In this work, by describing the procedures followed throughout the establishment of the Culture Collection, we present the LEGE CC and disclose its cyanobacterial diversity, including the first version of the catalog of strains. Strains were (re-)identified using a polyphasic approach (Komárek 2016) and their published and unpublished data surveyed. Additionally, we review biotechnologically-relevant information from LEGE CC strains, based on existing and newly obtained data, and make some considerations on the relation between biodiversity and chemodiversity for the discovery of natural compounds from cyanobacterial strains.

### 4.1.3 Material and Methods

Strain codes for all isolates at BBE were standardized by using the acronym LEGE followed by a five-digit number. The workflow followed during the establishment of the Culture Collection is depicted in Fig. 4.1. The figure shows the processes and methods used for researching and collecting secondary data and for generating primary data from the strains, and indicates the main outputs of these processes, which are presented in this study.



**Figure 4.1.** Diagram showing the workflow followed during the data gathering on the LEGE CC strains, the completed and expected outputs of the process and the planned updates (standard flowchart symbols were used). The LEGE CC website can be accessed at <http://lege.ciimar.up.pt>

#### *Literature and data survey*

A good number of isolates had been previously published using other strain names/codes or identifications. For that reason, all existing synonyms for a same strain were considered during the literature search and data survey. Strain synonyms and references where they appear are provided in the catalog (see Appendix A8). Strains having any type of data on natural products were recorded.

#### *Light microscopy and morphological characterization*

Morphological characteristics of LEGE CC strains were examined and microphotographed using a Leica DMLB light microscope coupled to a Leica ICC50 HD digital camera (Leica Microsystems, Wetzlar, Germany). Morphometric measurements were then performed using the image analysis software Leica Application Suite version 4.2.0 (Leica Microsystems). Strains were analyzed during the exponential phase of growth (i.e. two to three-week old cultures, depending on the strain; culture conditions for each strain can be found in the catalog, from Appendix A8). Each quantifiable morphological character was measured at least 20 times, along different positions of the slide preparation. These include size of vegetative, specialized or dormant cells, and of filaments or colonies.

Additionally, to evaluate the production of EPSs by the strains, early stationary-phase cultures (i.e. three to five-week old cultures, depending on the strain) were stained with 0.5% Alcian Blue solutions (Sigma A-3157), prepared either in 50% ethanol (v/v) or in 1% acetic acid (v/v) (Di Pippo et al. 2013). Cultures were also negatively stained using India ink (Micheletti et al. 2008). Images were acquired using the abovementioned equipment and software.

When relevant, other qualitative morphological features and distinguishing traits were recorded (e.g. the shape and arrangement of cells or filaments, the color of the cultures, the presence or absence of sheaths, motility, the existence of constrictions at the cross-wall of filaments, the formation of hormogonia and necridial cells).

#### *DNA extraction, PCR and sequencing*

Cells were harvested from log-phase cultures, and total genomic DNA (gDNA) of each strain was extracted using the commercial PureLink™ Genomic DNA Mini Kit (Invitrogen, Carlsbad, USA), according to the manufacturer's instructions provided for Gram-negative bacteria. The DNA integrity was confirmed with agarose gel electrophoresis using GelRed™ (Biotium, USA) staining. Cyanobacteria-specific primers CYA-106F and CYA-785R (Nübel et al. 1997; Muhling et al. 2008) were used for the amplification of a portion of the 16S rRNA gene. PCR reactions were performed in a final volume of 20 µL containing 1× Green GoTaq® Flexi Buffer, 2.5 mM MgCl<sub>2</sub>, 125.0 mM of each deoxynucleotide triphosphate, 1.0 µM of each primer, 0.5 U of GoTaq® Flexi DNA Polymerase (Promega, USA), 10 mg mL<sup>-1</sup> of bovine serum albumin (BSA) and 10–30 ng of template DNA, on a T-Professional Standard thermal cycler (Biometra, Germany). The PCR conditions were as follows: initial denaturation at 94 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 30 s and extension at 72 °C for 45 s, with a final extension step at 72 °C for 6 min. PCR products were separated with a 1.5% (w/v) agarose gel stained with GelRed™ (Biotium, USA) and DNA fragments with the expected size were excised from the gel and purified using the NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel, Germany), according to the manufacturer's instructions. Sequences were obtained either by directly sequencing the purified amplicons at GATC Biotech (Germany), or after cloning these into pGEM-T® Easy vector (Promega, Madison, WI). In the latter case, vectors containing inserts were then transformed into *Escherichia coli* TOP10 chemically competent cells (Invitrogen, San Diego, CA). Plasmid DNA was isolated using NZY Miniprep kit (NYZtech, Portugal) and sequenced at GATC Biotech using M13 primers. All nucleotide sequences were manually inspected for quality and assembled for each strain using the Geneious (v8.1.8) software package (Biomatters Limited, New Zealand). Novel sequences associated with this study were deposited in the GenBank database under the accession numbers KU951663 - KU951886.

### *Phylogenetic analyses*

Molecular-based analyses were conducted using the bioinformatics software package MEGA7 (Kumar et al. 2016). Two phylogenetic analyses based on 16S rRNA gene sequences were performed, one that reflects the overall cyanobacterial diversity present at LEGE CC and a second analysis that highlights the connection between such biodiversity and its associated chemodiversity. In both cases, sequences were aligned using the ClustalW algorithm (Thompson et al. 1994) and phylogenies were inferred by using the Maximum Likelihood (ML) method (Felsenstein 1981) based on the General Time Reversible model (Rodriguez et al. 1990), which was the nucleotide substitution model that best fitted the alignments data as evaluated by the corrected Akaike Information Criterion (Sugiura 1978). For both analyses also, a discrete Gamma distribution (+G) was used to model evolutionary rate differences among sites, while the rate variation model allowed for some sites to be evolutionarily invariable (+I). In the first case, the analysis involved 457 nucleotide sequences from LEGE CC strains and from: (1) Type strains (T) of Type species (i.e. cyanobacterial strains that were used to describe a new genus); (2) strains known to have the same phylogenetic placement as the Type species (t), when the sequence from the latter is not available; (3) Reference strains (R) from the Bergey's Manual of Systematic Bacteriology (Castenholz et al. 2001); and, (4) strains known to be included in the same phylogenetic cluster as the Reference strain (r), as mentioned in the Bergey's Manual (Castenholz et al. 2001). There was a total of 563 positions in the final dataset. The tree was rooted with the outgroup *Chloroflexus aurantiacus* J-10-fl (NR\_074263). In the second case, the phylogenetic analysis involved 165 nucleotide sequences from LEGE CC strains only and there was a total of 252 positions in the final dataset.

### *Strain identification*

By using data generated in this study, the taxonomic assignments of previously identified strains were reevaluated by a polyphasic methodology, i.e. based on morphological and phylogenetic data. The most recent classification, recommendations and advice for the identification of cyanobacteria (Komárek et al. 2014; Dvořák et al. 2015; Komárek 2016) were followed, namely the adoption of a conservative approach (Dvořák et al. 2015; Komárek 2016). Previously unidentified isolates were identified following the same procedures and principles. First, standard identification keys were used for the morphological-based identification of the strains (Komárek 2013; Komárek & Anagnostidis 1998, 2005). Then, each strain identification was compared with its phylogenetic placement (namely, assessing if the LEGE strain is closely-related to any Type strain), and with the recent taxonomic classification proposed by Komárek and co-authors (2014), at low (i.e. genus) and high (i.e. order) taxonomic levels. If existing,

taxonomic notes for a strain (e.g. incongruities between classification schemes) were added to the correspondent catalog sheet (Appendix A8).

#### 4.1.4 Results and Discussion

Three hundred and eighty-six cyanobacterial strains are included in the first version of the catalog of LEGE CC. For each particular strain, primary and secondary data collected in this study such as species identification, origin, morphometric information, morphological description and ecophysiological properties of the isolate, microphotographs, literature references, synonyms for the strain, accession numbers for sequences, etc., can be retrieved in the corresponding catalog sheet (Appendix A8) or be searched in the website database of the culture collection at <http://lege.ciimar.up.pt> (Appendix D).

##### *Strain characterization and polyphasic identification*

The morphological and molecular characterization exposed the wide diversity of LEGE CC strains (Figs. 4.2 and 4.3), being included in six orders (Fig. 4.3) and 46 genera (Table 4.1). Komárek et al. (2014) have recently proposed a new taxonomy classification for cyanobacteria. Based on phylogenetic systematics these authors have either erected new cyanobacterial orders or redefined the classical ones. For instance, unicellular or colonial cyanobacteria formerly included in the classical order Chroococcales (Komárek & Anagnostidis 1998) are now distributed in the new order Synechococcales and/or in the revisited Pleurocapsales (Komárek et al. 2014). The same is true to filamentous non-heterocystous cyanobacteria, which were traditionally included in the Oscillatoriales (Komárek & Anagnostidis 2005) and are now distributed in the redefined orders Chroococcales or Oscillatoriales (Komárek et al. 2014).

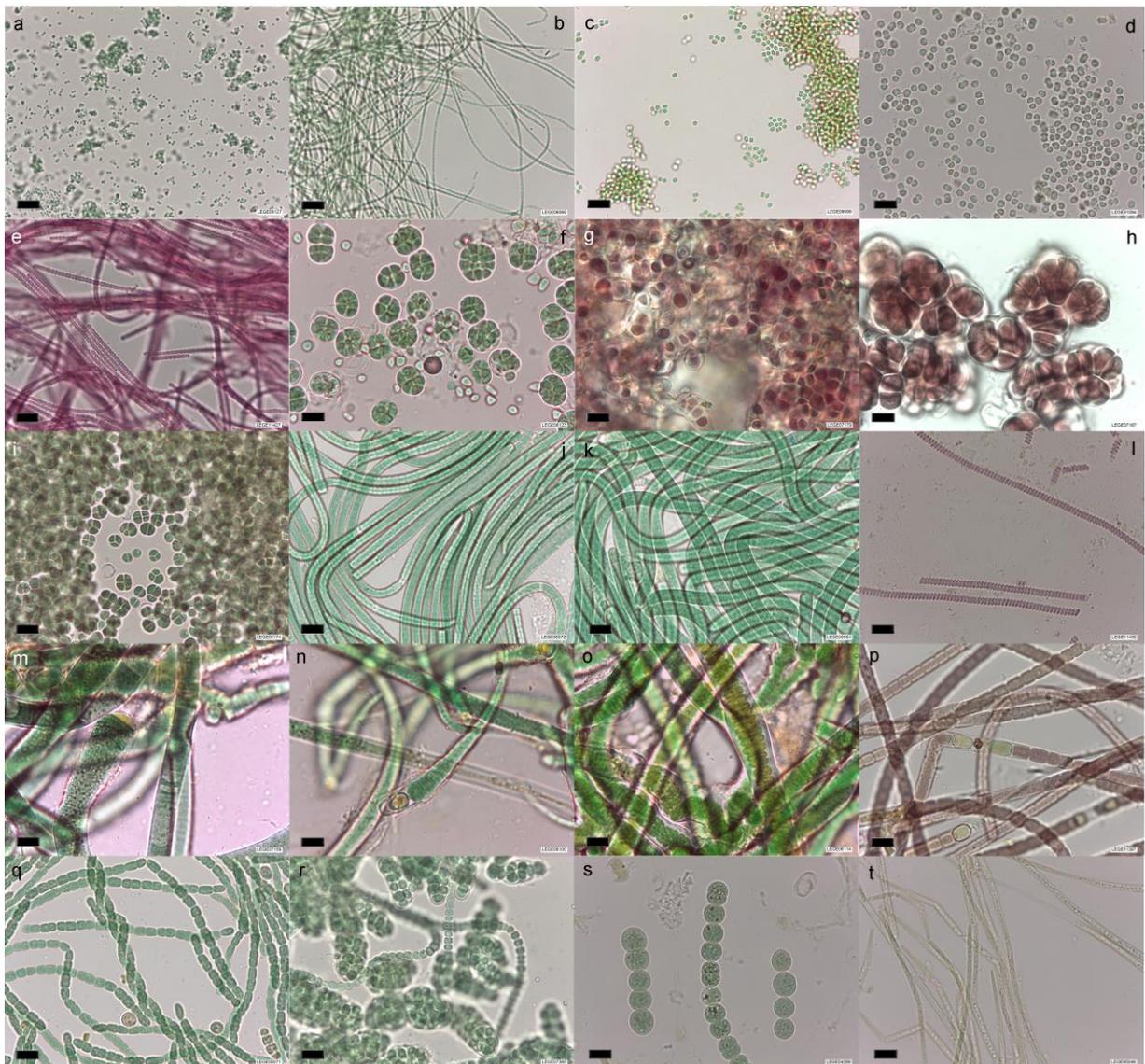
Accordingly, several LEGE CC strains that were previously assigned to those classical orders (e.g. Brito et al. 2012; Lopes et al. 2012) were now re-classified by using this new classification scheme (Komárek et al. 2014) and by assessing their phylogenetic position, as depicted in Fig. 4.3 (also provided in a high quality vector format, suitable to be enlarged; see Appendix A9). For this purpose, sequences from the same Reference strains included in the phylogeny performed by Komárek et al. (2014) were used in our analysis, which has permitted to map out the orders in the phylogenetic tree (Fig. 4.3). The abovementioned grouping of unicellular and filamentous non-heterocystous forms into new orders is illustrated by a selection of LEGE CC strains included in Fig. 4.2. Colonial forms that divide by multiple fission (Fig. 4.2, g-i) and heterocystous strains (Fig. 4.2, m-t) from the LEGE CC were found to be part of the Pleurocapsales and Nostocales clades, respectively (Fig. 4.3).

**Table 4.1** Number of cyanobacterial strains, by genera, in LEGE CC (386 isolates in total).

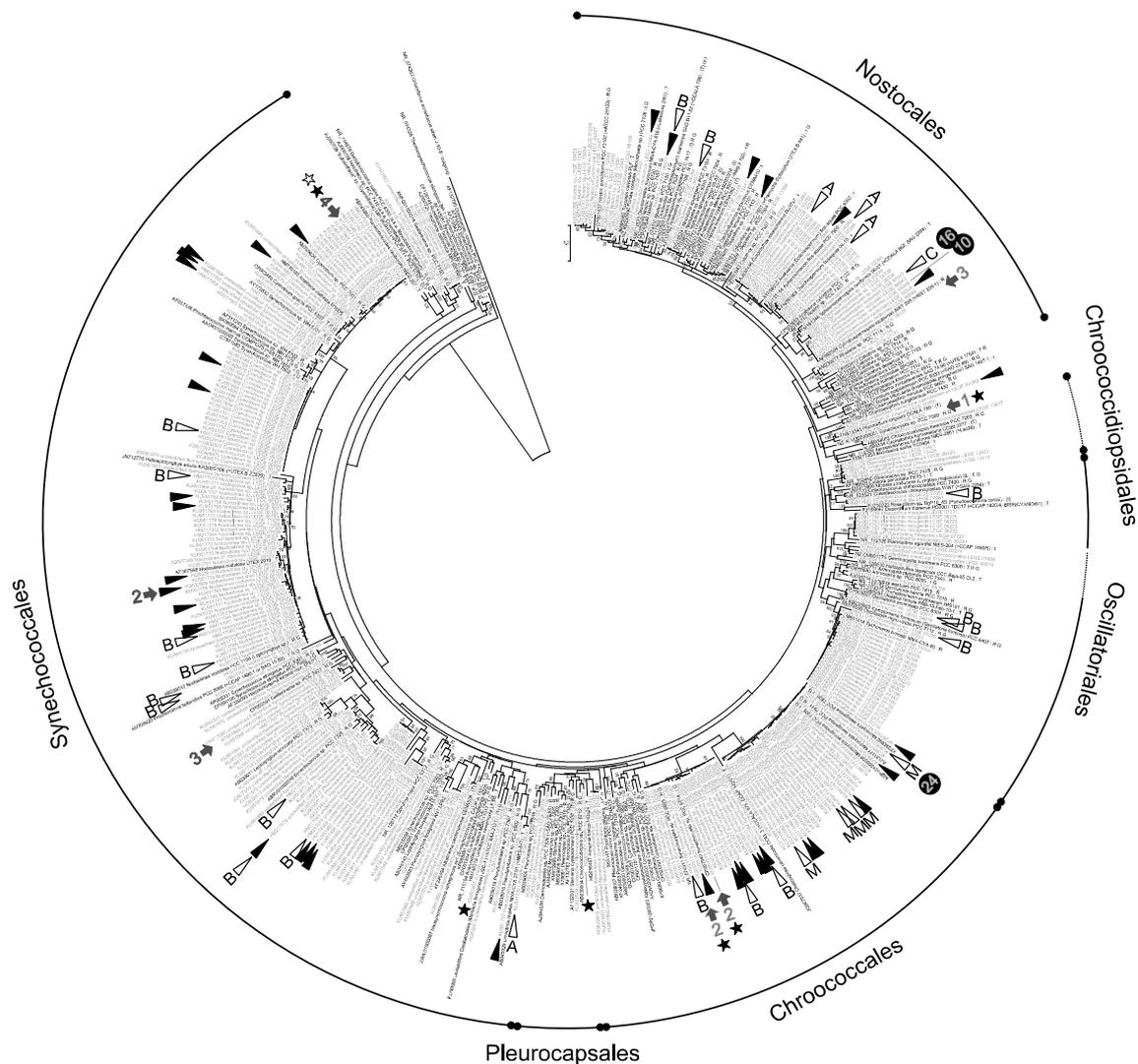
| Genus                     | Number of strains | Genus                              | Number of strains |
|---------------------------|-------------------|------------------------------------|-------------------|
| <i>Alkalinema</i>         | 1                 | <i>Nostoc</i>                      | 14                |
| <i>Anabaena</i>           | 3                 | <i>Oculatella</i>                  | 1                 |
| <i>Calothrix</i>          | 2                 | <i>Oxynema</i>                     | 2                 |
| <i>Chroococcidiopsis</i>  | 1                 | <i>Phormidesmis</i>                | 5                 |
| <i>Chroococcopsis</i>     | 2                 | <i>Phormidium</i>                  | 6                 |
| <i>Chrysosporum</i>       | 1                 | <i>Planktothrix</i>                | 10                |
| <i>Coleofasciculus</i>    | 2                 | <i>Plectonema</i>                  | 2                 |
| <i>Cuspidothrix</i>       | 4                 | <i>Pseudanabaena</i>               | 3                 |
| <i>Cyanobacterium</i>     | 1                 | <i>Rivularia</i>                   | 3                 |
| <i>Cyanobium</i>          | 48                | <i>Roholtiella</i>                 | 1                 |
| <i>Cylindrospermopsis</i> | 6                 | <i>Romeria</i>                     | 4                 |
| <i>Desmonostoc</i>        | 1                 | <i>Schizothrix</i>                 | 1                 |
| <i>Dolichospermum</i>     | 11                | <i>Scytonema</i>                   | 2                 |
| <i>Fortiea</i>            | 1                 | <i>Sphaerospermopsis</i>           | 4                 |
| <i>Geitlerinema</i>       | 4                 | <i>Spirulina</i>                   | 1                 |
| <i>Gloeocapsa-like</i>    | 1                 | <i>Synechococcus</i>               | 12                |
| <i>Gloeocapsopsis</i>     | 4                 | <i>Synechocystis</i>               | 21                |
| <i>Halomicronema</i>      | 1                 | <i>Tolypothrix</i>                 | 1                 |
| <i>Hyella</i>             | 1                 | <i>Tychonema</i>                   | 21                |
| <i>Leptolyngbya</i>       | 14                | unidentified Chroococcidiopsidales | 4                 |
| <i>Limnoraphis</i>        | 1                 | unidentified Chroococcales         | 5                 |
| <i>Limnothrix</i>         | 1                 | unidentified Nostocales            | 9                 |
| <i>Microcoleus</i>        | 1                 | unidentified Oscillatoriales       | 4                 |
| <i>Microcystis</i>        | 37                | unidentified Pleurocapsales        | 2                 |
| <i>Myxosarcina</i>        | 1                 | unidentified Synechococcales       | 32                |
| <i>Nodosilinea</i>        | 44                | unidentified cyanobacterium        | 20                |
| <i>Nodularia</i>          | 2                 |                                    |                   |

Interestingly, the filamentous cyanobacterium *Plectonema* cf. *radiosum* LEGE 06114 (Fig. 4.2, o), which lacks heterocysts and akinetes, exhibits discoid cells and rounded apical cells and shows visible sheaths and double false-branching (Brito et al. 2012), is phylogenetically placed within the Nostocales (Fig. 4.3). *Plectonema* is traditionally classified in the Oscillatoriales as it lacks specialized cells (Komárek & Anagnostidis 2005; Komárek et al. 2014), but its taxonomy is debatable and requires revision (Komárek & Anagnostidis 2005). For instance, as observed with *Plectonema* cf. *radiosum* LEGE 06114, some *Plectonema* species exhibit double false-branching similar to those found in Nostocales genera (e.g. *Scytonema* (Komárek 2013)) and could be transferred to this order according to Komárek and Anagnostidis (2005). Information on these and other (apparent) taxonomic incongruities, peculiarities or doubts that may have

arisen after the polyphasic identification of LEGE CC strains were included in the catalog sheet of the corresponding strain, as taxonomic notes (Appendix A8).



**Figure 4.2.** Example of morphological diversity among cyanobacterial isolates from LEGE CC. Strains belong to the orders: (a-b) Synechococcales; (c-e) Chroococcales; (f) Chroococciopsiales; (g-i) Pleurocapsales; (j-l) Oscillatoriales; (m-t) Nostocales. Identifications are as follows: (a) *Cyanobium* sp. LEGE 06127; (b) *Nodosilinea* sp. LEGE 06069; (c) *Synechocystis salina* LEGE 06099; (d) *Microcystis aeruginosa* LEGE 91094, a microcystin producer strain highly used in the literature (see also Fig. 3); (e) unidentified filamentous Chroococcales LEGE 11427; (f) *Gloeocapsopsis crepidinum* LEGE 06123; (g) *Hyella* sp. LEGE 07179; (h) *Chroococcopsis* sp. LEGE 07187; (i) *Chroococciopsis* sp. LEGE 06174; (j) cf. *Oxynema acuminatum* LEGE 06072; (k) *Phormidium* sp. LEGE 00064; (l) cf. *Spirulina* sp. LEGE 11439; (m) *Rivularia* sp. LEGE 07159; (n) *Calothrix* sp. LEGE 06100; (o) *Plectonema* cf. *radiosum* LEGE 06114; (p) *Tolypothrix* sp. LEGE 11397; (q) *Nodularia* sp. LEGE 06071; (r) *Nostoc* sp. LEGE 07365; (s) *Dolichospermum flosaquae* LEGE 04289, an anatoxin-a producer strain; (t) *Cylandrospermopsis raciborskii* LEGE 95046, a non-cylandrospermopsin producer often used in the literature (see also Fig. 3). Scale bars represent 10 µm.



**Figure 4.3.** Circular ML tree ( $-lnl = 25944.6863$ ) of 16S rRNA gene sequences illustrating the phylogenetic diversity of LEGE CC strains (in gray), their placement at the order level, and some traits or information relevant for biotechnological purposes. 152 sequences from reference material were included to disclose the cyanobacterial “Tree of Life” (T or t stand for type strains designated as representing type species, R or r for reference strains *sensu* Bergey’s Manual; and G for genome sequences available; see also M&M section for details). Accession numbers for all sequences are shown. Only bootstrap support values over 50% are given. Black arrowheads indicate strains capable of producing good amounts of EPS. White arrowheads denote strains producing the following cyanotoxins, as demonstrated by analytical chemistry methods: A - anatoxin-a; B - BMAA; C – cylindrospermopsin; and, M – microcystin. Arrows point to strains used to isolate and elucidate the structure of the following secondary metabolites: 1 - hierridin B; 2 - portoamides; 3 - bartolosides; and, 4 - dehydroabiatic acid. Black stars indicate strains having (or soon will have) their genome sequenced, and the white star stands for a strain that has a submitted patent application. Black circles and numbers within refer to highly used strains and to the number of times they appear in the literature, respectively.

The modern polyphasic approach (Komárek 2016) has and will continue to result in important changes for the taxonomy of cyanobacteria. Traditional genera or species, especially those with little phenotypic differentiation, very often exhibit polyphyly in phylogenetic studies (see Dvořák et al. (2015), for a review). Such findings suggest that extensive taxonomic revisions of those taxa are in need (Komárek et al. 2014; Dvořák et al. 2015; Komárek 2016). As a consequence, the number of new genera that are being described using polyphasic taxonomy is growing

rapidly, many such genera representing earlier entangled, cryptic taxa that have emerged from traditional genera (Dvořák et al. 2015; Komárek 2016). Given the current status of taxonomy, and as implicitly recommended by Dvořák et al. (2015), we have adopted a conservative approach for the identification of LEGE CC strains at low taxonomic levels. The availability and inclusion of sequences from Type strains in the phylogenetic analysis (Fig. 4.3) was essential to accurately identify the strains, namely to ascertain if they could belong to recently proposed genera not covered by the classification keys used (Komárek 2013; Komárek & Anagnostidis 1998, 2005). Therefore, previous morphology-based identifications of the strains were not considered if the phylogeny indicated that the strains belong to such recent genera, or if they were phylogenetically placed away from the holotype in question (i.e. Type strain) (Fig. 4.3). Applying these criteria resulted in 76 LEGE CC strains remaining unidentified since it was not possible to achieve an unequivocal identification at the genus level, even if in most cases it was possible to achieve an assignment at the order level (Table 4.1). On the other hand, 79 strains were identified as belonging to 14 recently-described genera by means of polyphasic taxonomy (see also Table 4.1). These consist of one *Alkalinema* (Vaz et al. 2015), one *Chrysochloris* (Zapomělová et al. 2009), two *Coleofasciculus* (Siegesmund et al. 2008), four *Cuspidothrix* (Rajaniemi et al. 2005), one *Desmonostoc* (Hrouzek et al. 2013), eleven *Dolichospermum* (Wacklin et al. 2009), one *Halomicronema* (Abed et al. 2002), one *Limnoraphis* (Komárek et al. 2013), forty-four *Nodosilinea* (Perkerson et al. 2011), one *Oculatella* (Zammit et al. 2012), two *Oxynema* (Chatchawan et al. 2012), five *Phormidesmis* (Turicchia et al. 2009), one *Roholtiella* (Bohunická et al. 2015), and four *Sphaerospermopsis* (Zapomělová et al. 2009) strains.

Well represented genera at LEGE CC include the picocyanobacterium *Cyanobium* (48 strains; Fig. 4.2a), the filamentous non-heterocystous *Nodosilinea* (44; Fig. 4.2b), the bloom forming *Microcystis* (37, including both microcystin and non-microcystin producers; Fig. 4.2d), the unicellular *Synechocystis* (21) and the filamentous non-heterocystous *Tychonema* (21). In addition to microcystin (MC)-producing *Microcystis aeruginosa* isolates, LEGE CC comprises other toxic strains (Fig. 4.3) such as the anatoxin-a (ANA-a) producers *Dolichospermum* spp. LEGE 00240, 00241 and 04289 and *Limnothrix* sp. LEGE 00237, the cylindrospermopsin (CYN) producer *Cylindrospermopsis raciborskii* LEGE 97047, as well as several strains, belonging to different taxa, that produce  $\beta$ -methylamino-L-alanine (BMAA), a toxin shown to be widespread among cyanobacteria (Cox et al. 2005; Cianca et al. 2012). Besides toxins, other secondary metabolites are known to be produced by LEGE CC strains. Indeed, some of these were used to isolate novel and known bioactive metabolites (Fig. 4.3) (Leão et al. 2010, 2013b, 2015; Costa et al. 2016).

LEGE CC aims to value its cyanobacterial diversity in a way that can be perceived by others, namely by stakeholders from the biotechnology sector. As such, strains are characterized in

order to highlight features that may have interest from an applied point of view. As depicted from the qualitative evaluation made by different staining techniques (Supplementary Data, Fig. 4.7), several LEGE CC strains produce considerable amounts of EPSs (Fig. 4.3), a feature that may have biotechnological applications. For instance, cyanobacterial EPSs can be used for heavy metal removal from contaminated waters (Pereira et al. 2011), as was already demonstrated for one of our strains, *Synechocystis* sp. LEGE 00032 (Ribeiro et al. 2008). Also, six strains (Fig. 4.3) have had their genomes sequenced and these will be made publicly available, following curation. One such strain, *Cyanobium* sp. LEGE 06113, has been included on a submitted patent application for a promising anti-malarial compound. Some strains held in LEGE CC have an earthy odor, something that may indicate the presence of odiferous metabolites such as 2-methylisoborneol or geosmin (Giglio et al. 2010), two compounds that pose problems in drinking water supply systems. This qualitative data was included in the catalog of strains (Appendix A8).

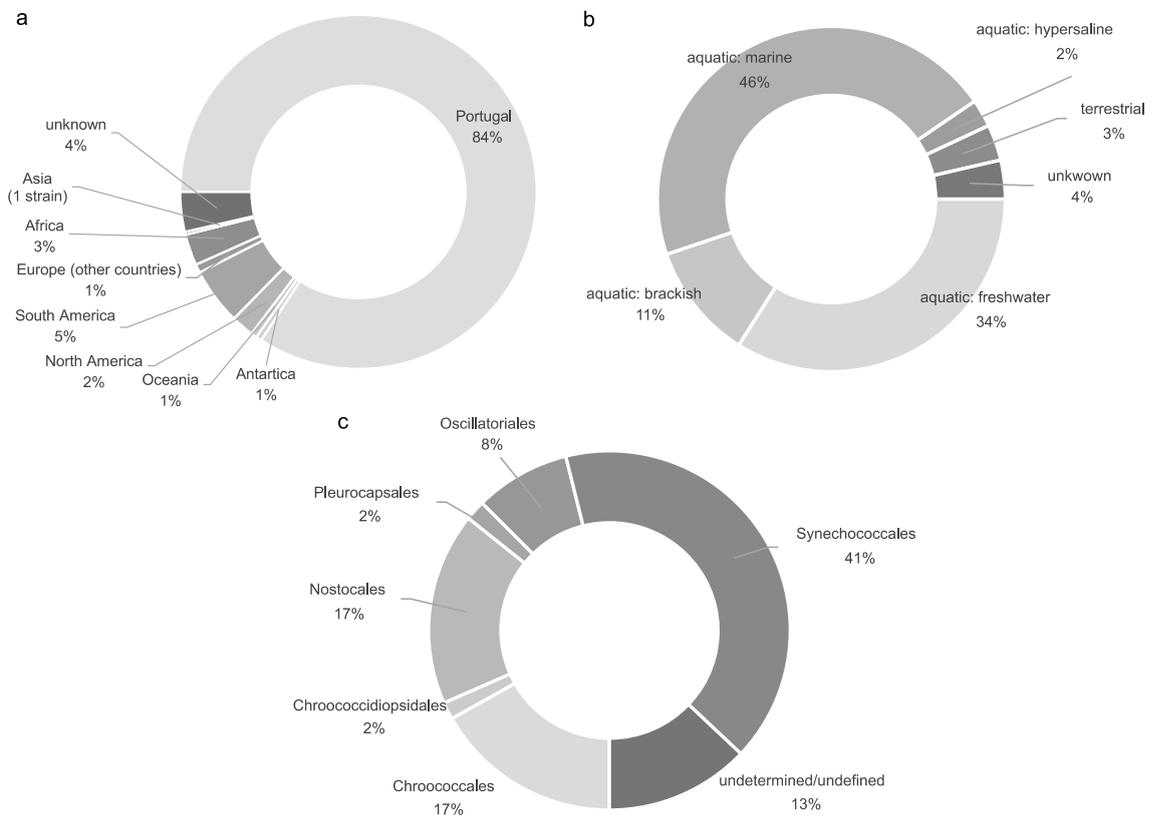
#### *General statistics*

The 386 strains that comprise the first version of the LEGE CC catalog were isolated from samples mainly collected in Portugal (84%), including Madeira and Azores Islands (Fig. 4.4, a). There are also strains from South (5%) and North (2%) America, Africa (3%), from other European countries (1%), Oceania (1%), Antarctica (1%), and Asia (one strain).

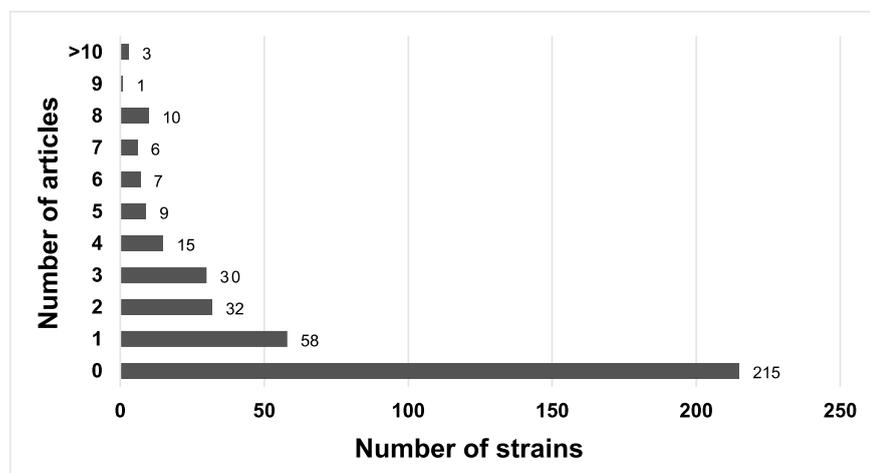
In relation to the habitat, LEGE CC strains were mainly collected from aquatic environments (Fig. 4.4, b), including marine (46%), freshwater (34%), brackish (11%), and hypersaline (2%) environments, while some strains are of terrestrial origin (3%). Concerning taxonomy (Fig. 4.4, c), LEGE CC strains are distributed by the orders Synechococcales (41%), Chroococcales (17%), Nostocales (17%), Oscillatoriales (8%), Pleurocapsales (2%), and Chroococciopsidales (2%) (see also Fig. 4.3).

Several LEGE CC strains have been used in academia, most of them in research related to cyanobacterial natural products, as underlined by data available in the literature (Fig. 4.5). Indeed, in May 2016, 171 strains (44% of the total) had some sort of data available in published journal articles (Fig. 4.5), from which 163 (42%) concerned natural products, including toxins (see also Fig. 4.3 and Fig. 4.6). The available information on LEGE CC strains was found to be disseminated through 98 different journal articles.

The three most frequently reported LEGE CC strains were found to be included in 10 or more journal articles (Figs. 4.3 and 4.5). These are the MC-producing strain *Microcystis aeruginosa* LEGE 91094 (Fig. 4.2, d), the CYN-producing strain *Cylindrospermopsis raciborskii* LEGE 97047, and *Cylindrospermopsis raciborskii* LEGE 95046, a non-CYN producer (Fig. 4.2, t).



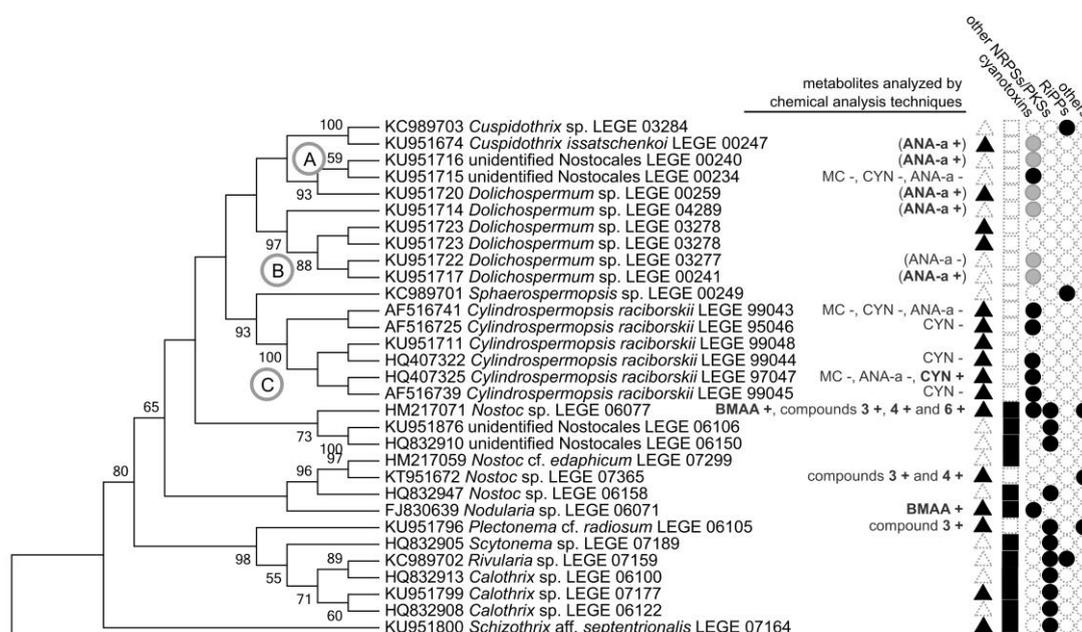
**Figure 4.4.** General statistics for the LEGE CC strains concerning (a) geographical origin, (b) source/type of habitat, and (c) taxonomic classification at the order level.



**Figure 4.5.** Number of LEGE CC strains included in journal articles.

### LEGE CC strains and their (potential) chemodiversity

Since the main research lines of BBE are ecotoxicology and the discovery of new natural products, in particular those with biotechnological potential, it is not surprising that a considerable fraction of LEGE CC strains have been used for their potential production of bioactive secondary metabolites. In fact, 165 cyanobacterial LEGE CC strains, representing 43% of the total number of strains, have some associated data (either published or not) concerning the production of natural products or information on biological activity of their constituents (Fig. 4.6). The phylogenetic relationships among these strains and associated data are depicted in the unrooted tree shown in Fig. 4.6. LEGE CC strains have the potential (e.g. presence of genes involved in the biosynthesis of secondary metabolites) or the effective capacity to produce a myriad of chemical compounds (Fig. 4.6, see also Leão et al. 2013a; Martins et al. 2013; Brito et al. 2015).



**Figure 4.6.** ML cladogram (-lnl = 3431.5512) for 165 LEGE CC strains having available data related to natural products. Capital letters in the tree highlight clades encompassing close-related strains for which the production of some of the following specific metabolites were detected (+) or not (-): Cyanotoxins - ANA-a, anatoxin-a; BMAA,  $\beta$ -Methylamino-L-alanine; CYN, cylindrospermopsin; MC, microcystin. Bioactive compounds - 1, portoamides; 2, bartolosides; 3, dehydroabiatic acid; 4, abietic acid; 5, hierridin B; 6, anabaenopeptins A and D. Notice that the production (+) or absence of production (-) of the different compounds were confirmed by analytical techniques such as HPLC, LC-MS or NMR. Metabolites between parentheses and symbols in gray indicate unpublished data. Symbols indicate the existence of data on: toxicity, bioactivity or allelopathy assays ( $\blacktriangle$ ); screening of metabolites by MALDI-TOF Mass Spectrometry or by LC-MS analysis coupled with molecular networking ( $\blacksquare$ ); cyanotoxins ( $\bullet$ , first column); other than cyanotoxins nonribosomal peptide synthetases, polyketide synthases or hybrid NRPS-PKS ( $\bullet$ , second column); ribosomally synthesized and post-translationally modified peptides ( $\bullet$ , third column); and other family of compounds such as terpenes, glycolipids, etc. ( $\bullet$ , fourth column). To get the data on a particular strain, literature references can be found in the corresponding catalog sheet (Appendix B).



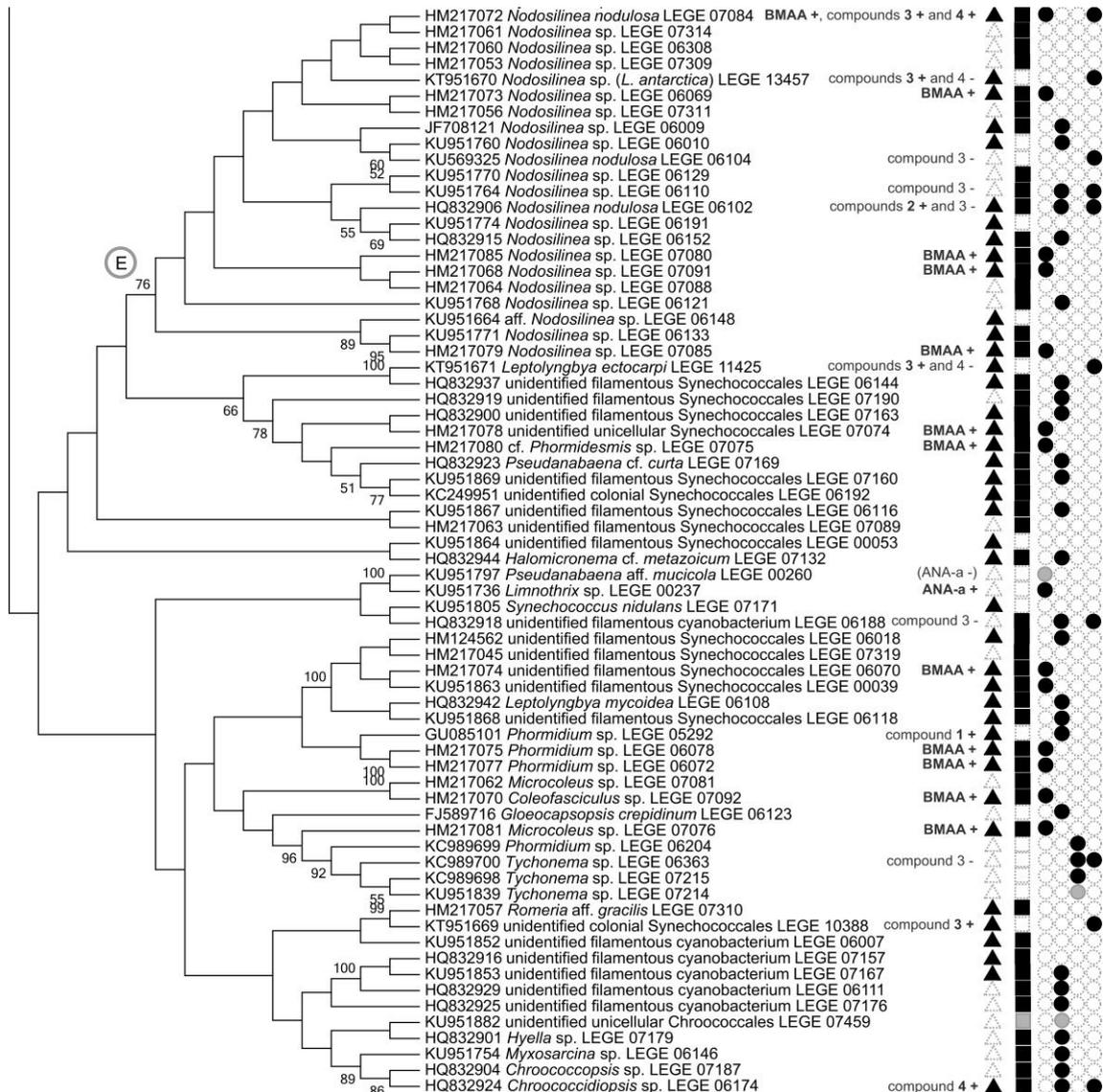


Figure 4.6. (continued)

By screening LEGE CC strains, Costa et al. (2016) soon demonstrated that this and one other terpenoid, the abietic acid, are present in a wide range of cyanobacteria (Fig. 4.6, compounds 3 and 4). In the same study, it was also shown that in some cases the two compounds could not be detected in strains closely related to the diterpenoid-producing cyanobacteria. The same pattern can be observed in different cyanobacterial clades (A-E) highlighted in Fig. 4.6, for different metabolites studied by analytical methods. For instance, regarding toxins, there are closely-related LEGE CC strains assigned as ANA-a producers and non-producers, in clade A and B, CYN producers and non-producers in clade C, and MC producers and non-producers in clade D. Closely-related strains that produced or did not produce the diterpenoids are included in clade E. Of course, a metabolite can remain undetected if it is being produced at low levels, below the limit of detection of the analytical technique. It can also happen that, under some conditions (e.g. lack of environmental stimuli), a cyanobacterium does not produce a particular

metabolite despite possessing the biosynthetic pathway to produce it (Watanabe & Oishi 1985; Boopathi & Ki 2014). It is also possible that the biosynthetic machinery is inactive (e.g. due to gene mutation events) (Leikoski et al. 2012; Vestola et al. 2014). Comparative genomics studies on diverse cyanobacterial taxa have demonstrated that closely related strains (i.e. at the subspecies level) may present high levels of genome divergence (Rocap et al. 2003; Shih et al. 2013; Bombar et al. 2014; Calteau et al. 2014), containing or not functionally active genes (or gene clusters) linked to the production of natural products (Shih et al. 2013; Calteau et al. 2014; Sinha et al. 2014; Dittmann et al. 2015). All these issues have important implications for the discovery of natural compounds from cyanobacteria. In particular, these indicate that, for an exploration of the full potential of these microorganisms as a source of natural products, bioprospection should be ideally conducted strain by strain rather than taxonomically- or phylogenetically-guided (Dittmann et al. 2015).

#### 4.1.5 Conclusions

As a result of the herein reported establishment of a culture collection, LEGE CC is now a member of the World Federation for Culture Collections (WFCC-WDCM #1089), and is committed to the best-practices described in WFCC guidelines (WFCC, 2010). It is now also part of the European Marine Biological Resource Centre (EMBRC), a pan-European infrastructure for marine biological sciences.

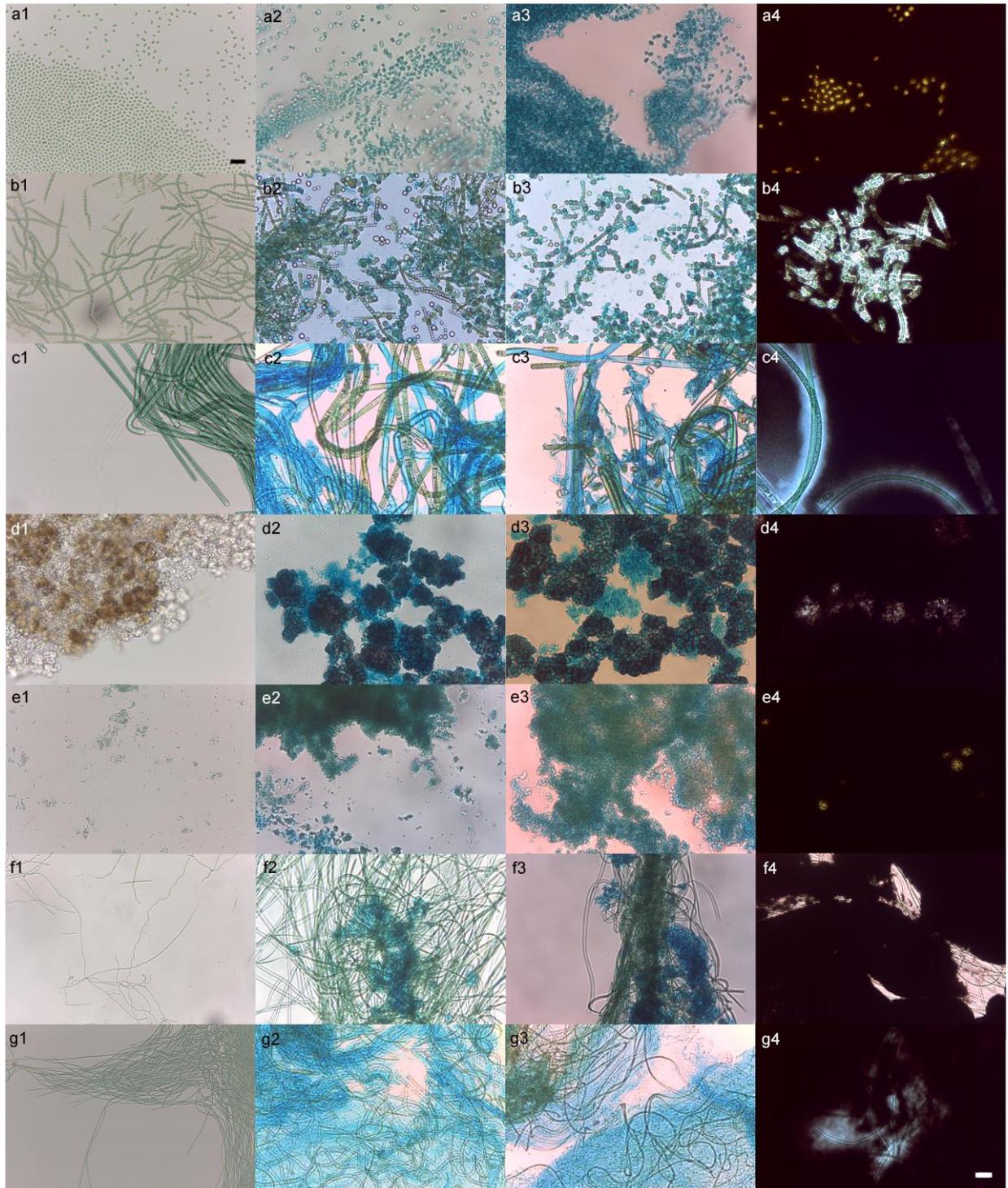
The cyanobacterial diversity that currently makes up the LEGE CC is an increasingly important biological resource, either from the taxonomic point of view or from a biotechnological perspective. Its diversity can be searched through the website of the collection, at <http://lege.ciimar.up.pt/> (see Appendix D). Possible biotechnological applications for LEGE CC strains and their bioproducts were described in several studies (e.g. Lopes et al. 2011; Leão et al. 2013a, 2013b; Costa et al. 2013; Almeida et al. 2015; Costa et al. 2015) and include anti-cancer, anti-viral, anti-microbial, or anti-biofouling capabilities. Even though the use of phylogenetic diversity data is a good way for directing the strains selection for natural products screening, we clearly illustrate that natural product discovery programs should consider a strain-by-strain assessment.

As happens with other mBRCs, LEGE CC is devoted to adding value to its biological resources and providing reliable biological material and associated information to potential stakeholders. We set the path for such goals in this study, by beginning to link the biodiversity and chemical diversity of LEGE CC strains. Soon, at least six strains will have their genomes sequenced, annotated and made publicly available and one strain is featured in a patent application for anti-malarial compounds.

#### **4.1.6 Acknowledgements**

VR gratefully acknowledges financial support from a Fundação para a Ciência e a Tecnologia (FCT) fellowship SFRH/BD/80153/2011. The Project NOMORFILM has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 634588. Structured Program of R&D&I INNOVMAR - Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER-000035, Research Line NOVELMAR), funded by the Northern Regional Operational Program (NORTE2020) through the European Regional Development Fund (ERDF). PNL acknowledges funding from FCT through contract (IF/01358/2014).

### 4.1.7 Supplementary Data



**Figure 4.7.** Examples of strains evaluated for EPSs production. Rows (a-g) refer to different isolates, and columns (1-4) to different types of staining techniques. (a) *Microcystis aeruginosa* LEGE 91353; (b) *Nostoc* sp. LEGE 06077; (c) *Planktothrix mougeotii* LEGE 06225; (d) *Myxosarcina* sp. LEGE 06146; (e) *Synechococcus nidulans* LEGE 07171; (f) unidentified filamentous *Synechococcales* LEGE 06018; (g) unidentified filamentous cyanobacterium LEGE 00060. (1) no staining; (2) stained with 0,5% Alcian Blue (w/v) in 50% ethanol; (3) stained with 0,5% Alcian Blue (w/v) in acetic acid; (4) negative staining with India ink. Scale bars represent 20 µm.

#### 4.1.8 References

- ABED, R., GARCIA-PICHEL, F. & HERNÁNDEZ-MARINÉ, M. 2002. Polyphasic characterization of benthic, moderately halophilic, moderately thermophilic cyanobacteria with very thin trichomes and the proposal of *Halomicronema excentricum* gen. nov., sp. nov. *Archives of Microbiology*, 177, 361-370.
- ABED, R. M. M., DOBRETISOV, S. & SUDESH, K. 2009. Applications of cyanobacteria in biotechnology. *Journal of Applied Microbiology*, 106, 1-12.
- AFONSO, T. B., COSTA, M. S., REZENDE DE CASTRO, R., FREITAS, S., SILVA, A., SCHNEIDER, M. P. C., MARTINS, R. & LEÃO, P. N. 2016. Bartolosides E–K from a marine coccoid cyanobacterium. *Journal of Natural Products*, 79, 2504-2513.
- ALMEIDA, J., FREITAS, M., CRUZ, S., LEÃO, P., VASCONCELOS, V. & CUNHA, I. 2015. Acetylcholinesterase in biofouling species: characterization and mode of action of cyanobacteria-derived antifouling agents. *Toxins*, 7, 2739-2756.
- ANTUNES, A., STACKEBRANDT, E. & LIMA, N. 2016. Fueling the bio-economy: European culture collections and microbiology education and training. *Trends in Microbiology*, 24, 77-79.
- BOHUNICKÁ, M., PIETRASIAK, N., JOHANSEN, J. R., GÓMEZ, E. B., HAUER, T., GAYSINA, L. A. & LUKEŠOVÁ, A. 2015. *Roholtiella*, gen. nov. (Nostocales, Cyanobacteria)—a tapering and branching cyanobacteria of the family Nostocaceae. *Phytotaxa*, 197, 84.
- BOMBAR, D., HELLER, P., SANCHEZ-BARACALDO, P., CARTER, B. J. & ZEHR, J. P. 2014. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *The ISME Journal*, 8, 2530-2542.
- BOOPATHI, T. & KI, J.-S. 2014. Impact of environmental factors on the regulation of cyanotoxin production. *Toxins*, 6, 1951-1978.
- BRITO, Â., GAIFEM, J., RAMOS, V., GLUKHOV, E., DORRESTEIN, P. C., GERWICK, W. H., VASCONCELOS, V. M., MENDES, M. V. & TAMAGNINI, P. 2015. Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*, 9, 218-226.
- BRITO, Â., RAMOS, V., SEABRA, R., SANTOS, A., SANTOS, C. L., LOPO, M., FERREIRA, S., MARTINS, A., MOTA, R., FRAZÃO, B., MARTINS, R., VASCONCELOS, V. & TAMAGNINI, P. 2012. Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Systematic and Applied Microbiology*, 35, 110-119.
- CALTEAU, A., FEWER, D. P., LATIFI, A., COURSI, T., LAURENT, T., JOKELA, J., KERFELD, C. A., SIVONEN, K., PIEL, J. & GUGGER, M. 2014. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics*, 15, 977.
- CASTENHOLZ, R. W., WILMOTTE, A., HERDMAN, M., RIPPKA, R., WATERBURY, J. B., ITEMAN, I. & HOFFMANN, L. 2001. Phylum BX. Cyanobacteria. In *Bergey's Manual of Systematic Bacteriology, Volume One: The Archaea and the Deeply Branching and Phototrophic Bacteria*, D. R. Boone, R. W. Castenholz, G. M. Garrity (eds.), Springer, pp. 473-599.
- CHATCHAWAN, T., KOMÁREK, J., STRUNECKÝ, O., ŠMARDA, J. & PEERAPORNPIHAL, Y. 2012. *Oxynema*, a new genus separated from the genus *Phormidium* (Cyanophyta). *Cryptogamie, Algologie*, 33, 41-59.
- CIANCA, R. C. C., BAPTISTA, M. S., LOPES, V. R. & VASCONCELOS, V. M. 2011. The non-protein amino acid  $\beta$ -N-methylamino-L-alanine in Portuguese cyanobacterial isolates. *Amino Acids*, 42, 2473-2479.
- COSTA, M., GARCIA, M., COSTA-RODRIGUES, J., COSTA, M., RIBEIRO, M., FERNANDES, M., BARROS, P., BARREIRO, A., VASCONCELOS, V. & MARTINS, R. 2013. Exploring bioactive properties of marine cyanobacteria isolated from the Portuguese coast: high potential as a source of anticancer compounds. *Marine Drugs*, 12, 98-114.
- COSTA, M. S., COSTA, M., RAMOS, V., LEÃO, P. N., BARREIRO, A., VASCONCELOS, V. & MARTINS, R. 2015. Picocyanobacteria from a clade of marine cyanobium revealed bioactive potential against microalgae, bacteria, and marine invertebrates. *Journal of Toxicology and Environmental Health, Part A*, 78, 432-442.
- COSTA, M. S., REGO, A., RAMOS, V., AFONSO, T. B., FREITAS, S., PRETO, M., LOPES, V., VASCONCELOS, V., MAGALHÃES, C. & LEÃO, P. N. 2016. The conifer biomarkers dehydroabietic and abietic acids are widespread in Cyanobacteria. *Scientific Reports*, 6, 23436.
- COX, P. A., BANACK, S. A., MURCH, S. J., RASMUSSEN, U., TIEN, G., BIDIGARE, R. R., METCALF, J. S., MORRISON, L. F., CODD, G. A. & BERGMAN, B. 2005. Diverse taxa of cyanobacteria

- produce  $\beta$ -N-methylamino-L-alanine, a neurotoxic amino acid. *Proceedings of the National Academy of Sciences*, 102, 5074-5078.
- DI PIPPO, F., ELLWOOD, N. T. W., GISMONDI, A., BRUNO, L., ROSSI, F., MAGNI, P. & DE PHILIPPIS, R. 2013. Characterization of exopolysaccharides produced by seven biofilm-forming cyanobacterial strains for biotechnological applications. *Journal of Applied Phycology*, 25, 1697-1708.
- DITTMANN, E., GUGGER, M., SIVONEN, K. & FEWER, D. P. 2015. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends in Microbiology*, 23, 642-652.
- DVOŘÁK, P., POULÍČKOVÁ, A., HAŠLER, P., BELLI, M., CASAMATTA, D. A. & PAPINI, A. 2015. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. *Biodiversity and Conservation*, 24, 739-757.
- ERWIN, P. M., LÓPEZ-LEGENTIL, S. & SCHUHMANN, P. W. 2010. The pharmaceutical value of marine biodiversity for anti-cancer drug discovery. *Ecological Economics*, 70, 445-451.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17, 368-376.
- HROUZEK, P., LUKESOVA, A., MARES, J. & VENTURA, S. 2013. Description of the cyanobacterial genus *Desmonostoc* gen. nov. including *D. muscorum* comb. nov. as a distinct, phylogenetically coherent taxon related to the genus *Nostoc*. *Fottea*, 201-213.
- KOMÁREK, J. 2013. Cyanoprokaryota. 3. Teil/3rd Part: Heterocytous Genera, In *Süßwasserflora von Mitteleuropa Bd. 19/2*, Büdel, B., Gärtner, G., Krienitz, L. & Schager, I M. (eds.), Springer Spektrum. p. 1131.
- KOMÁREK, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, 51, 346-353.
- KOMÁREK, J. & ANAGNOSTIDIS, K. 1998. Cyanoprokaryota, 1. Teil/1st Part: Chroococcales, In *Süßwasserflora von Mitteleuropa, Bd 19/1*, Ettl, H., Gärtner, G., Heynig, H. & Mollenhauer, D. (eds.), Gustav Fischer Verlag. p. 548.
- KOMÁREK, J. & ANAGNOSTIDIS, K. 2005. Cyanoprokaryota. 2. Teil/2nd Part: Oscillatoriales, In *Süßwasserflora von Mitteleuropa Bd. 19/2*, Büdel, B., Krienitz, L., Gärtner, G. & Schagerl, M. (eds.), Elsevier/Spektrum. p. 759.
- KOMÁREK, J., ZAPOMELOVA, E., SMARDA, J., KOPECKY, J., REJMANKOVA, E., WOODHOUSE, J., NEILAN, B. A. & KOMARKOVA, J. 2013. Polyphasic evaluation of *Limnographis robusta*, a water-bloom forming cyanobacterium from Lake Atitlan, Guatemala, with a description of *Limnographis* gen. nov. *Fottea*, 39-52.
- KOMÁREK, J., KASTOVSKY, J., MARES, J. & JOHANSEN, J. R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870-1874.
- LEÃO, P. N., PEREIRA, A. R., LIU, W. T., NG, J., PEVZNER, P. A., DORRESTEIN, P. C., KONIG, G. M., VASCONCELOS, V. M. & GERWICK, W. H. 2010. Synergistic allelochemicals from a freshwater cyanobacterium. *Proceedings of the National Academy of Sciences*, 107, 11183-11188.
- LEÃO, P., RAMOS, V., GONÇALVES, P., VIANA, F., LAGE, O., GERWICK, W. & VASCONCELOS, V. 2013. Chemoecological screening reveals high bioactivity in diverse culturable Portuguese marine cyanobacteria. *Marine Drugs*, 11, 1316-1335.
- LEÃO, P. N., COSTA, M., RAMOS, V., PEREIRA, A. R., FERNANDES, V. C., DOMINGUES, V. F., GERWICK, W. H., VASCONCELOS, V. M. & MARTINS, R. 2013. Antitumor activity of Hierridin B, a cyanobacterial secondary metabolite found in both filamentous and unicellular marine strains. *PLoS ONE*, 8, e69562.
- LEÃO, P. N., NAKAMURA, H., COSTA, M., PEREIRA, A. R., MARTINS, R., VASCONCELOS, V., GERWICK, W. H. & BALSUS, E. P. 2015. Biosynthesis-assisted structural elucidation of the bartolosides, chlorinated aromatic glycolipids from cyanobacteria. *Angewandte Chemie International Edition*, 54, 11063-11067.
- LEIKOSKI, N., FEWER, D. P., JOKELA, J., ALAKOSKI, P., WAHLSTEN, M. & SIVONEN, K. 2012. Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus *Microcystis*. *PLoS ONE*, 7, e43002.
- LOPES, V. R., RAMOS, V., MARTINS, A., SOUSA, M., WELKER, M., ANTUNES, A. & VASCONCELOS, V. M. 2012. Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries. *Marine Environmental Research*, 73, 7-16.

- LOPES, V. R., SCHMIDTKE, M., HELENA FERNANDES, M., MARTINS, R. & VASCONCELOS, V. 2011. Cytotoxicity in L929 fibroblasts and inhibition of herpes simplex virus type 1 Kupka by estuarine cyanobacteria extracts. *Toxicology in Vitro*, 25, 944-950.
- MARGESIN, R. & SCHINNER, F. 2001. Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles*, 5, 73-83.
- MARTINS, J., LEÃO, P., RAMOS, V. & VASCONCELOS, V. 2013. N-terminal protease gene phylogeny reveals the potential for novel cyanobactin diversity in cyanobacteria. *Marine Drugs*, 11, 4902-4916.
- MARTINS, R., PEREIRA, P., WELKER, M., FASTNER, J. & VASCONCELOS, V. M. 2005. Toxicity of culturable cyanobacteria strains isolated from the Portuguese coast. *Toxicon*, 46, 454-464.
- MICHELETTI, E., PEREIRA, S., MANNELLI, F., MORADAS-FERREIRA, P., TAMAGNINI, P. & DE PHILIPPIS, R. 2008. Sheathless mutant of cyanobacterium *Gloeotheca* sp. strain pcc 6909 with increased capacity to remove copper ions from aqueous solutions. *Applied and Environmental Microbiology*, 74, 2797-2804.
- MÜHLING, M., WOOLVEN-ALLEN, J., MURRELL, J. C. & JOINT, I. 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *The ISME Journal*, 2, 379-392.
- NÜBEL, U., GARCIA-PICHEL, F. & MUYZER, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, 63, 3327-3332.
- OSSWALD, J., RELLÁN, S., GAGO-MARTINEZ, A. & VASCONCELOS, V. 2009. Production of anatoxina by cyanobacterial strains isolated from Portuguese fresh water systems. *Ecotoxicology*, 18, 1110-1115.
- PEREIRA, S., MICHELETTI, E., ZILLE, A., SANTOS, A., MORADAS-FERREIRA, P., TAMAGNINI, P. & DE PHILIPPIS, R. 2010. Using extracellular polymeric substances (EPS)-producing cyanobacteria for the bioremediation of heavy metals: do cations compete for the EPS functional groups and also accumulate inside the cell? *Microbiology*, 157, 451-458.
- PERKERSON III, R. B., JOHANSEN, J. R., KOVÁČIK, L., BRAND, J., KAŠTOVSKÝ, J. & CASAMATTA, D. A. 2011. A unique pseudanabaenalean (cyanobacteria) genus *Nodosilinea* gen. nov. based on morphological and molecular data. *Journal of Phycology*, 47, 1397-1412.
- RAJANIEMI, P., KOMÁREK, J., WILLAME, R., HROUZEK, P., KAŠTOVSKÁ, K., HOFFMANN, L. & SIVONEN, K. 2005. Taxonomic consequences from the combined molecular and phenotype evaluation of selected *Anabaena* and *Aphanizomenon* strains. *Algological Studies*, 117, 371-391.
- RIBEIRO, J. P. N., LOPES, J. A. & FACHINI, A. 2008. Cleansing contaminated seawaters using marine cyanobacteria: evaluation of trace metal removal from the medium. *International Journal of Environmental Analytical Chemistry*, 88, 701-710.
- ROCAP, G., LARIMER, F. W., LAMERDIN, J., MALFATTI, S., CHAIN, P., AHLGREN, N. A., ARELLANO, A., COLEMAN, M., HAUSER, L., HESS, W. R., JOHNSON, Z. I., LAND, M., LINDELL, D., POST, A. F., REGALA, W., SHAH, M., SHAW, S. L., STEGLICH, C., SULLIVAN, M. B., TING, C. S., TOLONEN, A., WEBB, E. A., ZINSER, E. R. & CHISHOLM, S. W. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, 424, 1042-1047.
- RODRÍGUEZ, F., OLIVER, J. L., MARÍN, A. & MEDINA, J. R. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, 142, 485-501.
- SAKER, M. L. & EAGLESHAM, G. K. 1999. The accumulation of cylindrospermopsin from the cyanobacterium *Cylindrospermopsis raciborskii* in tissues of the Redclaw crayfish *Cherax quadricarinatus*. *Toxicon*, 37, 1065-1077.
- SHIH, P. M., WU, D., LATIFI, A., AXEN, S. D., FEWER, D. P., TALLA, E., CALTEAU, A., CAI, F., TANDEAU DE MARSAC, N., RIPPKA, R., HERDMAN, M., SIVONEN, K., COURSI, T., LAURENT, T., GOODWIN, L., NOLAN, M., DAVENPORT, K. W., HAN, C. S., RUBIN, E. M., EISEN, J. A., WOYKE, T., GUGGER, M. & KERFELD, C. A. 2012. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, 110, 1053-1058.
- SIEGESMUND, M. A., JOHANSEN, J. R., KARSTEN, U. & FRIEDL, T. 2008. *Coleofasciculus* gen. nov. (cyanobacteria): morphological and molecular criteria for revision of the genus *Microcoleus* Gomont. *Journal of Phycology*, 44, 1572-1585.
- SINGH, R. K., TIWARI, S. P., RAI, A. K. & MOHAPATRA, T. M. 2011. Cyanobacteria: an emerging source for drug discovery. *The Journal of Antibiotics*, 64, 401-412.
- SINHA, R., PEARSON, L. A., DAVIS, T. W., MUENCHHOFF, J., PRATAMA, R., JEX, A., BURFORD, M. A. & NEILAN, B. A. 2014. Comparative genomics of *Cylindrospermopsis raciborskii* strains with differential toxicities. *BMC Genomics*, 15, 83.

- SMITH, D., MCCLUSKEY, K. & STACKEBRANDT, E. 2014. Investment into the future of microbial resources: culture collection funding models and BRC business plans for biological resource centres. *SpringerPlus*, 3, 81.
- SUGIURA, N. 1978. Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7, 13-26.
- THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680.
- TURICCHIA, S., VENTURA, S., KOMÁRKOVÁ, J. & KOMÁREK, J. 2009. Taxonomic evaluation of cyanobacterial microflora from alkaline marshes of northern Belize. 2. Diversity of oscillatoriacean genera. *Nova Hedwigia*, 89, 165-200.
- VASCONCELOS, V. M., SIVONEN, K., EVANS, W. R., CARMICHAEL, W. W. & NAMIKOSHI, M. 1995. Isolation and characterization of microcystins (heptapeptide hepatotoxins) from Portuguese strains of *Microcystis aeruginosa* Kutz. emend Elekin. *Archiv fuer Hydrobiologie*, 134, 295-305.
- VAZ, M. G. M. V., GENUARIO, D. B., ANDREOTE, A. P. D., MALONE, C. F. S., SANT'ANNA, C. L., BARBIERO, L. & FIORE, M. F. 2014. *Pantanalinema* gen. nov. and *Alkalinema* gen. nov.: novel pseudanabaenacean genera (Cyanobacteria) isolated from saline-alkaline lakes. *International Journal of Systematic and Evolutionary Microbiology*, 65, 298-308.
- VESTOLA, J., SHISHIDO, T. K., JOKELA, J., FEWER, D. P., AITIO, O., PERMI, P., WAHLSTEN, M., WANG, H., ROUHIAINEN, L. & SIVONEN, K. 2014. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. *Proceedings of the National Academy of Sciences*, 111, E1909-E1917.
- WACKLIN, P., HOFFMANN, L. & KOMAREK, J. 2009. Nomenclatural validation of the genetically revised cyanobacterial genus *Dolichospermum* (Ralfs ex Bornet et Flahault) comb. nova. *Fottea*, 9, 59-64.
- WATANABE, M. F. & OISHI, S. 1985. Effects of environmental factors on toxicity of a cyanobacterium (*Microcystis aeruginosa*) under culture conditions. *Applied and Environmental Microbiology*, 49, 1342-1344.
- WFCC, 2010. *Guidelines for the Establishment and Operation of Collections of Cultures of Microorganisms*, World Federation for Culture Collections, p. 19.
- WIJFFELS, R. H., KRUSE, O. & HELLINGWERF, K. J. 2013. Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Current Opinion in Biotechnology*, 24, 405-413.
- ZAMMIT, G., BILLI, D. & ALBERTANO, P. 2012. The subaerophytic cyanobacterium *Oculatella subterranea* (Oscillatoriales, Cyanophyceae) gen. et sp. nov.: a cytomorphological and molecular description. *European Journal of Phycology*, 47, 341-354.
- ZAPOMĚLOVÁ, E., JEZBEROVÁ, J., HROUZEK, P., HISEM, D., ŘEHÁKOVÁ, K. & KOMÁRKOVÁ, J. 2009. Polyphasic characterization of three strains of *Anabaena reniformis* and *Aphanizomenon aphanizomenoides* (cyanobacteria) and their reclassification to *Sphaerospermum* gen. nov. (incl. *Anabaena kisseleviana*). *Journal of Phycology*, 45, 1363-1373.

## 5. General Discussion

Prior to the introduction of molecular methods, much of what was known about the diversity and distribution of cyanobacteria was derived from microscopy-based studies, involving the observation of either natural samples or laboratory-grown cultures. However, most strains are brought into culture after a strong selection process during culture enrichment and isolation, and often do not represent the natural population of origin. The combination of traditional and molecular techniques, also known as polyphasic approach, is now the preferred approach in the description of cyanobacterial communities (Garcia-Pichel 2008) or isolates (Komárek 2016). Though being increasingly adopted, the implementation of this approach is not free from difficulties (Lee et al. 2014). For instance, by using BLASTn, special care should be given to the existence of incorrectly identified 16S rRNA gene sequences in the public repositories. On the other hand, there are also an increasingly number of unidentified sequences, most of them deriving from culture-independent studies.

In the future, in parallel with a better knowledge of ecological characteristics and geographic distribution of cyanobacteria, the application of genome based analyses will be instrumental for improving current identification methods and classification systems. Recently, technological innovation driven by genome programs led to substantial decreases in the cost of DNA sequencing, and to significant increases in processing speed (Goodwin et al. 2016). This has resulted in several thousand complete bacterial genomes sequenced. For cyanobacteria (e.g. Shih et al. 2013), the number has already exceeded 370, although many are from phylogenetically closely related picocyanobacteria (see subsection 2.1). It's been more than 10 years since Konstantinidis and Tiedje (2005) published their seminal paper on average nucleotide identity (ANI), a whole-genome approach to delineate archaeal and bacterial species. ANI has received wide attention from the bacteriologist community and its applicability to the taxonomy of these organisms is becoming increasingly evident (Arahal 2014). It is therefore envisaged that genome-based taxonomy will be the standard for cyanobacteria, too. These advances in sequencing technology have also greatly impacted the field of environmental microbiology (or microbial ecology). For instance, NGS technologies are now very commonly used to study the taxonomic diversity and metabolic potential of microbial communities, in metagenomics studies (Escobar-Zepeda et al. 2015; Zhou et al. 2015; see also subsection 2.3). The bottleneck has shifted from the sequencing step to other methodological aspects, namely “big data” processing, integration and analysis, and the lack of reliable, reference sequences for identification (Mizrahi-Man et al. 2013; Nguyen et al. 2016). While the first problem is being

overcome by developing and improving bioinformatics tools and computational algorithms for NGS data analysis, the second obstacle should be surmounted by the recognition and incorporation of sequences from more and more diverse reference strains in dedicated databases (e.g. SILVA, Greengenes or the Ribosomal Database Project) (see also subsection 2.3). Here again, the importance of a list of relevant, taxonomy-curated cyanobacterial strains, with information on the availability of genome sequences (subsection 2.1), and of recognized Type strains deposited in Culture Collections (subsections 2.2. and section 4).

It was supposed (Pace 1997) that molecular-based methods, whose contribution for better understanding of microbial diversity is indisputable, would ultimately substitute the need for isolation of cultures. However, as stated by Garcia-Pichel (2008), there is a “need for cultivated anchors [...] in the process of determining [cyanobacterial] diversity, distribution, and function in the natural environment”. At least for cyanobacteria, rapidly it became clear that it is by the existence of isolates that it is possible to make ecological inferences from molecular sequences data. Molecular data have to correlate with identities or functional properties established on isolates themselves (Garcia-Pichel 2008). In this NGS era, the relevance of such isolates that could be used as anchoring points for interpretation of eDNA sequences is even more significant. As is also shown in subsection 2.3 (e.g. Fig. 2.19 and Appendix A2), there is an increasing number of unnamed branches in the cyanobacterial tree of life, for which, having no cultivated isolates, we had no idea about their identification or possible role in nature (Garcia-Pichel 2008).

The study of isolated strains, now included in LEGE CC (section 4), led to a fuller comprehension of the true diversity of cyanobacterial life present in temperate, intertidal habitats (subsection 2.2; see also Brito et al. 2012), and in tropical, coastal lagoons mats (subsection 2.3). The further characterization of these strains has either put in evidence their ecological relevance as N<sub>2</sub>-fixers (subsection 2.2), or their potential as a source of bioactive compounds (subsections 2.2 and 2.3; see also Brito et al. 2012, 2015). The fact that some of the coastal isolates from Portugal (subsection 2.2) exhibit N<sub>2</sub> fixation capabilities is in accordance with findings from Shiozaki et al. (2015), who demonstrated that N<sub>2</sub> fixation estimates for temperate coastal regions may have been underestimated, and thus should deserve further investigation. The polyphasic study of the mats collected in the Brazilian hypersaline lagoons (subsection 2.3) has revealed a rich, morphologically and genetically distinct cyanobacterial biodiversity. The study have exposed the relevance of using complementary methods (i.e. morphological- and molecular-based identifications, and culture-dependent and –independent approaches) to attain a deeper understanding of the cyanobacterial diversity present in complex samples. Cyanobacteria from

these ecosystems are still understudied regarding the bioprospecting for natural compounds. The study shed some light on this issue, since it has shown that 6 out of 9 of the obtained isolates possess PKS genes. With the proposal of 5 new marine taxa (3 genera and 2 species) (subsection 2.2), and although being just a small step, this thesis is contributing towards the Herculean task of reorganizing the taxonomy of cyanobacteria (Komárek 2016), something that was only possible by means of strains in cultures. Overall, as is being demonstrated in other microbiological fields (e.g. Dubourg et al. 2013; Lagier et al. 2015), all these findings reinforce the importance of isolation of strains for the study of unexplored microbial bio- or chemo-diversity. Indeed, isolated strains allow a better characterization of microbial organisms, enabling individualized, ecological and functional analyses (e.g. assessment of N<sub>2</sub> fixation) or taxonomic descriptions, and allow downstream investigations regarding biotechnological applications (see, for instance, section 4).

Taxonomic classification is the only method for better comprehend the biological diversity and to explore knowledge gained from it (Komárek 2016). Modern polyphasic approaches are now enabling more accurate cyanobacterial systematics (Komárek et al. 2014), which are critical to provide a framework for the activities of all cyanobacteriologists. However, the decline in the number of microbial systematists (Smith 2003; Dvorák et al. 2015) is making this process relatively slow, who are needed to form a sound base for molecular technologies and to aid in identifying, and characterizing the cyanobacterial diversity. Progress is necessary, considering that the number of known cyanobacterial species is severely underestimated (Nabout et al. 2013). In this context, CCs dedicated to cyanobacteria must work together to make the best use of new technologies and new systematics for the re-identification of their strains and to contribute to the description of the thousands of cyanobacteria yet to be discovered (Smith 2003; Sherwood et al. 2015). Studies of old, nonliving type material (e.g. herbarium specimens) by molecular methods as that promoted by Palinska et al. (2006) may be extremely valuable for taxonomic purposes. For Oren (2015), this kind of methodology may lead to the harmonization of traditional, morphology-based, and modern molecular taxonomic approaches for the cyanobacteria. By recognizing this, the CyanoType database (subsection 2.1) also anticipates the inclusion of exsiccata strains.

In spite of this, polyphasic identifications of cyanobacteria are not absolutely essential for toxicity risk assessment and water management. This is because taxonomic identification does not fully correlate with toxin production (Moreira et al. 2014; see also section 3). Genetic methods, however, have significantly increased our understanding of the biosynthesis of cyanotoxins and of the distribution of genes involved in their production (e.g. Sivonen & Börner 2008; Dittmann

et al. 2013), allowing the development of molecular-based methods for the detection of potentially toxigenic strains *in situ* (section 3). Instead of the 16S rRNA gene, these methods preferentially target the genes involved in cyanotoxins biosynthesis, demonstrating applicability for the environmental service sector, such as in water monitoring (section 3). Of course, the basic material for the development of these molecular methods were cyanobacterial isolates available in CCs. Such strains are previously obtained from toxic blooms and characterized in terms of toxin production capacity by chemical identification of the toxicants, as occurred for toxic strains held in LEGE CC (section 4). By recognizing the fields of application and the molecular tools being used in cyanobacteria monitoring programs, section 3 demonstrates that human risks concerning the consumption of BGAS are being overlooked, although clear evidences exist (e.g. Saker et al. 2005; Vichi et al. 2012). This has to do with the fact that these products based on edible cyanobacteria (or algae) are very often contaminated with other microorganisms, including toxic cyanobacteria (Vichi et al. 2012; Gomes 2013). In LEGE CC are deposited 11 cyanobacterial strains that were isolated from different brands of *Chorella* and *Spirulina* dietary supplements, by Dina Gomes (Gomes 2013). Those strains belong to different genera from the orders Chroococciopsidales, Nostocales and Synechococcales (see also section 4 and Appendix A8). Worth of notice is that brands often advertise their products as being 100% pure. This apparent lack of supervision and control by competent authorities may be due to the lack of legal force. Indeed, the commerce of BGAS is highly unregulated, although many countries are making laws and complementary regulations to address the health risks of nutraceuticals (Ohmori & Ehira 2013). As mentioned in section 3, there are risks that the edible organisms in BGAS does not conform to the species identification (Görs et al. 2010). There is also the risk that different strains from the same species possess different capabilities for the synthesis of bioactive compounds, most of them not characterized in terms of toxicity, e.g. the study of Shih et al. (2013) shows that closely related *Arthrospira* isolates have different number of gene clusters involved in the synthesis of secondary metabolites. That is why Wessels (2012) argues that, for approval of a beneficial microorganism for food and feed, it would be justified that “the authorities require documentation for both safety and efficacy on the level of strains, and not just species”. Due to the potential health risks for consumers, I have developed a set of three interlinked protocols (i.e. SOPs) for the early detection of potentially toxic cyanobacteria in BGAS (Appendix A5). These SOPs, which were primarily designed to be used by environmental and food laboratory technicians, stresses again the importance of the availability of authenticate strains known to produce cyanotoxins (as those in LEGE CC; see section 4), to be used as positive controls. Such toxic strains should serve also as reference materials in quality control measures, for making sure that analytical methods are being properly followed (e.g. in spiked controls; see Appendix A5).

Several works stress the importance of microbial strains for the developing bioeconomy, especially those from high quality standards collections (e.g. Smith 2003; Smith et al. 2014; Stackebrandt et al. 2015; Antunes et al. 2016). Well described microbial resources are expected to play a key role in driving economic growth (Smith et al. 2014). In line with this, cyanobacterial strains from LEGE CC are being widely characterized, and their biotechnological potential being evaluated by different approaches (section 4). Promising strains are being selected and further characterized, e.g. through genome sequencing and discovery and structure elucidation of bioactive compounds (section 4). But despite their importance, most CCs worldwide face challenges in sustaining their activities over the long term, due to unstable and unsustainable financing schemes (Smith et al. 2014). Also, the rightful collection, use, sharing or supply of biological (including genetic) resources is high on the agenda of both European and international legislation (e.g. Convention of Biological Diversity and the Nagoya Protocol), which imposes additional constraints for mBRCs. The WFCC and, in Europe, the European Culture Collection Organisation (ECCO) have had a key role in supporting their members in this particular. Pan-European network platforms such as the Microbial Resource Research Infrastructure (MIRRI) or EMBRC are also devoted to help mBRCs in relation to operational, certification/accreditation, and financial issues, as well as in the implementation of (new) European and international legislation (Smith et al. 2014). From the user/consumer side, these networks enable an improved access to enhanced quality microorganisms in an appropriate legal framework and to resource-associated data in a more interoperable way (Stackebrandt et al. 2015). LEGE CC is registered in WFCC (WDCM #1089) and is also a partner in EMBRC, as part of CIIMAR (section 4), while is envisaging to become a member of ECCO.

## 5.1 Final remarks

Either to better understand the role of microbes in the environment, to study their taxonomy and physiology, or just to make use of them to explore their bioproducts, microbiologists historically have relied on the availability of isolates in CCs. However, the vast majority of microorganisms in the environment remain unknown and uncultivated (Yarza et al. 2014), and cyanobacteria are no exception (Nabout et al. 2013; Dvořák et al. 2015). This thesis constitutes a further contribution to the comprehension of the full extent of cyanobacterial diversity. By taking the strains from LEGE CC as an exemplification, it also underpin the value of available cyanobacterial isolates for taxonomic, environmental or applied purposes.

The aim of microbial culture collections is not only to achieve long-term viability but also to ensure genotypic and phenotypic integrity of the organisms, something that is absolutely

essential for the authentication of previous findings (such those exposed in section 4, for LEGE CC strains). So, in addition to the isolation and cultivation of strains, it is highly advisable to pursue an adequate preservation that prevent changes in morphological, physiological and genetic traits. For cyanobacteria, cryopreservation is the most appropriate way to achieve it, and to ensure absolute genotypic and phenotypic stability it should be performed as soon as the microorganism is isolated (Day 2014; Prakash et al. 2013). In the near future, this procedure shall be finalized for LEGE CC strains (see also Rastoll et al. 2013).

Besides the discussed findings, important outcomes of this thesis were the founding of LEGE CC (section 4), the establishment of its associated searchable, online database (Appendix D), and the creation of the catalog of strains (Appendix A8). Also, the development of molecular-based protocols that can be used in management systems for food/dietary supplements quality and safety assurance (Appendix A5). Another outcome was the creation of a curated, online database where data and metadata for relevant cyanobacterial strains, from the taxonomic or phylogenetic point of view, can be easily found (Appendix C).

As a final thought, I close with a quote by James Staley (2010): “Microbial taxonomy is a field that is blessed with an exciting future: a biospheric cornucopia of microorganisms that is ripe for harvesting and from which microbiologists can more fully discern the taxonomic, genetic, physiological, ecological and biotechnological richness of microbial life”.

## 5.2 References

- ANTUNES, A., STACKEBRANDT, E. & LIMA, N. 2016. Fueling the bio-economy: European culture collections and microbiology education and training. *Trends in Microbiology*, 24, 77-79.
- ARAHAL, D. R. (2014). Whole-genome analyses: average nucleotide identity. *Methods in Microbiology*, 41, 103-122.
- BRITO, Â., RAMOS, V., SEABRA, R., SANTOS, A., SANTOS, C. L., LOPO, M., FERREIRA, S., MARTINS, A., MOTA, R., FRAZÃO, B., MARTINS, R., VASCONCELOS, V. & TAMAGNINI, P. 2012. Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Systematic and Applied Microbiology*, 35, 110-119.
- BRITO, Â., GAIFEM, J., RAMOS, V., GLUKHOV, E., DORRESTEIN, P. C., GERWICK, W. H., VASCONCELOS, V. M., MENDES, M. V. & TAMAGNINI, P. 2015. Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*, 9, 218-226.
- DITTMANN, E., FEWER, D. P. & NEILAN, B. A. 2013. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiology Reviews*, 37, 23-43.
- DAY, J.G. 2014. Cryopreservation of cyanobacteria. In *Cyanobacteria: An Economic Perspective, 1st edition*, N.K. Sharma, A.K. Rai, L.J. Stal (eds.), John Wiley & Sons, pp. 319-327.
- DUBOURG, G., LAGIER, J. C., ARMOUGOM, F., ROBERT, C., HAMAD, I., BROUQUI, P. & RAOULT, D. 2013. The proof of concept that culturomics can be superior to metagenomics to study atypical stool samples. *European Journal of Clinical Microbiology & Infectious Diseases*, 32, 1099-1099.
- DVOŘÁK, P., POULÍČKOVÁ, A., HAŠLER, P., BELLI, M., CASAMATTA, D. A. & PAPINI, A. 2015. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. *Biodiversity and Conservation*, 24, 739-757.

- ESCOBAR-ZEPEDA, A., DE LEÓN, A. V. P. & SANCHEZ-FLORES, A. 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6, 348.
- GARCÍA-PICHEL, F. 2008. Molecular ecology and environmental genomics of cyanobacteria. In *The Cyanobacteria: Molecular Biology, Genomics and Evolution*, A. Herrero, E. Flores (eds.), Caister Academic Press, pp. 59-87.
- GOMES, D. 2013. *Pesquisa de cianobactérias, cianotoxinas e bactérias patogénicas em suplementos alimentares (Clorela e Spirulina)*. MSc Thesis. ICBAS, University of Porto.
- GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333-351.
- GÖRS, M., SCHUMANN, R., HEPPELE, D. & KARSTEN, U. 2010. Quality analysis of commercial *Chlorella* products used as dietary supplement in human nutrition. *Journal of Applied Phycology*, 22, 265-276.
- KOMÁREK, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, 51, 346-353.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J. R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86, 295-335.
- KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005. Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology*, 187, 6258-6264.
- LAGIER, J.-C., HUGON, P., KHELAIFIA, S., FOURNIER, P.-E., LA SCOLA, B. & RAOULT, D. 2015. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews*, 28, 237-264.
- LEE, E., RYAN, U. M., MONIS, P., MCGREGOR, G. B., BATH, A., GORDON, C. & PAPANINI, A. 2014. Polyphasic identification of cyanobacterial isolates from Australia. *Water Research*, 59, 248-261.
- MIZRAHI-MAN, O., DAVENPORT, E. R. & GILAD, Y. 2013. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE*, 8, e53608.
- MOREIRA, C., RAMOS, V., AZEVEDO, J. & VASCONCELOS, V. 2014. Methods to detect cyanobacteria and their toxins in the environment. *Applied Microbiology and Biotechnology*, 98, 8073-8082.
- NABOUT, J. C., DA SILVA ROCHA, B., CARNEIRO, F. M. & SANT'ANNA, C. L. 2013. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodiversity and Conservation*, 22, 2907-2918.
- NGUYEN, N.-P., WARNOW, T., POP, M. & WHITE, B. 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms and Microbiomes*, 2, 16004.
- OHMORI, M. & EHIRA, S. 2013. *Spirulina*: an example of cyanobacteria as nutraceuticals. In *Cyanobacteria: An Economic Perspective, 1st edition*, N.K. Sharma, A.K. Rai, L.J. Stal (eds.), John Wiley & Sons, pp., 103-118.
- OREN, A. 2015. Comments on "Taxonomy of cyanobacteria: a contribution to consensus approach", by KA Palinska and W. Surosz (*Hydrobiologia* 740: 1-11, 2014, doi: 10.1007/s10750-014-1971-9). *Hydrobiologia*, 758, 1.
- PACE, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, 276, 734-740.
- PALINSKA, K. A., THOMASIU, C. F., MARQUARDT, J. & GOLUBIC, S. 2006. Phylogenetic evaluation of cyanobacteria preserved as historic herbarium exsiccata. *International Journal of Systematic and Evolutionary Microbiology*, 56, 2253-2263.
- PRAKASH, O., NIMONKAR, Y. & SHOUCHE, Y. S. 2013. Practice and prospects of microbial preservation. *FEMS Microbiology Letters*, 339, 1-9.
- RASTOLL, M. J., OUAHID, Y., MARTÍN-GORDILLO, F., RAMOS, V., VASCONCELOS, V. & DEL CAMPO, F. F. 2013. The development of a cryopreservation method suitable for a large cyanobacteria collection. *Journal of Applied Phycology*, 25, 1483-1493.
- SAKER, M. L., JUNGBLUT, A. D., NEILAN, B. A., RAWN, D. F. K. & VASCONCELOS, V. M. 2005. Detection of microcystin synthetase genes in health food supplements containing the freshwater cyanobacterium *Aphanizomenon flos-aquae*. *Toxicon*, 46, 555-562.
- SHERWOOD, A. R., CARLILE, A. L., VACCARINO, M. A. & JOHANSEN, J. R. 2015. Characterization of Hawaiian freshwater and terrestrial cyanobacteria reveals high diversity and numerous putative endemics. *Phycological Research*, 63, 85-92.
- SHIOZAKI, T., NAGATA, T., IJICHI, M. & FURUYA, K. 2015. Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences*, 12, 4751-4764.

- SHIH, P. M., WU, D., LATIFI, A., AXEN, S. D., FEWER, D. P., TALLA, E., CALTEAU, A., CAI, F., TANDEAU DE MARSAC, N., RIPPKA, R., HERDMAN, M., SIVONEN, K., COURSIN, T., LAURENT, T., GOODWIN, L., NOLAN, M., DAVENPORT, K. W., HAN, C. S., RUBIN, E. M., EISEN, J. A., WOYKE, T., GUGGER, M. & KERFELD, C. A. 2012. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, 110, 1053-1058.
- SIVONEN, K. & BÖRNER, T. 2008. Bioactive compounds produced by cyanobacteria. In *The Cyanobacteria: Molecular Biology, Genomics and Evolution*, A. Herrero, E. Flores (eds.), Caister Academic Press, pp. 159-197.
- SMITH, D. 2003. Culture collections over the world. *International Microbiology*, 6, 95-100.
- SMITH, D., MCCLUSKEY, K. & STACKEBRANDT, E. 2014. Investment into the future of microbial resources: culture collection funding models and BRC business plans for biological resource centres. *SpringerPlus*, 3, 81.
- STACKEBRANDT, E., SCHÜNGEL, M., MARTIN, D. & SMITH, D. 2015. The microbial resource research infrastructure MIRRI: strength through coordination. *Microorganisms*, 3, 890-902.
- STALEY, J. T. 2010. Comprehending microbial diversity: the fourth goal of microbial taxonomy. *Bulletin of BISMIS*, 1, 1-5.
- VICHI, S., LAVORINI, P., FUNARI, E., SCARDALA, S. & TESTAI, E. 2012. Contamination by *Microcystis* and microcystins of blue-green algae food supplements (BGAS) on the Italian market and possible risk for the exposed population. *Food and Chemical Toxicology*, 50, 4493-4499.
- WESSELS, S. 2012. Safety and regulation of microorganisms added to the food and feed chains, including probiotics—introduction and overview. In *Beneficial Microorganisms in Agriculture, Food and the Environment: Safety Assessment and Regulation*, Sundh, I., Wilcks, A., Goettel, M. S. (eds.), CAB International, pp. 12-26.
- YARZA, P., YILMAZ, P., PRUESSE, E., GLÖCKNER, F. O., LUDWIG, W., SCHLEIFER, K.-H., WHITMAN, W. B., EUZÉBY, J., AMANN, R. & ROSSELLÓ-MÓRA, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, 12, 635-645.
- ZHOU, J., HE, Z., YANG, Y., DENG, Y., TRINGE, S. G. & ALVAREZ-COHEN, L. 2015. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio*, 6, e02288-14.

## Appendices

### Appendix A

Out-of-format data such as large figures/tables, worksheets and the catalog of strains are provided in digital format only. They can be accessed directly via the following hyperlinks:

[Appendix A1](#). from Chapter 2.1 – Main phylogenetic tree

[Appendix A2](#). from Chapter 2.3 – Rectangular ML tree

[Appendix A3](#). from Chapter 2.3 – Circular ML tree, with labels.

[Appendix A4](#). from Chapter 2.3 – Metadata for sequences: accession numbers, BLAST results, etc

[Appendix A5](#). from Chapter 3 – Standard operating procedures (SOPs)

[Appendix A6](#). from Chapter 3 – Table 3.1 in complete form

[Appendix A7](#). from Chapter 3 – Table 3.2 in complete form

[Appendix A8](#). from Chapter 4 – Catalog of strains of LEGE CC

[Appendix A9](#). from Chapter 4 – Circular ML tree

Or alternatively by following the link:

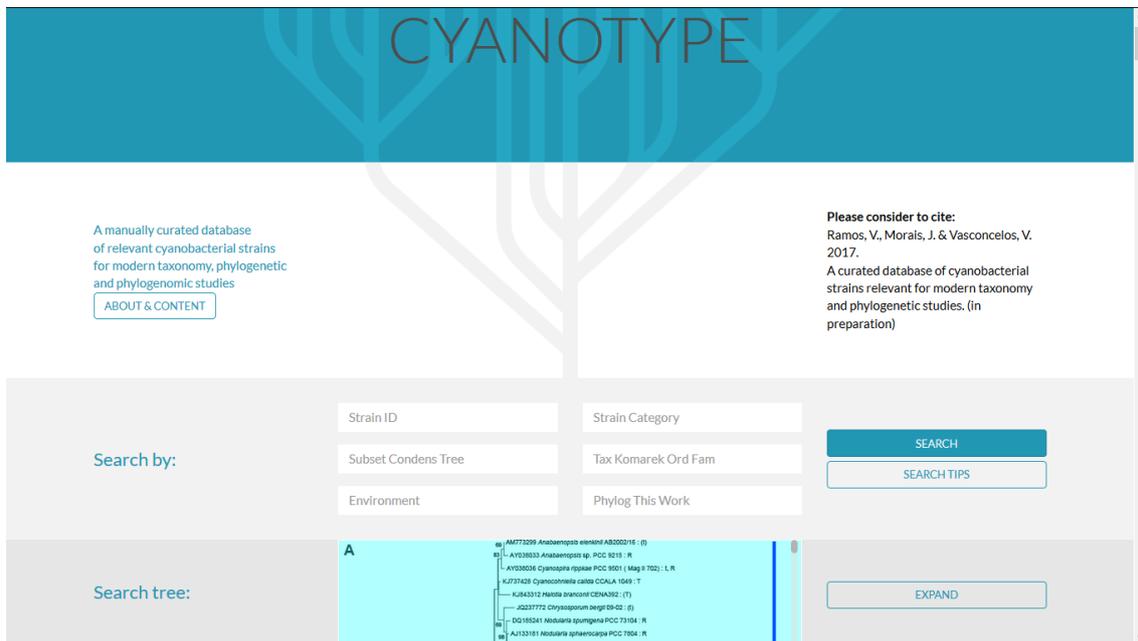
<http://informatica.ciimar.up.pt/bbe/>

## Appendix B

The Dataset from CyanoType (.csv file) and the sequence alignments (fasta format) and phylogenetic trees (newick format) will be made available at FigShare, after publication.

The [Dataset \(.xls file\)](#) can also be accessed via the hyperlink or by following the link:  
<http://informatica.ciimar.up.pt/bbe>

## Appendix C



Screenshot of the main page of the Cyanotype online database:

<http://lege.ciimar.up.pt/cyanotype/>

## Appendix D



Screenshot of the main page of the LEGE Culture Collection website:

<http://lege.ciimar.up.pt/>