UNIVERSITY OF CALIFORNIA, SAN DIEGO

Novel Biodiversity of Natural Products-producing Tropical Marine Cyanobacteria

A Dissertation submitted in partial satisfaction of the requirement for the degree

Doctor of Philosophy

in

Oceanography

by

Niclas Engene

Committee in Charge:

Professor William H. Gerwick, Chair
Professor James W. Golden
Professor Paul R. Jensen
Professor Brian Palenik
Professor Gregory W. Rouse
Professor Jennifer E. Smith

2011

The Dissertation of Niclas Engene is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

DEDICATION

This dissertation is dedicated to everyone that has been there when I have needed them the most.

TABLE OF CONTENTS

Signature Page ……..……... iii

Dedication ……...……..…….. iv

Table of Contents ……...……..……. v

List of Figures ……..…… ix

List of Tables ……...……..….. xii

List of Abbreviations ……....……..…… xiv

Acknowledgments ...…….……..…..… xvii

Vita ……..…….....… xx

Abstract ……..….. xxiv

Chapter I.Introduction ……..……..…… 1

Natural Products as Ancient Medicine ……..…….…….……. 2

Marine Natural Products ……..…….…….…….. 3

Cyanobacterial Natural Products ……..…….……. 10

Current Perspective of the Taxonomic Distribution of Natural Products in Marine Cyanobacteria ……..…….…….…. 11

Phylogenetics to Comprehend and Uncover Cyanobacterial Biodiversity ……..… 16

Rationale for Thesis Research ……..…….…..…… 17

Overview of Thesis Chapters ……..…….……..…….. 18

References ……..…….……..……..... 22

Chapter II. Phylogeny-guided Isolation of Ethyl Tumonoate A from the Marine Cyanobacterium cf. *Oscillatoria margaritifera* 26

Abstract 27

Introduction 28

Materials and Methods 30

Results and Discussion 38

Conclusions 52

Acknowledgments 54

Appendix 55

References 64

Chapter III. 16S rRNA Gene Heterogeneity in the Filamentous Marine Cyanobacterial Genus *Lyngbya* 68

Abstract 69

Introduction 71

Materials and Methods 74

Results 80

Discussion 98

Acknowledgments 104

Appendix 105

References 111

Chapter IV. Intra-genomic 16S rRNA Gene Heterogeneity in Cyanobacterial Genomes 113

Abstract 114

Introduction 115

Materials and Methods 116

Results and Discussion 117

Conclusions 132

Acknowledgments 133

References 134

Chapter V. Underestimated Biodiversity as a Major Explanation for the Perceived Rich Secondary Metabolite Capacity of the Cyanobacterial Genus *Lyngbya*..... 136

Abstract 137

Introduction 138

Materials and Methods 141

Results 148

Discussion 167

Acknowledgments 172

References 173

Chapter VI. *Moorea* gen. nov. (Cyanobacteria), a Tropical Marine Cyanobacterial Genus Rich in Bioactive Secondary Metabolites 177

Abstract 178

Introduction ……..……. 179

Materials and Methods ……..…… 180

Results and Discussion ……..……..……..…… 184

Acknowledgments ……...…...…….....……..…… 201

Appendix …….……...……... 202

References ……....…….……..……... 209

Chapter VII. Taxonomic Distribution of Marine Cyanobacterial Natural Products

…..……..……..... 212

Abstract ……...……..……..……... 213

Introduction ……...……..……... 214

Materials and Methods ……..…… 216

Results and Discussion ……..……..……..…… 218

Conclusions ……..……..……..….. 237

Acknowledgments ……...…...…….....……..…… 239

Appendix …….……..……..….... 243

References ……....…….……..... 243

Chapter VIII. Conclusion ……...……...……..…… 253

LIST OF FIGURES

Figure I.1: Ancient natural products ……..……..... 2

Figure I.2: Marine natural products ……..……..... 5

Figure I.3: Current perception of taxonomic distribution of natural products in marine cyanobacteria ……..…… 14

Figure I.4: Taxonomic distribution of marine cyanobacterial natural products reported during 2010-2011 …….……... 15

Figure II.1: Underwater pictures and microphotographs of cf. *Oscillatoria margaritifera* NAC8-46 from Curaçao …….…….. 39

Figure II.2: Phylogenetic inference of the tumonoic acid-producing cf. *Oscillatoria margaritifera* strains from Curaçao ……..…..... 43

Figure II.3: Molecular structures of Ethyl tumonoate A 47

Figure II.4: Mosher's esters of Ethyl tumonoate A 49

Figure II.5. HR-ESI-MS of Ethyltumonate A ……..……..…..... 56

Figure II.6: ^{1}H-NMR of Ethyltumonate A …….…… 57

Figure II.7: ^{13}C-NMR spectrum of Ethyltumonate A ……..…….. 58

Figure II.8: COSY spectra of Ethyltumonate A …..……..……. 59

Figure II.9: HSQC spectra of Ethyltumonate A …..……..……. 60

Figure II.10: HMBC spectra of Ethyltumonate A …….…… 61

Figure II.11: ^{1}H-NMR of *S*-MTPA ester of Ethyltumonate A …….……. 62

Figure II.12: ^{1}H-NMR of *R*-MTPA ester of Ethyltumonate A ……..… 63

Figure III.1: Underwater pictures and microphotographs of *Lyngbya* specimens ……. 85

Figure III.2: Phylogenetic tree of the order Oscillatoriales ……..…..… 89

Figure III.3: Phylogenetic tree of the "tropical marine *Lyngbya* lineage" ….....… 91

Figure III.4: Average pair-wise sequence divergence between 16S rRNA genes ……. 97

Figure IV.1: Phylogenetic relationships of the 59 cyanobacterial strains with sequenced genomes …….…….....… 120

Figure V.1: Geographic map of Curaçao ……...……...…… 154

Figure V.2: Secondary metabolites attributed to the genus *"Lyngbya"* obtained from Curaçao …….…….....……..…… 158

Figure V.3: Phylogenetic inferences for "*Lyngbya"* specimens from Curaçao 163

Figure V.4: Scanning/transmission electron micrograph (SEM/TEM) of the surface of a cultured "*L. sordida*" 3L filament colonized by heterotrophic bacteria …….…….……..……. 166

Figure VI.1: Phylogenetic inferences *Moorea* diversification based on SSU (16S) rRNA nucleotide sequences …….....…..…….…… 185

Figure VI.2: Phylogenomic inferences of *Moorea* gen. nov …… 187

Figure VI.3: Phylogenetic inferences of the *Moorea* lineage ……... 189

Figure VI.4: Phylogenetic analyses of *Lyngbya* and *Moorea* diversification based on 16S-23S Internal transcribed spacer (ITS) regions ……...…… 191

Figure VI.5: Morphological characterization of *Moorea* gen. nov ……..……...… 193

Figure VI.6: Microphotographs of cyanobacterial filaments obtained by transmission electron microscopy (TEM) ……...…… 194

Figure VI.7: Phylogenetic analyses of *Lyngbya* and *Moorea* diversification based on RNA polymerase gamma-subunit (*rpo*C1) nucleotide sequences ...……..…… 206

Figure VI.8: Photomicrographs of cyanobacterial filaments obtained by epifluorescent microscopy ……...……..…… 207

Figure VI.9: Microphotographs of cyanobacterial filaments obtained by transmission electron microscopy (TEM) …….....…… 208

Figure VII.1: Evolutionary tree of marine NP-producing cyanobacteria …….…… 225

Figure VII.2: Highlight of the lineage including tropical marine "*Lyngbya*" and tropical marine "*Symploca*" ……...……..…… 227

Figure VII.3: Natural products from *Moorea* gen. nov . …….……. 228

Figure VII.4: Highlight of the "*Trichodesmium-Oscillatoria*" lineage ……...…..…… 230

Figure VII.5: Natural products from *Pseudomoorea* gen. nov . …….……. 231

Figure VII.6: Natural products from tropical marine *Symploca* . …….……. 233

Figure VII.7: Highlight of the "*Leptolyngbya*" lineage …...…… 236

Figure VII.8: Evolutionary relationships of the credneramide-producing cyanobacterial strains ……...……..…… 242

LIST OF TABLES

Table I.1: Current clinical status of marine natural products ……. 7

Table I.2: Taxonomic and environmental distribution of secondary metabolites isolated from cyanobacteria 12

Table II.1: Morphological characterization of cf. *Oscillatoria margaritifera* strains from Curaçao 41

Table II.2: Secondary metabolites dereplication 45

Table II.3: NMR spectroscopic data for Ethyl tumonoate A …….. 48

Table III.1: Morphological characterization of *Lyngbya* specimens …….. 82

Table III.2: *Lyngbya* colonies with multiple copies of the 16S rRNA gene of variable sequence ……...…….……..…… 93

Table III.3: Collection data for *Lyngbya* specimens …… 106

Table IV.1: Ribosomal operon redundancy and intra-genomic SSU rRNA gene heterogeneity in cyanobacterial genomes …… 124

Table V.1: Collection data of cyanobacterial specimens from Curaçao ……...…… 143

Table V.2: Primers used for PCR-amplification of secondary metabolite encoding genes ……..…..…….…… 146

Table V.3: Morphological characterization of *Lyngbya* specimens from Curaçao 149

Table V.4: Characterization of secondary metabolites ……..… 156

Table V.5: Detection of biosynthetic pathways and secondary metabolites from *Lyngbya* specimens ……..…..……...…… 160

Table VI.1: Genomic and biochemical characteristics of *Moorea* strains ……..….. 197

Table VI.2: Collection data for *Moorea* specimens ……..…… 203

Table VII.1: True taxonomic identification of natural products-producing marine cyanobacteria ….……..……..…….....…..…… 219

Table VIII.2: Taxonomy and secondary metabolite production of *Symploca* specimens ….…….…….....…..……..…… 245

Table VIII.3: Collection data for *Symploca* specimens …….....…..……. 241

LIST OF ABBREVATIONS

AIC	akaike information criterion
ATCC	American type culture collection
BIC	bayesian information criterion
DDBJ	DNA data bank of Japan
DMAP	dimethylaminopyridine
DMSO	dimethyl sulfoxide
DT	decision-theoretic
EMBL	European molecular biology laboratory
ESI	electrospray ionization
FAB	fast atom bombardment
FBS	fetal bovine serum
FDVA	fluoro-2,4-dinitrophenyl-5-L-valine amide
FT-IR	Fourier transform infrared
GARLI	genetic algorithm for rapid likelihood inference
gCOSY	gradient correlation spectroscopy
gHMBC	gradient heteronuclear multiple bond correlation
gHSQC	gradient heteronuclear single quantum coherence
HAB	harmful algae blooms
HPLC	high-pressure liquid chromatography
HR	high resolution
IACUC	Institutional Animal Care and Use Committee

ICBG	International Cooperative Biodiversity Group
IC_{50}	50 % inhibitory concentration
ICA	intact single-cells analysis
IR	infrared
LC/MS	liquid chromatography / mass spectrometry
MAFFT	multiple alignment fast fourier transform
MALDI	matrix-assisted laser desorption ionization
Mbp	mega base-pairs
MCMC	metropolis-coupled Markov chain Monte Carlo
MDA	multiple displacement amplification
MHz	megahertz
ML	maximum likelihood
MP	maximum parsimony
MS	mass spectrometry
MT	methyl transferase
MTT	3-(4,5-dimethylthiazol-2-Yl)-2,5-diphenyltetrazolium bromide
NMR	nuclear magnetic resonance
NCBI	National center for biotechnology information
NCI	National cancer institute
NED	naphthylenediamine dihydrochloride
NO	nitric oxide
NP	natural product

NRPS	non-ribosomal peptide synthetase
PCC	Pasteur culture collection
PCR	polymerase chain reaction
PKS	polyketide synthase
pINV	proportion of invariable sites
PNG	Papua New Guinea
RaxML	randomized accelerated maximum likelihood
RP	reversed-phase
rpoC1	RNA polymerase gene gamma subunit
rrn	ribosomal operons
rrs	ribosomal genes
SAM	*S*-adenosyl-L-methionine
SEM	scanning electron microscopy
SSU	small ribosomal subunit
TBR	tree-bisection-reconnection
TE	thioesterase
TEM	transmission electron microscopy
TLC	thin layer chromatography
TOF	time of flight
UV	ultraviolet
VLC	vacuum liquid chromatography

ACKNOWLEDGMENTS

The research described in this thesis is far from the work of solely one person, but rather represents the collective efforts of colleagues, friends and collaborators over several years. There is therefore an extensive list of people to thank for their help and support.

First and foremost, I would like to sincerely show my gratitude to my Ph.D. advisor Prof. William Gerwick for his mentorship over the years. Bill has been motivating, inspiring, and thought-provoking. I would also like to thank my outstanding Ph.D. committee members: Prof. Brian Palenik, Prof. Greg Rouse, Prof. Paul Jensen, Prof. Jennifer Smith, and Prof. Jim Golden for all their research advice, encouraging support, and inspiring discussions. In addition, I also appreciate my former research advisors Prof. Richard Lewis at the University of Queensland, Gisela Concepcion Prof. at the University of the Philippines, and Prof. William Fenical for encouragement and guidance in the early stages of my research career.

My research has been made both possible and stimulating due to my wonderful colleagues in the Gerwick lab. All lab-mates (graduate students, post-docs, and visiting researchers) past and present have been helpful and have added to my research experience. A specific thanks to Emily Monroe, Karla Malloy, and Francisco Villa for all their help, especially for proof-reading my written work. Much of this work has been possible because of friends and colleagues in the Fenical lab, the Moore lab, and the Jensen lab as part of the Center for Marine Biomedicine and Biotechnology (CMBB). I particularly appreciate the help of Chris Kauffman and Kelley Freel for their help with

bacterial culturing. I appreciate my collaborators at the National Center for Microscopy and Imaging Research (NCMIR), in particular Erin Rottacker, Tom Dering, and Mark Ellisman. I would also like to thank Jiří Komárek and Jan Kaštovský at the University of South Bohemia for all taxonomic help, and Ed Esquenzi and Pieter Dorrestein for help with MALDI-TOF-MS. I sincerely appreciate the help of Becky Burrola, Denise Darling, Josh Reeves, Krystle Shertz and everyone else at the graduate office that has helped me over the years.

Most of all I would like to thank my wonderful family, my parents, brothers and Emily for always being there.

The text of II, in full, is published material as it appears in: Engene, N., Choi, H., Esquenazi, E., Byrum, T., Villa, F. A., Cao, Z., Murray, T. F., Dorrestein, P. C., Gerwick, L., Gerwick, W. H. (2011). Phylogeny-guided isolation of Ethyl tumonoate A from the marine cyanobacterium cf. *Oscillatoria margaritifera*. *J Nat Prod* (in press). The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

The text of III, in full, is published material as it appears in: Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

The text of IV, in part, is published material as it appears in: Engene, N., Gerwick W. H. (2011). Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes.

Fottea 1: 17-24. The dissertation author was the primary author and conducted the research which forms the basis for this chapter.

The text of V, in full, is published material as it appears in: Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2011). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya. Environ Microbiol* 13: 1601-1610. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

The text of VI, in full, is published material as it appears in: Engene, N., Rottacker, E. C., Kaštovský, K., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., Gerwick, W. H. (2011). *Moorea producta* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J Syst Evol Microbiol* (In press). The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

The text of VII, in full, is the manuscript draft to be submitted to an academic journal as it will appear: Engene, N., Gerwick, W. H. Evolution and taxonomic distribution of marine cyanobacterial natural products. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

VITA

EDUCATION

2006-2011	Doctor of Philosophy in Oceanography Scripps Institution of Oceanography, University of California, San Diego. *Advisor: William Gerwick*
2005-2006	Staff Research Associate II Scripps Institution of Oceanography, University of California, San Diego. *Advisor: William Gerwick*
2005	Master of Science in Chemistry University of Gothenburg/Scripps Institution of Oceanography, University of California, San Diego. *Advisor: William Fenical*
2004-2005	Master of Science in Marine Biology University of Gothenburg/Marine Science Institute, University of the Philippines. *Advisor: Giselle Concepcion*
2004	Undergraduate Researcher Institute of Molecular Bioscience, University of Queensland *Advisor: Richard Lewis*
2000-2004	Bachelor of Science in Chemistry University of Gothenburg
1999-2004	Bachelor of Science in Marine Biology University of Gothenburg

AWARDS AND FELLOWSHIPS

2011	Smithsonian Marine Science Network Fellowship
2011	Phycological Society of America (PSA) Grants-In-Aid Award
2010	Lynn Brady Travel Award
2010	Paul Silva student grant
2009	ZoBell Award
2008	Edna Bailey Sussman Fellowship
2007	Halliday Student Field Research Award
2006	Wallenbergska Foundation Fellowship
2006	Swedish-American Foundation Fellowship
2005	Willinska Foundation Fellowship
2004	Exchange student scholarship (University of Queensland)
2003	Adelbertska Foundation

PUBLICATIONS

Engene, N., Gerwick, W. H. Evolution and taxonomic distribution of marine cyanobacterial natural products. *Nat prod rev* (In prep).

Engene, N., Rottacker, E. C., Kaštovský, J. H., Komárek, J., Gerwick W. H. (2011). *Moorea producta* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J Syst Evol Microbiol* (In press)

Engene, N., Choi, H., Esquenazi, E., Byrum, T., Villa, F., Murray, T., Dorrestein, P. C., Gerwick, L. G., Gerwick, W.H. (2011). Phylogeny-guided isolation of Ethyl tumonoate A from the marine cyanobacterium *Oscillatoria margaritifera*. *J Nat Prod* (In press)

Leão, P., Engene, N., Antunes, A., Gerwick, W. H., Vasconcelos, V. M. The chemical ecology of cyanobacteria. *Nat Prod Rep* (In prep)

Malloy, K. L., Suyama, T. L., Engene, N., Debonsi, H., Cao, Z., Matainaho, T., Murray, T. F., Gerwick, W. H. Credneramides A and B: neuromodulatory phenethylamine and isopentylamine derivatives of a vinyl chloride-containing fatty acid from a Papua New Guinea collection of *Lyngbya* sp. *J Nat Prod* (In review)

Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2011). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya*. *Environ Microbiol* 13: 1601-1610.

Mevers, E., Liu, W-T., Engene, N., Byrum, T., Dorrestein, P. C., Gerwick, W. H. (2011). Cytotoxic Veraguamides, alkynyl bromide containing cyclic depsipeptides from an unusual marine cyanobacterium cf. *Oscillatoria margaritifera. J Nat Prod* 74: 928-936.

Engene, N., Gerwick W.H. (2011). Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. *Fottea* 1: 17-24.

Pereira, A. R., Etzbach, L., Engene, N., Müller, R., Gerwick, W. H. (2011). Unusual molluscicidal chemotypes from a new Palmyra Atoll cyanobacterium. *J Nat Prod* 74: 1175-1181.

Engene, N., Coates, R. C., Gerwick W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601.

Malloy, K. L., Villa, F., Engene, N., Matainaho, T., Gerwick, L. G., Gerwick, W. H. (2011). Malyngamide 2, a bioactive metabolite from the Papua New Guinea marine cyanobacterium *Lyngbya sordida. J Nat Prod* 74: 95-98.

Pereira, A. R., Cao, Z., Engene, N., Soria-Mercado, I. E., Murray, T. F., Gerwick, W. H. (2010). Palmyrolide A, a new neuroactive macrolide from a consortium of two Palmyra Atoll cyanobacteria. *Org Lett* 12: 4490-4493.

Choi, H., Engene, N., Smith, J. E., Preskitt, L. B., Gerwick, W. H. (2010). Crossbyanols A–D, toxic brominated polyphenyl ethers from the Hawai‘ian bloom-forming cyanobacterium *Leptolyngbya crossbyana*. *J Nat Prod* 73: 517-522.

Tidgewell, K., Engene, N., Byrum, T., Valeriote, F. A., Gerwick, W. H. (2010). Diversification of a modular natural product pathway: production of Apratoxins F and G, two cytotoxic cyclic depsipeptides from a Palmyra collection of *Lyngbya bouillonii*. *ChemMedChem* 11: 1-11.

Choi, H., Pereira, A. R., Cao, Z., Shuman, C. F., Engene, N., Byrum, T., Matainaho, T., Murray, T. F., Mangoni, A., Gerwick, W. H. (2010). The Hoiamides, structurally-intriguing neurotoxic lipopeptides from Papua New Guinea marine cyanobacteria. *J Nat Prod* 73: 1411-1421.

Gutierrez, M., Tidgewell, K., Capson, T., Engene, N., Almanza, A., Manfred, J., Jarg, S., Gerwick, W. H. (2010). Malyngolide dimer, a bioactive symmetric cyclodepside from the Panamanian marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73: 709-711.

Taniguchi, M., Nunnery, J., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., Gerwick, W. H. (2010). Palmyramide A, a cyclic depsipeptide from a Palmyra Atoll collection of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73: 393-398.

Clark, B. R., Engene, N., Teasdale, M. E., Rowley, D. C., Matainaho, T., Valeriote, F. A., Gerwick, W. H. (2008). Natural products chemistry and taxonomy of the marine cyanobacterium *Blennothrix cantharidosmum*. *J Nat Prod* 71: 1530-1537.

Simmons, T. L., Engene, N., Urena, L., Romero, L. I., Ortega-Barria, E., Gerwick, L., Gerwick, W. H. (2008). Viridamides A and B, Lipodepsipeptides with antiprotozoal activity from the marine cyanobacterium *Oscillatoria nigro-viridis*. *J Nat Prod* 71: 1544-1550.

Gutierrez, M., Suyama, T., Engene, N., Wingerd, J., Matainaho, T., Gerwick, W. H. (2008). Apratoxin D, a potent cytotoxic cyclodepsipeptide from Papua New Guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*. *J Nat Prod* 71: 1099-1103.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Gerwick, W. H., Byrum, T., Carland, T., R. Coates, C., Engene, N., Esquenazi, E., Gerwick, L., Grindberg, R., Johnson, M., Jones, A., Malloy, K., Nunnery, J., Pereira, A., Soria, I., Sorrels, C., Taniguchi, M., Tidgewell, K., Villa, F., Vining,

O., Dorrestein, P., Gu, L., Sherman, D. H. (2008). Integrating chemical and biochemical approaches to natural products drug discovery from marine cyanobacteria, Proceedings from the New Developments in Drug Discovery on Natural Products and Traditional Medicines. NIPER, India, 33-43.

Simmons, T. L., Coates, R. C., Clark, B. R., Engene, N., Gonzalez, D., Esquenazi, E., Dorrestein, P. C., Gerwick. W. H. (2007). Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *PNAS*. 105: 4587-4594.

ABSTRACT OF THE DISSERTATION

Novel Biodiversity of Natural Products-producing Tropical Marine Cyanobacteria

by

Niclas Engene

Doctor of Philosophy in Oceanography

University of California, San Diego, 2011

Professor William H. Gerwick, Chair

During the last three decades, tropical marine cyanobacteria have emerged as an extraordinarily prolific source of promising biomedical natural products (NPs). Creative endeavors have been used to explore this novel chemical diversity, but lagging behind is the recognition and description of the biological diversity responsible for these NPs. Instead of being recognized as unique taxa, tropical marine NP-producing groups have been identified, with few exceptions, based on classification systems tied to morphospecies of terrestrial and freshwater specimens from temperate regions. This lack of proper classification systems is primarily due to the fact that tropical marine cyanobacteria have only recently been explored. As a result, our current perspective of

the taxonomic origin and distribution of NPs in marine cyanobacteria is extraordinarily incomplete.

A major aspect of this thesis research has focused on providing a better understanding of how NPs are distributed among different cyanobacterial groups so as to improve the efficiency of future investigations. The initial perception was that bioactive secondary metabolites are unequally distributed among different taxonomic groups with a few groups being responsible for the majority of the isolated NPs. However, based on polyphasic characterization of globally distributed populations, many of the most chemically prolific groups were found to constitute polyphyletic groups. Moreover, a large proportion of the secondary metabolites attributed to these groups are in fact produced by morphologically similar but evolutionarily distant cyanobacteria. I argue that this morphological resemblance of different cyanobacterial lineages is a major reason why some cyanobacterial groups have such an apparent richness of secondary metabolites.

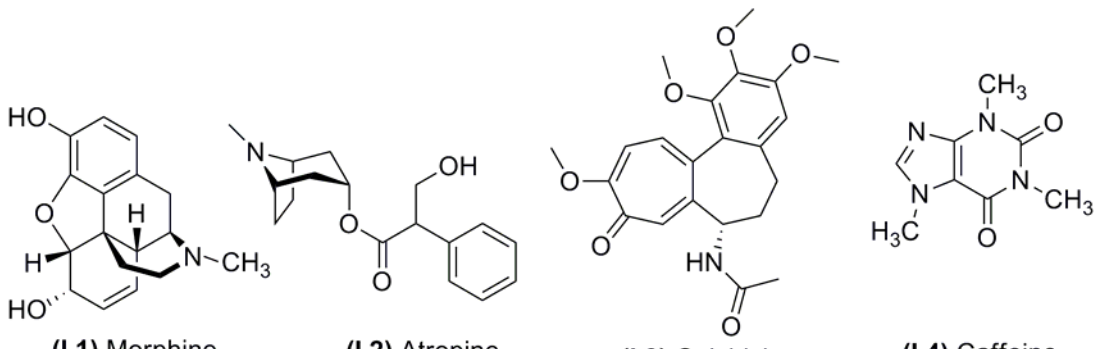
In my efforts to provide taxonomic clarity as well as to better guide future natural product drug discovery investigations, I have proposed a revision of the NP-rich genus *Lyngbya* on the basis of phylogenetic, genomic, secondary metabolism and ultrastructural comparisons with the genus reference strain PCC 7419^{T}. In conclusion, this proposed revision of "tropical marine *Lyngbya*"as a new genus (*Moorea* gen. nov.) highlights the underestimated biodiversity of tropical marine cyanobacteria.

Chapter I

Introduction

I.1 Natural Products as Ancient Medicine

The exploitation of biologically active secondary metabolites for useful applications, including therapeutic drugs, is far from new. Since ancient times, nature has been recognized as an important source of potential drugs; examples of early uses and benefits of natural products for human can be found in most major civilizations (Newman *et al*., 2000; Constantino *et al*., 2004). The use of natural products for medical purposes slowly developed into the scientific field of pharmacognosy. However, this was not until the early 1800's when the active principles from plants were isolated and characterized. Among the first active principles isolated were morphine **(I.1)**, atropine **(I.2)**, colchicine **(I.3)**, and caffeine **(I.4)** (Newman *et al*., 2005)**.** Morphine, atropine, and caffeine are today listed as core medicines in the World Health Organization's "Essential Drugs List", which is a list of minimum medical needs for basic health care systems (WHO, 2005).



(I.1) Morphine **(I.2)** Atropine **(I.3)** Colchicine **(I.4)** Caffeine

Figure I.1: Some examples of natural products isolated in the early 18th century.

Natural products are outstanding in the diversity of their chemical structures and biological activities. In contrast, the chemical diversity produced by the pharmaceutical industry using methods such as synthetic combinatorial chemistry, seldom shows as potent or diverse biological activities (Constantino *et al*., 2004; Berdy, 2005; Newman & Cragg, 2007). The advantage with natural products is that they have been developed and perfected upon millions of years of evolutionary pressure to be biologically active (Berdy, 2005).

I.2 Marine Natural Products

Today, perhaps the major frontier in the ongoing search for novel natural products is the marine environment. The world's oceans provide a tremendous resource for the discovery of potential therapeutic agents. The oceans cover over 70% of the earth's surface and represent more than 95% of the biosphere by volume (Hamann *et al*., 2003; Schumacher Hamann *et al*., 2011). All life forms originated from the ocean, and the ocean is the only environment that possesses all living phyla (Hamann *et al*., 2003). More than half of the total number of global species (>2,000,000) are present in the ocean, making it far more diverse and biologically rich than its terrestrial counterpart, in both biomass and number of species (Hamann *et al*., 2003). This rich biological resource has long provided man with a major food source, however, only during very recent times has the ocean been recognized in a biomedical sense.

The era of marine drug discovery began in the 1950's, with the isolation of the nucleosides, spongouridine **(I.5)** and spongthymidine (**I.6)** from the marine sponge

Tethya crypta (Bergman & Feeney, 1951; Bergman & Burks, 1955). These metabolites served as drug lead structures for the antiviral drug cytarbine, ara-A and the anticancer drug vidarbine, ara-A (Hamann *et al*., 2003; Mayer *et al*., 2010; Waters *et al*., 2010). The next few decades of marine NP drug discovery were relatively slow, with less than 200 marine NPs being published by 1974. However, novel compounds were isolated at an exponential rate over the next 20 years, and by the year 1993, 7,000 different marine NPs had been discovered (Faulkner, 2000). Today, a little over 60 years since Bergman discovered the arabinose-based nucleosides, more than 22,000 novel marine natural products have been discovered (Blunt *et al*., 2010).

As a result of these continuing discoveries, a total of four marine-derived natural products have reached the market as pharmaceutical drugs. Cytarbine (Cytosar-U®, Depocyt®) and vidarbine (Vira-A®) were followed by the polypeptide ω-conotoxin MVIIA (Ziconotide or Prialt®; **I.43**), which passed the final Phase III clinical trials and was launched as a new drug against pain in 2004 (Miljanich *et al*., 2003; Mayer *et al*., 2010). More recently, in 2007 the tunicate-derived trabectedin (Yondelis®; **I.24**) was approved by the European Agency for the Evaluation of Medicinal Products (EMEA) for the treatment of soft tissue sarcoma and was approved in 2009 for ovarian carcinoma (Mayer *et al*., 2010). In addition, the halichondrin B-analog Eribulin (Halaven®) was approved in November 2010 for treatment of metastatic breast cancer (USFDA, 2010).

The clinical pipeline contains many other promising marine-derived drugs currently in trials. A total of thirteen marine NPs or NP-derivatives are found in various stages of clinical development (Mayer *et al*., 2010; Waters *et al*., 2010). These marine-derived NPs are summarized in Figure I.2 and Table I.1.

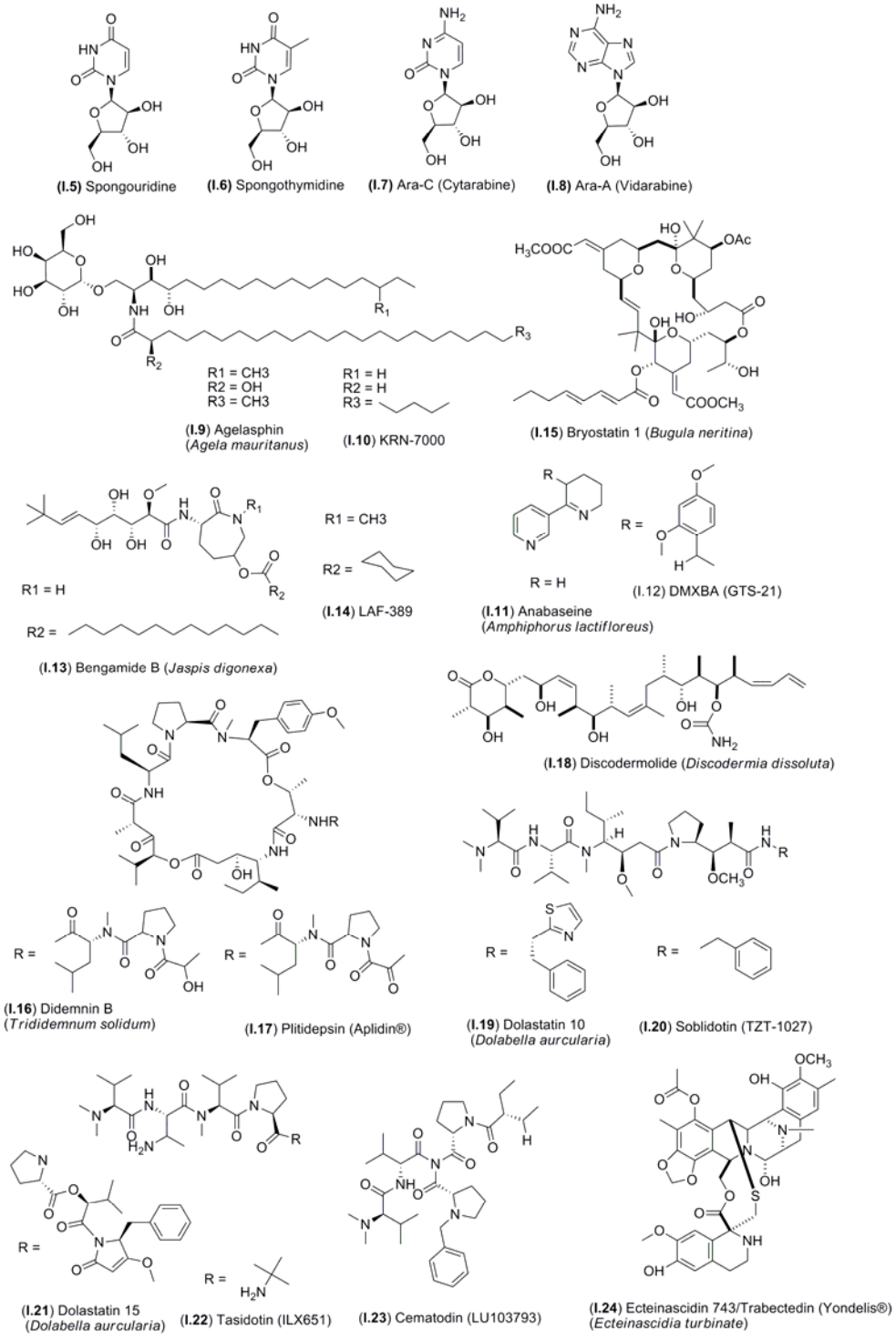


Figure I.2. Marine natural products approved by the Food and Drug Administration (FDA) or the European Agency for the Evaluation of Medicinal Products (EMEA) or in various stages of clinical trials.

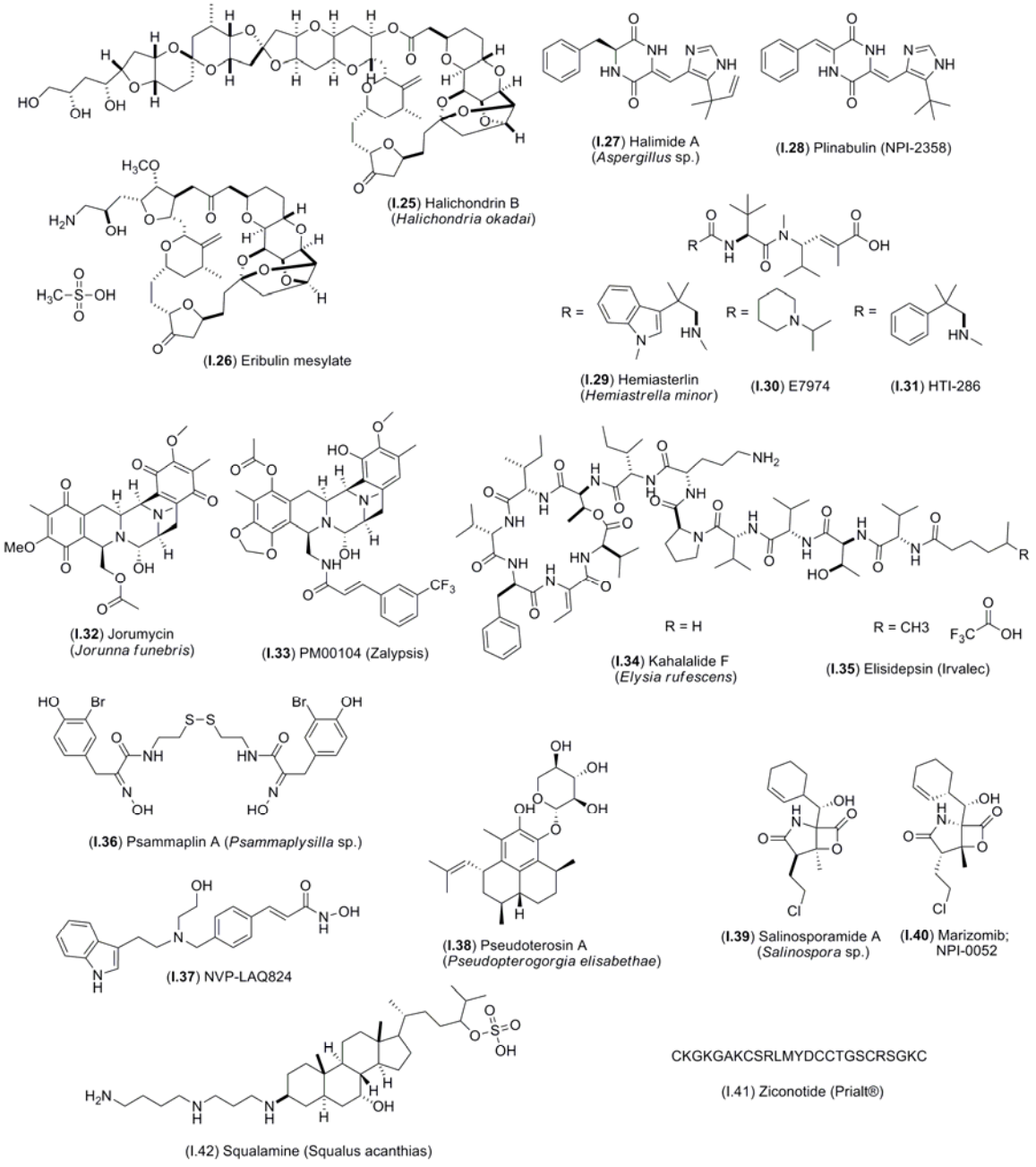


Figure I.2. Continued

Table I.1: Current Clinical Status of Marine natural products.

Original Natural Product	Compound name (Trademark)	Source Organism	Molecular target	Current Status in 2011 (Disease)
Agelasphin (**I.9**)	KRN-7000 (**I.10**)	*Agelas mauritianus* (sponge)	Vα24+NKT cell activation	Phase I (Cancer & hepatitits C)
Anabaseine (**I.11**)	DMXBA; GTS-21 (NA) (**I.12**)	*Amphiphorus lactifloreus* (worm)	Nicotinic acetylcholine receptor agonist	Phase II (Alzheimers/Schizophrenia)
Bengamide B (**I.13**)	LAF-389 (**I.14**)	*Jaspis digonoxea* (sponge)	Methionine aminopeptidase	Discontinued 2002
Bryostatin 1 (**I.15**)	(NA)	*Bugula neritina* (bryozoan)	Protein kinase C modulator	Phase I (Cancer)/ Phase II (Alzheimers)
Didemnin B (**I.16**)	Plitidepsin (Aplidin®) (**I.17**)	*Tridemnum solidum* (tunicate)	Under investigation	Phase II (Cancer)
Discodermolide (**I.18**)	NA	*Discodermia dissoluta* (sponge)	Tubulin	Discontinued 2005
Dolastatin 10 (**I.19**)	Soblidotin; TZT-1027 (NA) (**I.20**)	*Dolabella auricularia* (mollusk)	Tubulin	Phase III (Cancer)

Table I.1: Continued

Original Natural Product	Compound name (Trademark)	Source Organism	Molecular target	Current Status in 2011 (Disease)
Dolastatin 15 (**I.21**)	Tasidotin; ILX651 (**I.22**)	*Dolabella auricularia* (mollusk)	Tubulin	Phase II (Cancer)
Dolastatin 15	Cemadotin; LU103793 (**I.23**)	*Dolabella auricularia* (mollusk) / *Symploca* sp. (cyanobacterium)	Tubulin	Phase I/II (Cancer) (Discontinued 2004)
Ecteinascidin 743 (**I.24**)	Trabectedin; ET-743 (Yondelis®)	*Ecteinascidia turbinata* (tunicate)	Tubulin	Approved by EMEA (Cancer)
Halichondrin B (**I.25**)	Eribulin mesylate; (Halaven®) (**I.26**)	*Halichondria okadai* (sponge)	Tubulin	Approved by FDA (Cancer)
Halimide (**I.27**)	Plinabulin: NPI-2358 (**I.28**)	*Aspergillus* sp. *(*fungus)	Tubulin	Phase II (Cancer)
Hemiasterlin (**I.29**)	E7974 (NA) (**I.30**)	*Hemiastrella minor* (sponge)	Tubulin	Phase I (Cancer)
Hemiasterlin	HTI-286 (**I.31**)	*Cymbastella* sp. (sponge)	Tubulin	Discontinued 2005
Jorumycin (**I.32**)	PM00104 (Zalypsis®) (**I.33**)	*Jorunna funebris* (nudibranch)	DNA-binder	Phase II (Cancer)

Table I.1: Continued

Original Natural Product	Compound name (Trademark)	Source Organism	Molecular target	Current Status in 2011 (Disease)
Kahalalide F (**I.34**)	Elisidepsin (Irvalec®) (**I.35**)	*Elysia rufescens* (mollusk)	Lysosomes / erbB pathway	Phase II (Cancer)
Psammaplin A (**I.36**)	NVP-LAQ824 (**I.37**)	*Psammaplysilla* sp. (sponge)	HDAC/DNMT	Discontinued 2005
Pseudoterosins A-D (**I.38**)	NA	*Pseudopterogorgia elisabethae* (octocoral)	Under investigation	Anti-inflammatory/wound healing
Salinosporamide A (**I.39**)	Marizomib; NPI-0052 (NA) (**I.40**)	*Salinospora tropica* (marine bacterium)	Proteasome inhibitor	Phase I (Cancer)
Spongothymidine (**I.6**)	Vidarbine, Ara-A (Vira-A®) (**I.8**)	*Cryptotethya crypta* (sponge)	Nucleotide mimic	Approved by FDA (Antiviral)
Spongouridine (**I.5**)	Cytarabine; Ara-C (Cytosar-U®, Depocyt®) (**I.7**)	*Cryptotethya crypta* (sponge)	Nucleotide mimic	Approved by FDA, EMEA (Cancer)
Spirulosine	ES-285	*Spisula polynyma* mollusk)	*Rho*-GTP inhibitor	Discontinued 2007 (Cancer)
Ω-conotoxin MVIIA	Ziconotide (Prialt®) (**I.43**)	*Conus magus* (cone snail)	N-type calcium channel blocker	Approved by FDA, EMEA (Analgesic)

Abbreviations: FDA – Food and Drug Administration, EMEA – European Agency for the Evaluation of Medicinal Products

I.4 Cyanobacterial Natural Products

Cyanobacteria ("blue-green algae") are considered the most ancient group of oxygenic photosynthetic organisms (Graham & Wilcox, 2000). Over an estimated three billion years, this bacterial phylum has adapted to most habitats on the planet and is among the most abundant and geographically widespread group of prokaryotes known (Whitton, 1992; Stanley, 2004). Cyanobacteria, both in marine and freshwater environments, have recently attracted much attention due to their extraordinary capacities to produce diverse and highly bioactive secondary metabolites (Gerwick *et al*., 2008; Nunnery *et al*., 2010; Tidgewell *et al*., 2010; Tan *et al*., 2010). These bioactive molecules are often potent toxins, causing hazardous harmful algal blooms (HABs) (Osborne *et al*., 2008; Golubic *et al*., 2010). In most aquatic environments, these toxic HABs are increasing globally in frequency and size by alarming rates and represent hazards to both human health and natural ecosystems (Paul *et al*., 2005; Golubic *et al*., 2010).

Despite their hazardous toxicity, many cyanobacterial secondary metabolites also have promising natural products potential for a broad spectrum of pharmaceutical applications such as anticancer, anti-inflammatory, antibacterial and anti-infective therapeutic agents (Gerwick *et al*., 2008; Tidgewell *et al*., 2010; Tan, 2010; Nunnery *et al*., 2010; Villa & Gerwick, 2010). The economic value of anti-cancer drugs alone from marine sources has been estimated to be between $563 billion and 5.69 trillion (Erwin *et al*., 2010). In addition, many of these cyanobacterial secondary metabolites also have other potential commercial applications such as insecticides, algaecides, and herbicides (Berry *et al*., 2008).

I.4 Current Perspective of the Taxonomic Distribution of Natural Products in Marine Cyanobacteria

To date, a total of 533 natural products have been reported from marine cyanobacteria (Table I.2). The taxonomic distribution of these secondary metabolites is remarkably uneven (Figure I.2). First, these 533 NPs are attributed to a total of only 13 different genera. Second, over 90% of all these molecules are attributed to only five different genera (Figure I.2). This uneven taxonomic distribution can be considered rather remarkable when considering the creative sampling endeavors used to obtain these cyanobacteria from geographically and environmentally diverse habitats (Taniguchi *et al*., 2009). However, the trend of attributing NPs to *Lyngbya* appears to continue. Almost 75% of all secondary metabolites isolated from marine cyanobacteria during 2010 were attributed to collections of *Lyngbya* (Figure I.3).

Table I.2: Taxonomic and environmental distribution of secondary metabolites isolated from cyanobacteria.[1]

Genera	Habitat	Number of secondary metabolites
Anabaena	Fresh-water	28
Aphanizomenon	Fresh-water	5
Aulosira	Terrestrial (soil)	1
Calothrix	Terrestrial (soil)	2
Cylindrospermopsis	Fresh-water	4
Cylindrospermum	Fresh-water/terrestrial	2
Fischerella	Fresh-water	4
Geitlerinema	Marine	3
Leptolyngbya	Marine	6
Lyngbya	Marine	326
Microcystis	Fresh-water	50
Microcoleus	Marine	8
Nodularia	Fresh/brackish-water	3
Nostoc	Fresh-water/terrestrial	84
Oscillatoria	Marine	82
Phormidium	Marine	18
Plectonema	Fresh-water	2
Prochloron	Marine	1
Prochlorothrix	Fresh-water	3
Rivularia	Marine	7
Schizothrix	Marine	39
Scytonema	Fresh-water/terrestrial	14
Stigonema	Fresh-water/terrestrial	4
Symploca	Marine	26
Synecococcus	Marine	3
Synechocystis	Marine	13

Table I.2: Continued

Genera	Habitat	Nr. secondary Metabolites
Tolypothrix	soil/terrestrial	25
Trichodesmium	Marine	1
Westiellopsis	Terrestrial	1
Westiella	Soil	2

[1] Data obtained from MararinLit, 2011

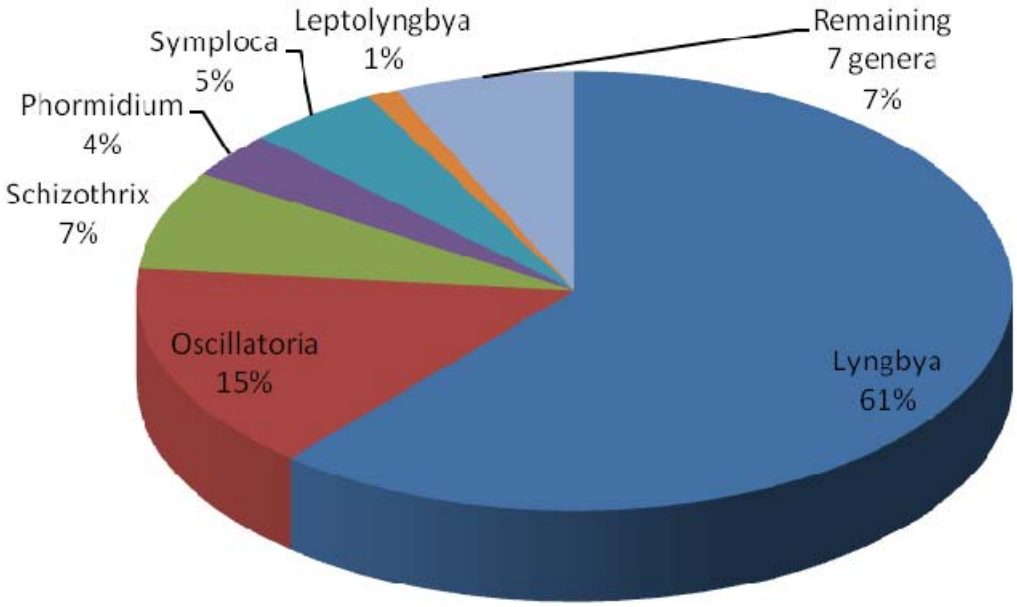


Figure I.3: Current perception of taxonomic distribution of NPs (n = 533) in marine cyanobacteria. The diagram reveals the published taxonomy of all marine cyanobacteria available in the MarinLit database, 2011. According to these statistics, the primary NP-producers are *Lyngbya* (61.2%), *Oscillatoria* (15.4%), *Schizothrix* (7.3%), *Symploca* (4.9%) and *Phormidium* (3.4%). The remaining 6.6% have been isolated from the following genera: *Synechocystis* (2.4%), *Microcoleus* (1.5%), *Rivularia* (1.3%), *Leptolyngbya* (1.1%), *Geitlerinema* (0.6%), *Prochloron* (0.2%), and *Trichodesmium* (0.2%).

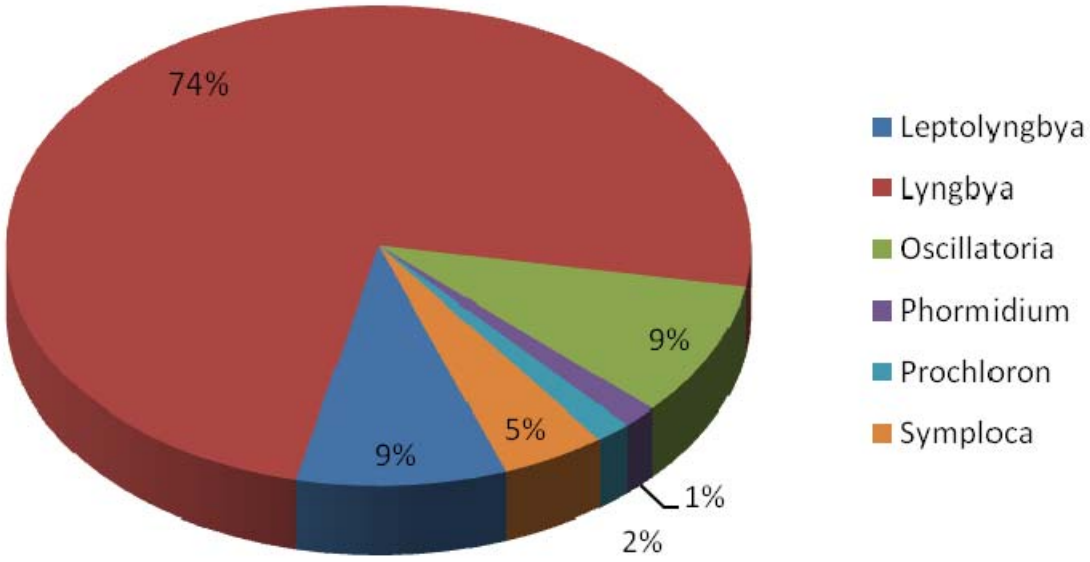


Figure I.4: Taxonomic distribution of marine cyanobacterial NPs reported during 2010-2011. Here, the reported NP-producers are *Lyngbya* (74%), *Oscillatoria* (9%), *Leptolyngbya* (9%), *Symploca* (5%), *Prochloron* (2%), and *Phormidium* (1%).

I.5 Phylogenetics to Comprehend and Uncover Cyanobacterial Biodiversity

Historically, the taxonomic classification of cyanobacteria has been based primarily upon morphological characterizations (Anagnostidis & Komárek, 1988). The first comprehensive classification system was constructed based on morphology of European field-collected specimens, which formed the basis of modern era cyanobacterial systematics (Geitler, 1932). However, many of the traditional phenotypic traits used for taxonomic characterization have been shown to be variable and influenced by environmental factors (Sumina, 2006). More recently, cyanobacteria have largely been accepted to represent a bacterial phylum, and a more comprehensive Bacteriological Code of systematics that incorporates biochemical, genetic, or physiological data has consequently been proposed (Stanier, 1978; Ripka, 1979; Oren, 2004).

More importantly, both the Bacteriological Code and the traditional Botanical Code have begun incorporating the evolutionary histories and phylogenetic relationships as a fundamental frame-work to establish biologically informative classification systems (Castenholz, 2002; Komárek, 2005; Komárek, 2006). Therefore, to uncover and understand the biodiversity of tropical marine cyanobacteria these specimens need to be placed in phylogenetic perspective with the type-strains used to establish our current taxonomic systems.

The most recently explored habitats have shown remarkable degrees of novel cyanobacterial biodiversity, *e.g.* polar regions (Casamatta *et al*., 2005; Strunecký *et al*., 2010), alkaline marshes (Fiore *et al*., 2007), hyperarid polar desert (Pointing *et al*., 2009), and caves (Lamprinou *et al*., 2011). However, the tropical marine benthos has been by far the most chemically prolific habitat in regard to cyanobacterial NPs (Gerwick *et al*.,

2008). Paradoxically, this remarkable chemical diversity of tropical marine cyanobacteria has not been matched with a comparative extent of novel biological diversity.

I.6 Rationale for Thesis Research

During the last three decades, tropical marine cyanobacteria have emerged as an extraordinarily prolific source of promising biomedical natural products. These bioactive secondary metabolites are typically reported from an exceptionally limited number of cyanobacterial genera. This unequal distribution, with certain groups being vastly overrepresented in chemical richness, may lead to the basic assumption that chemical diversity significantly exceeds biological diversity.

The overarching objective of my Ph.D. research is to increase our understanding of the evolution and taxonomic distribution of bioactive secondary metabolites in marine cyanobacteria. The central hypothesis is that the morphology-based taxonomic systems used to classify marine cyanobacteria greatly underestimates cyanobacterial biodiversity and that this misclassification is the primary reason for the unequal distribution of their secondary metabolites.

In addition, as these systems primarily are designed for temperate, freshwater or terrestrial cyanobacteria, I predict that many of these tropical marine cyanobacteria may represent novel and undescribed taxonomic groups. Taxonomic revisions on the basis of evolutionarily relationships are likely to give better appreciation for cyanobacterial biodiversity as well as provide a more efficient cataloging of secondary metabolite producing cyanobacterial strains. Along these lines, phylogenetics may also provide a

reliable predictor for metabolic capacities and a powerful guide for future natural product discovery endeavors.

I.7 Overview of Thesis Chapters

Understanding the taxonomic distribution of bioactive secondary metabolites in marine cyanobacteria is fundamental to comprehend how these metabolic pathways have evolved. Chapter II investigates the correlation between secondary metabolites and phylogenetic positions of the producing strain. The hypothesis in this chapter is that the biosynthetic pathways encoding the natural products evolve vertically. This hypothesis was tested in a case-study in which the phylogenetic relationships of cyanobacterial strains were used to predict the biosynthetic capacities for and consequent production of specific secondary metabolites. This phylogeny-guided isolation approach resulted in the discovery of the new tumonoic acid derivative, ethyl tumonoate A, with anti-inflammatory activity and inhibitory activity of calcium oscillation in cortical neurons. The text of II, in full, is published material as it appears in: Engene, N., Choi, H., Esquenazi, E., Byrum, T., Villa, F. A., Cao, Z., Murray, T. F., Dorrestein, P. C., Gerwick, L., Gerwick, W. H. (2011). Phylogeny-guided isolation of ethyl tumonoate A from the marine cyanobacterium cf. *Oscillatoria margaritifera*. *J Nat Prod*. The dissertation author was the primary author and conducted the research which forms the basis for this chapter.

In Chapter III I tested the hypothesis that taxonomic groups prolific in bioactive secondary metabolites are polyphyletic or evolutionarily diverse, since this could explain some of the inequality in secondary metabolite distribution. I showed that the NP-rich

genus *Lyngbya,* in fact, formed a polyphyletic group composed of multiple evolutionarily unrelated lineages. In addition, analysis of clonal *Lyngbya* cultures and multiple displacement amplified (MDA) single-cell genomes revealed that *Lyngbya* genomes contain two 16S rRNA gene copies, and that these typically were of variable sequence. Furthermore, intra-genomic and inter-species 16S rRNA gene heterogeneity were found to be of approximately the same magnitude. The text of III, in full, is published material as it appears in: Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

The objective in Chapter IV was to expand on the findings in Chapter III and investigate the extent of intra-genomic 16S rRNA gene heterogeneity in cyanobacteria. A bioinformatics comparison of all currently available sequenced genomes revealed that cyanobacteria commonly contain multiple and variable ribosomal operons. However, we conclude that cyanobacterial intra-genomic 16S rRNA gene heterogeneity generally has a relatively small impact on species delineation and inference of evolutionary histories of cyanobacteria. The text of IV, in full, is published material as it appears in: Engene, N., Gerwick W. H. (2011). Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. *Fottea* 1: 17-24. The dissertation author was the primary author and conducted the research which forms the basis for this chapter.

Chapter V tested the hypothesis that the polyphyly found in Chapter II could represent a major underlying explanation for the perceived unequal taxonomic distribution of secondary metabolites. Specimens of *Lyngbya* from various environmental

habitats around the chemo-diversity hot-spot Curaçao were analyzed for both their biosynthetic capacities to produce secondary metabolites as well as their actual metabolite production. The results of this study, in conjunction with Chapter III's phylogenetic investigations of the various NP-producing *Lyngbya* specimens, supported the hypothesis that polyphyly of morphologically similar cyanobacteria is the major explanation for the perceived chemical richness of the genus *Lyngbya*. These results further underscored the need to revise the taxonomy of this group of cyanobacteria. The text of V, in full, is published material as it appears in: Engene, N., Choi, H., Esquenazi, E., Rottacker, E.C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2011). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya. Environ Microbiol* 13: 1601-1610. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

In Chapter VI we tested if polyphyletic and evolutionarily distant cyanobacterial groups are also biologically and chemically different, and subsequently in need of revision for improved identification and classification. The tropical marine *Lyngbya* lineage was analyzed by morphological, phylogenetic, metabolomic and genomic comparisons with the genus type-strain PCC 7419^{T} to evaluate whether this group should represent a new generic entity. As a result, the genus *Lyngbya* was revised and the generic entity *Moorea* gen. nov. was described as a major tropical marine and NP-rich lineage. The text of VI, in full, is published material as it appears in: Engene, N., Rottacker, E. C., Kaštovský, K., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., Gerwick, W. H. (2011). *Moorea producta* gen. nov., sp. nov. and *Moorea bouillonii*

comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J Syst Evol Microbiol* (In press). The dissertation author was the primary author and conducted the majority of the research, which forms the basis for this chapter.

Chapter VII is a broader phylogenetic investigation of tropical marine cyanobacteria. This chapter concludes that the vast majority of NP-producing marine cyanobacteria represent novel and undescribed biological diversity, and that our current perception of the taxonomic distribution of natural products in marine cyanobacteria is incorrect and incomplete. Chapter VII is currently being prepared for journal submission in 2011. The dissertation author was the primary author and conducted the majority of the research, which forms the basis for this chapter.

Chapter VIII summarizes the findings of the thesis research and provides conclusions and future perspectives on the biodiversity of NP-rich marine cyanobacteria.

I.8 References

Anagnostidis, K., Komárek, J. (1988). Modern approach to the classification system of the Cyanophytes 3: Oscillatoriales. Arch. Hydrobiol. 80, *Algological Studies* 50: 327-472.

Berdy, J. (2005). Bioactive Microbial Metabolites. *J Antibiot* 58: 1-26.

Bergmann, W., Feeney, R. J. (1951). Contibutions to the study of marine products. XXXII. The nucleosides of sponges. *J Org Chem* 16: 981-987.

Bergmann, W., Burke, D. C. (1955). Marine products. XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. *J Org Chem* 20: 1501-1507.

Blunt, J. W., Copp, B. R., Munro, M. H. G., Northcote, P. T., Prinsep, M. R., & Faulkner D. J. (2003). *Nat Prod Rep* 20: 1.

Casamatta, D. A., Johansen, J. R., Vis, M. L., Broadwater, S. T. (2008). Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cynaobacteria). *J Phycol* 41: 421-438.

Castenholz, R. W. (2001). Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 473-553. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Choi, D. H., Noh, J. H., Lee, C. M., Rho, S. (2008). *Rubidibacter lacunae* gen. nov., sp. nov., a unicellular, phycoerythrin-containing cyanobacterium isolated from seawater of Chuuk lagoon, Micronesia. *Int J Syst Evol Microbiol* 58: 2807-11.

Cragg, G. M., Newmann, D. J. (2009). Nature: a vital source of leads for anticancer drug Development. *Phytochem Rev* 8: 313–331.

Constantino, V., Fattorusso, E., Menna, M., Taglilatela-Scafati, O. (2004). Chemical diversity of bioactive marine natural products: an illustrative study. *Curr Med Chem* 11: 1671-1692.

Faulkner, J. D. (2000). Marine Pharmacology. Antonie van Leeuwenhoek 77**:** 135–145.

Fiore, M. F., Sant'Anna, C. L., Azevedo, M. T. D., Komárek, J., Kaštovský, J., Sulek, J., Lorenzi, A. S. (2007). The cyanobacterial genus *Brasilonema*, gen. nov., a molecular and phenotypic evaluation'. *J Phycol* 43: 789-798.

Geitler, L. (1932). *Cyanophyceae*. In Rabenhorst, L. [Ed.]. Kryptogamen-Flora von Deutschland, Österreich und der Schweiz, vol. XIV. Leipzig: Akademische Verlag (1985 reprint: Königstein: Koeltz Scientific Books) 1027-1068.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Hamann, M. T. (2003). Enhancing Marine Natural Products structural Diversity and Bioactivity through Semisynthesis and Biocatalysis. *Curr Pharm Design* 9: 879-889.

Komárek, J., Anagnostidis, K. (2005). In *Süsswasserflora von Mitteleuropa* 19/2. Elsevier/Spektrum: Heidelberg.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae* 21: 349-375.

Komárek J., Hauer T. (2011). CyanoDB.cz - On-line database of cyanobacterial genera. - Word-wide electronic publication, Univ. of South Bohemia & Inst. of Botany AS CR, http://www.cyanodb.cz

Lamprinou, V., Hernández-Mariné, M., Canals, T., Kormas, K., Economou-Amilli, A., Pantazidou, A. (2011). Two new stigonematalean cyanobacteria: Iphinoe *spelaeobios* gen. nov., sp. nov. and *Loriellopsis cavernicola* gen. nov., sp. nov. from Greek and Spanish caves. Morphology and molecular evaluation. *Int J Syst Evol Microbiol* PMID: 21257695

Li, Z., Brand, J. (2007). *Leptolyngbya nodulosa* sp. nov. (Oscillatoriaceae), a subtropical marine cyanobacterium that produces a unique multi-cellular structure. *Phycologia* 46: 396-401.

Marine literature database (http://www.chem.canterbury.ac.nz/marinlit/ marinlit.shtml)

Mayer, A. M. S., Glaser, K. B., Cuevas, C., Jacobs, R. S., Kern, W., Little, R. D., McIntosh, J. M., Newman, D. J., Potts, B. C., Shustler, D. E. (2010). The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol Sci* 31: 255-265.

Miljanich, G. P. (2004). Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr Med Chem* 11: 3029-3040.

Newman, D. J., Cragg G. M., Snader, K. M. (2000). The influence of natural products upon drug discovery. *Nat Prod Rep* 17: 215-234

Newman, D. J., Cragg, G. M., Snader, K. M. (2003). The influence of natural products upon drug discovery. *Nat Prod Rep* 17: 215-234.

Oren, A. (2004). A proposal for further integration of the cyanobacteria under the Bacteriological Code. *Int J Syst Evol Microbiol* 5: 1895–1902.

Pointing, S. B., Chan, Y., Lacap, D. C., Lau, M. C. Y., Jurgens, J. A., Farrell, R. L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci USA* 106: 19964-19969.

Proksch, P., Ebel, R., Edrada, R. A., Schupp, P., Lin, W. H., Wray, S. V., Steube, K. (2003). Detection of pharmacologically active natural products using ecology. Selected examples from Indopacific marine invertebrates and sponge-derived fungi. *Pure Appl Chem* 75: 343-352.

Rasoulouniriana, D., Siboni, N., Ben-Dov, E., Kramarsky-Winter, E., Loya, Y., Kushmaro, A. (2009). *Pseudoscillatoria coralii* gen. nov., sp. nov., a cyanobacterium associated with coral black band disease (BBD). *Dis Aquat Organ* 16: 91-6.

Rippka, R., Deruelles, J., Waterbury, J. W., Herdman, M., Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. *Microbiol* 111: 1-61.

Schumacher, M., Kelkel, M., Dicato, M., Diederich, M. (2011). Gold from the sea: Marine compounds as inhibitors of the hallmarks of cancer. *Biotechnol Adv* doi:10.1016/j.physletb.2003.10.071

Simmons, T. L., Andrianasolo, E., McPhail, K., Flatt, P., Gerwick, W. H. (2005). Marine natural products as anticancer drugs. *Mol Cancer Ther* 4: 333-342.

Simmons, T. L., Gerwick, W. H. (2008). In *Oceans and Human Health*, Walsh, P., Solo-Gabriele, H., Fleming, L. E., Smith, S. L., Gerwick, W. H. Eds.; Elsevier: New York, New York. Ch. 22, p. 431-452.

Stanier, R. Y., Sistrom, W. R., Hansen, T. A. (1978). Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. *Int J Syst Bacteriol* 28: 335–336.

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 Biol* 33: 1419-1428.

Sumina, E. L. (2006). Behavior of filamentous cyanobacteria in laboratory culture. *Microbiology* 75: 459-464.

Taniguchi, M., Nunnery, J., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., Gerwick, W.H. (2010). Palmyramide A, a cyclic depsipeptide from a Palmyra atoll collection of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73: 393-398.

USFDA (2010-11-15). "FDA approves new treatment option for late-stage breast cancer". Press release. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm233863.htm. Retrieved November 15, 2010.

World Health Organization. (2005). WHO model list of essential medicines. http://whqlibdoc.who.int/hq/2005/a87017_eng.pdf.

Waters, A. L., Hill, R. T., Place, A. R., Hamann, M. T. (2010). The expanding role of marine microbes in pharmaceutical development. 21: 780-6.

Chapter II

Phylogeny-Guided Isolation of Ethyl Tumonoate A from the Marine Cyanobacterium cf. *Oscillatoria margaritifera*

Abstract

The evolutionary relationships of cyanobacteria, as inferred by their SSU (16S) rRNA genes, were used as predictors of their potential to produce varied secondary metabolites. The evolutionary relatedness in geographically distant cyanobacterial specimens was then used as a guide for the detection and isolation of new variations of predicted molecules. This phylogeny-guided isolation approach for new secondary metabolites was tested in its capacity to direct the search for specific classes of new natural products from Curaçao marine cyanobacteria. As a result, we discovered ethyl tumonoate A (**1**), a new tumonoic acid derivative with anti-inflammatory activity and inhibitory activity of calcium oscillations in neocortical neurons.

II.1 Introduction

Cyanobacteria form a monophyletic bacterial phylum extraordinarily rich in bioactive secondary metabolites (Tan, 2007; Gerwick *et al*., 2008). A number of these metabolites are potent toxins associated with harmful algal blooms (Sharp *et al*., 2009). Ironically, many of these same bioactive molecules have simultaneously been discovered to have a variety of potential pharmaceutical applications (Gerwick *et al*., 2008). In order to support and enhance the search for novel natural products it is valuable to have a proper understanding of the taxonomy of the secondary metabolite-producing cyanobacteria (Gerwick, 1989). The traditional system of classifying cyanobacteria, which is based on phenotypical observations, has recently shown major incongruities with phylogenetic classifications where evolutionarily informative housekeeping genes are analyzed (Wilmotte *et al*., 2001; Komárek *et al*., 2006). This is partly a result of significant morphological plasticity among cyanobacteria, even at the genus level (Sumina *et al*., 2006). Furthermore, cyanobacterial classification systems are largely founded on a relatively limited number of morphological characters, which often appear similar due to convergent evolution (Sumina *et al*., 2006). Thus, the recent inclusion of phylogenetic analysis in the classification of these microorganisms has led to a more accurate taxonomic understanding of these cyanobacterial groups (Hoffman *et al*., 2005; Komárek & Anagnostidis, 2005). These developments have benefitted not only the field of taxonomy, but also natural products drug discovery efforts. For example, phylogenetic approaches have been used to: (i) accurately identify natural product-producing strains (Simmons *et al*., 2008a); (ii) analyze the microbial diversity in natural product producing

assemblages (Simmons *et al*., 2008b; Taniguchi *et al*., 2009); (iii) recollect specific natural product producing strains from the environment (Simmons *et al*., 2008b); and (iv) distinguish cyanobacterial chemotypes (Tidgewell *et al*., 2010).

According to basic evolutionary principles, metabolic and biosynthetic pathways are evolving between geographically isolated or distant populations as an adaptation to new environments. This concept of biogeographical diversification corresponds to a major underlying rationale and approach in the search for new natural products (Taniguchi *et al*., 2009; Tidgewell *et al*., 2010). While in fresh-water cyanobacteria diverse secondary metabolites, such as microcystins and cylindrospermopsins, have been reported from phylogenetically unrelated genera, the production of secondary metabolites in marine cyanobacteria as well as the distribution of their biosynthetic pathways has been linked with the phylogenetic positions of the secondary metabolite production (Thacker & Paul, 2004; Sharp *et al*., 2009). However, to date the two concepts of evolutionary relatedness and geographic diversification have not been effectively combined for predictive exploration of new natural products from marine cyanobacteria.

The major focus of this study was to use evolutionary relationships of marine cyanobacteria inferred by their SSU (16S) rRNA genes as predictors of their potential to produce various different secondary metabolites. The evolutionary relatedness of geographically distant populations was then used as a guide for the detection and isolation of new variations of predicted molecules. Furthermore, because new geographic locations likely represent unique environments with different evolutionary pressures, such analogs may possess different biological activities. This phylogeny-guided isolation approach for new secondary metabolites was tested for its capacity to direct the search for

specific classes of new natural products from Curaçao marine cyanobacteria. As a result, we discovered ethyl tumonoate A (**1**), a new tumonoic acid derivative with anti-inflammatory activity in murine macrophage cells and inhibitory activity of calcium oscillations in neocortical neurons.

II.2 Materials and Methods

II.2.1 General Experimental Procedures

Optical rotations were measured on a JASCO P-2000 polarimeter and IR spectra on a ThermoElectron Nicolet IR100 FT-IR spectrometer. NMR spectra were recorded with chloroform as an internal standard (δ_C 77.2, δ_H 7.26) on a Bruker 600 MHz spectrometer (600 and 150 MHz for ^{1}H and ^{13}C NMR, respectively) equipped with 1.7 mm MicroCryoProbe. LR- and HR-ESI-MS were obtained on a ThermoFinnigan LCQ Advantage Max mass detector and Thermo Scientific LTQ-XL Orbitrap mass spectrometer, respectively. LCMS analysis was carried out on a Finnigan LCQ Advantage Max spectrometer with a Finnigan Surveyor HPLC system equipped with Thermo Finnigan Surveyor PDA plus detector. HPLC was performed using a Waters 515 pumps and Waters 996 photodiode array detector.

II.2.2 Sampling and Taxonomic Characterization

Cyanobacterial strains were collected by SCUBA or snorkeling from the following four sites along the leeward coast of Curaçao, Netherlands Antilles: NAC8-45 (Marie Pampoen; harbor inlet; 12º 05’47.22 N, 68º 54’54.54 W; 2−3 m depth), NAC8-46

(Caracas baii; coral reef; 12º 04’31.48 N, 68º 51’49.37 W; 8−10 m depth), NAC8-54 (Pierbaai reef; coral reef; 12º 05’40.07 N, 68º 54’47.55 W; 8−10 m depth), and NAC8-55 (Jan Theil baii; coral reef; 12º 04’33.80 N, 68º 52’54.55 W; 3−4 m depth). Specimens were cleaned from macro-flora/fauna under a dissecting scope. Algal tissue (ca. 200 mg) was preserved for genetic analysis in 10 mL RNA*later*® (Ambion), for chemical analysis in seawater/EtOH (1:1) at -20 ºC, and in seawater filtered through 0.2 µm Acrodisc® Syringe filters (PALL Life Sciences) for culturing and morphological analysis. Morphological characterizations were performed using an Olympus IX51 epifluorescent microscope (100X) equipped with an Olympus U-CMAD3 camera. Taxonomic identification of cyanobacterial specimens was performed in accordance with modern phycological systems (Castenholz, 2001; Komárek & Anagnostidis, 2005).

II.2.3 Polymerase Chain Reaction (PCR) and Cloning

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Inc.) following the manufacturer’s specifications. DNA concentration and purity were measured on a DU® 800 spectrophotometer (Beckman Coulter). The 16S rRNA genes were PCR-amplified from isolated DNA using the cyanobacteria-specific primers, 106F and 1509R, as previously described (Nübel *et al*., 1997). The PCR reaction volumes were 25 µL containing 0.5 µL (~50 ng) of DNA, 2.5 µL of 10 x PfuUltra IV reaction buffer, 0.5 µL (25 mM) of dNTP mix, 0.5 µL of each primer (10 µM), 0.5 µL of PfuUltra IV fusion HS DNA polymerase and 20.5 µL dH_2O. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 2 min at 95 °C, 25 cycles of amplification: 20 sec at 95 °C, 20 sec at 50 °C and 1.5 min at

72 °C, and final elongation for 3 min at 72 °C. PCR products were purified using a MinElute® PCR Purification Kit (Qiagen) before subcloned using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) following the manufacturer's specifications. Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced with M13 primers. The gene sequences are available in the DDBJ/EMBL/GenBank databases under acc. No.: GU724196, GU724197, GU724207 and GU724208.

II.2.4 Phylogenetic Inference

Gene sequences were aligned bi-directionally using the L-INS-i algorithm in MAFFT 6.717. A total of 1,378 bp (310 parsimony informative sites) of the 16S rRNA gene were analyzed without data exclusion. The evolutionary distant unicellular cyanobacterium *Gloeobacter violaceus* PCC 7421^{T} (NC005125) was included as an out-group. Representative type-strains (^{T}) were selected from *Bergey's Manual* systems (Castenholz, 2001). Phylogenetic analyses were compared using the Maximum Likelihood (ML), Bayesian inference and Maximum parsimony (MP) methods. Appropriate nucleotide substitution models were compared and selected using uncorrected/corrected Akaike Information Criterion (AIC/AICc), Bayesian Information Criterion (BIC), and the Decision-theoretic (DT) in jModelTest 0.1.1. Bayesian analysis was conducted using MrBayes 3.1 (Ronquist & Huelsenbeck, 2003). The AIC1, AIC2, DIC and BIC criteria all selected GTR+I+G as the optimum model. The Markov chains, one cold and three heated were run for 3,000,000 generations. The Maximum Likelihood (ML) inference was performed using PhyML v2.4.4 (Guindon & Gascuel, 2003). The

analysis was run using the GTR+I+G model with 1,000 bootstrap replicates. The maximum parsimony (MP) analysis was performed in PAUP* 4.0b10 using a heuristic search through the branch-swapping tree-bisection-reconnection (TBR) algorithm with the addition of 10,000 random replicates to find the most parsimonious tree. Bootstrap support was obtained from 1,000 replicates.

II.2.5 Isolation and Structure Elucidation

Biomass of each cyanobacterial specimen was extracted exhaustively with CH_2Cl_2–MeOH (2:1). The extracts were fractionated by silica gel VLC using a stepwise gradient from 100% hexanes to 100% MeOH in nine fractions (A−I) of increasing polarity. VLC-fractions were purified over a 500 mg C-18 Sep-Pak followed by RP HPLC (Phenomenex Jupiter 10 µm C18, 300 Å, 10×250 mm, 85% CH_3OH/H_2O at 3 mL/min). For MALDI-TOF-MS, each specimen (5-10 µg wet wt.) was extracted with 1 µg/µL of matrix solution (70 $mg \cdot mL^{-1}$ alpha-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid (1:1), 750 µL CH_3CN, 248 µL dH_2O, 2 µL trifluoroacetic acid) in 96-well plastic plates for 20−30 sec. One µL of matrix extract was deposited on the well of a Bruker Microflex MSP 96 Stainless Steel Target Plate and run on a Bruker Microflex mass spectrometer equipped with flexControl 3.0.

II.2.6 Ethyl Tumonoate A (1)

Pale yellow oil; $[\alpha]_D^{25}$ -77.5 (*c* 1.0, $CHCl_3$); IR (neat) ν_{max} 1733, 1630 cm^{-1}; ^1H, ^{13}C and 2D NMR data, see Table II.2; HRESIMS *m/z* $[M+Na]^+$ 390.2617 (calcd for $C_{21}H_{37}NO_4Na$, 390.2620).

II.2.7 Absolute Configuration of the Proline Residue in Ethyl Tumonoate A (1) by Marfey's Method

Ethyl tumonoate A (100 μg) was treated with 100 μL of 6 N HCl at 110 °C for 30 min. The reaction products obtained following lyophilization of the crude reaction mixture were dissolved in 100 μL of H_2O and re-lyophilized. The dried hydrolysate was dissolved in 100 μL of 1 M $NaHCO_3$ and then 25 μL of 1% L-FDVA (1-fluoro-2,4-dinitrophenyl-5-L-valine amide) in acetone was added. The solution was vortexed and incubated at 40 °C for 60 min. The reaction was quenched by the addition of 25 μL of 2 N HCl, diluted with 100 μL of MeOH, and then a 10 μL aliquot was analyzed by LC-MS using the following RP HPLC conditions: [(HP Lichrosphere 100 RP-18 column, 5.0 μm, 4.0×125 mm) with a stepped gradient elution of 0.1% trifluoroacetic acid in H_2O (eluent A) and 100% CH_3CN (eluent B); gradient program 0−5 min, B, 30%, 5−25 min; B, 30−70%, flow rate, 500 μL/·min^{-1}]. The Marfey derivatives of authentic D- and L-Pro eluted at 15.88 and 13.74 min, respectively. The Marfey derivative from acid hydrolysates of **1** was eluted at 13.47 min, and coinjection with the authentic L-Pro derivative gave a single peak.

II.2.8 Absolute Configuration of the 3-Hydroxy Group of Ethyl Tumonoate A (1) by Mosher's Method

Dried compound **1** (1 mg) was dissolved in 200 μL of anhydrous pyridine, and a catalytic amount of DMAP (4-dimethylaminopyridine) and an excess amount of (*R*)-MTPA-Cl were added. The reaction vial was maintained at room temperature for 24 h,

and the reaction progress was monitored by NP TLC. The (*S*)-MTPA ester of **1** was isolated by preparative NP TLC with a developing solvent of hexanes–EtOAc (1:1) and an eluent of 100% EtOAc. Using the same procedure with (*S*)-MTPA-Cl, the (*R*)-MTPA ester of **1** was also obtained. The absolute configuration of the 3-hydroxy group was determined by ^{1}H NMR analysis of the (*R*)/(*S*)-MTPA esters (Figure II.11-II.12).

II.2.9 3-(*S*)-MTPA Ester of Ethyl Tumonoate A (1)

Pale yellow oil; ^{1}H NMR ($CDCl_3$, 600 MHz) δ_H 5.79 (1H, t, J = 7.2 Hz, H-5), 5.61 (1H, d, J = 10.6 Hz, H-3), 4.15 (2H, m, H-1″), 4.14 (1H, m, H-2′), 3.56 (1H, m, H-5′a), 3.47 (1H, m, H-5′b), 2.96 (1H, dq, J = 10.6, 6.9 Hz, H-2), 2.07 (2H, m, H-6), 1.94 (1H, m, H-4′a), 1.93 (1H, m, H-3′a), 1.86 (1H, m, H-3′b), 1.77 (1H, m, H-4′b), 1.64 (3H, s, 4-Me), 1.37 (2H, dd, J = 13.3, 6.9 Hz, H-7), 1.32–1.21 (8H, m, H-8, H-9, H-10, H-11), 1.25 (3H, t, J = 7.1 Hz, H-2″), 1.01 (3H, d, J = 7.0 Hz, 2-Me), 0.87 (3H, t, J = 7.1 Hz, 3H); LR ESIMS *m/z* 583.94 $[M+H]^+$, 606.17 $[M+Na]^+$.

II.2.10 3-(*R*)-MTPA Ester of Ethyl Tumonoate A (1)

Pale yellow oil; ^{1}H NMR ($CDCl_3$, 600 MHz) δ_H 5.72 (1H, t, J = 6.9 Hz, H-5), 5.50 (1H, d, J = 10.6 Hz, H-3), 4.36 (1H, dd, J = 7.9 Hz, H-2′), 4.18 (2H, m, H-1″), 3.59 (1H, m, H-5′a), 3.52 (1H, m, H-5′b), 2.95 (1H, dq, J = 10.7, 6.9 Hz, H-2), 2.05 (2H, m, H-6), 1.98 (1H, m, H-3′a), 1.96 (1H, m, H-4′a), 1.91 (1H, m, H-3′b), 1.80 (1H, m, H-4′b), 1.49 (3H, s, 4-Me), 1.36 (2H, m, H-7), 1.33–1.20 (8H, m, H-8, H-9, H-10, H-11), 1.27 (3H, t, J = 7.1 Hz, H-2″), 1.03 (3H, d, J = 6.9 Hz, 2-Me), 0.87 (3H, t, J = 6.6 Hz, 3H); LR ESIMS *m/z* 583.82 $[M+H]^+$, 606.11 $[M+Na]^+$.

II.2.11 Nitric oxide Assay

Cells from the mouse macrophage cell line RAW264.7 (ATCC; Manassas, VA) were cultured in DMEM with 4 mM L-glutamine and 4.5 $g \cdot L^{-1}$ glucose supplemented with 10% fetal bovine serum (FBS), penicillin, and streptomycin. Unless otherwise stated RAW264.7 cells were seeded in 96-well plates (5×10^4 cells/well), and after settling for 1 day they were stimulated with 3 $\mu g \cdot mL^{-1}$ LPS in the absence or presence of ethyl tumonate A (1 to 10 $\mu g \cdot mL^{-1}$) for 24 h in triplicate wells at 37 °C with 5% CO_2.The generation of NO was assessed in the supernatant of cell cultures by quantification of nitrite using the Griess reaction (Green *et al*., 1982). In brief, 50 µL of each supernatant were added to 96-well plates together with 50 µL 1% sulfanilamide in 5% phosphoric acid plus 50 µL 0.1% naphthylenediamine dihydrochloride NED in H_2O, and the optical density was measured at 570 nm. The IC_{50} value, the sample concentration resulting in 50% inhibition of NO production, was determined using non-linear regression analysis (% nitrite versus concentration).

Cytotoxicity was measured in NCI H-460 human lung carcinoma cells using the MTT assay (Green *et al*., 1982). Cells were seeded in 96-well plates at 3.3×10^4 $cells \cdot mL^{-1}$ in 180 µL of RPMI 1640 medium with 10% FBS. After overnight recovery, the test compounds were dissolved in DMSO and diluted into medium without FBS and then added to the wells (tested at final concentrations of 30 and 3 $\mu g \cdot mL^{-1}$). After 48 h, cell viability was determined by MTT staining.

II.2.12 Neocortical Neuron Culture

Primary cultures of neocortical neurons were obtained from embryonic day 16 Swiss-Webster mice. Briefly, pregnant mice were euthanized by CO_2 asphyxiation, and embryos were removed under sterile conditions. Neocortices were collected, stripped of meninges, minced by trituration with a Pasteur pipette and treated with trypsin for 25 min at 37 °C. The cells were then dissociated by two successive trituration and sedimentation steps in soybean trypsin inhibitor and DNase containing isolation buffer, centrifuged and resuspended in Eagle's minimal essential medium with Earle's salt (MEM) and supplemented with 1 mM L-glutamine, 10% FBS, 10% horse serum, 100 $IU \cdot mL^{-1}$ penicillin and 0.10 $mg \cdot mL^{-1}$ streptomycin, pH 7.4. Cells were plated onto poly-L-lysine-coated 96-well (9 mm) clear-bottomed black-well culture plates (Costar) at a density of 1.5×10^5 cells/well. Cells were then incubated at 37 °C in a 5% CO_2 and 95% humidity atmosphere. Cytosine arabinoside (10 μM) was added to the culture medium on day 2 after plating to prevent proliferation of nonneuronal cells. The culture media was changed on days 5 and 7 using a serum-free growth medium containing Neurobasal Medium supplemented with B-27, 100 $IU \cdot mL^{-1}$ penicillin, 0.10 $mg \cdot mL^{-1}$ streptomycin, and 0.2 mM L-glutamine. Neocortical cultures were used in experiments between 8−13 days *in vitro* (DIV). All animal use protocols were approved by the Institutional Animal Care and Use Committee (IACUC).

II.2.13 Intracellular Ca^{2+} Monitoring

Neocortical neurons grown in 96-well plates were used for $[Ca^{2+}]_i$ measurements at 12−13 DIV. Briefly, the growth medium was removed and replaced with dye loading buffer (50 μL/well) containing 4 μM fluo-3 and 0.04% pluronic acid F-127 in Locke's

buffer (8.6 mM Hepes, 5.6 mM KCl, 154 mM NaCl, 5.6 mM glucose, 1.0 mM $MgCl_2$, 2.3 mM $CaCl_2$, 0.0001 mM glycine, pH 7.4). After 1 h incubation in dye loading buffer, the neurons were washed four times in fresh Locke's buffer (200 μL/well) using an automated cell washer (BioTek instrument) and transferred to a FLEXstation II (Molecular Devices). The final volume of Locke's buffer in each well was 150 μL. Cells were excited at 485 nm and Ca^{2+}-bound Fluo-3 emission was detected at 535 nm.

II.2.14 Data Analysis

Time-response and concentration-response graphs were generated using Graphpad Prism software (Graphpad Software). The EC_{50} values were determined by non-linear regression analysis using a logistic equation.

II.3 Results and Discussion

II.3.1 Taxonomic Identification and Phylogenetic Inference

In October 2008 the tropical depression Omar made landfall on the Caribbean island of Curaçao bringing torrential rainfall and beach erosion (Meterological service of Netherlands Antilles, 2008). As a probable result of this disruptive weather system, extensive black, mat-forming cyanobacterial blooms emerged at several sites along the island's leeward coast (Figure II.1). Four geographically dispersed populations of this cyanobacterium (NAC8-45, NAC8-46, NAC8-54 and NAC8-55) were collected and morphologically and phylogenetically compared. All four specimens fell under the criteria of *Lyngbya* according to traditional morphology-based classification systems.

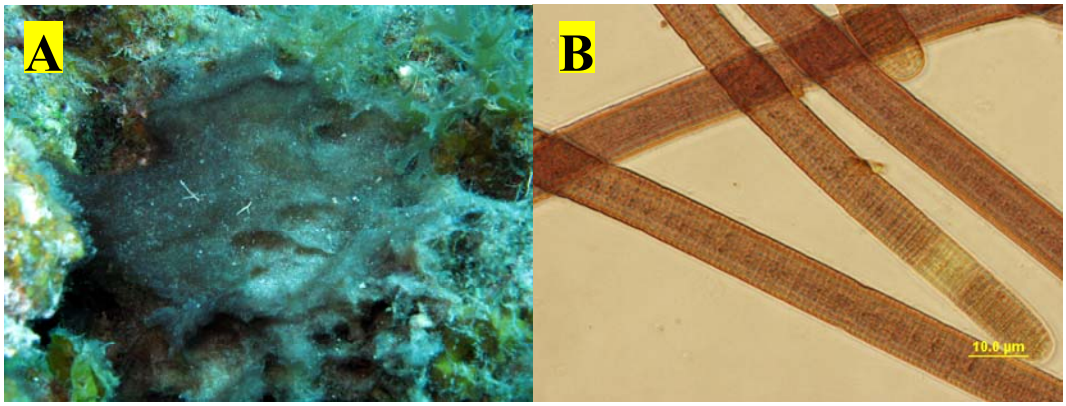


Figure II.1: (A) Underwater pictures of cf. *Oscillatoria margaritifera* NAC8-46 from 8−10 m in Caracas baii, Curaçao. (B) Photomicrograph (400X) of filaments of cf. *O. margaritifera* NAC8-46.

The SSU (16S) rRNA genes, however, revealed a relatively close phylogenetic relation with the *Oscillatoria* reference strain PCC 7515 as well as other marine *Oscillatoria* species (Figure II.2). The genera *Lyngbya* and *Oscillatoria* are evolutionarily distant, but share a number of morphological features and are therefore often mistaken for each other (Castenholz, 2001). As a consequence, the four specimens were re-classified as members of the genus *Oscillatoria* based on their phylogenetic position in relation to the *Oscillatoria* reference strain. The specimens were more specifically identified as *Oscillatoria margaritifera* (Kützing *ex* Gomont, 1892) based on a polyphasic combination of phylogenetic and morphological characters (Figure II.2; Table II.1).

Table II.1: Morphological Characterization of cf. *Oscillatoria margaritifera* Strains from Curaçao.[1]

Strain	Thallus	Sheaths	Filament width (µm) [2]	Cell width (µm) [2]	Cell length (µm) [2]	Cross-wall Constriction [3]	Apical cells
NAC8-45	Mats/ clumps	Thin, barely visible	20.3±0.7	16.8±0.8	2.0±0.1	>0.5 (>3.0%)	slightly rounded with thickened cell walls
NAC8-46	Mats/ clumps	Thin, barely visible	22.8±0.7	20.2±1.0	1.9±0.6	>0.5 (>2.5%)	slightly rounded with thickened cell walls
NAC8-54	Mats/ clumps	Thin, barely visible	22.4±0.4	20.8±0.9	2.1±0.5	>0.5 (>2.4%)	slightly rounded with thickened cell walls
NAC8-55	Mats/ clumps	Thin, barely visible	38.0±0.7	35.4±1.0	3.1±0.3	>0.5 (1.4%)	slightly rounded with thickened cell walls

[1] Species assigned by morphology in accordance with current taxonomic systems.
[2] The size measurements presented in this table are averages in µm. Filament diameters are the average of three filament measurements and cell measurements the average of ten adjacent cells of three filaments.
[3] Amount of cell cross-wall constriction in µm and percentage of total cell width.

Because tropical marine *Oscillatoria* are also related to the genus *Trichodesmium*, the paraphyly of *Oscillatoria* suggests the need for revision of tropical marine *Oscillatoria* as a distinct and separate generic entity; hence the definition cf. *Oscillatoria margaritifera* (Figure II.2).

Interestingly, the cf. *O. margaritifera* strains were phylogenetically most closely related (p-distance = 99.3%) to a tumonoic acid-producing *Blennothrix cantharidosmum* PNG05-4 from Papua New Guinea (Figure II.2) (Clark *et al*., 2008). The morphologically similar genus *Blennothrix* has been shown to share a close evolutionary history with the genus *Oscillatoria*, indicating that taxonomic delimitation of these two genera is also needed (Abed *et al*., 2006; Clark *et al*., 2008). Other closely related strains include the Panamanian venturamide-producing *Oscillatoria* sp. PAB-21 and the viridamide-producing *Oscillatoria nigro-viridis* strain 3LOSC from Curaçao. Phylogenetic inference suggests that PAB-21 and 3LOSC strains diverged as a prior evolutionary event into a more distant basal lineage. This basal group also includes the microcolin-producing strain LP16 which originally was characterized as a *Lyngbya polychroa* (Sharp *et al*., 2009). Herein, LP16 will be assumed to be a species of *Oscillatoria* because of its clear phylogenetic nesting within the *Oscillatoria* clade. The original tumonoic acid-producer, also described as a *Lyngbya*, was likely also misidentified because of the lack of phylogenetic inference (Harrigan *et al*., 1999).

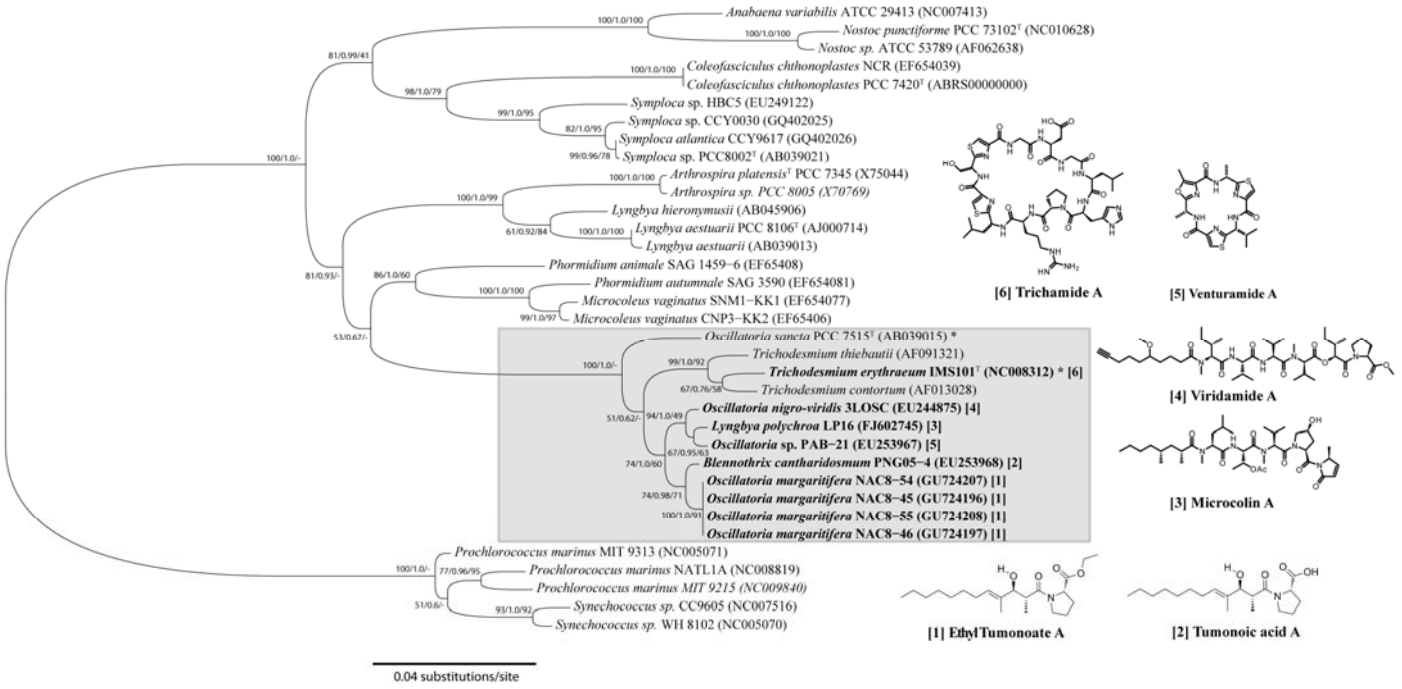


Figure II.2: Maximum-likelihood (PhyML) phylogenetic inference of the cf. *Oscillatoria margaritifera* strains based on the SSU (16S) rRNA gene. The evolutionarily distant *Gloeobacter violaceus* PCC 7421 was used as outgroup. A lineage of the unicellular order *Chroococcales* and a lineage of the heterocystous order Nostocales as well as several genera of the order Oscillatoriales, were added to place the *Oscillatoria* lineage in a broader taxonomic perspective. Representative reference strains were selected from *Bergey's Manual of Systematics* and are distinguished with asterisks. The *Oscillatoria* lineage is defined on the basis of the phylogenetic nesting with the *Oscillatoria* reference strain PCC 7515^{T} and is highlighted. Specimens are designated as taxa, strain and acc. Nr. in brackets. The secondary metabolites produced by the *Oscillatoria* species are numbered next to the *Oscillatoria* lineage. Corresponding secondary metabolite producing strains are numbered with the equivalent number. The statistical supports are indicated at each node according to the maximum-likelihood (PhyML), Bayesian Inference (MrBayes), Maximum Parsimony (MP) methods. The scale bar is equivalent to 0.04 substitutions per nucleotide position.

II.3.2 Natural Products Screening

On the basis of the close evolutionary relatedness with the tumonoic acid-producing *B. cantharidosmum*, the cf. *O. margaritifera* strains were predicted to have a high likelihood of possessing the genes encoding for the biosynthesis of these or closely related molecules. Therefore, these four cf. *O. margaritifera* strains were screened for tumonoic acid-type molecules by a combination of LC-ESI-MS and MALDI-TOF-MS. All four cf. *O. margaritifera* populations displayed molecular ions corresponding to the known tumonoic acids A−C and E−G, as well as their methyl esters (methyl tumonoates A and B). The identities of these were confirmed by comparison of retention times and MS/MS fragmentation patterns with the authentic compounds (Table II.2).

Table II.2: Secondary Metabolites Dereplication.

Secondary metabolite	MALDI-TOF-MS [1]	(+) ESI [2]	RT [3]	ESI-MS/MS [4]
Tumonoic acid A ($C_{19}H_{33}NO_4$)	ND	340.1 (100), 341.1 (20), 342.2 (5) $[M+H]^+$, 362.2 $[M+Na]^+$, 701.2 $[2M+Na]^+$	16.15	322.1 ($C_{19}H_{32}NO_3$)
Tumonoic acid B ($C_{28}H_{47}NO_8$)	$[M+Na]^+$ 548. (100), 549.2 (20), 550.2 (5)	$[M+H]^+$ 526.2, $[M+Na]^+$ 548.3, $[2M+Na]^+$ 1073.2	32.93	508.2 ($C_{28}H_{46}NO_7$) 322.1 ($C_{19}H_{32}NO_3$)
Tumonoic acid C ($C_{27}H_{46}NO_8$)	$[M+Na]^+$ 534.2 (100), 535.2 (25), 536.2 (7)	$[M+H]^+$ 512.0, $[M+Na]^+$ 534.2, $[2M+Na]^+$ 1045.1, $[2M+2Na]^+$ 1068.2	34.76	494.8 ($C_{27}H_{44}NO_7$)
Tumonoic acid F ($C_{21}H_{38}NO_5$)	ND	$[M+H]^+$ 384.1 (100), 385.1 (25), 386.1 (5), $[M+Na]^+$ 406.2, $[2M+Na]^+$ 789.2	16.79	366.1 ($C_{21}H_{36}NO_4$), 348.1 ($C_{21}H_{38}NO_3$), 324.2 ($C_{19}H_{34}NO_3$)
Methyl tumonoate A ($C_{29}H_{49}NO_8$)	ND	$[M+H]^+$ 353.9 (100), 355.0 (20), 356.0 (5) $[M+2Na]^+$ 729.2	21.61	336.1 ($C_{29}H_{47}NO_7$)
Methyl tumonoate B ($C_{29}H_{49}NO_8$)	$[M+Na]^+$ 562.2 (100), 563.2 (25)	$[M+H]^+$ 540.1, $[M+Na]^+$ 562.3, $[2M+Na]^+$ 1101.0	35.67	522.3 ($C_{29}H_{48}NO_7$)

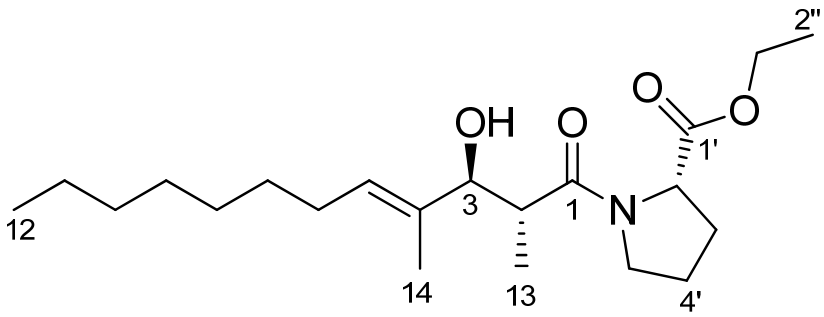
[1] *m/z* values of the secondary metabolites detected by MALDI-TOF-MS.
[2] *m/z* values of the secondary metabolites detected by ESI-MS in positive mode.
[3] Retention time (min) of detected molecules on LC-ESI-MS chromatogram using the conditions described in Experimental Procedures.
[4] Fragment ion of the secondary metabolites detected on LC-ESI-MS/MS.

II.3.3 Structure Elucidation

The substantial geographic distance between the related Pacific *B. cantharidosmum* PNG05-4 and the Caribbean cf. *O. margaritifera* strains inspired further investigation of their organic extracts by LC-ESI-MS for new variations of the tumonoic acids. A major constituent with a molecular weight of *m/z* 368 $[M+H]^+$ was detected in all four cf. *O. margaritifera* populations and database searches failed to suggest its identity.

Three midpolar vacuum liquid chromatography (VLC)-fractions of NAC8-46 contained the metabolite, and these were subsequently subjected to RP HPLC to yield a pale yellow oil in a 3.0% yield (**1**, 55.1 mg). HR-ESI-MS gave a peak at *m/z* 390.2617 $[M+Na]^+$ indicating a molecular formula of $C_{21}H_{37}NO_4$. The IR spectrum showed characteristic absorption bands at 1630 cm^{-1} and at 1733 cm^{-1}, indicating the presence of amide and ester groups in the molecule, respectively. The ^1H and ^{13}C NMR spectra of **1** showed an olefinic methine (δ_H 5.34, t; δ_C 129.0), an α proton of an amino acid (δ_H 4.39, dd; δ_C 58.7), an oxygenated methine (δ_H 4.04, d; δ_C 79.9) and an oxygenated methylene (δ_H 4.08, m; δ_C 61.0). The overlapped methylenes (δ_H 1.16−1.25) and methyls (δ_H 0.78, t; 0.99, d) suggested a branched fatty acid chain. Based on 2D NMR data analysis, all of the protons and carbons of a proline residue, an ethyl group and two terminal regions of a fatty acid were assigned. HMBC correlations from H_2-1" to C-1' indicated that the ethyl group was attached to the proline via an ester linkage. The HMBC correlations from the H-2, H_3-13 and H_2-5' to C-1 revealed that the proline and the fatty acid were linked through an amide bond. Three remaining methylenes, having similar ^1H and ^{13}C chemical

shifts (δ_H 1.18, m; δ_C 29.3, δ_H 1.17, m; δ_C 29.2, δ_H 1.16, m; δ_C 31.8), were assigned as linking the two termini of the fatty acid. A 1D NOE correlation between H_3-14 and H_2-6 indicated the *E* geometry of the double bond. The H-2 pentet (δ_H 2.67, dq, *J* = 7.2, 7.2 Hz) in the fatty acid chain suggested that H_3-13 and the 3-OH substituent were of *anti*-configuration (Table II.3) (Paik *et al*., 1994; Harrigan *et al*., 1999; Clark *et al*., 2008). The small difference in the ^{13}C NMR chemical shifts of C-3′ from C-4′ ($\Delta\delta_{C3'-C4'}$=4.3 ppm) indicated that proline peptide bond was in the *trans* conformation (Table II.3) (Dorman *et al*., 1973).



Ethyl tumonoate A (**1**)

Figure II.3: Molecular structures of Ethyl tumonoate A (**1**).

Table II.3: NMR Spectroscopic Data for Ethyl Tumonoate A (**1**) in $CDCl_3$ at 600 MHz (^1H) and 150 MHz (^{13}C).

Position	**δ_C**	**δ_H multi (*J* in Hz)**	**COSY**	**HMBC [1]**
Fatty acyl group				
1	174.9			
2	41.1	2.67, dq (7.2, 7.2)	2-Me, 3	1, 3, 4, 13
3	79.9	4.04, d (7.2)	2	1, 2, 4, 5, 13, 14
4	134.2			
5	129.0	5.34, t (7.1)	6	3, 6, 7, 14
6	27.6	1.92, m	5, 7	4, 5, 7, 8
7	29.4	1.25, m	6	5
8	29.3	1.18, m		
9	29.2	1.17, m		
10	31.8	1.16, m		
11	22.7	1.18, m	12	
12	14.1	0.78, t (7.0)	11	10, 11
13	14.4	0.99, d (7.2)	2	1, 2, 3
14	11.4	1.52, s		3, 4, 5
3-OH		3.32, brs		
Proline				
1′	172.2			
2′	58.7	4.39, dd (8.6, 4.0)	3′	1′, 3′, 4′, 5′
3′	29.1	2.10, m, 1.88 m	2′, 4′	1′, 2′, 4′, 5′
4′	24.8	1.96, m , 1.90 m	3′, 5′	2′, 3′, 5′
5′	47.0	3.57, m	4′	2′, 3′, 4′
Ethyl group				
1″	61.0	4.08, m	2″	1′, 2″
2″	14.1	1.16, t (7.3)	1″	1″

[1] From ^1H to the indicated ^{13}C.

The absolute configuration of proline in ethyl tumonoate A (**1**) was determined by acid hydrolysis followed by derivatization with Marfey's reagent fluorodinitrophenyl-5-L-valine amide (FDVA) and subsequent LCMS analysis, and revealed that it was of L configuration.[24] The absolute configuration of the 3-OH group was determined as *S* by derivatization with Mosher's reagent (*R*/*S*-MTPA-Cl) followed by ^{1}H NMR data analysis (Figure II.4) (Dale *et al*., 1969). Therefore the absolute configuration of the 2,4-dimethyl-3-hydroxydodec-4-enoic acid was 2*R*,3*S*.

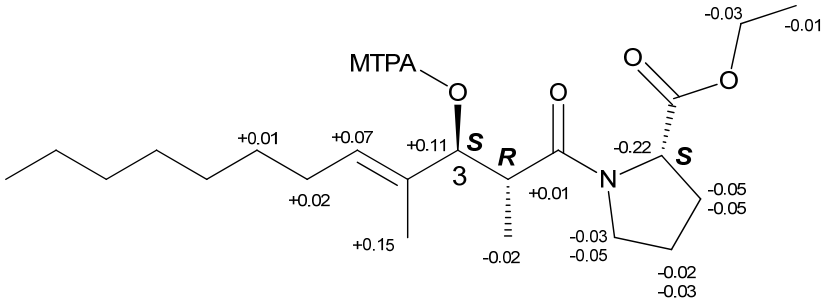


Figure II.4: $\Delta\delta_{S-R}$ values around C-3 between the Mosher's esters of ethyl tumonoate A (**1**).

II.3.4 MALDI-TOF-MS Analysis

Live filaments from cultured specimens of cf. *O. margaritifera* were screened by MALDI-TOF-MS for methyl and ethyl tumonoates to confirm that these molecules were real secondary metabolites and not isolation artifacts. Molecular weights of *m/z* 368 $[M+H]^+$ and *m/z* 562.2 $[M+Na]^+$ were detected, corresponding to ethyl tumonoate A and methyl tumonoate B, respectively. Methyl esters of NRPS-type metabolites have been shown to be produced in other bacteria by an unusual *S*-adenosyl-L-methionine (SAM)-dependent methyl-transferase (Weinig *et al*., 2003). Additionally, naturally-occurring methyl esters have been found as natural products in other cyanobacterial collections (Simmons *et al*., 2008). From the MALDI-TOF MS experiments described above and this precedence, we conclude that the ethyl group of compound **1** and methyl group of methyl tumonoate B are true natural products of cf. *O. margaritifera*.

II.3.5 Bioassays

Ethyl tumonoate A (**1**) shares structural resemblance with a number of other cyanobacterial secondary metabolites, such as the viridamides and the microcolins; the latter are derived from phylogenetically related cyanobacteria (Figure II.2). These molecules are composed of 8−12 carbon-long fatty acid chains connected with amino acid moieties, a structure type predicted to be biosynthesized by mixed polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) clusters. Because of the structural similarities of the molecules as well as the evolutionary relatedness of the organisms, these secondary metabolites are likely derived from homologous PKS/NRPS pathways. However, marine cyanobacteria have often been found to contain multiple,

genetically different biosynthetic pathways (Jones & Monroe, 2011). Thus, the fact that the venturamides are structurally distinct from the aforementioned secondary metabolites is most likely explained by their biosynthesis involving paralogous genetic pathways. The venturamides appear structurally more similar to trichamide, a metabolite isolated from the related cyanobacterium *Trichodesmium erythraeum* IMS 101^{T}.

The structural resemblance of the tumonoic acids to bacterial homoserine lactones stimulated investigations into the quorum sensing activities of these molecules (Clark *et al*., 2008). In fact, several of the tumonoic acids moderately inhibit cell-to-cell communication, resulting in a reduction in bioluminescence in *Vibrio harveyi* (Clark *et al*., 2008). Similarly, the potent immunosuppressive activity in the structurally and evolutionary related microcolins A and B was a rationale for testing ethyl tumonoate A for these activities (Zhang *et al*., 1997). Ethyl tumonoate A (**1**) showed *in vitro* anti-inflammatory activity in the RAW264.7 murine macrophage cell-based nitric oxide assay with an IC_{50} of 9.8 μM (3.6 μg/mL) with little or no cytotoxicity. Additionally, only a low level of toxicity was observed at concentrations of 10 μg/mL or higher against H-460 human lung tumor cells. The low toxicity of ethyl tumonoate A (**1**) corresponds with the lack of toxicity previously reported for related tumonoic acid analogs (Harrigan *et al*., 1999; Clark *et al*., 2008).

The original VLC fractions were also screened for their ability to activate or suppress calcium flux in neocortical neurons. VLC fractions E and F produced calcium influx at high concentrations (0.05 mg/mL), while fractions E, F, and G suppressed spontaneous calcium oscillations at lower concentrations. Because these fractions predominately contained the tumonoic acids, these metabolites were isolated and

individually tested as pure compounds in the same assay. While ethyl tumonoate A at 10 μM produced a nearly complete inhibition of Ca^{2+} oscillations, tumonoic acids A and F showed partial inhibition at this concentration, and tumonoic acid G was completely inactive.

II.4 Conclusions

Evolutionary relationships inferred from common house-keeping genes have had an enormous impact on cyanobacterial taxonomy and systematics (Komárek & Anagnostidis, 2005). Moreover, phylogenetics has been increasingly informative for natural products research in the characterization, comparison and recollection of natural products-producing organisms. In this study, we demonstrated a direct correlation between phylogenetics and class of secondary metabolite production in marine cyanobacteria. It should be noted that in fresh-water cyanobacteria, secondary metabolites such as microcystins and cylindrospermopsins, have been reported from phylogenetically unrelated genera (Stüken & Jakobsen, 2010). However, the origin of these toxins' biosynthetic pathways remains unclear as it has not been established whether they have been laterally transferred or have evolved from an ancient ancestral pathway followed by gene loss in non-producing taxa (Stüken & Jakobsen, 2010). The direct correlation found in marine cyanobacteria allows for the prediction of orthologous biosynthetic pathways and production of specific secondary metabolites. Herein, we applied this knowledge to investigate related but geographically distant cyanobacterial populations for new bioactive secondary metabolites. This phylogeny-guided isolation

approach led to the discovery of ethyl tumonoate A, a new tumonoic acid derivative with anti-inflammatory and ion modulatory activities.

II.5 Acknowledgment

We dedicate this work to J. Komárek for his contributions to cyanobacterial taxonomy. We gratefully acknowledge the government of Curaçao for collection permits and the CARMABI Research Station for assistance and research support. We also thank J. Nunnery for assistance with the cyanobacterial collections. This research was supported by the Halliday award (SIO) and NIH Grant NS 053398.

The text of II, in full, is in press as it appears in: Engene, N., Choi, H., Esquenazi, E., Byrum, T., Villa, F. A., Cao, Z., Murray, T. F., Dorrestein, P. C., Gerwick, L., Gerwick, W. H. (2011). Phylogeny-guided Isolation of Ethyl Tumonoate A from the Marine Cyanobacterium cf. *Oscillatoria margaritifera*. *J Nat Prod* The dissertation author was the primary investigator and author of this paper.

II.6 Appendix - Supporting Information for Chapter II

Table of Contents of Appendix

Figure II.5: HR-ESI-MS of Ethyl tumonate A.

Figure II.6: ^{1}H-NMR of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.

Figure II.7: ^{13}C-NMR spectrum of Ethyl tumonate A, $CDCl_3$, 150 MHz, 256 scans.

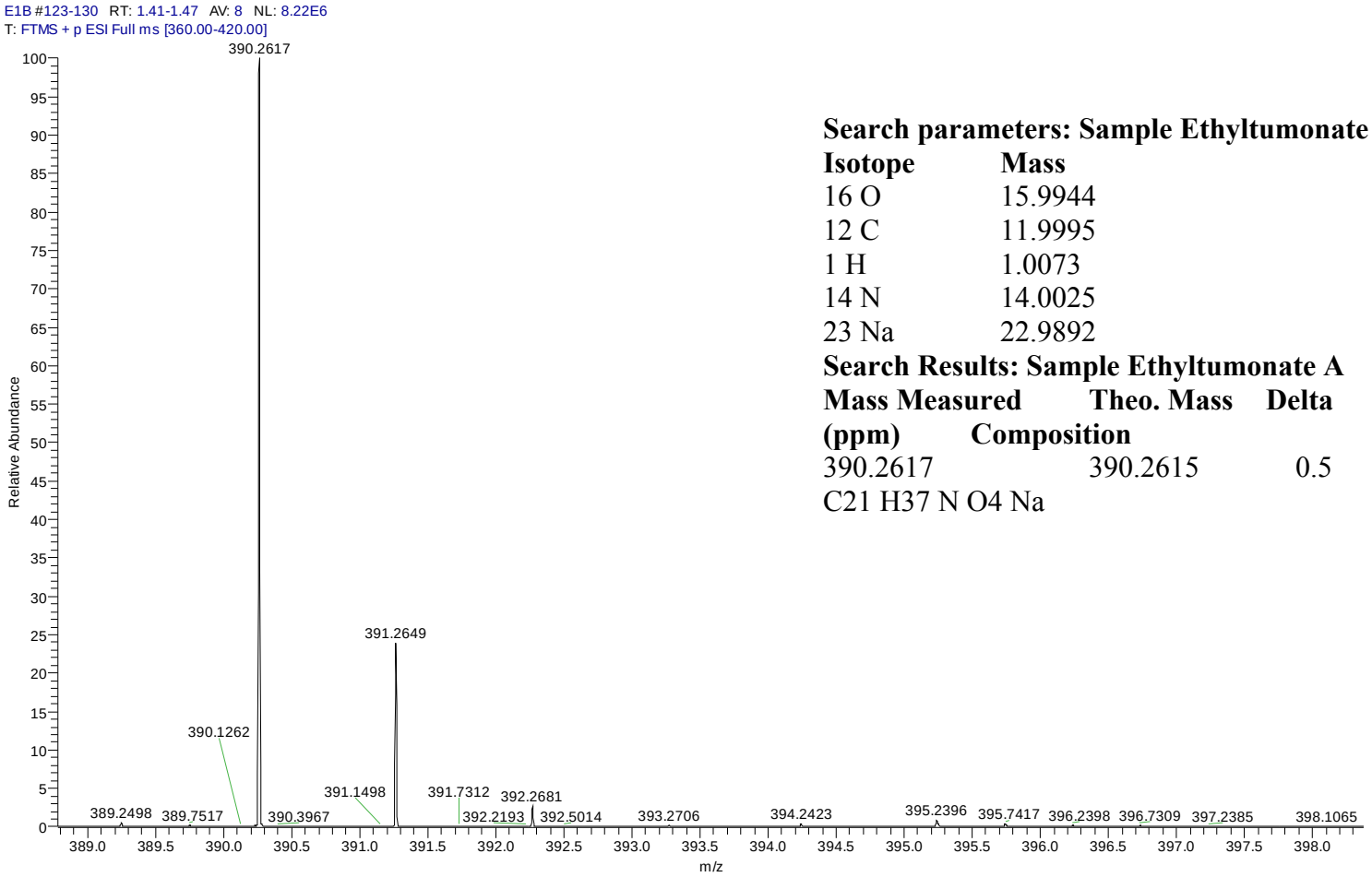
Figure II.8: COSY spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

Figure II.9: HSQC spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

Figure II.10: HMBC spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

Figure II.11: ^{1}H-NMR of *S*-MTPA ester of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.

Figure II.12: ^{1}H-NMR of *R*-MTPA ester of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.



Search parameters: Sample Ethyltumonate A

Isotope	Mass
16 O	15.9944
12 C	11.9995
1 H	1.0073
14 N	14.0025
23 Na	22.9892

Search Results: Sample Ethyltumonate A

Mass Measured	Theo. Mass	Delta (ppm)	Composition
390.2617	390.2615	0.5	C21 H37 N O4 Na

Figure II.5: HR-ESI-MS of Ethyl tumonate A.

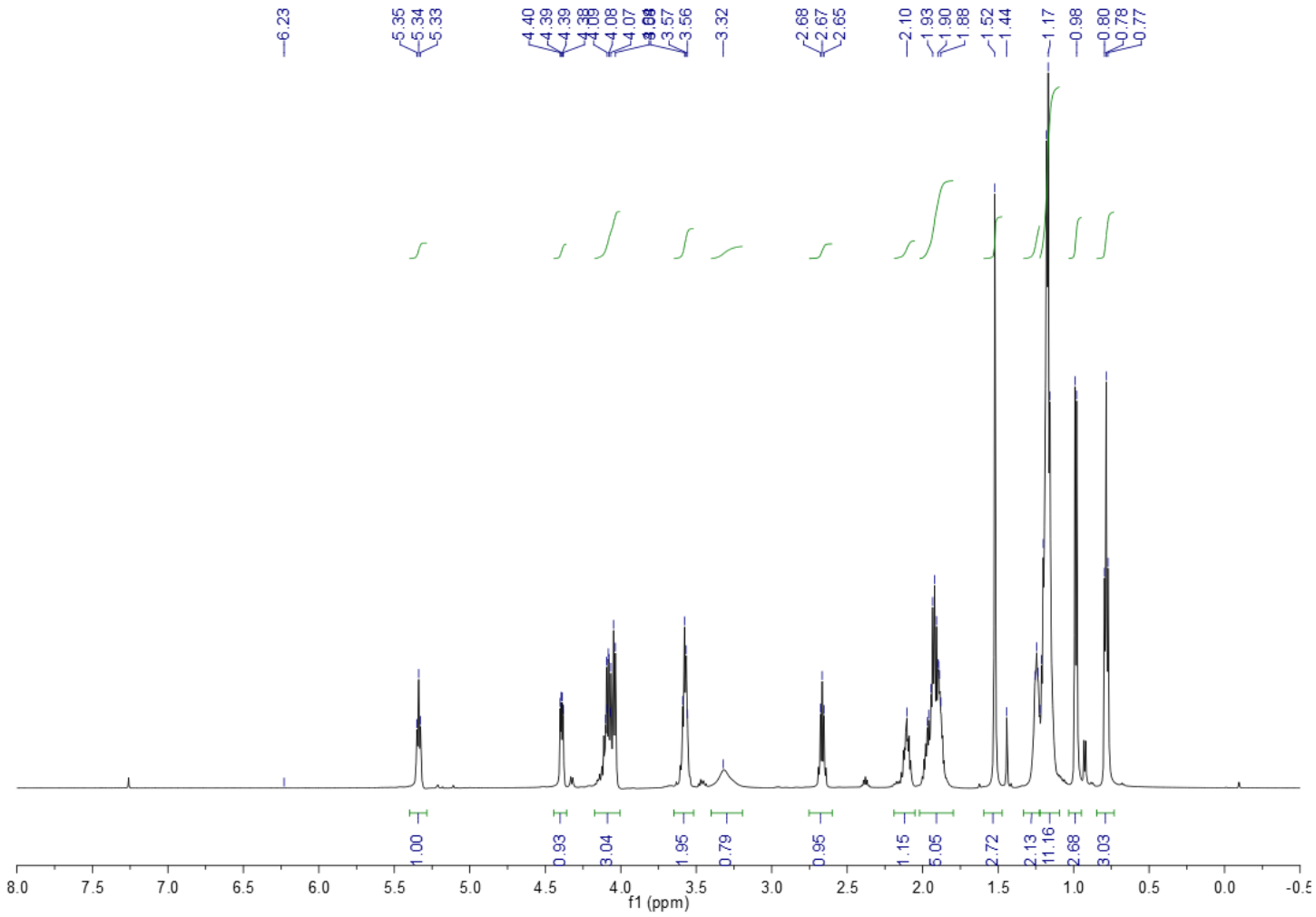


Figure II.6: ^{1}H-NMR of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.

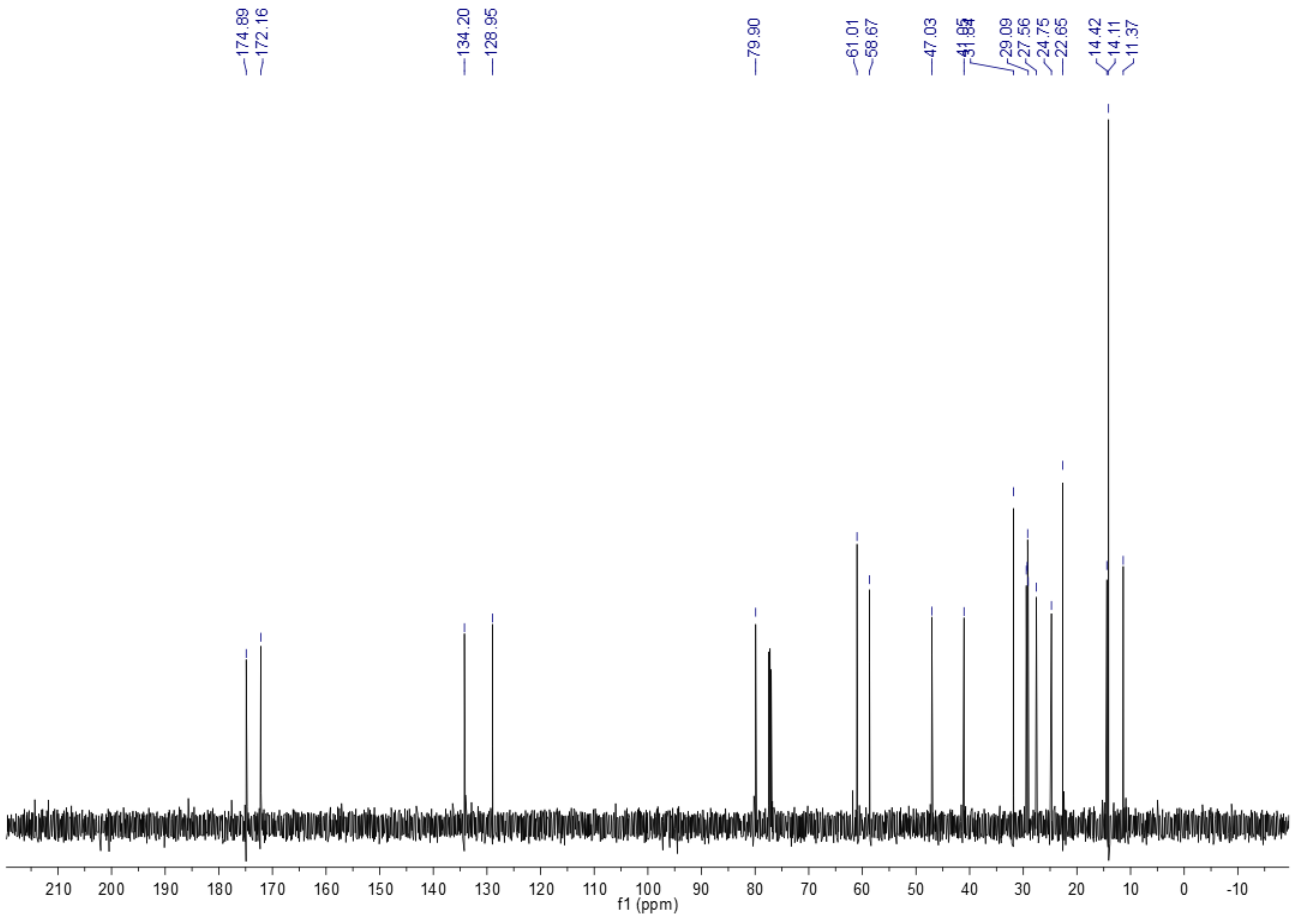


Figure II.7: ^{13}C-NMR spectrum of Ethyl tumonate A, $CDCl_3$, 150 MHz, 256 scans.

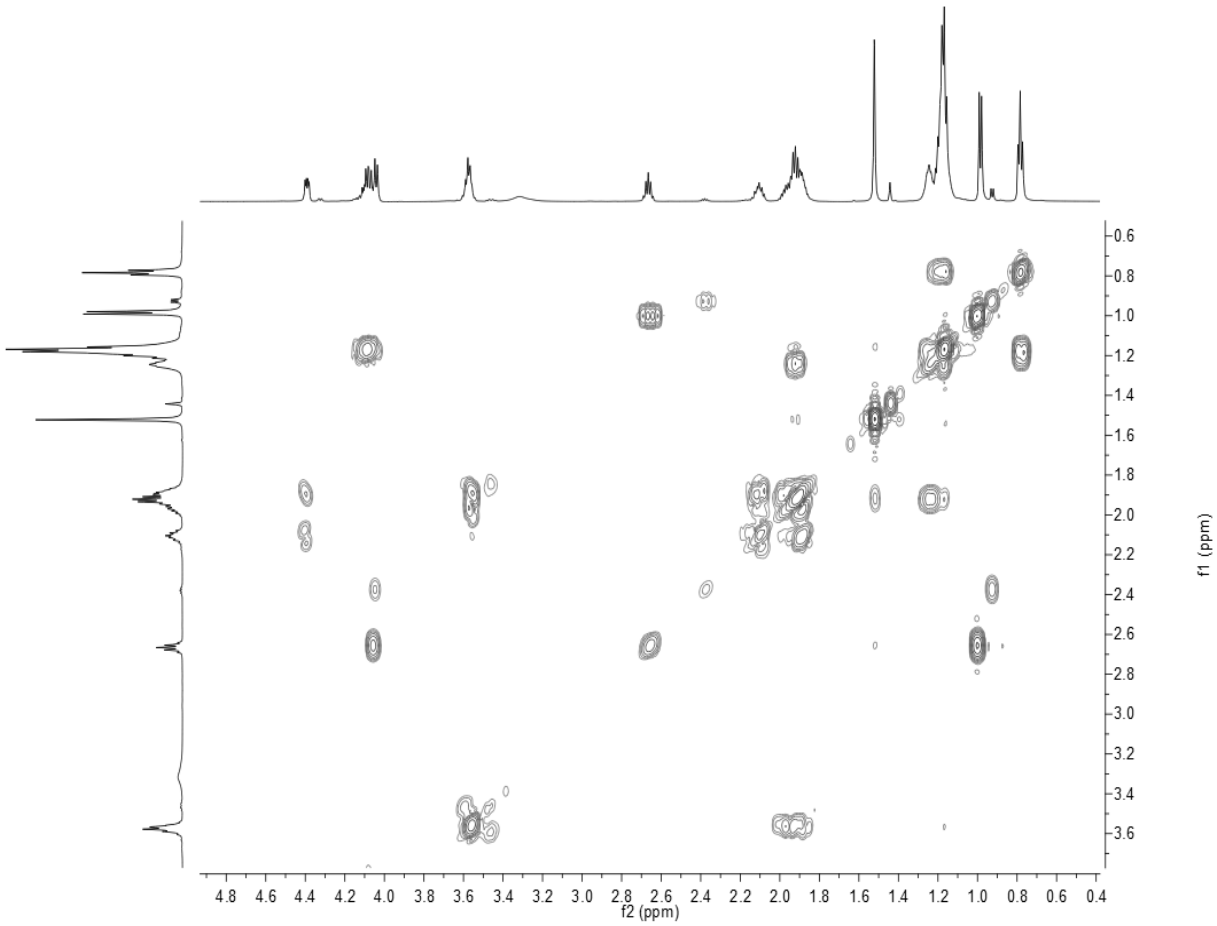


Figure II.8: COSY spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

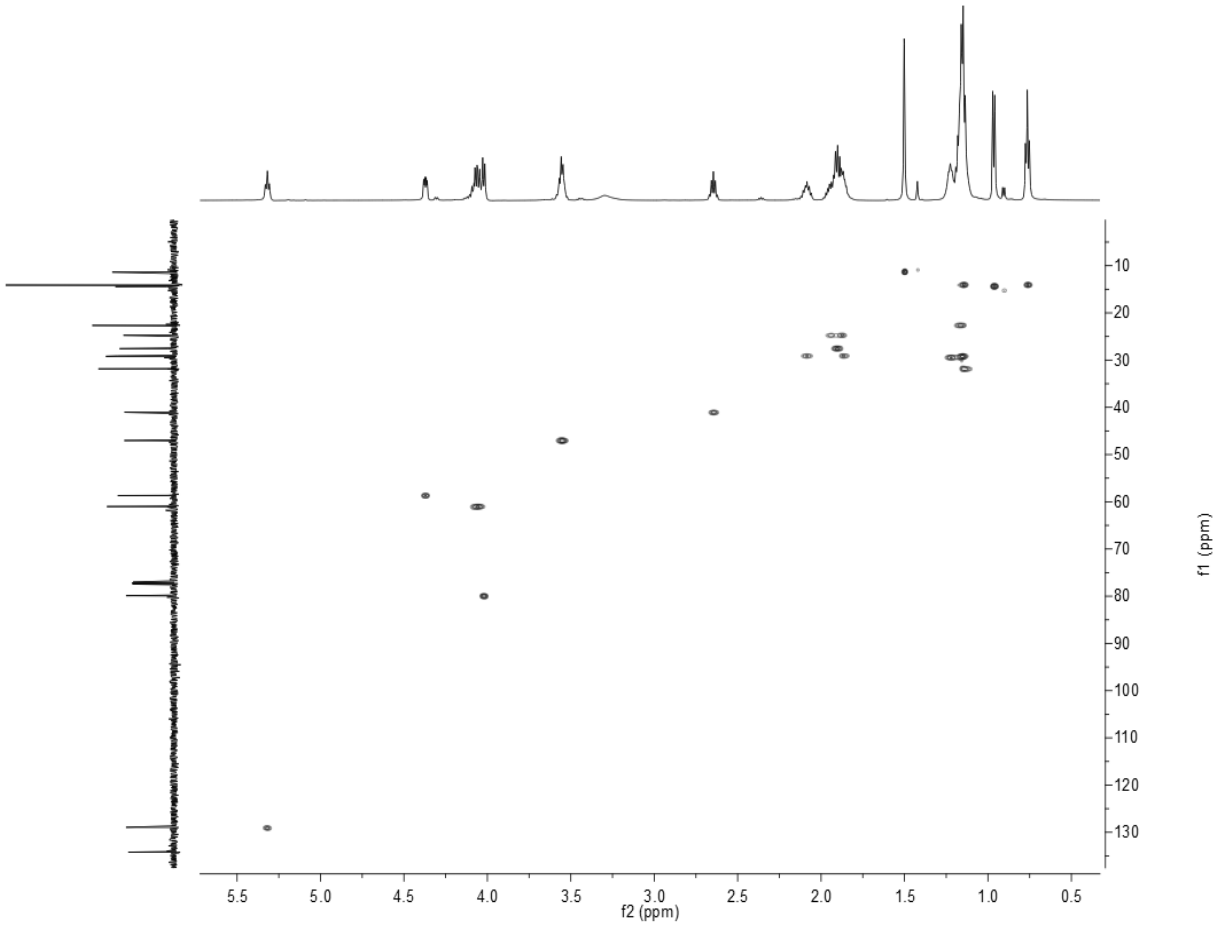


Figure II.9: HSQC spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

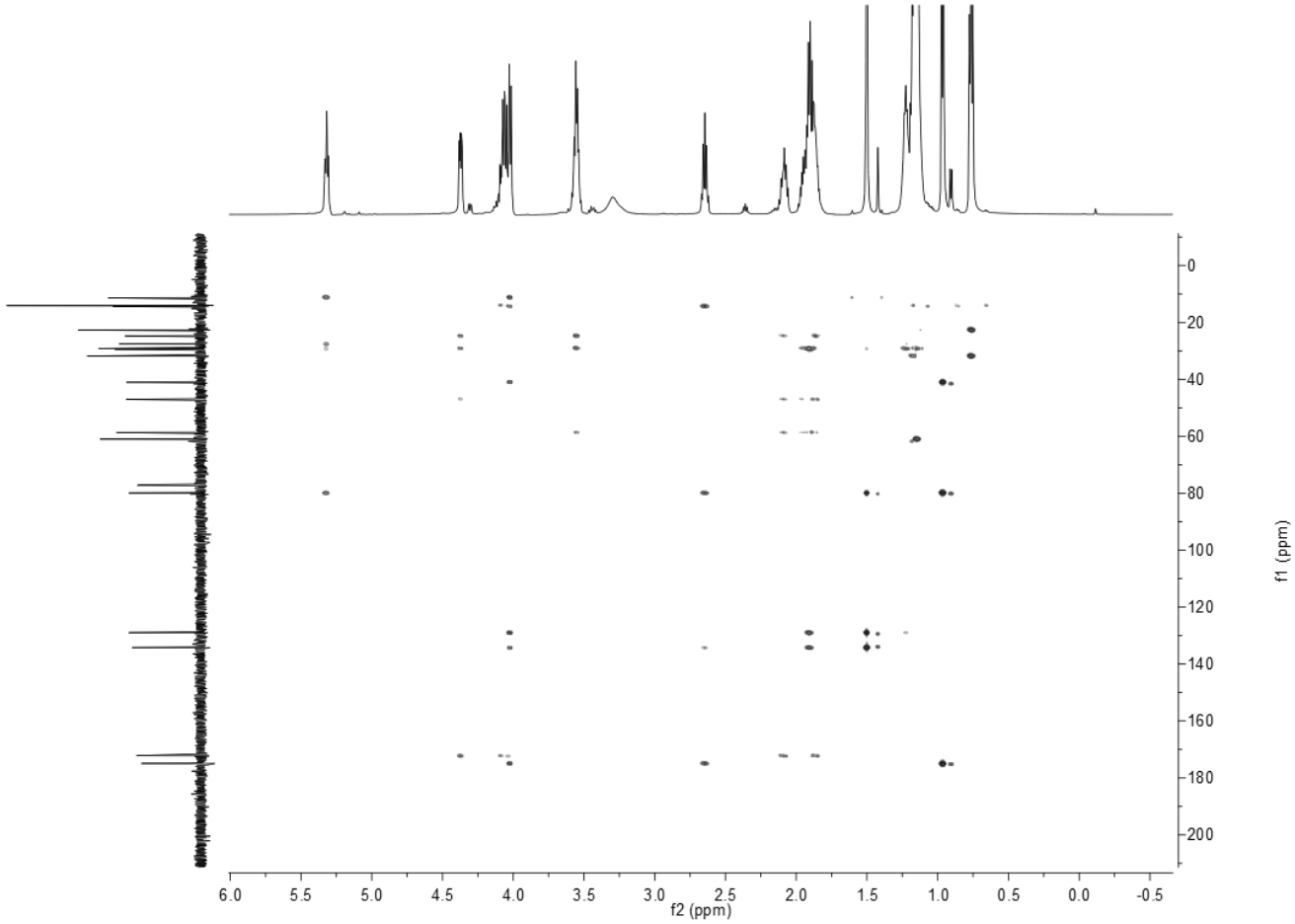


Figure II.10: HMBC spectra of Ethyl tumonate A, $CDCl_3$, 600 MHz, 16 scans.

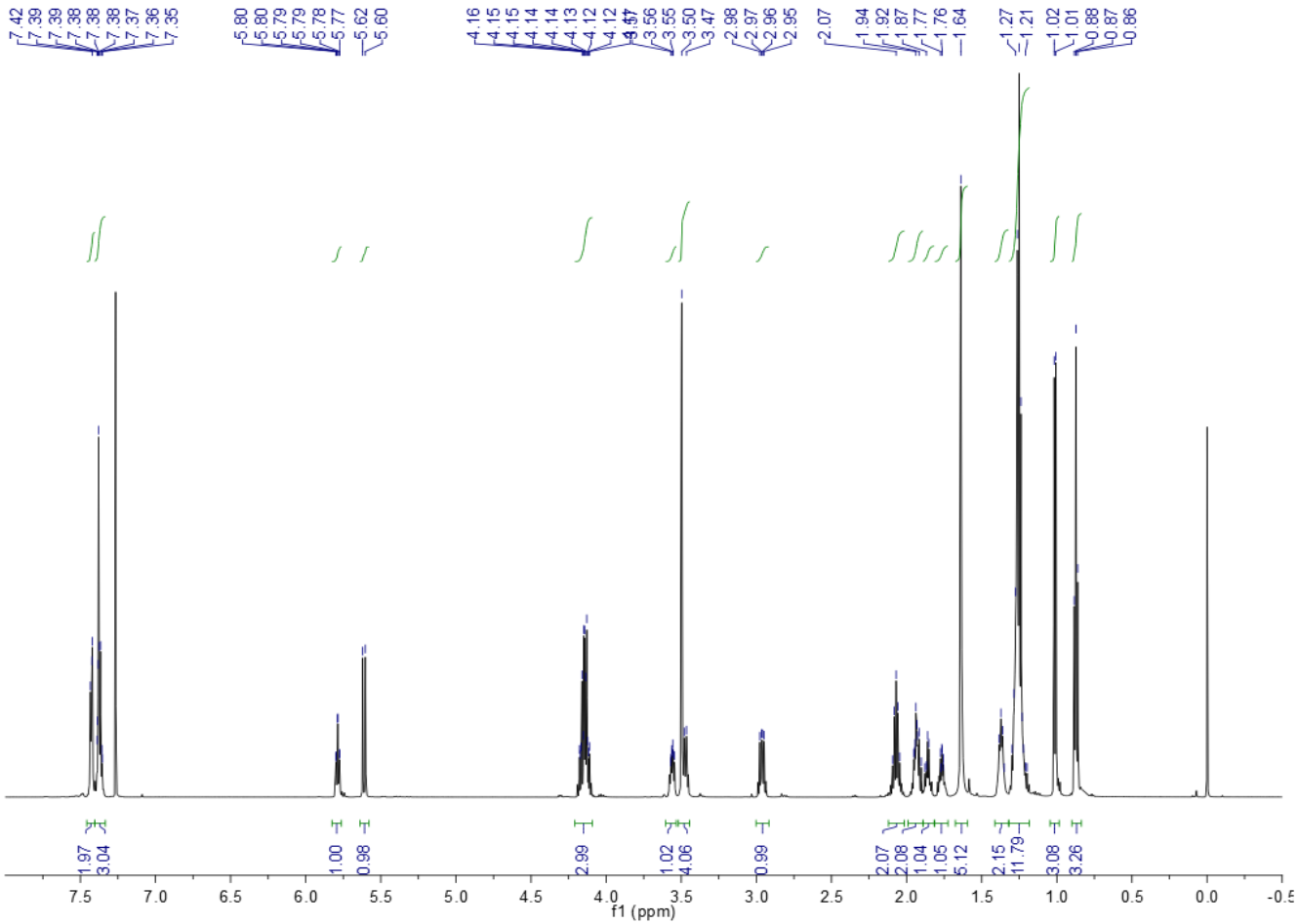


Figure II.11: ^{1}H-NMR of *S*-MTPA ester of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.

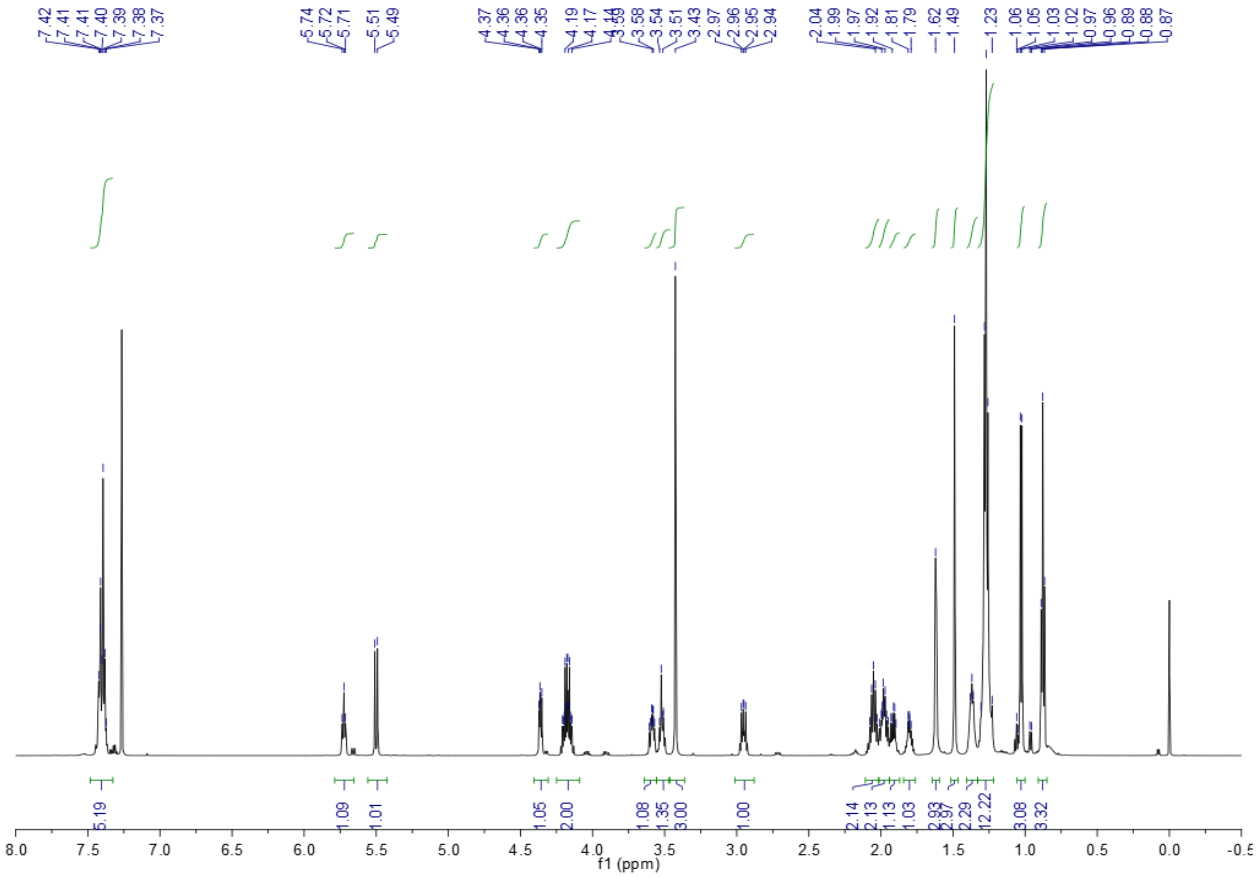


Figure II.12: ^{1}H-NMR of *R*-MTPA ester of Ethyl tumonate A, $CDCl_3$, 600 MHz, 128 scans.

II.7 References

Abed, R. M., Palinska, K. A., Camoin, G. Golubic, S. (2006). Common evolutionary origin of planktonic and benthic nitrogen-fixing oscillatoriacean cyanobacteria from tropical oceans. *FEMS Microbiol Lett* 260: 71-177.

Castenholz, R. W. (2001). Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 473-553. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Clark, B. R., Engene, N., Teasdale, M. E., Rowley, D. C., Matainaho, T., Valeriote, F. A., Gerwick W. H. (2008). Natural products chemistry and taxonomy of a *Blennothrix* species of marine cyanobacterium. *J Nat Prod* 71: 1530-1537.

Dale, J. A., Dull, D. L., Mosher, H. S. (1969). α-Methoxy-α-trifluoromethylphenylacetic acid, a versatile reagent for the determination of enantiomeric composition of alcohols and amines. *J Org Chem* 34: 2543-2549.

Dorman, D. E., Borvey, F. A. (1973). Proton coupled carbon-13 magnetic resonance spectra. Simple amides. *J Org Chem* 38: 1719-1722.

Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2011). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya*. *Environ Microbiol* 13: 1601-1610.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Gerwick, W. H. (1989). Drug discovery from marine organisms. In *Ocean Agenda 21: Passages to the Pacific Century*; Oregon State University Sea Grant; pp 59-65.

Green, L. C., Wagner, D. A., Glogowski, J., Skipper, P. L., Wishnok, J. S., Tannenbaum, S. R. (1982). Analysis of nitrate, nitrite, and [15N]-nitrate in biological fluids. *Anal Biochem* 126: 131-138.

Guindon, S., Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *System Biol* 52: 696-704.

Harrigan, G. G., Luesch, H., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Biggs, J., Park, P. U., Paul, V. J. (1999). Tumonoic acids, novel metabolites from a

cyanobacterial assemblage of *Lyngbya majuscula* and *Schizothrix calcicola*. *J Nat Prod* 62: 464-467.

Hoffman, L., Komárek, J. & Kaštovskŷ, J. (2005). System of cyanoprokaryotes (cyanobacteria) - state in 2004. *J Arch Hydrobiol* 159: 95-115.

Jones, A. C., Monroe, E. A., Podell, S., Hess, W., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R., Yates III, J. R., Reinhardt, R., Kube, M., Burkart, M., Allen, E. E., Dorrestein, P. C., Gerwick, W. H., Gerwick, L. (2011). Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* 108: 8815-8820.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae* 21: 349-375.

Komárek, J., Anagnostidis, K. (2005). In *Süßwasserflora von Mitteleuropa* 19/2; Büdel, B., Gärtner, G., Krienitz, L., Schagerl, M., Eds. Gustav Fischer: Jena.

Meterological service of Netherlands Antilles and Aruba (http://www.weather.an/reports/documents/EffectsofOmar.pdf).

Marfey, P. (1984). Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res Commun* 49: 591−596.

Mosman, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J Immunol Methods* 65: 55-63.

Nübel, U., Garcia-Pichel, F. & Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63: 3327-3332.

Paik, S., Carmeli, S., Cullingham, J. Moore, R. E. Patterson, G. M. L. Tius, M. A. (1994). Mirabimide E, an unusual N-acylpyrrolinone from the blue-green alga *Scytonema mirabile*: structure determination and synthesis. *J Am Chem Soc* 116: 8116-8125.

Ronquist, F., Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxf)* 12: 1572-1574.

Sharp, K., Arthur, K. E., Gu, L., Ross, C., Harrison, G., Paul, V. J. (2009). Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* species (Cyanobacteria: Oscillatoriales) from reefs of southeastern Florida. *Appl Environ Microbiol* 75: 2879-2888.

Simmons, T. L., Engene, N., Ureña, L. D., Romero, L. I., Ortega-Barría, E., Gerwick, L., Gerwick, W. H. (2009). Viridamides A and B, lipodepsipeptides with anti-protozoan activity from the marine cyanobacterium *Oscillatoria nigro-viridis*. *J Nat Prod* 71:1544-1550.

Simmons, T. L., Coates, R. C., Clark, B. R., Engene, N., Gonzalez, D., Esquenazi, E., Dorrestein, P. C., Gerwick, W. H. (2008). Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proc Natl Acad Sci USA* 105: 4587-4594.

Stüken, A., Jakobsen, K. S. (2010). The cylindrospermopsin gene cluster of *Aphanizomenon* sp. strain 10E6: organisation and recombination. *Microbiol* 156: 2438-2451.

Sumina, E. L. (2006). Behavior of filamentous cyanobacteria in laboratory culture. *Microbiol* 75: 459-464.

Tan, L. T. (2007). Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochem* 68: 954-979.

Taniguchi, M., Nunnery, J. K., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., Gerwick, W. H. (2009). Palmyramide A, a cyclic depsipeptide from a Palmyra Atoll collection of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73: 393-398.

Thacker, R. W., Paul, V. J. (2004). Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales). *Appl Env Microbiol* 70: 3305-3312.

Tidgewell, K., Engene, N., Byrum, T., Media, J., Doi, T., Valeriote, F. A., Gerwick, W. H. (2010). Evolved diversification of a modular natural product pathway: apratoxins F and G, two cytotoxic cyclic depsipeptides from a Palmyra collection of *Lyngbya bouillonii*. *J ChemBioChem* 11: 1458-1466.

Weinig, S., Hecht, H-J., Mahmud, T., Müller, R. (2003). Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. *Chem Biol* 10: 939-952.

Wilmotte, A., Herdman, M. (2001). Phylogenetic Relationships Among Cyanobacteria Based on 16S rRNA Sequences. In *Bergey's Manual of Systematic Bacteriology*. 1:487-493. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Zhang, L-H., Longley, R. E., Koehn, F. E. (1997). Antiproliferative and immunosuppressive properties of microcolin A, a marine-derived lipopeptide. *Life Sci* 60: 751-762.

Chapter III

16S rRNA Gene Heterogeneity in the Filamentous Marine Cyanobacterial Genus *Lyngbya*

Abstract

The SSU (16S) rRNA gene was used to investigate the phylogeny of the cyanobacterial genus *Lyngbya* as well as examined for its capacity to discriminate between different marine species of *Lyngbya*. We show that *Lyngbya* forms a polyphyletic genus composed of multiple marine, brackish, and fresh-water lineages. In addition, we found morphological and genetic evidence that *Lyngbya* spp. often grow in association with other microorganisms, in particular smaller filamentous cyanobacteria such as *Oscillatoria*, and propose that these associated microorganisms have led to extensive phylogenetic confusion in identification of *Lyngbya* spp. At the species level the phylogenetic diversity obtained from the comparison of 16S rRNA genes exceeded morphological diversity in *Lyngbya*. However, the expectation that this improved phylogeny would be useful to species and sub-species identification was eliminated by the fact that phylogenetic species did not correlate in any respect with the species obtained from current taxonomic systems. In addition, phylogenetic identification was adversely affected by the presence of multiple gene copies within individual *Lyngbya* colonies. Analysis of clonal *Lyngbya* cultures and multiple displacement amplified (MDA) single-cell genomes revealed that *Lyngbya* genomes contain two 16S rRNA gene copies, and that these typically are of variable sequence. Furthermore, intra-genomic and inter-species 16S rRNA gene heterogeneity were found to be of approximately the same magnitude. Hence, the intra-genomic heterogeneity of the 16S rRNA gene overestimates

the microdiversity of different strains, and does not accurately reflect speciation within cyanobacteria, including the genus *Lyngbya*.

III.1 Introduction

Cyanobacteria (Cyanophytes) constitute a monophyletic group within the domain Bacteria (Castenholz, 2001). However, as a phylum, they are morphologically diverse, ranging from microscopic unicellular organisms to macroscopic mats of filamentous colonies (Castenholz, 2001). Despite the large size of some of these colonies, morphological identification can often be challenging due to deceiving physical features such as coloration or variable growth forms (Sumina, 2006). Traditionally, cyanobacteria were classified according to phycological taxonomic systems using botanical nomenclature under the International Code of Botanical Nomenclature (Oren, 2004). This phycological system consists of a combination of phenotypic and ecological characters (Komárek & Anagnostidis, 2005). Relatively recent acceptance of cyanobacteria as a prokaryote phylum within the domain Bacteria has required the characterization of cyanobacteria according to the International Code of Nomenclature of Prokaryotes (Stanier, 1978). Thus, from a historical perspective the taxonomy of this group of organisms has been described by two separate systems, the traditional phycological system and a modern bacteriological system (Oren, 2004).

In adherence with the bacteriological code the use of the 16S rRNA gene has become common in the identification and cataloging of cyanobacteria, either as the sole method of identification or in combination with morphological, phenotypic, and ecological characterization (Castenholz, 2001). Correlations have been uncovered between SSU (16S) rRNA genes and whole-genome similarity, further underscoring the

reliability of 16S rRNA gene use in determining phylogeny (Clarridge, 2004). The unicellular baeocyte-forming Pleurocapsales and heterocyst-forming cyanobacteria are examples of phylogenetically and morphologically coherent groups (Wilmotte *et al*., 2001, Tomitani *et al*., 2006). This correlation between morphology and 16S rRNA gene sequence has also been shown to apply relatively well with some genera of the filamentous, non-heterocystous order Oscillatoriales (Subsection III) including: *Planktothrix* (Komárek 2006), *Trichodesmium* (Wilmotte et al. 2001), *Blennothrix* (Abed et al. 2006), *Oscillatoria* (Simmons et al. 2008), *Tychonema* (Komárek 2006), *Arthrospira* (Komárek 2006), and *Spirulina* (Abed et al. 2003). However, the majority of traditional genera in this order have been revealed to be heterogeneous and consequently in need of revision (Komárek 2006).

Another strong argument for the molecular taxonomic approach is that genotypic diversity typically exceeds phenotypic diversity. It has therefore been suggested that molecular-phylogenetics allows for much more distinct and fine discrimination between taxonomic groups than morphological methods (Clarridge, 2004). In principal, the higher level of genetic variation compared with morphological variation has been proposed to allow cataloging at a sub-species level (Moffitt *et al*., 2001).

Underlying the usage of the 16S rRNA gene in determining microbial identity and diversity is the assumption that this gene only occurs as one copy per genome, or if several gene copies are present, that they are identical in sequence. With regard to bacteria, this assumption may not always apply because they typically have several different copies of their ribosomal operons (rrn) and consequently several 16S rRNA genes. For example, the number of rrn copies ranges from single copies up to 15 in some

bacteria with an average number of 4.2 (Case *et al*., 2007). Moreover, in 56-62% of bacterial genomes the 16S rRNA genes in at least two of their rrn's differ from one another (Coenye & Vandamme, 2003; Case *et al*., 2007). In rare cases, genomes with multiple operons have heterogeneity in all of their 16S rRNA gene copies. For example, *Bacillus subtilis* and *Clostridium perfringens* both have ten rrn copies wherein all ten 16S rRNA genes are different from one another (Acinas *et al*., 2004). Intra-genomic 16S rRNA gene sequence heterogeneity is sometimes present in more than 5% of the nucleotide positions and has been reported as high as 6.8% in *Halosimplex carlsbadense* (Boucher *et al*., 2004).

In the present study, the genus *Lyngbya* Agardh ex Gomont was used as a model taxonomic group to study the phylogenetic capacity of 16S rRNA genes to describe microdiversity in cyanobacteria. The genus *Lyngbya* was selected primarily for its prominent ecological role in marine ecosystems. *Lyngbya* is important to oxygen production and nitrogen fixation in the ocean and is a group rich in bioactive secondary metabolites (Hoffmann, 1994; Gerwick *et al*., 2008). In addition to its ecological importance, the genus *Lyngbya* represents a morphologically cohesive taxonomic group, which has remained a conserved genus since first being described in Gomont's *Monographie des Oscillariées* in 1892 (Gomont, 1892; Geitler, 1932; Desikachary, 1959; Komárek, 2005). The relative morphological and ecological ease with which *Lyngbya* specimens can be identified and the extensive, well-described key-systems available for *Lyngbya* spp. are factors which favor the use of *Lyngbya* as a model taxon for investigating cyanobacterial phylogeny (Komárek, 2005). Nevertheless, errors and oversimplifications in phylogenic assignments have created considerable confusion in the

literature for this genus. For example, multiple 16S rRNA gene sequences varying in 1.9% of their nucleotide positions have been identified in a single *Lyngbya* specimen from Guam (Luesch *et al*., 2002), polyphyletic groupings have been reported from identical *Lyngbya* species (Thacker *et al*., 2004; Paul *et al*., 2005; Sharp *et al*., 2009), and different *Lyngbya* species identified by morphology have been observed with identical 16S rRNA genes (Gutiérrez *et al*., 2008). The primary purpose of this study was to clarify the phylogeny of the genus *Lyngbya* and investigate the capacity of the 16S rRNA gene to delineate different *Lyngbya* species that are well differentiated by morphological techniques. The presence of intra-genomic 16S rRNA gene heterogeneity was also investigated through *Lyngbya* single-cells by the application of multiple displacement amplification (MDA) of genomic DNA. MDA-reactions using the polymerases of the bacteriophage phi29 allow the amplification of femtograms to micrograms of DNA under isothermal conditions (Lasken, 2007). Finally, the magnitude of the intra-species and intra-genomic heterogeneity were compared and the impact on fine-scale phylogenetic resolution considered.

III.2 Materials and Methods

III.2.1 Cyanobacterial Collection

A total of 87 *Lyngbya* specimens were collected by SCUBA from Papua New Guinea (PNG), Curaçao Las Palmas (L), and Jamaica Hector's Bay (JHB) (Table III.3). Samples of each *Lyngbya* collection were directly cleaned from associated macro fauna

and washed with seawater filtered through a 0.2 μm Acrodisc® Syringe filter (PALL Life Sciences Inc.) in Petri dishes. Genetic samples were preserved in RNA*later*® (Ambion Inc.) in 10 mL Falcon™ tubes and stored at -20 ºC. Live specimens were brought back to the laboratory for culturing and morphological characterization in vented 50 mL BD Falcon™ tissue culture flasks with 20 mL filtered sea-water. Morphological voucher specimens were preserved in (70%) ethanol.

III.2.2 Morphological Characterization

Taxonomic identification of *Lyngbya* spp. was performed in accordance with bacteriological systems (Castenholz *et al*., 2005), traditional (Desikachary, 1959; Geitler, 1932) and current phycological systems (Komárek *et al*., 2005) as well as relevant taxonomic literature (Hoffmann *et al*., 1991; Hoffmann, 1994). Morphological characterizations were performed using an Olympus IX51 epifluorescent microscope (100x objective) equipped with an Olympus U-CMAD3 camera. In this study a colony was defined as an individual, separate, and distinct mass of tissue, and was observed closely by microscopy in order to establish uniform filament morphology. Heterotrophic bacteria associated with cyanobacterial sheaths were stained with a single drop of VECTASHIELD HardSet Mounting Medium with DAPI (Vector Laboratories) and visualized at 100x using a 360 nm excitation and 460 nm emission.

III.2.3 Isolation and Culturing of Clonal Strains

Field collections of cyanobacteria were carefully rinsed with autoclaved SWBG11-media and visible macroorganisms mechanically removed with sterile tweezers

under an Olympus VMZ dissecting microscope. Cultures contaminated with diatoms were treated with GeO_2 (250 $mg \cdot L^{-1}$). Clonal strains were derived from phototactically isolated single-filaments on (0.5%) agar plates with SWBG-11. The isolated *Lyngbya* filaments were cultured in SWBG-11 medium at 28 °C with 35 $g \cdot L^{-1}$ Instant Ocean (Aquarium Systems Inc, Mentor, OH, USA). The cultures were kept in a 16 h light/8 h dark cycle with a light intensity of ~7 µmol photon $s^{-1} \cdot m^{-2}$ provided by 40W cool white fluorescent lights. Scaled-up cultures for DNA extraction were produced using 6 well cell culture clusters (Corning Inc., Lowell, MA, USA) and 125 mL Erlenmeyer flasks with changes of medium every 21 days.

III.2.4 Genomic DNA Extraction and MDA of Single-cell Genomes

DNA was extracted from 40 mg of cyanobacterial filaments rinsed with dH_2O using the Wizard® Genomic DNA Purification Kit (Promega Corporation) following the manufacturer's specifications. The isolated DNA was further purified using a Genomic-tip 20/G kit (Qiagen Inc.). DNA concentration and purity was measured on a DU® 800 spectrophotometer (Beckman Coulter Inc.). Single-cells of *Lyngbya* were obtained by cutting cyanobacterial filaments using two razor blades into roughly 0.1-200 mm sections. Filaments were sonicated in a 75D bath-sonicator (VWR International) for 30 sec and passed through 300 µm wire mesh filters. Single-cells were isolated from the filtrate using a MM3-All micromanipulator (World Precision Instruments) using pulled capillary tubes with lumen sizes ranging between 60-200 µm in diameter. The single-cells were washed twice in 0.5 µL sterile SWBG-11 medium and twice in 0.5 µL dH_2O before transfer into 0.2 mL PCR-tubes. DNA was amplified from single-cells by MDA

using the REPLI-g® Mini Kit (Qiagen), following the manufacturer's specifications. All MDA reactions were performed in 50 µL reaction volume for 16 h at 30 °C.

II.2.5 PCR and Cloning

The 16S rRNA genes were PCR-amplified using the cyanobacterial-specific primers 106F and 1509R (Nübel *et al*., 1997) and the rpoC1 genes with the degenerate primers LrpoC1-F (CYTGYTTNCCYTCDATDATRTC) and LrpoC1-R (YTNAARCCNGARATGGAYGG). The PCR reaction volumes were 25 µL containing 0.5 µL (50 ng) of DNA, 2.5 µL of 10 x PfuUltra IV reaction buffer, 0.5 µL of dNTP mix (25 mM each of dATP, dTTP, dGTP, and dCTP), 0.5 µL of each primer (10 µM), 0.5 µL of PfuUltra IV fusion HS DNA polymerase and 20.5 µL dH_2O. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 2 min at 95 °C, 25 cycles of amplification: 20 sec at 95 °C, 20 sec at 50 °C and 1.5 min at 72 °C, and final elongation for 3 min at 72 °C. PCR-products were analyzed on a (1%) agarose-gel in SB-buffer and visualized by EtBr-staining. PCR products were subcloned using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) into the pCR®-Blunt IV TOPO® vector, and then transformed into TOPO® cells and cultured on LB-kanamycin plates. Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced with pCR®-Blunt IV TOPO® vector specific primers M13F and M13R. Sequencing of the 16S rRNA genes middle regions were improved using the internal primers 359F and 781R (Nübel *et al*., 1997). The gene sequences are available in the GenBank/EMBL/DDBJ databases under accession numbers: U492878, EU492877, EU315909, FJ041298-FJ041309, FJ147300-FJ147305, FJ151509-FJ151533, FJ157181-

FJ157183, FJ172665, and FJ154877-FJ154878 for *Lyngbya* 16S rRNA genes; EU24875, FJ157181, and 125592 for *Lyngbya*-associated cyanobacteria 16S rRNA genes; FJ214725, FJ217387, FJ231748, and FJ231749 for *Lyngbya* rpoC1 genes.

II.2.6 *In silico* Analysis of Cyanobacterial Gene Sequences

The *Lyngbya* gene sequences were obtained up to October 2008 from the National Center for Biotechnology Information (NCBI) web page (http://www.ncbi.nlm.nih.gov). The percent sequence divergence was determined by the pair-wise distance calculation using the Maximum Composite Likelihood method in Mega 4.0 (Tamura *et al*., 2007). The secondary RNA structures were predicted by CLC Combined Workbench 3.5.2 (CLC bio.). Mutation types and domains of the 16S rRNA genes were determined by superimposing their secondary structures on the SSU model for *Escherichia coli* J01695 (Cannone *et al*., 2002). Gene sequence anomalies, including chimeric sequences, were predicted using the Pintail software with the cut-off size set at >600 bp (Ashelford *et al*., 2005), and were manually confirmed by comparison of (NJ) phylogenetic trees for different regions (>300 bp) of the sequences.

II.2.7 Phylogenetic Analysis

The *Lyngbya* 16S rRNA genes were aligned with 69 cyanobacterial gene sequences using ClustalW XXL in MEGA 4.0 with standard gap opening and extension penalties without gaps (Tamura *et al*., 2007). The evolutionary histories of the cyanobacterial 16S rRNA genes were inferred using the Neighbor-Joining (NJ), Maximum Parsimony (MP), and Minimum Evolution (ME) algorithms in MEGA 4.0 as

well as the Bayesian (MrBayes) and maximum likelihood (ML) methods using TOPALi v2.5 (Milne *et al*., 2009). All algorithms were perfomed with 1,000 bootstrap repetitions. The evolutionary distances were computed using the Maximum Composite Likelihood method. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1 (Nei *et al*., 2000). All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option) for a total of 534 bp of the Oscillatoriales genera and 865 bp of the *Lyngbya* spp.

II.3 Results

II.3.1 Morphological and Ecological Characterization of *Lyngbya*

In total, the morphology of 40 *Lyngbya* specimens (16 *L. bouillonii* Hoffmann et Demoulin, 20 *L. sordida* Gomont ex Gomont (synonym: *L. polychroa*), and 4 *L. majuscula* Harvey ex Gomont) were analyzed in detail (Table III.1). *L. bouillonii* formed characteristic reddish, non-gelatinous colonies that were irregular but generally 10 cm in diameter and 1 cm thick. The colony morphology of *L. bouillonii* was always mat-like and tenaciously attached to the surrounding substrates (Fig. 1b). The substrate was usually coral, *e.g.* Staghorn coral (*Acropora cervicornis*), but was occasionally composed of dead materials such as rocks or wood. *L. bouillonii*, in contrast to *L. sordida* and *L. majuscula*, was often found in association with the snapping shrimp *Alpheus frontalis*. *L. sordida* formed upright turf or tuft colony morphologies that measured up to a few centimeters in height, and was often found epiphytic on macro fauna or coral (Figure III.1A). The coloration of *L. sordida* specimens was highly varied, ranging from dark-red to greenish-brown. *L. majuscula* strains typically formed tangled, mucilaginous mats (Figure III.1C). The *L. majuscula* mats were generally very large, at times reaching 20 cm in height. *L. majuscula* was typically retrieved from shallower depths and on sand or rocks compared with the other *Lyngbya* species. Chromatic variation in each *Lyngbya* species was found to be highly influenced by environmental factors such as light intensity (i.e. depth), and therefore, was unreliable for identification purposes.

The cylindrical filaments were relatively similar in shape among all species being straight or slightly waved. *L. majuscula* strains typically had wider [39.1 ± 13.6 μm (n =

number counted = 90)] and shorter cells [2.7 ± 0.9 μm (n = 90)] with a smaller cell length/width ratio [0.06 ± 0.03 (n = 90)] than the other two *Lyngbya* species. The trichomes of *L. majuscula* were also wider [45.1 ± 16.6 μm (n = 9)] and, in contrast to both *L. bouillonii* and *L. sordida,* were not constricted at the cell walls. *L. bouillonii* and *L. sordida* have relatively similar cell morphology and cellular dimensions, but were delineated by distinct colony morphologies. The cells of *L. bouillonii* [cell width 23.9 ± 2.2 μm (n = 480), cell length 6.7 ± 1.9 μm (n = 480)] were narrower and shorter than those of *L. sordida* [cell width 27.8 ± 7.0 μm (n = 570), cell length 7.2 ± 3.2 μm (n = 570)]. However, the average cell length/width ratios were similar in both species [*L. bouillonii:* 0.3 ± 0.07 (n = 480) and *L. sordida*: 0.3 ± 0.1 (n = 570)] and they both had constricted cross-walls. The trichome widths were 26.5 ± 2.1 μm (n = 48) in *L. bouillonii* and 31.9 ± 8.9 μm (n = 57) in *L. sordida*. All analyzed *Lyngbya* strains had very similarly shaped terminal cells which were slightly rounded and lacked calyptra.

Table III.1: Morphological Characterization of *Lyngbya* Specimens.

Species (strain)	Colony Morphology	Color	Trichome width[1]	Cell width[1]	Cell length[1]	Width/ length	Cell wall constr-iction[2]	Apical Cells	Calyptra
bouillonii PNG5-198	Tenacious mat	Red-brown	23.8	19.1	4.2	0.2	++	Rounded	-
bouillonii PNG6-41	Tenacious mat	Dark-red	26.7	23.6	4.7	0.2	++	Rounded	-
bouillonii PNG6-42	Tenacious mat	Red	27.9	26.2	7.7	0.3	++	Rounded	-
bouillonii PNG6-47	Tenacious mat	Red	25.8	22.7	4.3	0.2	++	Rounded	-
bouillonii PNG6-50	Tenacious mat	Red	27.3	23.4	5.4	0.2	++	Rounded	-
bouillonii PNG6-59	Tenacious mat	Red	27.3	25.8	9.3	0.4	++	Rounded	-
bouillonii PNG6-62	Tenacious mat	Red	26.2	24.1	7.2	0.3	++	Rounded	-
bouillonii PNG6-65	Tenacious mat	Dark-red	27.3	23.8	8.3	0.3	++	Rounded	-
bouillonii PNG7-6	Tenacious mat	Dark-red	26.2	24.9	6.9	0.3	++	Rounded	-
bouillonii PNG7-14	Tenacious mat	Red	26.3	24.2	4.9	0.2	++	Rounded	-
bouillonii PNG7-22	Tenacious mat	Red	25.7	22.3	6.7	0.3	++	Rounded	-
bouillonii PNG7-29	Tenacious mat	Red	30.1	27.5	8.3	0.3	++	Rounded	-

Table III.1: Continued

Species (strain)	Colony Morphology	Color	Trichome width[1]	Cell width[1]	Cell length[1]	Width/ length	Cell wall constr-iction[2]	Apical Cells	Calyptra
sordida PNG5-192	Turf	Brown	36.4	30.0	7.4	0.2	++	Rounded	-
sordida PNG5-194	Tufts	Red-brown	57.0	44.0	15.0	0.3	++	Rounded	-
sordida PNG6-2	Turf	Red	23.2	20.1	6.4	0.3	++	Rounded	-
sordida PNG6-3	Turf	Green	15.0	14.5	3.5	0.2	+	Rounded	-
sordida PNG5-194	Tufts	Red-brown	57.0	44.0	15.0	0.3	++	Rounded	-
sordida PNG6-9	Turf/tufts	Dark-green	33.0	30.1	3.9	0.1	+	Rounded	-
sordida PNG6-34	Turf/tufts	Dark-red	25.6	21.7	5.9	0.3	++	Rounded	-
sordida PNG6-38	Tufts	Red	26.1	24.0	6.5	0.2	++	Rounded	-
sordida PNG6-45	Turf/tufts	Green	35.0	32.1	7.2	0.2	++	Rounded	-
sordida PNG6-46	Tufts	Brown	36.4	29.9	6.0	0.2	++	Rounded	-
sordida PNG6-48	Turf	Brown	32.9	30.0	7.5	0.3	++	Rounded	-
sordida PNG6-51	Turf/tufts	Green	35.4	32.1	8.3	0.3	++	Rounded	-
sordida PNG6-52	Turf/tufts	Red	35.2	33.6	9.4	0.3	++	Rounded	-
sordida PNG6-54	Turf/tufts	Brown	32.0	23.3	3.7	0.2	++	Rounded	-
sordida PNG6-57	Turf/tufts	Green	30.0	26.0	2.2	0.01	++	Rounded	-

Table III.1: Continued

Species (strain)	Colony Morphology	Color	Trichome width[1]	Cell width[1]	Cell length[1]	Width/ length	Cell wall constr-iction[2]	Apical Cells	Calyptra
sordida PNG6-68	Turf	Red	27.1	25.1	7.5	0.3	++	Rounded	-
sordida PNG6-69	Tufts	Brown	37.7	31.2	10.0	0.3	++	Rounded	-
sordida PNG7-4	N/A	Green	28.2	26.5	6.2	0.2	+	Rounded	-
sordida PNG7-13	Tufts	Red	19.8	17.5	6.4	0.4	+	Rounded	-
majuscula PNG5-1913	Mat-like tufts	Red-brown	26.0	23.4	2.4	0.1	-	Rounded	-
majuscula PNG6-24	Mat-like	Blackish	53.3	47.0	1.9	0.04	-	Rounded	-
majuscula PNG5-223	Mat-like	Brown	55.9	47.0	2.5	0.05	-	Rounded	-

[1] The size measurements presented in this table are averages in µm. Filament diameters are the average of three filament measurements and cell measurements the average of ten adjacent cells of three filaments.
[2] Size of cell wall constriction: (++) distinct (5-10%), (+) slightly present (1-5%), (-) absent (0-1%) in percentage of cell width.
[3] Measurements performed on strains in culture.
[4] The strain 3L has been reported as *L. majuscula* in the literature. However, the cells were distinctly constricted at the cross-walls and the cells relatively wide. Thus, the description fits better with that of *L. sordida*.

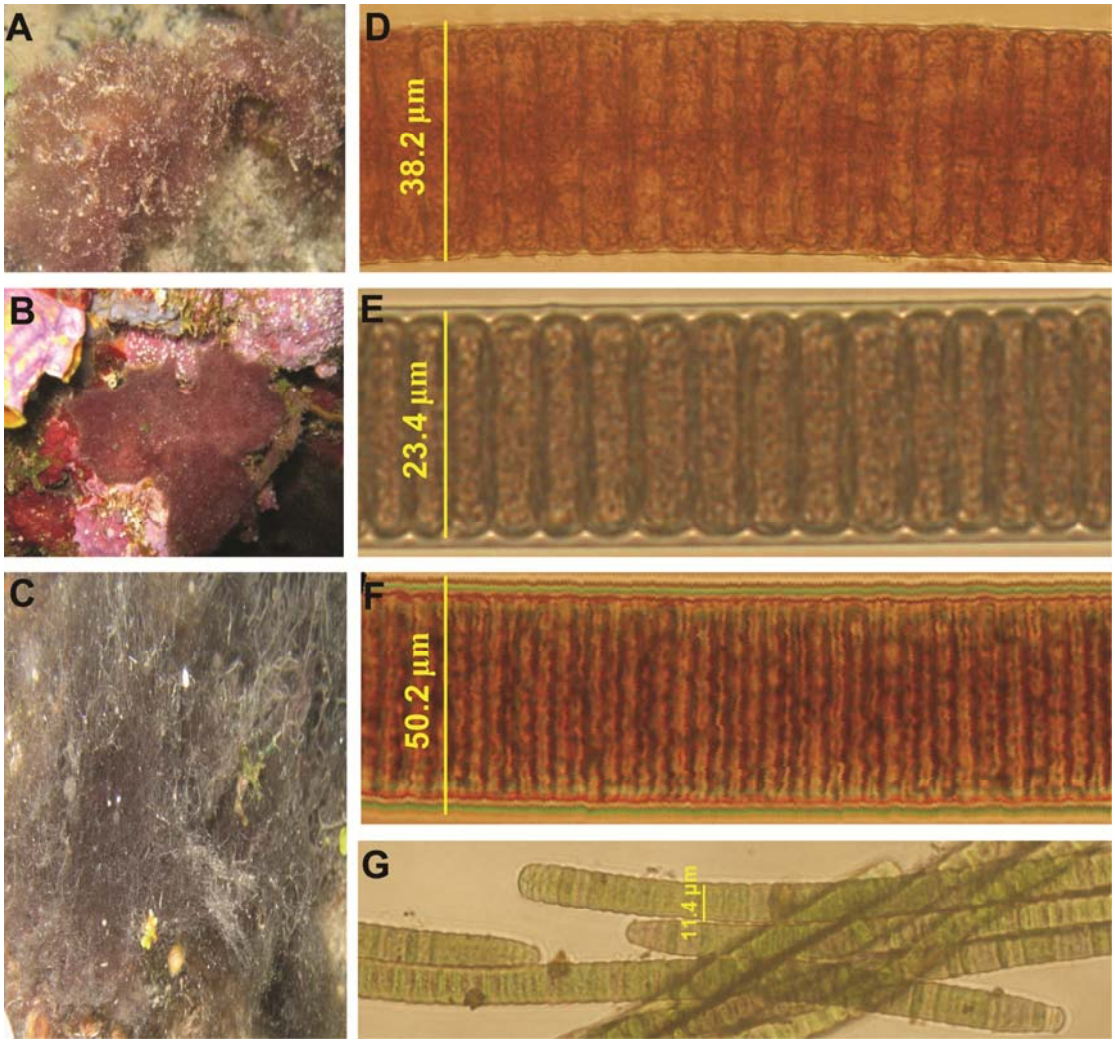


Figure III.1: Underwater pictures of (**A**) *L. sordida* PNG6-52, (**B**) *L. bouillonii* PNG5-198 firmly attached to surrounding coral, and (**C**) *L. majuscula* PNG6-221 growing in an assemblage with *Oscillatoria nigro-viridis* PNG6-24 (FJ157181). Photomicrographs (x60) of (**D**) *L. sordida* PNG6-52 (filament width: 38.2 µm), (**E**) *L. bouillonii* PNG5-198 (filament width: 23.4 µm), (**F**) *L. majuscula* PNG6-221 (filament width: 50.2 µm), and (**G**) *Oscillatoria nigro-viridis* PNG6-24 (filament width: 11.4 µm). The microscopic pictures are adjusted to fit the frame (see scale bars) and the scales for the underwater pictures are approximated as shown.

II.3.2 *Lyngbya*-associated Microorganisms

Partial 16S rRNA gene sequences were PCR-amplified from 36 of the *Lyngbya* field specimens using cyanobacterial-specific primers and clone libraries were constructed. The number of clones sequenced ranged between 1 and 10 clones and libraries from individual *Lyngbya* specimens revealed considerable 16S rRNA gene heterogeneity. However, combined BLAST-analysis and phylogenetic tree construction methods revealed that 42 (44%) of a total of 95 gene sequences obtained from the 36 *Lyngbya* specimens belonged to associated microorganisms other than *Lyngbya*. Particularly common were gene sequences belonging to various Gram-negative heterotrophic bacteria such as α-Proteobacteria (*Caulobacter*, *Thalassobius*, and *Rhodospirillaceae*), β-Proteobacteria (*Achromobacter*), γ-proteobacteria (*Pseudomonas* and *Oceanospirillum*), and Bacterioidetes (*Flavobacterium* and *Flexibacter*).

In addition to these heterotrophic microbial taxa, smaller filamentous cyanobacteria were often present in association with the *Lyngbya* filaments. These associated cyanobacteria were identified both microscopically and by gene sequence analysis as belonging to the genera *Oscillatoria, Spirulina, Geitlerinema*, or other Oscillatoriales. Gene sequences for *O. nigro-viridis* were particularly abundant in several of the *Lyngbya* colonies (*e.g. O. nigro-viridis* 3LOSC (acc. No. EU24875), *O. nigro-viridis* PNG6-24 (acc. No. FJ157181), and *O. nigro-viridis* PNG5-223 (acc. No. FJ157182) were isolated from *L. sordida* 3L (acc. No. FJ151527), *L. majuscula* PNG6-221 (acc. No. FJ356669), and *L. majuscula* PNG5-223 (acc. No. U492878), respectively; Figure III.2).

II.3.3 Gene Sequence Anomalies

All PCR-amplified *Lyngbya* gene sequences were carefully screened for anomalies, such as chimeric sequences or sequencing errors. These analyses revealed that approximately 6% of the PCR-amplified gene sequences contained some form of anomaly. In this study, all gene sequences containing anomalies were discarded and new clone libraries established from the same sample to assure correct genetic information.

The 72 *Lyngbya* 16S rRNA genes present in GenBank were also screened for sequence anomalies to investigate the frequency of incorrect gene information for this genus present in public databases. Thirty nine (54%) of the GenBank obtained gene sequences were excluded from the examination due to short length (<600 bp). One of the 33 sequence (3%) clearly contained an anomaly (most likely a sequence error but possibly a chimeric sequence) and an additional three sequences (9%) were suspected to contain chimeric anomalies (*i.e*. the percent sequence difference between the analyzed sequence with a query sequence were >9 and ≤11 using Pintail software). It should be noted that 66% of the analyzed gene sequences in GenBank were only slightly greater than the minimally accepted length of 600 bp, and thus, the short length of these sequences may affect the reliability of these analyses.

II.3.4 Phylogenetic Resolution of the 16S rRNA Gene

The 16S rRNA gene sequences determined in this study for 36 tropical, marine *Lyngbya* specimens formed a phylogenetically tight group with more than 98.6%

sequence identity (this tropical, marine *Lyngbya* lineage is compressed in Figure II.2 and unfolded in Figure II.3).

A phylogram was constructed to provide a phylogenetic perspective for these *Lyngbya* specimens in relation to the 72 *Lyngbya* strains available in GenBank as well as twelve major genera within the order Oscillatoriales (Figure II.2). Five different phylogenetic algorithms, ME, NJ, ML, MP, and Bayesian inference were used and compared to infer the evolutionary history and relationships among these cyanobacterial groups. Overall tree topology and bootstrap values were well conserved in all five analyses (annexed Table 2 in Figure II.2). Atleast five different lineages of *Lyngbya* could be observed in the phylogenetic tree in Figure III.2 and, thus, the genus *Lyngbya* clearly represent a polyphyletic group.

Among the 72 16S rRNA sequences from GenBank 13 (16%) were phylogenetically related to the tropical, marine *Lyngbya* lineage, 30 (38%) to the halophilic/brackish/freshwater lineage, 27 (34%) to epiphytic cyanobacteria common to *Lyngbya*, and 2 (3%) to associated heterotrophic or eukaryotic microorganisms.

Table 2. Percent bootstrap support

Node	Lineage	ME	NJ	MP	B	ML
A	*Trichodesmium*	90	89	86	100	64
B	*Oscillatoria*	93	89	86	97	53
C	*Microcoleus*	92	91	71	86	46
D	*Planktothrix*	99	99	98	100	80
E	*Arthrospira*	99	98	100	100	80
F	*Lyngbya*	80	71	62	100	54
G	*Symploca*	87	86	64	100	60
H	*Lyngbya*	99	99	100	99	99
I	*Geitlerinema*	99	99	53	100	64
J	*Limnothrix*	99	99	100	100	80
K	*Leptolyngbya*	99	99	99	100	78
	Mean:	94.2 (±6.4)	92.6 (±8.7)	83.5 (±17.9)	98.4 (±4.2)	68.9 (±15.8)

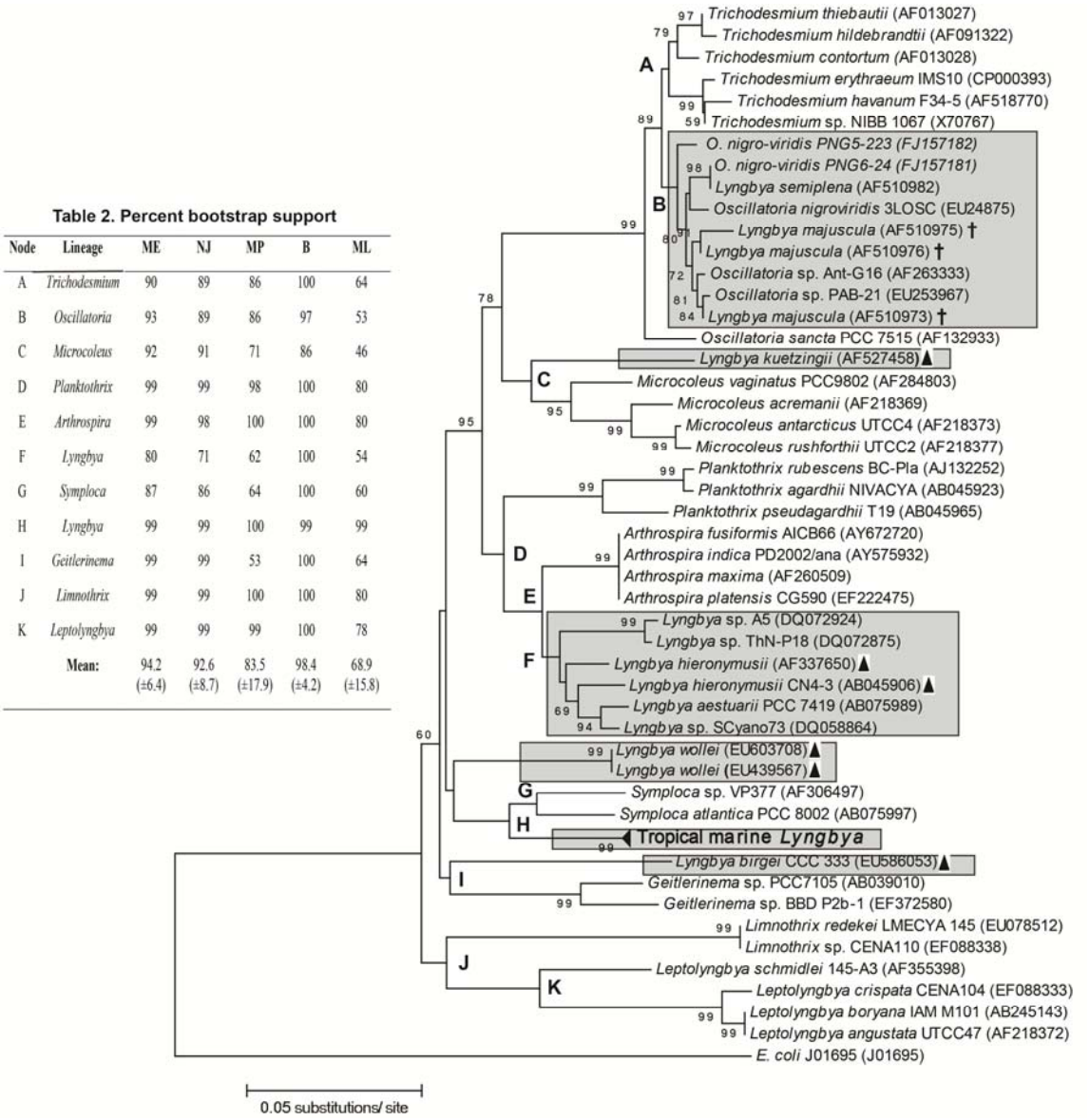


Figure III.2: Polyphyletic evolution of the cyanobacterial genus *Lyngbya* based on 16S rRNA gene sequences (1,370 bp) with other genera of the order Oscillatoriales (Subsection III). Representative type-strains are shown by (^{T}). Six different *Lyngbya* lineages are present (enclosed in boxes). The true *Lyngbya* lineage (including type-strain PCC 7419^{T}) is shown with an arrow. Gene sequences annotated as *Lyngbya* from GenBank clading in the "*Oscillatoria* lineage" and proposed to belong to *Oscillatoria* are denoted with a dagger (†). Note that the genus *Trichodesmium* nests within the genus *Oscillatoria* and needs to be revised (Appendix 13). The marine *Lyngbya* clade representing new *Lyngbya* sequences from this study is compressed and highlighted in bold. Fresh-water *Lyngbya* are designated with a triangle (▲). Statistical supports for the major nodes are shown at the annexed Table 2.

The tropical, marine *Lyngbya* lineage was analyzed in greater detail using the ME/Bayesian-methods with 865 nucleotide positions (Figure III.3). The sequence divergence within the marine *Lyngbya* clade was low, ranging from 0-1.4% with an average of 0.4% ± 0.2. The phylogenetic diversification exceeded the morphological diversification by a ratio of 3:29 (*i.e.* the three morphological species resulted in 29 different 16S rRNA gene sequences). In seven of the 29 nodes there were multiple *Lyngbya* strains present (Figure III.3, node I-VII). Interestingly, morphologically different species of *Lyngbya* occurred in five of the seven nodes which contained multiple identical 16S rRNA gene sequences (marked with asterisks in node I, III, IV, V, and VII in Figure III.3).

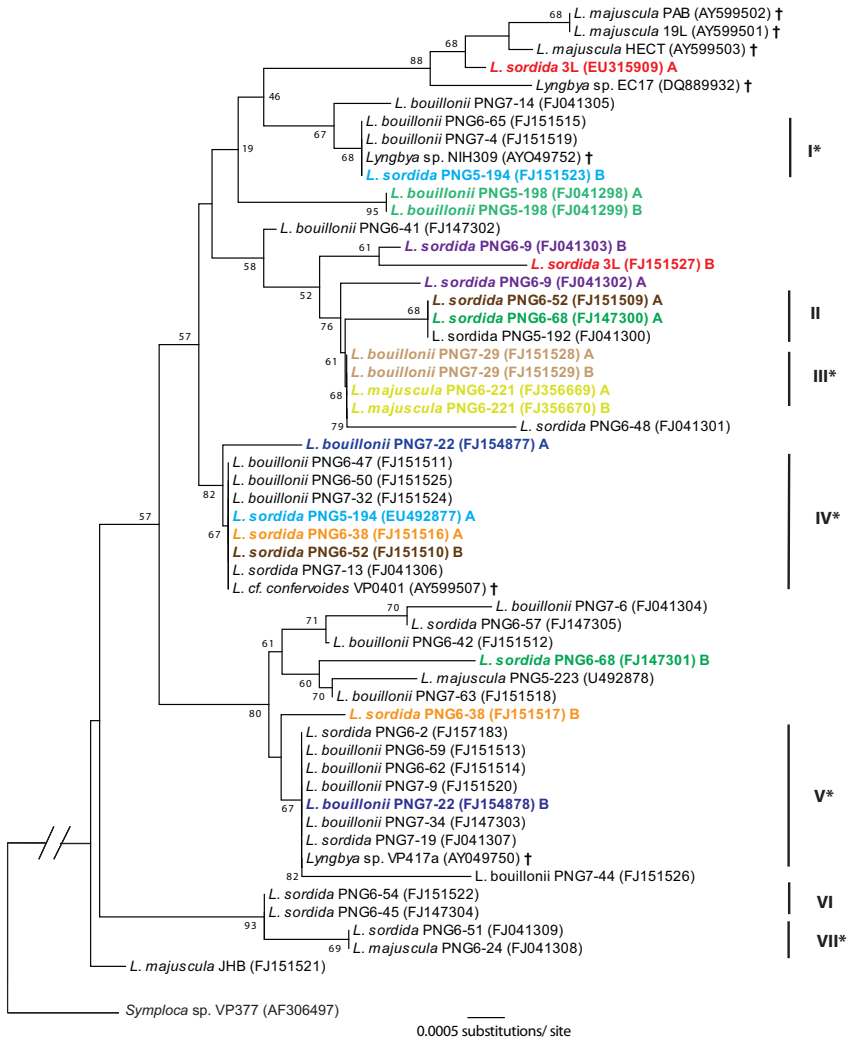


Figure III.3: Phylogenetic tree of the "marine *Lyngbya* lineage" with the outgroup *Symploca* sp. VP377 pruned for clarity. The phylogram includes all *Lyngbya* specimens analyzed in this study together with *Lyngbya* 16S rRNA gene sequences obtained from GenBank [the following reference strains are designated with a dagger; *L. majuscula* 19L (AY599501), *L. majuscula* PAB (AY599502), *L. majuscula* HECT (AY599503), *Lyngbya* sp. VP417a (AYO49750), *L. confervoides* VP0401 (AY599507), *Lyngbya* sp. NIH309 (AYO49752) and *Lyngbya* sp. EC17 (DQ889932)]. Variable intra-genomic gene copies designated as A or B are bolded and identified by the same color. Note that certain intra-genomic gene copies (e.g. *L. bouillonii* PNG5-198 A and B) appear to be identical in the tree due to truncation of gene sequences and exclusion of variable regions. Terminal nodes with multiple identical 16S rRNA gene sequences are designated as I-VII and those nodes with identical 16S rRNA gene sequences for morphologically different *Lyngbya* species are designated with an asterisk. Bootstrap-support using the ME method is shown at the main branches.

II.3.5 Intra-colonial 16S rRNA Gene Heterogeneity in Environmental Specimens

Multiple differing copies of the *Lyngbya* 16S rRNA genes were observed in ten of the environmental *Lyngbya* colonies (Table III.2; Figure III.3). In this study, when multiple clones from a single environmental colony were analyzed, 90% revealed intra-colonial sequence heterogeneity in the 16S rRNA gene. The number of variations found between clones from a colony was limited to a maximum of two differing types, a number that remained stable even after the analysis of up to 30 different clones from a single colony. The sequence divergence of intra-colonial 16S rRNA gene variations varied between 0.1-1.1% with an average of 0.4% ± 0.3 (Table III.2). The presence of multiple phylogenetic strains within a single-colony was investigated by collecting and creating clone libraries from three different sections of colonies of *L. bouillonii* PNG7-22 and *L. bouillonii* PNG7-29. Both *L. bouillonii* PNG7-22 and *L. bouillonii* PNG7-29 revealed gene copies A and B in the gene libraries derived from the three different sections of each colony.

Table III.2: *Lyngbya* Colonies with Multiple Copies of the 16S rRNA Gene of Variable Sequence.

Strain	16S rRNA gene copy A	16S rRNA gene copy B	Percent sequence divergence between copy A and copy B [1]
L. bouillonii PNG5-198	FJ041298	FJ041299	0.2
L. bouillonii PNG7-22	FJ154877	FJ154878	0.4
L. bouillonii PNG7-29	FJ151528	FJ151529	0.3
L. sordida PNG5-194	EU492877	FJ151523	0.2
L. sordida PNG6-9	FJ041302	FJ041303	0.3
L. sordida PNG6-38	FJ151516	FJ151517	0.2
L. sordida PNG6-52	FJ151509	FJ151510	0.4
L. sordida PNG6-68	FJ147300	FJ147301	0.5
L. majuscula PNG6-221	FJ356669	FJ356670	0.1
L. sordida 3L	EU315909	FJ151527	1.1

[1] The percent sequence divergent between the gene copies from sequencing of 1378 bp.

II.3.6 16S rRNA gene and rpoC1 Gene Heterogeneity in Clonal *Lyngbya* Strains

Single-filaments were isolated from the following four *Lyngbya* specimens (*L. majuscula* JHB, *L. majuscula* PNG6-221, *L. sordida* 3L, and *L. bouillonii* PNG5-198) and grown into clonal cultures. 16S rRNA (1378 bp) and rpoC1 (877 bp) gene-libraries were constructed in parallel for each strain and multiple clones (>17) were sequenced from each library to thoroughly investigate the intra-strain genetic diversity. However, despite the establishment of monoclonal cultures, none of the strains were axenic. All four strains revealed morphological evidence, and three of them genetic evidence, of associated heterotrophic bacteria. DAPI-staining was used to visualize the associated heterotrophic bacteria which were entirely localized to the sheaths surrounding the filaments.

L. sordida strain 3L displayed two different 16S rRNA gene variations, rrn A and rrn B. These two gene variations varied in 14 (1.1%) nucleotide positions (1378 bp) of the sequenced genes. Secondary structure analysis revealed that the 14 nucleotide substitutions were all located in loop-regions of the 16S rRNA gene and that the overall RNA structure was conserved in both gene variations. *L. bouillonii* PNG5-198 also showed two different 16S rRNA gene variations, rrn A and rrn B. The gene sequences in these clones varied in only two nucleotides, and thus, had a sequence divergence of 0.3%. One of the substitutions was located in a hairpin-loop and the other (C↔T) in a stem-region which resulted in a U-G bond. Thus, both nucleotide substitutions conserved the overall secondary structure of the gene product.

The clone libraries of *L. majuscula* strain PNG6-221 also revealed two different 16S rRNA gene variations, rrn A and rrn B. The two gene variations A and B differed in

a single nucleotide substitution which was located in a stem-region and resulted in a U-G bond in the final RNA product. In contrast to the above three cases, *L. majuscula* strain JHB showed only one 16S rRNA gene sequence among 17 different sequenced clones from its 16S rRNA gene library.

By contrast, all four *Lyngbya* strains (JHB, PNG6-221, 3L, and PNG5-198) each displayed only one copy of the rpoC1 gene (acc. No.: FJ214725, FJ217387, FJ231748, and FJ231749, respectively) from the sequencing of ten clones from each gene library, thus confirming their clonal nature. Indeed, a preliminary survey of sequenced cyanobacterial genomes found only a single copy of the rpoC1 gene in each (See Thesis Chapter IV). The average rpoC1 gene sequence divergence between the four rpoC1 genes of the four *Lyngbya* strains ranged from 2.2-3.3% with a mean of 2.7 ± 0.6%. A preliminary (ME) phylogenetic inference of the four clonal strains based on these few rpoC1 gene sequences delineated the different *Lyngbya* species.

II.3.7 16S rRNA Gene Heterogeneity in Single-cell MDA-genomes

In order to evaluate genetic variability of the 16S rRNA gene at the genome level in our *Lyngbya* isolates, single-filaments constituting between 12-20 neighboring cells of the *Lyngbya* strains, 3L, PNG5-198, and PNG6-221, were individually isolated and selected for genomic amplification by MDA. The 16S rRNA genes were then PCR-amplified from these MDA's and 16S rRNA gene libraries constructed for each MDA. MDA-amplification was repeated on isolated single-cells of each *Lyngbya* strain to evaluate whether genetic variations occurred on a cellular level or between neighboring cells.

MDA-DNA from a single-filament of *L. sordida* 3L displayed the two different gene copies A and B that had previously been observed in the cultures. Furthermore, the genome of a single-cell of *L. sordida* 3L was MDA-amplified and the 16S rRNA genes were PCR-amplified. Both gene copy A and B were observed in two individual single-cell MDA's. Single-filament MDA-DNA of *L. bouillonii* PNG5-198 revealed both 16S rRNA gene copies A and B previously observed in the clonal culture. Three separate MDA-amplified genomes of single-cells from *L. bouillonii* PNG5-198 confirmed the presence of both 16S rRNA gene copies A and B in the genome. MDA-DNA obtained from a single-filament of *L. majuscula* PNG6-221, as well as one from three single-cell MDA's from this same strain, revealed the presence of both 16S rRNA gene copies A and B. In summary, the sequence divergence between multiple 16S rRNA genes in *Lyngbya* single-cell genomes ranged between 0-1.1% with a mean of 0.4% ± 0.4.

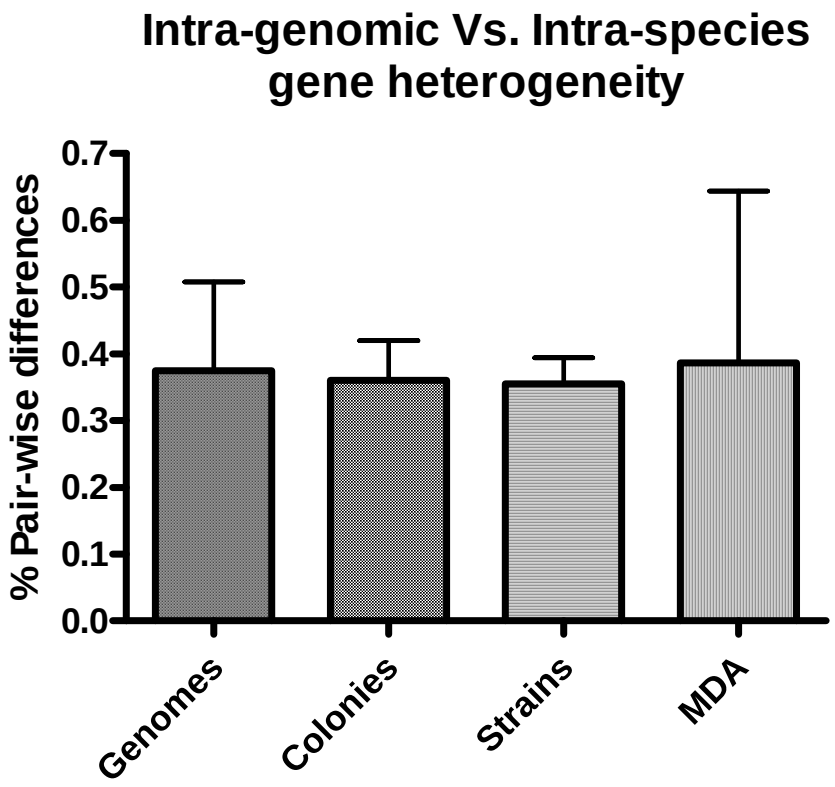


Figure III.4: Average pair-wise sequence divergence between 16S rRNA genes: (**A**) between different ribosomal operons within the same sequenced cyanobacterial genomes, (**B**) within individual *Lyngbya* colonies, *i.e.* environmental specimens (Table II.2), (**C**) between different strains and species of *Lyngbya* and (**D**) within single-cell multiple displacements amplified (MDA) DNA.

II.4 Discussion

The SSU rRNA gene is generally a highly useful tool for the identification and cataloging of cyanobacteria or other microorganisms (Wilmotte & Herdman, 2001). The 16S rRNA gene was utilized in this study to investigate the phylogeny of the biologically and biomedically important cyanobacterial genus *Lyngbya* and to examine the utility of this gene to discriminate between different marine species of *Lyngbya*. Several factors were initially pinpointed as potential challenges in this molecular-phylogenetic approach. Firstly, environmental colonies of macroscopic *Lyngbya* were found to exist in close association with a diverse and complex flora of microorganisms. Many of the most common heterotrophic Gram-negative bacterial groups in *Lyngbya* colonies also form the predominant components in surrounding seawater, as well as in filter-feeding marine organisms such as sponges and tunicates, and some may be present as symbionts (Taylor *et al*., 2007). However, other groups of bacteria appear to form specific associations with *Lyngbya*. For example, smaller filamentous cyanobacteria such as *Oscillatoria nigro-viridis* grow in association with *Lyngbya* and are also rich in bioactive secondary metabolites (Simmons *et al*., 2008). In this regard, it is intriguing to speculate that species of these two genera might form specific relationships that broaden their spectrum of defensive chemicals.

In this study, we showed that a large number of 16S rRNA gene sequences retrieved from environmental *Lyngbya* specimens, as well as sequences present in public databases, are genetically closely related to *Oscillatoria* or other common *Lyngbya*-

associated cyanobacteria. Moreover, we found from our sequencing study that in every case where a 16S rRNA gene of a *Lyngbya*-associated cyanobacterium was obtained from a *Lyngbya* colony (*e.g.* an *Oscillatoria* 16S rRNA sequence), a *Lyngbya* 16S rRNA gene was obtained as well. Thus, it is likely that many of these publicly accessible *Lyngbya* 16S rRNA gene sequences are actually gene sequences derived from *Oscillatoria* spp. which coexist with *Lyngbya*. Alternatively, some species of *Oscillatoria* morphologically resemble *Lyngbya*, and thus may have lead to misidentifications between these two genera. In this regard, some studies have found sequence divergences greater than 8% and sometimes as high as 14% in morphologically closely related *Lyngbya* species, such as *L. majuscula* and *L. bouillonii* (Paul *et al*., 2005, Thacker *et al*., 2004, Sharp *et al*., 2009). However, these gene sequences are almost identical to those of *Oscillatoria* spp. (*i.e*. the percent sequence divergence range between 0 and 1.7). Thus, the potential confusion of *Lyngbya* 16S rRNA genes with those of epiphytic cyanobacteria or misidentification provides likely explanations for the high sequence divergence reported between closely related *Lyngbya* species. Moreover, approximately three percent of the 16S rRNA genes submitted to GenBank as *Lyngbya* display clear phylogenetic relationships with heterotrophic bacterial or eukaryotic groups often found in association with *Lyngbya* species, further illustrating the confusion that is produced from these incorrect annotations in the public databases.

The complex gene pools of closely related microbial taxa, such as those of *Lyngbya* colonies with associated cyanobacteria and heterotrophic bacteria, adds to the possibility of PCR-errors resulting in a relatively high frequency of chimeric sequences

(Ashelford *et al*., 2005). It has been estimated that approximately five percent of gene sequences submitted to GenBank contain anomalies (DeSantis *et al*., 2006). This frequency is comparable to the six percent of anomalous sequences observed in the *Lyngbya* specimens of this study that are attributed to either PCR-artifacts or sequence errors. The overall result of the above factors, in combination with the submission of incorrectly analyzed gene sequences, has been to produce a large number of mischaracterized genes in gene databases and the literature, and a resulting confusion for the phylogenetics of this microbial group.

The 16S rRNA genes retrieved from public databases revealed *Lyngbya* as a polyphyletic genus with two distinct *Lyngbya* clades in addition to the aforementioned *Oscillatoria* sequences. Members of these different *Lyngbya* lineages are morphologically similar but are ecologically distinct. The clade composed of tropical, marine *Lyngbya* species such as *L. majuscula, L. sordida, L. confervoides,* and *L. bouillonii*, while the temperate clade is comprised by marine, brackish, or fresh-water species such as *L. aestuarii* and *L. hieronymusii*. The phylogenetic distance between the tropical *Lyngbya* species and the temperate *Lyngbya* species suggests that these two clades may represent different genera altogether. Members of this latter clade are phylogenetically closely related to the genus *Arthrospira*, which may be indicative that they derive from a recent common ancestor. Several other fresh-water *Lyngbya* species (e.g. *L. birgei* CCC 333, *L. wollei* Carmichael/Alabama, *L kuetzingii* Schmidle) appear not to clade with either of these two lineages, but rather appear as “loner” sequences in the phylogenetic analysis of the *Oscillatoriales*. This may be the result of these latter

collections coming from divergent geographically and presumably genetically isolated locations. Therefore, our analysis indicates that a revision of the genus *Lyngbya* is required and that phylogenetically unique lineages may have evolved in different environments.

The presence of multiple copies of the 16S rRNA gene has been rarely reported in cyanobacteria. This may be because most molecular-taxonomic identifications are typically limited in the number of clones selected for sequencing based upon the assumption that each specimen is represented by a single 16S rRNA gene sequence. From the analysis presented in this paper, reports of single gene sequences in filamentous cyanobacteria likely represent limitations in clone sampling rather than the absence of multiple gene variations. Furthermore, our study shows that 16S rRNA gene heterogeneity observed within colonies was evenly distributed in different sections of cyanobacterial colonies. This strongly suggests that these gene variations are explained by multiple copies of variable sequence within the same cyanobacterial genome (i.e. intra-genomic gene heterogeneity), rather than colonies being composed of multiple strains of different phylogeny.

In the current study, 16S rRNA gene heterogeneity was systematically investigated from environmental colonies to single-cell genomes. Most speculations concerning intra-genomic gene heterogeneity have previously been based on *in silico* investigations of sequenced bacterial genomes and have included a very limited number of cyanobacterial genomes (Coenye *et a*l., 2003; Acinas *et al*., 2004; Case *et al*., 2007). To our knowledge, this is the first study to investigate 16S rRNA gene heterogeneity on a

cellular level through the application of MDA as a method of amplifying the genomes of single-cells. Our hypothesis that the gene variations observed in individual environmental colonies were intra-genomic was supported by the fact that the number of gene copies remained stable upon additional sequencing of clone libraries from environmental specimens, from clonal cultures, and from single-cell MDA's.

In *Lyngbya* single-cell MDA genomes, the intra-genomic sequence divergence between different 16S rRNA gene copies ranged between 0% and 1.1% with an average divergence of 0.4%. Most importantly to phylogenetics within this genus, this gene sequence divergence was of a similar magnitude as between different morphological species of *Lyngbya*. Thus, it appears that in *Lyngbya* the average number of nucleotide substitutions between 16S rRNA genes within a genome is roughly equal to the number of nucleotide substitutions between morphological species. Using bioinformatics approaches, similar correlations between intra-genomic and inter-species heterogeneity have also been observed in other Gram-negative bacteria such as α-Proteobacteria (Coenye *et al*., 2003).

Genotypic diversity exceeded phenotypic diversity of the *Lyngbya* spp. by approximately 10-fold. Initially, it was thought that this 10-fold greater phylogenetic resolution would provide a sub-grouping system within each morphological species of *Lyngbya*. However, the phylogenetic species obtained through the 16S rRNA gene comparison did not correlate with the well differentiated morphological species of *Lyngbya*. While it is possible that these phylogenetic species could represent true taxonomic groupings, thus challenging the current phycological taxonomic system of

Lyngbya differentiation at the species level, we argue against the use of a phylogenetic system based on 16S rRNA gene sequences at the species level. This is based on the fact that several strains contained multiple intra-genomic 16S rRNA gene sequence variants that grouped with different morphological species, thus clearly showing the unreliability of 16S rRNA gene sequences for species level identification of these filamentous cyanobacteria.

In contrast to the 16S rRNA gene which occurred in multiple variable copies in *Lyngbya* spp., the rpoC1 gene could be found only as single copies. This protein-encoding gene was used to confirm that the single-filament derived cultures were truly clonal. Because the rpoC1 gene is present as a single-copy, it may be a good candidate for future explorations of fine-scale species delineation in filamentous cyanobacteria. In support of this assumption, a preliminary rpoC1 gene phylogenetic analysis appeared to delineate between the three *Lyngbya* species, *L. bouillonii, L, sordida* and *L. majuscula*. Furthermore, the intra-strain rpoC1 gene sequence variations analyzed herein were approximately five times larger than those of ribosomal genes. This higher genetic variation may be utilized to increase phylogenetic resolution and to delineate *Lyngbya* at the species level as has been implemented previously in the unicellular genera *Prochlorococcus* and *Synechococcus* (Toledo *et al*., 1997; Ma *et al*., 2004).

II.5 Acknowledgments

We thank L. Gerwick with helpful discussions on diverse molecular biology issues, and P. Jensen and A. Prieto-Davo for phylogenetic discussions. In addition, we thank B. Palenik for insights of cyanobacterial phylogeny and the design of the degenerate primers for the rpoC1 gene. We also gratefully acknowledge the governments of Curaçao, Jamaica, and Papua New Guinea for permission to make these cyanobacterial collections. Financial support for this study came from the Swedish-American Foundation (NE) and the NIH (NS053398 and CA100851).

The text of III, in full, is published material as it appears in: Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601. The dissertation author was the primary author and directed and supervised the research, which forms the basis for this chapter.

II.6 Appendix - Supporting information for chapter III

Table of Contents of Appendix

Table III.3: Collection data for *Lyngbya* specimens.

Table III.3: Collection Data for *Lyngbya* Specimens.

Species (strain)	PNG region	Location	GPS Coordinates		Collection Date[1]	Depth (m)
			latitude	longitude		
L. bouillonii PNG5-198	New Ireland	Pigeon Island	4° 16.063'S	152° 20.266'E	05.05.18	10
L. bouillonii PNG6-1	Milne Bay	Kwato Island	10° 36.990'S	150° 37.396'E	06.04.20	10
L. bouillonii PNG6-41	Milne Bay	Deacon's reef	10° 15.612'S	150° 44.878'E	06.04.24	4
L. bouillonii PNG6-42	Milne Bay	Deacon's reef	10° 15.612'S	150° 44.878'E	06.04.24	16
L. bouillonii PNG6-47	Milne Bay	Bentley Bay	10° 13.051'S	150° 36.110'E	06.04.25	10
L. bouillonii PNG6-50	Milne Bay	Bentley Bay	10° 13.051'S	150° 36.110'E	06.04.25	<3
L. bouillonii PNG6-59	Milne Bay	Sewa Bay	09° 58.998'S	150° 57.359'E	06.04.27	17
L. bouillonii PNG6-62	Milne Bay	Wadu Point	10° 06.277'S	150° 57.673'E	06.04.27	13
L. bouillonii PNG6-65	Milne Bay	Gallows Reef	10° 16.579'S	151° 11.492'E	06.04.28	7
L. bouillonii PNG7-4	Milne Bay	Sponge Heaven	10° 15.836'S	150° 40.193'E	07.05.02	17
L. bouillonii PNG7-6	Milne Bay	Sponge Heaven	10° 15.836'S	150° 40.193'E	07.05.02	17
L. bouillonii PNG7-14	Milne Bay	Triple Crown	9° 32.491'S	150° 16.715'E	07.05.04	N/A
L. bouillonii PNG7-19	Milne Bay	Shark Hill	8° 32.364'S	148° 98.104'E	07.05.05	10
L. bouillonii PNG7-22	Milne Bay	Hanging Gardens	5° 25.566'S	05° 06.754'E	07.07.12	17
L. bouillonii PNG7-29	Milne Bay	Restof Island	5° 25.575'S	150° 06.234'E	07.07.13	N/A
L. bouillonii PNG7-32	New England	Restof Island	5° 25.575'S	150° 06.234'E	07.07.13	12

Table III.3: Continued

Species (strain)	PNG region	Location	GPS Coordinates latitude	GPS Coordinates longitude	Collection Date[1]	Depth (m)	Species (strain)
L. bouillonii PNG7-44	New England	N/A	5° 26.192'S	150° 40.813'E	07.07.14	13	
L. bouillonii PNG7-46	New England	N/A	5° 26.192'S	150° 40.813'E	07.07.14	<2	
L. bouillonii PNG7-51	New England	N/A	5° 26.192'S	150° 40.813'E	07.07.14	24	
L. bouillonii PNG7-63	New England	N/A	5° 26.926'S	150° 47.897'E	07.07.16	24	
L. bouillonii PNG7-79	Milne Bay	N/A	5° 19.588'S	150° 18.034'E	07.07.18	22	
L. sordida PNG5-192	New Ireland	Pigeon Island	4° 16.063'S	152° 20.266'E	05.05.18	10	
L. sordida PNG5-194	New Ireland	Pigeon Island	4° 16.063S	152° 20.266'E	05.05.18	10	
L. sordida PNG6-2	Milne Bay	Kwato Island	10° 36.990'S	150° 37.396'E	06.04.20	15	
L. sordida PNG6-3	Milne Bay	Kwato Island	10° 36.990'S	150° 37.396'E	06.04.20	2	
L. sordida PNG6-9	Milne Bay	Miska's Rock	10° 40.294'S	150° 41.226'E	06.04.20	12	
L. sordida PNG6-34	Milne Bay	Sponge Heaven	10° 15.836'S	150° 40.193'E	06.04.23	1	
L. sordida PNG6-38	Milne Bay	Deacon's reef	10° 15.612'S	150° 44.878'E	06.04.23	13	
L. sordida PNG6-43	Milne Bay	Deacon's reef	10° 15.612'S	150° 44.878'E	06.04.24	2	
L. sordida PNG6-45	Milne Bay	Deacon's reef	10° 15.612'S	150° 44.878'E	06.04.24	<1	
L. sordida PNG6-46	Milne Bay	Bentley Bay	10° 13.051'S	150° 36.110'E	06.04.25	1	
L. sordida PNG6-48	Milne Bay	Bentley Bay	10° 13.051'S	150° 36.110'E	06.04.25	7	

Table III.3: Continued

Species (strain)	PNG region	Location	GPS Coordinates latitude	GPS Coordinates longitude	Collection Date[1]	Depth (m)	Species (strain)
L. sordida PNG6-52	Milne Bay	Bentley Bay	10° 13.051'S	150° 36.110'E	06.04.25	22	
L. sordida PNG6-54	Milne Bay	Kapa point	10° 14.055'S	150° 49.873'E	06.04.26	<1	
L sordida PNG6-57	Milne Bay	Kapa point	10° 14.055'S	150° 49.873'E	06.04.26	3	
L. sordida PNG6-63	Milne Bay	Sewa Bay	09° 58.998'S	150° 57.359'E	06.04.27	10	
L. sordida PNG6-68	Milne Bay	Gallows Reef	10° 16.579'S	151° 11.492'E	06.04.28	<3	
L. sordida PNG6-69	Milne Bay	Gallows Reef	10° 16.579'S	151° 11.492'E	06.04.28	N/A	
L. sordida PNG7-2	Milne Bay	Gonu-bala-bala	10° 41.115'S	150° 40.808'E	07.05.01	7	
L. sordida PNG7-13	Milne Bay	Wahoo Wall	10° 15.177'S	150° 47.034'E	07.05.03	17	
L. sordida PNG7-19	Milne Bay	Shark Hill	8° 32.364'S	148° 48.104'E	07.05.05	10	
Lyngbya sp. PNG7-60	New Ireland	N/A	5° 26.926'S	150° 47.897'E	07.07.16	N/A	
L. majuscula PNG5-1913	New Ireland	Credner Islands	4° 14.105'S	152° 25.605'E	05.05.19	<2	
L. majuscula PNG6-24	New Ireland	Dudawali Bay	10° 17.274'S	151° 00.390'E	06.04.22	5	
L. majuscula PNG5-223	Milne Bay	New Ireland coast	3° 40.620'S	152° 20.406'E	05.05.22	9	
L. sordida 3L	Curaçao	CARMABI	N/A	N/A	91.12.12	>2	
L. majuscula JHB	Jamaica	Hector Bay	N/A	N/A	96.08.22	2	

[1] Dates indicated by year/month/day, (N/A) Data not available.

References

Abed, R. M., Golubic, S., Garcia-Pichel, F., Camoin, G. F. Sprachta, S. (2003). Characterization of microbialite-forming cyanobacteria in a tropical lagoon: Tikehau Atoll, Tuamotu, French Polynesia. *J Phycol* 39: 862-873.

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* 186: 2629–2635.

Agardh, C. A. (1898) In *Species, genera et ordines algarum*.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., Weightman, A. (2005). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Env Microbiol* 71: 7724-7736.

Cannone, J. J., Subramanin, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pnde, N., Schang, Z., Yu, N., Gutell R. R. (2002). The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*. 3: 1471-2105.

Case, R. J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W. F. & Kjelleberg S. (2007). Use of 16S rRNA and rpoB Genes as molecular markers for microbial ecology studies. *Appl Env Mic Biol* 73: 278-288.

Castenholz, R. W. (2001). Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 473-553. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Clarridge, J. E. III. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17: 840-862.

Coenye, T., Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* 228: 45-49.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Env Microbiol* 72: 5069-5072.

Desikachary, T. V. (1959). In *Cyanophyta. Indian council of cultural research*, New Delhi, India.

Geitler, L. (1932). *Cyanophyceae*. In Rabenhorst, L. [Ed.]. *Kryptogamen-Flora von Deutschland, Österreich und der Schweiz, vol. XIV*. *Leipzig: Akademische Verlag* (1985 reprint: Königstein: Koeltz Scientific Books) 1027-1068.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Gomont, M. M. (1892). In *Monographie des Oscillariées (Nostocacées homocystées). Ann Sci Nat Bot Ser* 16:91-264.

Gutiérrez, M., Suyama, T. L., Engene, N., Wingerd, J. S., Matainaho, T., Gerwick, W. H. (2008). Apratoxin D, a potent cytotoxic cyclodepsipeptide from Papua New Guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*. *J Nat Prod* 71: 1099-1103.

Hoffmann, L., Demoulin, V. (1991). Marine Cyanophyceae of Papua New Guinea. II. *Lyngbya bouillonii* sp. nov., a remarkable tropical reef-inhabiting blue-green alga. *Belg J Bot* 124: 82-88.

Hoffmann, L. (1994). Marine Cyanophyceae of Papua New Guinea. VI. The genus *Lyngbya*. S.L. *Belg J Bot* 127: 79-86.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae* 21: 349-375.

Komárek, J., K. Anagnostidis. (2005). In *Süsswasserflora von Mitteleuropa*. 19/2. Elsevier/Spektrum, Heidelberg, Germany, pp. 1-759.

Lasken, R. S. (2007). Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr Opin Microbiol* 10: 510-516.

Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J. (2002). New apratoxins of marine cyanobacterial origin from Guam and Palau. *Bioorg Med Chem* 10: 1973-1978.

Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J. (2002). Structurally diverse new alkaloids from Palauan collections of the apratoxin-producing marine cyanobacterium *Lyngbya* sp. *Tetrahedron* 58: 7959-7966.

Ma, Y., Jiao, N.Z., Zeng, Y.H. (2004). Natural community structure of cyanobacteria in the South China Sea as revealed by rpoC1 gene sequence analysis. *Lett Appl Microbiol* 39: 353-358.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. F., Wright, F. (2009). TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25: 126-127.

Moffitt, M. C., Blackburn, S. I., Neilan, B. A. (2001). rRNA sequences reflect the ecophysiology and define the toxic cyanobacteria of the genus *Nodularia*. *Int J Syst Evol Micbiol* 51: 505-512.

Nagle, D. G., Paul, V. J. (1999). Production of secondary metabolites by filamentous tropical marine cyanobacteria: ecological functions of the compounds. *J Phycol* 35: 1412-1421.

Nei, M., Kumar, S. (2000). In *Molecular evolution and phylogenetics*. Oxford University Press, New York.

Nübel, U., Garcia-Pichel, F., Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63: 3327-3332.

Oren, A. (2004). A proposal for further integration of the cyanobacteria under the Bacteriological Code. *Int J Syst Evol Microbiol* 5: 1895–1902.

Paul, V. J., Thacker, W., Banks, K., Golubic, S. (2005). Benthic cyanobacterial bloom impacts the reefs of South Florida (Broward County, USA). *Coral Reefs* 24: 693-697.

Simmons, T. L., Engene, N., Ureña, L. D., Romero, L. I., Ortega-Barría, E., Gerwick, L., Gerwick, W. H. Viridamides A and B, lipodepsipeptides with anti-protozoan activity from the marine cyanobacterium *Oscillatoria nigro-viridis. J Nat Prod* 71: 1544-1550.

Sharp, K., Arthur, K. E., Gu, L., Ross, C., Harrison, G., Gunasekera, S. P., Meickle, T., Matthew, S., Luesch, H., Thacker, R. W., Sherman, D. H., Paul, V. J. (2009). Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* (Cyanobacteria: *Oscillatoriales*) from reefs of Southeast Florida. *Appl Env Microbiol* 75: 2879–2888.

Stanier, R. Y., Sistrom, W. R., Hansen, T. A. (1978). Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. *Int J Syst Bacteriol* 28: 335–336.

Suda, S., Liu, Y., He, J., Hu, Z., Hiroki, M., Watanabe, M. M. (1998). Morphological, biochemical and. physiological characteristics of *Lyngbya hieronymusii*. *var*. *hieronymusii* (Oscillatoriales, Cyanobacteria). *Phycol Resarch* 46: 51-55.

Sumina, E. L. (2006). Behavior of filamentous cyanobacteria in laboratory culture. *Microbiology* 75: 459-464.

Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599.

Tan, L. T. (2007). Bioactive Natural Products from Marine Cyanobacteria for Drug Discovery. *Phytochem* 68: 954-979.

Taylor, M. W., Radax, R., Steger, D., Wagner, M. (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* 71: 295-347.

Thacker, R. W., Paul, V. J. (2004). Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales). *Appl Env Microbiol* 70: 3305-3312.

Toledo, G., Palenik, B. (1997). *Synechococcus* diversity in the California current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Env Microbiol* 63: 4298-4303.

Tomitani, A., Knoll, A. H., Cavanaugh, C. M., Ohno, T. (2006). The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A* 103: 5442-5447.

Wilmotte, A., Herdman, M. (2001). Phylogenetic Relationships Among Cyanobacteria Based on 16S rRNA Sequences. In *Bergey's Manual of Systematic Bacteriology*. 1:487-493. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Chapter IV

Intra-genomic 16S rRNA Gene Heterogeneity in Cyanobacterial Genomes

Abstract

The ability of the small ribosomal subunit (16S) rRNA to infer fine-scale phylogenetic relationships is, in principal, impaired by the presence of multiple and variable gene copies within the same genome. This chapter investigated the extent of intra-genomic 16S rRNA gene heterogeneity in cyanobacteria. Using bioinformatics, all available sequenced cyanobacterial genomes were screened for microheterogeneity between their paralogous ribosomal genes. As a result, cyanobacteria were found to commonly contain multiple ribosomal operons and the numbers of copies were relatively proportional to genome size. Moreover, intra-genomic paralogous 16S rRNA gene copies often contain point-mutations that were validated by secondary structure modeling to be true point mutations rather than sequencing errors. Although microheterogeneity between paralogous 16S rRNA genes is relatively common in cyanobacterial genomes, the degree of sequence divergence is relatively low. We conclude that cyanobacterial intra-genomic 16S rRNA gene heterogeneity generally has a relatively small impact on species delineation and inference of evolutionary histories of cyanobacteria.

IV.1 Introduction

In consistency with the unifying biological theory of evolution, a framework of phylogenetic relationships is necessary for establishing informative and reliable taxonomic systems (Pleijel & Rouse, 2000). In the transition from morphology to phylogeny based classification systems, it is crucial to carefully determine optimum evolutionary metrics as well as their capacities to infer evolutionary relationships. In this regard, the small ribosomal subunit (SSU) rRNA gene has long been considered the "gold standard" in interpreting evolutionary relationships, and the 16S rRNA gene has consequently been embraced in cyanobacterial classification systems (Woese *et al*., 1990; Wilmotte & Herdman, 2001; Hoffman *et al*., 2005).

A recognized limitation in using the 16S rRNA genes has been that ribosomal genes often occur in multiple copies in microbial genomes and, importantly, that paralogous gene copies can vary in sequence between the different operons (Acinas *et al*., 2004; Case et al. 2007). Recent studies have detected intra-genomic gene heterogeneity of the 16S rRNA gene in cyanobacteria of the genus *Lyngbya* (Engene *et al*., 2010 = Chapter III). In this cyanobacterial genus the genetic variation between the paralogous 16S rRNA gene copies was of equivalent magnitude to sequence variation between different morphological species. This degree of intra-genomic gene heterogeneity challenges the ability of this gene to reliably distinguish between different species, let alone between different sub-species. However, the frequency and amount of general intra-genomic gene heterogeneity in cyanobacteria is, to date, unclear. *In silico* investigations of sequenced bacterial genomes have been conducted, but these studies

included a limited number of cyanobacterial genomes (Coenye & Vandamme, 2003; Acinas *et al*., 2004; Case *et al*., 2007). This study was aimed at assessing general heterogeneity in cyanobacterial 16S rRNA genes through a bioinformatic screening of sequenced cyanobacterial genomes. Moreover, the implication of potential heterogeneity on fine-scale phylogenetic resolution was considered.

IV.2 Materials and Methods

Cyanobacterial genomes were obtained from the National Center for Biotechnology information (NCBI) Microbial Genome Database web pages (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html). The BLASTn 2.2.18 alignment algorithm available at the NCBI web page was used to screen and obtain gene sequences as well as to determine percent sequence divergence among them. Multiple sequence alignments were performed using the L-INS-I algorithm in MAFFT 6.717 and manually refined (Katoh & Toh, 2008). Secondary RNA structures were predicted by the CLC Combined Workbench 3.5.2 (CLC bio). Mutation types and domains of the 16S rRNA genes were determined by superimposing their secondary structures on the SSU model for *Escherichia coli* strain J01695 (Cannone *et al*., 2002). A total of twelve house-keeping genes (*DNA-G, FRR, rpsB, NusA, PGK, PyrG, rpoB, rpsC, rpl2, rpl3, rpl4,* and *TSF*) were downloaded and concatenated for phylogenetic inference of the cyanobacterial strains. Appropriate nucleotide substitution model (WAG + I + G) was selected using Akaike information criterion (AIC/AIC_C) and Bayesian information criterion (BIC) in jModeltest. The Maximum likelihood (ML) inference was performed

using RaxML. The analysis was run using the WAG + I + G model (selected by AIC and BIC criteria) assuming a heterogeneous substitution rates and gamma substitution of variable sites (proportion of invariable sites (pINV) = 0.265, shape parameter (α) = 0868, number of rate categories = 4). Bootstrap resampling was performed on 500 replicates.

IV.3 Results and Discussion

IV.3.1 Redundancy of Ribosomal Genes

To gain insight into the redundancy of cyanobacterial ribosomal genes, all 59 publicly accessible cyanobacterial genomes were assessed for the number of ribosomal genes (rrs) and their clustering into ribosomal operons (rrn) (Table IV.1). In the vast majority (>97%) of the cyanobacterial rrn's, the ribosomal genes were organized in the typical bacterial 16S-23S-5S gene configuration. The number of rrn's varied between one to four copies with an average of 1.8 ± 0.8 copies per genome. This is relatively low compared with the 4.2 rrn's present in an average bacterial genome (Case *et al*. 2007). A plausible explanation for the average low copy number of ribosomal operons in cyanobacterial genomes was the fact that 45 out of the 59 sequenced genomes belonged to the unicellular order Chroococcales and among the genomes that possessed one rrn per genome, more than 77% belonged to these unicellular forms (Figure IV.1). Unicellular cyanobacteria are mainly pelagic, and similar trends of smaller rrn copy numbers are also observed in the pelagic filamentous *Arthrospira maxima* CS-328, *Arthrospira* sp. PCC 8005 and *Arthrospira platensis* Paraca, which live predominately in the open ocean environment. The open ocean represents a relatively constant habitat that allows

organisms to specialize in their environment. Specialized organisms normally have less need for rapid adjustments and consequently possess a lower number of rrn copies compared with more generalized organisms (Klappenbach *et al*., 2000). Moreover, Chroococcales have relatively small size genomes (mean: 3.2 ± 1.5 Mbp), ranging from 1.6 Mbp in *Prochlorococcus marinus* MIT 9301 to *Acaryochloris marina* MBIC11017 with an abnormally large genome size of 6.5 Mbp. Thus, their lower rrn copy numbers may be a consequence of the smaller genome sizes of these unicellular cyanobacteria.

In filamentous cyanobacteria, the genomes increase in size and were found to typically contain multiple rrn copies. For example, the average genome size for filamentous forms of cyanobacteria was 6.1 ± 1.6 Mbp and contained an average of 2.4 ± 1.3 rrn's. The order Oscillatoriales (n = 7 strains) has an average genome size of 6.8 ± 1.2 Mbp and contained an average of 1.7 ± 1.2 rrn's, while the heterocystous order Nostocales (n = 7 strains) has an average genome size of 5.5 ± 1.7 Mbp and contained an average of 3.0 ± 1.2 rrn's. Smaller genomes with few ribosomal operons may, therefore, be a consequence of specialized organisms living in a uniform environment.

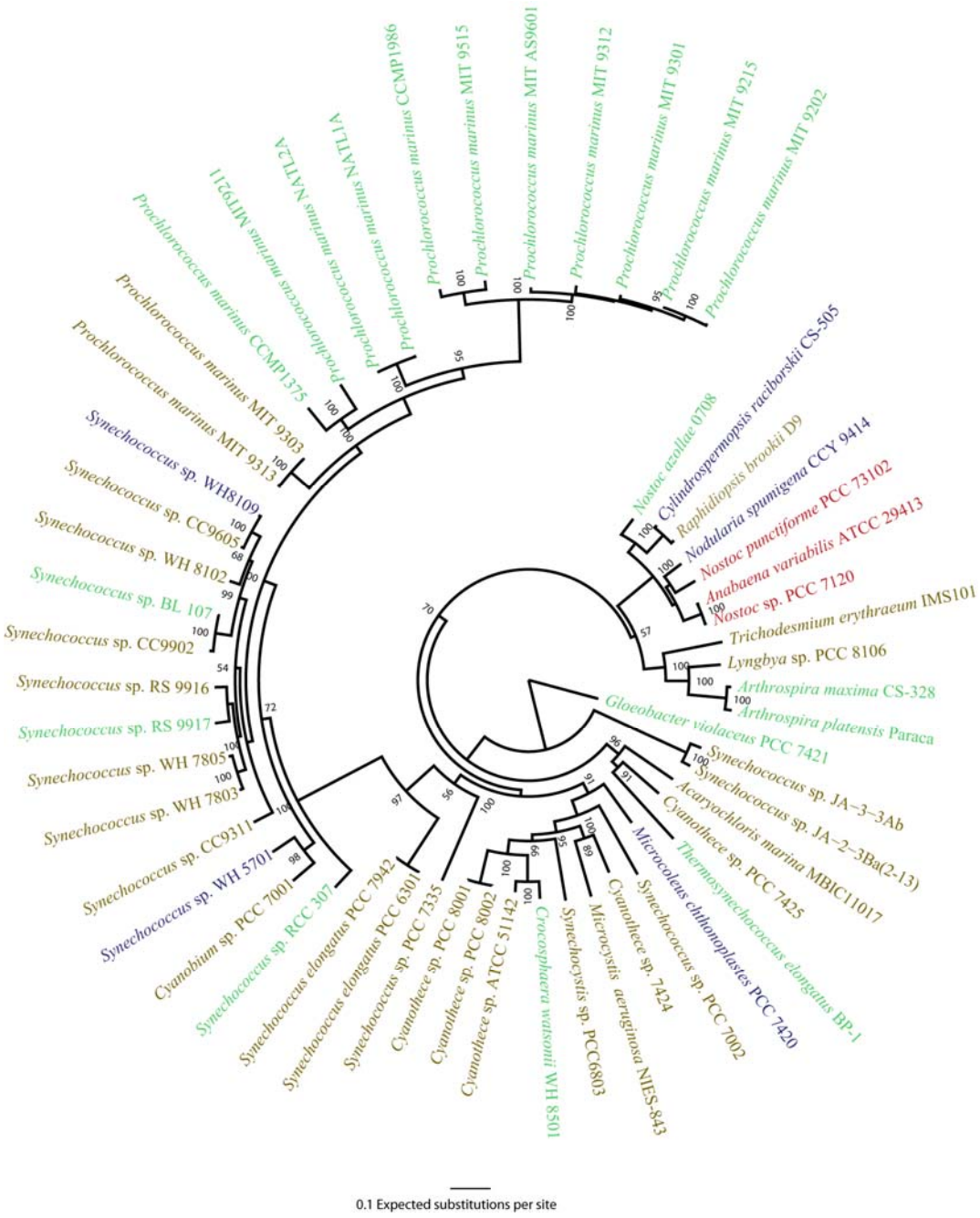


Figure IV.1: Phylogenetic relationships of the 59 cyanobacterial strains with sequenced genomes. The number of paralogous 16S rRNA gene copies within each genome is displayed in the following colors: green = 1 copy, brown = 2 copies, blue = 3 copies and red = 4 copies. The phylogenetic inference was performed by concatenation of twelve house-keeping genes using RaxML. The scale bar is equivalent to 0.1 expected substitutions per nucleotide position.

IV.3.2 16S rRNA Gene Heterogeneity

In total, 62.7% of all cyanobacterial genomes and 64.3% of filamentous forms contained more than one ribosomal operon (Table IV.1). Among these 37 cyanobacterial genomes with multiple rrn's, more than one third (35.1%) of the genomes displayed sequence divergence between at least two of their 16S rRNA gene copies. As these assessments are based on bioinformatic analyses of sequenced genomes, it is important to note that assembly programs typically assemble closely related (<6%) sequence reads together and form consensus sequences. The formation of consensus sequences has been found to overlook sequence variations between paralogous 16S rRNA gene copies, as observed in the unpublished genome assembly of *Lyngbya majuscula* 3L, and thus, genome sequences likely underestimate their microheterogeneity (E. Monroe, personal communication).

The intra-genomic sequence divergence between the 16S rRNA genes of different rrn's ranged from 0% to 0.6% with an average divergence of 0.2% in all cyanobacteria. This value is relatively low compared with other gram-negative bacteria (Case *et al*., 2007; Rastogi *et al*., 2009). The intra-genomic gene sequence divergence in filamentous cyanobacteria was slightly higher with 0.3% divergence, which could correlate to variety between the higher numbers of copies. However, there also appears to be a trend between genomes with higher gene sequence divergence and secondary metabolite diversity. For example, the microcystin-producing *Microcystis aeruginosa* NIES-843 and the hepatotoxic *Nodularia spumigena* CCY 9414 both have relatively high rates of divergence between their 16S rRNA genes (0.3% and 0.5% sequence divergence, respectively). A potential reason for this large divergence may be that the genomes of

such cyanobacteria undergo higher rates of recombination, which ultimately results in a richer diversity of secondary metabolites as well as a higher rate of intra-genomic gene heterogeneity in house-keeping genes.

IV.3.3 Mutation Types in Intra-genomic 16S rRNA Gene Variations

The mutation types and locations between paralogous 16S rRNA genes were examined by superimposing the sequenced genes on established secondary structure models. This was undertaken to examine the likelihood that the nucleotide substitutions in sequenced genomes were true mutations and not artifacts of genome sequencing. Summarizing the 41 mutations responsible for intra-genomic 16S rRNA gene heterogeneity in sequenced cyanobacterial genomes, 61% were nucleotide substitutions, 37% were nucleotide deletions and only 2% were nucleotide insertions. Among the 24 nucleotide substitutions, 79% were transitional mutations and 21% were transversional mutations (Table IV.1). Transitional mutations are usually point-mutations which result from errors caused by DNA-polymerases during replication and are estimated to constitute approximately two thirds of all nucleotide substitutions (Collins & Jukes, 1994).

The majority of the nucleotide substitutions were located in the ribosomal loop-regions, either in hairpin-loops (44%) or in interior-loops (31%) of the gene helices. The mutations occuring in stem-regions (25%) were typically limited to cytosine to thymine substitutions and only occurred if the nucleotide on the complementary DNA strand was a guanine. The resulting uracil and guanine base pairing is energetically allowed, even though it is slightly more constraining than the cytosine and guanine Watson-Crick base-

pairing (Gautheret *et al*., 1995). Thus, the observed nucleotide substitutions in these genomic 16S rRNA genes were typically the result of thermodynamically allowed gene mutations. The secondary structures of ribosomal genes are essential for ribosome assemblage and ribosomal protein interactions (Van de Peer *et al*., 1996). Therefore, the conserved secondary structures observed in these cyanobacterial 16S rRNA gene sequences are unlikely to affect ribosomal function, and are thus tolerated without strong selection pressure. On the basis of the types of mutations and their locations, we predict that all or a large proportion of the observed nucleotide variations are true mutations rather than artifacts from genome sequencing.

Table IV.1: Ribosomal Operon Redundancy and Intra-genomic SSU rRNA gene Heterogeneity in Cyanobacterial Genomes.

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Prochlorococcus (n = 13)									
P. marinus MIT 9301	CP000576	1.6	1	-	1498	-	-	-	-
P. marinus MIT 9202	ACDW0000	1.7	1	-	1441	-	-	-	-
P. marinus AS 9601	CP000551	1.7	1	-	1465	-	-	-	-
P. marinus CCMP1986	CP000951	1.7	1	-	1465	-	-	-	-
P. marinus MIT 9211	CP000878	1.7	1	-	1465	-	-	-	-
P. marinus MIT 9312	CP000111	1.7	1	-	1465	-	-	-	-
P. marinus MIT 9215	CP000825	1.7	1	-	1465	-	-	-	-
P. marinus MIT 9515	CP000552	1.7	1	-	1465	-	-	-	-
P. marinus CCMP1375	AE017126	1.8	1	-	1465	-	-	-	-
P. marinus NATL1A	CP000553	1.9	1	-	1465	-	-	-	-
P. marinus NATL2A	CP000095	1.9	1	-	1451	-	-	-	-
P. marinus MIT9313	BX548175	2.4	2	0	1465	-	-	-	-

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
P. marinus MIT9303	CP000554	2.7	2	0	1401	-	-	-	-
Synechococcus (n = 18)									
Synechococcus sp. WH 8109	ACNY0000	2.1	3	0	1456	-	-	-	-
P. marinus MIT9303	CP000554	2.7	2	0	1401	-	-	-	-
Synechococcus sp. RCC 307	CT978603	2.2	1	-	1498	-	-	-	-
Synechococcus sp. CC9902	CP000097	2.2	2	0	1479	-	-	-	-
Synechococcus sp. BL107	AATZ0000	2.3	1	-	1479	-	-	-	-
Synechococcus sp. WH 8102	BX548020	2.4	2	2	1462/1464	0.1	2	2d	HL, IL
Synechococcus sp. WH 7803	CT971583	2.4	2	0	1497	-	-	-	-

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Synechococcus sp. CC 9605	CP000110	2.5	2	0	1440	-	-	-	-
Synechococcus sp. RS9917	AANP0000	2.6	1	-	1448	-	-	-	-
Synechococcus sp. WH 7805	AAOK0000	2.6	2	2	1498/1497	0.1	1	d	IL
Synechococcus sp. CC 9311	CP000435	2.6	2	2	1477	0.1	1	ti	HL
Synechococcus sp. RS 9916	AAUA0000	2.7	2	2	1498	0.1	1	d	IL
S. elongatus PCC 6301	AP008231	2.7	2	0	1489	-	-	-	-
S. elongatus PCC 7942	CP000100	2.7	2	0	1490	-	-	-	-
Synechococcus sp. JA-3-3Ab	CP000239	2.9	2	0	1480	-	-	-	-

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Synechococcus sp. PCC 7002	AB015058	3.0	2	0	1452	-	-	-	-
Synechococcus sp. WH 5701	AANO0000	3.0	3	0	1447	-	-	-	-
Synechococcus sp. JA-2-3B'a(2-13)	CP000240	3.0	2	0	1479	-	-	-	-
Synechococcus sp. PCC 7335	AB015062	6.0	2	0	1448	-	-	-	-
Cyanothece (n = 7)									
Cyanothece sp. PCC 8801	CP001287	4.7	2	0	1479	-	-	-	-
Cyanothece sp. PCC 8802	CP001701	4.7	2	2	1479	0.1	1	ti	HL
Cyanothece sp. ATCC 51142	CP000806	4.9	2	2	1489	0.1	1	tv	S

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Cyanothece sp. PCC 7822	ABVE0000	5.7	2	0	1492	-	-	-	-
Cyanothece sp. PCC 7425	CP001344	5.4	2	0	1474	-	-	-	-
Cyanothece sp. CCY 0110	AAXW0000	5.9	2	0	1488	-	-	-	-
Cyanothece sp. PCC 7424	CP001291	5.9	2	0	1483	-	-	-	-
Other Chroococcales (n = 7)									
Thermosynechococcus elongatus BP-1	BA000039	2.6	1	-	1491	-	-	-	-
Cyanobium sp. PCC 7001	ABSE0000	2.8	2	0	1448	-	-	-	-
Synechocystis sp. PCC 6803	BA000022	3.6	2	0	1489	-	-	-	-

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Gloeobacter violaceus PCC 7421	BA000045	4.7	1	-	1484	-	-	-	-
Microcystis aeruginosa NIES-843	AP009552	5.8	2	2	1477	0.3	4	4ti	2IL, 2HL
Crocosphaera watsonii WH 8501	AADV0000	6.2	1	-	1408	-	-	-	-
Acaryochloris marina MBIC11017	CP000828	6.5	2	0	1500	-	-	-	-
Oscillatoriales (n = 7)									
Arthrospira platensis Paraca	ACKSK0000	5.0	1	-	1483	-	-	-	-

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Arthrospira maxima CS-328	ABYK0000	6.0	1	-	1482	-	-	-	-
Arthrospira sp. PCC 8005	ADDH0000	6.1	1	-	1482	-	-	-	-
Arthrospira maxima CS-328	ABYK0000	6.0	1	-	1482	-	-	-	-
Arthrospira sp. PCC 8005	ADDH0000	6.1	1	-	1482	-	-	-	-
Oscillatoria sp. PCC 6506	CACA00000000	6.7	1	-	1480	-	-	-	
Lyngbya aestoarii PCC 8106	AAVU0000	7.0	2	2	1493/1494	0.3	4	3ti, i	2HL, 2IL
Microcoleus chthonoplastes PCC7420	ABRS0000	8.7	4	3	1482	0.6	9	2ti, tv, 6d	3HL, S*, 4IL

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Nostoc azollae 0708	ACIR0000	5.0	1	-	1492	-	-	-	-
Anabaena variabilis ATCC 29413	CP000117	6.4	4	0	1488	-	-	-	-
Trichodesmium erythraeum IMS101	CP000393	7.8	2	0	1482	-	-	-	-
Nostocales (n = 7)									
Raphidiopsis brookii D9	ACYB0000	3.2	2	2	1487	0.1	1	ti	S*
Nostoc punctiforme PCC 7120	BA000019	6.4	4	2	1489	0.1	1	ti	5’

Table IV.1: Continued

Species (strain)	Acc. Nr.	Genome size (Mbp)	SSU Nr.[1]	Variable SSU	Gene Length (bp)	P-distance (%)	Hetero. Positions	Mutation types[2]	Structural location[3]
Nostoc punctiforme PCC 73102	CP001037	8.2	4	2	1489	0.1	2	2ti	2S*
Cylindrospermopsis raciborskii CS-505	ACYA0000	3.9	3	0	1487	-	-	-	-
Nodularia spumigena CCY 9414	AAVW0000	5.3	3	3	1563/1566/1485	0.5	13	5ti, 3tv, 5d	3IL, 2S*, HL, 4C, 3S

Abbreviations: (-) Uninformative, (NA) unavailable.
[1] Number of rRNA operons per genome.
[2] Mutation types: tv = transversion mutation, ti = transition mutation, I = insert, d = deletion.
[3] Structural location is based on superimposing of the gene to the secondary structure of *E. coli* J01695. Helix position: HL = hairpin-loop, IL = interior-loop, S = stem-region, S* = substitution in stem-region resulting in a U-G bond, C = substitution in stem-region with a complementary substitution in the opposite strand, 5'-end is in the gene's single stranded upstream region, 3'-end refers to the single stranded downstream region.

IV.4 Conclusions

Cyanobacteria, like most bacteria, frequently have multiple copies of their ribosomal operons, and the number of these appears to be reasonably correlated with genome size of the organism. Among cyanobacteria with multiple ribosomal gene copies, variations between gene copies are relatively frequent and may be underestimated due to limitations of genome assembly programs. Despite the relatively frequent occurrence of intra-genomic 16S rRNA gene heterogeneity, the extent of sequence divergence is typically quite small (mean = 0.2%) and generally much less than in other groups of bacteria (Case *et al*., 2007). Furthermore, the average cyanobacterial intra-genomic 16S rRNA gene heterogeneity (mean = 0.2%) is of an order less than the 3% gene sequence divergence typically used to delineate microbial species (Tindall *et al*., 2010). It is, however, important to note that much higher degrees of intra-genomic gene heterogeneity (>1%) have been reported in filamentous marine cyanobacteria (Engene et al. 2010). Moreover, single nucleotide substitutions can, in principal, have large impacts on the interpretations of DNA fingerprinting methods if these mutations are located in restriction sites (Roudière *et al*., 2007). Intra-genomic gene heterogeneity should, therefore, not be neglected when inferring phylogenetic relationships of cyanobacteria, especially on an infra-species level. However, we argue that the general degree of microheterogeneity in paralogous 16S rRNA gene copies of cyanobacteria is relatively small and, thus, has only minor impact on the inference of phylogenetic relationships and evolutionary histories.

IV.5 Acknowledgements

We gratefully thank S. Podel and E. Monroe for bioinformatics advice and discussions. Financial support for this study came from the Swedish-American Foundation (NE), NIH grants (CA-108874 and NS053398).

The text of IV, in part, is published material as it appears in: Engene, N., Gerwick W.H. (2011). Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. *Fottea* 1: 17-24. The dissertation author was the primary author and conducted the majority of the research which forms the basis for this chapter.

IV.6 References

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., Polz. M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacter* 186: 2629-2635.

Cannone, J. J., Subramanin, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pnde, N., Schang, Z., Yu, N., Gutell R. R. (2002). The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3: 1471-2105.

Case, R. J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W. F., Kjelleberg, S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Env Mic Biol* 73: 278-288.

Coenye, T., Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* 228: 45-49.

Collins, D. W., Jukes, T. H. (1994). Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* 20: 386-96.

Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601.

Gautheret, D., Konings, D., Gutell, R. R. (1995). G.U base pairing motifs in ribosomal RNA. *RNA* 1: 807-814.

Hoffman, L., Komárek, J., Kaštovskŷ, J. (2005). System of cyanoprokaryotes (cyanobacteria) - state in 2004. *J Arch Hydrobiol* 159: 95-115.

Katoh, K., Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9, 286-298.

Klappenbach, J. A., Dunbar, J. M., Schmidt, T. M. (2000). rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. *Appl Env Mic Biol* 66: 1328-1333.

Pleijel, F., Rouse, G. W. (2000). Least-inclusive taxonomic unit: a new taxonomic concept for biology. *Proc R Soc Lond* 267: 627-630.

Rappe, M. S., Connon, S. A., Vergin, K. L., Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418: 630-633.

Rastogi, R., Wu, M., DasGupta, I., Fox, G. E. (2009). Visualization of ribosomal RNA operon copy number distribution. *BMC Microbiol* 25: 208.

Roudière, L., Lorto, S., Tallagrand, E., Marchandin, H., Jeannot, J.-L., Jumas-Bilak, E. (2007). Molecular fingerprint of bacterial communities and 16S rDNA intra-species heterogeneity: a pitfall that should be considered. *Pathologie Biologie* 55: 434-440.

Tindall, B. J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 60: 249-266.

Van de Peer, Y., Chapelle, S., De Wachter, R. (1996). The European database on small subunit ribosomal RNA. *Nucleic Acids Res* 24: 3381-3391.

Wilmotte, A., Herdman, M. (2001). Phylogenetic Relationships Among Cyanobacteria Based on 16S rRNA Sequences. In *Bergey's Manual of Systematic Bacteriology*. 1:487-493. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Woese, C. R., Kandler, O., Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Nati Acad Sci USA* 87: 4576-4579.

Chapter V

Underestimated Biodiversity as a Major Explanation for the Perceived Rich Secondary Metabolite Capacity of the Cyanobacterial Genus *Lyngbya*

Abstract

Marine cyanobacteria are prolific producers of bioactive secondary metabolites responsible for harmful algal blooms as well as rich sources of promising biomedical lead compounds. The current study focused on obtaining a clearer understanding of the remarkable chemical richness of the cyanobacterial genus *Lyngbya*. Specimens of *Lyngbya* from various environmental habitats around Curaçao were analyzed for their capacity to produce secondary metabolites by genetic screening of their biosynthetic pathways. The presence of biosynthetic pathways was compared with the production of corresponding metabolites by LC-ESI-MS^2 and MALDI-TOF-MS. The comparison of biosynthetic capacity and actual metabolite production revealed no evidence of genetic silencing in response to environmental conditions. On a cellular level, the metabolic origin of the detected metabolites was pinpointed to the cyanobacteria, rather than the sheath-associated heterotrophic bacteria, by MALDI-TOF-MS and multiple displacement amplification of single-cells. Finally, the traditional morphology-based taxonomic identifications of these *Lyngbya* populations were combined with their phylogenetic relationships. As a result, polyphyly of morphologically similar cyanobacteria was identified as the major explanation for the perceived chemical richness of the genus *Lyngbya*, a result which further underscores the need to revise the taxonomy of this group of biomedically important cyanobacteria.

V.1 Introduction

Natural products discovery programs focused on the cyanobacterial genus *Lyngbya* continue to yield an extraordinary diversity of biologically active secondary metabolites (Gerwick *et al.*, 2008; Tidgewell *et al.*, 2010). Marine, tropical forms of *Lyngbya* in particular have been very affluent in this regard and have yielded over 260 different secondary metabolites (>40% of all reported marine cyanobacterial molecules) (MarinLit, 2010). Even more remarkable is the fact that a total of 196 different secondary metabolites have been reported from a single species, *Lyngbya majuscula*.

The natural biological roles of these secondary metabolites are often ascribed to defensive functions with many having potent toxicity (Thacker *et al.*, 1997; Capper *et al.*, 2006). Exploitation of these bioactivities has yielded a number of important natural products with therapeutic potential (Tan, 2007; Gerwick *et al.*, 2008). Hence, a clear understanding of the extent and origin of this chemical diversity in *Lyngbya* is important to enhancing the discovery of new natural products (NP) as well as for predicting some types of harmful algae blooms (HAB).

Despite nearly 30 years of investigation of cyanobacteria for their unique secondary metabolites, the full extent of the biosynthetic capacities of these microorganisms is still largely unknown. Transcriptional expression of cyanobacterial secondary metabolite pathways have been shown to be influenced by environmental factors (Kaebernick *et al.*, 2000; Shalev-Malul *et al.*, 2008; Sorrels *et al.*, 2009). Differential expression of secondary metabolites depending on environmental conditions has also been highlighted as an alternative explanation for the reported chemical diversity

from genetically related *Lyngbya* populations at different collection sites (Thacker & Paul, 2004). Unfortunately, at the present time there is little genomic information available for these largely marine, filamentous cyanobacteria, so a true assessment of their capacity for natural products biosynthesis remains hidden (NCBI Microbial Genomes).

The identification of the vast majority of chemically interrogated strains of cyanobacteria has mainly been founded on traditional morphology-based taxonomic systems. These "morpho-species" of the genus *Lyngbya* are traditionally defined as filamentous non-heterocystous cyanobacteria with discoid cells enclosed within distinct sheaths (Castenholz, 2001; Komárek & Anagnostidis, 2005). The recent inclusion of phylogenetics into taxonomic classification, however, greatly enhances understanding of the relatedness of cyanobacteria and has led to major revisions of traditional genera (Hoffman *et al.*, 2005). Along these lines the genus *Lyngbya* has been recognized as a polyphyletic group (Sharp *et al.*, 2009; Engene *et al.*, 2010). Moreover, secondary metabolites described from *Lyngbya* morpho-types have been reisolated from distantly related phylogenetic lineages (Sharp *et al.*, 2009). However, the true taxonomic identities of these different *Lyngbya* lineages have not yet been clarified in reference to type-strains.

An alternative explanation for the distribution of secondary metabolites among evolutionary distinct *Lyngbya* populations could be that microorganisms associated with the cyanobacterial filaments are responsible for the production of the secondary metabolites. *Lyngbya* populations often form extensive mat-like colonies which are frequently colonized with complex microbial communities of diverse heterotrophic

bacteria or other epiphytic cyanobacteria (Engene *et al.*, 2010). Dissection of cyanobacterial assemblages has shown that in some cases cyanobacteria which are epiphytic on *Lyngbya* can also produce bioactive secondary metabolites (Simmons *et al.*, 2008). Moreover, the thick polysaccharide sheaths enveloping *Lyngbya* filaments provide a haven for heterotrophic bacteria (Simmons *et al.*, 2007), and these may also contribute to the metabolites reported from field collections of this genus. Indeed, some of the natural products isolated from *Lyngbya* specimens structurally resemble those of heterotrophic bacteria (Graber & Gerwick, 1998; Burja *et al.*, 2001). Thus, due to the microbial complexity of *Lyngbya* colonies, the true metabolic origin of secondary metabolites isolated from *Lyngbya* populations remains uncertain.

An excellent example of the extraordinary secondary metabolite diversity found in marine *Lyngbya* derives from the Caribbean island of Curaçao. Despite its rather small coast line (total land area: 444 km^2), *Lyngbya* populations from various sites around the leeward half of the island (<40 coast line miles) have yielded at least sixteen novel molecules (Gerwick *et al.*, 1994; Orjala & Gerwick, 1994; Orjala *et al.*, 1995; Orjala & Gerwick, 1997; Wu *et al.*, 1997; Graber & Gerwick, 1998; Wu *et al.*, 2000). Considering that the shallow-water margin along most of Curaçao's perimeter is narrow, the microenvironment found there may well represent the greatest *Lyngbya* chemo-diversity examined to date.

In the current study, cyanobacteria corresponding with the *Lyngbya* morpho-type (herein referred to as "*Lyngbya*") were sampled from diverse environments around Curaçao. Phylogenetic inferences of the conserved SSU (16S) rRNA genes were included to give a more complete understanding of the relatedness among these populations and

insight into how secondary metabolite production was distributed between different lineages. Genetic and chemical approaches were combined to thoroughly explore the biosynthetic capacity of different samples, and comparison between the biosynthetic capacity versus the actual biosynthetic production of discreet indicator natural products allowed for hypotheses concerning variations arising from differential regulation of gene expression. Additionally, specimens from diverse habitats were compared to evaluate if environmental influences might impact the observed distribution of these secondary metabolites. Finally, we also examined at the cellular level whether the secondary metabolites ascribed to *Lyngbya* are truly produced by this cyanobacterium, or if they are produced by microorganisms associated with its surfaces and sheath material.

The data from these various analyses, including gene regulation, metabolic origin and taxonomic revision, were synthesized and integrated into a model that provides a better understanding of the underlying mechanisms behind secondary metabolite diversity in marine cyanobacteria. As a result, polyphyly of morphologically similar cyanobacteria was identified as the major explanation for the perceived chemical richness of the genus, and further underscores the need to revise the taxonomy of this group of biomedically important cyanobacteria.

V.2 Materials and Methods

V.2.1 Sampling and Characterization

Cyanobacterial specimens were collected by SCUBA from 12 sites along the leeward coast of Curaçao, Netherlands Antilles (Figure V.1). Morphological

identification of the specimens was performed in accordance with modern taxonomic systems (Castenholz, 2003; Komárek & Anagnostidis, 2005). Specimens were directly cleaned from meio/macro-fauna under a dissecting scope in seawater filtered through 0.2 µm Acrodisc® Syringe filters for culturing and morphological analysis. Additional biomass (~200 mg) was preserved for genetic analysis in 10 mL RNA*later*® (Ambion Inc.) and for chemical analysis in seawater/EtOH (1:1) at -20 ºC (Table V.1). Light microscopy was performed using an Olympus IX51 epifluorescent microscope equipped with an Olympus U-CMAD3 camera.

Samples for scanning electron microscopy (SEM) were placed on indium-tin-oxide glass slides that had been coated with 0.1% polyethylenimine to facilitate adhesion. The samples were then fixed in 2.5% glutaraldehyde buffered in 1X PBS for 30 min followed by a secondary fix of 2% osmium tetroxide (OsO_4) for 15 min. Dehydration was done in a graded EtOH series. Samples were dried in a Balzers critical-point dryer with liquid carbon dioxide as the transition fluid and then sputter-coated with gold palladium using a Polaron E5100 SEM Coating. A Hitachi SU6600 Field Emission SEM was used to view the samples. Samples for transmission electron microscopy (TEM) were fixed overnight in a 4% formaldehyde and 1% glutaraldehyde solution (4F:1G) buffered in 1X PBS followed by a secondary fix of 2% OsO_4 for 1-2 h. Dehydration was done in a graded EtOH series. Samples were then embedded in Spurr's resin and left to polymerize for 48 h. Thin sections (70 nm) were obtained using the Reichart Ultracut E ultramicrotome and picked up on 75 mesh copper grids. The grids were subsequently stained with uranyl acetate and Sato lead (Hanaichi *et al*., 1986). A JEOL 1200EX TEM was used to view and record images of these samples.

Table V.1: Collection Data of Cyanobacterial Specimens from Curaçao.

Strain	Thallus morphology	Color	Location	Environment	Depth (m)	Biomass (g)	Yield (mg)(%)
3L	Hair-like tufts	Dark-red	Piscadera baii	Reef	>2	3.3	417 (12.6%)
NAC8-18	Hair-like tufts	Brownish	Playa Kalki	Reef	5-10	0.1	19.4 (19.4%)
NAC8-45	Mats/clumps	Dark-brown	Marie Pampoen	Harbour	2-3	38.8	1373 (3.6%)
NAC8-46	Mats/clumps	Dark-red	Caracas baii	Reef	8-10	267.7	2598 (1.0%)
NAC8-47	Hair-like tufts	Dark-red	Schlange baii	Reef	3-4	<0.5*	*
NAC8-48	Hair-like tufts	Dark-red	Seru Boca baii	Mangrove	>2	6.0	1121 (18.7%)
AC8-49	Hair-like tufts	Dark-red	Seru Boca channel	Rocks	>1	4.8	163 (3.4%)
NAC8-50	Tufts	Red-brown	Playa Barbara	Beach	>1	13.2	1154 (8.7%)
NAC8-51	Hair-like tufts	Blackish	Caracas baii Island	Rocks	>1	7.1	936 (13.2%)
NAC8-52	Hair-like tufts	Dark-brown	Brakkeput Abou	Harbour	>1	1.0	237 (2.4%)
NAC8-53	Hair-like strings	Blackish	Blauw baii	Reef	3-4	0.8	70.5 (8.8%)
NAC8-54	Mats/clumps	Dark-brown	Pierbaai reef	Reef	8-10	9.2	1154 (12.5%)
NAC8-55	Tufts	Dark-brown	Jan Theil baii	Reef	3-4	19.5	1603 (8.2%)

* Not sufficient biomass for extraction

V.2.2 Genomic DNA-extraction and Multiple Displacement Amplification (MDA)

Cyanobacterial filaments were cleaned and pretreated using TE (10 mM Tris; 0.1M EDTA; 0.5% SDS; 20 $\mu g \cdot mL^{-1}$ RNase)/lysozyme (1 $mg \cdot mL^{-1}$) at 37 °C for 30 min followed by incubation with proteinase K (0.5 $mg \cdot mL^{-1}$) at 50 °C for 1 h. Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Inc.) following the manufacturer's specifications. DNA concentration and purity was measured on a DU® 800 spectrophotometer (Beckman Coulter). Single-cells were isolated using a MM3-All micromanipulator (World Precision Instruments Inc.) as previously described (Engene *et al.*, 2010; Chapter III). DNA was amplified from single-cell genomes using the REPLI-g® Mini Kit (Qiagen Inc.) following the manufacturer's specifications. All MDA reactions were performed in 50 µL reaction volume for 16 h at 30 °C.

V.2.3 Screening of Biosynthetic Genes

PCR-primers were designed using Primer3Plus (Untergasser *et al.*, 2007) (Table V.2). The PCR reaction volumes were 25 µL containing 0.5 µL (~50 ng) of DNA, 2.5 µL of 10x PfuUltra IV reaction buffer, 0.5 µL (25 mM) of dNTP mix, 0.5 µL of each primer (10 µM), 0.5 µL of PfuUltra IV fusion HS DNA polymerase and 20.5 µL dH_2O. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 2 min at 95 °C, 25 cycles of amplification: 20 sec at 95 °C, 20 sec at 50 °C and 1.5 min at 72 °C, and final elongation for 3 min at 72 °C. PCR products were purified using a MinElute® PCR Purification Kit (Qiagen) before being subcloned using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) following the

manufacturer's specifications. Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced with M13 primers. Putative biosynthetic genes were compared with the original barbamide (AF516145) and curacin (AY652953) pathways as well as a putative, unpublished biosynthetic pathway for carmabin obtained from a partly sequenced genome of "*L. sordida"* strain 3L (only sequences with >99% sequence identity were accepted). The gene sequences are available in the DDBJ/EMBL/GenBank databases under acc. Nr. GU724195-GU724208.

Table V.2: Primers Used for PCR-amplification of Secondary Metabolite Encoding Genes.

Gene	Gene product	Secondary metabolites	Primer	Sequence (5'-3')	Gene size (bp)	Gene coverage % (bp)
*phy*H	α-kG-dependent halogenase	curacin	phyH-41F	ACCAACCTATCCCTGGCAAC	1615	55.0 (889)
			phyH-1593R	GCAGCGGTTGTAAGGTTTTC		
*bar*B1	halogenase	barbamide	barB1-23F	TGAAAACTAGGGCAAAAGCTG	903	94.7 (855)
			barB1-878R	TTCCCCTGCTACCAGAACAG		
*bar*B2	halogenase	barbamide	barB2-28F	AACTATCACTGCTGAACCCAGA	894	94.7 (847)
			barB2-875R	TTCCCCAGCCACTAGAACAG		
*car*MT1-MT	dimethyl transferase	carmabin	MT-MT-37F	AGGATGCTATTTTGGCGAGA	1412	97.4 (1375)
			MT-MT-1401R	TGACTATCCGAATTGCCTCA		
*car*TE	thioesterase	carmabin	carTE 32F	TCCATGGCTTCGATAGTTCC	633	94.8 (600)
			carTE 630R	AGCAGCAGCAGTCCCTAAA		
16S rRNA	General cyanobacterial SSU	-	CYA106F	CGGACGGGTGAGTAACGCGTGA	1403	>95%
			CYA1509R	GGTTACCTTGTTACGACTT		

V.2.4 Phylogenetic Inference

Gene sequences were aligned bi-directionally using the L-INS-i algorithm in MAFFT 6.717. A total of 1 378 bp (310 parsimony informative sites) of the 16S rRNA gene were analyzed without data exclusion. The evolutionary distant unicellular cyanobacterium *Gloeobacter violaceus* PCC 7421^{T} (NC005125) was included as an out-group. Representative type-strains (^{T}) were selected from *Bergey's Manual* (Castenholz, 2001). Pair-wise sequence divergences were calculated in PAUP* 4.0b10. Appropriate nucleotide substitution models were compared and selected using uncorrected/corrected Akaike Information Criterion (AIC/AICc), Bayesian Information Criterion (BIC), and the Decision-theoretic (DT) in jModelTest 0.1.1. The GTR+I+G model was selected by AIC/AICc/BIC/DT criteria assuming a heterogeneous substitution rates and gamma substitution of variable sites (proportion of invariable sites (pINV) = 0.292, shape parameter (α) = 0.316, number of rate categories = 4). The Maximum likelihood (ML) inference was performed using GARLI 1.0 for the GTR+I+G model with 1,000 bootstrap re-sampling. Bayesian analysis was conducted using MrBayes 3.1. Four Metropolis-coupled MCMC chains (one cold and three heated) was run for 10,000,000 generations. MCMC convergence was determined using AWTY and the first 1,000,000 generations (10%) were discarded as burn-in and the following data set was sampled with a frequency of every 1,000 generations. The maximum parsimony (MP) analysis was performed in PAUP* 4.0b10 using a heuristic search through the branch-swapping tree-bisection-reconnection (TBR) algorithm with the addition of 10,000 random replicates to find the most parsimonious tree. Bootstrap support was obtained from 1,000 replicates.

V.2.5 Secondary Metabolite Detection

Algal biomass (~1 g) of each specimen was exhaustively extracted with CH_2Cl_2–MeOH (2:1). The extract was dried under vacuum and the dried residues were re-dissolved in MeOH at a concentration of 1 $mg \cdot mL^{-1}$. Each sample (10 µL) was injected into an LC ESI-MS system (LCQ Advantage Max spectrometer and UV-profiles by Surveyor PDA plus detector, Thermo Finnigan) and separated on an RP HPLC column (HP Lichrosphere 100 RP-18, 4 × 125 mm, 5.0 µm) with step gradient elution of 0.1% formic acid in water (eluent A) and 100% ACN (eluent B). Gradient program: 0-5 min, B, 45%; 5-55 min, B, 45-100%; 55-65 min, B, 100%; flow rate, 700 $\mu L \cdot min^{-1}$. The column temperature was kept at 30 °C. The MS and MS^2 spectra and retention time of each peak were recorded using the positive ion detection mode. For MALDI-TOF-MS, 5-10 µg (wet wt) of each specimen were extracted with 1 $\mu L \cdot \mu g^{-1}$ matrix solution (70 $mg \cdot mL^{-1}$ alpha-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid (1:1), 750 µL acetonitrile, 248 µL dH_2O, 2 µL trifluoroacetic acid) in a 96-well plastic plate for 20-30 sec. One µL of matrix extract was deposited on a well of a Bruker Microflex MSP 96 Stainless Steel Target Plate and run on a Bruker microflex™ mass spectrometer equipped with flexControl 3.0. Identification of secondary metabolites required support of predicted isotope patterns, corresponding MS^2 fragmentations and conserved retention times (RT).

V.3 Results

V.3.1 Sampling and Taxonomic Identification

A total of 12 "*Lyngbya"* populations were obtained from a variety of environments ranging from a harbor/dock habitat, a sandy beach, an exposed reef and a salt water mangrove (Figure V.1). All 12 specimens were initially identified as species of "*Lyngbya"* based on the following traditional morphological criteria: (i) filamentous (width >6 μm), (ii) lack of specialized cells (*e.g.* no heterocysts or akinetes), (iii) isopolar trichomes, (iv) discoid cells and (v) presence of sheaths (Table V.3). Six of the specimens were identified as "*L. majuscula"* and the other six as "*L. sordida"* based on the cell width to cell length dimensions and the amount of cross-wall constriction between the cells (Table V.3).

Table V.3. Morphological Characterization of "*Lyngbya*" Specimens from Curaçao.

Species (strain)	**Thallus**	**Filament**	**Filament width**	**Sheath width**	**Cell width**	**Cell length**	**Length/ width**	**Cross-wall constriction**	**Granulation**	**Apical Cells**
sordida (3L)	hair-like tufts, dark-red	thick, colorless sheaths	62.6 (±2.4)	4.3 (±3.6)	58.3 (±5.7)	7.6 (±1.7)	0.13 (±0.02)	8.9% (± 0.9)	large, evenly distributed	rounded, w/o calyptra or thickened cell wall
sordida (NAC8-18)	hair-like tufts, brownish	brown, straight long (>1 cm), thin sheaths	18.7 (±0.6)	3.4 (±0.7)	15.4 (±0.3)	1.9 (±0.4)	0.12 (±0.03)	4.7% (± 0.3)	finely evenly granulated	annotated, convex w/o calyptra or thickened cell wall
majuscula (NAC8-45)	mats with erect clumps, blackish	reddish, barely visible sheaths	20.3 (±0.7)	3.4 (±0.4)	16.8 (±0.8)	2.0 (±0.1)	0.12 (±0.01)	>3.0% (± 0.2)	small Granula	blunt, slightly rounded w/o calyptra

Table V.3: Continued

Species (strain)	Thallus	Filament	Filament width	Sheath width	Cell width	Cell length	Length/ width	Cross-wall constriction	Granulation	Apical Cells
majuscula (NAC8-46)	mats with erect clumps, blackish	Red, short (>500µm) filaments	22.8 (±0.7)	2.6 (±1.0)	20.2 (±1.0)	1.9 (±0.6)	0.09 (±0.02)	>2.5% (± 0.1)	granulation at cross-walls	rounded, flat w/o calyptra
majuscula (NAC8-47)	hair-like tufts, dark-red	brown, (1-5 cm), straight, thick, sheaths	35.6 (±1.9)	2.8 (±1.2)	32.9 (±1.1)	2.3 (±0.2)	0.07 (±0)	>1.5% (±0.06)	Finely granulated, evenly	rounded (9 µm) w/o calyptra or thickened cell wall
sordida (NAC8-48)	hair-like tufts, dark-red	red-brown, (1-5 cm), straight, thick sheaths	56.2 (±7.3)	4.7 (±0.4)	51.5 (±7.5)	10.8 (±1.0)	0.21 (±0.03)	13.2% (±3.0)	large, evenly distributed	rounded w/o calyptra or thickened cell wall

Table V.3: Continued

Species (strain)	Thallus	Filament	Filament width	Sheath width	Cell width	Cell length	Length/ width	Cross-wall constriction	Granulation	Apical Cells
sordida (NAC8-49)	hair-like tufts, dark-red	red, 10-2 cm, straight, colorless sheaths	55.6 (±2.2)	9.8 (±1.4)	45.8 (±2.3)	10.5 (±1.5)	0.23 (±0.05)	13.3 % (±2.8)	large, evenly distributed	rounded, w/o calyptra or thickened cell wall
majuscula (NAC8-50)	tufts, red-brown	wine-red, straight or slightly curved, with barely visible sheaths	30.1 (±4.2)	3.4 (±1.1)	26.7 (±3.4)	4.5 (±0.9)	0.2 (±0)	>1.9% (±0.2)	granulated at cells cross-walls	blunt or slightly rounded w/o calyptra or thickened cell wall
sordida (NAC8-51)		dark-red, (1-5 cm) straight thick	52.4 (±3.8)	13.7 (±2.6)	38.7 (±2.3)	4.5 (±0.6)	0.12 (±0.02)	6.9% (±0.04)	large, evenly distributed	rounded w/o calyptra

Table V.3: Continued

Species (strain)	Thallus	Filament	Filament width	Sheath width	Cell width	Cell length	Length/ width	Cross-wall constriction	Granulation	Apical Cells
sordida (NAC8-52)		red, 1-5 cm straight	37.6 (±4.3)	3.3 (±0.8)	34.3 (±3.5)	5.8 (±0.9)	0.17 (±0.01)	7.9% (±2.5)	large, evenly distributed	rounded w/o calyptra
sordida (NAC8-53)	red-brown	red, long (1-5 cm) straight	73.4 (±15.8)	7.7 (±1.9)	65.6 (±15.0)	9.0 (±3.0)	0.14 (±0.02)	6.5% (±1.4)	large, evenly distributed	rounded no calyptra
majuscula (NAC8-54)		wine-red, curved, sheaths	22.4 (±0.4)	1.6 (±1.3)	20.8 (±0.9)	2.1 (±0.5)	0.10 ±0.03)	>2.4% (±0.1)	granulated at cross-walls	rounded, w/o calyptra
majuscula (NAC8-55)		brown-red	38.0 (±0.7)	2.7 (±0.4)	35.4 (±1.0)	3.1 (±0.3)	0.09 (±0.01)	>0.5 (1.4%)	granulated	rounded, w/o calyptra

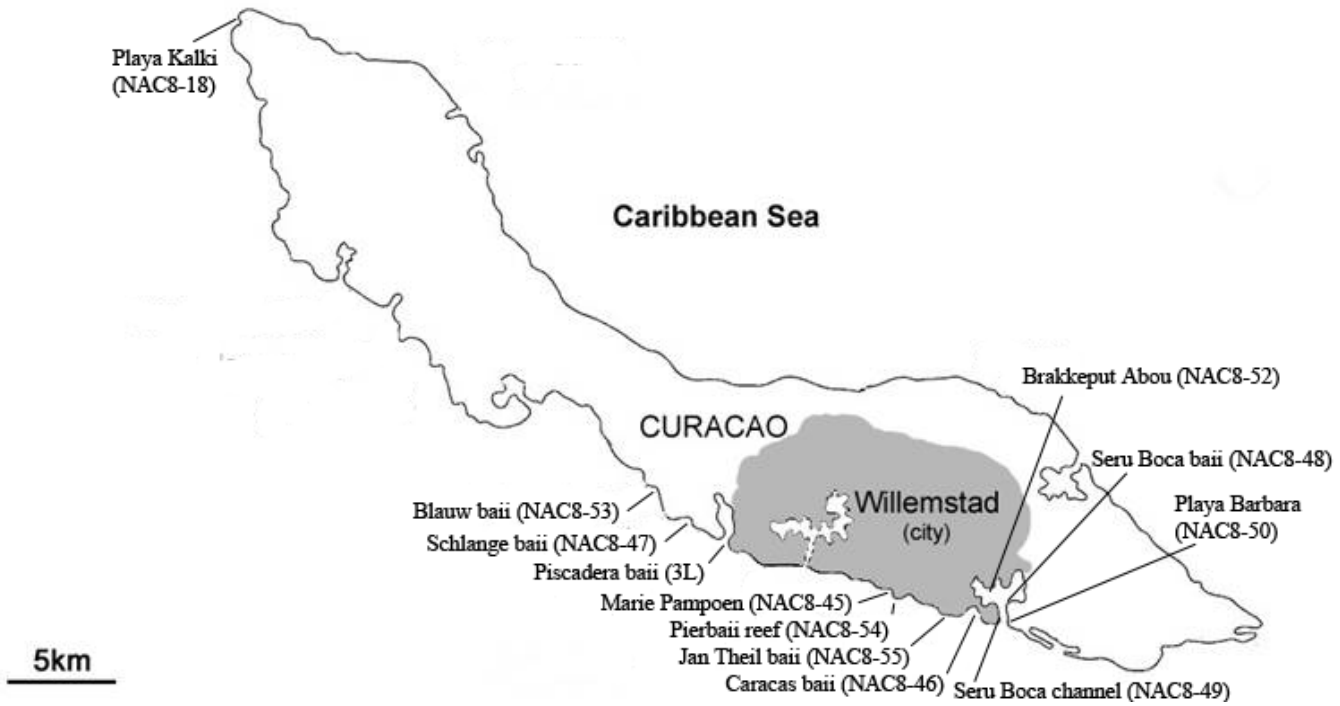


Figure V.1: Geographic map of Curaçao indicating the various collection sites for the current study. Environmental descriptions of each collection site are available in Table V.1.

V.3.2 Distribution of Expressed Secondary Metabolites

A combination of MALDI-TOF-MS of live filaments and LC ESI-MS of crude extracts was used to screen each specimen for known secondary metabolites (Table V.4; Figure V.2). Identities of proposed metabolites were confirmed by a combination of ESI-MS^2, LC retention times (RT) and compared against these features for secondary metabolites of reference strains PNG05-4 (tumonoic acid producing) and 3L (curacin/barbamide/carmabin producing) (Table V.4).

A total of six of the twelve "*Lyngbya"* strains (NAC8-47, NAC8-48, NAC8-49, NAC8-51, NAC8-52, and NAC8-53) contained a suite of bioactive secondary metabolites composed of the crustacean toxins curacin A, curazole, and curacin D, the molluscicidal agents barbamide and dechlorobarbamide, and the lipopeptides carmabins A-B. Additionally, four of the remaining six "*Lyngbya"* strains (NAC8-45, NAC8-46, NAC8-54 and NAC8-55) produced tumonoic acids A-C and F as well as the methyl and ethyl esters, methyl tumonoates A, B and ethyl tumonoate A. No secondary metabolites cataloged in the MarinLit (2010) database were detected from either of the remaining two strains (NAC8-18 and NAC8-50).

Table V.4: Characterization of Secondary Metabolites.

Secondary metabolite	MALDI-TOF-MS[1]	(+) ESI-MS[2]	RT[3]	ESI-MS/MS[4]
curacin A ($C_{23}H_{35}NOS$)	$[M+H]^+$ 374.2 (100), 375.2 (25), 376.2 (5)	$[M+H]^+$ 374.2 (100), 375.2 (25)	43.05	342.2 ($C_{22}H_{32}NS$),
curazole ($C_{23}H_{33}NOS$)	$[M+H]^+$ 372.2 (100), 373.2 (25)	$[M+H]^+$ 372.2 (100), 374.2 (25)	44.02	340.2 ($C_{22}H_{30}NS$)
curacin D ($C_{22}H_{33}NOS$)	ND	$[M+H]^+$ 360.2 (100), 361.2 (25)	40.46	260.4 ($C_{16}H_{22}NS$)
barbamide ($C_{20}H_{23}Cl_3N_2O_2S$)	$[M+H]^+$ 460.9, 462.9	$[M+H]^+$ 460.9 (100), 462.9 (100), 464.9 (35)	25.49	218.9 ($C_{12}H_{14}N_2S$)
dechlorobarbamide ($C_{20}H_{24}Cl_2N_2O_2S$)	ND	$[M+H]^+$ 426.9 (100), 428.9 (77), 430.9 (16)	20.91	218.8 ($C_{12}H_{14}N_2S$)
carmabin A ($C_{40}H_{57}N_5O_6$)	$[M+Na]^+$ 726.4 (100), 727.4 (40), 728.4 (7)	$[M+H]^+$ 704.2, $[M+Na]^+$ 726.5	39.64	341.2 ($C_{22}H_{31}NO_2$)
carmabin B ($C_{40}H_{59}N_5O_7$)	$[M+H]^+$ 722.4	$[M+H]^+$ 722.3, $[M+Na]^+$ 744.5	22.51	359.2 ($C_{22}H_{33}NO_3$)
tumonoic acid A ($C_{19}H_{33}NO_4$)	ND	$[M+H]^+$ 340.1, $[M+Na]^+$ 362.2	16.15	322.1 ($C_{19}H_{32}NO_3$)
tumonoic acid B ($C_{28}H_{47}NO_8$)	$[M+Na]^+$ 548.2 (100), 549.2 (20), 550.2 (5)	$[M+H]^+$ 526.2, $[M+Na]^+$ 548.3	32.93	508.2 ($C_{28}H_{46}NO_7$)

Table V.4: Continued

Secondary metabolite	MALDI-TOF-MS[1]	(+) ESI-MS[2]	RT[3]	ESI-MS/MS[4]
tumonoic acid C ($C_{27}H_{45}NO_8$)	$[M+Na]^+$ 534.2 (100), 535.2 (25), 536.2 (7)	$[M+H]^+$ 512.0, $[M+Na]^+$ 534.2	34.76	494.8 ($C_{27}H_{44}NO_7$)
tumonoic acid F ($C_{21}H_{38}NO_5$)	ND	$[M+H]^+$ 384.1	16.79	366.1 ($C_{21}H_{37}NO_4$)
methyl tumonoate A ($C_{20}H_{35}NO_4$)	ND	$[M+H]^+$ 353.9	21.61	336.1 ($C_{20}H_{34}NO_3$)
methyl tumonoate B ($C_{29}H_{49}NO_8$)	$[M+Na]^+$ 562.2 (100), 563.2 (25)	$[M+H]^+$ 540.1, $[M+Na]^+$ 562.3,	35.67	522.3 ($C_{29}H_{48}NO_7$)
ethyl tumonoate A ($C_{21}H_{35}NO_4$)	$[M + Na]^+$ 562.2	$[M+H]^+$ 368.0	24.84	350.1 ($C_{21}H_{34}NO_3$)

[1] *m/z* values of the secondary metabolites detected by MALDI-TOF-MS.
[2] *m/z* values of the secondary metabolites detected by ESI-MS in positive mode.
[3] Retention time (min) of detected molecules on LC-ESI-MS chromatogram using the conditions described in Experimental Procedures.
[4] Fragment ion of the secondary metabolites detected on LC-ESI-MS/MS.

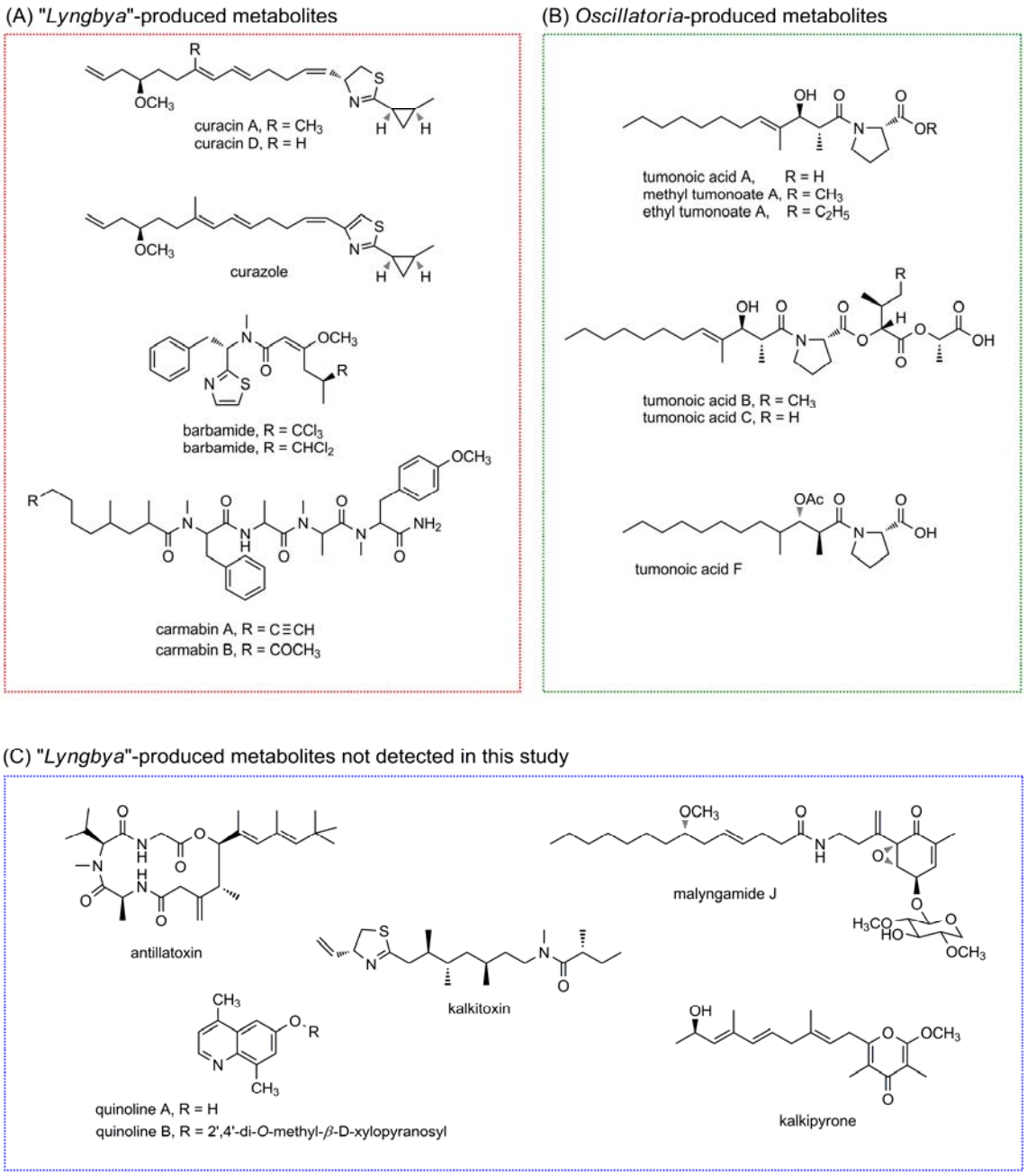


Figure V.2: Secondary metabolites attributed to the genus *"Lyngbya"* obtained from Curaçao. Box A shows "*Lyngbya"* metabolites detected in this study. Box B represents molecules previously attributed to "*Lyngbya"*, but herein shown to be produced by *Oscillatoria* spp. Box C represents secondary metabolites previously isolated from "*Lyngbya"* populations from Curaçao but which could not be detected in this study.

V.3.3 Genetic Capacity for Secondary Metabolite Production

All "*Lyngbya"* specimens were screened by PCR for genetic markers of specific secondary metabolite pathways in order to verify their biosynthetic capacity to produce these molecules (Table V.5). The following biosynthetic gene markers were analyzed from genomic-DNA: (i) *phy*H encoding the α-ketoglutarate dependent halogenase of the curacin A pathway, (ii) *bar*B1 and *bar*B2 encoding the BarB1 and BarB2 halogenases of the barbamide pathway, (iii) the tandem methyl transferases (*car*MT1-MT2) methylating the *N*,*O*–dimethyl tyrosine of carmabin A, and (iv) the thioesterase (*car*TE) of the carmabin A pathway (Chang *et al*., 2002; Chang *et al*., 2004). In one case, the DNA extracted from a single filament of "*L. majuscula"* NAC8-47 was amplified by multiple displacement amplification (MDA) prior to PCR-screening, due to the limited amount of biomass obtained from the field collection.

This PCR-screening and sequencing effort revealed that all six of the curacins/barbamide/carmabins-producing specimens contained the three pathways listed above (*phy*H [curacin A], *bar*B1 and *bar*B2 [barbamide], and *car*MT1-MT2 and *car*TE [carmabin A]), and that these were identical or nearly identical to sequences present in the reference strain 3L (p-distance = <99%). The MDA-DNA of "*L. majuscula"* NAC8-47 contained genes from the barbamide pathway and the carmabin A pathway, but not the curacin pathway. It should be noted that MDA typically amplifies only *ca*. 50-70% of a genome (Lasken, 2007). None of the remaining specimens contained any of these biosynthetic pathways.

Table V.5: Detection of Biosynthetic Pathways and Secondary Metabolites from "*Lyngbya*" Specimens.

	phyH (% Id[1])	**Cura-cin A**	**Cura-zole**	**Cura-Cin D**	**barB1 (% Id[1])**	**barB2 (% Id[1])**	**Barb-amide**	**Dechloro barbamide**	**carMT-MT (% Id[1])**	**carTE (% Id[1])**	**Carm-abin A**	**Carm-Abin B**
3L	100	++	+	+	100	100	++	+	100	100	+	-
NAC8-48	100	++	+	+	100	100	++	+	100	100	+	-
NAC8-49	100	++	+	+	99.5	99.8	++	-	99.9	99.9	+	+
NAC8-51	100	++	+	+	99.8	100	++	-	100	100	+	+
NAC8-52	100	+	+	+	100	100	++	-	99.8	99.9	+	+
NAC8-53	99.7	+	+	+	99.8	100	+	-	100	100	+	-
NAC8-47	ND**	+*	+*	NA*	99.7**	99.8**	+*	-*	100**	99.7**	+*	+*

	Tumonoic acid				Methyl tumonoate		Ethyl tumonoate
	A	**B**	**C**	**F**	**A**	**B**	**A**
NAC8-45	++	++	++	+	+	++	++
NAC8-46	++	++	+	+	+	++	++
NAC8-54	++	++	+	+	+	++	++
NAC8-55	++	++	++	+	+	++	++
NAC8-18	-	-	-	-	-	-	-
NAC8-50	-	-	-	-	-	-	-

Abbreviations: ND - not detected. NA - no data available (*i.e.* due to overlapping matrix ions).

[1] Percent sequence identity with the original biosynthetic pathways.

(++) main or major constituent, (+) minor constituent or trace compound estimated based on ESI abundance, (-) not detected.

(*) Secondary metabolites detection based only on MALDI-TOF-MS. (**) Genes screened using MDA genomes.

V.3.4 Phylogenetic Inference of "*Lyngbya*" strains

Phylogenetic inference of the 16S rRNA genes (1372 bp; ~95% of the gene coverage) revealed that the 12 specimens formed three distinct and distantly related lineages (Figure V.3). All five curacins/barbamide/carmabins producing specimens: NAC8-47, NAC8-48, NAC8-49, NAC8-51, NAC8-52, and NAC8-53 shared >99% sequence identity with other tropical marine "*Lyngbya"* species, including the previously described curacin/barbamide/carmabin producing strain 3L from Curaçao (Rossi *et al.*, 1997).

Despite nesting with other tropical marine "*Lyngbya"* specimens this lineage lacked a type-strain associated with it. The closest related genera are *Symploca* (type-strain = PCC 8002^{T}; mean p-distance = 6.1%) and *Coleofasciculus* (type-strain = PCC 7420^{T}; mean p-distance = 6.7%). More importantly, this tropical marine "*Lyngbya"* lineage is evolutionarily highly distant from other *Lyngbya* specimens, including the genus type-strain PCC 7419^{T} (mean p-distance = 9.4%).

The remaining six specimens (NAC8-18, NAC8-45, NAC8-46, NAC8-50, NAC8-54 and NAC8-55) claded with the morphologically similar genera *Oscillatoria* and *Trichodesmium*, including the type-strains *Oscillatoria sancta* PCC 7515^{T} and *Trichodesmium erythraeum* IMS 101^{T} (Figure V.3). The tumonoic acid-producing strains (NAC8-45, NAC8-46, NAC8-50, NAC8-54 and NAC8-55) formed a clade with the Pacific tumonoic acid-producing strain PNG05-4. The strain NAC8-18 was most closely related to the viridamide-producing 3L-Osc from Curaçao and the venturamide-producing PAB-21 *Oscillatoria* spp from the Caribbean coast of Panama. Interestingly, the strain NAC8-50 grouped with marine species of the planktic genus *Trichodesmium.*

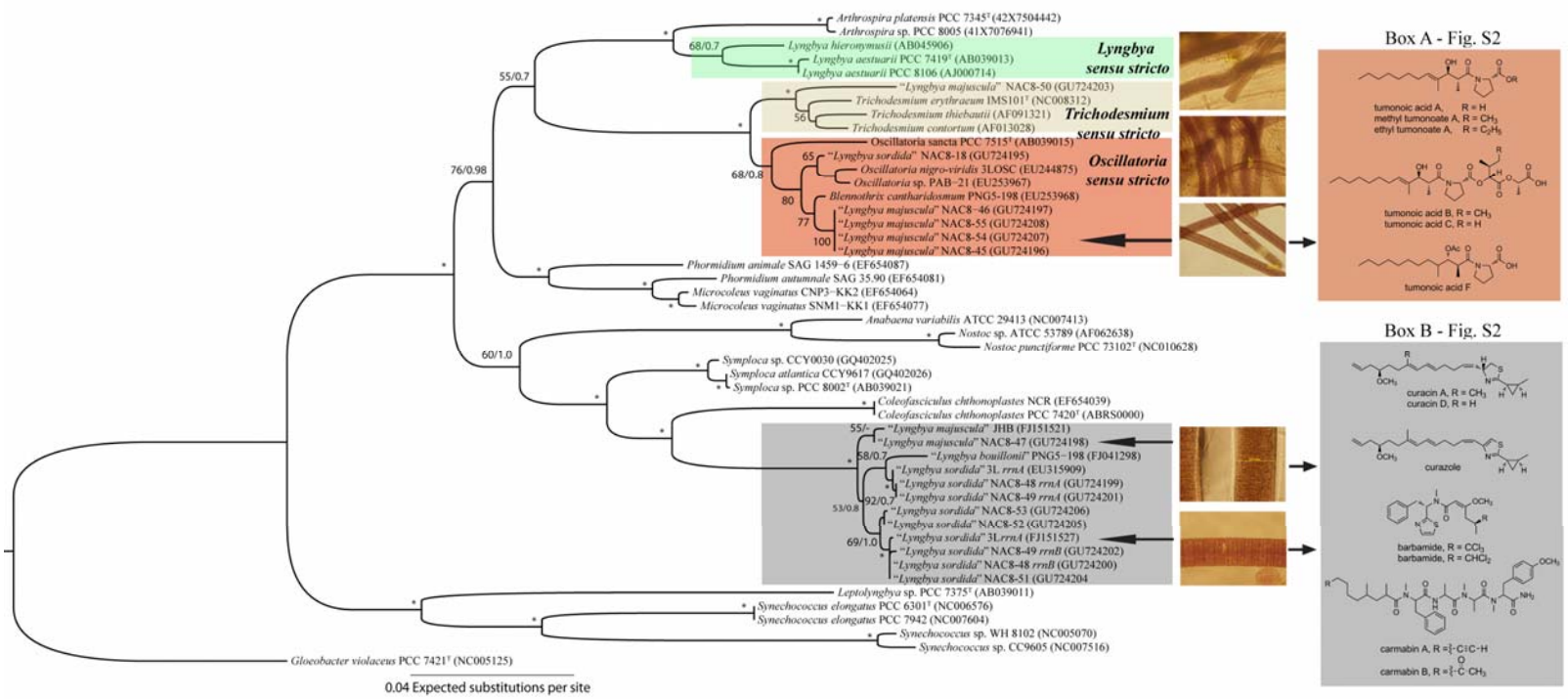


Figure V.3: Phylogenetic inferences for "*Lyngbya"* specimens from Curaçao based on the SSU (16S) rRNA genes. All specimens are indicated as species, strains and with GenBank accession numbers in brackets. Appropriate type-strains (^{T}) were obtained from *Bergey's Manual* for representative genera. Specimens corresponding with *Lyngbya* morphologically, but which are phylogenetically unrelated to the genus type-strain PCC 7419^{T} are designated as "*Lyngbya*". Lineages for the morphologically similar genera *Lyngbya*, *Trichodesmium*, and *Oscillatoria* are shown in green, brown, and red boxes, respectively. The "*Lyngbya*" lineage boxed in blue lacks a representative reference strain and needs to be considered a new generic entity. Microphotographs of the different phylotypes are indicated with arrows as well as the secondary metabolites they produce. Support at important nodes are indicated as bootstrap and posterior probability for the Maximum likelihood (GARLI)/Bayesian inference (MrBayes). Well supported nodes (>80% bootstrap and 0.9 posterior probability) are indicated with asterisks (*). The scale bar is equivalent to 0.04 substitutions per nucleotide position.

V.3.5 Cellular Origin of Secondary Metabolites and their Biosynthetic Genes

Scanning electron microscopy (SEM) of cultured filaments from the reference strain *"Lyngbya"* 3L revealed a rich diversity of heterotrophic bacteria on the exterior surface of the polysaccharide sheath (Figure V.4). Transmission electron microscopy (TEM) of filament sections showed that these bacteria were indeed restricted to these exterior surfaces and that the interior space and cells were free from associated bacteria (Figure V.4B). The enveloping sheath was removed from a filament of strain 3L, two single-cells were individually isolated by micro-manipulation and their genomes amplified by multiple displacement amplification (MDA). Multiple copies (>10) of the 16S rRNA gene were PCR-amplified using general bacterial primers and sequenced from each single-cell MDA-DNA to verify that the MDA reaction had only amplified cyanobacterial DNA and not that from any associated heterotrophic bacteria.

All ten 16S rRNA gene sequences from each single-cell MDA-DNA had 100% gene sequence identity with either of the two 16S rRNA gene sequences from *"Lyngbya"* 3L. Tropical marine *Lyngbya* genomes have been found to often contain two different ribosomal operons with variable 16S rRNA gene sequences possessing up to 1.1% divergence (Engene *et al.*, 2010). Thus, the two 16S rRNA gene copies found in 3L is assumed to belong to paralogous ribosomal operons (*i.e*. *rrn* A and *rrn* B).Furthermore, the DNA from one single-cell revealed the presence of the biosynthetic genes *phyH,* *bar*B1, *bar*B2, *car*MT1-MT2 and *TE* while the other single-cell MDA DNA contained only *bar*B1 and *bar*B2.

Intact cell MALDI-TOF-MS (ICM) was also performed on additional isolates of single cells from *"Lyngbya"* 3L, and revealed metabolites with molecular weights of *m/z* 374.2 $[M+H]^+$, *m/z* 460.9 $[M+H]^+$ and *m/z* 726.4 $[M+H]^+$, corresponding to curacin A, barbamide and carmabin A, respectively.

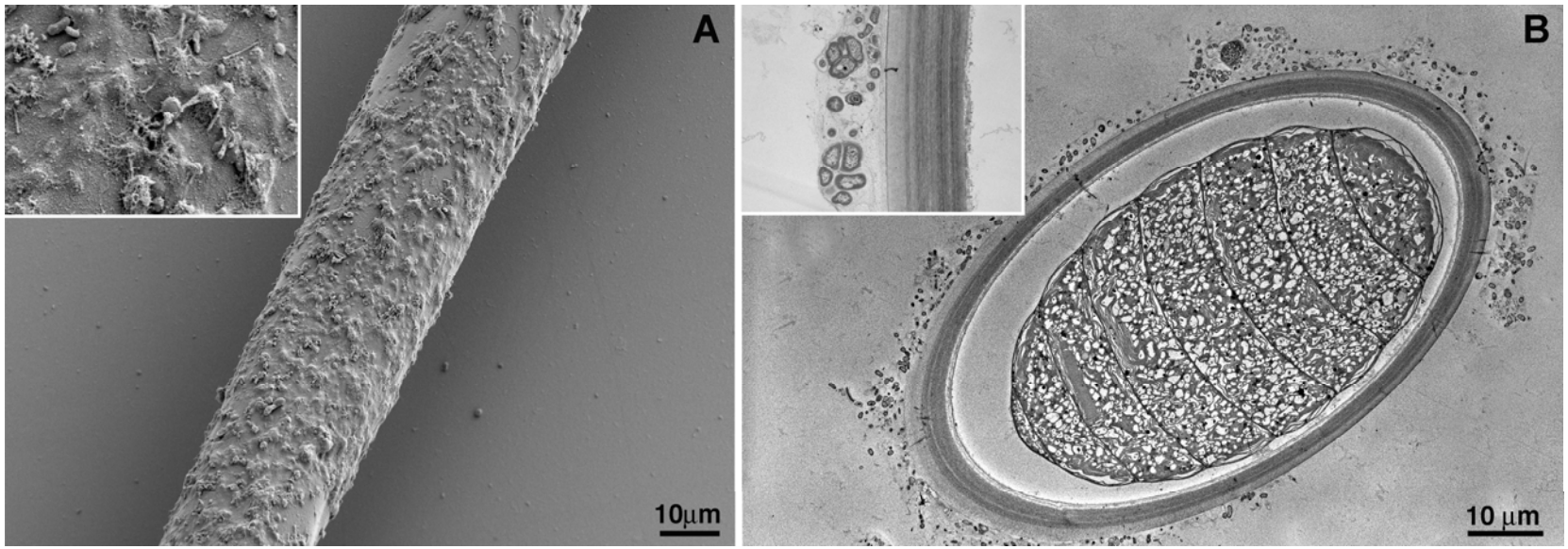


Figure V.4: (**A**) Scanning electron micrograph (SEM) of the surface of a cultured "*L. sordida*" 3L filament colonized by heterotrophic bacteria. (**B**) Transmission electron micrograph (TEM) of a cross section of cultured "*L. sordida*" 3L. Note that heterotrophic bacteria are only visibly present on the exterior surface of the filaments' polysaccharide sheaths.

V.4 Discussion

Cyanobacteria represent one of the most ancient and biologically diverse groups of organisms on earth (Komárek, 2005; Rasmussen *et al*., 2008). Thus, the extraordinary diversity of secondary metabolites is likely a reflection of this biodiversity. Yet, despite this diversity, over 90% of marine cyanobacterial-derived secondary metabolites have been isolated from a total of ten different genera (MarinLit, 2010). This imbalance in secondary metabolite distribution is best exemplified with the genus *Lyngbya*, which according to our current understanding, is responsible for the production of over 40% of all marine cyanobacterial secondary metabolites (Tidgewell *et al*., 2010). The extreme secondary metabolite diversity of marine *Lyngbya* was exemplified by the populations surveyed in this study from the island of Curaçao.

Although twelve different secondary metabolites were detected in various "*Lyngbya"* populations surveyed in this study, the majority of the molecules previously isolated from marine "*Lyngbya"* from Curaçao were not identified in any of the current specimens. Previously isolated molecules that were not observed in any of the current specimens included malyngamides H-L (Wu *et al.*, 1997), antillatoxin A (Orjala *et al.*, 1995), quinones A and B (Orjala & Gerwick, 1997), kalkipyrone (Graber & Gerwick, 1998) and kalkitoxin (Wu *et al.*, 2000). Thus, the true metabolic origin of the molecules that were not detected in this study can only be speculated upon. For example, a close structural resemblance has been noted between kalkipyrone and secondary metabolites of heterotrophic bacteria (Graber & Gerwick, 1998), and thus, their production from

ephemeral populations of associated heterotrophic bacteria may explain their variable isolation from environmental samples of cyanobacteria.

However, for those natural products detected in the current study, we were able to demonstrate both the presence of the biosynthetic pathways and the detection of secondary metabolites from single cells of "*Lyngbya"*, thereby firmly establishing that these are products of cyanobacterial genes and biochemical processes. Multiple displacement amplification (MDA) and sequencing analysis revealed that the curacin A, barbamide and carmabin A biosynthetic pathways were all present in the cyanobacterial genomes. This was matched by analysis of intact single-cells by MALDI-TOF-MS which demonstrated that these secondary metabolites were physically present in the cyanobacterial cells. To our knowledge, this is the first direct combined genetic and chemical proof of secondary metabolite biosynthesis on a microbial single-cell level.

Moreover, this correlation between biosynthetic capacity and biosynthetic expression was a prominent feature, despite sampling from diverse and variable environments. In addition to the conserved expression levels found in different environments, all secondary metabolites were produced in subsequent culture conditions (Table V.1). Thus, in this model, the surrounding environments appear to have little impact on the genetic expression of these secondary metabolites and consequently, in the distribution of secondary metabolites among different "*Lyngbya"* populations.

The initial classification of the various "*Lyngbya"* species in this study was based solely on traditional morphology-based criteria, as this has been the predominant foundation for identification of secondary metabolite producing cyanobacteria to date. However, the 16S rRNA phylogenetic analysis revealed that these specimens formed

distinct and evolutionarily distant lineages. A corresponding polyphyletic grouping has previously been described for the genus "*Lyngbya"* where these different "*Lyngbya"* lineages were assumed to represent species within the same genus (Sharp *et al.*, 2009). Because the 16S rRNA gene is relatively slowly evolving, an uncorrected genetic divergence of approximate ten percent corresponds to a relatively long period of evolution (Wilmotte & Herdman, 2001). In further support of the evolutionarily distance between these different "*Lyngbya"* lineages is the fact that heterocystous-forming cyanobacteria of the order Nostocales nests between the different lineages. This branching point represents an evolutionary event that has been estimated to have occurred between 2,450 and 2,100 mega-annums ago (Tomitani *et al*., 2006). Thus, these lineages are clearly evolutionarily distinct and need to be considered as different generic entities rather than species of the same genus. The fact that these unrelated lineages possess highly comparable morphological features likely results from convergent evolution, perhaps as a result of their occupying similar ecological niches.

An alternative hypothesis to homoplasy may simply be limitations in the number and degree of distinguishing characters that form the basis for our traditional morphology-based taxonomic systems. Cyanobacterial taxonomy has been predominantly founded upon temperate soil or freshwater specimens. Therefore, strains from recently explored environments, such as tropical marine, have typically been identified based on their morphological similarities with previously described taxa.

In this study, we show that none of the "*Lyngbya"* specimens were related to the genus type-strain. Instead, some of the specimens were related to the morphologically similar genera *Oscillatoria* and *Trichodesmium*, and should be grouped with these

taxonomic entities. These findings underscore the need to reevaluate the morphological characters differentiating these other genera from *Lyngbya*. By contrast, the curacin/barbamide/carmabins-producing "*Lyngbya*" specimens formed an evolutionarily distant lineage without any related type-strain. This lineage clearly represents a novel cyanobacterial group that has been positioned with the genus *Lyngbya* solely because of morphological similarities. However, this lineage is phylogenetically distinguished from the genus *Lyngbya* and needs to be described as a new generic entity.

The different evolutionarily paths of these lineages have resulted in different metabolic capacities as is shown in their produced secondary metabolites. By grouping morphologically similar but evolutionarily distant specimens together, these groups have become extensively overrepresented in their perceived chemical richness, and this explains the imbalance in secondary metabolite distribution ascribed above. In this study, we show that polyphyly is the major reason for the misconception that *Lyngbya* is such a secondary metabolite-rich group. Ongoing analyses of secondary metabolite-producing cyanobacteria corresponding with the "*Lyngbya"* morpho-type from other geographic regions support this hypothesis. Moreover, to our knowledge, no secondary metabolites have been isolated from specimens related to the genus type-strain.

In conclusion, we argue that knowledge of phylogenetic relationships is essential for cyanobacterial systematics as well as for developing a clear understanding regarding the distribution of secondary metabolites. In this chapter, we showed that "*Lyngbya*" is a polyphyletic group and that bioactive secondary metabolites attributed to "*Lyngbya*" are actually produced by morphologically similar but phylogenetically distant lineages. A corollary to this conclusion is that it is the morphological resemblance of different

cyanobacterial groups that has contributed to the perception that *Lyngbya* is so remarkably rich in secondary metabolites. Thus, taxonomic clarification and revision of polyphyletic cyanobacterial lineages is essential for developing an accurate understanding of the distribution of bioactive secondary metabolites, and can be used to direct targeted and more efficient natural products discovery programs in the future.

V.5 Acknowledgments

We gratefully acknowledge the government of Curaçao and the Carmabi research station for support and permits for this research project. We also thank M. Vermeij and J. K. Nunnery for aid with the collection as well as R. V. Grindberg for help selecting genetic markers for the biosynthetic pathways. We would also like to express appreciation for the guidance and advice provided by T. Deerinck with regard to electron microscopy studies. The research was generously supported by the Halliday award (SIO) and NIH CA CA100851 and Sea Grant SG-100-TECH-N. Facilities used at the National Center for Microscopy and Imaging Research were supported by NIH through grant number P41-RR004050 to M. H. Ellisman.

The text of chapter V, in full, is published material as it appears in: Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2011). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya. Environ Microbiol* 13: 1601-1610. The dissertation author was the primary author and directed and supervised the research, which forms the basis for this chapter.

V.6 References

Burja, A. M., Banaigs, B., Abou-Mansour, E., Burgess, J. G., Wright, P. C. (2001). Marine cyanobacteria-a prolific source of natural products. *Tetrahedron* 46: 9347-9377.

Capper, A., Cruz-Rivera, E., Paul, V. J., Tibbetts, I. R. (2006). Chemical deterrence of a marine cyanobacterium against sympatric and non-sympatric consumers. *Hydrobiologia* 553: 319-326.

Castenholz, R. W. (2001). Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 473-553. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Chang, Z., Flatt, P. M., Gerwick, W. H., Nguyen, V-A., Willis, C. L., Sherman, D. H. (2002). The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296: 235-247.

Chang. Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., Gerwick WH. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67: 1356-1367.

Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601.

Erhard, M., von Döhren, H., Jungblut, P. (1997). Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry. *Nat Biotechnol* 15: 906-909.

Gerwick, W. H., Proteau, P. J., Nagle, D. G., Hamel, E., Blokhin, A., Slate, D. L. (1994). Structure of curacin A, a novel antimitotic, antiproliferative and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. *J Org Chem* 59: 1243-1245.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Graber, M. A., Gerwick, W. H. (1998). Kalkipyrone, a toxic γ-pyrone from an assemblage of the marine cyanobacteria *Lyngbya majuscula* and *Tolypothrix* sp. *J Nat Prod* 61: 677-680.

Hanaichi, T., Sato, T., Hoshino, M., Mizuno, N. (1986). Proceedings XIth Int. Congress on Electron Microscopy, Kyoto.

Hoffman, L., Komárek, J., Kaštovskŷ, J. (2005). System of Cyanoprokaryotes. (Cyanobacteria) - State in 2004. *Arch Hydrobiol* 159: 95-115.

Kaebernick, M., Neilan, B. A., Borner, T., Dittmann, E. (2000). Light and the transcriptional response of the microcystin biosynthesis gene cluster. *Appl Environ Microbiol* 66: 3387-3392.

Komárek, J., Anagnostidis, K. (2005). In Süsswasserflora von Mitteleuropa 19/2. Elsevier/Spektrum: Heidelberg.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae* 21: 349-375.

Lasken, R.S. (2007). Single-cell genomic sequencing using multiple displacement amplification. *Curr Opin Microbiol* 10: 510-516.

Marine literature database (http://www.chem.canterbury.ac.nz/marinlit/marinlit.shtml)

Nübel, U., Garcia-Pichel, F., Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63: 3327-3332.

Orjala, J., Gerwick, W. H. (1994). Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula. J Nat Prod* 59: 427-430.

Orjala, J., Gerwick, W. H. (1997). Two quinoline alkaloids from the Caribbean cyanobacterium *Lyngbya majuscula*. *Phytochemistry* 45: 1087-1090.

Orjala, J., Nagle, D. G., Hsu, V. L., Gerwick, W. H. (1995). Antillatoxin: an exceptionally ichthyotoxic cyclic lipopeptide from the tropical cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 117: 8281-8282.

Rasmussen, B., Fletcher, I. R., Brocks, J. J., Kilburn, M. R. (2008). Reassessing the first appearance of eukaryotes and cyanobacteria. *Nat Lett* 455: 1101-1104.

Rossi, J. V., Roberts, M. A., Yoo, H-D., Gerwick, W. H. (1997). Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9: 195-204.

Sharp, K., Arthur, K. E., Gu, L., Ross, C., Harrison, G., Gunasekera, S. P., Meickle, T., Matthew, S., Luesch, H., Thacker, R. W., Sherman, D. H., Paul, V. J. (2009). Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* species (Cyanobacteria: Oscillatoriales) from reefs of southeastern Florida. *Appl Environ Microbiol* 75: 2879-2888.

Simmons, T. L., Coates, R. C., Clark, B. R., Engene, N., Gonzalez, D., Esquenazi, E., Dorrestein, P. C., Gerwick, W. H. (2007). Biosynthetic origin of natural products isolated from marine microorganism–invertebrate assemblages. *Proc Natl Acad Sci U S A* 105: 4587-4594.

Simmons, T. L., Engene, N., Urena, L. D., Romero, L. I., Ortega-Barrıa, E., Gerwick, G. L., Gerwick, W. H. (2008). Viridamides A and B, lipodepsipeptides with antiprotozoal activity from the marine cyanobacterium *Oscillatoria nigro*-v*iridis. J Nat Prod* 71: 1544-1550.

Shalev-Malul, G., Lieman-Hurwitz, J., Viner-Mozzini, Y., Sukenik, A., Gaathon, A., Lebendiker, M., Kaplan, A. (2008). An AbrB-like protein might be involved in the regulation of cylindrospermopsin production by *Aphanizomenon ovalisporum. Environ Microbiol* 10: 988-999.

Sorrels, C. M., Proteau, P. J., Gerwick, W. H. (2009). Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol* 75: 4861-4869.

Tan, L. T. (2007). Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochem* 68: 954-979.

Taori, K., Liu, Y., Paul, V. J., Luesch, H. (2009). Combinatorial strategies by marine cyanobacteria: symplostatin 4, an antimitotic natural dolastatin 10/15 hybrid that synergizes with the coproduced HDAC inhibitor largazol. *ChemBioChem* 10: 1634-1639.

Tidgewell, K., Clark, B. R., Gerwick, W. H. (2010). The natural products chemistry of cyanobacteria. Compr Nat Prod Chem. Pergamon Press: Oxford.

Thacker, R. W., Nagle, D. G., Paul, V. J. (1997). Effects of repeated exposures to marine cyanobacterial secondary metabolites on feeding by juvenile rabbitfish and parrotfish. *Mar Ecol Prog Ser* 147: 21-29.

Thacker, R. W., Paul, V. J. (2004). Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales). *Appl Environ Microbiol* 70: 3305-3312.

Tomitani, A., Knoll, A. K., Cavanaugh, C. M., Ohno, T. (2006). The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A* 103: 5442-5447.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J. A. M. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35: 71-74.

Wilmotte, A., Herdman, M. (2001). Phylogenetic Relationships Among Cyanobacteria Based on 16S rRNA Sequences. In *Bergey's Manual of Systematic Bacteriology*. 1:487-493. Edited by Boone, D. R., Castenholz, R. W., Garrity, G. M. New York, USA: Springer.

Wu, M., Milligan, K. E., Gerwick, W. H. (1997). Three new malyngamides from the marine cyanobacterium *Lyngbya majuscula*. *Tetrahedron* 53: 15983-15990.

Wu, M., Okino, T., Nogle, L. M., Marquez, B. L., Williamson, R. T., Gerwick, W. H. *et al.,* (2000). Structure, synthesis, and biological properties of kalkitoxin, a novel neurotoxin from the marine cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 122: 12041-12042.

Chapter VI

Moorea producta gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., Tropical Marine Cyanobacteria Rich in Bioactive Secondary Metabolites

Abstract

The filamentous cyanobacterial genus *Moorea* gen. nov. is a cosmopolitan pan-tropical group abundant in the marine benthos. *Moorea* is photosynthetic (containing phycocyanin, phycoerythrin, allophycocyanin, and chlorophyll *a*), but non-diazotrophic (lacks heterocysts and nitrogenase reductase genes). The cells (discoid and 25-80 μm wide) are arranged in long filaments (<10 cm in length) and often form extensive mats or blooms in shallow-water. The cells are surrounded by thick polysaccharide sheaths covered by a rich diversity of heterotrophic microorganisms. A distinctive character of this genus is its extraordinarily rich production of bioactive secondary metabolites. This is matched by genomes rich in polyketide synthase (PKS) and non-ribosomal protein synthetase (NRPS) biosynthetic genes which are dedicated to secondary metabolism. The encoded natural products are sometimes responsible for harmful algae blooms (HABs) and, due to morphological resemblance with the genus *Lyngbya*, this group has often been incorrectly cited in the literature. The type species for *Moorea producta* gen. nov., sp. nov. is 3L^{T} and *Moorea bouillonii* comb. nov. is PNG05-198^{T}.

VI.1 Introduction

Benthic filamentous marine cyanobacteria from the tropics have been of increasing biomedical interest due to their extraordinary richness in bioactive secondary metabolites (Tidgewell *et al*., 2010). Many of these natural product (NP) molecules are potent toxins responsible for harmful algal blooms (HABs), and thus are hazardous to humans as well as near shore environments. At the same time, some of these cyanobacterial toxins and other natural products have properties of potential benefit to human health as pharmaceutical leads (Golubic *et al*., 2010). Surprisingly, a majority of these unique NPs have been ascribed as being produced by a single genus, *Lyngbya*, and a preponderance of these come from a single species, *L. majuscula* (Liu & Rein, 2010). However, an unfortunate consequence of using traditional morphology-based taxonomic systems in these identifications has been that cyanobacteria of many recently explored biological frontiers (*e.g.* tropical marine environments) have been forced into existing morphological groupings, and thus, the true biodiversity of this group has been greatly underestimated (Casamatta *et al*. 2005; Engene *et al*., 2011).

The proposed cyanobacterial genus *Moorea* gen. nov. is a cosmopolitan, pan-tropical group abundant in the marine benthos. Strains of *Moorea* gen. nov. have often been incorrectly classified as the cyanobacterial genus *Lyngbya* due to morphological similarities between the two groups (Engene *et al*., 2011). This misidentification of *Moorea* as *Lyngbya* has been a source of considerable taxonomic confusion as well as the major reason for the perceived chemical richness of the genus *Lyngbya* (Engene *et al*., 2011). Herein, we firmly differentiate between these two phylogenetically distinct groups

and describe *Moorea* as a novel generic entity (gen. nov.). This description and naming of *Moorea* gen. nov. was performed under the provisions of the International Code of Botanical Nomenclature. However, because cyanobacteria are prokaryotes, the Bacteriological Code is important in these definitions, and thus we have considered and accommodated this latter code to the highest degree possible in this genus revision.

VI.2 Materials and Methods

VI.2.1 Sampling and Culturing

A total of 51 geographically distributed populations of the genus *Moorea* were included in this taxonomic revision (geographic data for *Moorea* specimens is available as Supplementary Table S1 in IJSEM Online). Field collections of cyanobacteria were carefully rinsed with autoclaved SWBG-11-media and visible macroorganisms mechanically removed with sterile tweezers under an Olympus VMZ dissecting microscope. Clonal, non-axenic strains were derived from phototactically isolated single-filaments on (0.5 %) agar plates with SWBG-11 and cultured in SWBG-11 medium at 28 °C with 33 g l^{-1} Instant Ocean salt (Aquarium Systems). The cultures were kept at a light intensity of 7 µmol photon s^{-1} m^{-2} (light: dark = 16: 8 h). Two isolated strains, *M. producta* $3L^{T}$ and *M. bouillonii* PNG5-198^{T}, were deposited to the Canadian Phycological Culture Centre (CPCC) and the national marine phytoplankton collection (CCMP) as reference strains. Additionally, the *Lyngbya* reference strain PCC 7419 was obtained from the Pasteur Culture Collection (PCC) for biological comparison.

VI.2.2 Microscopy and Characterization

Light microscopy was performed using an Olympus IX51 epifluorescent microscope (1000X) equipped with an Olympus U-CMAD3 camera. Samples for scanning electron microscopy (SEM) were placed on indium-tin-oxide glass slides that had been coated with 0.1% polyethylenimine to facilitate adhesion. The samples were then fixed in 2.5% glutaraldehyde buffered in 1X PBS for 30 min and a secondary fix of 2% aqueous osmium tetroxide for 15 min. Dehydration was achieved with a graded EtOH series. The samples were then critical point dried and sputter coated with gold palladium. A Hitachi SU6600 SEM was used to view the samples.

Samples for transmission electron microscopy (TEM) were prepared using high pressure freezing and subsequent freeze substitution (HPF-FS). The filaments were cut into < 0.5 mm pieces and placed into specimen holders with a drop of cryoprotectant hexadecane. The samples were frozen using a Bal-Tec HPM 010 high pressure freezing machine. Freeze substitution was done using a Leica EM AFS machine. Samples were transferred to pre-cooled 1% glutaraldehyde with 0.2% tannic aced in anhydrous acetone and left at -90 °C for 36 h, then washed with acetone three times for 15 min each and subsequently transferred to 1% osmium textroxide with 0.1% uranyl acetate in acetone and held for 24 h. The temperature was raised to -60 °C, -30 °C, and 0 °C - held for 24 h between each step. At 0 °C the samples were washed with acetone three times for 15 min each and transferred into 50% Durcopan in acetone and held for 12 h. Once samples warmed to room temperature they were embedded in Durcopan and left to polymerize for 48 h. Thin sections (70 nm) were obtained using a Reichart Ultracut E and then placed on formvar coated 75 and 200 mesh copper grids. The grids were subsequently stained with

uranyl acetate and Sato lead. A JEOL 1200FX TEM was used to view the samples. Photosynthetic pigments were characterized as previously described (Tandeau de Marsac & Houmard, 1988).

VI.2.3 DNA-isolation, PCR-amplification and Cloning

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's specifications. DNA concentration and purity was measured on a DU® 800 spectrophotometer (Beckman Coulter). The PCR reaction volumes were 25 µl containing 0.5 µl (~50 ng) of DNA, 2.5 µl of 10 x PfuUltra IV reaction buffer, 0.5 µl (25 mM) of dNTP mix, 0.5 µl of each primer (10 µM), 0.5 µl of PfuUltra IV fusion HS DNA polymerase and 20.5 µl dH_2O. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 2 min at 95 °C, 25 cycles of amplification: 20 sec at 95 °C, 20 sec at 50 °C and 1.5 min at 72 °C, and final elongation for 3 min at 72 °C. PCR products were purified using a MinElute® PCR Purification Kit (Qiagen) before subcloning into the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen). Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced bidirectionally with M13 vector-primers as well as internal primers. The gene sequences are available in the DDBJ/EMBL/GenBank databases (see Supplementary Table S1, available in IJSEM Online).

VI.2.4 Phylogenetic Inference

The SSU (16S) rRNA genes of all 51 *Mooreae* specimens were included in the analysis. Representative reference strains were selected from *Bergey's Manual*

(Castenholz, 2001). The unicellular *Gloeobacter violaceus* PCC 7421 (GenBank acc. Nr. NC005125) was included as an evolutionary distant out-group. All gene sequences were aligned using the L-INS-I algorithm in MAFFT 6.717 (Katoh & Toh, 2008). The alignment was visually compared and refined using the SSU secondary structures model of *Escherichia coli* J01695 (Cannone *et al*., 2002) without data exclusion. The multiple sequence alignments are available in the TreeBASE database (http://www.treebase.org) under the submission ID 11599. Pair-wise sequence divergences were calculated in PAUP* 4.0b10. Appropriate nucleotide substitution models were compared and selected using uncorrected/corrected Akaike Information Criterion (AIC/AICc), Bayesian Information Criterion (BIC), and the Decision-theoretic (DT) in jModeltest 0.1.1 (Posada, 2008). The Maximum likelihood (ML) inference was performed using GARLI 1.0 (Zwickl, 2006). The analysis was run using the GTR+I+G model assuming a heterogeneous substitution rates and gamma substitution of variable sites (proportion of invariable sites (pINV) = 0.450, shape parameter (α) = 0.449, number of rate categories = 4). Bootstrap resampling was performed on 1000 replicates. Bayesian analysis was conducted using MrBayes 3.1 (Ronquist & Huelsenbeck, 2003). Four Metropolis-coupled MCMC chains (one cold and three heated) were run for 10,000,000 generations. MCMC convergence was determined using AWTY and the first 1,000,000 generations (10 %) were discarded as burn-in and the following data sets were sampled with a frequency of every 1000 generations. The maximum parsimony (MP) analysis was performed in PAUP* 4.0b10 using a heuristic search through the branch-swapping tree-bisection-reconnection (TBR) algorithm with the addition of 10,000 random replicates to find the most parsimonious tree. Bootstrap support was obtained from 1000 replicates.

VI.2.5 Genomic Analysis

Draft genomes from the *M. producta* strain 3L^T (GenBank acc. Nr. AEPQ01000000) and the *M. bouillonii* strain PNG05-198^T have recently been obtained and were used for phylogenomic and functional genomics comparison. Phylogenomic inference was performed bioinformatically on the basis of the *DNA-G*, *FRR, rpsB*, *NusA*, *PGK*, *PyrG*, *rpoB*, *rpsC*, *rpl2*, *rpl3*, *rpl4*, and *TSF* genes. These gene sequences were downloaded from all 59 publically available cyanobacterial genomes and concatenated for phylogenetic inference with the two *Moorea* genomes. Maximum likelihood (RaxML) inference was performed on the WAG+I+G model assuming heterogeneous substitution rates and gamma substitution of variable sites [proportion of invariable sites (pINV) = 0.265, shape parameter (α) = 0868, number of rate categories = 4] with a bootstrap re-sampling of 500 replicates.

VI.3 Results and Discussion

VI.3.1 Molecular-phylogenetic Analysis

Phylogenetic inference based on the SSU (16S) rRNA gene revealed that the *Moorea* lineage was evolutionary distinct and distant from the *Lyngbya sensu stricto* (reference strain = PCC 7419; p-distance = 9.2%) (Fig. 1(a). The *Moorea* lineage was nested between the closest related genera *Symploca* (reference strain = PCC 8002; p-distance = 6.1%) and *Coleofasciculus* (reference strain = PCC 7420; p-distance = 6.9%).

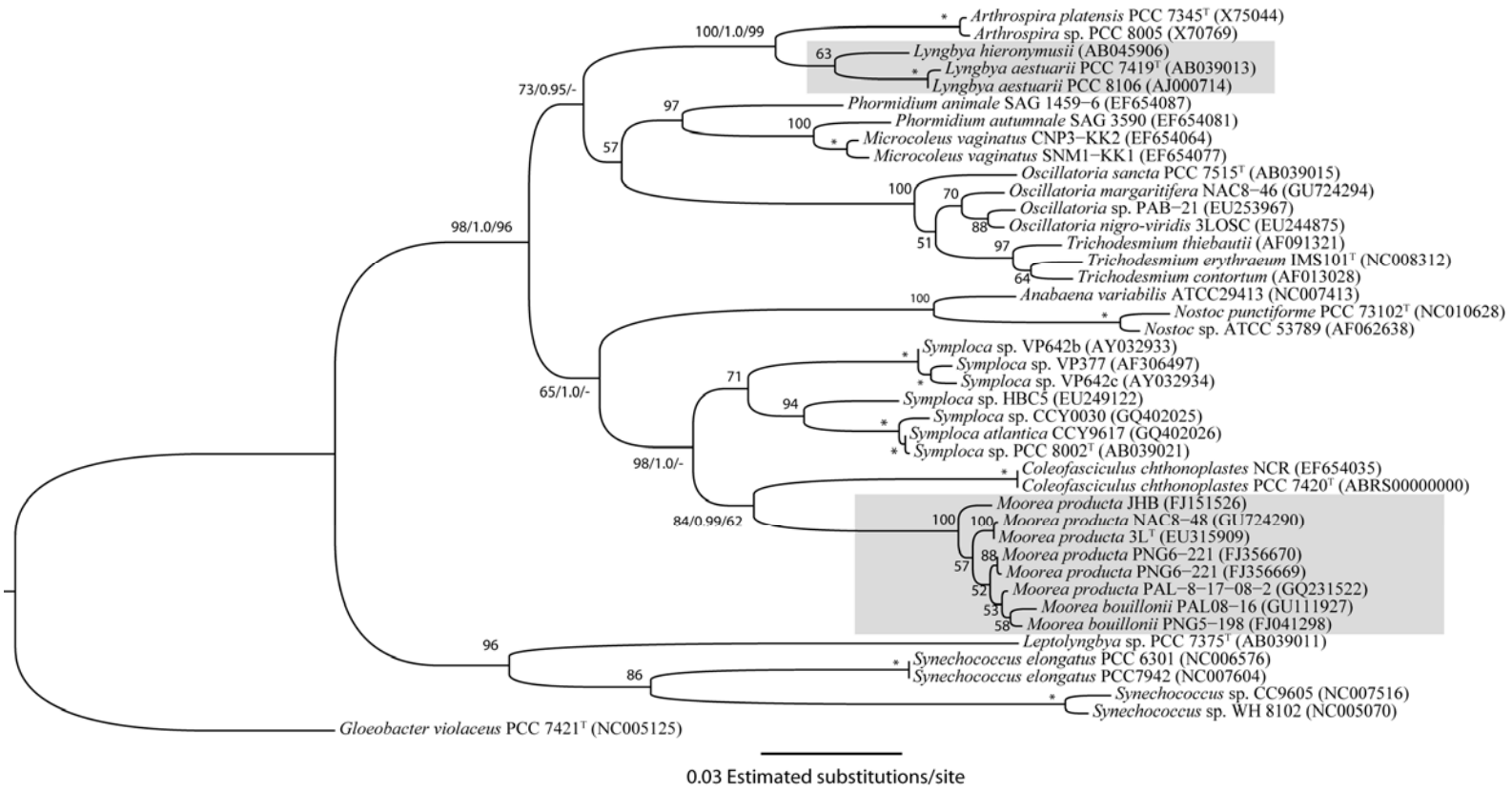


Figure VI.1: Phylogenetic inferences (GARLI) of *Lyngbya* and *Moorea* diversification based on SSU (16S) rRNA nucleotide sequences. The specimens are indicated as species, strain, and accession numbers in brackets. Type-strains were obtained from *Bergey's Manual.* The support values are indicated as boot-strap and posterior probability for the maximum likelihood, Bayesian inference, and maximum parsimony. The scale bar is indicated at 0.04 expected nucleotide substitutions per site using the GTR+I+G substitution model.

The phylogenetic positions of *Moorea* and evolutionary distances with *Lyngbya* were corroborated by analysis of the RNA polymerase gamma subunit (*rpoC1*) gene (evolutionarily tree of the *rpoC1* gene is available as Supplementary Fig. S1). An additional 12 other protein-coding genes (*DNA-G*, *FRR, rpsB*, *NusA*, *PGK*, *PyrG*, *rpoB*, *rpsC*, *rpl2*, *rpl3*, *rpl4*, and *TSF*) were selected from the *Moorea* genome drafts and the evolutionary histories of these genes were individually constructed and compared with all available sequenced genomes of cyanobacteria. As a result, each protein-coding gene showed an evolutionary history relatively congruent to that of the 16S rRNA gene phylogram. All 12 phylogenetically informative genes were concatenated for a more robust phylogenomic inference. The combined genes supported the phylogenetic distance between *Lyngbya* (*i.e.* PCC 8106) and *Moorea* (*i.e.* $3L^T$ and PNG5-198^T) as well as the overall evolutionary history of the phylum (a phylogenomic inference is available as Supplementary Fig. S2).

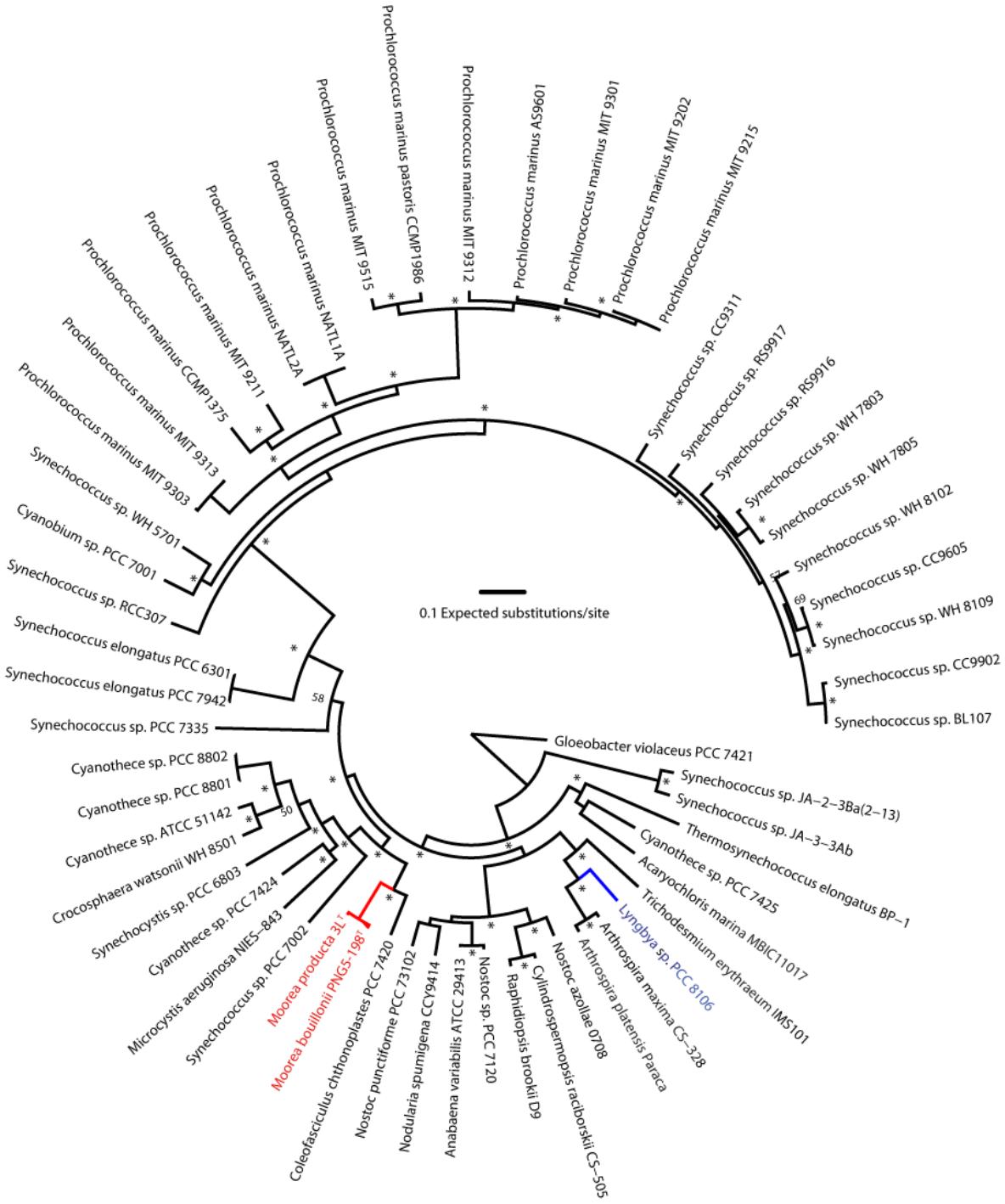


Figure VI.2: Phylogenomic inferences of 57 cyanobacterial genomes using the maximum likelihood method (RaxML). The analysis is based on the concatenation of twelve protein coding genes (>200 amino acids long).

On a sub-generic level, the *Moorea* specimens formed a tight clade with low interior sequence divergence (p-distance: mean = 0.5 %; max. = 1.4 %) (Fig. 1(b). This high DNA bar-coding gap of the *Moorea* clade of more than 12 times further supports the exclusivity of this clade and the need to distinguish it from neighboring genera [Fig. 1(a)]. However, the genomes of *Moorea* specimens contain multiple and variable copies of their 16S rRNA genes (Engene *et al*., 2010), and this relatively high level of intra-genomic gene heterogeneity in combination with the low sub-generic sequence divergence makes the 16S rRNA gene inadequate for speciation. The lack of phylogenetic resolution for species delineation was further indicated by low statistical node support at the terminal nodes and incongruence using different phylogenetic methods [Fig. 1(b)].

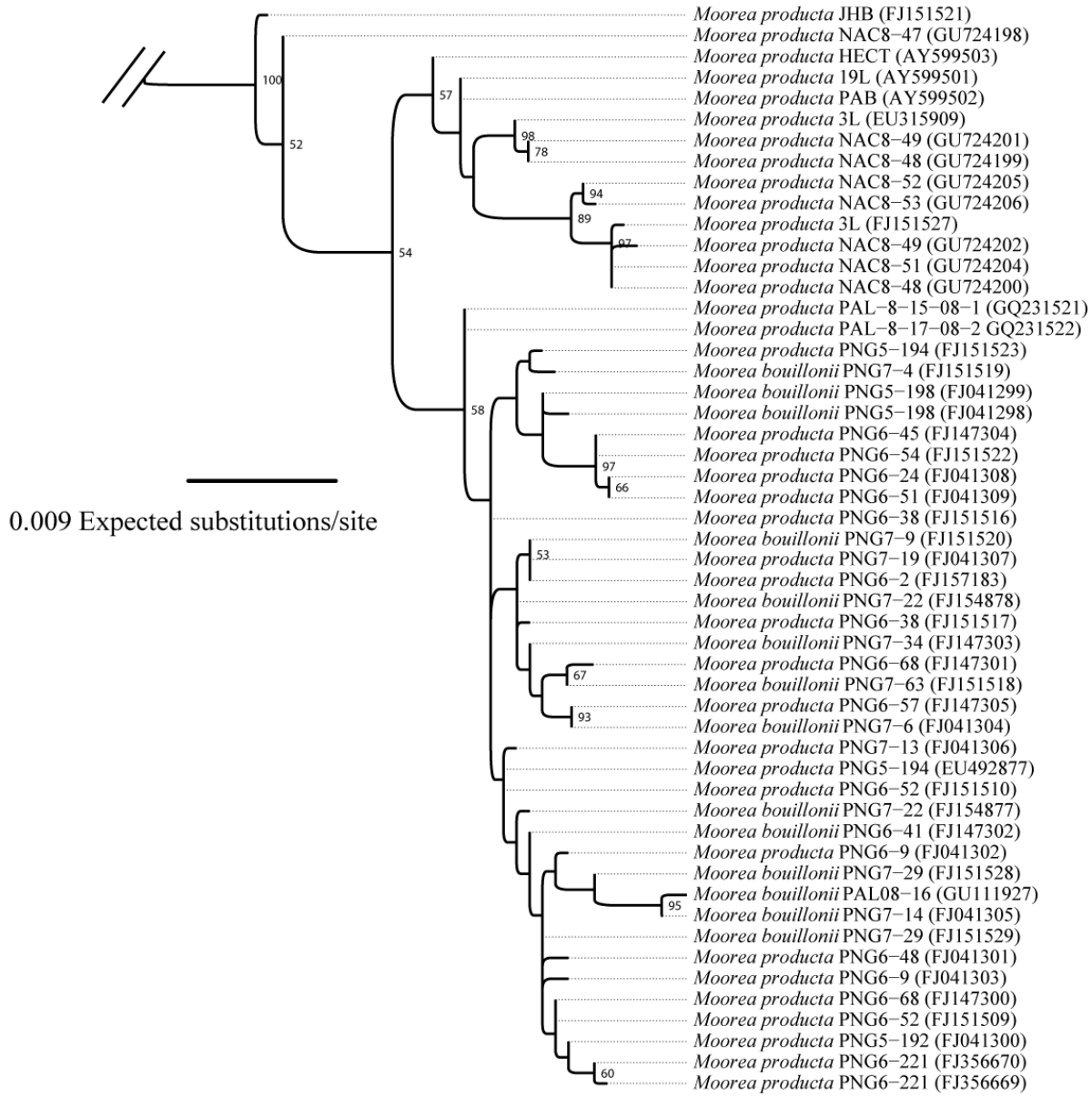


Figure VI.3: Phylogenomic inferences of the *Moorea* lineage

The less conserved internal transcribed spacer (ITS) region linking the 16S and 23S ribosomal gene has been proposed to be taxonomically more informative on a sub-generic level, and has often been used for species delineation in cyanobacteria (Otsuka *et al*., 1999; Boyer *et al*., 2001; Gugger *et al*., 2004; Siegesmund *et al*., 2008). Primer-sites on the adjacent 16S- and 23S rRNA genes were used to PCR-amplify the 16S-23S ITS-regions of 41 *Moorea* specimens. However, the 16S-23S ITS-regions were, in accordance with the 16S rRNA genes, present in multiple and variable gene copies, and thus, this gene region was also not able to definitively distinguish between *Moorea* specimens (Supplementary Fig. S3). In addition, the intra-genomic sequence heterogeneity of the 16S-23S ITS-region was found to influence structurally informative domains, such as the D1-D1' helix and the Box-B, secondary structures which are frequently used for taxonomic delineation (Boyer *et al*., 2001). In the case of *Moorea*, we argue that the 16S-23S ITS-regions are not able to further resolve species delineation.

Figure VI.4: Maximum-likelihood (PhyML) phylogenetic analyses of *Lyngbya* and *Moorea* diversification based on 16S-23S Internal transcribed spacer (ITS) regions.

VI.3.2 Morphological Characterization

Morphologically, the *Moorea* specimens were composed of long isopolar filaments enclosed in thick exopolysaccharide sheaths with discoid cells arranged in trichomes (Fig. 2). The exteriors of the sheaths were consistently covered by a rich fauna of heterotrophic bacteria and other microorganisms (Fig. 2). The two *Moorea* species *M. producta* and *M. bouillonii* had distinctively different colony morphologies. *M. producta* has often been reported in Literature as either *Lyngbya majuscula* or *L. sordida*. The primary reason for combining tropical marine *L. majuscula* and *L. sordida* into a single species, *Moorea producta*, was variability in morphological characters. *M. bouillonii* was, in contrast to *L. majuscula* and *L. sordida*, initially described from tropical marine environments and will consequently keep its species nomenclature in order to preserve taxonomic stability (Hoffman *et al*., 1994). Environmental specimens of *M. bouillonii* always formed characteristic reddish cobweb-like mats firmly attached to surrounding substrate and each colony was also always found with an associated snapping shrimp cf. *Alpheus frontalis* (Fig. 2).

The ultrastructure of *Moorea* cells contained a high degree of compartmentalization and were rich in intrathylacoidal spaces (widened thylacoids) (Fig. 3). The thylakoid membranes were arranged parallel to the cell walls. The filaments were surrounded by thick (2-3 μm) firm and laminated sheaths.

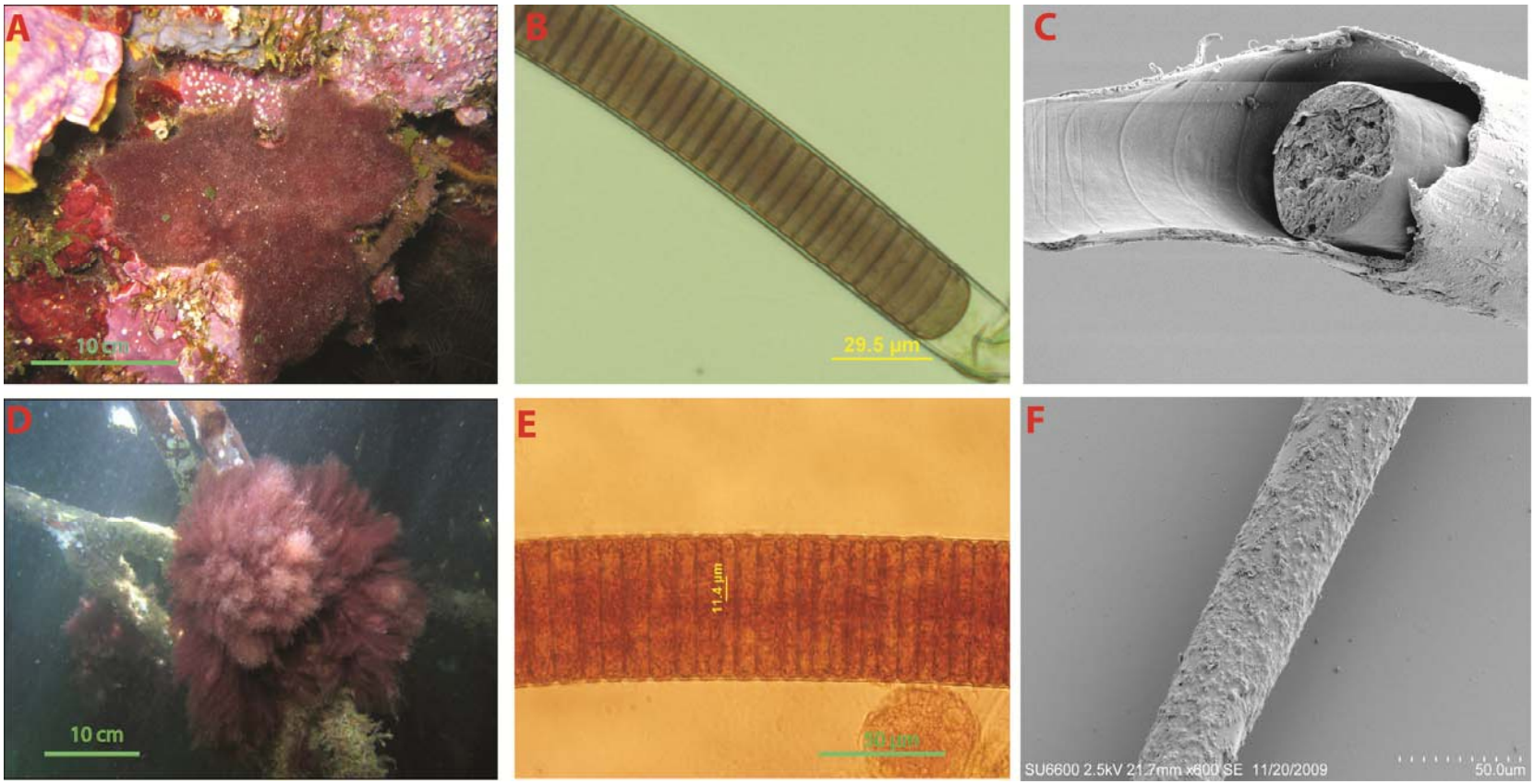


Figure VI.5: Morphological characterization of *Moorea* gen. nov. **(A-C)** *M. bouillonii* PNG05-198^{T} and **(D-E)** *M. producta* 3L^{T}. **(A)** Underwater pictures of *M. bouillonii* at 10 m depth forming a characteristic cob-web mat firmly attached to the surrounding corals. **(D)** Tuft colony morphology of *M. producta* growing on shallow-water mangrove roots. The microscopic pictures are adjusted to fit the frame (see scale bars) and the scales for the underwater pictures are approximated as shown. Microphotographs of cyanobacterial filaments (400X) of **(B)** *M. bouillonii* and **(E)** *M. producta* and scanning electron microscopy (SEM) **(C)** *M. bouillonii* and **(F)** *M. producta.*

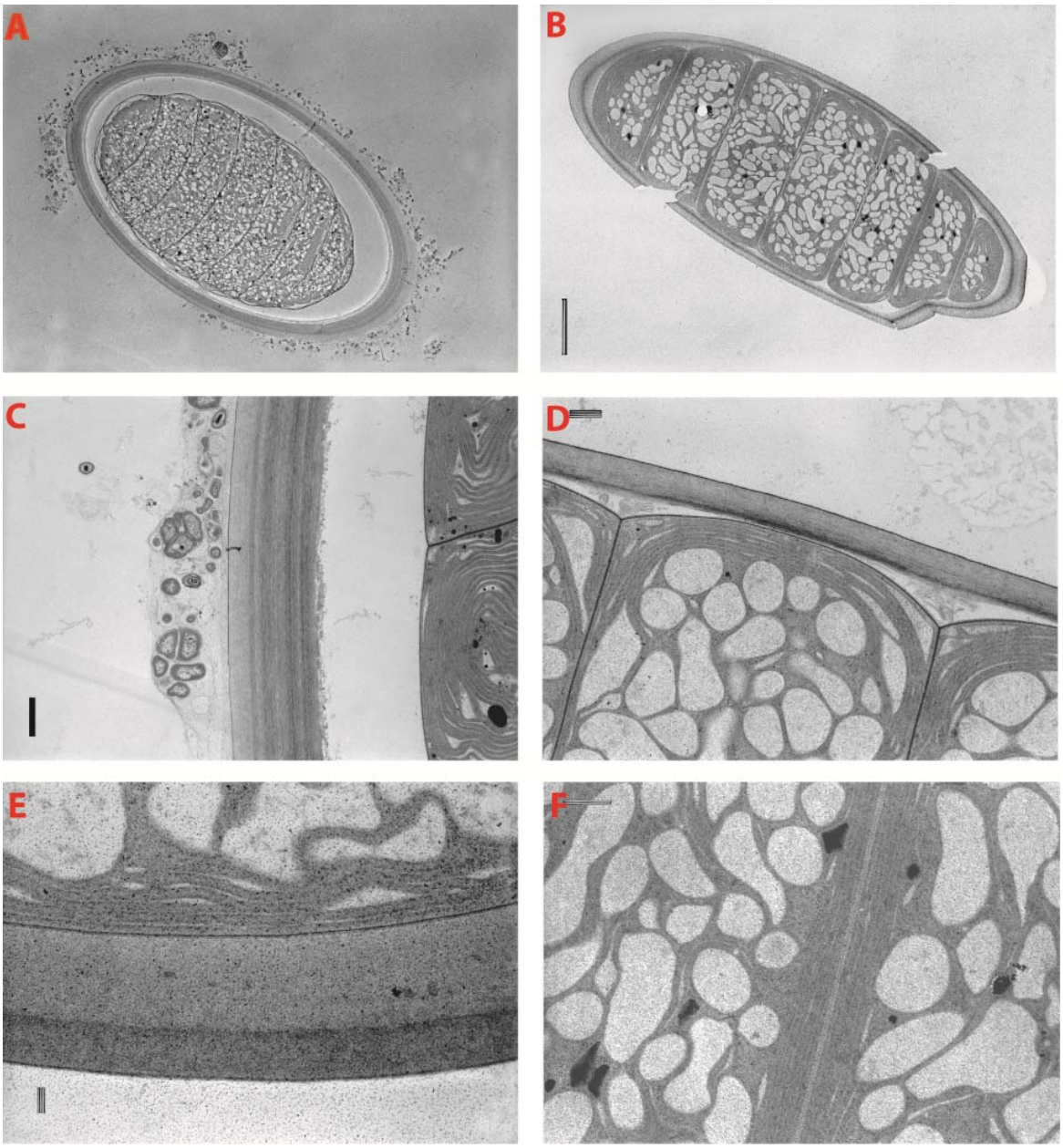


Figure VI.6: Microphotographs of cyanobacterial filaments obtained by transmission electron microscopy (TEM). Filament transections of: **(A)** *Moorea producta* 3L^{T} and **(B)** *M. producta* JHB: polysaccharide sheaths and thylakoid arrangements in: **(C)** *M. producta* 3L^{T} with heterotrophic bacteria on the exterior; **(D)** *M. producta* JHB*;* **(E)** polysaccharide sheath of *M. producta* 3L^{T}; **(F)** thylakoid arrangements in adjacent cells in *M. producta* JHB.

VI.3.3 Biogeography

Geographically, *Moorea* is a widely distributed group abundant in tropical marine regions (see Supplementary Table S1, available in IJSEM Online). The latitudinal distribution of this group, according to current sampling and records, ranges approximately between the Tropic of Cancer and the Tropic of Capricorn. The most northern reported collection of *Moorea* is Florida (26º 04'N) just north of the Tropic of Cancer (Sharp *et al*., 2008). The habitats of *Moorea* include diverse shallow-water marine environments such as coral reefs, sandy beaches and mangroves. While *M. producta* is a cosmopolitan species and has been found pantropically in shallow marine waters, *M. bouillonii* has only been reported from tropical Pacific locations (see Supplementary Table S1).

VI.3.4 Biochemical Analysis

Biochemically, extracts from the three *Moorea* strains 3L^T, PNG5-198^T, and JHB showed UV absorptions at 565, 620, 650, and 665 nm, corresponding to the photosynthetic pigments phycocyanin, phycoerythrin, allophycocyanin and chlorophyll *a*, respectively (Table 1). In addition to these basic cyanobacterial photosynthetic pigments, all three *Moorea* strains contained at least two structurally unique bioactive secondary metabolites, as characterized by LCMS and NMR (Table 1).

VI.3.5 Comparative Genomics

The DNA G+C content of *M. producta* 3L^T and *M. bouillonii* PNG05-198^T were 41.0 mol % and 42.3 mol %, respectively (Table 1), which were comparable to other

filamentous cyanobacteria (mean DNA G + C content = 41.2 mol %). The genome size of *M. producta* 3L^T (8.5 Mbp) was larger than the average genome of filamentous cyanobacteria (6.1 Mbp) and the second largest after the evolutionarily related *Coleofasciculus* PCC 7420 (genome size = 8.7 Mbp). The relatively large genome of *M. producta* 3L^T was reflected in a high abundance of protein-coding genes (7,415 compared with the average copy number of protein-coding genes in filamentous cyanobacteria of 5,468 copies). A potential reason for the large genome is the relatively large number of genes involved in the biosynthesis of bioactive secondary metabolites. For example, genome analysis of the strain 3L^T has revealed that approximately three percent of its genome contains polyketide synthase (PKS), non-ribosomal protein syntetase (NRPS), or other biosynthetic genes dedicated to secondary metabolism (Jones & Monroe *et al*., 2011). The partial genome of *M. bouillonii* PNG05-198^T also contained multiple copies of PKS and NRPS genes with high identity to biosynthetic genes involved in secondary metabolite production. The genome of *M. producta* 3L^T has been shown to lack genes involved in nitrogen fixation (Jones & Monroe *et al*., 2011). This was further supported by BLAST-search (E-value = 1e-5) of the *M. bouillonii* PNG05-198^T genome, which also lacked the *nifHDK* genes necessary for nitrogen-fixation. The presence of *nif*-genes in the closely related genera *Coleofasciculus* and *Symploca* suggests a loss of these vertically inherited genes in *Moorea* as a relatively recent evolutionary event, which further supports the delineation of *Moorea* as an exclusive genus.

Table VI.1: Genomic and Biochemical Characteristics of *Moorea* Strains.

Characteristics	***Moorea producta*** **3L^{T}**	***Moorea bouillonii*** **PNG5-198^{T}**
Genome size	8.5 Mbp	ND
DNA G + C content	41 mol %	42 mol %
Protein-coding genes	7,415	ND
Nitrogen fixation genes	absent	Absent
Secondary metabolites	cur/car/bar	apr/lbn
Secondary metabolites genes	PKS/NRPS	PKS/NRPS
Chlorophyll (Chl *a*)	Chl *a*	Chl *a*
Phycobiliproteins	PC/PE/AP	PC/PE/AP

Abbreviations: ND – not determined; cur – curacins; car – carmabins; bar – barbamide; apr – apratoxins; lbn – lyngbyabellins; Chl *a* – Chlorophyll a; PC – phycocyanin; PE – phycoerythrin; AP – allophycocyanin.

VI.4 Description of *Moorea* gen. nov.

Moorea gen. nov. (Mo.o.re'a. N.L. fem. n. Moorea)

Diagnosis: Filamenta solitaria vel in fasciculis irregularis, ad macroscopica, ad <10 cm longa, 25-65 (82) µm lata, not divaricata nec ramosa. Trichoma cylindrical, ad dissepimenta not vel paucim constrincta, aeruginosa, olivacea vel rubentes. Vaginae firmae, plus minusve tenues vel paucim dilatatae, externe saepe mucosae, sine colore, paucim lamellosae. Cellulae semper distincte brevior quam latae (20-55 (70) x (2) 3-10 µm), discoidae; cellula apicalis late rotundata. Reproductio hormogoniis necridiis separatur. Heterocytae akinetaeque carentes.

Typus generis: Moorea producta spec. nova

Etymologia: ad honorem in memoriam of Prof. Richard E. Moore nominata

Descriptions: Large filamentous cyanobacteria common in tropical marine oceans abundant on coral reefs, rocks or mangroves at depths ranging between 0.3-30 m. Filaments are unbranched, <10 cm in length and the diameters of the filaments are wide 25-65 (82) µm in width. Trichomes cylindrical, not attenuated towards ends, constricted or not constricted on crosswalls, surrounded by thick (3-5 µm) and distinct polysaccharide sheaths. The sheaths are typically covered by a rich diversity of mucous (often containing heterotrophic bacteria and other microorganisms). The cells are discoid, always shorter than wide 20-55 (70) µm broad and (2) 3-10 µm length. The trichomes contain necridic cells separating the trichomes into hormogonia. The terminal cells of the filaments and those of the hormogonias are rounded. *Moorea* is non-diazotrophic and the

filaments lack heterocysts or other specialized cells. *Moorea* is photosynthetic and contain phycobiliproteins (phycocyanin, phycoerythrin, and allophycocyanin) and chlorophyll *a*. *Moorea* strains are often rich in bioactive secondary metabolites typically biosynthesized by PKS, NRPS, or mixed PKS/NRPS pathways.

Type species: *Moorea producta* sp. nov. ($3L^T$)

Etymology: *Moorea* gen. nov. (Mo.o.re'a. N.L. fem. n. Moorea; in memoriam of Prof. Richard E. Moore).

VI.4.1 Description of *Moorea producta* sp. nov.

Thalus caespitosus vel prostratus, coloratus, rubescens ad viride-fuscus. Filamenta 30-67 (82) μm crassa. Vaginae sine colore, plus minusve tenues, 1-2 (12) μm latae, paucim lamellosae. Trichomata rubra vel praecipue olivaceae, cylindrica, apicem not attenuata, ad dissepimenta constricta (25) 30-65 (70) μm lata. Cellulae 3-7 μm longae, cellula apicalis rotundata, calyptra nulla.

G+C contentus = 41,2 %.

Holotypus: cultura 3L, in PCC deposita; exemplum conservatum in herbario Musei Moravici Brno (BRNM/HY 2364) depositum; icona typical: figura nostra 2.

Habitatio: ad radices arborum mangrovis, ad oras Antillarum Hollandicum, in profunditate 2-3 m.

Etymologia: contentuu multis producti chemicis

The thallus morphology range from tuft to extensive mats. The coloration is highly varied, ranging from dark-red to greenish-brown. Filaments 30-67 (82) μm width.

Sheaths are colourless, thin (1-2 μm, in extreme situation 12 μm broad), slightly lamellose. Trichomes are cylindrical, attenuated on the end, constricted on the cell walls, cells (25) 30- 65 (70) μm width and 3- 7 μm long. Apical cells rounded, without calyptra.

The G+C content of the genomic DNA of the type strain is 41.2 mol %. The type strain, $3L^T$ was isolated from coral rubble and rocks at 2-3 m depth in Curaçao, Netherlands Antilles.

Ethymology: producta = rich in metabolic products

Holotypus: cultura 3L, in PCC collection; dried material deposited at the herbarium of the Moravian Museum Brno (BRNM/HY 2364); icona typica Fig. 2

This species has often been incorrectly cited in the literature as *L. majuscula* and *L. sordida*.

VI.4.2 Description of *Moorea bouillonii* comb. nov.

Basionym: *Lyngbya bouillonii* (Hoffmann et Demoulin *Belg J Bot* 124**:** 85, 1991)

(Basic description see Hoffman & Demoulin, 1991).

The colony morphology is mat-like and tenaciously attached to surrounding substrate. The colonies are found in association with the snapping shrimp cf. *Alpheus frontalis*. Coloration fluctuates between brownish-red and dark red depending on depth. The G+C content of the genomic DNA of the type strain is 42.3 mol %. The type strain, PNG05-198^T, was isolated from coral reefs at a depth of 10 m in New Ireland, Papua New Guinea.

VI.4 Acknowledgements

We thank the governments of Papua New Guinea, Curaçao, Panama, Jamaica and the Palmyra Atoll Research Consortium for permission to collect *Moorea* specimens. We are also grateful to J. P. Euzéby for help with valid nomenclature. We also acknowledge the generous funding from Sea Grant R/NMP-103EPD and Grant MSM6007665801 by the Ministry of Education of Czech Republic. Some of the work reported here was carried out at the National Center for Microscopy and Imaging Research, which is supported by NIH grant P41-RR004050 to M. H. Ellisman. This manuscript and the naming of *Moorea* are dedicated to the late R. E. Moore (U. Hawaii) for his landmark contributions to natural products discovery from tropical marine cyanobacteria.

The text of II, in full, is published material as it appears in: Engene, N., Rottacker, E. C., Kaštovský, K., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., Gerwick, W. H. (2011). *Moorea producta* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J Syst Evol Microbiol* (In press). The dissertation author was the primary author and directed and supervised the research, which forms the basis for this chapter.

VI.5 Appendix - Supporting Information for Chapter VI

Table of Contents of Appendix

Table VI.2: Collection data for *Moorea* specimens.

Figure VI.7: Phylogenetic analyses of *Lyngbya* and *Moorea* diversification based on RNA polymerase gamma-subunit (*rpo*C1) nucleotide sequences.

Figure VI.8: Photomicrographs of cyanobacterial filaments obtained by epifluorescent microscopy.

Figure VI.9: Microphotographs of cyanobacterial filaments obtained by transmission electron microscopy (TEM).

Table VI.2: Collection data for *Moorea* specimens.

Species	Strain	Acc. Nr.	Origin	GPS Coordinates		Depth
				latitude	longitude	(m)
bouillonii	PNG5-198	FJ041298-9	Papua New Guinea	4° 16.063'S	152° 20.266'E	10
bouillonii	PNG6-41	FJ147302	Papua New Guinea	10° 15.612'S	150° 44.878'E	4
bouillonii	PNG6-42	FJ151512	Papua New Guinea	10° 15.612'S	150° 44.878'E	16
bouillonii	PNG6-47	FJ151511	Papua New Guinea	10° 13.051'S	150° 36.110'E	10
bouillonii	PNG6-50	FJ151525	Papua New Guinea	10° 13.051'S	150° 36.110'E	<3
bouillonii	PNG6-59	FJ151513	Papua New Guinea	09° 58.998'S	150° 57.359'E	17
bouillonii	PNG6-62	FJ151514	Papua New Guinea	10° 06.277'S	150° 57.673'E	13
bouillonii	PNG6-65	FJ151515	Papua New Guinea	10° 16.579'S	151° 11.492'E	7
bouillonii	PNG7-4	FJ151519	Papua New Guinea	10° 15.836'S	150° 40.193'E	17
bouillonii	PNG7-6	FJ041304	Papua New Guinea	10° 15.836'S	150° 40.193'E	17
bouillonii	PNG7-9	FJ151520	Papua New Guinea	10° 15.177'S	150° 47.034'E	17
bouillonii	PNG7-14	FJ041305	Papua New Guinea	9° 32.491'S	150° 16.715'E	N/A
bouillonii	PNG7-19	FJ041307	Papua New Guinea	8° 32.364'S	148° 98.104'E	10
bouillonii	PNG7-22	FJ154877-8	Papua New Guinea	5° 25.566'S	05° 06.754'E	17
bouillonii	PNG7-29	FJ151528-9	Papua New Guinea	5° 25.575'S	150° 06.234'E	N/A
bouillonii	PNG7-32	FJ151524	Papua New Guinea	5° 25.575'S	150° 06.234'E	12

Table VI.2: Continued

Species	Strain	Acc. Nr.	Origin	GPS Coordinates		Depth
				latitude	longitude	(m)
bouillonii	PNG7-44	FJ151526	Papua New Guinea	5° 26.192'S	150° 40.813'E	13
bouillonii	PNG7-63	FJ151518	Papua New Guinea	5° 26.926'S	150° 47.897'E	24
producta	PNG5-192	FJ041300	Papua New Guinea	4° 16.063'S	152° 20.266'E	10
producta	PNG5-194	FJ151523	Papua New Guinea	4° 16.063S	152° 20.266'E	10
producta	PNG6-2	FJ157183	Papua New Guinea	10° 36.990'S	150° 37.396'E	15
producta	PNG6-9	FJ041302-3	Papua New Guinea	10° 40.294'S	150° 41.226'E	12
producta	PNG6-38	FJ151516-7	Papua New Guinea	10° 15.612'S	150° 44.878'E	13
producta	PNG6-45	FJ147304	Papua New Guinea	10° 15.612'S	150° 44.878'E	<1
producta	PNG6-48	FJ041301	Papua New Guinea	10° 13.051'S	150° 36.110'E	7
producta	PNG6-51	FJ041309	Papua New Guinea	10° 13.051'S	150° 36.110'E	17
producta	PNG6-52	FJ15159-10	Papua New Guinea	10° 13.051'S	150° 36.110'E	22
producta	PNG6-54	FJ151522	Papua New Guinea	10° 14.055'S	150° 49.873'E	<1
producta	PNG6-57	FJ147305	Papua New Guinea	10° 14.055'S	150° 49.873'E	3
producta	PNG6-68	FJ147300-1	Papua New Guinea	10° 16.579'S	151° 11.492'E	<3
producta	PNG7-13	FJ041306	Papua New Guinea	10° 15.177'S	150° 47.034'E	17
producta	PNG7-19	FJ041307	Papua New Guinea	8° 32.364'S	148° 48.104'E	10

Table VI.2: Continued

Species	Strain	Acc. Nr.	Origin	GPS Coordinates		Depth
				latitude	longitude	(m)
producta	PNG6-221	FJ356669-70	Papua New Guinea	10° 17.274'S	151° 00.390'E	5
producta	PNG5-223	EU492878	Papua New Guinea	3° 40.620'S	152° 20.406'E	9
producta	PNG5-194	EU492877	Papua New Guinea	3° 49.626′S	152° 26.017′E	15
producta	3L	EU31590	Curaçao	N/A	N/A	>2
producta	19L	AY599501	Curaçao	N/A	N/A	N/A
producta	NAC8-47	GU724198	Curaçao	N/A	N/A	
producta	NAC8-48	GU724199-0	Curaçao	N/A	N/A	>2
producta	NAC8-49	GU724201-2	Curaçao	N/A	N/A	>1
producta	NAC8-51	GU724204	Curaçao	N/A	N/A	>1
producta	NAC8-52	GU724205	Curaçao	N/A	N/A	>1
producta	NAC8-53	GU724206	Curaçao	N/A	N/A	3-4
producta	JHB	FJ151521	Jamaica	N/A	N/A	2
bouillonii	PAL08-16	GU111927	Palmyra	5° 52.021'N	162° 03.813'W	10-15
bouillonii	PAL-8-10-09-1	-	Palmyra	5° 52.021'N	162° 03.813'W	N/A
producta	PAL-8-15-08-1	GQ231521	Palmyra	5° 52.021'N	162° 03.813'W	1-2
producta	PAL-8-17-08-2	GQ231522	Palmyra	5° 52.021'N	162° 03.813'W	1-2
producta	PAB	AY599502	Panama	N/A	N/A	N/A

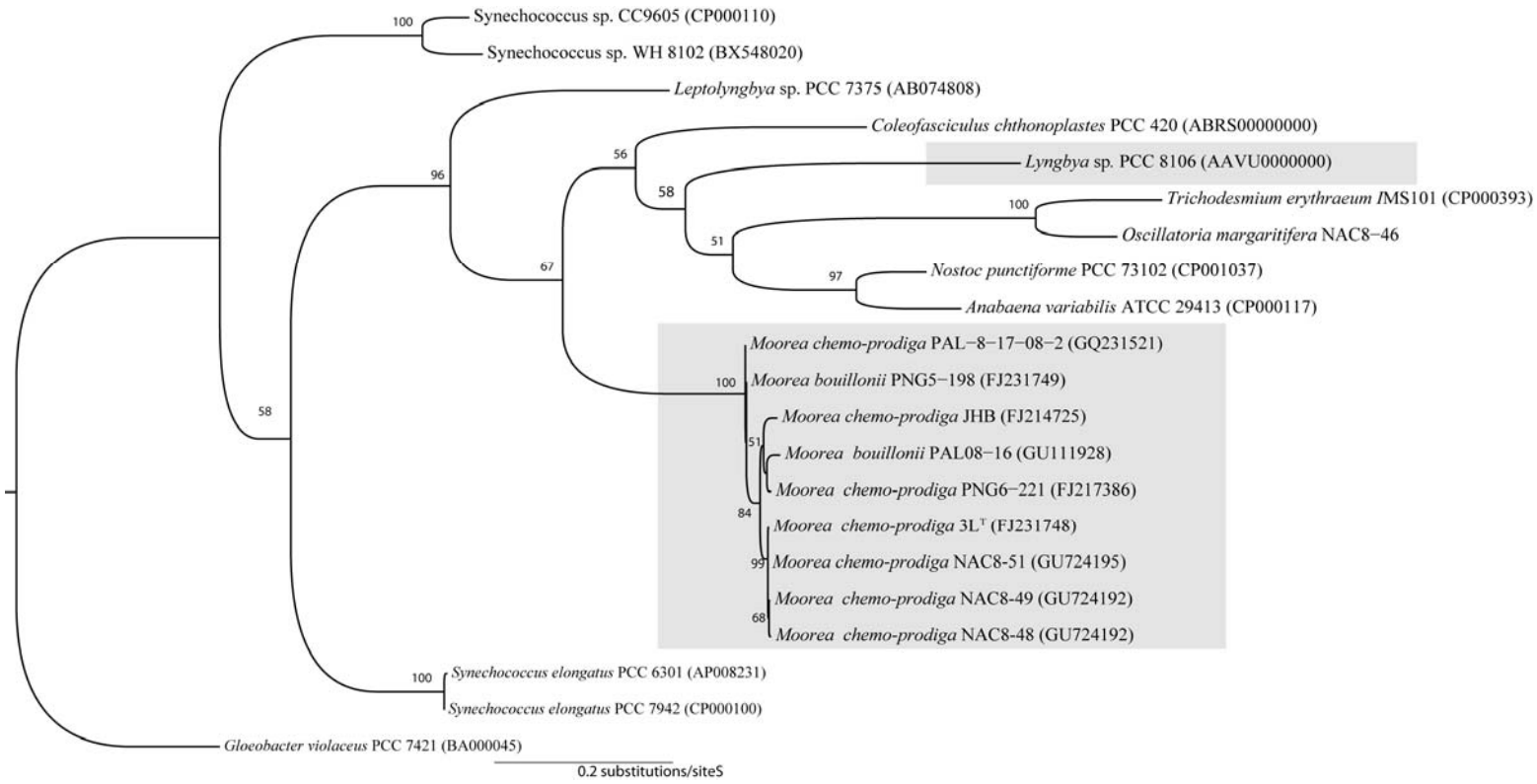


Figure VI.7: Maximum-likelihood (PhyML) phylogenetic analyses of *Lyngbya* and *Moorea* diversification based on RNA polymerase gamma-subunit (*rpo*C1) nucleotide sequences. The specimens are indicated as species, strain, and accession numbers in brackets. Type-strains were obtained from *Bergey's Manual.* The support values are indicated as boot-strap and posterior probability for the maximum likelihood, Bayesian inference, and maximum parsimony. The scale bar is indicated at 0.04 expected nucleotide substitutions per site using the GTR+I+G substitution model.

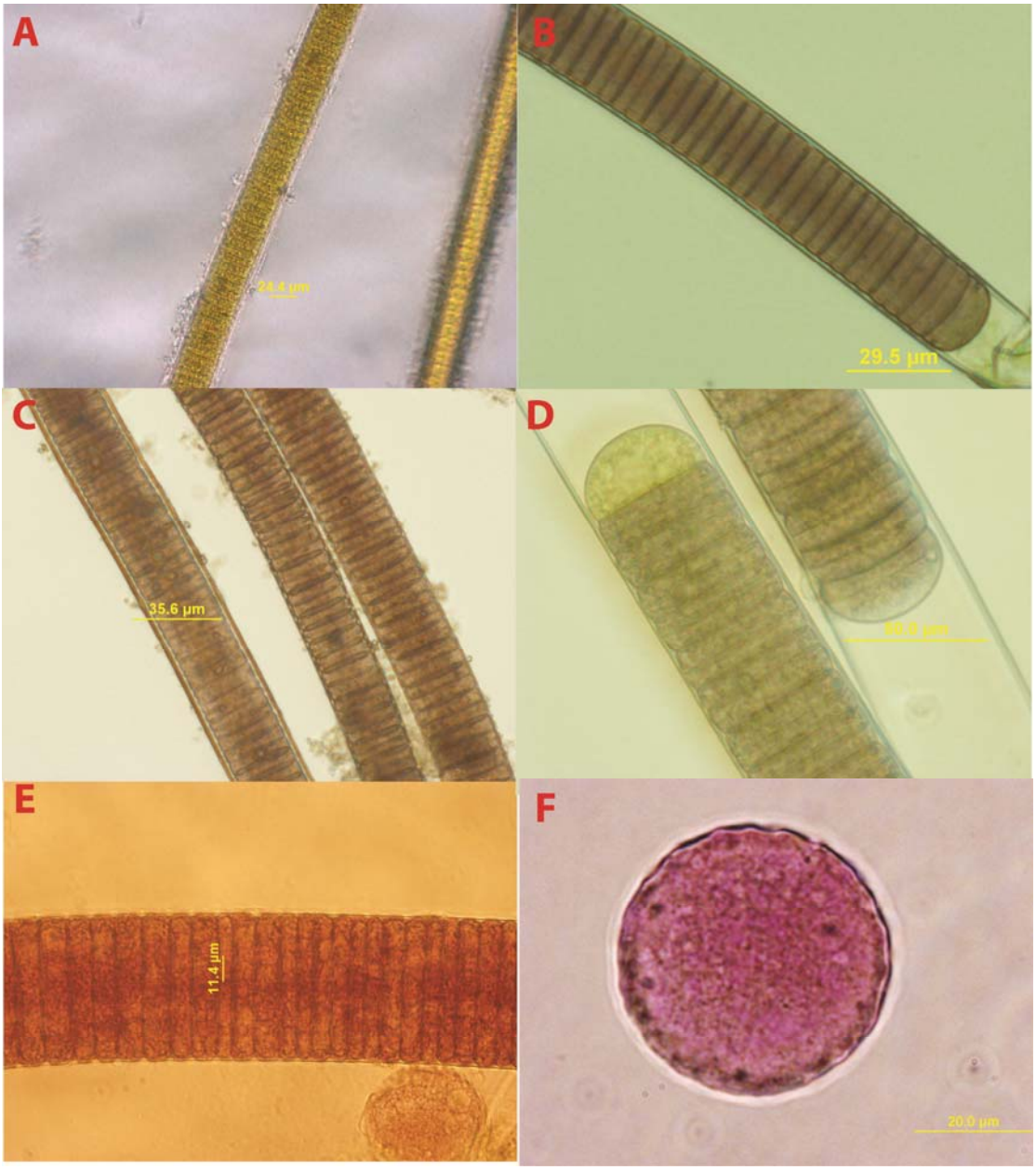


Figure VI.8: Photomicrographs of cyanobacterial filaments obtained by epifluorescent microscopy (40X). **(A)** *Lyngbya aestuarii* PCC 7419^{T}; **(B)** *Moorea bouillonii* PNG05-198^{T}: **(C)** *M. majuscula* PNG6-221, **(D)** *M. majuscula* JHB, **(E)** *M. sordida* 3L^{T} **(F)** *M. sordida* 3L^{T} single-cell.

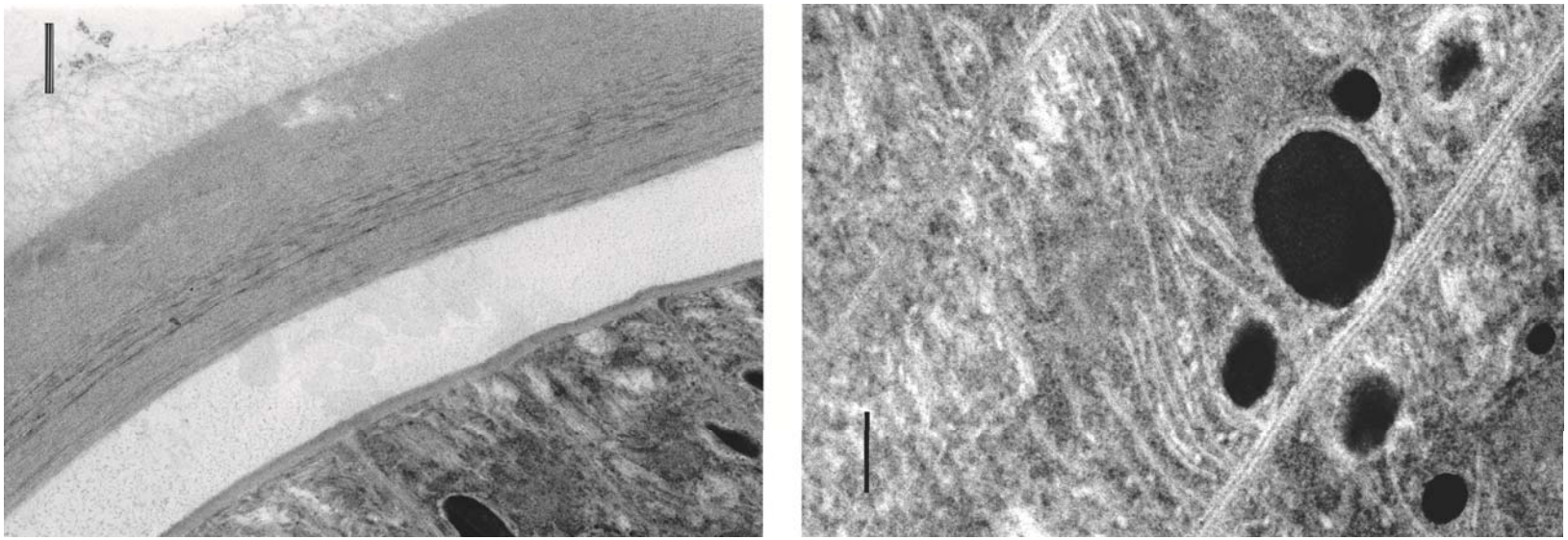


Figure VI.9: Microphotographs of cyanobacterial filaments obtained by transmission electron microscopy (TEM). polysaccharide sheaths of *Lyngbya* sp. PCC7419^{T} and thylakoid membranes and carboxysomes of *Lyngbya* sp. PCC7419^{T}.

VI.6 References

Boyer, S. L., Flechtner, V. R. & Johansen, J. R. (2001). Is the 16S–23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Mol Biol and Evol* 18: 1057-1069.

Cannone, J. J., Subramanin, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pnde, N., Schang, Z., Yu, N.,Gutell R. R. (2002). The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3: 1471-2105.

Casamatta, D. A., Johansen, J. R., Vis, M. L., Broadwater, S. T. (2005). Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cynaobacteria). *J Phycol* 41: 421-438.

Castenholz, R. W. (2001). Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 473-553. Edited by Boone, D.R., Castenholz, R.W., & Garrity, G.M. New York, USA: Springer.

Engene N., Coates R. C., Gerwick W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601.

Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya*. *Environ Microbiol* 13: 1601-1610.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277–284.

Golubic, S., Abed, R. M. M., Palinska, K., Pauillac, S., Chinain, M., Laurent, D. (2010). Marine toxic cyanobacteria: diversity, environmental responses and hazards. *Toxicon* 56: 836-841.

Gugger, M., Molica, R., Le Berre, B., Dufour, P., Bernard, C., Humbert, J-F. (2005). Genetic diversity of *Cylindrospermopsis* strains (Cyanobacteria) isolated from four continents. *Appl Env Microbiol* 71: 1097-1100.

Hoffmann, L., Demoulin, V. (1991). Marine Cyanophyceae of Papua New Guinea. II. *Lyngbya bouillonii* sp. nov., a remarkable tropical reef-inhabiting blue-green alga. *Belg J Bot* 124: 82-88.

Katoh, K., Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9: 286-298.

Komárek, J., Anagnostidis, K. (2005). *Süsswasserflora von Mitteleuropa*, 19/2 edn. Heidelberg, Germany: Elsevier/Spektrum.

Jones, A. C.*, Monroe, E. A.* (*co 1st authors), Podell, S., Hess, W., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R., Yates III, J. R., Reinhardt, R., Kube, M., Burkart, M., Allen, E. E., Gerwick, W. H., Gerwick, L. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* 108: 8815-8820.

Liu, L., Rein, K. S. (2010). New peptides isolated from *Lyngbya* species: a review. *Mar Drugs* 8: 1817-1837.

Nübel, U., Garcia-Pichel, F., Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63: 3327-3332.

Otsuka, S., Suda, S., Li, R., Watanabe, M., Oyaizu, H., Matsumoto S., Watanabe, M. M. (1999). Phylogenetic relationships between toxic and non-toxic strains of the genus Microcystis based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol Lett* 172: 15-21.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256.

Ronquist, F., Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxf)* 12: 1572-1574.

Sharp, K., Arthur, K. E., Gu, L., Ross, C., Harrison, G., Paul, V. J. (2009). Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* species (Cyanobacteria: Oscillatoriales) from reefs of southeastern Florida. *Appl Environ Microbiol* 75: 2879–2888.

Tandeau de Marsac, N., Houmard, J. (1988). Complementary chromatic adaption: physiological conditions and action spectra. *Methods Enzymol* 167: 318-328.

Tidgewell, K., Clark, B. R., Gerwick, W. H. (2010). The natural products chemistry of cyanobacteria. In *Comprehensive Natural Products II Chemistry and Biology* vol 2, pp.141–188. Edited by Mander, L. & Lui, H.-W. Oxford: Elsevier.

Weckesser, J., Jürgens, U. J. (1988). Cell walls and external layers. *Methods Enzymol* 167: 173-187.

Zwickl, D. J. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Thesis, The University of Texas at Austin.

Chapter VII

Taxonomic Distribution of Marine Cyanobacterial Natural Products

Abstract

In this chapter we undertook a phylogenetic approach to understand the evolution and taxonomic distribution of NPs in marine cyanobacteria. The producing marine cyanobacterial specimens of a total of 108 different NPs were phylogenetically investigated and placed in perspective with respective type-strains to speculate their true taxonomy. As a result only one single NP-producing cyanobacterial specimen, the trichamide-producing *Trichodesmum erythraeum,* was phylogenetically closely related to the genera that they have been assigned. All remaining NP-producing specimens were either evolutionarily distant or polyphyletic with their type-strains that they have been assigned. The simple conclusion of these results is that our current morphology-based classification systems have underestimated cyanobacterial biodiversity and skewed the perceived taxonomic distribution of NPs in marine cyanobacteria.

VII.1 Introduction

Cyanobacteria are considered the most ancient group of oxygenic photosynthetic organisms (Graham & Wilcox, 2000). Over the course of 3 billion years, this bacterial phylum has adapted to most habitats on the planet and is among the most abundant and geographically widespread group of prokaryotes known (Whitton, 1992; Stanley, 2004).

In recent years, tropical marine cyanobacteria have attracted much attention due to their extraordinary capacities to produce diverse and highly bioactive secondary metabolites (Gerwick *et al*., 2008; Nunnery *et al*., 2010; Tidgewell *et al*., 2010; Tan *et al*., 2010). These bioactive molecules are often potent toxins, causing harmful algal blooms (HABs) (Osborne *et al*., 2008; Golubic *et al*., 2010). In tropical marine environments, these toxic HABs are increasing globally in frequency and size by alarming rates and represent hazards to both human health and natural ecosystems (Paul *et al*., 2005; Golubic *et al*., 2010). Cyanotoxins are also a major factor responsible for black band disease (BBD) contributing to coral decline in the wider Caribbean, the Indo-Pacific, the Red Sea, and the Great Barrier Reef (Gantar *et al*., 2009). Cyanotoxins have recently been linked with diseases and toxicities in larger marine animals (Osborne *et al*., 2008). For example, the tumor promoting compounds lyngbyatoxin A and debromoaplysiatoxin produced by *Lyngbya majuscula* have been correlated with fibropapillomatosis, a potentially fatal neoplastic disease, in sea turtles (Arthur *et al*., 2008). Debromoaplysiatoxin has also been detected in the dorsa of manatees with ulcerative dermatitis from feeding on *Lyngbya*-dominated mats (Harr *et al*., 2008).

Despite their hazardous toxicity, many cyanobacterial secondary metabolites also have promising natural products potential for a broad spectrum of pharmaceutical applications such as anticancer, anti-inflammatory, antibacterial and anti-infective therapeutic agents (Gerwick *et al*., 2008; Tidgewell *et al*., 2010; Tan, 2010; Nunnery *et al*., 2010; Villa & Gerwick, 2010). The economic value of anti-cancer drugs alone from marine sources has been estimated to be between $563 billion and 5.69 trillion (Erwin *et al*., 2010). In addition, many of these cyanobacterial secondary metabolites have other potential commercial applications such as insecticides, algaecides, and herbicides (Berry *et al*., 2008).

Unfortunately, for both HAB monitoring and NP discoveries, the taxonomy of tropical marine cyanobacteria is in a state of chaos. This is primarily due to the lack of proper classification systems since tropical marine cyanobacteria have only recently been explored. Instead of recognizing these tropical marine groups as unique taxa, their identification has, with few exceptions, relied on classification systems tied to morphospecies of terrestrial and freshwater specimens from temperate regions (Komárek *et al*., 2006).

In this chapter I will discuss how current morphology-based classification systems have underestimated cyanobacterial biodiversity and skewed the perceived taxonomic distribution of NPs in marine cyanobacteria. Herein, all phylogenetically investigated NP-producing cyanobacteria will be analyzed by placing them in perspective with respective type-specimens to speculate on their true taxonomy. In summary, this study will attempt a cladistic approach to understand the evolution and taxonomic distribution of NPs in marine cyanobacteria.

VII.2 Materials and Methods

VII.2.1 DNA-isolation, PCR-amplification and Cloning

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's specifications. DNA concentration and purity was measured on a DU® 800 spectrophotometer (Beckman Coulter). The PCR reaction volumes were 25 µL containing 0.5 µL (~50 ng) of DNA, 2.5 µL of 10 x PfuUltra IV reaction buffer, 0.5 µL (25 mM) of dNTP mix, 0.5 µL of each primer (10 µM), 0.5 µL of PfuUltra IV fusion HS DNA polymerase and 20.5 µL dH_2O. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 2 min at 95 °C, 25 cycles of amplification: 20 sec at 95 °C, 20 sec at 50 °C and 1.5 min at 72 °C, and final elongation for 3 min at 72 °C. PCR products were purified using a MinElute® PCR Purification Kit (Qiagen) before subcloning into the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen). Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced bidirectionally with M13 vector-primers as well as internal primers. The gene sequences are available in the DDBJ/EMBL/GenBank databases (Table VII.1).

VII.2.2 Phylogenetic Inference

The gene sequences were obtained from the National Center for Biotechnology Information (NCBI) web page (www.ncbi.nlm.nih.gov). Reference strains were selected from *Bergey's Manual* (Castenholz, 2002) and the type-strains selected from CyanoDB

(Komárek & Hauer, 2011). Uncorrected gene sequence divergence (*i.e.* p-distance) was determined by the pair-wise distance calculation without model selection in Mega 4.0 (Tamura et al. 2007). Chimeric gene sequences were predicted using the Pintail software with the cut-off size set at >600 bp (Ashelford *et al*., 2005), and were manually confirmed by comparison of (NJ) phylograms for different regions (>300 bp) of the sequences. The unicellular *Gloeobacter violaceus* PCC 7421 (GenBank acc. Nr. NC005125) was included as an evolutionary distant out-group.

Gene sequences were aligned using the L-INS-I algorithm in MAFFT 6.717 (Katoh & Toh, 2008). The best-fitting nucleotide substitution models optimized by Maximum Likelihood (ML) were compared and selected using uncorrected/corrected Akaike Information Criterion (AIC/AICc), Bayesian Information Criterion (BIC), and the Decision-theoretic (DT) in jModeltest 0.1.1 (Posada, 2008). The evolutionary histories of the cyanobacterial genes were inferred using Maximum likelihood (ML) and Bayesian inference algorithms. The ML inference was performed using GARLI 1.0 with >1,000 bootstrap-replicates (Zwickl, 2006). Bayesian inference was conducted using MrBayes 3.1 with four Metropolis-coupled MCMC chains (one cold and three heated) ran for 3-10,000,000 generations (Ronquist & Huelsenbeck, 2003). The MCMC convergence was detected by AWTY (Nylander *et al*., 2008).

VII.2.3 Secondary Metabolite Detection

Algal biomass (~1 g) of each specimen was exhaustively extracted with CH_2Cl_2–MeOH (2:1). The extract was dried under vacuum and the dried residues were re-dissolved in MeOH at a concentration of 1 $mg \cdot mL^{-1}$. Each sample (10 µL) was injected

into an LC ESI-MS system (LCQ Advantage Max spectrometer and UV-profiles by Surveyor PDA plus detector, Thermo Finnigan) and separated on an RP HPLC column (HP Lichrosphere 100 RP-18, 4 × 125 mm, 5.0 μm) with step gradient elution of 0.1% formic acid in water (eluent A) and 100% ACN (eluent B). Gradient program: 0-5 min, B, 45%; 5-55 min, B, 45-100%; 55-65 min, B, 100%; flow rate, 700 $\mu L \cdot min^{-1}$. The column temperature was kept at 30 °C. The MS and MS^2 spectra and retention time of each peak were recorded using the positive ion detection mode. Identification of secondary metabolites required support of predicted isotope patterns, corresponding MS^2 fragmentations and conserved retention times (RT).

VII.3 Results and Discussion

SSU (16S) rRNA gene sequences of NP-producing marine cyanobacteria were either sequenced from genetically preserved samples or obtained from publically available databases. As a result, 16S rRNA gene sequences for the producing strains of a total of 94 different NPs were available for phylogenetic inference (Table VII.1). Phylogenetic inferences were performed to place the NP-producing strains in evolutionarily perspective with respective available type/reference strains (Figure VII.1).

As a result, multiple NP-producing polyphyletic groups, including *Lyngbya*, *Oscillatoria*, *Leptolyngbya*, and *Synechococcus*, are evolutionarily unrelated to their respective type/reference-strains and, thus, represent new groups (highlighted with blue boxes in Figure VII.1).

Table VII.1: True Taxonomic Identification of Natural Products-producing Marine Cyanobacteria.

Natural products	**Published Identification**	**p-distance (%)**	**Phylogenetically-revised identification**	**Strain**	**GenBank Acc Nr.**
Almiramides A-C	*Lyngbya majuscula*	9.7	*Moorea producta*	PAP29JUN071	NYS
Apratoxin A-C (**19-21**)	*Lyngbya majuscula*	9.7	*Moorea bouillonii*	PNG5-198	FJ041298
Apratoxins D	*Lyngbya sordida*	9.5	*Moorea producta*	PNG5-194	U492877
Apratoxins F-G (**22-23**)	*Lyngbya bouillonii*	9.8	*Moorea bouillonii*	PAL08-16	GU111927
Barbamide (**12**)	*Lyngbya majuscula*	9.7	*Moorea producta*	3L	GU724199
Bastimolide	*Lyngbya majuscula*	8.3	*Pseudomoorea tropica*	PAB10FEB101	NYS
Carmabins A-B (**7-8**)	*Lyngbya majuscula*	9.7	*Moorea producta*	3L	GU724199
Coibamide A (**32**)	*Leptolyngbya* sp.	11.8	gen. nov	PAC-10-03	Unpublished
Coibacins A-D	*Lyngbya majuscula*	8.3	*Pseudomoorea tropica*	PAC18FEB101	NYS
Credneric acid	*Lyngbya majuscula*	8.1	*Pseudomoorea tropica*	PAC17FEB103	NYS

Table VII.1: Continued

Natural products	Published identification	p-distance (%)	Phylogenetically-revised identification	Strain	GenBank Acc Nr.
Crossbyanols A-D (**68-71**)	*Leptolyngbya crossbyana*	7.4	gen. nov.	HI09-1	GU111930
Curacin A-C (**1, 2**)	*Lyngbya majuscula*	9.7	*Moorea producta*	3L	GU724199
Curacin D	*Lyngbya majuscula*	9.6	*Moorea producta*	PAL0817082	GQ231522
	Lyngbya polychroa	9.7	*Moorea producta*	LP5	FJ602753
Curazole	*Lyngbya majuscula*	9.7	*Moorea producta*	3L	GU724199
Cyanolide A	*Lyngbya bouillonii*	9.7	*Moorea bouillonii*	PNG5-198	FJ041298
Dechlorobarbamide (**13**)	*Lyngbya majuscula*	9.7	*Moorea producta*	3L	GU724199
Dolastatin 10 (**24**)	*Symploca hydnois*	5.1	gen. nov.	PNG14JUL076	NYS
Dolastatin 12 (**47**)	*Lyngbya majuscula/ Schizothrix calcicola*	8.3	*Pseudomoorea tropica*	NAC8-46	GU724197
Dragonamide A-B (**9-10**)	*Lyngbya majuscula*	9.6	*Moorea producta*	PAP28JUl061	NYS
Dragonamide C-D	*Lyngbya polychroa*	9.7	*Moorea producta*	LP5	FJ602753*
Dudawalamide E	*Lyngbya majuscula*	9.7	*Moorea producta*	PNG6-221a	FJ356669/FJ157181

Table VII.1: Continued

Natural products	Published identification	p-distance (%)	Phylogenetically-revised identification	Strain	GenBank Acc Nr.
Hectochlorin (**16**)	*Lyngbya majuscula*	11.5	*Moorea producta*	JHB	FJ151521
Hermanamide	*Lyngbya penicilliformis*	8.3	gen. nov.	PAP17JUL101	NYS
Hoiamides A-C (**27-29**)	*Phormidium gracile/ Lyngbya majuscula*	7.2/8.1	gen. nov./ *Pseudomoorea tropica*	PNG06-65[a]	HM072001/ HM072002
Honaucin	*Leptolyngbya crossbyana*	7.4	gen. nov.	HI09-1	GU111930
Indianone	*Lyngbya majuscula*	5.1	gen. nov.	GTR 6 III 95-1	AF510978*
Jamaicamides A-C (**3**)	*Lyngbya majuscula*	11.5	*Moorea producta*	JHB	FJ151521
Kimbeamide	*Symploca hydnois*	5.1	gen. nov.	PNG14JUL076	NYS
Leptolyngolide	*Leptolyngbya* sp.	11.8	gen. nov	PAC-10-03	Unpublished
Lyngbyabellin A (**17**)	*Lyngbya majuscula*	9.7	*Moorea producta*	PNG6-221	FJ356669
Lyngbyabellin B (**18**)	*Lyngbya majuscula*	9.7	*Moorea bouillonii*	PNG5-198	FJ041298
Lyngbyapeptin A	*Lyngbya bouillonii*	9.9	*Moorea bouillonii*	VP417	AY049751

Table VII.1: Continued

Natural products	Published identification	p-distance (%)	Phylogenetically-revised identification	Strain	GenBank Acc Nr.
Lyngbyastatin 1, 3 (**48-49**)	*Lyngbya majuscula*	8.3	*Pseudomoorea tropica*	PNG05-4	EU253968
Lyngbyatoxin A	*Lyngbya majuscula*	9.4	*Moorea producta/ Pseudomoorea tropica*	PNG6-221a	FJ356669/FJ157181
Majusculamide A-B, D	*Lyngbya majuscula*	9.4	*Moorea producta/ Pseudomoorea tropica*	PNG6-221a	FJ356669/FJ157181
Malyngamide C (**6**)	*Lyngbya majuscula*	9.7	*Moorea producta*	PAL0817082	GQ231522
Medusamide	*Oscillatoria* sp.	3.0	*Trichodesmium* sp. nov	PAC18FEB101	NYS
Methyltumonoate A-B (**42-43**)	*Lyngbya majuscula/ Schizothrix calcicola*	8.3	*Pseudomoorea tropica*	NAC8-46	GU724197
Microcollins A-B (**42-43**)	*Lyngbya majuscula*	10.0	*Pseudomoorea tropica*	LP16	FJ602745
Palau'imide	*Lyngbya bouillonii*	9.9	*Moorea bouillonii*		AY049751*

Table VII.1: Continued

Natural products	Published identification	p-distance (%)	Phylogenetically-revised identification	Strain	GenBank Acc Nr.
Palmyramide A (**11**)	*Lyngbya majuscula*	9.7	*Moorea producta*	PAL0817082	GQ231522
Palmyrolide A (**Y**)	*Leptolyngbya/Oscillatoria*	10.9/3.9	gen. nov./ *Oscillatoria* sp.*	PAL08-3[a]	HM585025/ HM585026
Phormidolide (**67**)	*Phormidium* sp.	-	*Leptolyngbya* sp.*	ISB-3N94-8PLP	Unpublished
Pitiamide A (**46**)	*Lyngbya majuscula/ Microcoleus* sp.	10.4	*Pseudomoorea tropica*	1	AF510982*
Portoamide A-D (**63-66**)	*Oscillatoria* sp.	9.8	gen. nov.	LEGE 05292	GU085101
Santacruzamate	*Symploca* sp.	5.3	gen. nov.	PAC19FEB101	NYS
Symplostatin 1, 2 (**25-26**)	*Symploca hydnois*	5.1	gen. nov.	VP377	AF306497*
Synechobactins A-C	*Synechococcus* sp.	8.4	gen. nov.	PCC 7002	AJ000716
Toronamide	*Symploca* sp.	5.5	gen. nov.	PAB10FEB102	NYS

Table VII.1: Continued

Natural products	Published identification	p-distance (%)	Phylogenetically-revised identification	Strain	GenBank Acc Nr.
Trichamide	*Trichodesmium erythraeum*	0	*Trichodesmium erythraeum*	IMS101	NC008312
Tumonoic acids A-C (**33-35**)	*Lyngbya majuscula/Schizothrix calcicola*	8.3	*Pseudomoorea tropica*	NAC8-46	GU724197
Tumonoic acids D-I (**36-41**)	*Blennothrix cantharidosmum*	-	*Pseudomoorea tropica*	PNG05-4	EU253968
Ulongamide A-F	*Lyngbya* sp.	9.7	*Moorea producta*	NIH309	AY049752*
Venturamide A (**52**)	*Oscillatoria* sp.	3.4	*Pseudomoorea tropica*	PAP01APR058	EU253967
Veraguamide A-C, H-L (**53-60**)	cf. *Oscillatoria margaritifera*		*Pseudomoorea tropica*	PAC17FEB102	HQ900689
Viridamides A-B (**50-51**)	*Oscillatoria nigro-viridis*	3.2	*Pseudomoorea epiphyta*	3LOSC	EU244875
Wewakazole	*Lyngbya majuscula*	9.7	*Moorea producta*	PNG7-22	FJ154877

* Obtained from GenBank.

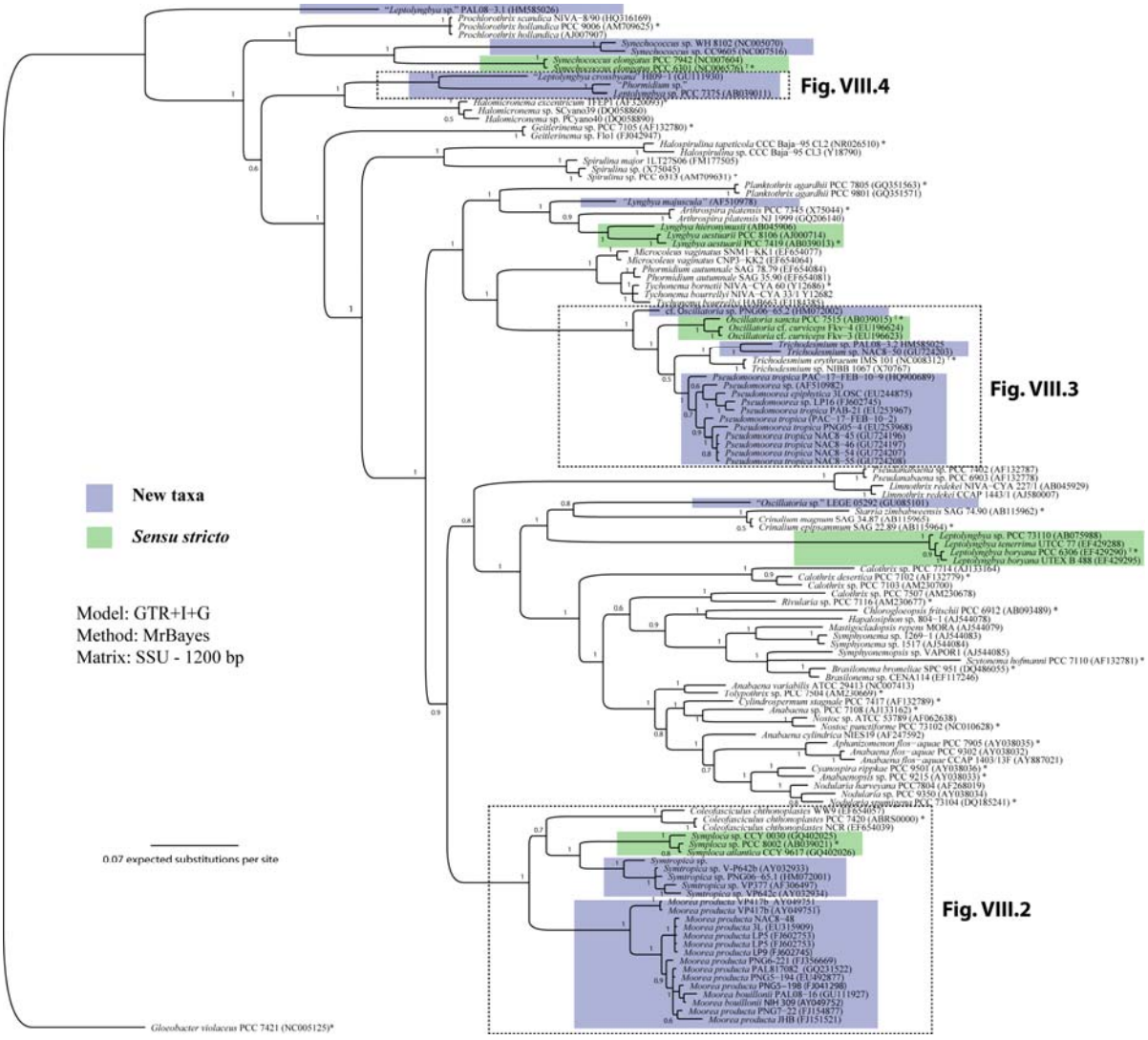


Figure VII.1: Phylogenetic inference of marine NP-producing cyanobacteria. The NP-producing groups are placed in perspective with other cyanobacterial groups, in particular their corresponding reference strains (highlighted with an asterisks) or type-strains (^{T}). The reference strains are obtained from *Bergey's Manual* and the type-strains from CyanoDB (Castenholz, 2002; Komárek & Hauer, 2011). Recognition of new genera are defined as either: (i) paraphyletic or polyphyletic groups or (ii) lineages with >5% 16S rRNA gene sequence divergence to corresponding reference/type-strain (for reference see Tindall *et al*., 2010). Proposed new taxa are highlighted with blue boxes while the true lineages are highlighted with green boxes. The phylogram is based on SSU (16S) rRNA gene sequences using bayesian method (MrBayes) and the support values are indicated as posterior probability at the nodes. The specimens are indicated as species, strain, and access number in brackets. The scale bar is indicated at 0.07 expected nucleotide substitutions per site.

VII.3.1 Natural Products from *Moorea* gen. nov.

A large proportion of the natural products attributed the genus *Lyngbya* were here shown to be produced by the "tropical marine *Lyngbya*", recently renamed and described as the genus *Moorea* in Chapter VI.

These natural products isolated from *Moorea* gen. nov. include the curacins A-D (**1**) and (**2**), curazole (GenBank acc. Nr. GU724199), the jamaicamides A (**3**), B (**4**) and C (**5**) (GenBank acc. Nr. FJ151521), malyngamide C (**6**) (GenBank acc. Nr. GQ231522), the carmabins A (**7**) and B (**8**) (GenBank acc. Nr. GU724199), dragonamide A (**9**) and (**10**), palmyramide A (**11**) (GenBank acc. Nr. GQ231522), barbamide (**12**) and dechlorobarbamide (**13**) (GenBank acc. Nr. GU724199), herbamides A (**14**) and B (**15**), hectochlorin (**16**) (GenBank acc. Nr. FJ151521), lyngbyabellin A (**17**) (GenBank acc. Nr. FJ35666), lyngbyabellin B (**18**) (GenBank acc. Nr. FJ041298), apratoxins A (**19**), B (**20**) and C (**21**) (GenBank acc. Nr. AY049751), apratoxins F (**22**) and G (**23**) (GenBank acc. Nr. GU111927).

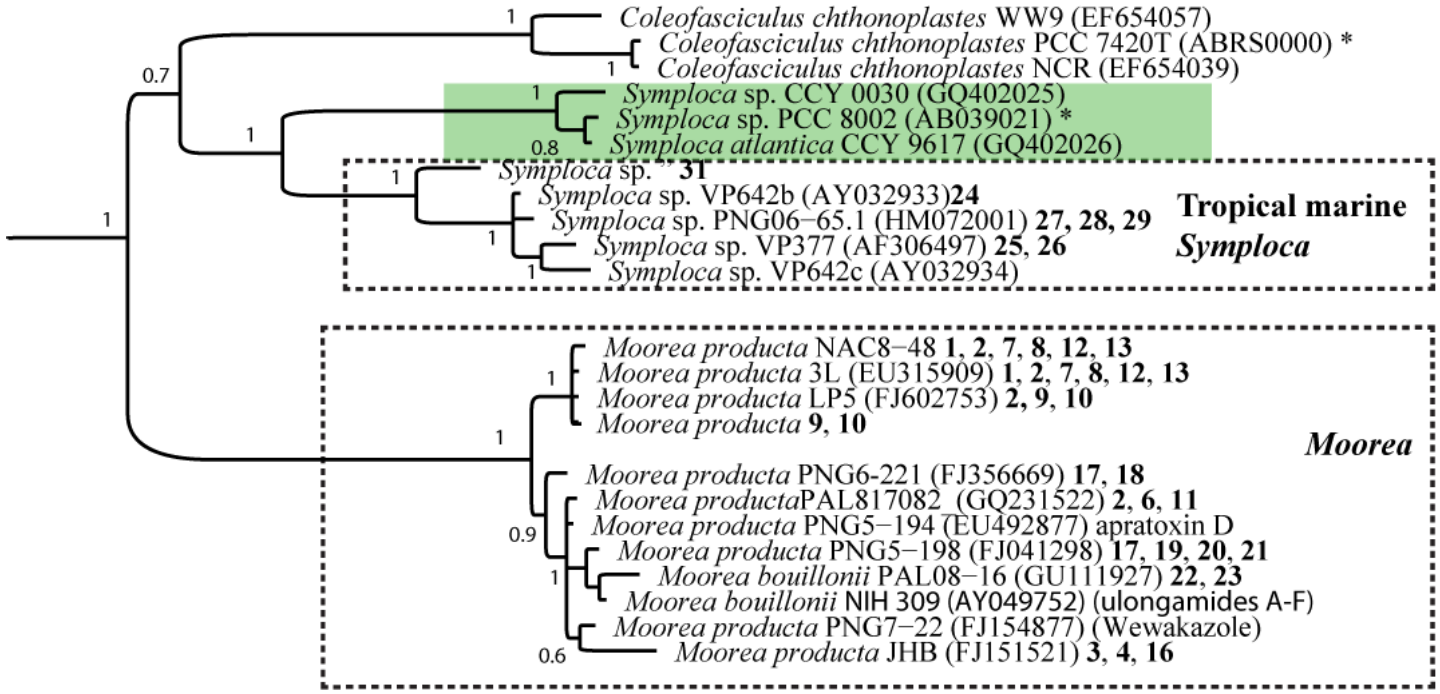


Figure VII.2: Highlight of the lineage including *Moorea* gen. nov. (formerly "tropical marine *Lyngbya*") and "tropical marine *Symploca*". The green box represents the genus *Symploca* (reference strain = PCC 8002; GenBank acc. Nr. AB039021). The bold numbers represent natural products isolated from corresponding specimens.

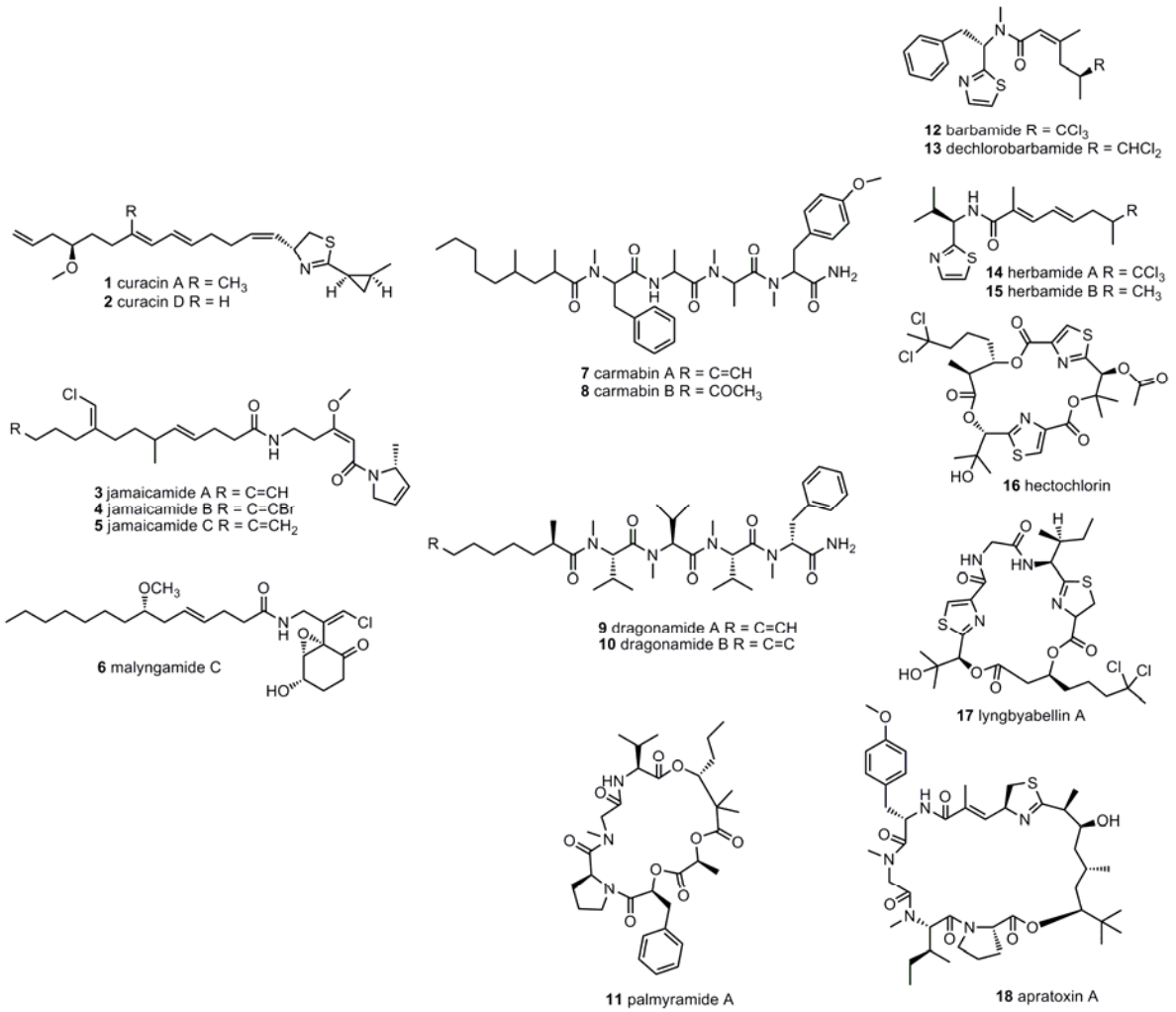


Figure VII.3: Natural products from *Moorea* gen. nov.

VII.3.2 Natural Products from *Pseudomoorea* gen. nov.

In this thesis, I distinguished *Pseudomoorea* gen. nov. (formerly "tropical marine *Oscillatoria*") from *Oscillatoria* because of phylogenetic, geographic, ecological, and biochemical differences between the two lineages. The morphological similarities between the genera *Lyngbya*, *Moorea*, *Pseudomoorea* and *Oscillatoria* have resulted in much taxonomic confusion between these different groups and many of their natural products have consequently been attributed to incorrect groups. For example, some of the molecules that have been ascribed to *Lyngbya* are phylogenetically show here to be produced by *Pseudomoorea* gen. nov. These molecules include: tumonoic acids A (**33**), B (**34**), C (**35**), D **(36**), E (**37**), F (**38**), G (**39**), H (**40**), and I (**41**)**,** methyl tumonoates A (**42**) and B (**43**)**,** microcolins A (**44**) and (**45**), pitiamide A (**46**), dolastatins 12 (**47**), lyngbyastatin 1 (**48**) and lyngbyastatin 3 (**49**).

Natural products from *Pseudomoorea* gen. nov. also include viridamide A (**50**) and B (**51**) (GenBank acc. Nr. EU244875), venturamide A (**52**) and B (**53**) (GenBank acc. Nr. EU253967), and veraguamide (**54-61**) (GenBank acc. Nr. HQ900689).

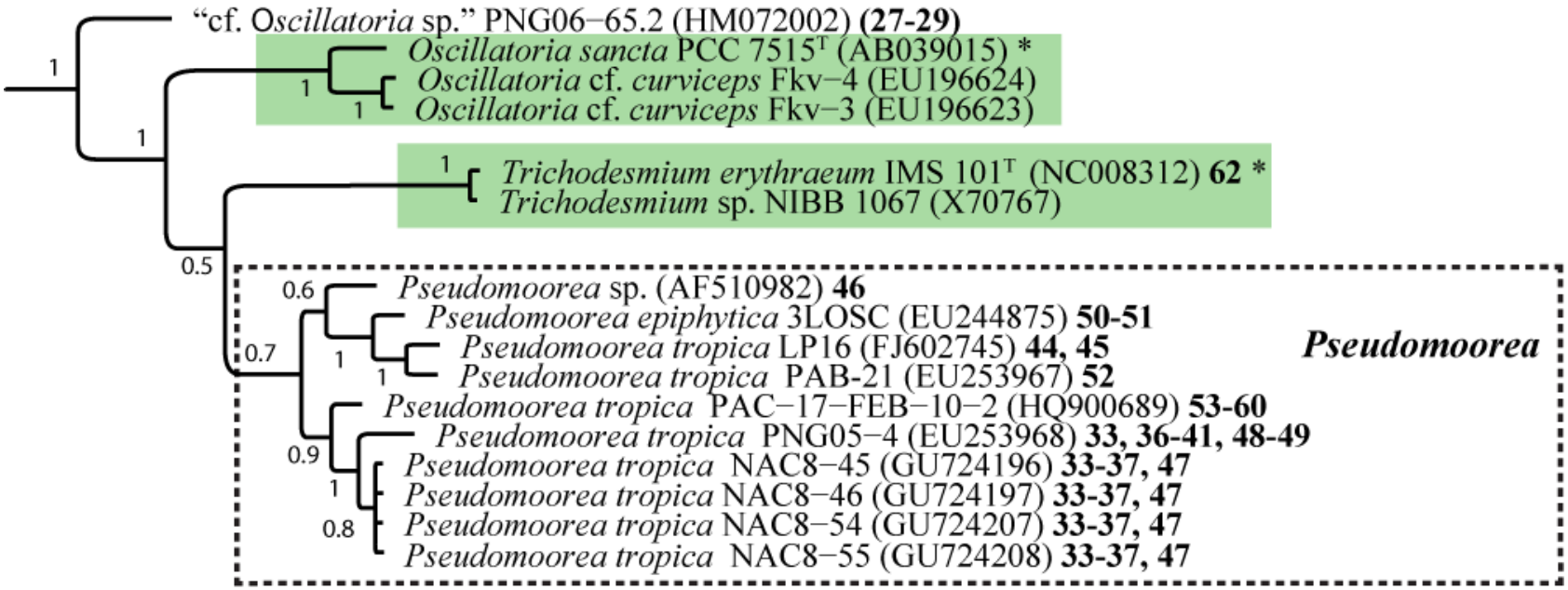


Figure VII.4: Highlight of the *Pseudomoorea* (formerly "tropical marine *Oscillatoria*") clade. The green boxes represent sensu stricto for the genera *Trichodesmium* (reference/type-strain = IMS 101^{T}; GenBank acc. Nr. NC008312) and *Oscillatoria* (reference/type-strain = PCC 7515; GenBank acc. Nr. AB039015). The bold numbers represent natural products isolated from corresponding specimens.

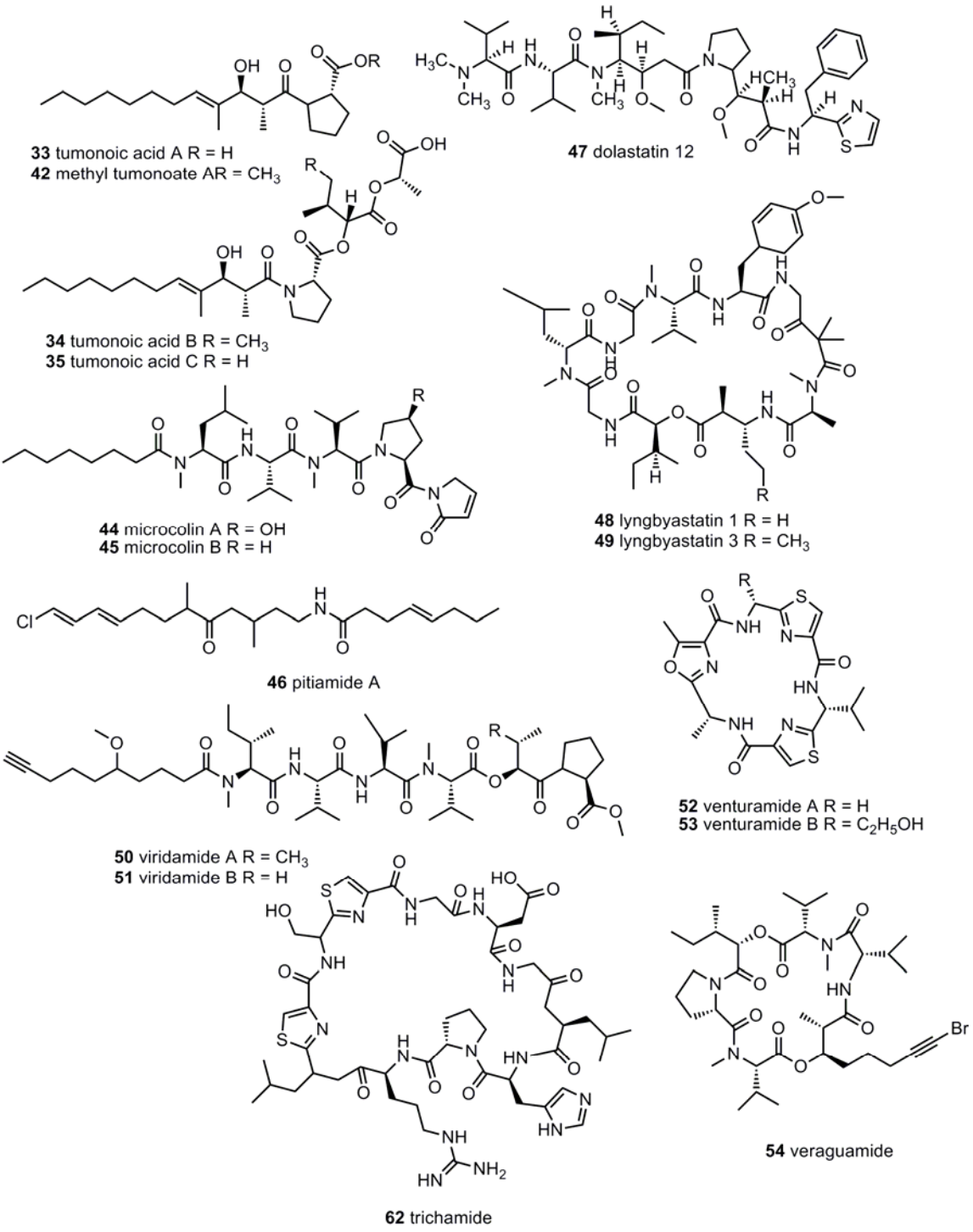


Figure VII.5: Natural products from *Pseudomoorea* gen. nov.

VII.3.3 Natural Products from "Tropical Marine *Symploca*"

"Tropical marine *Symploca*" specimens are here shown to form an independent clade, which is evolutionarily distant (p-distance = 5.4%) from the *Symploca* type/reference-strain (strain PCC 8002; GenBank acc. Nr. AB039021). The fact that this clade of "tropical marine *Symploca*" is also geographically and environmentally distinct from *Symploca* strongly suggests that "tropical marine *Symploca*" needs to be separated as a new genus.

This group of "tropical marine *Symploca*" has yielded compounds such as dolastatin 10 (**24**) (GenBank acc. Nr. AY032933), symplostatin 1 (**25**) and symplostatin 2 **26** (GenBank acc. Nr. AF306497), hoiamides A (**27**), B **(28**) and C **(29**) (GenBank acc. Nr. HM072001). Based on structural similarities to these compounds, belamide A (**32**) is also putatively produce by a strain related to the "tropical marine *Symploca*".

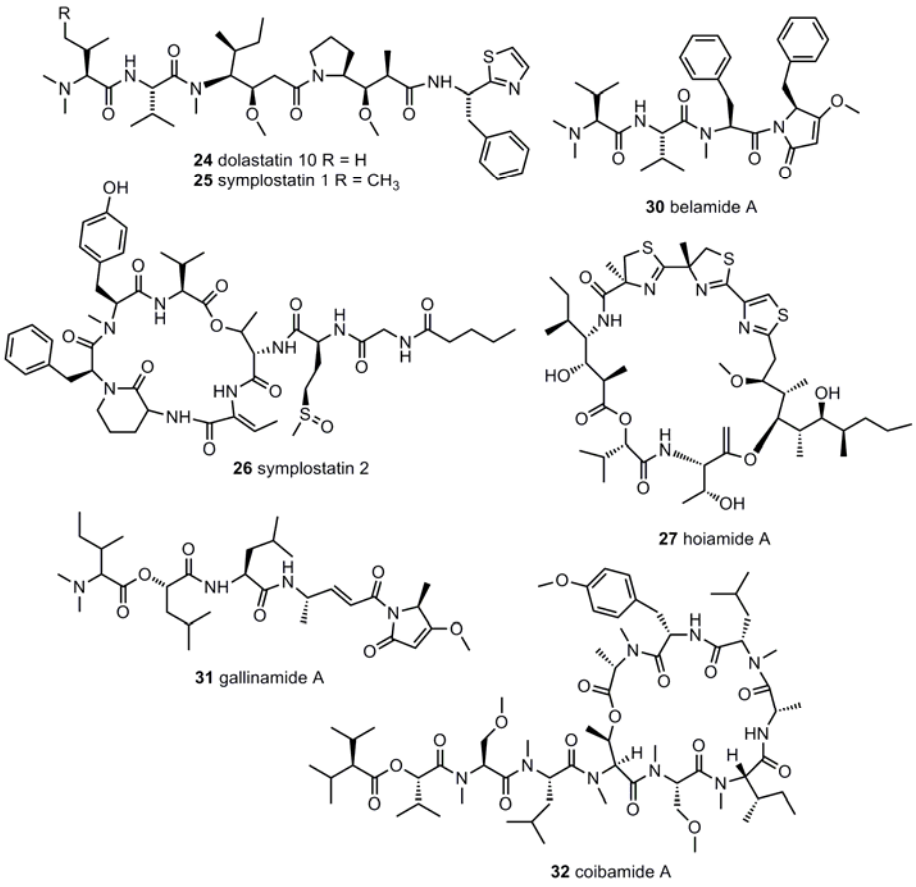


Figure VII.6: Natural products from "tropical marine *Symploca*"

Basal to this group (p-distance = 2.2%) is the coibamide (**31**)–producing cyanobacterium (previously identified as a *Leptolyngbya* sp.) and should, based on the relatively close evolutionary proximity, be included in this group of "tropical marine *Symploca*".

The fact that this specimen lacks the traditional, erect growth-morphology suggests that this could be a misleading character, which could have resulted in other misidentifications. Similar observations have been observed in hoiamide A (**27**)-producing specimens (initially identified as a *Phormidium* sp.) from Papua New Guinea (Pereira *et al*., 2009). The hoiamide A-producer was taxonomically corrected when the same molecule as well as the related analogs hoiamide B (**28**) and C (**29**) were isolated from a microscopically similar "tropical marine *Symploca* sp." (Choi *et al*., 2010). Similar taxonomic confusion probably explains why the cytotoxic symplostatin 1 (**10**) (GenBank acc. Nr. AF306497) was reported from both specimens of *Symploca* (Harrigan *et al*., 1998) and specimens of *Phormidium* (Salvador *et al*., 2010).

There are also natural products reported from specimens of *Schizothrix*, *e.g.* gallineamide A (Linnington *et al*., 2009) that primarily have been identified based on their colony morphologies. The superficial growth morphologies distinguishing *Schizothrix* from *Symploca* have likely resulted in misidentification, which would explain the apparent structural similarities between secondary metabolites isolated from these two groups (Linnington *et al*., 2009; Taori *et al*., 2009).

VIII.3.3 Natural Products from "Tropical Marine *Leptolyngbya*"

The phenotypic characters distinguishing the genera *Phormidium* and *Leptolyngbya* are often unclear, since both of these two groups are defined as fine-filamentous mats (Komárek *et al*., 2005). The phormidolide (**67**)-producing *Phormidium* sp. from Indonesia was here found to be most closely (p-distance = 4% 16S rRNA gene sequence divergence) related to the *Leptolyngbya* strain PCC 7335. Moreover, the growth-forms distinguishing these two genera from *Symploca* can also be misleading, as shown above with the hoiamides (**27-29**)-producing strains and the coibamide (**31**)–producing strains.

It should be noted that both *Phormidium* and *Leptolyngbya* are also polyphyletic groups (Fig. 5). One example of a *Leptolyngbya* lineage that preferably should be erected as a separate generic entity is the tropical marine *Leptolyngbya* (formerly *Phormidium*) *crossbyana*. Recently, this cyanobacterium has been associated with extensive blooms overgrowing reefs on Hawai'i (Smith *et al*., 2008). The potent toxins crossbyanols A-D (**68-71**) (GenBank acc. Nr. GU111930) with pharmaceutical potential were isolated from these blooms (Choi *et al*., 2010). However, this ecologically and biomedically important cyanobacterium lacks proper generic description.

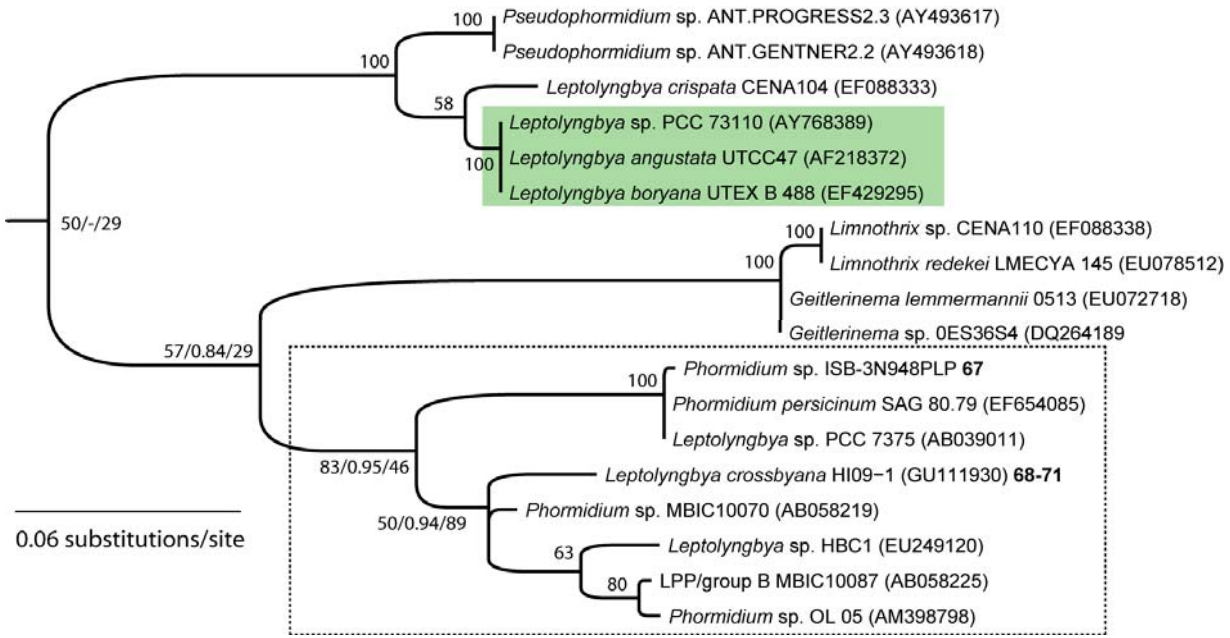


Figure VII.7: Highlight of the "*Leptolyngbya*" lineage. The green boxes represent the sensu stricto for the genus *Leptolyngbya* (reference/type-strain = UTEX B 488^{T}; GenBank acc. Nr. EF429295). The bold numbers represent the natural products isolated from these specimens.

VII.4 Conclusions

During the last three decades, tropical marine cyanobacteria have been a prolific source of natural products. Creative endeavors have been used to explore this novel chemical diversity, but lagging is the recognition and description of the biological diversity responsible for these NPs. Instead of being recognized as unique taxa, NP-producing groups have been identified, with few exceptions, based on classification systems tied to morphospecies of terrestrial and freshwater specimens from temperate regions. This is primarily due to a lack of proper classification systems since tropical marine cyanobacteria have only recently been explored. As a result, our current perspective of the taxonomic origin and distribution of NPs in marine cyanobacteria is extremely incomplete.

In this chapter, I speculate upon the taxonomy of NP-producing marine cyanobacteria based on their evolutionary relatedness. None of these NP-producing strains, with the exception of the trichamide-producing *Trichodesmum erythraeum,* were phylogenetically closely related to the genera to which they were originally assigned. This can be attributed to a lack of description of tropical marine cyanobacteria in general, which represents a relatively novel and unexplored biological frontier.

Over 90% of all marine-derived NPs have been attributed to the five cyanobacterial genera: *Lyngbya*, *Oscillatoria*, *Schizothrix*, *Symploca* and *Phormidium*. In this chapter, I placed some of the cyanobacteria producing the NPs attributed to these five groups in phylogenetic perspective with their reference-strains and speculate that no NPs have actually been isolated from any of these five genera. The common perception of

these genera as NP-rich groups is therefore incorrect for two different reasons: (1) the NPs ascribed to these genera are, in fact, produced by other, often undescribed, groups with phenotypic similarities and (2) taxonomic lumping due to misidentification greatly inflates the perceived NP-richness of these groups.

The true biodiversity of these NP-producing cyanobacteria is revealed using phylogenetic inferences, which is absolutely essential for monitoring and predicting environmentally important HABs as well as sustaining efficient and productive bioactive secondary metabolite discovery efforts for pharmaceutical applications.

VII.5 Acknowledgment

The text of VII, in part, is the manuscript draft to be submitted to an academic journal as it will appear: Engene, N., Gerwick, W. H. Evolution and taxonomic distribution of marine cyanobacterial natural products. The dissertation author was the primary investigator and author of this paper.

VII.6 Appendix - Supporting Information for Chapter VIII

Table of Contents of Appendix

Table VII.2: Taxonomy and secondary metabolite production of *Symploca* specimens.

Figure VII.5: Evolutionary relationships of the credneramide-producing cyanobacterial strains.

Table VII.2: Taxonomy and Secondary Metabolite Production of "*Symploca*" Specimens.

Strain	Original Identification	Lineage	Origin	Secondary Metabolites
PNG-14-JUL-07-6	*Schizothrix*	I	PNG	Kimbeamide
PNG-18-JUL-07-3	*Schizothrix*	I	PNG	Kimbeamide
PNG 4/28/06-1	*Phormidium*	I	PNG	Hoiamide A-B
PNG-5-19-05-7	*Symploca*	V	PNG	Hoiamide C
VP377	*Symploca*	I	Guam	Symplostatin 1
VP642b	*Symploca*	I	Palau	Dolastatin 10
VP642v	*Symploca*	I	Palau	Dolastatin 10
EHu-5-28-07-4	*Symploca*	II	Egypt	N/D
PAP-17-Jul-10-1A	*Lyngbya pencilliformis*	I	Panama	Hermanamide
PAP-17-Jul-10-2	*Symploca*	V	Panama	N/D
PAP-8-SEP-10-4	*Symploca*	stri	Panama	N/D
PAP-9-SEP-10-10	*Symploca*	stri	Panama	N/D
PAC-18-FEB-10-1	*Phormidium*	III	Panama	Bastimolide?
PAC-19-FEB-10-1	*Symploca*	V	Panama	Santacruzamate
PAC-10-03	*Leptolyngbya*	I	Panama	Coibamide
PAB-20-JUN-10-1	*Symploca*	V	Panama	Toronamide
NAC 12/21/08-3	*Symploca*	I	Curacao	Janthielamide

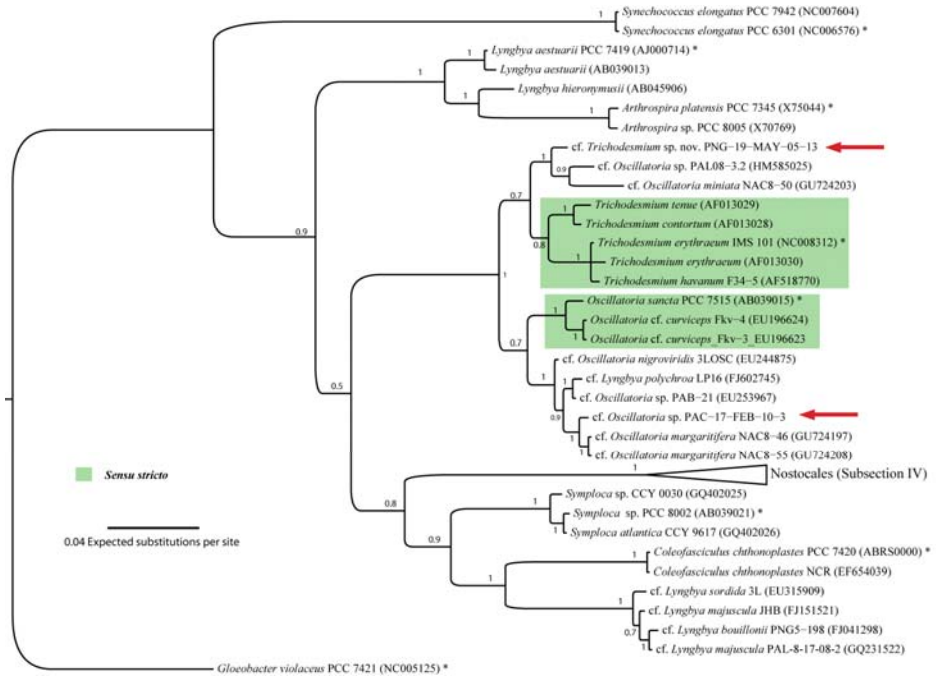


Figure VII.8: Evolutionary tree of the credneramide-producing strain PNG-05/19/05-13. The phylogenetic proximity with the genus *Trichodesmium* (reference/type-strains = *Trichodesmium erythraeum* IMS) suggests that this strain represents a new taxa within this genus, which would also include the related and morphologically-similar strains PAL08-3.2 (GenBank acc. Nr. HM585025) and NAC8-50 (GenBank acc. Nr. GU724203) previously putatively identified as cf. *Oscillatoria* sp. Representative reference-strains obtained from *Bergey's Manual* are highlighted with asterisk (*) and the *sensu stricto* for the genera *Trichodesmium* and *Oscillatoria* are highlighted with green boxes. In addition, to the strain PNG-05/19/05-13, credneric acid was also isolated from the Panamanian strain cf. *Oscillatoria* sp. PAC-17-FEB-10-3 (both credeneric acid-producing strains are highlighted with red arrows). The phylogram is based on SSU (16S) rRNA gene sequences using the bayesian (MrBayes) method and the support values are indicated as posterior probability at the nodes. The specimens are indicated as species, strain, and access number in brackets. The scale bar is indicated at 0.04 expected nucleotide substitutions per site.

VII.6 References

Ainslie, R. D., Barchi Jr., J. J., Kuniyoshi, M., Moore, R. E., Mynderse, J. S. (1985). Structure of malyngamide C. *J Org Chem* 50: 2859-2862.

Anagnostidis, K., Komárek, J. (1988). Modern approach to the classification system of the Cyanophytes 3: Oscillatoriales. Arch. Hydrobiol. 80, *Algological Studies*. 50: 327-472.

Arthur, K., Limpus, C., Balazs, G., Capper, A., Udy, J., Shaw, G., Keuper-Bennett, U., Bennett, P. (2008). The exposure of green turtles (*Chelonia mydas*) to tumour promoting compounds produced by the cyanobacterium *Lyngbya majuscula* and their potential role in the aetiology of fibropapillomatosis. *Harmful Algae* 7: 114-125.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., Weightman, A. (2005). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Env. Microbiol*. 71:7724-7736.

Balunas, M. J., Linington, R. G., Tidgewell, K., Fenner, A. M., Urea, L-D., Della Togna, G., Kyle,. E., Gerwick, W. H. (2010). Dragonamide E, a modified linear lipopeptide from *Lyngbya majuscula* with antileishmanial activity. *J Nat Prod* 73: 60-66.

Berry, J. P., Gantar, M., Perez, M. H., Berry, G., Fernando G. Noriega, F. G. (2008). Cyanobacterial toxins as allelochemicals with potential applications as a algaecides, herbicides and insecticides. *Mar Drugs* 6: 117-146.

Cardellina II, J. H., Marner, F. J., Moore, R. E. (1979). Malyngamide A, a novel chlorinated metabolite of the marine cyanophyte *Lyngbya majuscula*. *J Am Chem Soc* 101: 240-242.

Casamatta, D. A., Johansen, J. R., Vis, M. L., Broadwater, S. T. Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cynaobacteria). *J. Phycol*. 41: 421-438.

Castenholz, R.W. (2001) Phylum BX. Cyanobacteria oxygenic photosynthetic bacteria. Bergey's Manual of Systematic Bacteriology. Springer: New York, pp: 473-553.

Choi, H., Engene, N., Smith, J. E., Preskitt, L. B., Gerwick, W. H. (2010) Crossbyanols A-D, toxic brominated polyphenyl ethers from the Hawai'ian bloom-forming Cyanobacterium *Leptolyngbya crossbyana*. *J Nat Prod*. 23: 517-22.

Choi, H., Pereira, A. R., Cao, Z., Shuman, C. F., Engene, N., Byrum, T., Matainaho, T., Murray, T. F., Mangoni, A., Gerwick, W. H. (2010). The Hoiamides, structurally

intriguing neurotoxic lipopeptides from Papua New Guinea marine cyanobacteria. *J Nat Prod* 73: 1411-1421.

Clark, B. R., Engene, N., Teasdale, M. E., Rowley, D. C., Matainaho, T., Valeriote, F. A., Gerwick, W. H. (2008). Natural products chemistry and taxonomy of the marine cyanobacterium *Blennothrix cantharidosmum*. *J Nat Prod* 71: 1530-1537.

Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., Gerwick, W. H. (2004). Structure and biosynthesis of the Jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula. Chembiol* 11: 817-833.

Engene, N., Coates, R. C., Gerwick, W. H. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46: 591-601.

Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., Gerwick, W. H. (2010). Underestimated biodiversity as a major explanation for the perceived prolific secondary metabolite capacity of the cyanobacterial genus *Lyngbya. Environ Microbiol* 13: 1601-1610.

Esquenazi, E., Coates, C., Simmons, L., Gonzalez, D., Gerwick, W. H., Dorrestein, P. C. (2008). Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges *via* MALDI-TOF imaging. *Mol BioSyst* 4: 562-570.

Fiore, M.F., Sant'Anna, C.L., Azevedo, M.T.D., Komárek, J., Kaštovský, J., Sulek, J., Lorenzi, A.S. (2007). The cyanobacterial genus *Brasilonema*, gen. nov., a molecular and phenotypic evaluation'. *J Phycol* 43: 789-798.

Flatt, P. M., Gautschi, J. T., Thacker, R. W., Musafija-Girt, M., Crews, P., Gerwick, W. H. (2005). Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea* (Lamellodysidea) herbacea and symbiotic cyanobacterium *Oscillatoria spongeliae* by CARD-FISH analysis. *Mar Biol* 147: 761-774.

Geitler, L. (1932). *Cyanophyceae*. In Rabenhorst, L. [Ed.]. Kryptogamen-Flora von Deutschland, Österreich und der Schweiz, vol. XIV. Leipzig: Akademische Verlag (1985 reprint: Königstein: Koeltz Scientific Books) 1027-1068.

Gerwick, W. H., Coates, R. C., Engene, N., Gerwick, L. G., Grindberg, R., Jones, A., Sorrels, C. (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3: 277-284.

Golubic, S., Abed, R. M. M., Palinska, K., Pauillac, S., Chinain, M., Laurent, D. (2010). Marine toxic cyanobacteria: Diversity, environmental responses and hazards. *Toxicon* 56: 836-841.

Graber, M.A., Gerwick, W.H. (1998) Kalkipyrone, a toxic γ-pyrone from an assemblage of the marine cyanobacteria *Lyngbya majuscula* and *Tolypothrix* sp. *J Nat Prod* 61: 677-680.

Graham, L. E. Wilcox, L. W. (2000). Algae. Prentice Hall, Upper Saddle River, NJ. 640 pp.

Gunasekera, S. P., Ross, C., Paul, V. J., Matthew, M., Luesch, H. (2008). Dragonamides C and D, linear lipopeptides from the marine cyanobacterium *Lyngbya polychroa. J Nat Prod* 71: 887-890.

Gutierrez, M., Suyama, T., Engene, N., Wingerd, J., Matainaho, T., Gerwick, W. H. (2008). Apratoxin D, a potent cytotoxic cyclodepsipeptide from Papua New Guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*. *J Nat Prod* 71: 1099-1103.

Harr, K. E., Szabo, N. J., Cichra, M., Phlips, E. J. (2008). Debromoaplysiatoxin in *Lyngbya*-dominated mats on manatees (Trichechus manatus latirostris) in the Florida King's Bay ecosystem. *Toxicon* 52: 385-388.

Harrigan, G. G., Luesch, H., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Paul, V. J., Mooberry, S. M., Corbett, T. H., Valeriote, F. A. (1998). Symplostatin 1: A dolastatin 10 analogue from the marine cyanobacterium *Symploca hydnoides. J Nat Prod* 61: 1075-1077.

Harrigan, G. G., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Park, P. U., Biggs, J., Paul, V. J., Mooberry, S. M., Corbett, T. H., Valeriote, F. A. (1998). Isolation, structure determination, and biological activity of Dolastatin 12 and Lyngbyastatin 1 from *Lyngbya majuscula*/*Schizothrix calcicola* cyanobacterial assemblages. *J Nat Prod* 61: 1221-1225.

Harrigan, G. G., Luesch, H., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Biggs, J., Park, P. U., Paul, V. J. (1999). Tumonoic acids, novel metabolites from a cyanobacterial assemblage of *Lyngbya majuscula* and *Schizothrix calcicola. J Nat Prod* 62: 464-467.

Harrigan, G. G., Luesch, H., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Paul, V. J., Mooberry, S. M., Corbett, T. H., Valeriote, F. A. (1999). Symplostatin 2: a dolastatin 13 analogue from the marine cyanobacterium *Symploca hydnoides. J Nat Prod* 62: 655-658.

Hooper, G. J., Orjala, J., Schatzman, R. C., Gerwick, W. H. (1998). Carmabins A and B, new lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula. J Nat Prod.* 61: 529-33.

Johansen, J. R., Olsen, C. E., Lowe, R. L., Fučíková, K., Casamatta, D. A. (2008). *Leptolyngbya* species from selected seep walls in the Great Smoky Mountains National Park. *Algol Stud* 126: 21-36.

Ito, I., Butler, A. (2005). Structure of synechobactins, new siderophores of the marine cyanobacterium *Synechococcus* sp. PCC 7002. *Limnol Oceanogr* 50: 1918-1923.

Jones, A. C., Monroe, E. A., Podell, S., Hess, W., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R., Yates III, J. R., Reinhardt, R., Kube, M., Burkart, M., Allen, E. E., Dorrestein, P. C., Gerwick, W. H., Gerwick, L. (2011). Genomic Insights Into the Physiology and Ecology of the Marine Filamentous Cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* 108: 8815-8820.

Katoh, K. & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9, 286-298.

Klein, D., Braekman, J-C., Daloze, D., Hoffmann, L., Castillo, G., Demoulin, V. (1998). Lyngbyapeptin A, a modified tetrapeptide from *Lyngbya bouillonii* (Cyanophyceae). *Tetrahedron Lett* 40: 695-696.

Koehn, F. E., Longley, R. E., Reed, J. K. (1992). Microcolins A and B, new immunosuppressive peptides from the blue-green alga *Lyngbya majuscula*. *J Nat Prod* 55: 613-9.

Komárek, J., Anagnostidis, K. (2005). Süsswasserflora von Mitteleuropa 19/2. Elsevier/Spektrum: Heidelberg.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae* 21: 349-375.

Komárek J., Hauer T. (2011): CyanoDB.cz - On-line database of cyanobacterial genera. - Word-wide electronic publication, Univ. of South Bohemia & Inst. of Botany AS CR, http://www.cyanodb.cz

Lamprinou, V., Hernández-Mariné, M., Canals, T., Kormas, K., Economou-Amilli, A., Pantazidou, A. (2011). Two new stigonematalean cyanobacteria: Iphinoe spelaeobios gen. nov., sp. nov. and Loriellopsis cavernicola gen. nov., sp. nov. from Greek and Spanish caves. Morphology and Molecular Evaluation. *Int J Syst Evol Microbiol.* [Epub ahead of print].

Leao, P. N., Pereira, A. R., Liuc, W. T., Ngd, J., Pevznere, P. A., Dorrestein, P. C., Kanig, G. M., Vasconcelosa, M. T. S. D., Vasconcelosa, V. M., Gerwick, W. H. (2010). Synergistic allelochemicals from a freshwater cyanobacterium. *Proc Natl Acad Sci* 107: 11183-8.

Liu, L., Rein, K. S. (2010). New Peptides Isolated from *Lyngbya* Species: A Review. *Mar Drugs* 8: 1817-1837.

Luesch, H., Yoshida, W. Y., R. E. Moore, V. J. Paul., Mooberry, S. L. (2000). Isolation, Structure Determination, and Biological Activity of Lyngbyabellin A from the Marine Cyanobacterium *Lyngbya majuscula. J. Nat. Prod.*63: 611-615.

Luesch, H., Yoshida, W. Y., R. E. Moore, V. J. Paul. Corbett, T. H. (2001). Total structure determination of Apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula. J Am Chem Soc* 123: 5418-5423.

Luesch, H., Pangilinan, R., Yoshida, W. Y., R. E. Moore, V. J. Paul. (2001). Pitipeptolides A and B, new cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula J Nat Prod* 64: 304-307.

Luesch, H., Moore, R. E., Paul, V. J., Mooberry, S. L., Corbett, T. H. (2001). Isolation of Dolastatin 10 from the marine cyanobacterium *Symploca* species VP642 and total stereochemistry and biological evaluation of its analogue Symplostatin 1. *J Nat Prod* 64: 907-910.

Luesch, H., W. Y. Yoshida, R. E. Moore, V. J. Paul. (2002). New apratoxins of marine cyanobacterial origin from Guam and Palau. *Bioorg Med Chem* 10: 1973-1978.

Luesch, H., W. Y. Yoshida, R. E. Moore, V. J. Paul. (2002). Structurally diverse new alkaloids from Palauan collections of the apratoxin-producing marine cyanobacterium *Lyngbya* sp. *Tetrahedron* 58: 7959-7966.

Luesch, H., Williams, P. G., W. Y. Yoshida, R. E. Moore, V. J. Paul. (2002). Ulongamides A−F, new β-Amino acid-containing cyclodepsipeptides from Palauan collections of the marine cyanobacterium *Lyngbya* sp. *J Nat Prod* 65: 996-1000.

Marine literature database (http://www.chem.canterbury.ac.nz/marinlit/ marinlit.shtml)

Marner, F-J., Moore, R. E., Hirotsu, K., Clardy, J. (1977). Majusculamides A and B, two epimeric lipodipeptides from Lyngbya majuscula Gomont. *J Org Chem* 42: 2815-2819.

Márquez, B., Verdier-Pinard, P., Hamel, E., Gerwick, W. H. (1999). Curacin D, an antimitotic agent from the marine cyanobacterium *Lyngbya majuscula*. *Phytochem* 49: 2387-2389.

Márquez, B., Watts, K. S., Yokochi, A., Roberts, M. A., Verdier-Pinard, P., Jimenez, J. I., Hamel, E., Scheuer, P. J., Gerwick, W. H. (2002). Structure and absolute stereochemistry of Hectochlorin, a potent stimulator of actin assembly. *J Nat Prod* 65: 866-871.

McPhail, K. L., Correa, J., Linington, R. G., González, J., Ortega-Barría, E., Capson, T. L., Gerwick, W. H. (2007). Antimalarial linear lipopeptides from a Panamanian strain of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 70: 984-988.

Medina, R. A., Goeger, D. E., Hills, P., Mooberry, S. L., Huang, N., Romero, L. I., Ortega-Barria, E., Gerwick, W. H., McPhail, K. L. (2008). Coibamide A, a potent antiproliferative cyclic depsipeptide from the Panamanian marine cyanobacterium *Leptolyngbya* sp. *J Am Chem Soc* 130: 6324-6325.

Mevers, E., Liu, W-T., Engene, N., Byrum, T., Dorrestein, P. C., Gerwick, W. H. (2011). Cytotoxic Veraguamides, alkynyl bromide containing cyclic depsipeptides from an unusual marine cyanobacterium cf. *Oscillatoria margaritifera. J Nat Prod* 74: 928-936.

Milligan, K. E., Marquez, B. L., Williamson, R. T., Gerwick, W. H. (2000). Lyngbyabellin B, a toxic and antifungal secondary metabolite from the marine cyanobacterium *Lyngbya majuscula. J Nat Prod* 63: 1440-1443.

Moore, R. E., Blackman, A. J., Cheuk, C. E., Mynderse, J. S., Matsumoto, G. K., Clardy, J., Woodard, R. W., Craig, J. C. (1984). Absolute stereochemistries of the aplysiatoxins and oscillatoxin A. *J Org Chem* 49: 2484-2489.

Moore, R. E. (1982). Toxins, anticancer agents, and tumor promoters from marine prokaryotes. *Pure and applied chemistry* 54: 1919.

Montaser, R., Abboud, K. A., Paul, V. J., Luesch, H. (2011). Pitiprolamide, a proline-rich Dolastatin 16 analogue from the marine cyanobacterium *Lyngbya majuscula* from Guam. *J Nat Prod*74: 109–112.

Nagle, D. G., P. U. Park, V. J. Paul. (1997). Pitiamide A, a new chlorinated lipid from a mixed marine cyanobacterial assemblage. *Tetrahedron Lett* 38:6969-6972.

Nagle, D. G., Zhou, Y. D., P. U. Park, V. J. Paul. Rajbhandari, I., Duncan, C. J., Pasco, D. S. (2000). A new indanone from the marine cyanobacterium *Lyngbya majuscula* that inhibits hypoxia-induced activation of the VEGF promoter in Hep3B cells. *J Nat Prod* 63: 1431-3.

Nogle, L. M., Marquez, B. L., Gerwick, W. H. (2003). Wewakazole, a novel cyclic dodecapeptide from a Papua New Guinea *Lyngbya majuscula. Org Lett* 5: 3-6.

Nunnery, J. K., Mevers, E., Gerwick, W. H. (2010). Biologically active secondary metabolites from marine cyanobacteria. *Curr Opin Biotech* 21: 1-7.

Nylander, J. A. A., Wilgenbusch, J. C., Warren, D. L., Swofford, D. L. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics (Oxf.)* 15: 581-583.

Oren, A. (2004). A proposal for further integration of the cyanobacteria under the Bacteriological Code. *Int J Syst Evol Microbiol* 5: 1895-1902.

Orjala, J., Gerwick, W. H. (1994) Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula. J Nat Prod* 59: 427-430.

Orjala, J., Nagle, D. G., Hsu, V. L., Gerwick, W. H. (1995) Antillatoxin: an exceptionally ichthyotoxic cyclic lipopeptide from the tropical cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 117: 8281-8282.

Orjala, J., Gerwick, W. H. (1996). Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula. J Nat Prod* 59: 427-430.

Orjala, J., Gerwick, W.H. (1997). Two quinoline alkaloids from the Caribbean cyanobacterium *Lyngbya majuscula*. *Phytochem* 45: 1087-1090.

Osborne, N., Seawright, A., Shaw, G. (2008). Dermal toxicology of *Lyngbya majuscula*, from Moreton Bay, Queensland, Australia. *Harmful Algae* 7: 584-589.

Paul, V. J., Thacker, W., Banks, K., Golubic, S. (2005). Benthic cyanobacterial bloom impacts the reefs of South Florida (Broward County, USA). *Coral Reefs* 24: 693-697.

Pereira, A., Cao, Z., Murray, T. F., Gerwick, W. H. (2009). Hoiamide A, a sodium channel activator of unusual architecture from a consortium of two Papua New Guinea cyanobacteria. *Chem Biol* 16: 893-906.

Pereira, A. R., McCue, C., Gerwick, W. H. (2010a). Cyanolide A, a glycosidic macrolide with potent molluscicidal activity from the Papua New Guinea cyanobacterium *Lyngbya bouillonii. J Nat Prod* 73: 217-220.

Pereira, A. R., Cao, Z., Engene, N., Soria-Mercado, I. E., Murray, T. F., Gerwick, W. H. (2010b). Palmyrolide A, an unusually stabilized neuroactive macrolide from Palmyra atoll cyanobacteria. *Org Lett* 12: 4490-4493.

Pereira, A. R., Etzbach, L., Engene, N., Müller, R., Gerwick, W. H. (2011). Unusual molluscicidal chemotypes from a new Palmyra Atoll cyanobacterium. *J Nat Prod* 74: 1175-1181.

Pointing, S. B., Chan, Y., Lacap, D. C., Lau, M. C. Y., Jurgens, J. A., Farrell, R. L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci USA* 106: 19964-19969.

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256.

Rasoulouniriana, D., Siboni, N., Ben-Dov, E., Kramarsky-Winter, E., Loya, Y., Kushmaro, A. (2009). *Pseudoscillatoria coralii* gen. nov., sp. nov., a cyanobacterium associated with coral black band disease (BBD). *Dis Aquat Organ* 16: 91-6.

Ronquist, F., Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxf)* 12: 1572-1574.

Salvador L. A., Paul V. J., Luesch, H. (2010). Caylobolide B, a macrolactone from symplostatin 1-producing marine cyanobacteria *Phormidium* spp. from Florida. *J Nat Prod* 24: 1606-1609.

Sanchez, L. M., Lopez, D., Vesely, B. A., Della Togna, G., Gerwick, W. H., Kyle, D. E., Linington, R. G. (2010). Almiramides A-C: discovery and development of a new class of leishmaniasis lead compounds. *J Med Chem* 53: 4187-4197.

Sharp, K., Arthur, K. E., Gu, L., Ross, C., Harrison, G., Gunasekera, S. P., Meickle, T., Matthew, S., Luesch, H., Thacker, R. W., Sherman, D. H., Paul, V. J. (2009). Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* (Cyanobacteria: *Oscillatoriales*) from reefs of Southeast Florida. *Appl Env Microbiol* 75: 2879-2888.

Simmons, T. L. Gerwick, W. H. (2008). Anticancer drugs of marine origin, p. 431-452. *In* Oceans and human health. Elsevier, Academic Press, Burlington, Mass.

Simmons, T. L., Coates, R. C., Clark, B. R., Engene, N., Gonzalez, D., Esquenazi, E., Dorrestein, P. C., Gerwick. W. H. (2008) Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proc Natl Acad Sci USA* 105: 4587-4594.

Simmons, T. L., McPhail, K. L., Ortega-Barria, E., Mooberry, S. L., Gerwick, W. H. (2006). Belamide A, a new antimitotic tetrapeptide from a Panamanian marine cyanobacterium. *Tetrahed Lett* 47: 3387-3390.

Sitachitta, N., Márquez, B. L., Williamson, R. T., Rossi, J., Roberts, M. A., Gerwick, W. H. Nguyen, V-A., Willis, C. L. (2000). Biosynthetic pathway and origin of the chlorinated methyl group in Barbamide and Dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula. Tetrahed* 56: 9103-9113.

Smith, J. E., Kuwabara, J., Coney, J., Flanagan, K., Beets, J., Brown, D., Stanton, F., Takabayashi, M., Duplessis, S., Griesemer, B. K., Barnes, S., Turner, J. (2008). An unusual cyanobacterial bloom in Hawai`i. *Coral Reefs* 27: 851-851.

Stanier, R. Y., Sistrom, W. R., Hansen, T. A. (1978). Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. *Int J Syst Bacteriol* 28: 335-336.

Stanley, S.M. *Earth System History*, 2nd ed.; WH Freeman & Co.: New York, NY, USA, 2004; p. 263.

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 Biol* 33: 1419-1428.

Sudek, S., Haygood, M. G., Youssef, D. T. A., Schmidt, E. W. (2006). Structure of Trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. *Appl Env Microbiol* 72: 4382-4387.

Sumina, E. L. (2006). Behavior of filamentous cyanobacteria in laboratory culture. *Microbiol* 75: 459-464.

Swingley, W. D., Blankenship, R. E., Raymond, R. (2008). Integrating markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol Biol Evol* 25: 643-654.

Tan, L. T. (2010). Filamentous tropical marine cyanobacteria: a rich source of natural products for anticancer drug discovery. *J Appl Phycol* 5: 659-676.

Taniguchi, M., Nunnery, J., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., Gerwick, W. H. (2010). Palmyramide A, a cyclic depsipeptide from a Palmyra atoll collection of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73: 393-398.

Taori K., Liu Y., Paul, V. J., Luesch, H. (2009). Combinatorial strategies by marine cyanobacteria: symplostatin 4, an antimitotic natural dolastatin 10/15 hybrid that synergizes with the coproduced HDAC inhibitor largazole. *Chembiochem* 6: 1634-9.

Thacker, R. W., Paul, V. J. (2004). Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales). *Appl Environ Microbiol* 70: 3305-12.

Tidgewell, K., Clark, B.R., Gerwick, W.H. (2010) The natural products chemistry of cyanobacteria. Compr Nat Prod Chem Pergamon Press: Oxford.

Tidgewell, K., Engene, N., Byrum, T., Media, J., Valeriote, F. A., Gerwick, W. H. (2010). Diversification of a Modular Natural Product Pathway: Production of Apratoxins F and G, Two Cytotoxic Cyclic Depsipeptides from a Palmyra Collection of *Lyngbya bouillonii*. *ChemBioChem* 11: 1458-1466.

Villa, F. A., Gerwick, L. (2010). Marine natural product drug discovery: leads for treatment of inflammation, cancer, infections, and neurological disorders. *Immunopharmacol Immunotoxicol* 32: 228-237.

Whitton, B. A., Potts, M. (2000). The ecology of Cyanobacteria: the diversity in time and space. Kluwer Academic Publishers, 669 pp.

Williams, P. G., Moore, R. E., Paul, V. J. (2003). Isolation and structure determination of Lyngbyastatin 3, a Lyngbyastatin 1 homologue from the marine cyanobacterium *Lyngbya majuscula.* Determination of the configuration of the 4-Amino-2,2-dimethyl-3-oxopentanoic acid unit in Majusculamide C, Dolastatin 12, Lyngbyastatin 1, and Lyngbyastatin 3 from cyanobacteria. *J Nat Prod* 66: 1356-1363.

Williamson, R. T., Boulanger, A., Vulpanovici, A., Roberts, M. A., Gerwick, W. H. (2002). Structure and absolute stereochemistry of Phormidolide, a new toxic metabolite from the marine cyanobacterium *Phormidium* sp. *J Org Chem* 7: 7927-7936.

Wu, M., Milligan, K. E., Gerwick, W. H. (1997) Three new malyngamides from the marine cyanobacterium *Lyngbya majuscula*. *Tetrahed* 53: 15983-15990.

Wu, M., Okino, T., Nogle, L. M., Marquez, B. L., Williamson, R. T., Gerwick, W. H. *et al.,* (2000) Structure, synthesis, and biological properties of kalkitoxin, a novel neurotoxin from the marine cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 122: 12041-120.

Chapter VIII

Conclusions

VIII.1 Cyanobacterial House-keeping Genes as Biosynthetic Predictors

In recent years, common housekeeping genes have increasingly been applied to infer evolutionary histories and relationships among cyanobacteria. Although the conserved SSU rRNA gene has been well-recognized as a robust tool for such evolutionary inferences, its phylogenetic capability was further examined for marine cyanobacteria in Chapter III and Chapter IV by thorough comparison of phylograms obtained by the SSU rRNA gene with those of other housekeeping genes. These phylogenetic comparisons showed that the evolutionary interpretations of the SSU rRNA gene were highly consistent with other housekeeping genes, and consequently applicable for our phylogenetic investigations of tropical marine cyanobacteria. Additionally investigated was the presence of paralogous copies of the ribosomal genes within the same genome, as their variable sequences could skew phylogenetic inferences. These studies indicated that cyanobacterial genomes contain multiple and often variable ribosomal genes. However, our result also concluded that the degree of intra-genomic gene heterogeneity was relatively low and had little impact on phylogenetic interpretations. Therefore, we concluded that the SSU rRNA genes were phylogenetically representative and could be reliably utilized for taxonomic inference.

Both the Botanical and the Bacteriological Code are today largely in agreement that phylogenetics provide the necessary framework for establishing informative and reliable taxonomic systems. The evolutionary insights obtained from inferences of both house-keeping genes and the SSU rRNA genes were successfully utilized to rearrange our taxonomic systems for cyanobacteria. Herein, these new systems were successfully

tested in Chapter II to predict the biosynthetic capacity of marine cyanobacteria to produce specific groups of natural products based on their evolutionary relationships to other known NP-producers. The underlying principle of the case study presented was further strengthened by the fact that the original producers of the tumonoic acids had been taxonomically incorrectly identified, and thus their NP-potential could not have been forecasted based on their taxonomic identification. Moreover, the previously identified tumonoic acid-producing strains had been obtained from Guam and Papua New Guinea, respectively. Thus, production of tumonoic acids by the specimens from Curaçao could not have been predicted based on the geographic locations of the specimens.

The trend of a correlation between phylogenetic position and production of specific secondary metabolites was consistent with other examples throughout the thesis work. It is conceivable that genes involved in the biosynthesis of these natural products are being transferred horizontally between different groups of marine cyanobacteria or other microorganisms. However, based on the results of this thesis, lateral transfers appear to be rare evolutionary events and the primary mode of evolution of these biosynthetic genes occurs vertically. A vertical evolution of natural products underscores the rationale of phylogenetics as a biosynthetic predictor and clearly highlights the importance of a proper taxonomic understanding of NP-producing microorganisms. The observed correlation between phylogenetic position and production of specific secondary metabolites also suggests that these compounds could act as unique taxonomic markers for identification of marine cyanobacteria. However, correct taxanomic systems must be established for marine cyanobacteria prior to investigating the chemotaxonomic robustness of these secondary metabolites.

VIII.2 True Metabolic Origin of Natural Products

Among the many natural products and NP-producers investigated in this thesis work, a single example was found of incongruence between phylogenetic position and natural product production. The vinyl chloride-containing metabolite credneric acid was isolated from two cyanobacterial specimens belonging to the two different taxonomic groups *Trichodesmium* and *Oscillatoria*, respectively (Figure VII.5). The evolutionary distance between the two credneric acid-producing cyanobacteria and the fact that other related specimens did not produce this secondary metabolite obviously point to the possibility that the biosynthetic genes encoding this metabolite had recently been transferred horizontally. However, credneric acid was later reported from the tropical marine heterotrophic bacterium *Rhodopirellula baltica* isolated from the surface of eukaryotic macro algae. Thus, a speculative but likely possibility is that associated heterotrophic bacteria would be responsible for the production of the credneric acids isolated from the marine cyanobacteria.

Many secondary metabolites isolated from marine cyanobacteria have structural resemblance with those of heterotrophic bacteria. Because of the possibility that heterotrophic bacteria are responsible for the production of compounds attributed to cyanobacteria, a major aspect of this thesis work has been to determine the degree and location of heterotrophic bacteria associated with marine NP-producing cyanobacteria. All samples analyzed by scanning electron microscopy contained rich and diverse micro flora of heterotrophic bacteria and other microorganisms, which were localized to the surfaces of the thick polysaccharide sheaths enclosing the cyanobacterial filaments. In

our efforts to investigate the metabolic origin of cyanobacterial secondary metabolites, single-cells were separated by micromanipulation and genomic DNA was obtained from these single-cells by multiple displacement amplification (MDA). Sequencing analysis of cyanobacterial single-cell MDA-genomes revealed the presence of at least the curacin, barbamide, and carmabin biosynthetic pathways. This was matched by analysis of intact single-cells by MALDI-TOF-MS which demonstrated that these secondary metabolites were physically present in the cyanobacterial cells. To our knowledge, this is the first direct combined genetic and chemical proof of secondary metabolite biosynthesis on a microbial single-cell level. Despite the correlation found in this study between cyanobacterial biosynthetic capacity and NP-production, it is still possible that many other natural products, such as those that do not manifest in culture, are produced by heterotrophic bacteria associated with the cyanobacteria in the wild.

In addition, our studies show that cyanobacterial mats typically are composed of multiple different types of cyanobacteria and eukaryotic algae. These micromanipulation methods described above have provided excellent opportunities to dissect algal assemblages and analyze individual filaments. The combined MDA and MALDI-TOF-MS have allowed correlation of specific secondary metabolites and gene sequences with cyanobacterial components of these assemblages. As these dissection and analysis methods have provided essential insight into the metabolic origin of certain secondary metabolites, we hope they become standard in the field of natural products science.

VIII.3 Underestimated Biodiversity as an Explanation for the Skewed Perception of the Taxonomic Distribution of Natural Products in Marine Cyanobacteria

By far the most chemically prolific genus of marine cyanobacteria has been *Lyngbya*, as almost two thirds of all isolated cyanobacterial compounds are reported to be from various collections of this organism. One of the more remarkable examples of the extraordinary secondary metabolite diversity initially found in marine *Lyngbya* is the chemo-diversity found in *L. majuscula* populations around the Caribbean island of Curaçao. At least sixteen novel molecules have been isolated from *L. majuscula* collections along the leeward half of this small island with less than 40 miles of coast line. In Chapter V, I investigated the correlation of biosynthetic capacities and actual production of secondary metabolites with the phylogenetic relationships of *L. majuscula* from geographically and environmentally disperse sites around Curaçao. This study concluded that polyphyly of morphologically similar cyanobacteria was the major explanation for the perceived chemical richness of the genus *Lyngbya*. The phylogenetic investigations of *Lyngbya* reveal that this group is a highly polyphyletic group and composed of numerous evolutionarily distantly related lineages. These different phylogenetic lineages are geographically and ecologically distinct from the genus *sensu stricto*, a brackish water strain from Northern Europe, and are only lumped together on the basis of similar morphological characters.

In addition to *Lyngbya*, all major NP-rich groups of marine cyanobacteria, including *Oscillatoria*, *Symploca*, *Leptolyngbya*, *Synechococcus* and *Phormidium*, were found in Chapter VII to be polyphyletic and composed of multiple evolutionarily unrelated groups. The unfortunate consequence of this lumping together has been an underestimation of the true biodiversity of cyanobacteria and a poor understanding of the

taxonomic distribution of NPs in cyanobacteria. Ironically, no secondary metabolites have, from our results, been isolated from the *sensu stricto* of any of the aforementioned groups or from specimens truly related to the *sensu stricto*.

VIII.4 Understanding and Recognizing Novel Cyanobacterial Biodiversity

Undoubtedly, tropical marine cyanobacteria have provided a prolific source of natural products and exploration of unique habitats in new geographic locations has been used to explore this novel chemical diversity. However, the recognition and description of the biological diversity responsible for these natural products has been completely overlooked. Instead of being recognized as unique taxa, NP-producing groups have been identified, with few exceptions, based on classification systems tied to morphospecies of terrestrial and freshwater specimens from temperate regions. This is primarily due to a lack of proper classification systems since tropical marine cyanobacteria have only recently been explored. As a result, our current perspective of the taxonomic origin and distribution of NPs in marine cyanobacteria is incomplete.

In Chapter VII of this thesis, I place many of the NP-producing cyanobacteria in phylogenetic perspective with their reference-strains. Ironically, none of these NP-producing strains, with the exception of the trichamide-producing *Trichodesmum erythraeum,* were phylogenetically closely related to the genera that they have been assigned. The NPs ascribed to these genera are, in fact, produced by other often undescribed groups that are phenotypically similar.

The phylogenetic inference of these NP-producing cyanobacteria uncovers true biodiversity. This is absolutely essential for monitoring and predicting environmentally important HABs as well as sustaining efficient and productive bioactive secondary metabolite discovery efforts for pharmaceutical applications. In my efforts to provide taxonomic clarity as well as to better guide future natural product drug discovery investigations, I revised the NP-rich genus *Lyngbya* and proposed the new genus *Moorea* in Chapter VI. It is, however, important to stress that this taxonomic proposals represent recognition of new biodiversity rather than simply changes in nomenclature. More importantly, these two new groups only represent a small proportion of NP-producing marine cyanobacteria in need of taxonomic revision and description. However, as phylogeny-based systems are continually developed to facilitate the identification and classification of tropical marine cyanobacteria, allowing for more accurate correlations of NPs with their producing groups, we expect the progress of more targeted natural products discovery programs.