**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**
INSTITUTO DE BIOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E BIOLOGIA EVOLUTIVA - PPGBBE

**TESE DE DOUTORADO**

# Genômica estrutural e funcional de turfs e cianobactérias do Banco de Abrolhos (Bahia, Brasil)

**JULINE MARTA WALTER**

Rio de Janeiro
2016

# Genômica estrutural e funcional de turfs e cianobactérias do Banco de Abrolhos (Bahia, Brasil)

**JULINE MARTA WALTER**

Tese apresentada como requerimento parcial ao título de Doutor pelo Programa de Pós Graduação em Biodiversidade e Biologia Evolutiva da Universidade Federal do Rio de Janeiro.

Aprovada em: 19 de dezembro de 2016.

Prof. Dr. Fabiano Lopes Thompson _____

Profª. Drª. Sigrid Neumann Leitão _____

Prof. Dr. Ronaldo Bastos Francini-Filho _____

Prof. Dr. Carlos Eduardo Guerra Scharago _____

Prof. Dr. Osmindo Rodrigues Pires Júnior _____

Profª. Drª. Laura Silvia Bahiense da Silva Leite _____

Rio de Janeiro

2016

Programa de Pós-graduação em
**Biodiversidade & Biologia Evolutiva**
Instituto de Biologia - Universidade Federal do Rio de Janeiro

# C E R T I F I C A D O

Certificamos, para os devidos fins que **Juline Marta Walter**, completou sua carga horária de 450 horas exigidas neste curso e defendeu sua Tese de **Doutorado**, intitulada "Genômica estrutural e funcional de turfs e cianobactérias do banco de Abrolhos (Bahia, Brazil)", nesta data, tendo sido "Aprovada" fazendo jus ao grau de Doutora em Biodiversidade e Biologia Evolutiva, aguardando o registro de seu diploma.

Rio de Janeiro, 19 de dezembro de 2016.

Profa. Dra. Daniela M. Takiya
Coordenadora (SIAPE 1654008)
Programa de Pós-graduação em
Biodiversidade e Biologia Evolutiva
Instituto de Biologia, UFRJ

Programa de Pós-Graduação em Biodiversidade e Biologia Evolutiva
Instituto de Biologia – Universidade Federal do Rio de Janeiro
Prédio do CCS, Rua Prof. Rodolpho Paulo Rocco, S/N
Cidade Universitária - Rio de Janeiro, RJ, Brasil
CEP: 21941-617
ppgbbe@biologia.ufrj.br

W151g

Walter, Juline Marta
Genômica estrutural e funcional de turfs e cianobactérias do Banco de Abrolhos (Bahia, Brasil) / Juline Marta Walter. -- Rio de Janeiro, 2016.
185 f.

Orientador: Fabiano Lopes Thompson.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Instituto de Biologia, Programa de Pós Graduação em Biodiversidade e Biologia Evolutiva, 2016.

1. Microbiologia. 2. Ecologia Marinha. 3. Genômica. 4. Bioinformática. 5. Metagenômica. I. Thompson, Fabiano Lopes, orient. II. Título.

# Structural and functional genomics of turfs and cyanobacteria from the Abrolhos Bank (Bahia, Brazil)

**JULINE MARTA WALTER**

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor in the Graduate Program of Biodiversity and Evolutionary Biology at Federal University of Rio de Janeiro.

Adviser:
Prof. Dr. Fabiano Lopes Thompson

Doctoral Committee:
Prof. Dr. Ronaldo Bastos Francini-Filho (UFPB)
Prof. Drª. Sigrid Neumann Leitão (UFPE)
Prof. Dr. Osmindo Rodrigues Pires Júnior (UnB)
Prof. Dr. Carlos Eduardo Guerra Schrago (UFRJ)
Prof. Drª. Laura Silvia Bahiense da Silva Leite (COPPE, UFRJ)

Alternate Members:
Profª. Drª. Cristiane Carneiro Thompson (UFRJ)
Profª. Drª. Michelle Regina Lemos Klautau (UFRJ)

Rio de Janeiro
2016

This thesis is dedicated to my parents

# ACKNOWLEDGMENTS

I want to thank the many people who contributed with this thesis, particularly my adviser Dr. Fabiano L. Thompson, who gives me this unique opportunity, the partnership and knowledgement into the coral microcosms. I also would like to thank you for the hours you spent meticulously correcting chapters and papers. I also thank the Phyto research group, particularly Dr. Paulo S. Salomon, for his continuous assistance and enthusiastic encouragements. Special thanks are offered to Dr. Rodrigo L. Moura for inspiration and introducing me to this fascinating ecosystem, helping me in the fieldtrip. I extend special thanks to the crewmembers behind the *catamarã* for their hard work: Zá, Maurício, and Van. Thank to undergraduate students who spend many hours assisting in the isolation of the clonal cultures. Thank to Paulo Iiboshi for the maintenance of the cultures. I thank Dra. Cristiane C. Thompson for being always there, with her supportive and pleasant atmosphere. Many thanks to all my friends and colleagues at the Micro Lab: Adriana Fróes (Dri), Ana Carolina Soares, Ana Paula Moreira, Angélica, Arthur Lima, Bruno Sérgio, Diogo Tschoeke, Felipe Coutinho, Gizele Garcia, Giselle Cavalcanti, Gabriela Calegario, Graciela Dias, Gustavo Pitta, Lívia Vidal, Louise Oliveira, Luciana Leomil, Luciana Reis, Mariana Campeão, Pedro Meirelles, Raphael Silva, Rosa Koko, and many others. Special thanks go to my darling Lílian J. K. Mayerhofer. I could not have completed this project without her constant love and encouragements. Love you! I would also like to thank my Brazilian, Dutch and Italian colleagues and friends from Utrecht University for their high quality discussion during the time I spend there. The fruitful discussions with Prof. Dr. Jean Swings were highly acknowledged. Suggestions

from members of the Examining Evaluation greatly improved the manuscripts and thus, this Thesis. Thank to the contribution of Rede Abrolhos (Abrolhos Network), funded by the National Biodiversity System (SISBIOTA). I would like to thank the Abrolhos Marine National Park to facilitate the access to the reefs. Thank to the Biodiversity and Evolutionary Biology Graduate Program at Federal University of Rio de Janeiro. Thank to all PPGBBE Committee especially during the year (2014) that I was Vice-Representative of student body. Financial support was provided in part by the CAPES, and from the CNPq through grant 207751. Grant from the CNPq enabled me to spend one year (2015-2016) visiting a Bioinformatics laboratory in The Netherlands. Special thanks are offered to Dr. Bas Dutilh for their time, patience and advices while I was pretty new in the bioinformatics field. Last but not least, thanks to my whole family, been always the biggest support.

Rio de Janeiro, November 2016.

"With every drop of water you drink,

Every breath you take,

You're connected to the sea.

No matter where on Earth you live."

Sylvia Earle

- Marine biologist. She was the first woman chief scientist of the U.S. National Oceanic and Atmospheric Administration. -

# LIST OF FIGURES

# LIST OF TABLES

## Chapter II



## Chapter III

# ABBREVIATIONS AND SIMBOLS

| | |
|---|---|
| ºC | degrees Celsius |
| % | percent |
| $A_{260}$ | absorbance at 260 nanometres |
| $A_{280}$ | absorbance at 280 nanometres |
| bp | base pair(s) |
| DNA | deoxyribonucleic acid |
| g | gram(s) |
| Gb | giga-base pairs |
| h | hour(s) |
| ID | identification |
| Kb | kilo-base pairs |
| log | logarithm to the base 10 |
| M | molar |
| Mb | mega-base pairs |
| min | minute(s) |
| ml | millilitre(s) |
| mM | millimolar |
| ng | nanogram(s) |
| NGS | next generation sequencing |
| no. | number |
| nt | nucleotide |
| PCoA | principal coordinate analysis |
| PCR | polymerase chain reaction |
| RNA | ribonucleic acid |
| rpm | revolutions per minute |
| rRNA | ribosomal ribonucleic acid |
| s | second(s) |
| S.D. | stardard deviation |

| | |
|---|---|
| spp. | Species |
| TBE | tris-borate EDTA buffer |
| TE | tris EDTA buffer |
| µl | microliter(s) |
| µM | micromolar |

# PUBLICATIONS AND MANUSCRIPTS

Walter JM, Tschoeke DA, de Oliveira L, Leomil L, Tenório M, Meirelles, PM, Valle R, Coutinho R, Salomon PS, Thompson CC, Thompson FL. Taxonomic and functional metagenomic signature of turfs in the Abrolhos reef system (Brazil). **Plos One**. 11(8): e0161168, 2016. doi:10.1371/journal.pone.0161168

Walter JM, Coutinho FH, Dutilh BE, Thompson CC, Thompson FL. Linking Ecogenomics and Taxonomy of Cyanobacteria. **Frontiers in Microbiology**, *Under major revision* (*special research topic: "Cyanobacterial Blooms in a Changing World").*

Walter JM, Coutinho, FH, de Oliveira L, Leomil L, Hargreaves PI, Silva, BS, Fistarol GO, Miyashita K, Maruyama F, Salomon PS, Thompson CC, van Verk MC, Dutilh BE, Thompson FL. A novel, widespread turf-forming cyanobacterium (*Acrophormium turfae*) from the Abrolhos reef system has a vast secondary metabolite genomic repertoire. **The Isme Journal***, Under major revision.*

Silveira CB, Cavalcanti GS, Walter JM, Silva-Lima AW, Dinsdale E, Bourne D, Thompson CC, Thompson FL. Microbial processes driving coral reef organic carbon flow. **FEMS Microbiology Reviews**. fux018, 2017. doi:10.1093/femsre/fux018

Silva-Lima A, Walter JM, Garcia GD, Ramires N, Ank G, Meirelles PM, Nobrega AF, Silva-Neto ID, Moura RL, Salomon OS, Thompson CC, Thompson FL. Multiple *Symbiodinium* Strains Are Hosted by the Brazilian Endemic Corals *Mussismilia* spp. **Microbial Ecology**. 70(2):301-10, 2015. doi: 10.1007/s00248-015-0573-z

Coutinho FH, von Meijenfeldt, FAB, Walter JM, van Verk MC, Paranhos R, Dutilh BE, Thompson FL. Archaeal and Bacterial genomes discovered through metagenomics reveal microbial adaptations to the deep Atlantic ocean. *In prep*.

Tschoeke DA☯, Walter JM☯, de Oliveira L, Meirelles PM, Cavalganti G, Leomil L, Thompson CC, Coutinho R, Sergey D, Thompson FL. Metagenomic analysis and salinity adaptation of microbial mats from a complex of hypersaline coastal lagoons (Brazil). *In prep*. ☯The two first authors contributed equally to this work.

Walter JM☯, Lopes FAC☯, Lopes-Ferreira M, Vidal LM, Leomil L, Melo F, Azevedo GS, Oliveira RMS, Medeiros AJ, Melo ASO, Rezende CE, Tanuri A, Thompson FL. Harmful cyanobacteria in drinking water from a severely drought-impacted semi-arid region. **The Isme Journal***, Under review.* ☯The two first authors contributed equally to this work.

Fistarol GO, Walter JM, Hargreaves PI, Viana TV, Gomes PDF, Brito CL, Gregoracci G, Rua C, Rezende CE, Lopes FAC, Tschoeke DA, Thompson CC, Salomon PS, Thompson FL. Genetic Diversity of Chlorophyta from the Shallow Lakes System of Lençois Maranhenses National Park, Northeast Brazil. **PeerJ**, *Under review.*

# SCIENTIFIC EXPEDITION

*Abrolhos Bank: March, 2013* – Turf sampling for metagenomic analysis, cyanobacterial isolation. Diving-PAM measurements to estimate the photosynthetic rates of turfs. Diel sampling of seawater to evaluate changes in bacterial composition (metagenomic, cytometry, and nutrient analysis). Sampling corals species.

*Abrolhos Bank: October, 2013* – Turf sampling for metagenomic analysis, cyanobacterial isolation and pigments concentration analysis. Diving-PAM measurements to estimate the photosynthetic rates of turfs. Diel sampling of seawater to evaluate changes in bacterial composition (metagenomic, cytometry, and nutrient analysis). Sampling corals species.

*Abrolhos Bank: February, 2014* – Diel sampling of seawater to evaluate changes in bacterial composition (through metagenomic, cytometry, and nutrient analysis), and also in dissolved and particulate organic carbon and chlorophyll concentration. Diving-PAM measurements to estimate the photosynthetic rates of turfs and corals in controlled experiment. Sampling corals species.

*Paraíba State: September, 2016* – Water sampling for metagenomic analysis, cyanobacterial isolation and toxicity assays. Fixation for cytometry analysis, nutrient analysis, and chlorophyll concentration analysis.

# ABSTRACT

Coral reefs worldwide are actually in decline. At approximately 17 ºS, Abrolhos Bank emerge from the continental shelf of Bahia State (Brazil) as the largest coral reefs in the South Atlantic. In a context of modern ocean, changes in coral reefs are evident. Nowadays turfs representing greater than half (56.1%) the coverage of the entire benthic community in Abrolhos. Their presence is associated with a reduction in three-dimensional coral reef complexity and decreases the habitats available for reef biodiversity. Turf microbial definition and their functional roles are largely unclear. This thesis performed a structural and functional genomic analysis of turf assemblages and their newly cyanobacterial isolated strains, and advances into a wide genome-based taxonomy of the entire Cyanobacteria phylum, bringing a new, consistent and challenge proposal for classify them. A multidisciplinary approach was applied, since physiological analysis to (meta)genomics, taxonomy and ecological delineations. Here turfs were taxonomic- and functionally characterized, showing homogeneous functional core across the Abrolhos Bank, where a diversity of microbial guilds acting in synergism. Cyanobacteria strains composing turfs were isolated and their genomes demonstrated high potential for secondary metabolites biosynthesis and evidence of acclimation strategies, hint at the potential of *Acrophormium turfae* for turfs success in the reef system. Moreover, their phylogenomic reconstruction revealed huge incongruences with the current Cyanobacteria taxonomy, encouraging to expand the analysis and to use ecogenomics to delineate ecological niches preferences of Cyanobacteria and link them towards a new genome-based taxonomic classification. As result, three ecogenomic groups were identified at global level. From a total of 100 genomes, 31 new genera and 32 new species were circumscribed. We showed for the first time the genomic plasticity of turfs and possible ecological processes carried out by turfs and their isolated members in the Abrolhos reefs.

**Keywords:** Microbial Ecology / Coral Reefs / Abrolhos Bank / Metagenome / Ecological Delineation / Ecological Niches / Genomic Taxonomy / Genome-based Microbial Taxonomy / Turf / Cyanobacteria / Secondary Metabolites

# RESUMO

Recifes de corais em todos os oceanos estão atualmente em declínio. Em aproximadamente 17 ºS, o Banco de Abrolhos emerge a partir da plataforma continental do Estado da Bahia (Brasil) como o maior recife de corais do Atlântico Sul. Em um contexto de oceano moderno, mudanças nos recifes de corais são evidentes. Atualmente turfs representam mais da metade (56,1%) da cobertura total da comunidade bêntica em Abrolhos. Sua presença está associada à redução na complexidade tridimensional proporcionada pelos corais e à diminuição de *habitats* disponíveis para a biodiversidade nos recifes. A definição microbiana dos turfs e suas funções não são claras. Esta tese de doutorado realizou análise genômica estrutural e funcional das assembleias de turfs e de seus isolados de Cyanobacteria, e avançou em ampla análise baseada em genomas da taxonomia do filo Cyanobacteria, trazendo uma nova, consistente e desafiadora proposta para classificar essas bactérias. Estratégia multidisciplinar foi aplicada, desde análises da fisiologia, até estratégias metagenômicas e genômicas, além de taxonomia e delineamento ecológico. Neste trabalho, os turfs foram caracterizados taxonomica- e funcionalmente, mostrando homogênea estrutura funcional ao longo de recifes de Abrolhos, onde a diversidade de associações age em sinergia. Linhagens de Cyanobacteria que compõe os turfs foram isoladas e a análise de seus genomas demonstrou alto potencial para produção de metabólitos secundários e evidência de estratégias para aclimatação, sugerindo potencial de Acrophormium turfae na persistência de turfs nos recifes. Além disso, sua reconstrução filogenômica revelou alta incongruência com a atual taxonomia de Cyanobacteria, encorajando expandir as análises e usar ecogenômica para delinear nichos ecológicos preferenciais em Cyanobacteria, relacionando-os em direção à nova classificação taxonômica baseada em genomas. Como resultado, três grupos ecogenômicos foram identificados a nível global. De um total de 100 genomas, 31 novos gêneros e 32 novas espécies foram delimitados. Mostra-se pela primeira vez a plasticidade genômica dos turfs e possíveis processos ecológicos desempenhados pelos turfs e seus isolados nos recifes de Abrolhos.

**Palavras-chave:** Ecologia Microbiana / Recifes de Coral / Banco dos Abrolhos / Metagenoma / Delineamento Ecológico / Nichos Ecológicos / Taxonomia Genômica / Taxonomia Genômica Microbiana / Turf / Cyanobacteria / Metabólitos Secundários

# SUMMARY

# Chapter I

## GENERAL INTRODUCTION

# CORAL REEFS SIGNATURES AND THEIR MICROBIAL ECOLOGY

### Going back to the sea: coral reefs

Coral reefs are the largest structures built solely by animals (Cnidaria Phylum; Anthozoa Class), algae (mainly of Corallinales Order), and their bacterial communities associated (Bacteria and Archaea Domains). Although coral reefs occupy less than 0.1% of the ocean area (Spalding & Grenfell 1997), they harbor approximately one-fourth of all of marine species (Davidson 1998). Coral reefs have long been considered stenotolerant ecosystems, i.e., limited by a relatively narrow range of environmental conditions, such as warm, clear, shallow, and fully saline waters. More specifically, corals are broadly recognized thriving in tropical warm waters (around 20 ºC), between the latitudes 30 °N and 30 °S (Figure 1). High temperatures above 30 ºC may be lethal to the major coral species. The coral depth distribution ranging from 0 m to 100 m below sea surface, in clear water, where there is abundant sunlight available. Draw the line of environmental limits to coral reef distribution is a great challenge. Recent exploratory studies have showed the unexpected presence of pristine deep-sea coral gardens (Bo et al. 2015; Boavida et al. 2016; Dahlgren et al. 2016). Furthermore, an extensive reef system was also recently detected at the Amazon mouth, underneath the river plume (Moura et al. 2016).

Corals are among the first animals on the planet. They have been found in cnidarian-like fossils reefs as old as 500 million years, in the Ediacaran Period (ca 635 – 542 Ma) (Menon et al. 2013). However, corals similar to the modern colonial coral reefs varieties have evolved in the early Cenozoic (60 million years), growing constrained by sea level.

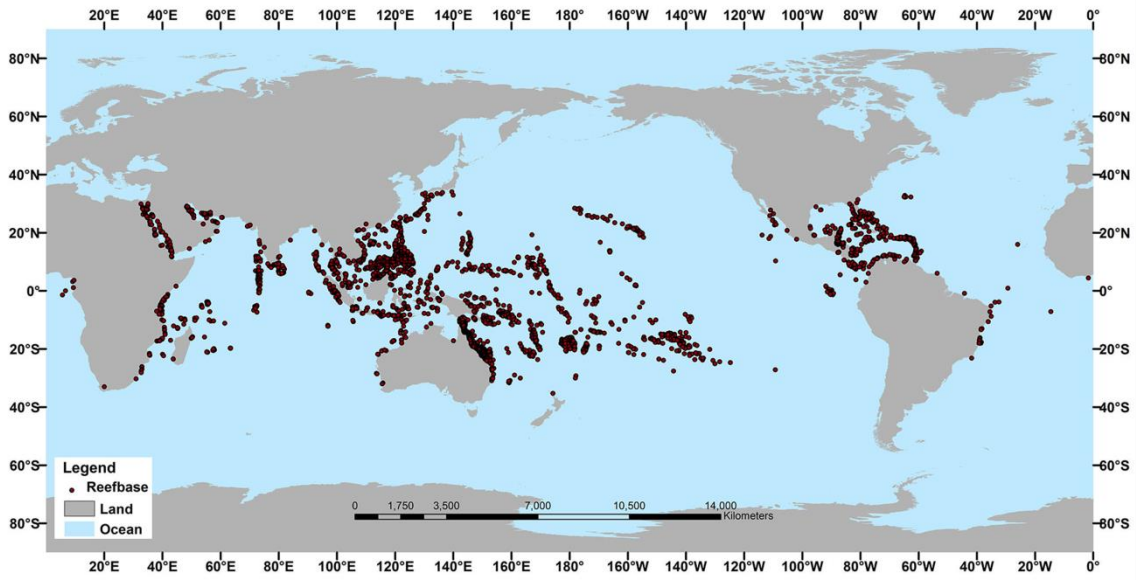



**Figure 3. Global distribution of coral reefs.** The majority of reef building corals are found within tropical and subtropical waters. These typically occur between 30 º north and 30 º south latitudes. The dots on this map show the location of major stony coral reefs of the world. Cold-water coral reefs are not represented. Source: <http://oceanservice.noaa.gov/>.

Scleractinian corals are key reef builders, providing most of the structural complexity in the reef ecosystem. The coral skeleton grow into a rock-like colony that forms the calcified structure of the coral reefs (Guest et al. 2012). Subsequent reef accretion contribute, changing the surrounding physical environment by increasing structural complexity and providing habitat for associated organisms (Bellwood & Hughes 2001; Wild et al. 2011). Corals are considered holobionts, because they acting as a host which hold several organisms in strongly association with them. One of these symbionts has an extremely importance: microalgae from the genus *Symbiodinium* (Dinoflagellata Phylum), also known as zooxanthellae. By photosynthesizing within the coral, *Symbiodinium* can provide more than 95% of the coral's metabolic requirements (Hoegh-Guldberg et al. 2007). The photosynthetic products from the symbionts, such as amino acids, sugars, complex carbohydrates, and small peptides, are selectively leaked across the host-symbiont barrier (Trench 1979). Zooxanthellae supply the coral

polyps with sufficient energy to build their protective calcium carbonate skeleton. Addicionally, zooplankton community has also an important role as food for corals.

The socioeconomic importance promoted by coral reefs reach further the tourism, fishing, building materials, pharmaceuticals and biochemical (Carté 1996). These habitats absorb 70 to 90 % of wind-generated wave energy (Wild et al. 2011) and support the coastal/shoreline protection, in the means that enables the formation of associated ecosystems (e.g., mangroves and seagrass beds), which are essential nursery habitats for reef fishes. Altogether, coral reefs are one of the most productive ecosystems on the planet (Hoegh-Guldberg 1999). Despite all these benefits, coral reefs are among the most severely threatened ecosystems, especially by rising sea temperatures.

The health of coral reefs is declining worldwide. Since the emergence of modern coral reefs, they have faced and persisted through a myriad of dramatic temperature and other environmental shifts (Pandolfi 1999). However, the last decades have had a huge increased in anthropogenic load effects. Additionally, studies over the last century showed that the rate of change in global temperature was two to three orders of magnitude higher than most changes observed in the geologic record over the past 420,000 years (Hoegh-Guldberg et al. 2007). Nearly all of the reefs have been heated above their maximum temperature threshold, and many have already lost a significant portion of their corals. Studies showed that approximately 30% to 40% of the world's coral reefs are already severely degraded (Hoegh-Guldberg et al. 2007; Hoegh-Guldberg 1999; Eakin et al. 2010; Burt et al. 2011; Guest et al. 2012; Pandolfi et al. 2005).

## Abrolhos reefs: the largest Bank of corals from South Atlantic

Abrolhos coral reefs were the study sites of this thesis. “Abrolhos” literally meaning “open your eyes” referring to the shallow reefs that make navigation dangerous. This name was gave by the first Portuguese navigators when reached the

southern coast of Bahia. Interesting, the same name was utilized in Western Australia, possibly because the same shallow reefs feature.

The Abrolhos Bank distribution (between the coordinates 17$^{o}$20-18$^{o}$10'S and 38$^{o}$35'-39$^{o}$20'W) extends along approximately 46,000 km$^{2}$ across the Eastern Brazilian continental shelf, in a distance of approximately 70 km from Bahia State shoreline (Figure 2). The remarkable extension provides to Abrolhos been the largest coral reefs in the South Atlantic and besides, harbor the largest continuous rhodolith bed in the world (Amado-Filho et al. 2012; Leão 1999), and also hold the unique mushroom-shaped coral pinnacles structures called 'chapeirões' (Leão et al. 2003). Indeed, these uncommon structures characterize Abrolhos reefs. Chapeirões were built following the ocean level, until stabilize and expand laterally on their tops creating large reef areas. They often attain the low water level. Hartt (1870, apud Kikuchi & Leão 1998) described these coral structures as "chapeirão" (pl. chapeirões) first. Further, considering the mesophotic zone, submerged pinnacles are detected across the mid and outer shelves at depths from 25 to 90 m. These structures are coalescent reefs with sinkhole-like valley known as 'buracas'.

Two arcs form these reefs. Whereas chapeirões are found in the outer arc, reefs of various shapes and dimensions are found in the coastal arc. Altogether, Abrolhos coral reefs complex consist to an extended bank harboring coral reefs (including Chapeirões shape), rhodoliths beds, surrounding buracas channels, sediments, sand shoals, and the Archipelago volcanic islands.

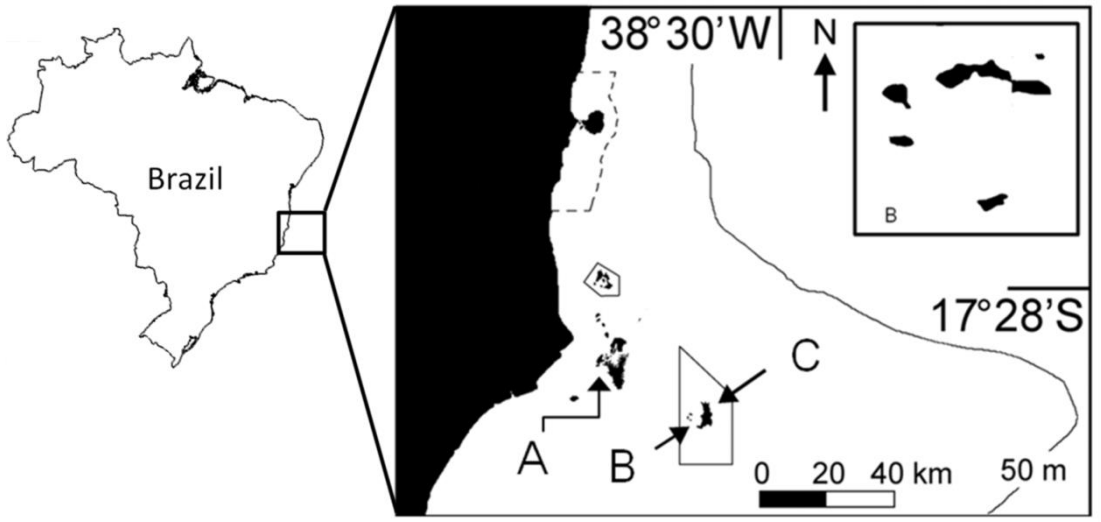



**Figure 4. Study site: the South Atlantic's largest coral reefs.** Arrows indicate the sampling sites of Abrolhos Bank used in this study. Sites: A) Pedra de Leste, B) Archipelago, C) Parcel dos Abrolhos.

The emergence and evolving of these reefs were from an isolated coral fauna of the Tertiary, and during the Quaternary the development of these coral reefs were strongly controlled by relative sea-level changes which affected the east coast of Brazil during this period (Leão 1996). At least two relative sea-level fluctuations occurred (Martin et al. 1996). One of them was around 123,000 years B.P. (before present), when sea-level reached 8 ± 2 meters above the present level (Martin et al. 1980), and the subsequent event extended up to 18,000 years B.P., when sea-level reached a minimum of 100-120 meters below the present level (Martin et al. 1980). The latest (regression of the sea-level) exposed the reef tops to marine erosion, dissolution and extensive bioerosion, and the reef communities dwelling thereon these tops experienced stress resulting, chiefly, from strong solar radiation and high levels of sedimentation (siliciclastic sediment influx from coastline) and water turbidity. Those environmental conditions must have exceeded the tolerance levels of most of the Brazilian coral species. Only endemic species of *Siderastrea stellata* and *Favia gravida* inhabit the shallow reef top pools, and are able to withstand this stressing environment.

Evidences indicate that a continuum reef had existed connecting northern Brazilian reefs to Caribbean ones, being this flux interrupted by the Amazon River changes course from Pacific to Atlantic Ocean. The huge continuum of reefs of the Atlantic Ocean was separated in hotspots across the shelf. Among them, Abrolhos reefs excels. Despite differences mainly related to reef morphology and depositional setting (Knoppers et al. 1999), Abrolhos Bank and Caribbean reef system share more than half of Scleractinian corals.

Brazilian coral reefs represent a priority area for biodiversity conservation in the Atlantic Ocean. Abrolhos Bank hold about 20 scleractinian coral species, among six are endemic (Leão et al. 2003). Such number of coral species may appear low, when comparing with Indo-Pacific (Hughes et al. 2002) and Caribbean reefs (Mumby et al. 2014). However, their relatively high endemism levels are concentrated in a small reef area (5% of West Atlantic reefs). The brain corals of the *Mussismilia* genus is endemic of Brazil, and the *M. braziliensis* species is endemic to the coast of Bahia and Espírito Santo States (Leão & Kikuchi 2005). While *Mussismilia* genus (*M. braziliensis*, *M. hispida*, and *M. hartii*) is the major endemic reef-building contributor (Leão & Kikuchi 2005), *Montastraea cavernosa* is the major reef-building species in southwestern Atlantic coral reefs (Francini-Filho et al. 2013). *Siderastrea* spp., *Favia gravida*, *Scolymia wellsi*, and *Madracis decactis* species are in a lesser extent also contributors for the reef framework (Leão & Kikuchi 2005), in different coverages considering pinnacles' tops and walls (Francini-Filho et al. 2013). In addition to factors as light, depth and bottom inclination, sedimentation rates are important drivers to establish the whole benthic assemblages' structure of Abrolhos reefs, since *Mussismilia* genus evolved to tolerate turbidity stress promoted by high sedimentation rate (Loiola et al. 2013), as well as *M. cavernosa* species (Lasker 1980) which is also abundant in mesophotic (30 – 150 m) clear water reefs (Western Atlantic locations, Kahng et al. 2010), suggesting low light levels adaptation.

Increasing of polymicrobial syndromes and diseases and also turf assemblages (phase-shifting phenomenon) on corals have been detected in Abrolhos Bank. Microbial loads on reefs are depicted in the next section.

## The microbial point-of-view in modern ocean context

The microbiological background is now emerging on the most evident and dramatic changes in coral reefs worldwide. Although coral syndromes and diseases were registered since the 1980s as a major cause of the loss of coral biodiversity and coral coverage worldwide (Harvell 2008; Harvell 2002; Rosenberg et al. 2007; Rosenberg & Loya 2004), only in 2005 they were detected occurring in Abrolhos Bank. Coral diseases also represent a threat to the reefs in the region (Francini-Filho et al. 2008). The most common diseases observed in Abrolhos are white plague, black band, red band, dark spot, aspergillosis and octocoral tissue necrosis, some of them considered polymicrobial syndromes (Francini-Filho et al. 2008). Super-heterotrophic bacteria, such as *Vibrionales*, *Rickettsiales* and *Neisseriales*, as well as members of the *Cytophaga–Flavobacterium–Bacteroides* complex are highly abundant in white plague-affected coral tissue (Garcia et al. 2013; Fernando et al. 2015).

Although the negative effects in corals are evident, the nature of the origin of these diseases and even the concept of disease or syndrome in this context is currently in discussion. However, seems to be indubitable that modern alterations (i.e. climate change) in the oceans are leading to an advantage to turfs over corals. Turfs are mixed assemblages consisting of microbes and algae that inhabit reef systems and act as an unified functional entity (Walter et al. 2016) (see Figure 3 to understand how they looks like).

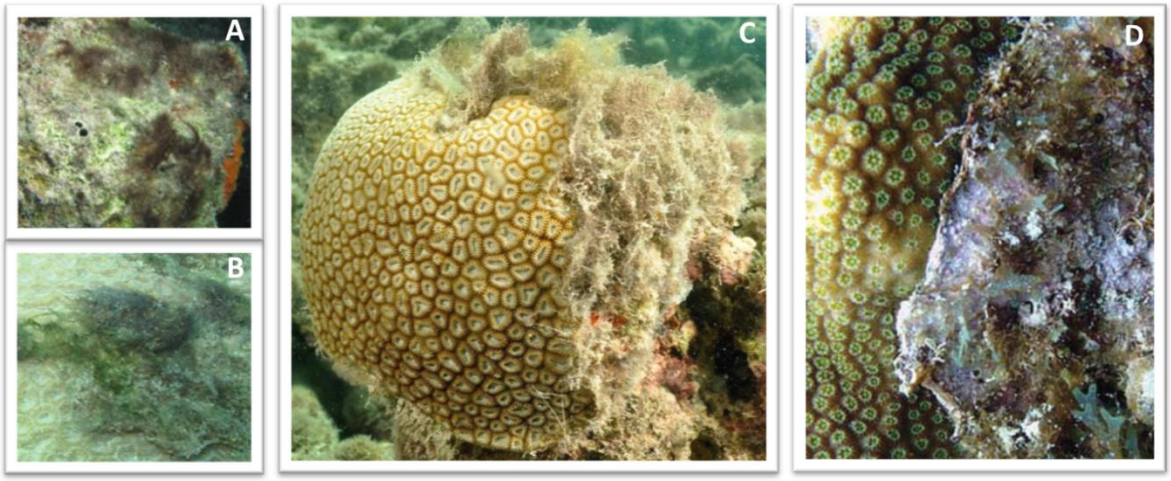



**Figure 3. Turf assemblages samples.** (A-C) Overview of turfs collected in the South Atlantic Ocean, Abrolhos reefs (Source: Walter, J.M. 2013); and (D) turfs sampling from Central America Ocean, Belize (Wild et al., 2014).

A marked turf spread has been observed over the last decade, and their presence potentially reflecting a phase shift phenomenon (Knowlton & Jackson 2008). In this process, a decrease in the abundance of herbivorous fish leads to an increase in turf cover (Dinsdale, Edwards, et al. 2008; Sandin et al. 2008). Additionally, the spread of turf can be promoted by an increase in temperature and nutrients as a result of local and global environmental changes (Bruce et al. 2012). Turfs release higher concentration of dissolved organic carbon (DOC) (Haas et al. 2013; Haas et al. 2011; Nelson et al. 2013), which stimulate rapid growth of heterotrophic microbes (Haas et al. 2011; Dinsdale et al. 2008) and may promote shifts towards potentially pathogenic microbial populations (Nelson et al. 2013), trigging microbialization of coral reef in a global distribution level, as showed by Haas and collegues (2016) (Figure 4).

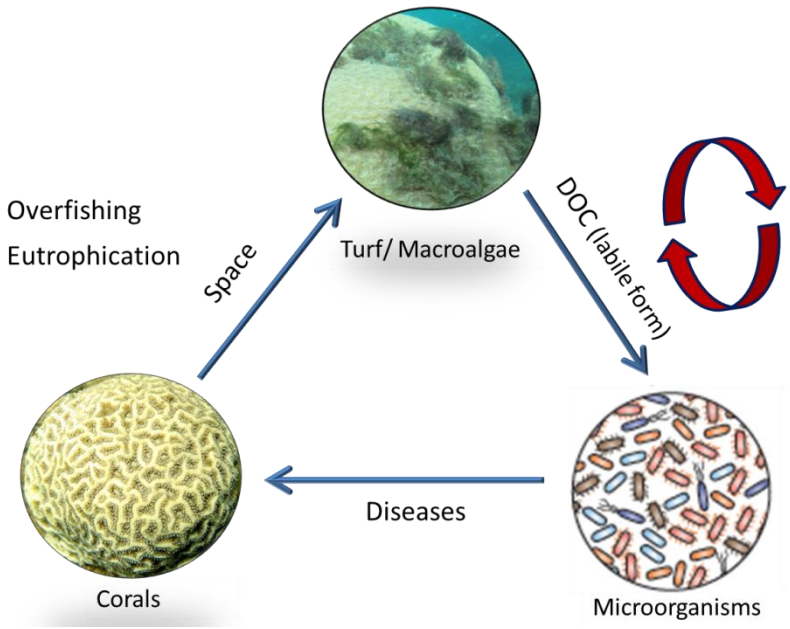



**Figure 4. Microbialization process of coral reefs.** Overfishing and eutrophication are facilitators, enhanced growth of turf/macroalgae, which gain competitive advantage over calcifying corals and coralline algae. The increased DOC available into the system support microrganisms grown. Then, the ecosystem trophic structure shift towards higher microbial biomass and energy use (Adaptated from Haas et al., 2016).

Turf assemblages increasing were detected in Abrolhos and incorporated in the periodical measurements and monitoring evaluations. Results showed the expressive abundance of turf on benthic community across the Abrolhos reefs (56.1%) (Francini-Filho et al. 2013) (Figure 5). The high turf abundance in Abrolhos Bank raises issues about their taxonomic and functional nature, and whether several types of turf microbiomes co-existing and how they are distributed and persist across reefs. These questions are addressed on the second chapter of this thesis.

Once previously observations and sampling showed that turfs have dense filamentous macro aspect, isolations allowed unicyanobacterial cultive of two representative strains. We showed that Cyanobacteria are among the most abundant microrganisms populations composing turfs in Abrolhos reefs. Genomes of the named *Achrophormium turfae* CCRM0081$^{T}$ and CCRM0082 strains are among the largest cyanobacterial genomes and they hold a vast potential repertoire of secondary

metabolites, which may be involved in important ecological processes in the reefs (e.g., diverse signaling pathways, resource competition, allelopathy, feeding deterrence, UV-protection, and even influencing the coral health) (third chapter).

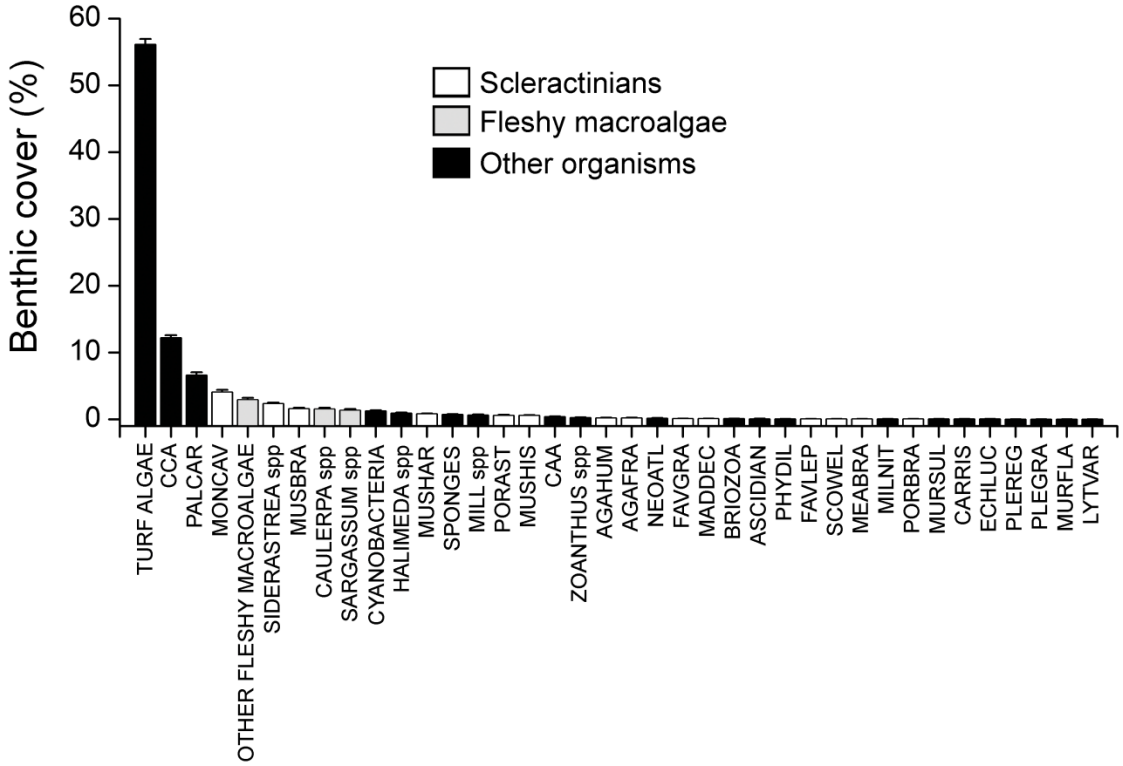



**Figure 5. Abundance of benthic organisms in the Abrolhos Bank.** Species codes are represented by the first three letters of genus name followed by first three letters of specific epithet, according to the source paper. Source: Francini-Filho et al. (2013).

## NGS technologies: revolutionizing microbial studies

Since its release in 2005, massively parallel sequencing technologies, collectively known as high-throughput next-generation sequencing (NGS), have been responsible for a revolution in the biological research (Zhang et al. 2011; Niedringhaus et al. 2011). It started with the introduction of second-generation sequencers, where platforms such as 454 from Roche; GA, MiSeq, and HiSeq from Illumina; SOLiD and Ion Torrent from Life Technologies; RS system from Pacific Bioscience; and Heliscope from Helicos Biosciences, are able to produce larger amount of data in a single run (high-throughput data) (Niedringhaus et al. 2011; Koren et al. 2012; Loman et al. 2012;

Miller et al. 2012) at a moderate cost (i.e., considering the drastic reduction in the price of genome sequencing and the huge quantity of genomic information) (Zhou et al. 2010) (Figure 6). It is particularly important for the decision to obtain greater sequencing coverage (i.e., deep sequencing with up to millions of reads per sample), especially if the objective is to focus on microbial diversity, distribution, and biogeography, where sampling rare taxa could be more important (Caporaso et al. 2012).

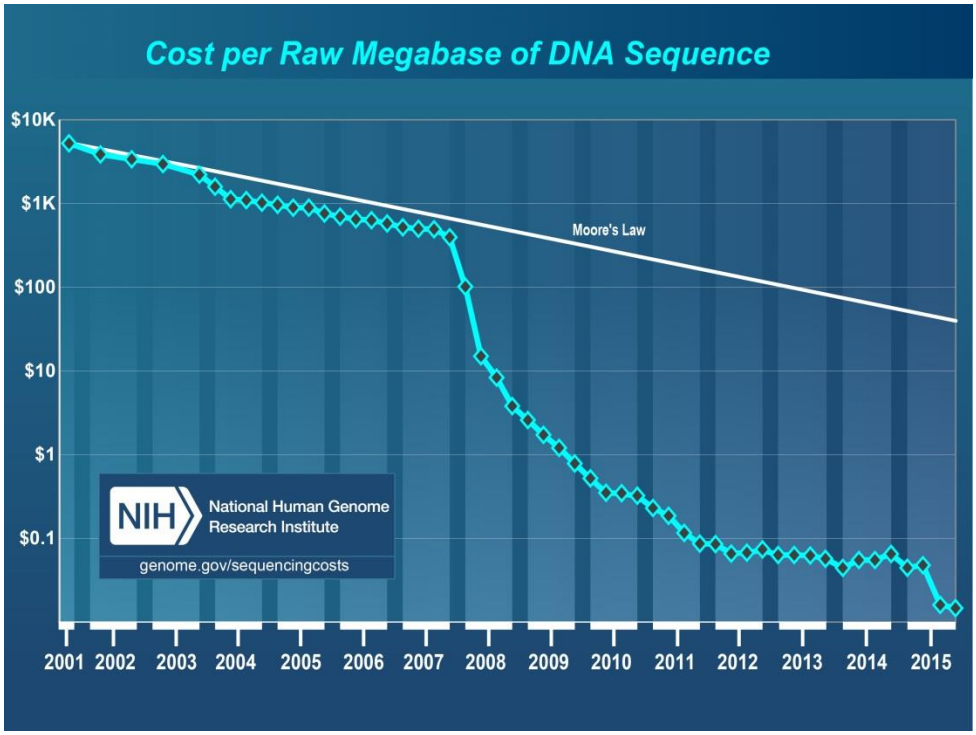



**Figure 6. Cost per raw Megabase of DNA sequence.** The cost of determining one megabase (Mb; a million bases) of DNA sequence of a specified quality. This graph does not reflect all of the costs associated with production-oriented DNA sequencing. Source: < https://www.genome.gov/sequencingcostsdata/>.

Each platform generally produces two types of data: 1) the short-read sequences and 2) the quality score values for each base in the read. The quality values are used to assess the sequence quality, trim reads, and remove low-quality bases. The sequencing technologies possibilities have brought excitement to the field of microbial genomics, leading researches interrogating microbial communities. Much has been done, but one of the big challenges of 21$^{st}$ century science is still understand the

structure, functions, activities and dynamics of microbial communities in natural environments (Zhou et al. 2015). And all these advances in NGS technologies have allowed significant breakthroughs in microbial studies, through the establishment of the field of “metagenomics”. Metagenomics is the study of entire communities on the basis of their genetic material from samples obtained directly from the environment, bypassing the requirement for obtaining pure cultures for sequencing (Cardenas and Tiedje, 2008; Hugenholtz and Tyson, 2008; Su et al., 2012; Riesenfeld et al. 2004). Metagenomics has evolved to address the issues of who is present in an environmental community, what they can do, what they are doing, how does it change, and how these microorganisms interact to sustain a balanced ecological niche. These questions are answer through examining the community structure, determining the genetic potential, determining gene expression and function, and looking at what changes occur over time or under different environmental pressures, respectively (Cardenas and Tiedje, 2008). Metagenomics creates a community biodiversity profile that can be further associated with functional composition analysis of known and unknown organism lineages (i.e., genera or taxa) (Tringe et al 2005).

The incredibly breakthroughs promoted by metagenomics in microbial ecology field was extremely important once, in any ecosystem, a plenty of microrganisms interact with each other to form complex networks whose behavior is hard to predict (Fuhrman 2009; Zhou et al. 2010). Addicting, microbial communities can be extremely diverse, and the majority of microrganisms in natural environments have not yet been cultivated (Quince et al. 2008; Kallmeyer et al. 2012). Then, sequence analysis of entire microbial communities creates an opportunity to discover a multitude of different bacterial species that may be either unique to environmental sources (Dubinsky et al., 2012), such as turf assemblages.

## New insights into the Cyano(bacterial) taxonomy

The knowledgment of the microbial diversity was passed by deeper improvements due the discoveries promoted by the new technologies. Now, the molecular-based methods have become an integral part of microbial ecology, basically from a culture-dependent to a -independent method. Even though, issues involving microbial systematics and their taxonomy have been standstill for long. However, despite the resistance of innovation, we saw in literature well-articulated calls for reform inside the microbial systematics (Hugenholtz et al. 2016; Gevers et al. 2005; Konstantinidis & Tiedje 2005; Gribaldo & Brochier-Armanet 2012; Sutcliffe et al. 2013; Garrity & Oren 2013).

The availability of complete genomes provides the opportunity to reconstruct events of genomic evolution through the analysis of entire functional classes (Koonin et al. 2000; Shi & Falkowski 2008), integrated to the knowledge of gene composition, distribution, identity, and also ecological patterns of strains. It is especially exciting for the called microbial dark matter (Rinke et al. 2013), consisting of the genomes of microbes and viruses that have never been seen before and which represent the most of the microbial sequences in all environments on earth. New methods based on metagenome assembly are now becoming available to retrieve genomes from the sequenced metagenomes (Peng et al. 2011; Namiki et al. 2012). Altogether, these new approaches allow us to extend the tree of life in a new version (Di Rienzi et al. 2013; Soo et al. 2014; Anantharaman et al. 2016; Spang et al. 2015; Hug et al. 2016).

The renewed classification for Cyanobacteria is also a big challenge. Their muddle taxonomic history is a reflection of their astonishing diversity (comprising unicellular and multicellular, photosynthetic and also non-photosynthetic, free living, symbiotic and predatory organisms – further many members with bi-phasic life cycles), as explored follows (Chapter IV).

## Cyanobacteria: a challenge group of microorganisms

Cyanobacteria form one of the major eubacterial phyla, dating from the Proterozoic Era (2,500-542 Mya). After their bacterial recognition classification (Stanier et al. 1978), molecular studies using different sequences and approaches e.g., 16S rRNA sequences (Woese, 1987), 23S rRNA and protein sequences (Schleifer & Ludwig 1989), conserved proteins sequences (Wu and Eisen 2008), shown convincingly that cyanobacterial phylogenetic trees form a coherent branch that is well separated from sister bacterial groups (Figures 7 and 8).

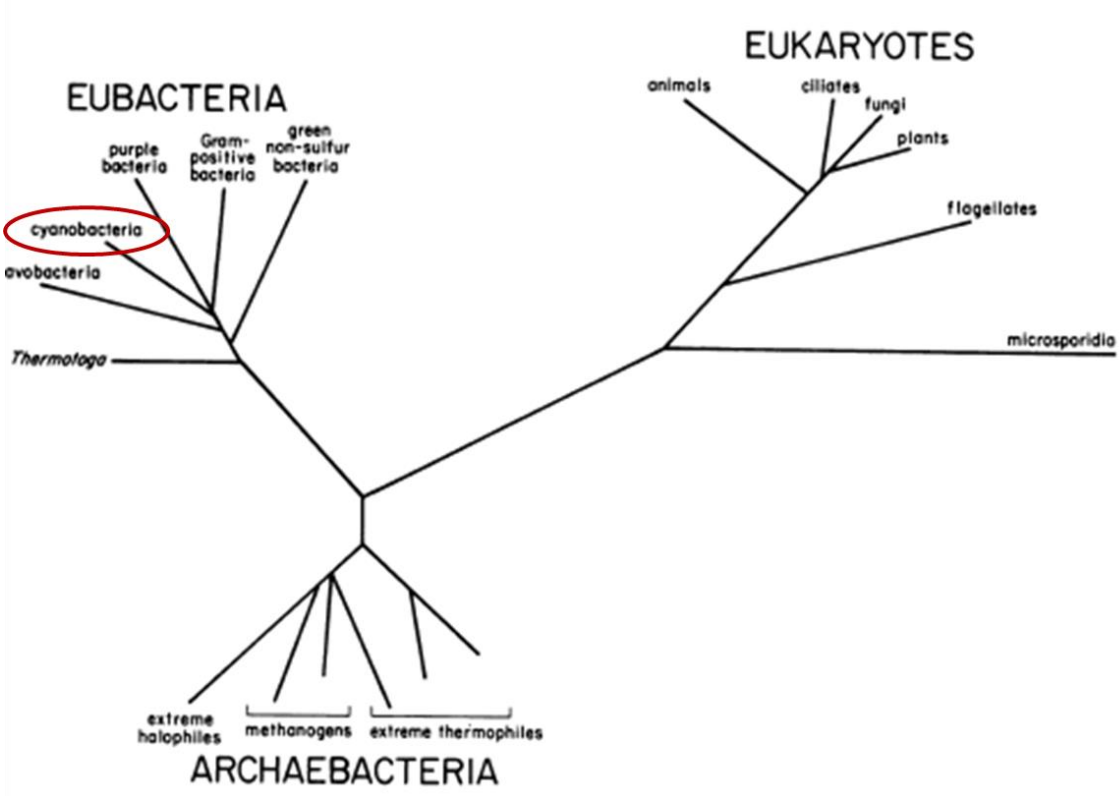



**Figure 7. Phylogenetic tree from Woese (1987) based on 16S rRNA sequences.**

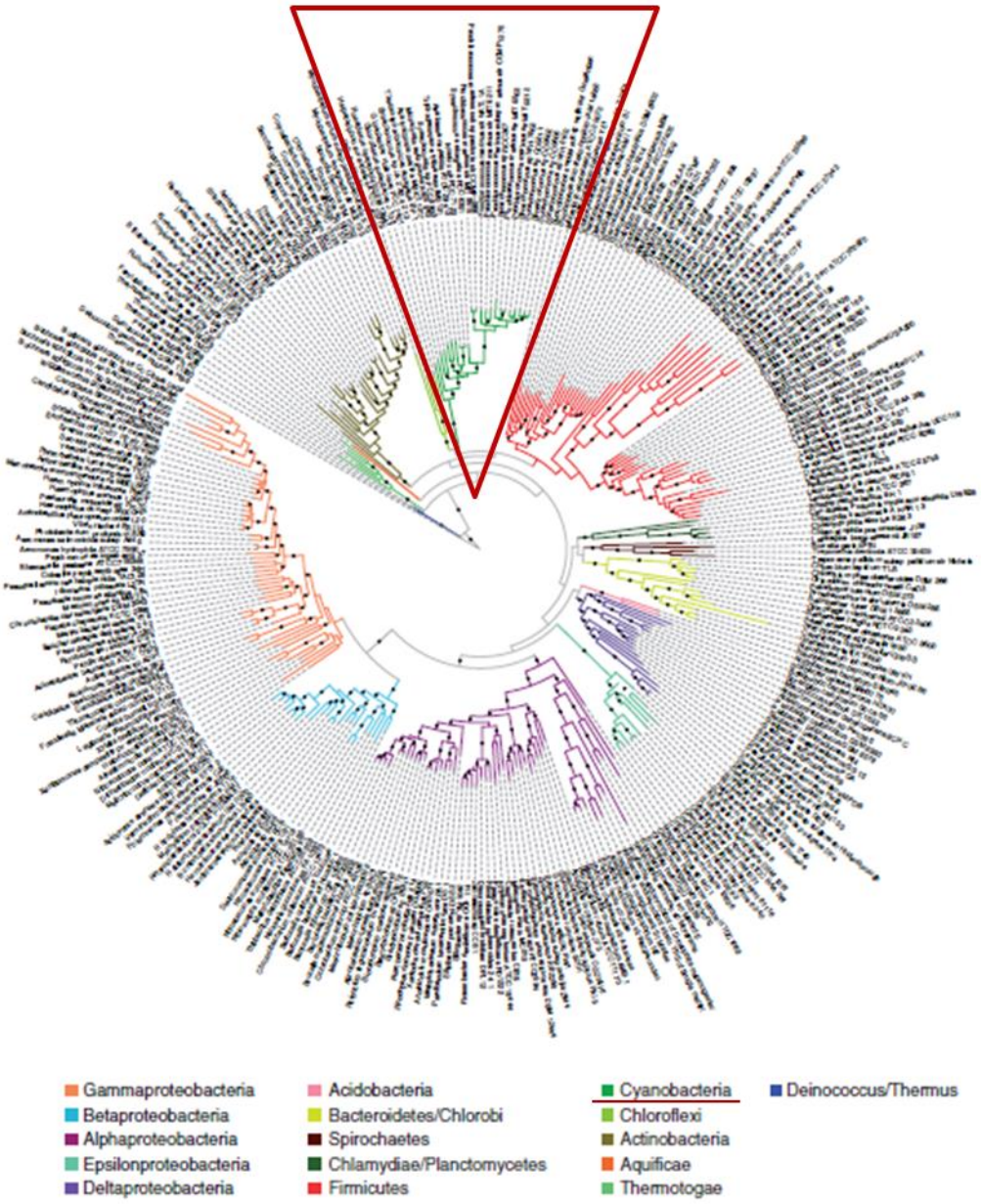



**Figure 8. Phylogenetic reconstruction based on 31 conserved proteins sequences.** Source: Wu and Eisen (2008).

The traditional classification of Cyanobacteria based on morphology is subjective (Boone & Castenholz 2001; Rippka et al. 1979). Thus, an unfortunate consequence of using traditional morphology-based taxonomic systems for cyanobacterial identifications has been that they are forced into existing morphological groupings and, therefore, the true biodiversity of this group has been greatly underestimated (Engene et al. 2013; Shih et al. 2013). It is especially worrisome for

new members of many recently explored biological frontiers (e.g., tropical marine environments). The wide undiscovered ecological diversification of cyanobacterial members reflects their muddled taxonomy, where phylogenetic studies indicate that all subsections and genera using this classification are polyphyletic (Gugger and Hoffmann 2004; Hugenholtz et al. 2016; Schirrmeister et al. 2011).

The evolutionary history and diversity subjects of Cyanobacteria are just beginning to gain an in-depth understanding (Shih et al. 2013). To settle this issue, genome sequencing comes to addicting to the unbalanced distribution of publicly available genomic information from the Cyanobacteria phylum. The availability of new cyanobacterial genomes generates the possibility to reconstruct the puzzle of the evolutionary history of this intriguing phylum.

Since the genome sequencing of newly cyanobacterial isolated strains from turfs revealed the needed to indicate their phylogenetic position, we compared with publicly available genomes. Therefore, it showed a deepest taxonomic incongruence. Expanding the genomic analysis to the entire Cyanobacteria phylum, we could detect the urgent needed to a taxonomic re-evaluation.

Aims of the present study are concerning about the turf signatures from Abrolhos reefs in a context of modern ocean traits, and the genomic characterization of Cyanobacteria strains isolated from them. Also, we proposed a new genome-based taxonomy for the entire Cyanobacteria phylum. The results are shown in subsequent chapters encompassing this document. In chapter 2, the taxonomic and functional metagenomic traits of turf microbiome are investigated, helping us to define the turf structure from Abrolhos reefs (chapter 2, in full, is a reprint of the material as it has been published in *Plos One* journal). In chapter 3, cyanobacterial strains living in turfs assemblages were isolated, allowing us to obtain genomes assembled and to investigate how the genomic content may allow these organisms enhance the turf persistence. An unexpected frequency of secondary metabolites was detected, as well as an acclimative respose to far-red light give us hint for that (chapter 3, in full, is a reprint of the material as it has been submitted in *The ISME Journal*). In chapter 4, we correlated taxonomic affiliation (phylogenomic) and niche occupancy (ecological delineation) for 100 cyanobacterial genomes. We re-evaluated the Cyanobacteria taxonomy, proposing in result a new scheme for the entire phylum, reorganizing them based on genomic metrics (chapter 4, in full, is a reprint of the material as it has been submitted in *Frontiers in Microbiology* journal). In chapter 5, we conclude all the results obtaining with this thesis work, and also shed some light on the future projects.

## References


Amado-Filho, G.M. et al., 2012. Rhodolith Beds Are Major CaCO3 Bio-Factories in the Tropical South West Atlantic. *PLoS ONE*, 7(4), p.e35171. Available at: http://dx.plos.org/10.1371/journal.pone.0035171.

Anantharaman, K. et al., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature communications*, 7, pp.1–11. Available at: http://dx.doi.org/10.1038/ncomms13219.

Bellwood, D.R. & Hughes, T.P., 2001. Regional-Scale Assembly Rules and Biodiversity of Coral Reefs. *Science*, 292(5521), pp.1532–1535. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1058635.

Bo, M. et al., 2015. Persistence of pristine deep-sea coral gardens in the Mediterranean Sea (SW Sardinia). *PLoS ONE*, 10(3), pp.1–21.

Boavida, J. et al., 2016. A well-kept treasure at depth: Precious red coral rediscovered in atlantic deep coral gardens (SW Portugal) after 300 Years. *PLoS ONE*, 11(1), pp.1–26.

Boone, D.R. & Castenholz, R.W., 2001. *Bergey's Manual of Systematic Bacteriology Volume One : The Archaea and the Deeply Branching and Phototrophic Bacteria* 2nd ed. G. Garrity, D. R. Boone, & R. W. Castenholz, eds., New York, NY: Springer-Verlag New York.

Bruce, T. et al., 2012. Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. S. A. Sandin, ed. *PloS one*, 7(6), p.e36687. Available at: http://dx.plos.org/10.1371/journal.pone.0036687 [Accessed July 16, 2014].

Burt, J., Al-Harthi, S. & Al-Cibahy, A., 2011. Long-term impacts of coral bleaching events on the world's warmest reefs. *Marine Environmental Research*, 72(4), pp.225–229. Available at: http://dx.doi.org/10.1016/j.marenvres.2011.08.005.

Caporaso, J.G. et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6(8), pp.1621–1624. Available at: http://dx.doi.org/10.1038/ismej.2012.8.

Cardenas, E. & Tiedje, J.M., 2008. New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology*, 19(6): 544-549.

Carté, B.K., 1996. Potential of Marine Biomedical Natural Products research and medical applications. *BioScience*, 46(4), pp.271–286.

Dahlgren, T. et al., 2016. Abyssal fauna of the UK-1 polymetallic nodule exploration area, Clarion-Clipperton Zone, central Pacific Ocean: Cnidaria. *Biodiversity Data Journal*, 4, p.e9277. Available at: http://bdj.pensoft.net/articles.php?id=9277.

Davidson, O.G., 1998. *The Enchanted Braid: Coming to Terms with Nature on the Coral Reef*, New York: Wiley.

Dinsdale, E. a, Edwards, R. a, et al., 2008. Functional metagenomic profiling of nine biomes. *Nature*, 452(7187), pp.629–632.

Dinsdale, E. a, Pantos, O., et al., 2008. Microbial ecology of four coral atolls in the

Northern Line Islands. *PloS one*, 3(2), p.e1584.

Dubinsky, E.A. et al., 2012. Application of phylogenetic microarray analysis to discriminate sources of fecal pollution. *Environmental Science & Technology*, 46 (8):4340-4347.

Eakin, C.M. et al., 2010. Caribbean corals in crisis: Record thermal stress, bleaching, and mortality in 2005. *PLoS ONE*, 5(11).

Engene, N. et al., 2013. Phylogenetic inferences reveal a large extent of novel biodiversity in chemically rich tropical marine cyanobacteria. *Applied and Environmental Microbiology*, 79(6), pp.1882–1888.

Fernando, S.C. et al., 2015. Microbiota of the Major South Atlantic Reef Building Coral *Mussismilia*. *Microb Ecol.* 69: 267. doi:10.1007/s00248-014-0474-6

Francini-Filho, R.B. et al., 2008. Diseases leading to accelerated decline of reef corals in the largest South Atlantic reef complex (Abrolhos Bank, eastern Brazil). *Marine Pollution Bulletin*, 56, pp.1008–1014.

Francini-Filho, R.B. et al., 2013. Dynamics of Coral Reef Benthic Assemblages of the Abrolhos Bank, Eastern Brazil: Inferences on Natural and Anthropogenic Drivers. *PLoS ONE*, 8(1), pp.1–12.

Fuhrman, J. a, 2009. Microbial community structure and its functional implications. *Nature*, 459(7244).

Garcia, G. et al., 2013. Metagenomic Analysis of Healthy and White Plague-Affected Mussismilia braziliensis Corals. *Microb Ecol*, 65: 1076-1086.

Garrity, G.M. & Oren, A., 2013. Response to Sutcliffe et al.: Regarding the International Committee on Systematics of Prokaryotes. *Trends in Microbiology*, 21(2), pp.53–55. Available at: http://dx.doi.org/10.1016/j.tim.2012.12.003.

Gevers, D. et al., 2005. Re-evaluating prokaryotic species. *Nat.Rev.Microbiol.*, 3(1740–1526 (Print)), pp.733–739. Available at: C:\Karsten\PDFs\Grundlagen-PDFs\Grund-2005\Gevers et al.-Re-evaluating prokaryotic species.pdf.

Gribaldo, S. & Brochier-Armanet, C., 2012. Time for order in microbial systematics. *Trends in Microbiology*, 20(5), pp.209–210.

Guest, J.R. et al., 2012. Contrasting patterns of coral bleaching susceptibility in 2010 suggest an adaptive response to thermal stress. *PLoS ONE*, 7(3), pp.1–8.

Gugger, M.F. & Hoffmann, L., 2004. Polyphyly of true branching cyanobacteria (Stigonematales). *International Journal of Systematic and Evolutionary Microbiology*, 54(2), pp.349–357.

Haas, A.F. et al., 2013. Influence of coral and algal exudates on microbially mediated reef metabolism. *PeerJ*, 1, p.e108. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3719129&tool=pmcentrez&rendertype=abstract.

Haas AF, Nelson CE, Wegley Kelly L, Carlson CA, Rohwer F, et al, 2011. Effects of Coral Reef Benthic Primary Producers on Dissolved Organic Carbon and Microbial Activity. Available at: http://coralreefecology.ucsd.edu/files/2010/09/Effects-of-Coral-Reef-Benthic-Primary-Producers-on-Dissolved-Organic-Carbon-and-Microbial-Activity1.pdf

[Accessed February 3, 2015].

Haas, A.F. et al., 2016. Global microbialization of coral reefs. *Nature Microbiology*, 16042, doi:10.1038/NMICROBIOL.2016.42

Harvell, C.D., 2002. Climate Warming and Disease Risks for Terrestrial and Marine Biota. *Science*, 296(5576), pp.2158–2162. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1063699.

Harvell, C.D., 2008. Emerging Marine Diseases — Climate Links and. *Science*, 1505(1999).

Hoegh-Guldberg, O., 1999. Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research*, 50(8), pp.839–866.

Hoegh-Guldberg, O. et al., 2007. Coral Reefs Under Rapid Climate Change and Ocean Acidification. *Science*, 318(5857), pp.1737–1742.

Hug, L.A. et al., 2016. A new view of the tree of life. *Nature Microbiology*, 1(5), p.16048. Available at: http://www.nature.com/articles/nmicrobiol201648.

Hugenholtz, P. & Tyson, G.W., 2008. Microbiology: Metagenomics. *Nature*, 455:481-483. doi:10.1038/455481a.

Hugenholtz, P., Skarshewski, A. & Parks, D.H., 2016. Genome-Based Microbial Taxonomy Coming of Age. *Cold Spring Harbor perspectives in biology*, p.cshperspect.a018085-. Available at: http://cshperspectives.cshlp.org/content/early/2016/03/17/cshperspect.a018085.abstract.

Hughes, T.P., Bellwood, D.R. & Connolly, S.R., 2002. Biodiversity hotspots , centres of endemicity, and the conservation of coral reefs. *Ecology Letters*, 5, pp.775–784.

Kahng, S.E. et al., 2010. Community ecology of mesophotic coral reef ecosystems. *Coral Reefs*, pp.1–21.

Kallmeyer, J. et al., 2012. From the Cover: Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences*, 109(40), pp.16213–16216.

Kikuchi, R.K.P. & Leão, Z.M.A.N., 1998. The effects of Holocene sea level fluctuation on reef development and coral community structure, northern Bahia, Brazil. *An. Acad. Bas. Ci.*, 70(2).

Knoppers, B. et al., 1999. Compartments of the pelagic system and material exchange at the Abrolhos Bank coral reefs, Brazil. *Archive Fishery Marine Res*, 47(2–3), pp.285–306.

Knowlton, N. & Jackson, J.B.C., 2008. Shifting baselines, local impacts, and global change on coral reefs. *PLoS biology*, 6(2), p.e54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2253644&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Konstantinidis, K.T. & Tiedje, J.M., 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), pp.2567–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=549018&tool=pmcentrez&rendertype=abstract.

Konstantinidis, K.T. & Tiedje, J.M., 2005. Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology*, 187(18), pp.6258–6264.

Koonin, E. V, Aravind, L. & Kondrashov, a S., 2000. The impact of comparative genomics on our understanding of evolution. *Cell*, 101(6), pp.573–576.

Koren, S. et al., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature biotechnology*, 30(7), pp.693–700. Available at: http://dx.doi.org/10.1038/nbt.2280.

Leão, Z.M.A.N., 1999. Abrolhos - The south Atlantic largest coral reef complex. In C. Schobbenhaus et al., eds. *Sítios Geológicos e Paleontológicos do Brasil*. Available at: http://sigep.cprm.gov.br/sitio090/sitio090english.htm.

Leão, Z.M.A.N., 1996. The coral reefs of Bahia: morphology, distribution and the major environment impacts. *Anais Academia Brasileira de Ciências*, 68, pp.439–452.

Leão, Z.M.A.N. & Kikuchi, R.K.P., 2005. A relic coral fauna threatened by global changes and human activities, Eastern Brazil. *Marine Pollution Bulletin*, 51(5–7), pp.599–611.

Leão, Z.M.A.N., Kikuchi, R.K.P. & Testa, V., 2003. *Corals and coral reefs of Brazil*,

Loiola, M., Oliveira, M.D.M. & Kikuchi, R.K.P., 2013. Tolerance of Brazilian brain coral Mussismilia braziliensis to sediment and organic matter inputs. *Marine Pollution Bulletin*, 77(1–2), pp.55–62. Available at: http://dx.doi.org/10.1016/j.marpolbul.2013.10.033.

Loman, N.J. et al., 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*, 30(5), pp.434–9. Available at: http://dx.doi.org/10.1038/nbt.2198.

Martin, L. et al., 1980. Texto explicativo para o mapa geológico do Quaternário costeiro do Estado da Bahia - Escala 1:250.000. COM. SME-BA, 57 p.

Martin, L. et al., 1996. Quaternary sea-level history and variation in dynamics along the central Brazilian coast: consequences on coastal plain construction. *Anais Academia Brasileira de Ciências*, 68, pp.303–354.

Menon, L.R., Mcilro, D. & Brasier, M.D., 2013. Evidence for Cnidaria-like behavior in ca. 560 Ma Ediacaran Aspidella. *Geology*, 41(8), pp.895–898.

Miller, J.M. et al., 2012. Short reads, circular genome: Skimming solid sequence to construct the bighorn sheep mitochondrial genome. *Journal of Heredity*, 103(1), pp.140–146.

Moura, R.L. et al., 2016. An extensive reef system at the Amazon River mouth - Supplementary Materials. *Science Advances*, 2(4), pp.e1501252–e1501252. Available at: http://advances.sciencemag.org/cgi/doi/10.1126/sciadv.1501252.

Mumby, P.J. et al., 2014. Caribbean Coral Reefs Ecological History and Biogeography a handbook for caribbean coral reef managers Caribbean Coral Reefs Ecological History and Biogeography. *Towards Reef Resilience and Sustainable Livelihoods: a handbook for Caribbean coral reef managers*, p.172.

Namiki, T. et al., 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res*, 40(20): e155. doi.org/10.1093/nar/gks678

Nelson, C.E. et al., 2013. Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *The ISME journal*, pp.1–18.

Niedringhaus, T.P. et al., 2011. Landscape of Next-Generation Sequencing Technologies, 83(12), pp.4327–4341.

Pandolfi, J.M. et al., 2005. Are U. S. coral reefs on the slippery slope to slime? *Science*, 307(October), pp.1725–1726.

Pandolfi, J.M., 1999. Response of Pleistocene Coral Reefs to Environmental Change Over Long Temporal Scales. *Integr Comp Biol*, 39(1):113-130. doi.org/10.1093/icb/39.1.113

Peng, Y. et al., 2011. Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics*, 27(13): i94-i101. doi.org/10.1093/bioinformatics/btr216

Quince, C., Curtis, T.P. & Sloan, W.T., 2008. The rational exploration of microbial diversity. *The ISME journal*, 2(10), pp.997–1006.

Di Rienzi, S.C. et al., 2013. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife*, 2013(2), pp.1–25.

Riesenfeld,C.S., Schloss,P.D. & Handelsman, J., 2004. Metagenomics: Genomic Analysis of Microbial Communities. *Annual Review of Genetics*, 38:525-552. doi.org/10.1146/annurev.genet.38.072902.091216.

Rinke, C. et al., 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459).

Rippka, R. et al., 1979. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Journal of General Microbiology*, 111(1), pp.1–61.

Rosenberg, E. et al., 2007. The role of microorganisms in coral health, disease and evolution. *Nature reviews. Microbiology*, 5(5), pp.355–62.

Rosenberg, E. & Loya, Y., 2004. *Coral Health and Disease*, Springer-Verlag Berlin Heidelberg.

Sandin, S.A. et al., 2008. Baselines and degradation of coral reefs in the Northern Line Islands. *PLoS ONE*, 3(2).

Schirrmeister, B.E. et al., 2011. Evolution of cyanobacterial morphotypes: Taxa required for improved phylogenomic approaches. *Communicative & integrative biology*, 4(4), pp.424–7.

Schleifer, K. & Ludwig, W., 1989. Phylogenetic relationships among bacteria. In B. Fernholm, K. Bremer, & Jornvall, eds. *The Hierarchy of Life*. Amsterdam: Elsevier.

Shi, T. & Falkowski, P.G., 2008. Genome evolution in cyanobacteria: the stable core and the variable shell. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), pp.2510–2515.

Shih, P.M. et al., 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3), pp.1053–8.

Soo, R.M. et al., 2014. An expanded genomic representation of the phylum yanobacteria. *Genome Biology and Evolution*, 6(5), pp.1031–1045.

Spalding, M.D. & Grenfell, A.M., 1997. New estimates of global and regional coral reef areas. , pp.225–230.

Spang, A. et al., 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551), pp.173–179. Available at: http://www.nature.com/doifinder/10.1038/nature14447.

Stanier, R.Y. et al., 1978. Proposal to Place the Nomenclature of the Cyanobacteria (Blue-Green Algae) Under the Rules of the International Code of Nomenclature of Bacteria. *International Journal of Systematic Bacteriology*, 28(2), pp.335–336.

Su, C. et al., 2012. Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Applied Microbiology and Biotechnology*, 93(3): 993–1003.

Sutcliffe, I.C. et al., 2013. A call to action for the International Committee on Systematics of Prokaryotes. *Trends in Microbiology*, 21(2), pp.51–52. Available at: http://dx.doi.org/10.1016/j.tim.2012.11.004.

Trench, R.K., 1979. The Cell Biology of Plant-Animal Symbiosis. *Annual Review of Plant Physiology*, 30:485-531. doi.org/10.1146/annurev.pp.30.060179.002413

Tringe, S.G. et al., 2005. Comparative Metagenomics of Microbial Communities. *Science*, 308(5721):554-557. doi:10.1126/science.1107851

Walter, J.M. et al., 2016. Taxonomic and Functional Metagenomic Signature of Turfs in the Abrolhos Reef System (Brazil). *PLoS ONE* 11(8):e0161168.

Wild et al., 2014. Turf algae-mediated coral damage in coastal reefs of Belize, Central America. PeerJ 2:e571.doi:10.7717/peerj.571

Wild, C. et al., 2011. Climate change impedes scleractinian corals as primary reef ecosystem engineers. *Marine and Freshwater Research*, 62, pp.205–215.

Woese, C.R., 1987. Bacterial Evolution. *Microbiol Rev.*,51(2):221–271.

Wu, M. & Eisen, J. a, 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome biology*, 9(10), p.R151.

Zhang, J. et al., 2011. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3): 95–109.

Zhou, J. et al., 2010. Functional molecular ecological networks. *MBio*, 1(4), pp.e00169-10. Available at: http://mbio.asm.org/content/1/4/e00169-10.short.

Zhou, J. et al., 2015. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. *CEUR Workshop Proceedings*, 1542(1), pp.33–36.

# Objectives

## Central Objective

The central goal of this work is to use broad approaches, mainly omics, in order to understand molecular processes involved in the persistence of turfs in Abrolhos Bank system, by means of Cyanobacteria emphasize analysis.

## Specific Aims

(i) Characterize taxonomically and functionally the microbial diversity of turfs;

(ii) Determine the basic metabolisms ans microbial processes of turfs from Abrolhos reefs;

(iii) Isolate and establish cultive of cyanobacterial strains from turfs;

(iv) Sequence, assembly, annotate and analyse the cyanobacterial genomes from turfs;

(v) Perform a comparative genomic analysis of the novel Cyanobacteria isolated from turfs in order to characterize the genetic diversity of their genomes and their potential metabolic;

(vi) Perform a comparative genomic analysis for the phylum Cyanobacteria, delineating their ecological niches;

and

(vii) Propose a novel cyanobacterial taxonomy based on genomic approach.

Chapter II

# TAXONOMIC AND FUNCTIONAL METAGENOMIC SIGNATURE OF TURFS IN THE ABROLHOS REEF SYSTEM (BRAZIL)

Published in Plos One (ISSN 1932-6203, Electronic; 1932-6203, Linking)
DOI:10.1371/journal.pone.0161168

Received: 22 April 2016 Accepted: 01 August 2016 Published: 22 August 2016
Publisher: San Francisco, CA: Public Library of Science

Juline M. Walter[1,2], Diogo A. Tschoeke[1,2], Pedro M. Meirelles[1], Louisi de Oliveira[1,2], Luciana Leomil[1,2], Márcio Tenório[3], Rogério Valle[4], Paulo S. Salomon[2,3], Cristiane C. Thompson[1], and Fabiano L. Thompson[1,2]

[1]Laboratory of Microbiology, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[2]Center of Technology - CT2, SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[3]Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[4]COPPE-Production Engineering Program, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

**Short Title:** Consistency of Turf in the Abrolhos Coral Reef

**Keywords:** Abrolhos Bank / Coral Reefs / Metagenome / Turf

## Abstract

Turfs are widespread assemblages (consisting of microbes and algae) that inhabit reef systems. They are the most abundant benthic component in the Abrolhos reef system (Brazil), representing greater than half the coverage of the entire benthic community. Their presence is associated with a reduction in three-dimensional coral reef complexity and decreases the habitats available for reef biodiversity. Despite their importance, the taxonomic and functional diversity of turfs remain unclear. We performed a metagenomics and pigments profile characterization of turfs from the Abrolhos reefs. Turf microbiome primarily encompassed Proteobacteria (mean 40.57% ± s.d. 10.36, N = 1.548,192), Cyanobacteria (mean 35.04% ± s.d. 15.5, N = 1.337,196), and Bacteroidetes (mean 11.12% ± s.d. 4.25, N = 424,185). Oxygenic and anoxygenic phototrophs, chemolithotrophs, and aerobic anoxygenic phototrophic (AANP) bacteria showed a conserved functional trait of the turf microbiomes. Genes associated with oxygenic photosynthesis, AANP, sulfur cycle (S oxidation, and DMSP consumption), and nitrogen metabolism ($N_2$ fixation, ammonia assimilation, dissimilatory nitrate and nitrite ammonification) were found in the turf microbiomes. Principal component analyses of the most abundant taxa and functions showed that turf microbiomes differ from the other major Abrolhos benthic microbiomes (i.e., corals and rhodoliths) and seawater. Altogether, these features suggest that turfs have a homogeneous functional core across the Abrolhos Bank, which holds diverse microbial guilds when comparing with other benthic organisms.

## Introduction

Turfs represent one of the most abundant benthic functional groups in the Atlantic, Caribbean, and Pacific reefs; however, there have been a limited number of studies characterizing the microbial composition of turfs [1]-[7]. Previous studies suggested that microbiomes play a central role in the functions of turfs. However, the majority of these studies were restricted to the Pacific region, and the composition of turf microbiomes remains unknown. As a functional group, turfs are characterized by filamentous and short, upright tufts rich in detritus that form a canopy less than 1 cm high over the substratum [8], [9]. In the Abrolhos Bank, turf is the most abundant group of benthic organisms on the reefs, having an average coverage of 56% (ca. ± s.d. 4.0) over the 8,844 $Km^2$ reef area [10]. A marked turf spread has been observed over the last decade, potentially reflecting a phase shift process [11]. In this process, a decrease in the abundance of herbivorous fish leads to an increase in turf cover [12], [13]. Additionally, the spread of turf can be promoted by an increase in temperature and nutrients as a result of local and global environmental changes [14]. The presence of potentially toxic photosynthetic and pathogenic microbes may also favor the spread of turfs in coral reefs, along with a potential increase in photosynthesis and heightened adaptation to changing environmental conditions [15]. For example, bacteria that are pathogenic to corals in the turf can kill the corals, which then serve as substrates for turf spreading [16]. In addition, turf overgrowth above corals can also lead to oxygen depletion, anoxia, and tissue death in corals [17], [18]. Turfs release higher concentration of dissolved organic carbon (DOC) [18]-[20], which stimulate rapid growth of heterotrophic microbes [19], [21] and may promote shifts towards copiotrophic and potentially pathogenic microbial communities [20]. However, previous studies have not addressed the functional metagenomic diversity of turfs. It is currently unclear whether turfs have a specific stable microbiome or several types of turf microbiomes co-existing in the same reef system.

Turf has been recently defined as a holobiont, or an assemblage of macrorganisms (mainly algae) and microrganisms, representing an unified functional entity [7]. Analogous to the coral holobiont concept, both host and microbiome produce extracellular products that permit the holobiont to closely interact [22]. In the coral holobiont, the host is a Cnidarian. Symbiotic microbes and *Symbiodinium* grow on the coral mucus and inside the coral cells. In contrast, a single host is not evident in turf [23]. The host could be a filamentous cyanobacterium due to their abundance, dimension (large macroscopic filaments), and copious production of exudates (dissolved organic matter) [5]-[7]. However, recent critiques of the holobiont concept have led to substantial turmoil [24] amongst researchers, suggesting that further evidence is required to prove that turf is a holobiont. For instance, the stability of turf assemblages across both different reef locations and seasons is unclear.

Previous researchers have suggested that cyanobacteria are the dominant microbial group in turf, representing more than half of the microbes in turf [25]-[27]. Turfs in the Mascarene Archipelago (Indian Ocean) are dominated by species of the genera *Hydrocoleum*, *Anabaena*, *Symploca*, *Leptolyngbya*, and *Lyngbya* [27]. Furthermore, cyanobacteria can play a crucial role in the physical structuring of turf due to their filamentous (turf-like) nature. Cyanobacteria significantly contribute to photosynthesis and N cycling in benthic communities.

In contrast with previous studies, Barott and co-workers [1] used 16S rRNA pyrosequencing to estimate that 7,700 different types of bacteria are associated with turf. Approximately 50% of the identified sequences were related to Proteobacteria. Turf also appeared to be a source of phototrophic bacteria (e.g., Rhodobacteraceae), acid-tolerant bacteria (e.g., Acidovorax, Lactobacillus), and potentially pathogenic bacteria (e.g., Vibrio and Bacteroidetes). A recent analysis of 38 turf samples from the Line Islands resulted in an estimated 18,065 Operational Taxonomic Units (OTUs) deemed to be stable symbionts via 16S rRNA pyrosequencing [7]. The most abundant

and stable turf symbionts are Alphaproteobacteria of the orders Rhodobacterales, Rhizobiales, and Rhodospirales [7]. Hester et al. introduced the stable and sporadic community concept, wherein bacterial communities are neither ubiquitous nor specific across turf samples [7], [28].

These previous studies inspired us to test the hypothesis that turfs have a conserved genomic and metabolic signature, with similar communities and functional gene repertoires spread across different Abrolhos reef locations and seasons. The assemblage of a diverse metagenome could explain why turfs are becoming a dominant component of reefs. In contrast with previous studies, we applied metagenomics to uncover the major types of metabolism found in turfs. Metagenomics has been used to investigate the microbial metabolic potential across a variety of marine and terrestrial environments [12], [29]-[31], whereas the majority of turf studies have relied on the diversity analysis of partial 16S rRNA sequences, which cannot realize the functional potential or diverse functional gene repertoire of turf (e.g., photosynthesis, nitrogen metabolism, and sulfur metabolism).

In the present study we tested the following hypotheses for the microbial composition of the turf from the Abrolhos reefs: H1) the taxonomic and H2) functional composition of the turf are conserved in space (three different locations, inside and outside of protected areas) and time (at two different seasons of the year, summer and winter); H3) the abundance of key genes (e.g. oxygenic and anoxygenic photosynthesis, and chemosynthesis) are different among turf and other benthic holobionts (corals and rhodoliths) and seawater. We also performed morphologic and pigment characterization of turf which enabled us to assess their healthy status. The functional diversity assessment obtained by metagenomics allows us to infer potential adaptive advantages that occur during the competition for space amongst the benthic organisms in reef systems. The conserved nature of the functional composition of turf

metagenomes observed in this study hints to a stable core of complementary functions in the turf.

## Materials and Methods

### Study site and sample collection

This study was conducted at the Abrolhos Bank (16°40′,19°40′S/39°10′, 37°20′W) at near-shore and off-shore locations. In total, 19 turf samples were collected by SCUBA divers from 10-15 m depths; eleven samples were collected on March 10, 2013 (austral summer) and eight samples were collected on October 15, 2013 (austral winter). The samples were collected from three reefs sites: (A) Pedra de Leste (PL, 39º2'00"W/17º46'00"S), (B) Parcel dos Abrolhos (PAB, 17°57′32.7″S/38°30′20.3″W), and (C) Archipelago/Mato Verde (AR, 17º96'43"S/38º70'06"W for March, and 17º57′76.4″S/38º41′73.7″W for October) (S1 Fig). The sampling design allowed for spatial (inner vs. outer shelf) and temporal comparisons. We selected locations inside and outside protected marine areas with different degrees of degradation to account for wide environmental variation. The areas located within Abrolhos National Park had higher coral coverage and higher fish biomass than those located outside the park [14]. To characterize the metagenomes and pigment profiles of the nineteen turf samples, we aimed to shed light on the features that enable this multispecies assembly system to be the most abundant benthic organisms in the Abrolhos Bank. Each turf sample was collected a minimum of 10 meters away from the previous. At each site, 2.3 cm2 samples of turf were carefully collected from the barren bottom using metal spatulas, without coral tissue scraping. The samples were stored in sterile polypropylene tubes and preserved in liquid nitrogen for further metagenomic and pigment analyses. In addition, turf samples collected in parallel were fixed with glutaraldehyde (1% final conc.) in seawater and further examined with light microscopy. The remaining samples were fixed in 4% formaldehyde to determine their ash-free dry weights. The sampling

was authorized by the Brazilian Environmental Agency, Instituto Chico Mendes de Conservação da Biodiversidade (SISBIO license no. 21811–1).

## DNA extraction and sequencing

The samples were separately ground in liquid nitrogen using a mortar and pestle, and the DNA was extracted with a 2% hexadecyl-trimethyl-ammonium bromide (CTAB) extraction buffer (pH 8.0) [0.5 M EDTA, 1 M Tris-HCl, 5 M NaCl, 2% CTAB], followed by a Chloroform:Isoamyl Alcohol (24:1, v/v) step. The disaggregated material was scraped into 2.0 mL microtubes containing preheated (65 °C) extraction buffer at a 1:5 ratio (0.5 mL). RNAse (10 mg mL–1, Sigma-Aldrich, St Louis, MO, USA) was added, and the tubes were incubated for 20 min at 55 °C. Next, Proteinase K (10 mg mL–1, Sigma-Aldrich, St Louis, MO, USA) was added, and the identical incubation conditions were provided as above. An equal volume of Chloroform:Isoamyl Alcohol mixture (24:1) was added to the extract and mixed by gentle inversion for 5 to 10 min to form a uniform emulsion. The mixture was centrifuged at 8,000 rpm for 10 min at room temperature. The aqueous phase was pipetted out gently, avoiding the interface. To the above solution, 3 M NaOAc and 0.6 of the total solution volume of cold isopropanol (-20 ºC) was added, and the tubes were incubated at -4 ºC for 20 min. The mixture was then centrifuged at 10,000 rpm for 30 min. The pellet was washed (2x) with 70% cold ethanol (-20 ºC), dried at room temperature, resuspended in 50 µL of TE buffer [10 mM Tris-HCl, 1 mM EDTA, pH 8.0], and stored at −20 °C. DNA extraction from the sample was carried out in duplicate. The solutions and buffers were autoclaved at 121 °C at 15 psi. The stock solutions of 10 mg mL–1 of RNase and Proteinase K were prepared according to the user's manual. The integrity of the DNA samples was evaluated using electrophoresis on 1% agarose gels with GelRedTM (Biotium Inc., Hayward, CA) to verify its quality, and the DNA purity was assessed with a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Accurate DNA quantification was obtained using a Qubit® 3.0 Fluorometer (Life Technologies-

Invitrogen, Carlsbad, CA, USA). The DNA libraries were generated using a Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The size distribution of the libraries was evaluated using a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA), and DNA quantification was obtained using 7500 Real Time PCR (Applied Biosystems, Foster City, CA, USA) and KAPA Library Quantification Kits (Kapa Biosystems, Wilmington, MA, USA). Paired-end sequencing (2 × 250 bp) was performed on a MiSeq machine (Illumina, San Diego, CA, USA).

## Bioinformatics and statistical analysis of metagenomes

The fastq files generated by Illumina sequencing were qualitatively evaluated using FASTQC (v.0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [32]. The sequences were preprocessed with PRINSEQ (v0.20.4, http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi) [33] to remove low quality DNA sequences (Phred score < 30), duplicates, and short sequences (< 35 bp). Paired-ended Illumina reads were merged using She-Ra software with default parameters and a quality metric score of 0.5 [34]. Sequence annotation was conducted via Metagenome Rapid Annotation using the Subsystem Technology (MG-RAST) webserver (http://metagenomics.nmpdr.org/) [35], using the following cut-off parameters: e-value lower of 1e-5, 60% of minimum sequence identity and at least 15 bp alignment length. Taxonomic annotation was performed using the GenBank database, (http://www.ncbi.nlm.nih.gov/) and functional annotation was completed using the SEED database [36]. The statistical analyses were performed with R version 3.0.3 [37], except where indicated. The abundance and multivariate figures were plotted with the ggplot2 and reshape packages [38], [39]. To test the hypotheses that the taxonomic (H1) and functional (H2) composition of the turf are conserved in space and time, and that the abundance of key genes (e.g., photosynthesis and chemosynthesis) are different among turf and other benthic holobionts (corals and rhodoliths) and seawater (H3), Permutational Multivariate Analysis of Variance

(PERMANOVA) was performed using the “adonis” function of Vegan package [40] (Bray-Curtis distances and 999 permutations).

A collection of 22 metagenomes corresponding to corals, rhodoliths and seawater were retrieved from MG-RAST: eight metagenomes from coral (healthy and diseased) [41], six from rhodolith [42], [43], and eight from seawater [14] (Table 1). All the metagenome samples were from the Abrolhos Bank (Table 1) and were annotated with same databases to diminish possible annotation biases.

**Table 1. General features of each metagenomic sample of the benthic organisms (turf, coral, and rhodolith) and seawater collected from the Abrolhos Bank.**

| MG-RAST ID | Sample Name | Sample Type / Coral Disease Status | Reef | PMID | Sampling date | Depth (m) | Lat. | Lon. | Technology of Sequencing | Size (reads) | Size (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4561212.3 | PL.1\|M | Turf | Pedra de Leste | this study | 10/03/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 172,944 | 31,448,043 |
| 4561207.3 | PL.2\|M | Turf | Pedra de Leste | this study | 10/03/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 29,495 | 5,111,519 |
| 4561206.3 | PL.3\|M | Turf | Pedra de Leste | this study | 10/03/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 554,519 | 95,157,293 |
| 4561211.3 | PL.4\|M | Turf | Pedra de Leste | this study | 10/03/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 198,979 | 35,519,165 |
| 4561210.3 | AR.1\|M | Turf | Archipelago (Mato Verde) | this study | 10/03/2013 | 12 | -17.964 | -38.702 | Illumina/Solexa | 262,886 | 46,552,396 |
| 4561205.3 | AR.2\|M | Turf | Archipelago (Mato Verde) | this study | 10/03/2013 | 12 | -17.964 | -38.702 | Illumina/Solexa | 280,308 | 51,419,363 |
| 4561203.3 | AR.3\|M | Turf | Archipelago (Mato Verde) | this study | 10/03/2013 | 12 | -17.964 | -38.702 | Illumina/Solexa | 659,796 | 168,993,630 |
| 4561213.3 | AR.4\|M | Turf | Archipelago (Mato Verde) | this study | 10/03/2013 | 12 | -17.964 | -38.702 | Illumina/Solexa | 575,503 | 99,195,357 |
| 4561208.3 | PAB.2\|M | Turf | Parcel dos Abrolhos | this study | 10/03/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 80,324 | 14,650,440 |
| 4561202.3 | PAB.3\|M | Turf | Parcel dos Abrolhos | this study | 10/03/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 102,782 | 26,888,594 |
| 4561209.3 | PAB.4\|M | Turf | Parcel dos Abrolhos | this study | 10/03/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 176,395 | 32,257,528 |
| 4564639.3 | PL.2\|O | Turf | Pedra de Leste | this study | 15/10/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 107,665 | 18,350,484 |
| 4564642.3 | PL.4\|O | Turf | Pedra de Leste | this study | 15/10/2013 | 12 | -17.783 | -39.051 | Illumina/Solexa | 186,167 | 33,060,641 |
| 4564647.3 | AR.1\|O | Turf | Archipelago (Portinho Norte) | this study | 15/10/2013 | 12 | -17.959 | -38.701 | Illumina/Solexa | 45,168 | 8,118,890 |
| 4564646.3 | AR.2\|O | Turf | Archipelago (Portinho Norte) | this study | 15/10/2013 | 12 | -17.959 | -38.701 | Illumina/Solexa | 624,907 | 152,157,722 |
| 4564644.3 | AR.3\|O | Turf | Archipelago (Portinho Norte) | this study | 15/10/2013 | 12 | -17.959 | -38.701 | Illumina/Solexa | 249,663 | 74,875,522 |
| 4564648.3 | PAB.1\|O | Turf | Parcel dos Abrolhos | this study | 15/10/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 535,455 | 134,542,669 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4564643.3 | PAB.2\|O | Turf | Parcel dos Abrolhos | this study | 15/10/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 151,800 | 29,152,445 |
| 4564645.3 | PAB.3\|O | Turf | Parcel dos Abrolhos | this study | 15/10/2013 | 12 | -17.998 | -38.671 | Illumina/Solexa | 586,084 | 146,060,007 |
| 4447483.3 | California | Water | California | 22679480 | 29/01/2009 | 12 | -18.102 | -38.591 | 454/Roche | 167,513 | 74,889,416 |
| 4447551.3 | PAB5 | Water | Parcel dos Abrolhos | 22679480 | 28/01/2009 | 20 | -17.959 | -38.506 | 454/Roche | 126,741 | 53,914,839 |
| 4447862.3 | Timbebas | Water | Timbebas | 22679480 | 27/01/2009 | 5.6 | -17.478 | -39.028 | 454/Roche | 149,734 | 58,969,826 |
| 4448427.3 | Sebastião Gomes | Water | Sebastião Gomes | 22679480 | 26/01/2009 | 4.7 | -17.912 | -39.129 | 454/Roche | 10,906 | 3,768,178 |
| 4453304.3 | Timbebas 2010 | Water | Timbebas | 22679480 | 26/02/2010 | 3.2 | -17.478 | -39.028 | 454/Roche | 67,439 | 27,057,181 |
| 4453305.3 | Pedra do Leste 2010 | Water | Pedra de Leste | 22679480 | 27/02/2010 | 4.7 | -17.784 | -39.051 | 454/Roche | 31,365 | 13,219,777 |
| 4453371.3 | PAB5 2010 | Water | Parcel dos Abrolhos | 22679480 | 23/02/2010 | 2.3 | -17.959 | -38.506 | 454/Roche | 79,476 | 29,639,075 |
| 4453372.3 | Sebastião Gomes 2010 | Water | Sebastião Gomes | 22679480 | 23/02/2010 | 5.6 | -17.912 | -39.129 | 454/Roche | 39,792 | 11,785,037 |
| 4477768.3 | R1 | Rhodolith | Buracas | 23985749 | 12/7/2010 | 27 | -17.914 | -37.909 | 454/Roche | 17,863 | 6,729,554 |
| 4477767.3 | R2 | Rhodolith | Buracas | 23985749 | 12/7/2010 | 27 | -17.914 | -37.909 | 454/Roche | 27,229 | 10,794,633 |
| 4477763.3 | R3 | Rhodolith | Buracas | 23985749 | 12/6/2010 | 43 | -17.856 | -38.12 | 454/Roche | 32,488 | 11,605,525 |
| 4477766.3 | R4 | Rhodolith | Buracas | 23985749 | 12/6/2010 | 43 | -17.856 | -38.12 | 454/Roche | 18,171 | 6,999,671 |
| 4478209.3 | R5 | Rhodolith | Buracas | 23985749 | 12/6/2010 | 43 | -17.856 | -38.12 | 454/Roche | 29,791 | 10,856,929 |
| 4477765.3 | R6 | Rhodolith | Buracas | 23985749 | 12/7/2010 | 51 | -17.914 | -37.909 | 454/Roche | 19,207 | 6,633,451 |
| 4463374.3 | SGS2 | Coral / Healthy | Sebastião Gomes | 23314124 | 23/02/2010 | 3.2 | -17.912 | -39.129 | 454/Roche | 27,275 | 10,566,518 |
| 4462249.3 | GiSGW2 | Coral / Diseased | Sebastião Gomes | 23314124 | 23/02/2010 | 3.2 | -17.912 | -39.129 | 454/Roche | 20,483 | 7,615,646 |
| 4463363.3 | SGW5 | Coral / Diseased | Sebastião Gomes | 23314124 | 23/02/2010 | 3.2 | -17.912 | -39.129 | 454/Roche | 29,048 | 11,075,785 |
| 4463368.3 | P5S2 | Coral / Healthy | Parcel dos Abrolhos | 23314124 | 23/02/2010 | 3.2 | -17.959 | -38.506 | 454/Roche | 40,671 | 16,401,293 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4463367.3 | P5S4 | Coral / Healthy | Parcel dos Abrolhos | 23314124 | 23/02/2010 | 5.6 | -17.959 | -38.506 | 454/Roche | 11,949 | 4,745,703 |
| 4463366.3 | P5S5 | Coral / Healthy | Parcel dos Abrolhos | 23314124 | 23/02/2010 | 5.6 | -17.959 | -38.506 | 454/Roche | 26,551 | 10,563,130 |
| 4463359.3 | P5W2 | Coral / Diseased | Parcel dos Abrolhos | 23314124 | 23/02/2010 | 5.6 | -17.959 | -38.506 | 454/Roche | 7,406 | 2,725,373 |
| 4463358.3 | P5W5 | Coral / Diseased | Parcel dos Abrolhos | 23314124 | 23/02/2010 | 5.6 | -17.959 | -38.506 | 454/Roche | 17,351 | 6,328,106 |

The metadata were retrieved from a MG-RAST server framework. The details of each collected sample are provided. The sampling reef sites are indicated, as well as their exact locations (latitude and longitude). The sampling depth (meters) and the sequencing technology used for each metagenome are listed. The total (bp) and read sizes given for the metagenomes are post quality-control values.

To verify if the samples from turf, rhodoliths, corals and seawater would group together according to their most important (see below) metagenomic (taxonomic and functional) features, a principal component analysis (PCA) was performed, using the "rda" function of Vegan package [40]. The goal of PCA is to explain as much of the variance as possible in the first few components, thus reducing the complexity of the data by combining related variables. We used the supervised Random Forest algorithm to select the variables (bacterial orders and subsystems) based on the "Mean Decreasing Accuracy" values [44] using the "randomForest" function of Random Forest R package [45]. The Variable Importance Measure Metric allows us to identify the most important variables for discriminating sample groups [46]. There are some approaches to estimate variables importance in separating samples into groups in supervised Random Forests. We have used mean decrease in accuracy to select the most important variables, which is determined in the error-calculating phase [46]. The purity of a node in a Random Forest is measured by the Gini index, so the mean decrease in Gini is a measure of variable contribution to the nodes and leaves homogeneity in the Random Forest. Random Forests are a robust approach for clustering metagenomes and metabolic processes in microbial communities from different environments [46]. To standardize the metagenome sizes, we present the metagenomic data as relative abundances (the number of sequences of a given taxa or subsystem of a metagenome divided by the total number of annotated sequences of this metagenome). Percentage data were transformed to arcsin($\sqrt{x}$) for multivariate analysis. P-values of $<0.05$ were considered statistically significant.

## Analysis of pigments and ash-free dry weights

To assess the composition of the photosynthetic components of the turf, we quantified the pigments and determined their ash-free dry weights in the turf samples collected in October, 2013. Chlorophylls *a, b* and $c_{1+2}$ (Chl*a*, Chl*b*, and Chl*c*, respectively), Pheophytin *a* (Pheo*a*), and Phycobiliproteins (PBPs) were measured. Chl*a*, Chl*b*, Chl*c*, and Pheo*a* were measured via spectrofluorometry using duplicate 0.2 g aliquots of turf material per sample. The pigments were extracted in 90% acetone:water. The cells were ruptured by grinding the material with a glass rod in the bottom of a glass test tube. The tubes were stored overnight at 4 ºC in the dark. After centrifugation, the fluorescence properties of the acetonic extracts were measured on a Varian Cary Eclipse® (wavelength accuracy ± 1.0 nm from 200-900 nm) spectrofluorometer (Agilent©, USA). The chlorophyll concentrations were assessed using a modified version of Neveux and Lantoine's [47] method, which was described in [48]. Data acquisition was performed by recording the fluorescence emission spectra for each of the 31 excitation wavelengths (3 nm increments from 390-480 nm). The emission spectra were recorded at 2 nm intervals from 615-715 nm and yielded 51 points for each spectrum. The pigment concentrations were estimated from the resulting 1,581 data points. A least squares approximation technique was used to discard negative solutions.

For the quantification of the PBPs, the turf samples were suspended in 5 ml of 20 nM sodium acetate buffer, pH 5.5 (supplemented with 3 mM sodium azide and 10 mM disodium EDTA). Turf cells were ruptured via sonication intercalated with freeze and thaw cycles. The crude extracts were treated with 1% (w/v) streptomycin sulfate for 30 min at 4 ºC and centrifuged at 10,000 g for 10 min at 4 ºC to precipitate the cellular debris. The phycoerythrin, phycocyanin, and allophycocyanin concentrations were calculated from measurements of their optical densities at 564, 620, and 650 nm in a spectrophotometer. The quantification of the PBPs was based on the equations described by [49] and [50]. To obtain the ash-free dry weight (AFDW) of the turf

samples, the material was dried at 60 ºC for 24 h, weighed, oxidized (ashed) in a muffle furnace at 450 ± 10 ºC for 4 h, cooled to room temperature in a desiccator, and then re-weighed.

## Results

### Morphological and pigment analyses

The turfs grew over the corals in the 19 samples collected at three locations. The turfs were phenotypically heterogeneous, exhibiting different colors (green, brown and red) and textures (S1 Fig). Microscopic analyses revealed that the turf assemblages were composed of filamentous, non-heterocystous cyanobacteria in a network with embedded seaweeds. Cyanobacteria were the structuring organisms in the turf, and morphotypes similar to the genera *Oscillatoria* and *Leptolyngbya* were commonly found. The minor turf components consisted of red (*Florideophyceae* members), green (filamentous *Bryopsis* sp. and flattened calcareous algae *Halimeda* sp., Ulvophyceae), and brown (fleshy *Dictyota* sp., Phaeophyceae) seaweeds. The Chlorophyll *a* per unit of turf area ranged from 7.2 (sample PAB.3|O) to 93.8 (sample PL.1|O) µg Chl*a* $cm^{-2}$ (average 37.2; median 24.3 µg Chl*a* $cm^{-2}$) (S1 Table).

Chlorophyll *b* accounted for a small fraction of the total chlorophyll concentration (average 8.5 µg Chl*b* $cm^{-2}$). Chlorophyll $c_{1+2}$ accounted for a minute amount of the total chlorophyll concentration (max. 3.2%). The AFDW of the turf samples averaged 0.29 mg $cm^{-2}$, ranging from 0.11 (sample PAB.3|O) to 0.53 (sample PAB.1|O) mg $cm^{-2}$. The Pheophytin *a*:Chlorophyll *a* ratio, a proxy for the degree of degradation in the photosynthetic community of the turf system, averaged 12.7%, ranging from 3.8% (representing the healthiest conditions) to 19.9% in the samples with the highest degree of chlorophyll degradation. Phycoerythrin was the main phycobiliprotein detected in the samples; its concentrations averaged 0.45 mg $ml^{-1}$, with the maximum value found in the AR.3|O sample (1.9 mg $ml^{-1}$), suggesting high abundance of cyanobacteria in turfs.

## Taxonomic assignment of turf metagenomic sequences

The taxonomic assignments indicated that bacteria contributed an average of 85.92% (total N = 3,558,854) of the sequences annotated, ranging from 49.5% to 97.8%. Proteobacteria (40.57%, ± s.d. 10.36, N = 1.548,192), Cyanobacteria (35.04%, ± s.d. 15.5, N = 1.337,196), and Bacteroidetes (11.12%, ± s.d. 4.25, N = 424,185) were dominant in the turf microbiome (Fig 1A). Eukarya contributed 5.78% of the sequences on average (total N = 170,580), ranging from 1.45% to 42.11% of the total of all the metagenomes. Among the 19 samples, the macroalgae Chlorophyta (Chlorophyceae and Ulvophyceae), Pheophyta (Pheophyceae), and Rhodophyta (Florideophyceae) corresponded to, on average (% ± standard deviation), only 6.84 ± 5.7, 1.13 ± 1.12, and 1.09 ± 2.02, respectively, of the total Eukarya sequences. The Archaea domain and viruses contributed less than 1% of the sequences (a total of approximately 16,000 for Archaea and 2,000 for viruses). Unassigned and unclassified sequences corresponded to 7.81% of the average (total N = 315,994) (see S2 Table).

Turf samples had a statistically indistinguishable taxonomic profile at order level based on MG-RAST annotation, confirming the study hypothesis H1. No statistical difference in turf metagenomes composition in different locations and seasons were detected (Fig 1A and S3 Table).

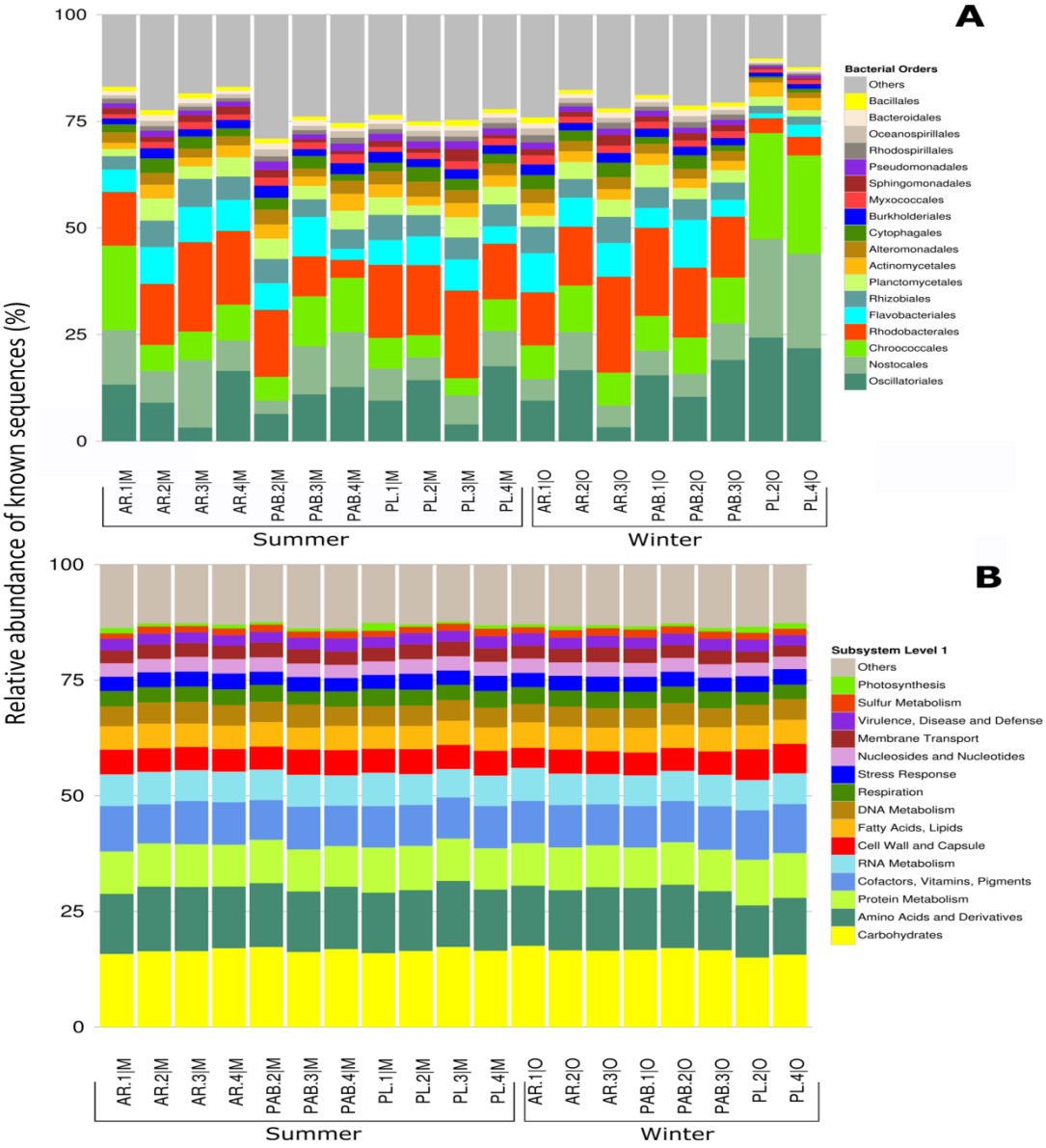



**Figure 1. Turf system taxonomic and functional profiles.**

A) The major orders and B) subsystems (SEED Level 1) found in the metagenomes of the turf samples from the Abrolhos reefs. The percent correspond to the abundance normalized by the total of known sequences, i.e., genes with defined function. See Results section for the contribution of the major metabolisms. Sulfur metabolism accounted for 1.52% of the total metagenomes (± s.d. 0.14, N = 26,716), whereas nitrogen metabolism accounted for 1.13% (± s.d. 0.1, N = 19,939), and photosynthesis was responsible for 0.77% of the total metagenomes (± s.d. 0.3, N = 13,508). Ammonia assimilation related genes represented greater than 50% (52.3% ± s.d. 4.2) of the total genes in the nitrogen metabolism subsystems of the turf metagenomes, whereas $N_2$ fixation genes represented 10.1% (± s.d. 5.7) of the total genes in this subsystem. Dissimilatory nitrate and nitrite ammonification corresponded to 18% ± s.d. 3.75 of the total genes in the nitrogen metabolism in this study). The sequences assigned as Clustering-based Subsystems (14.1%; N = 331,723 of the total) and Miscellaneous Subsystems (8.4%; N = 196,810 of total) were not included in the analysis.

The major taxa contributing to the core microbiome of these systems comprised typical aerobic anoxygenic photosynthesis (AANP) bacteria (e.g., Rhodobacteriales related to *Congregibacter*), Cyanobacteria (Oscillatoriales, Nostocales and Chroococcales), Flavobacteriales (related to *Maribacter*), Rhizobiales (related to *Roseobacter*, *Silicibacter* and *Dinoroseobacter*), and Planctomycetes (S4 Table).

The cyanobacterial sequences varied from approx. 19.8% (N = 7,036) in PL.3|M to approx. 72.25% (N = 58,906) in PL.2|O, whereas the Rhodobacterales sequences varied from approx. 4.4% (N = 3,383) in PL.2|O to approx. 22.5% (N = 37,695) in AR.3|O. The cyanobacteria sequences were related to *Trichodesmium* sp. (Oscillatoriales). Microscopic examination of the fixed samples (formaldehyde) revealed large quantities of cells that were phenotypically identified as *Leptolyngbya* sp., which is closely related to *Trichodesmium*.

## Metabolic diversity of turf metagenomes

According to the subsystems analysis, the turf metagenomes had a homogeneous profile, also confirming our hypothesis H2 (Fig 1B and S5 Table). The five most abundant subsystems (carbohydrates; amino acids and derivatives; protein metabolism; cofactors, vitamins, prosthetic groups, pigments; RNA metabolism) accounted for greater than 50% of all the classified metagenomic sequences (54.9% ± s.d. 3.5 for the five most abundant subsystems; N = 961,649 of the total) (see Fig 1B). Cofactors, vitamins, prosthetic groups and pigments made up the fourth most abundant subsystem; with the related folates and pterins (at level 2 of the classification according to the SEED database) being the fifth highest contributors (3.31% ± s.d. 0.17; N = 74,747), behind the level 2 core functions central carbohydrate metabolism, protein biosynthesis, and RNA processing and modification. Iron-sulfur metabolism involving the oxidative stress-related proteins of the YgfZ family (at SEED classification level 3)

was abundant in all the metagenomes (March: N = 16,216; 1.5% ± s.d. 0.106 of all the metagenomes; October: N = 17,656; 1.5% ± s.d. 0.104 of all the metagenomes).

The photosynthesis subsystem had a total contribution ranging from 0.45% (PL.3|M) to 1.3% (PL.2|M) of the entire dataset. Approx. 46.7% (± s.d. 12.7), 33.82% (± s.d. 2.19), and 19.46% (± s.d. 1.74) of the sequences within the photosynthesis subsystem (N = 19,783) belonged to Cyanobacteria, Bacteria, and Eukarya, respectively. A total of 13,204 reads belonging to various chloroplast genes (i.e., rbc, pet, psa, psb, ndh, LSU, and SSU) were identified as Cyanobacteria (52% ± s.d. 37.01), Eukarya (45% ± s.d. 34.3), and Bacteria (3% ± s.d. 4.05). These chloroplast genes represent oxygenic phototrophs (Fig 2A), whereas other light harvesting genes (e.g., bchl, puf, and rhodopsin-based phototrophy genes, N = 6,969) indicated the presence of anoxygenic phototrophy in turfs (Fig 2B). Chemolithotrophic sulfur oxidation genes (i.e., sox genes, 1.9%, N = 739) were also present in the turfs (Fig 2C).

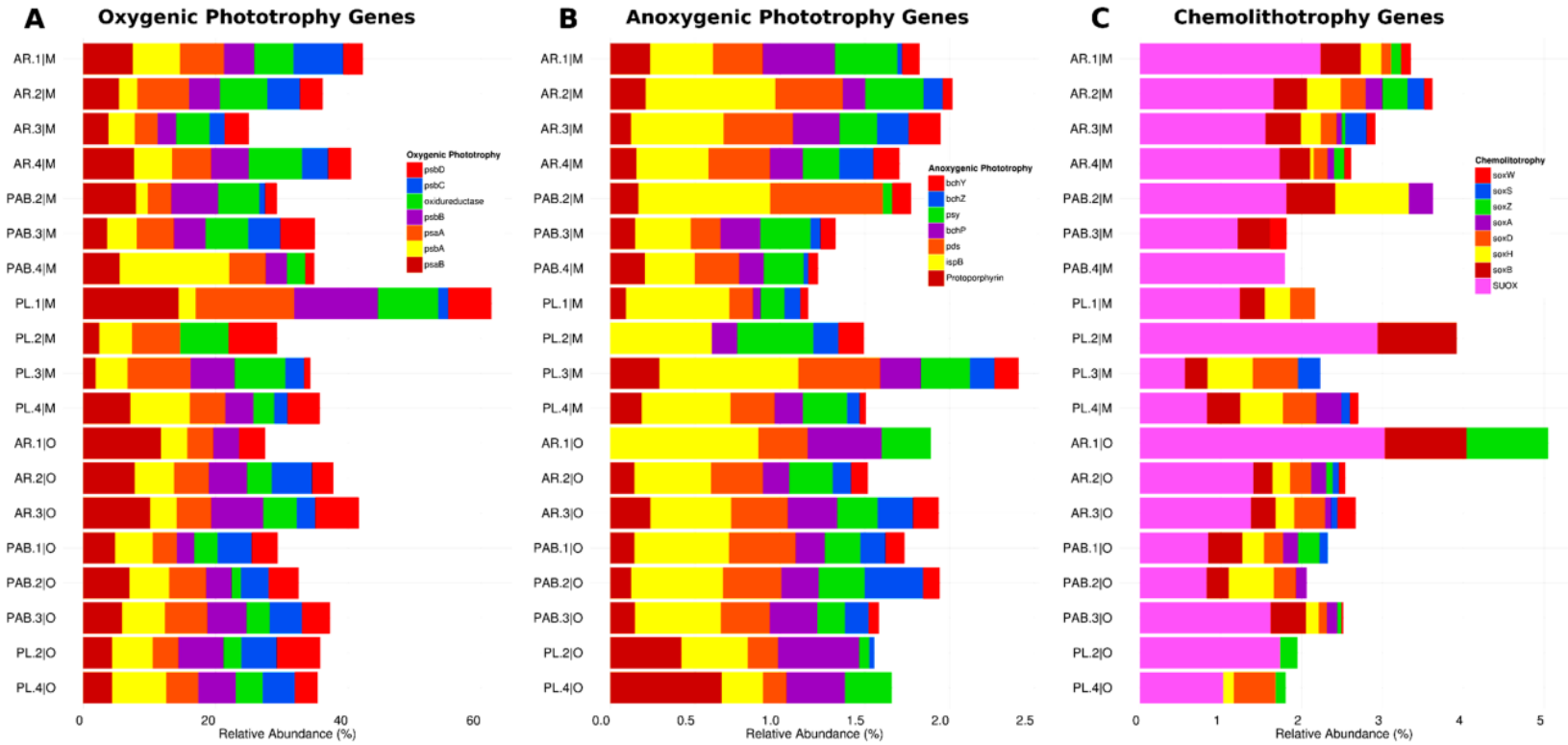



**Figure 2. Most abundant genes from the different lifestyles co-occurring in Abrolhos Bank turfs.**

Genes involved in A) oxygenic photosynthesis, B) anoxygenic photosynthesis, and C) chemolithotrophy from each turf metagenome. Relative abundance of annotated sequences. Genes related to oxygenic photosynthesis are most abundant. The values in panel A varies from 0-60%; B from 0-2.5%, and C from 0-5%.

Sequences related to bacterial secretion systems at the functional level (within the subsystem membrane transport) corresponded to 13,638 out of 51,555 sequences and were identified as T1SS – Type 1 (1.16% of the total membrane transport subsystem, N = 623), T2SS (9.3%, N = 4,978), T3SS (0.9%, N = 460), T4SS (7.1%, N = 3,793), T5SS (0.15%, N = 84), and T6SS (6.9%, N = 3,700) secretion systems. T3SS, T4SS, T6SS interact directly with the host eukaryotic cell membrane, whereas T1SS, T2SS and T5SS secrete proteins into the extracellular milieu. The iron acquisition and metabolism subsystems (N = 18,561) included 6,758 sequences associated with iron acquisition in Vibrio (N = 7,660 sequences were identified as TonB-dependent receptors), 588 sequences of the siderophore pyoverdine (related to *Pseudomonas*), and 1,256 sequences related to iron acquisition *Streptococcus*. Sequences in the virulence, disease and defense subsystem corresponded to 1.7% (N = 1,219; PL.4|O) to 2.2% (N = 1,116; PAB.3|M) of the total metagenomes. Beta-lactamase resistance genes (N = 2,064, 4.9%), multidrug efflux pumps (N = 1,078, 2.5%), multidrug resistance efflux pumps (N = 4,760, 11.2%), and fluoroquinolone resistance genes (N = 6,034, 14.2%), at SEED classification level 3, represented approximately 31.8% of all the genes in the virulence subsystem. Superoxide dismutase (1,039 sequences) and sulfur metabolism sequences (at least 101 different genes, comprising 11 categories of assimilatory and dissimilatory pathways) were also discovered. The major metabolic features are depicted (S2 Fig).

## Turf microbiomes differ from the other microbiomes of the major Abrolhos benthic organisms

Turfs, corals, rhodoliths, and water formed different groups using PCA based on the most abundant taxa (Fig 3A) and functions (Fig 3B). All the turfs formed a tight group clearly separated from other abundant benthic organisms, also confirming our hypothesis H3. The two first axes explained a large proportion of the taxonomic variation between the samples (43.59% for PC1 and 23.02% for PC2) (Fig 3A).

Nostocales and Oscillatoriales highly influenced the turf sample grouping (Fig 3A), in agreement with the RF results (S6 Table).

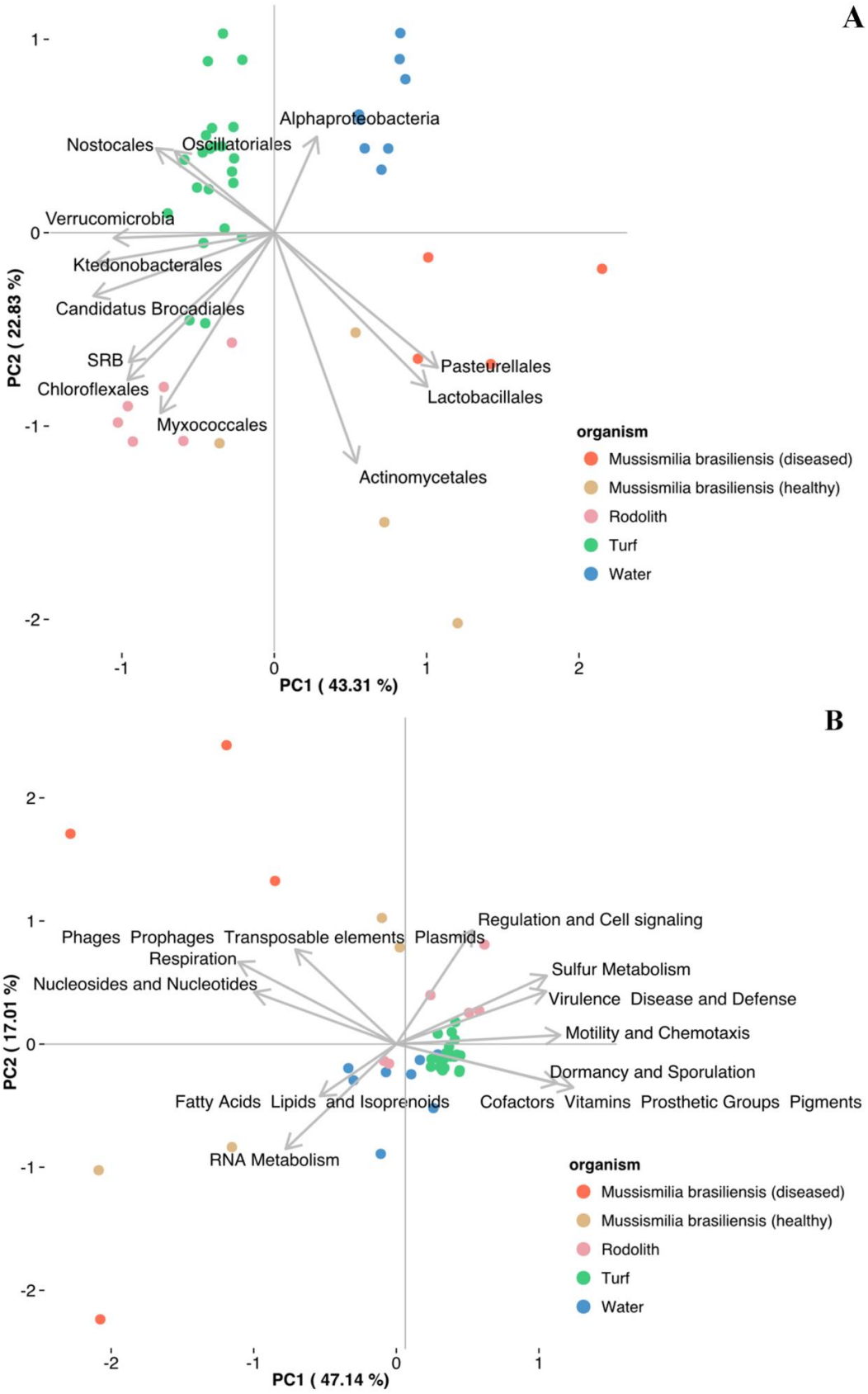


**Figure 3. Principal Component analysis of the 41 metagenomes using the top eleven variables identified from the Random Forest analysis.**

Turf taxonomic and functional potential compared with other abundant benthic organisms and seawater from Abrolhos environments. A) The major orders and B) subsystems (SEED Level 1) found in the metagenomes of the turf samples from Abrolhos reefs compared with coral and rhodolith metagenomic data. Principal Component Analysis using 11 Level 1 SEED Subsystems selected based on the "Mean Decreasing Accuracy" of the Random Forest analysis. Metagenomic relative abundance data in percentages were transformed to arcsin(√x). Previously published seawater, coral, and rhodolith metagenome datasets were used in this analysis (Bruce et al., 2012; Garcia et al., 2013; Cavalcanti et al., 2013, 2014, respectively).

Likewise, the two first axes explained a large proportion of the functional variation between the samples (47.14% for PC1 and 17.01% for PC2) (Fig 3B). The following important functional features were found: virulence, disease, and defense; sulfur metabolism; motility and chemotaxis; dormancy and sporulation; cofactors, vitamins, prosthetic groups and pigments (Fig 3B, for RF results, see S7 Table). The number of nitrogen metabolism metagenomic sequences in the turfs was higher than those of the corals, rhodoliths, and seawater from Abrolhos (S8 Table).

Additionally, ANOVA results showed that the abundance of genes involved in oxygenic photosynthesis, anoxygenic photosynthesis and chemosythesis (Fig 2) are different among turfs and other benthic holobionts (corals and rhodoliths) and seawater (Table 2 and S9 Table).

**Table 2. Genes abundance.**

The abundance of genes related to oxygenic and anoxygenic photosynthesis and chemosynthesis are different among turf and other benthic holobionts (corals and rhodoliths) and seawater (H3). ANOVA results of total bacterial abundance. DF, degrees of freedom; SS, sum of squares; MS, mean sum of squares. Bonferroni method was used to adjust p values.

| | DF | SS | MS | F value | *P* value | *P* value (adjusted) |
|---|---|---|---|---|---|---|
| **Oxygenic** | | | | | | |
| Model | 3 | 4572 | 1524.1 | 6.705 | 0.001 | 0.003 |
| Residuals | 37 | 8411 | 227.3 | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Anoxygenic** | | | | | | |
| Model | 3 | 31.950 | 10.651 | 4.875 | 0.006 | 0.018 |
| Residuals | 37 | 80.840 | 2.185 | | | |
| **Chemolithotrophy** | | | | | | |
| Model | 3 | 10974 | 3658 | 16.26 | <0.001 | <0.001 |
| Residuals | 37 | 8323 | 225 | | | |

## Discussion

This study highlights the taxonomic and functional metagenomic consistency of turf as the most dominant benthic component of the Abrolhos reefs. Our hypotheses (H1 and H2) that turfs have homogeneous microbial and functional compositions across space and time were confirmed. We also confirmed our hypothesis H3 as turfs had particular sets of genes involved in oxygenic and anoxygenic phothosynthesis and chemosynthesis not present in other benthic organisms (corals and rhodoliths) and seawater. The functional properties of the turf metagenomes hint a stable microbiome association. The association of cyanobacteria with aerobic, anaerobic, and AANP bacteria may increase turf stability and their competitive advantage (photosynthesis, nitrogen, and sulfur metabolism) in comparison with other autotrophic benthic components, such as corals and rhodoliths. The unexpected homogeneity in both taxonomy and function, despite their diversity in geographical origin, sampling time, biomass per unit of area, and pigments profile of their photosynthetic component (i.e., chlorophylls and phycoerythrin concentrations and pheophytin:Chl*a* ratios), corroborates this potential stable association. Furthermore, the metagenomic profiles of these systems clearly differed from those of other Abrolhos benthic organisms, which points toward the existence of a specific turf assemblage. The core microbiome of the turf consists of an assemblage of primary producers (both phototrophic and lithotrophic) and heterotrophic microbes possibly bound by complementary synergistic functions.

## Photosynthesis in turfs

Oxygenic (by cyanobacteria and algae) and anoxygenic photosynthesis (e.g., by certain Rhodobacterales, Rhizobiales, and Rhodospirillales), sulfide production from sulfate reduction (by sulfate-reducing bacteria, SRB), and microbial sulfide oxidation overlap and fluctuate by day and night, during which sulfate-reducing bacteria in the surface layer must tolerate oxygen exposure as a result of cyanobacterial photosynthesis during the day [51]. The co-occurrence of different photosynthetic organisms in turf allows them to use light under the typically diverse environmental conditions (i.e., high and low luminous incidence and water turbidity) of reef systems. Eukaryotic photosynthetic algae and cyanobacteria possess Chlorophyll *a*, which permits light capture at both 430 nm (blue light) and 680 nm (red light). Nevertheless, cyanobacteria utilize phycobilisomes (large pigment-protein complexes) to capture photons between the blue and red regions of the spectrum, which are not efficiently trapped by chlorophyll [52]. On the other hand, AANP bacteria possess bacteriochlorophyll, which allows for light harvest at ultraviolet (360 nm) and infrared wavelengths (805 nm and 870 nm). In general, zooxanthellate corals such as *Orbicella*, *Mussismilia,* and *Porites* have a narrower light usage range because they rely upon Chlorophyll *a* containing phototrophs (e.g., *Symbiodinium* and cyanobacteria) [53]. However, endolytic chlorophyll *d*-containing cyanobacteria have been found in association with dead corals [54]. Chlorophyll *d* allows for near infrared (700-740 nm) light harvesting and photosynthesis [55]. Large quantities of $O_2$ can be produced by turfs. Bacterial persistence in these systems is enabled through the production of enzymes (e.g., superoxide dismutase) that avoid oxidative stress by neutralizing reactive oxygen species such as hydrogen peroxide, organic peroxides, and superoxide [6], [56], [57].

## Sulfur cycle in turfs

Sulfate-reducing bacteria (e.g., *Desulfovibrio*, *Desulfobacterium*, and *Desulfobulbus*) have adapted to the oxic conditions of turf surface layers through motility and aggregate formation [51]. The sulfide produced by these bacteria as a respiration product can be utilized by AANP bacteria and cyanobacteria. The latter is commonly found in assemblages with continuous exposure to biogenically produced sulfide, in which fluctuations of anaerobic and aerobic conditions are common [58]. We have shown that turf microbiomes are rich in anaerobic and facultative anaerobic microbes, which corroborates previous studies [7]. Sulfide acts as an electron donor to anoxygenic photosynthesis in *Rhodobacteraceae* members and cyanobacteria. Additionally, sulfide serves as an assimilatory sulfur source and is highly toxic, reacting with various cytochromes, hemoproteins, and other compounds. These reactions inhibit the electron transport chain, blocking respiration as well as oxygenic and anoxygenic photosynthesis. The capacity of turfs to engage in S-cycle reactions indicates their competence to thrive under harsh conditions. Coral reefs provide a habitat enriched in dissolved and particulate organic material, such as amino acids, sugars and dimethylsulfoniopropionate (DMSP), which trigger the development of DMSP-degrading bacteria in turfs. Organic sulfur compounds, particularly DMSP and gas dimethyl sulfide (DMS), are important for the structuring of coral and turf-associated bacterial communities [1], [59], [60]. The present study indicates that sulfur metabolism is an important factor for distinguishing Abrolhos turfs from corals and rhodoliths. In contrast with corals, turfs tolerate higher levels of sulfide, which is a consequence of their core microbiome [58], [61].

Turfs appear to be more stable to environmental changes (e.g., nutrients and temperature) and display higher growth rates and metabolic potential than other reef organisms (e.g., corals and rhodoliths) [62]. Turfs can grow faster than calcifying organisms, broadly occupying benthic habitats, particularly outside of the National Abrolhos Marine Park areas, while corals and rhodoliths grow only a few millimeters

per year [63]. In addition, turfs can obtain limited nutrients in the environment for growth, e.g., iron, through secreted bacterial siderophores [64], [65]-[67]. Antimicrobial production and antimicrobial resistance can modulate the homeostasis of microbial populations because they are also involved in scavenging and the uptake of nutrients [65].

Phototrophic organisms, represented by non-heterocystous and heterocystous cyanobacteria, are the most abundant components of turfs, suggesting that they play an important role in the nitrogen cycle of Abrolhos reef systems. Indeed, the number of nitrogen metabolism metagenomic sequences in the turfs was higher than those of the corals, rhodoliths, and seawater from Abrolhos. Ammonia assimilation related genes represented greater than 50% (52.3% ± s.d. 4.2) of the total genes in the nitrogen metabolism subsystems of the turf metagenomes, whereas $N_2$ fixation genes represented 10.1% (± s.d. 5.7) of the total genes in this subsystem. Ammonia can be produced in turfs via dissimilatory nitrate and nitrite ammonification (18% ± s.d. 3.75 of the total genes in the nitrogen metabolism in this study), a process carried out by the sulfate-reducing bacteria (such as *Desulfovibrio*, *Desulfobacterium*, and *Desulfobulbus*) found in this study and several types of Proteobacteria [68].

## Conclusion

The turfs of the Abrolhos Bank have characteristic microbial communities and conserved metabolic profiles revealed by metagenomics, particularly regarding the coupling between oxygenic photosynthesis, AANP, ammonia assimilation, $N_2$ fixation, and the S cycle. The turf microbiome is involved in the core functions of the turfs and may promote the proliferation of this dominant benthic component in the Abrolhos Bank.

## DNA deposition

All metagenomic data are held at http://metagenomics.anl.gov (Metagenomics RAST Server), according the following MG-RAST IDs: 4561212.3, 4561207.3, 4561206.3, 4561211.3, 4561210.3, 4561205.3, 4561203.3, 4561213.3, 4561208.3,

4561202.3, 4561209.3, 4564639.3, 4564642.3, 4564647.3, 4564646.3, 4564644.3, 4564648.3, 4564643.3, 4564645.3; and fasta files are available at https://marinebiodiversity.lncc.br/files/index.php/s/3HqRMATuGDQWDU6 (fthompson.6.1) (Brazilian Marine Biodiversity Database, BaMBa).

## Acknowledgements

The authors thank Fernando Moraes for his support in the field. This paper is part of the D. Sc. requirements of JMW at the Biodiversity and Evolutionary Biology Graduate Program of the Federal University of Rio de Janeiro (UFRJ).

## Author Contributions

Conceived and designed the experiments: JMW PSS FLT. Performed the experiments: JMW DAT PMM LO LL. Analyzed the data: JMW DAT PMM MT PSS FLT. Contributed reagents/materials/analysis tools: LO LL RV PSS CCT FLT. Wrote the paper: JMW DAT PMM LO PSS FLT.

## References

[1] Barott K, Rodriguez-Brito B, Janouškovec J, Marhaver K, Smith J, Keeling et al. (2011) Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral Montastraea annularis. Environmental Microbiology 13: 1192-1204.

[2] Barott K, Rodriguez-Mueller B, Youle M, Marhaver KL, Vermeij MJA, Smith JE, et al. (2012) Microbial to reef scale interactions between the reef-building coral Montastraea annularis and benthic algae. Proc R Soc B 279, 1655–1664. doi:10.1098/rspb.2011.2155

[3] Roff G. and Mumby P. (2012) Global disparity in the resilience of coral reefs. Trends in Ecology & Evolution 27: 404-413.

[4] Sweet M, Burn D, Croquer A, Leary P. (2013) Characterisation of the Bacterial and Fungal Communities Associated with Different Lesion Sizes of Dark Spot Syndrome Occurring in the Coral Stephanocoenia intersepta. PLoS ONE 8(4): e62580.

[5] Egan S, Harder T, Burke C, Steinberg P, Kjelleberg S, Thomas T (2013) The seaweed holobiont: understanding seaweed-bacteria interactions. FEMS Microbiology Reviews 37: 462-476.

[6] Hollants J, Leliaert F, De Clerck O, Willems A. (2013) What we can learn from

sushi: a review on seaweed-bacterial associations. FEMS Microbiol Ecol 83: 1-16.

[7] Hester E, Barott K, Nulton J, Vermeij M, Rohwer F. (2016) Stable and sporadic symbiotic communities of coral and algal holobionts. The ISME Journal. 10(5):1157-69.

[8] Goldberg WM (2013). The Biology of Reefs and Reef Organisms, University of Chicago Press.

[9] Sheppard CRC, Davy SK, Pilling GM (2009). The Biology of Coral Reefs (Biology of Habitats), OUP Oxford.

[10] Francini-Filho R, Coni E, Meirelles P, Amado-Filho G, Thompson F, Pereira-Filho G. et al. (2013) Dynamics of Coral Reef Benthic Assemblages of the Abrolhos Bank, Eastern Brazil: Inferences on Natural and Anthropogenic Drivers. PLoS ONE 8: e54260.

[11] Knowlton, N. and Jackson, J. (2008) Shifting Baselines, Local Impacts, and Global Change on Coral Reefs. PLoS Biology 6: e54.

[12] Dinsdale E, Edwards R, Hall D, Angly F, Breitbart M., Brulc, J. et al. (2008) Functional metagenomic profiling of nine biomes. Nature 455: 830-830.

[13] Sandin S, Smith J, DeMartini E, Dinsdale E, Donner S, Friedlander A et al. (2008) Baselines and Degradation of Coral Reefs in the Northern Line Islands. PLoS ONE 3: e1548.

[14] Bruce T, Meirelles P, Garcia G, Paranhos R, Rezende C, de Moura R, et al. (2012) Abrolhos Bank Reef Health Evaluated by Means of Water Quality, Microbial Diversity, Benthic Cover, and Fish Biomass Data. PLoS ONE 7: e36687.

[15] Barott K, Williams G, Vermeij M, Harris J, Smith J, Rohwer F, Sandin S (2012) Natural history of coral-algae competition across a gradient of human activity in the Line Islands. Marine Ecology Progress Series 460: 1-12.

[16] Barott K, Rohwer F (2012) Unseen players shape benthic competition on coral reefs. Trends in Microbiology 20: 621-628.

[17] Haas A, Gregg A, Smith J, Abieri M, Hatay M, Rohwer F (2013) Visualization of oxygen distribution patterns caused by coral and algae. PeerJ 1: e106.

[18] Haas A, Nelson C, Rohwer F, Wegley-Kelly L, Quistad S, Carlson C, et al. (2013)

Influence of coral and algal exudates on microbially mediated reef metabolism. PeerJ 1: e108.

[19] Haas AF, Rohwer F, Carlson C, Nelson C,Wegley L, Leichter J, Smith JE. 2011. Effects of coral reef benthic primary producers on dissolved organic carbon and microbial activity. PLoS ONE 6:e27973.

[20] Nelson CE, Goldberg SJ, Kelly LW, Haas AF, Smith JE, Rohwer F, Carlson CA. 2013. Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. The ISME Journal 7:962–979.

[21] Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F,Wegley L, Hatay M, Hall D, Brown E, Haynes M, Krause L, Sala E, Sandin SA, Thurber RV,Willis BL, Azam F, Knowlton N, Rohwer F. 2008. Microbial ecology of four coral atolls in the Northern Line Islands. PLoS ONE 3(2):e1584.

[22] Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The role of microorganisms in coral health, disease and evolution. Nature Reviews Microbiology 5: 355-362.

[23] Casadevall A, Pirofski LA (2015) What is a host? Incorporating the microbiota into the damage-response framework. Infect Immun 83(1):2-7.

[24] Moran NA, Sloan DB (2015) The Hologenome Concept: Helpful or Hollow? PLoS Biol 13(12): e1002311.

[25] Stal LJ (1995) Physiological ecology of cyanobacteria in microbial mats and other communities. New Phytol 131:1-32.

[26] Connell S, Foster M, Airoldi L (2014) What are algal turfs? Towards a better description of turfs. Marine Ecology Progress Series 495: 299-307.

[27] Echenique-Subiabre I, Villeneuve A, Golubic S, Turquet J, Humbert J, Gugger M (2014) Influence of local and global environmental parameters on the composition of cyanobacterial mats in a tropical lagoon. Microb Ecol 69: 234-244.

[28] Ainsworth TD, Krause L, Bridge T, Torda G, Raina JB, Zakrzewski M, et al. (2015) The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. The ISME Journal 9 2261–2274, doi:10.1038/ismej.2015.39

[29] Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. (2012) Cross-biome metagenomic analyses of soil microbial communities and their

functional attributes. Proc Natl Acad Sci USA 109 21390–21395.

[30] Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. (2012) Human gut microbiome viewed across age and geography. Nature 486(7402):222–227.

[31] Mason O U, Hazen TC, Borglin S, Chain PS, Dubinsky EA, Fortney JL, et al. (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. ISME Journal 6:1715–1727.

[32] Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

[33] Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863-864.

[34] Rodrigue S, Materna A, Timberlake S, Blackburn M, Malmstrom R, Alm E, et al. (2010) Unlocking Short Read Sequencing for Metagenomics. PLoS ONE 5: e11840.

[35] Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: 386.

[36] Overbeek R (2005) The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Research 33: 5691-5702.

[37] R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Available at: https://www.r-project.org/.

[38] Wickham H (2007) Reshaping Data with the reshape Package. Journal of Statistical Software 21(12):1-20.

[39] Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer Publishing Company.

[40] Oksanen J, Kindt R., O'Hara B (2005) Vegan: R functions for vegetation ecologists. Available at: http//cc.oulu.fi/jarioksa/softhelp/vegan.html.

[41] Garcia G, Gregoracci G, de O. Santos E, Meirelles P, Silva G, Edwards R, et al. (2013) Metagenomic Analysis of Healthy and White Plague-Affected Mussismilia

braziliensis Corals. Microb Ecol 65: 1076-1086.

[42] Cavalcanti G, Gregoracci G, Longo L, Bastos A, Ferreira C, Francini-Filho R, et al. (2013) Sinkhole-like structures as bioproductivity hotspots in the Abrolhos Bank. Continental Shelf Research 70: 126-134.

[43] Cavalcanti G, Gregoracci G, dos Santos E, Silveira C, Meirelles P, Longo L, et al. (2014) Physiologic and metagenomic attributes of the rhodoliths forming the largest CaCO3 bed in the South Atlantic Ocean. The ISME Journal 8: 52-62.

[44] Breiman L (2001) Random Forests. Machine Learning 1–33.

[45] Liaw A, Wiener M (2002) Classification and regression by randomForest. R news 2:18–22.

[46] Dinsdale E (2013) Multivariate analysis of functional metagenomes. Front Gene 2 4:41.

[47] Neveux J, Lantoine F (1993) Spectrofluorometric assay of chlorophylls and phaeopigments using the least squares approximation technique. Deep Sea Research Part I: Oceanographic Research Papers 40: 1747-1765.

[48] Tenário M, Le Borgne R, Rodier M, Neveux J (2005) The impact of terrigeneous inputs on the Bay of Ouinné (New Caledonia) phytoplankton communities: A spectrofluorometric and microscopic approach. Estuarine, Coastal and Shelf Science 64: 531-545.

[49] Bennett A (1973) Complementary chromatic adaptation in a filamentous blue-green alga. The Journal of Cell Biology 58: 419-435.

[50] Bryant D, Guglielmi G, de Marsac N, Castets A, Cohen-Bazire G (1979) The structure of cyanobacterial phycobilisomes: a model. Archives of Microbiology 123: 113-127.

[51] Teske A, Ramsing NB, Habicht K, Fukui M, Küver J, Jørgensen BB, et al. (1998) Sulfate-Reducing Bacteria and their activities in cyanobacterial mats of Solar Lake (Sinai, Egypt). Applied and Environmental Microbiology, 64(8), 2943–2951.

[52] Steunou AS, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E, et al. (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proc. Natl Acad. Sci. USA, 103: 2398–2403.

[53] Lesser M (2004) Discovery of Symbiotic Nitrogen-Fixing Cyanobacteria in Corals. Science 305: 997-1000.

[54] Behrendt L, Larkum A, Norman A, Qvortrup K, Chen M, Ralph P, et al. (2010) Endolithic chlorophyll d-containing phototrophs. The ISME Journal 5: 1072-1076.

[55] Chen M, Schliep M, Willows R, Cai Z, Neilan B, Scheer H (2010) A Red-Shifted Chlorophyll. Science 329: 1318-1319.

[56] Egan S, Fernandes N, Kumar V, Gardiner M, Thomas T (2014) Bacterial pathogens, virulence mechanism and host defence in marine macroalgae. Environ Microbiol 16: 925-938.

[57] Fernandes N, Case R, Longford S, Seyedsayamdost M, Steinberg P, Kjelleberg S, et al. (2011) Genomes and Virulence Factors of Novel Bacterial Pathogens Causing Bleaching Disease in the Marine Red Alga Delisea pulchra. PLoS ONE 6: e27387.

[58] Richardson L, Miller A, Broderick E, Kaczmarsky L, Gantar M, Stanic D, et al. (2009) Sulfide, microcystin, and the etiology of black band disease. Diseases of Aquatic Organisms 87: 79-90.

[59] Raina J, Dinsdale E, Willis B, Bourne D (2010) Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? Trends in Microbiology 18: 101-108.

[60] Raina J, Tapiolas D, Forêt S, Lutz A, Abrego D, Ceh J, et al. (2013) DMSP biosynthesis by an animal and its role in coral thermal stress response. Nature 502: 677-680.

[61] Weber M, de Beer D, Lott C, Polerecky L, Kohls K, Abed R, et al. (2012) Mechanisms of damage to corals exposed to sedimentation. Proceedings of the National Academy of Sciences 109: E1558-E1567.

[62] Welker M, Von Döhren H (2006) Cyanobacterial peptides - Nature's own combinatorial biosynthesis. FEMS Microbiology Reviews 30: 530-563.

[63] Amado-Filho G, Moura R, Bastos A, Salgado L, Sumida P, Guth A, et al. (2012) Rhodolith Beds Are Major CaCO3 Bio-Factories in the Tropical South West Atlantic. PLoS ONE 7: e35171.

[64] Cordero O, Ventouras L, DeLong E, Polz M (2012) Public good dynamics drive

evolution of iron acquisition strategies in natural bacterioplankton populations. Proceedings of the National Academy of Sciences 109: 20059-20064.

[65] Cordero O, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, et al. (2012) Ecological Populations of Bacteria Act as Socially Cohesive Units of Antibiotic Production and Resistance. Science 337: 1228-1231.

[66] Costa T, Felisberto-Rodrigues C, Meir A, Prevost M, Redzej A, Trokter M, Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nature Reviews Microbiology 13: 343-359.

[67] Faraldo-Gómez J, Sansom M (2003) Acquisition of siderophores in Gram-negative bacteria. Nature Reviews Molecular Cell Biology 4: 105-116.

[68] Yoon S, Cruz-García C, Sanford R, Ritalahti K, Löffler F (2014) Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO3−/NO2− reduction pathways in Shewanella loihica strain PV-4. The ISME Journal 9: 1093-1104.

## Supplementary Information

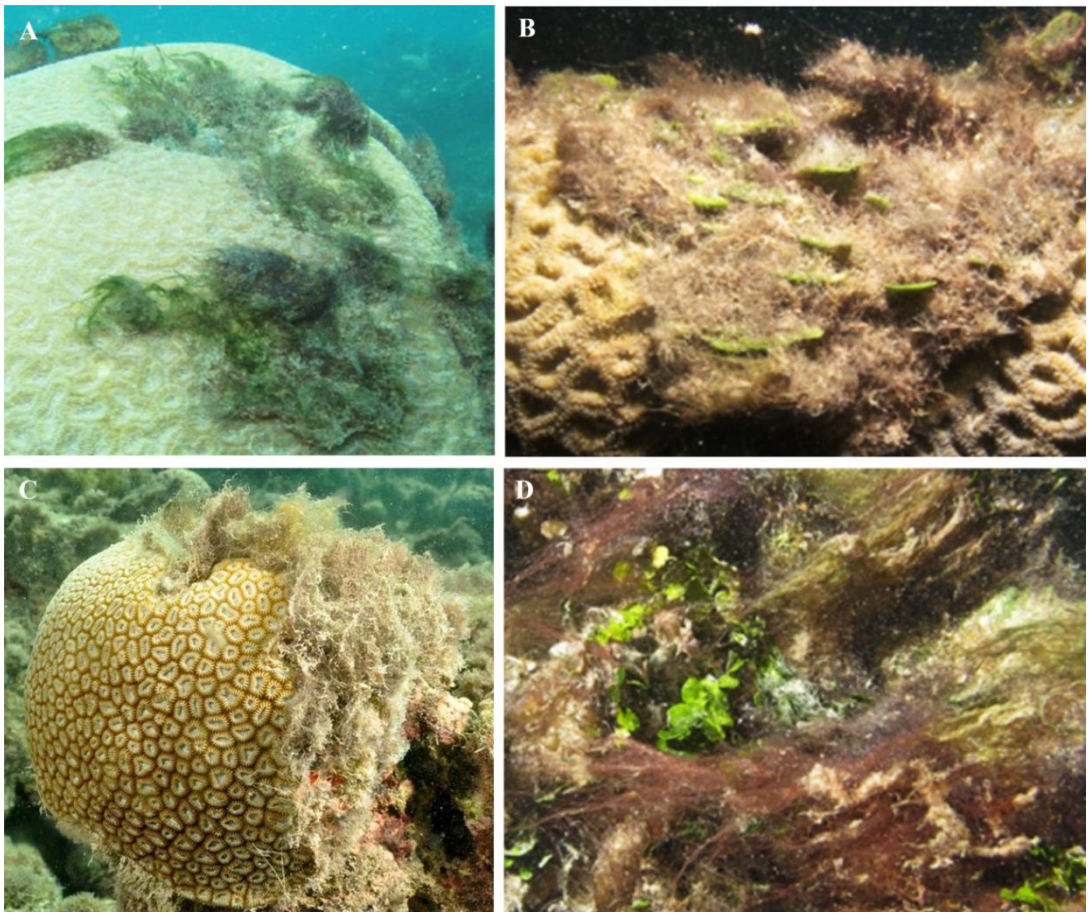



**Figure S1. Underwater pictures of turfs.**

A, B, C, turf growing over *Mussismilia* corals. D, turf growing over *Orbicella* coral.

doi:10.1371/journal.pone.0161168.s001

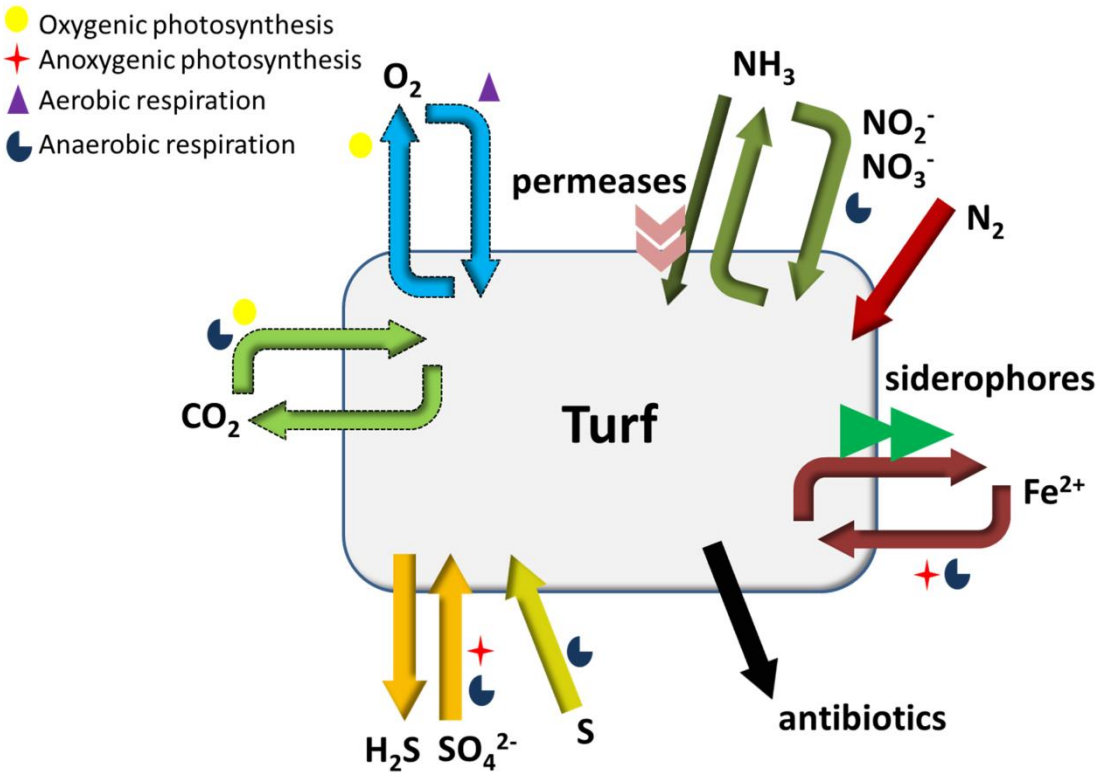



**Figure S2. Scheme of turf system and major metabolisms.**

Conceptual model presenting the major metabolisms acting in turf from Abrolhos reefs, which are not a common function in other holobionts (e.g., sulphate reduction and anoxygenic photosynthesis absent in healthy corals).
doi:10.1371/journal.pone.0161168.s002

**Table S1. Turf pigment profile.**

Pigment concentration relative to area in the samples of turf assemblages collected in October 2013, ordered by hierarchical clustering. Chl: Chlorophyll; Phae: Phaeophytin; AFDW: ash-free dry weight.

doi:10.1371/journal.pone.0161168.s003

| Samples | Chl *a* (µg/cm2) | Phae (µg/cm2) | AFDW (mg/cm2) | Chl *a*/AFDW (µg /g) | Phae/Chl *a* (%) |
|---|---|---|---|---|---|
| **PAB.3\|O** | 7.2 | 1.2 | 109 | 66 | 16.1 |
| **AR.2\|O** | 47.1 | 9.4 | 321 | 147 | 19.9 |
| **PAB.1\|O** | 57.0 | 4.3 | 532 | 107 | 7.5 |
| **AR.3\|O** | 15.0 | 2.0 | 449 | 33 | 13.0 |
| **PAB.2\|O** | 15.8 | 2.3 | 254 | 62 | 14.6 |
| **PL.1\|O** | 93.8 | 3.6 | 198 | 473 | 3.8 |
| **AR.1\|O** | 24.3 | 3.3 | 134 | 181 | 13.6 |
| **max** | 93.8 | 9.4 | 532 | 473 | 19.9 |
| **min** | 7.2 | 1.2 | 109 | 33 | 3.8 |
| **average** | 37.2 | 3.7 | 285 | 153 | 12.7 |
| **median** | 24.3 | 3.3 | 254 | 107 | 13.6 |

**Table S2. General features of the turf assemblage metagenomes.**

Approximately 7 Gbp and 11.63 million reads (raw sequences) were obtained from turfs at different locations of the Abrolhos Bank. Approximately 6.5 million reads (11 samples) and 5.13 million reads (eight samples) were obtained from the turfs collected in March (summer) and October (winter) 2013, respectively. After quality control, the metagenomes contained 58,990 to 1,172,168 reads, with an average read length of 162 to 344 nt. The metagenomic sequences had high taxonomic coverage according to the rarefaction curve.

doi:10.1371/journal.pone.0161168.s004

| | Metagenome (MG) sequences | | | | Taxonomy Genbank - Domain (normalized by each metagenome) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | MG raw size bp | Total number of sequences (raw) | MG size (Post QC) bp | Total number of sequences (Post QC) | Bacteria (%) | Eukaryota (%) | Archaea (%) | Virus (%) | Others/ Unclassified/ Unassigned (%) |
| **MARCH** | | | | | | | | | |
| **PL.1** | 34,382,205 | 178,457 | 31,448,043 | 172,944 | 76.76 | 10.64 | 0.34 | 0.12 | 12.14 |
| **PL.2** | 6,227,368 | 32,310 | 5,111,519 | 29,495 | 84.09 | 3.19 | 0.53 | 0.08 | 12.11 |
| **PL.3** | 106,200,288 | 614,521 | 95,157,293 | 554,519 | 49.48 | 42.11 | 0.18 | 0.08 | 8.15 |
| **PL.4** | 40,050,776 | 205,111 | 35,519,165 | 198,979 | 84.14 | 2.67 | 0.49 | 0.05 | 12.65 |
| **AR.1** | 54,612,957 | 280,551 | 46,552,396 | 262,886 | 80.20 | 4.26 | 0.28 | 0.03 | 15.22 |
| **AR.2** | 56,685,765 | 290,201 | 51,419,363 | 280,308 | 82.00 | 4.12 | 0.43 | 0.06 | 13.39 |
| **AR.3** | 195,751,729 | 677,377 | 168,993,630 | 659,796 | 80.94 | 1.56 | 0.34 | 0.04 | 17.12 |
| **AR.4** | 110,216,192 | 593,480 | 99,195,357 | 575,503 | 81.68 | 3.34 | 0.36 | 0.04 | 14.59 |
| **PAB.2** | 16,810,356 | 83,490 | 14,650,440 | 80,324 | 82.77 | 4.41 | 0.47 | 0.04 | 12.31 |
| **PAB.3** | 35,270,546 | 111,285 | 26,888,594 | 102,782 | 82.40 | 2.61 | 0.36 | 0.03 | 14.59 |
| **PAB.4** | 36,562,110 | 185,369 | 32,257,528 | 176,395 | 82.38 | 3.33 | 1.18 | 0.16 | 12.94 |
| **OCTOBER** | | | | | | | | | |
| **PL.2** | 20,899,397 | 113,592 | 18,350,484 | 107,665 | 97.76 | 1.45 | 0.26 | 0.05 | 0.48 |
| **PL.4** | 35,773,050 | 195,651 | 33,060,641 | 186,167 | 97.21 | 1.96 | 0.31 | 0.05 | 0.47 |
| **AR.1** | 9,751,818 | 47,689 | 8,118,890 | 45,168 | 91.84 | 7.13 | 0.46 | 0.20 | 0.37 |
| **AR.2** | 169,513,303 | 642,298 | 152,157,722 | 624,907 | 95.90 | 3.25 | 0.39 | 0.08 | 0.38 |
| **AR.3** | 89,320,369 | 258,981 | 74,875,522 | 249,663 | 92.89 | 6.30 | 0.35 | 0.05 | 0.41 |
| **PAB.1** | 151,894,049 | 547,928 | 134,542,669 | 535,455 | 97.01 | 2.20 | 0.38 | 0.04 | 0.37 |
| **PAB.2** | 32,456,431 | 156,556 | 29,152,445 | 151,800 | 96.14 | 3.09 | 0.32 | 0.05 | 0.40 |
| **PAB.3** | 165,067,882 | 602,539 | 146,060,007 | 586,084 | 96.87 | 2.27 | 0.41 | 0.05 | 0.40 |

**Table S3. The taxonomic composition of turf metagenomes are statistically indistinguishable in different locations and seasons (H1).**

Adonis (PERMANOVA) results of taxonomic composition of turf metagenomes (bacterial order level) based on Bray-Curtis distances with 999 permutations. MS, mean sum of squares; SS, sum of squares. D.f., degrees of freedom; SS, sum of squares; MS, mean sum of squares.
doi:10.1371/journal.pone.0161168.s005 [*See Online Material*]

**Table S4. Taxonomic contribution to the turf composition from the Abrolhos Bank, indicated by collection period and reef location.**

The percent correspond to the abundance normalized by the total of known sequences A) Contributions of the major phyla in the composition of turf metagenomes. B) Proteobacteria and cyanobacteria: Contributions of the most abundant closest species identified of Proteobacteria and cyanobacteria in each reef and collection period. C) Contributions of archaea and viruses (closest species) in each reef and collection period from the Abrolhos Bank.
doi:10.1371/journal.pone.0161168.s006 [*See Online Material*]

**Table S5. The functional composition of turf metagenomes are statistically indistinguishable in different locations and seasons (H2).**

Adonis (PERMANOVA) results of functional composition of turf metagenomes (SEED Level 1 Subsystems) based on Bray-Curtis distances with 999 permutations. MS, mean sum of squares; SS, sum of squares. D.f., degrees of freedom; SS, sum of squares; MS, mean sum of squares.
doi:10.1371/journal.pone.0161168.s007 [*See Online Material*]

**Table S6. Variable importance (taxonomic Order level) determined by the unsupervised Random Forest analysis ranking of Taxonomy.**
doi:10.1371/journal.pone.0161168.s008 [*Online Material*]

**Table S7. Variable importance (level 1 Subsystems) determined by the unsupervised Random Forest analysis ranking of Function.**
doi:10.1371/journal.pone.0161168.s009 [*See Online Material*]

**Table S8. Comparison of the different types of metabolism in the turf, coral, rhodolith, and seawater samples from the Abrolhos Bank.** The percentage, abundance, minimum, maximum, average and standard deviation are provided for each sample, as well as for photosynthesis, nitrogen, and sulfur metabolism.
doi:10.1371/journal.pone.0161168.s010 [*See Online Material*]

**Table S9. Turkey Honest Significant Differences (HSD) pos hoc test results of H3 ANOVA.** Diff., difference in the observed means.
doi:10.1371/journal.pone.0161168.s011 [*See Online Material*]

# Chapter III

## A NOVEL, WIDESPREAD TURF-FORMING CYANOBACTERIUM (*Acrophormium turfae*) FROM THE ABROLHOS REEF SYSTEM HAS A VAST SECONDARY METABOLITE GENOMIC REPERTOIRE

*Manuscript under revision, The ISME Journal*

Juline M. Walter[1,2], Felipe H. Coutinho[1,2], Louisi de Oliveira[1], Luciana Leomil[1], Paulo I. Hargreaves[1], Beatriz S. Silva[3], Giovana O. Fistarol[1], Kazuo Miyashita[4], Fumito Maruyama[5], Paulo S. Salomon[3,6], Cristiane C. Thompson[1], Marcel C. van Verk[7], Bas E. Dutilh[2,8*] and Fabiano L. Thompson[1,6*]

[1]Laboratory of Microbiology, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[2]Radboud University Medical Centre, Radboud Institute for Molecular Life Sciences, Centre for Molecular and Biomolecular Informatics (CMBI), Nijmegen, The Netherlands.

[3]Marine Phytoplankton Laboratory, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[4]Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Japan.

[5]Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan.

[6]Center of Technology - CT2, SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[7]Plant-Microbe Interactions, Bioinformatics, Faculty of Science, Utrecht University, Utrecht, The Netherlands.

[8]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands.

*Corresponding Authors:
Av. Carlos Chagas Filho. 373, CCS-IB-Biomar, Lab. de Microbiologia, Bloco A3 (Anexo), sl. 102, Cidade Universitária, Rio de Janeiro, RJ, Brasil, CEP 21941-599, T/F: +552139386567, E-mail: fabianothompson1@gmail.com (FLT); and Padualaan 8, Room Z509, Hugo R. Kruyt-building, Bioinformatics, Faculty of Science, Utrecht University, Utrecht, The Netherlands, 3584 CH, T +31302534212, E-mail: bedutilh@gmail.com (BED).

**Running Title:** Filamentous Cyanobacteria from Abrolhos Bank

**Keywords:** Microbial Ecology / Cyanobacteria / Turf / Coral Reefs / Abrolhos Bank / Secondary Metabolites / Photoacclimation Strategy

## Abstract

Turfs are among the major benthic components of reef systems worldwide. The nearly complete genome sequences, basic physiological characteristics, and phylogenomic reconstruction of the two novel, phycobiliprotein-rich filamentous Cyanobacteria strains isolated from turf assemblages from the Abrolhos Bank (Brazil) are reported. Containing approx. 8 Mb in size, both *Acrophormium turfae* gen.nov. CCMR0081$^{T}$ and CCMR0082 genomes have a broader photosynthetic pigment set, exhibits an uncommon acclimative response, and have a significantly higher number and diversity of secondary metabolite synthesis genes, which NRPS, PKS and NRPS/PKS hybrid are the most abundant potential producers that all open opportunities for reef systems occupancy. The genomic and physiologic repertoires of the novel cyanobacterium hint at important features that allow them a widespread distribution in reef systems.

## Introduction

Cyanobacteria are among the most diverse groups of prokaryotic organisms, comprising unicellular and multicellular, photosynthetic and non-photosynthetic, and free-living and symbiotic forms. They are the earliest oxygenic photosynthetic organisms and correspond to a relevant global biomass on the order of $10^{14}$ g C (Garcia-Pichel *et al.*, 2003), accounting for 20–30% of Earth's photosynthetic productivity (Luque *et al.*, 2001; Muramatsu and Hihara 2012; Watanabe *et al.*, 2014; Waterbury *et al.*, 1979). Coupled with their photosynthetic abilities and roles to global carbon, Cyanobacteria are ecologically important for their contributions to nitrogen fixation (Capone *et al.*, 1997) as well as carbonate precipitation (Dupraz *et al.*, 2009), regulating the essential cycles for all higher life.

The immense diversity within this group of microorganisms is also reflected in the extent of their oligopeptides (i.e. secondary metabolites) (Micallef *et al.*, 2014), and in a number of evolved acclimation strategies to sense and respond to changing nutrient and light conditions (Grossman, 2003), which have aided species survival in diverse and highly competitive ecological niches. The oligopeptides are modular enzymes known as polyketide synthases (PKSs, a family of multi-domain enzymes), nonribosomal peptide synthases (NRPSs, large multisubunit enzymes) or hybrids thereof (Welker and Von Döhren 2006), which many are potent toxins that cause deleterious effects to humans and wildlife in aquatic systems (Cheung *et al.*, 2013). When conditions are limiting cell growth (i.e. diffuse/low-light or excess irradiation), Cyanobacteria are able to largely reshape their photoapparatus, as altering the whole chlorophyll and phycobiliproteins content, and adjust their photosystem ratio.

Turfs are dense filamentous mats that represent one of the most abundant benthic functional groups in Atlantic, Caribbean, and Pacific reefs (Sweet *et al.*, 2013; Walter *et al.*, 2016; Wild *et al.*, 2014). Cyanobacteria are among the most abundant microorganisms in turfs from the Abrolhos reefs (Bahia, Brazil) (Walter *et al.*, 2016) and play important physical, structural and functional roles. However, the genomic and

physiological properties of Abrolhos turf-forming Cyanobacteria are not known. Coral reefs worldwide have experienced a phase-shifting phenomenon (Knowlton and Jackson 2008) whereby the coral cover is decreasing and the turf/algae cover are increasing. The major causes of this phase-shift are overfishing and coral disease. Despite ample understanding of the ecologic relevance of turfs in reef systems worldwide, the pathogenic (toxic) potential of turf-forming Cyanobacteria has not been studied.

The aim of the present study was to determine the genomic and physiological potential of two novel representative filamentous Cyanobacteria (CCMR0081$^T$ and CCMR0082) from the Abrolhos turfs. First, clonal cultures of two novel representative filamentous Cyanobacteria (CCMR0081$^T$ and CCMR0082) were established, and growth rates and photosynthetic pigments were determined. A single contig representing the nearly complete genome of each strain was obtained, and the phylogenetic and taxonomic positions of the two novel strains were determined. The two novel genomes were analysed thoroughly, including detailed pigment, acclimative response, and secondary metabolite gene content determination. Finally, comparison of the two genomes with 19 previously published turf metagenomes (Walter *et al.,* 2016) allowed us to determine the frequency of occurrence of these two novel strains in turfs from different locations in the Abrolhos Bank.

## Material and Methods

### Source of organism, isolation and cultivation

The cyanobacteria strains reported herein were isolated from turf samples growing over the corals from the Abrolhos reefs (17°57'32.7"S, 38°30'20.3"W), Brazil, collected in March and October 2013 by scuba diving (approx. 10 m), under a federal government license (SISBIO no. 27147-2). Samples were kept at ambient temperature and protected from direct sunlight with the flask lids untightened; these flasks were transported to the laboratory for isolation. Upon arrival, turf samples were enriched with

f/2 medium (Guillard 1975; Guillard and Ryther 1962) prepared with 0.22 um filtered Abrolhos seawater, and kept at 22 °C under a 16/8-h light/dark cycle and light intensity of 80 µmol photons $s^{-1}$ $m^{-2}$. Light was supplied with cool white fluorescent lamps. As a first cleaning step, cyanobacterial filaments were manually picked from the turf samples and rinsed with five serial transfers through drops of sterile f/2 medium and incubated under the conditions described above for 10 days. The filaments were fragmented by sonication (Ultrasonic Processor/Cell Disruptor, 130-Watt Cole-Parmer Instrument Co., USA), for 1 min. alternated 5s on/off pulses at 20% power. Single fragments of filaments were sorted by Fluorescence Activated Cell Sorting (FACS) in a MoFlo (Dako-Cytomation) flow cytometer into the wells of 96-well microplates containing f/2 medium. Cultures that grew in the wells within 4 to 6 weeks were transferred to fresh culture medium. The two strains reported herein were added to the Culture Collection of Microorganisms at the Federal University of Rio de Janeiro (CCMR) under the codes CCMR0081$^T$ and CCMR0082. The strains were cryopreserved in liquid $N_2$ using 5% DMSO for long-term storage. Cultures were maintained as an unicyanobacterial but nonaxenic culture. Axenic cultures of filamentous cyanobacteria are difficult to obtain due to the heterotrophic bacteria strongly associated at the sheath (Paul *et al.*, 2014).

## Growth rate estimates

Both cyanobacterial strains were grown in 100 mL tissue-culture bottles in f/2 medium made with Abrolhos seawater at an irradiance of 50 µE m-2 s-2 with a 16 h:8 h light:dark cycle at 24 and 26°C. Filaments from seven-day old cultures of each strain were homogenized by gentle mechanical disruption in a sonicator (Ultrasonic Processor/Cell Disruptor, 130-Watt Cole-Parmer Instrument Co., USA, 1 min. alternated 1s on/off pulses at 10% power). The homogeneous filament suspensions were diluted with fresh f/2 medium and 150 µL each were transferred (4 replicates for each strain and temperature) to the wells of 96-well polystyrene microplates. The plates were incubated using the same irradiance and temperature conditions as above.

Growth in each well was followed by daily *in vivo* chlorophyll fluorescence measurements for five days in a spectrofluorometer (Varian Cary Eclipse, $\lambda_{ex}$ 450 nm, $\lambda_{em}$ 678 nm). Specific growth rates (μ, $d^{-1}$) were calculated as the slope between the Ln of chlorophyll fluorescence vs. time in the linear portion of the growth curves. Generation times (G, d) were calculated as Ln2/μ.

## Cell-size measurements

Filaments of both strains CCMR0081$^T$ and CCMR0082 harvested from 10-day-old cultures were observed in an upright microscope (Olympus BX51) at 1,000x magnification. The linear dimensions (length and width) of 50 cells from each strain were measured using a calibrated ruler printed on the microscope's ocular.

## *In vivo* absorption spectra

Cultures of both strains were grown at 24°C as explained above for the growth rate estimates. *In vivo* absorption spectra were obtained by scanning homogeneous seven-day-old cyanobacterial filament suspensions from 400 to 700 nm in a dual beam spectrophotometer (Femto Cirrus 80ST) with 1 cm optical path quartz cuvettes using f/2 medium as a blank. Spectra were normalized at the chlorophyll *a* blue peak (442 nm).

## Chromatic adaptation evaluation

The cyanobacterial strains CCMR0081$^T$ and CCMR0082 were continuously grown in the laboratory under white fluorescent light lamp (25°C, 16-8h photoperiod), a condition commonly used for maintaining Cyanobacteria, in f/2 media and then growth under Far-Red light (FRL, λ > 700 nm) to evaluate chromatic adaptation (730 nm LED Light 30 μmol photons/$m^2$/s, 12-12h photoperiod) for 21 days. Absorption spectra of CCMR0081$^T$ and CCMR0082 cells suspended in a growth medium were measured with a UV-VIS spectrometer (UV-2450, SHIMADZU, Kyoto, Japan).

## Photosynthetic pigments quantification

Cultures of both strains were grown at 24°C as explained above for the growth rate estimates. Biomass from triplicate, 10-day-old cultures of each strain were split in two approximately equal parts and the wet weight of each aliquot was determined. Filaments of both aliquots from each culture were harvested by filtration onto 25 mm GF/F glass-fibre discs. Biomass retained in one set of GF/F discs was extracted overnight in 90% acetone for chlorophyll quantification by spectrofluorometry as described elsewhere (Tenório *et al.*, 2005). The other filter set was extracted in 20 nM sodium acetate buffer, pH 5.5, supplemented with 3 mM sodium azide and 10 mM disodium EDTA. Cells were ruptured by sonication (Ultrasonic Processor/Cell Disruptor, 130-Watt Cole-Parmer Instrument Co., USA, 5 min, 90% power) in an ice bath intercalated with freeze and thaw cycles. The crude extracts were treated with 1% (w/v) streptomycin sulphate for 30 min at 4°C and centrifuged at 10,000 g for 10 min to precipitate the cellular debris. Phycobiliprotein (phycoerythrin, phycocyanin and allophycocyanin) concentrations were calculated from measurements of optical densities at 565, 620, and 650 nm in a spectrophotometer (Femto Cirrus 80ST) using the equations described in Bennett and Bogorad (1973) and Bryant *et al.*, (1979). Results were expressed as mass of pigment per wet weight of cyanobacterial biomass.

## DNA extraction and library construction

Strains CCMR0081$^T$ and CCMR0082 were grown in 250 ml Erlenmeyer flasks in f/2 medium to obtain cyanobacterial biomass for DNA extraction (Figure S1). Genomic DNA was extracted using the Dneasy Plant Mini kit (Qiagen Inc., Valencia, CA, USA). The integrity of the DNA samples was evaluated using electrophoresis on a 1% agarose gel with GelRedTM (Biotium Inc., Hayward, CA, USA). The quality and purity of the DNA were assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The quantity and quality of the DNA in each

sample used to construct the paired-end library were 317.3 ng/µl (A260/280 2.0 and A260/230 2.3) for CCMR0081$^{T}$ and 162.4 ng/µl (A260/280 2.0 and A260/230 1.6) for CCMR0082.

Accurate DNA quantification was obtained using the Qubit® 3.0 Fluorometer (Life Technologies-Invitrogen, Carlsbad, CA, USA). The DNA libraries were built using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The size distribution of the libraries was evaluated using a 2100 Bioanalyser (Agilent, Santa Clara, CA, USA), and quantification was obtained using a 7500 Real-Time PCR Thermocycler (Applied Biosystems, Foster City, CA, USA) and the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, MA, USA).

## Acquisition of whole-genome sequences

Genomic DNAs were sequenced in an Illumina MiSeq paired-end flow cell (San Diego, CA, USA) (2 × 250-bp read length) at the Laboratory of Microbiology, Federal University of Rio de Janeiro, Brazil. The sequencing coverage values were calculated based on the number of total reads generated from each sample in the sequencing runs by considering that a percentage of the sequences belong to heterotrophic bacteria associated with the cyanobacterial sheath, even the isolation and cultivation were considered insightful steps. The fastq files generated by Illumina sequencing were checked for quality in both the forward and reverse directions with FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Andrews 2010). Based on the sequence profiles, the files were filtered for trimming ends, sequence quality and minimal length using the FASTX-Toolkit version 0.0.12 (http://hannonlab.cshl.edu/fastx_toolkit/) and PRINSEQ version 0.20.4 - *Preprocessing and Information of Sequences* (http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi) (Schmieder and Edwards 2011).

Sequence data were assembled using two assemblers. The genome assembly was performed using both the A5-Miseq pipeline (Coil, Jospin, and Darling 2014; Tritt

*et al.*, 2012) and Velvet version 1.2.10 (Zerbino and Birney 2008) assemblers. The latter assembler used raw sequence data that was previously stringently trimmed using odd k-mer lengths (k-mer 89 for CCMR0081$^{T}$, and k-mer 81 for CCMR0082). The A5 is a pipeline which automatically proceeds to data cleaning, error correction, contig assembly, scaffolding and quality control (Coil *et al.*, 2014; Tritt *et al.*, 2012). The results were merged, and binning was applied based on the frequency of k-mers. The generated clusters were proceeding visualization with Emergent Self-Organizing Maps (ESOMs) (Dick *et al.*, 2009). The *Database and Visualization Tool for the Next-gen Data* – GAP5 (Bonfield and Whitwham 2010) was used to join contigs to obtain improved scaffolds using the *tg_index.sh* script to create a database suitable for editing. Further, the *St. Petersburg genome assembler* - SPAdes version 3.5.0 (Bankevich *et al.*, 2012) tool was used to re-assemble the contigs generated again from the raw data. Finally, the results were clustered again based on frequency of k-mers, and the clusters were visualized with ESOM. The assembly statistics were estimated using the *Quality Assessment Tool for Genome Assemblies* - QUAST version 2.3 (Gurevich *et al.*, 2013). The completeness of all genomes was assessed with CheckM (Parks *et al.*, 2015). The raw sequencing data were deposited in the sequencing read archive (SRA) with their accession numbers [*accession number CCMR0081; accession number CCMR0082*]. The whole genome sequences have been deposited at the National Center for Biotechnology Information (NCBI) GenBank database under the accession numbers: [*accession number CCMR0081; accession number CCMR0082*]. The PCC7375 genome was retrieved from the database GEBA (*Genomic Encyclopedia of Bacteria and Archaea*) for comparison analysis. Genome completeness was accessed with CheckM (Parks *et al.,* 2015) (Table 1).

**Table 1. Features of *Acrophormium* strains included in this study.**

The completeness score were determinate by CheckM tool.

| Species | Strain ID | NCBI Project ID | Ecosystem Type | Genome Size (Mb) | % mol GC | N. Scaffolds | Completeness score |
|---|---|---|---|---|---|---|---|
| *Acrophormium turfae* | CCMR0081 T | *processing* | Turf, Abrolhos coral reef, Brazil (marine) | 8.41 | 47.32 | 1 | 99.18 |
| *Acrophormium turfae* | CCMR0082 | *processing* | Turf, Abrolhos coral reef, Brazil (marine) | 8.16 | 47.4 | 1 | 99.46 |
| *Acrophormium splendidus* | PCC 7375 * | NZ_ALVN 00000000.1 | Plankton, Woods Hole, Massachusetts, USA (marine, coastal) | 9.42 | 47.62 | 5 | 99.73 |

* Formerly known as Leptolyngbya sp. PCC 7375. Reference: Shih *et al*., 2013.

## Genome annotation and genomic taxonomy analysis

Genomes were annotated using Prokka version 1.11 (Seemann 2014) with default settings. The clustered regularly interspaced short palindromic repeat (CRISPR) Finder (Grissa, Vergnaud, and Pourcel 2007) was used to identify CRISPR arrays. The genomic similarity between the three strains was determined by Average Amino Acid Identity (AAI) (Konstantinidis and Tiedje 2005), Average Nucleotide Identity (ANI), Dinucleotide Signature (Karlin, Mrázek, and Campbell 1997) and *in silico* DNA-DNA Hybridization (DDH) or Genome-to-Genome Distance (GGD) (Auch, von Jan, *et al.*, 2010; Auch, Klenk, and Göker 2010; Meier-Kolthoff *et al.*, 2013). The GGD was compared using the Genome-to-Genome Distance Calculator tool, version 2.1 (Meier-Kolthoff *et al.*, 2013) (http://ggdc.dsmz.de/). The Cgview comparison tool (Grant *et al.*, 2012) was used to create a genome-wide homology map.

## Phylogenomic analysis

To establish the phylogenomic position of the new strains, a phylogenetic tree (multilocus sequence analysis, MLSA) was constructed using the concatenation of 31 conserved protein sequences previously validated as phylogenetic markers (Shih *et al.*, 2013; Wu and Eisen 2008). A Maximum Likelihood tree was constructed using RaxML v. 7 (Stamatakis 2006) and the Dayhoff+G likelihood model. One thousand bootstrap

replications were calculated to evaluate the relative support of the branches. Trees were visualized with FigTree, version 1.4.2 (Rambaut 2015). *Gloeobacter violaceus* PCC 7421 was set as the outgroup.

## Functional SEED categories annotation

To identify the gene function-relatedness for each genome, we used the first sequence of each cluster of homologous genes generated for the pangenome as a query in the Diamond tool, version 0.7.9 (Buchfink, Xie, and Huson 2014). Using the SEED database (new release70 data set), Diamond created a new database and searched for homology between the query and the SEED database. The homology search was performed for all proteins from the homologous clusters. An e-value of < $1e^{-5}$ was considered and the top hit was defined as the putative function of each cluster. Further, we associated each protein with specific roles and subsystems, i.e., in terms of biological process, molecular function, and cellular composition. Joining the related SEED category in the whole pan-genome table yielded the number of each protein in all genomes analysed. Homologues of certain key genes of interest were searched as BLAST queries of homologues from phylogenetically close protein sequences from the NCBI against the entire genome.

## Evaluations of secondary metabolites

We used the *Antibiotics & Secondary Metabolite Analysis Shell* – antiSMASH, version 3.0.3 (Weber *et al.*, 2015) tool to predict the potential secondary metabolite biosynthesis gene clusters. The number of gene clusters for ribosome-dependent peptides, nonribosomal encoded peptides and polyketide biosynthetic pathways for each analysed genome were generated.

## Abundance profiles

To estimate the relative abundance of these two genomes (CCMR0081$^T$ and CCMR0082) across Abrolhos coral reef habitats, their genomes were used as a reference database against the raw reads from 19 turf metagenomes (which were

sampled in parallel with the cyanobacterial isolation) (Walter *et al.,* 2016). These metagenomes were mapped using Bowtie2 (Langmead and Salzberg 2012) with the *--very-sensitive-local* and *-a* options. Correction for ambiguous reads (i.e., those that can be mapped to more than one genome) were performed as described by Iverson *et al.* (2012).

## Results

### Cell size and growth characteristics

Cells were of similar sizes (strain CCMR0081$^T$: 2.4 ± 0.3 µm long; 1.3 ± 0.1 µm wide, N = 50, CCMR0082: 2.7 ± 0.5 µm long, 1.3 ± 0.1 µm wide, N = 50) and arranged in unbranched filaments. In batch cultures without stirring, filaments grew attached to the walls of the flasks or as buoyant, macroscopic agglomerates (Figure S1). The specific growth rates of strains CCMR0081$^T$ and CCMR0082 were in the range of 0.08 to 0.15 $d^{-1}$ (Figure S2A), representing generation times of 8.6 and 4.5 d, respectively. The strain CCMR0081$^T$ showed higher growth rates than strain CCMR0082 at both 24°C and 26°C. Both strains grew slightly faster at 26°C compared to 24°C.

### Optical properties and pigment composition

The *in vivo* absorption spectra of both cyanobacterial strains were very similar (Figure S2B), with a strong peak corresponding to phycoerythrin at 550-552 nm and the typical peaks of chlorophyll *a* in the blue (436 nm) and red (676 nm) regions of the spectrum. Chlorophyll *a* per wet weight was 119 ± 8 and 591 ± 34 µg $g^{-1}$ for strains CCMR0081$^T$ and CCMR0082, respectively. Phycobiliprotein (PC+APC+PE) content per wet weight was 6,323 ± 461 and 9,256 ± 743 µg g wet $weight^{-1}$ for strain CCMR0081$^T$ and CCMR0082, respectively. Phycoerythrin was the dominant phycobiliprotein in both strains (41 ± 6% for strain CCMR0081$^T$ and 64 ± 1% CCMR0082). Phycobiliprotein : chlorophyll ratios ranged from 54 ± 7 for strain CCMR0081$^T$ to 15 ± 7 for strain CCMR0082.

As shown in Figure 1A-B, both cells presents similar absorption peaks when cultivated under white light, such as Chl *a* at 437 and 678 nm, carotenoids at 450–530 nm, phycoerythrin (PE) at 567 nm and allophycocyanin (APC) around 630 nm. Indeed, in white light CCMR0081$^T$ and CCMR0082 cells are red/brownish in colour due to accumulation of PE. Additionally they synthesize carotenoids, and produce Chl *a* as the sole Chl pigment. However, under FRL (Figure 1A-B), each strain responds differently. While CCMR0082 producing more PE, CCMR0081$^T$ producing more APC.

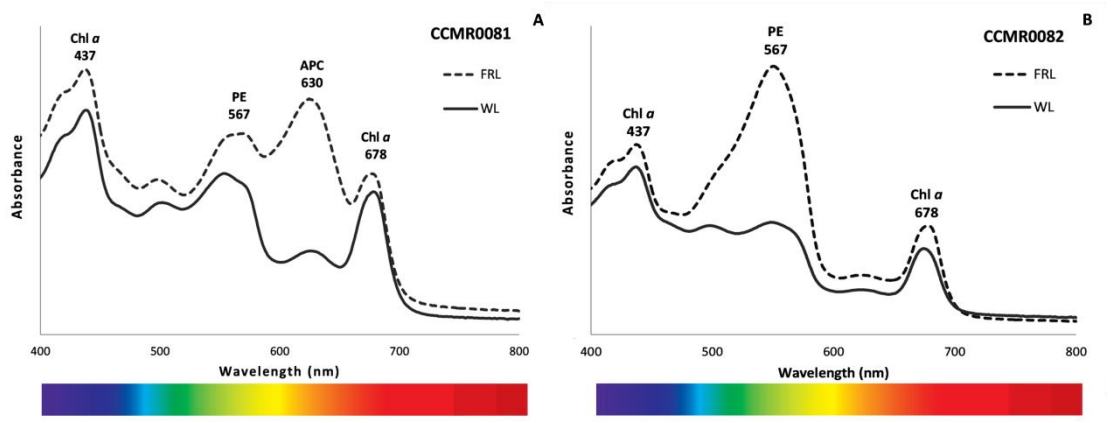



**Figure 1. Absorption spectra for *Acrophormium turfae* strains.**

Comparison of absorption spectra for *A. turfae* CCMR0081$^T$ (A) and CCMR0082 (B) cells grown in white light (WL, solid black line), and far-red light (FRL, dotted black line).

## Main features of whole genome sequences

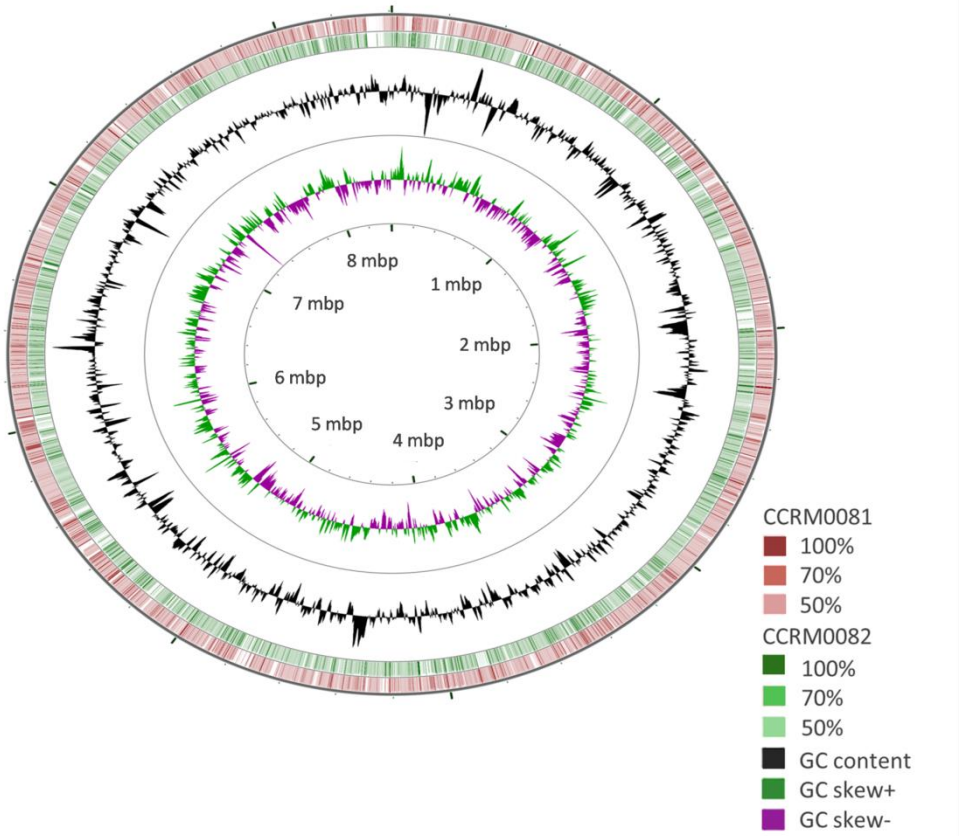
A total of 7.2 million reads for CCMR0081$^T$ (150 X coverage) and 5.7 million reads for CCMR0082 (136 X coverage) were obtained. Two genomes maps were obtained by binning of tetranucleotide frequency with ESOM (Figure S3A). The nearly complete genome sequences of the two new filamentous cyanobacteria were assembled and the basic statistics were obtained (Tables 1 and S1). We obtained a single non-circularized contig for each novel isolate (Figure 2A). CCMR0081$^T$ has a genome size of 8.41 Mb (average GC content of 47.3%), while CCMR0082 had a genome size of 8.16 Mb (average GC content of 47.4%). The genome size of their

closest relative PCC7375 is 9.42 Mb (average GC content of 47.6%). The number of predicted protein-coding genes were 7,276 and 7,713 for CCMR0081$^T$ and CCMR0082, respectively, while the abundance of total RNAs genes was 66 and 61, respectively. Both novel cyanobacterial genomes had over 99% completeness, as does PCC7375.

## Taxonomic assignment of the novel *Cyanobacteria*

The AAI and GGD between the novel isolates CCMR0081$^T$ and CCMR0082 were 97% and 95.5%, respectively (Figure 2B). The phylogenetic reconstruction based on MLSA (31 conserved protein sequences) demonstrated that the novel strains CCMR0081$^T$ and CCMR0082 share 99% MLSA identity and only 94% identity towards PCC7375 (Figure 3). Therefore, there is sufficient genomic and phylogenetic evidence to propose the creation of a new genus and species to encompass these two new strains, which we have named *Acrophormium turfae*.

A




B

| AAI / AIA / GGD / Karlin | | | |
|---|---|---|---|
| | CCRM0081 | CCRM0082 | PCC 7375 |
| CCRM0081 | 100 / 100 / 100 / 0 | 97 / 98.9 / 95.2 / 0.000656 | 92 / 95.2 / 62.9 / 0.003787 |
| CCRM0082 | | 100 / 100 / 100 / 0 | 92 / 95.1 / 62.6 / 0.003688 |
| PCC 7375 | | | 100 / 100 / 100 / 0 |

**Figure 2. Whole genome comparison.**

A) Whole-genome identity map generated through Cgview comparison tool. Protein sequences of each strain were compared against the CCRM0081 genome through blastx. The reference used was the major genome assembled in one contig. Rings represent from outermost to innermost: 1) CCRM0082 genes; 2) PCC7375 genes; 3) G+C Content; 4) G+C Skew. B) Values of AAI, ANI, GGD, and genome distance (Karlin signature) disposed in a matrix format.

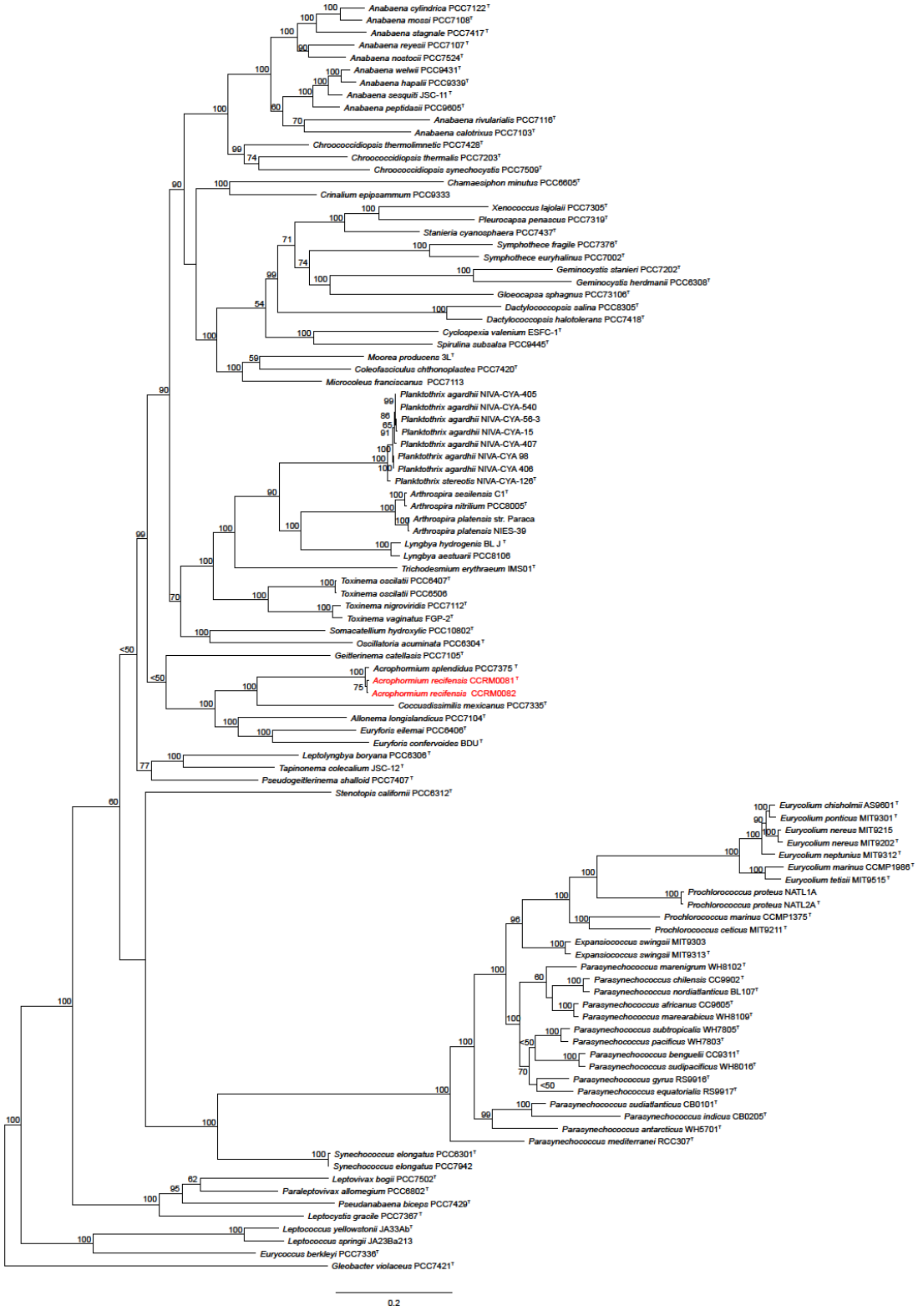



**Figure 3. Phylogenetic position of *Acrophormium turfae* species.**

Multilocus sequence analysis (MLSA) phylogenetic reconstruction was built though ML using the Dayhoff+G likelihood model by RaxML tool. Tree was inferred from a set of conserved marker genes of 100 genomes. The numbers at the nodes indicate bootstrap values as percentages greater than 50 %. Bootstrap tests were conducted with 1,000 replicates. The unit of measure for the scale bars is the number of

nucleotide substitutions per site. The *Gloeobacter violaceus* PCC 7421 sequence was designated as outgroup.

## Formal description of *Acrophormium* gen. nov.

*Acrophormium* (A.cro.phor.mium. Gr. *acro*, high, top; Gr. *phorm*, basket, mat, tuft; *-um*, sense of pertaining to; *Acrophormium*, living on the top of the mat/turf assemblage).

Description of *Acrophormium turfae* (preliminarly identified as *Leptolyngbya*-like, CCRM Collection): this species is characterized by β-carboxysome. Type strain is CCRM0081$^T$ which was isolated from the turf of Abrolhos Coral Reef, Brazil. This species forms turf assemblages growing over the corals. The genome of this strain contains 8,41 Mbp (GC = 47,32%) harboring 7,276 coding DNA sequences.

## Annotation of the *Acrophormium turfae* genomes

The functional roles for the *Acrophormium* spp. genomic sequences were categorized (Figure S4) and the photosynthetic potential of two novel *Acrophormium turfae* was analysed (Figure S5A-B). Genes related to light-harvesting related that comprise the phycobilisomes were abundant in CCMR0081$^T$ (N = 50) and CCMR0082 (N = 36), including the α, β subunits, the linker gene of phycoerythrin (N = 19 for CCMR0081$^T$ and N = 16 for CCMR0082), and the α, β subunits and small rod linker gene of phycocyanin (N = 18 for CCMR0081$^T$; N = 9 for CCMR0082). The uptake and utilization of nitrogen, phosphate and iron were detected (Figure S5C). In total, 84 genes involved in phosphate metabolism were identified in CCMR0081$^T$ and 73 genes in CCMR0082 (Figure S5C). The genetic machinery responsible for cellular phosphate incorporation includes alkaline phosphatase genes and the supplementary genes for the functioning of the *pho* regulon. The uptake and utilization of nitrogen source genes *nif*B, *nif*H, *nif*S, *nif*T, *nif*U, *nif*W, *nif*X and *nif*Z were present in both novel isolates (Figure S5C). For iron metabolism, 32 genes were identified in CCMR0081$^T$, and 29 genes in CCMR0082, compared to only 15 genes in PCC7375. Sequences related to

the chemotaxis regulator (CheY) (N = 21) and other chemotaxis-involved genes (*che*W-V, *che*A, *che*B, *che*R) were found in both novel strains. The type IV pilus/fimbrial assembly and prepilin peptidase (*pil*A, *pil*B, *pilC, pilD, pilE*, *pil*N, *pil*O, *pil*Q, *pil*T) genes were also found in both genomes. The CCMR0081$^T$ and CCMR0082 strains possess chemotaxis-related genes, including the genetic machinery encoding the CheA-CheY signal transduction pathway coupled to flagella rotation or pili extension, attachment and retention; these components allow the cell to move towards chemoattractants or away from chemorepellents (Wadhams and Armitage 2004). Homology search analyses indicated the top-30 core genome for CCMR0081$^T$, CCMR0082 and PCC7375 (Table S2).

## Secondary metabolites genes

The two novel isolates have more gene clusters than PCC7375 (N = 24 for CCMR0081$^T$, N = 20 for CCMR0082, and N = 12 for PCC7375). Genes coding for polyketides synthase (PKS, mainly Type I) and nonribosomal peptide synthase (NRPS) biosynthetic pathways were most abundant in the genomes of the novel isolates (Figure 4). Genes coding for polyketides, nonribosomal peptides, terpenes, bacteriocins, cyanobactins, lassopeptides and arylpolyenes were found in the novel strains. CCMR0081$^T$ and CCMR0082 had approx. three-fold more PKS and approx. seven-fold more hybrid NRPS/PKS than PCC7375. CCMR0081$^T$ has 24 gene clusters; among these were seven PKS and eight hybrids (PKS/NRPS) (Figure 4). Compared to a diverse panel of previously examined cyanobacterial genomes (Shih *et al.,* 2013), CCMR0081$^T$ has the second highest number of NRPS/PKS clusters (Table S3).

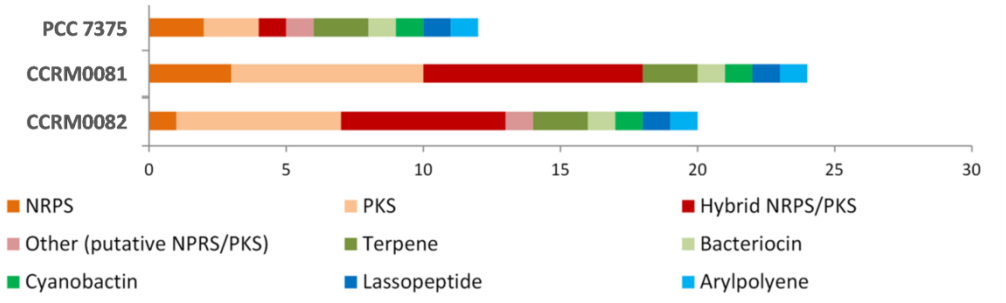



**Figure 4. Distribution of the secondary metabolites across the *Acrophormium* spp.**

Occurrence of the nonribosomal peptide, polyketide gene clusters, and gene clusters involved in ribosome-dependent synthesis of diverse peptides. Hybrids are described as resulted molecules containing nonribosomal and ribosomal traits.

## Abundance of *Acrophormium turfae* across turf metagenomes

The two *Acrophormium turfae* genomes represent a significant portion of turf metagenomes (Figure 5). Reads of the two novel genomes described here were widespread among turf metagenomes (ranging from ~1% to 22%) and were more abundant in the metagenomes from the Parcel dos Abrolhos (PAB) and Archipelago (AR) sites than the Pedra de Leste (PL) site (except for the PL1|O sample). Metagenome AR.3|O had the highest number of hits in both new genomes (Figure 5). These results showed that *A. turfae* is widely distributed across the Abrolhos reefs.

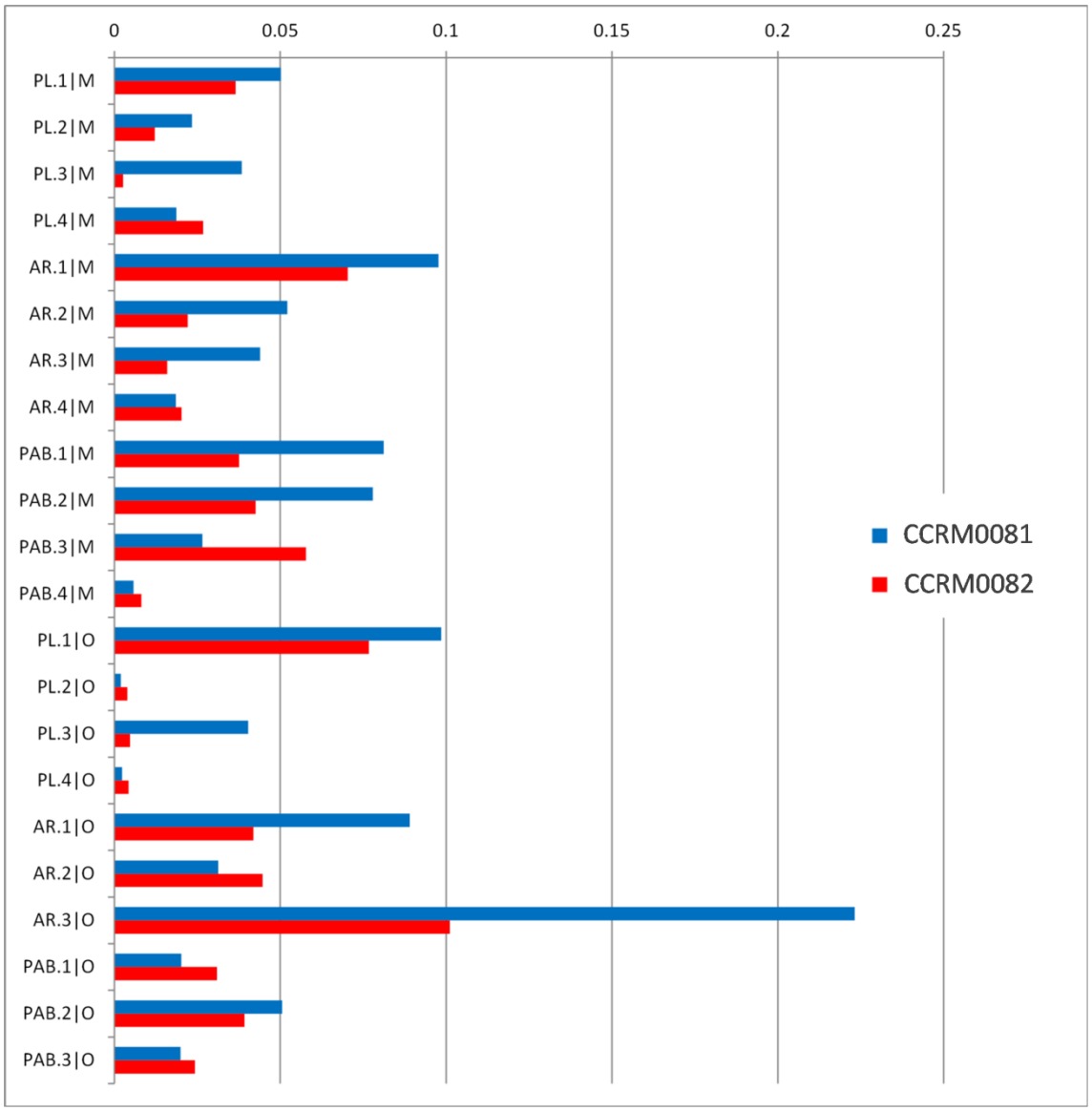



**Figure 5. Abundance and distribution profiles of the novel cyanobacterial strains across turfs metagenomes from Abrolhos coral reefs.**

Ambiguous reads were corrected as describing in Materials and Methods.

## Discussion

### Genome size and complexity of *Acrophormium turfae*

The formation of turf is an important phenotype displayed by this novel Cyanobacteria. Turfs are among the most successful benthic components of reef systems (Francini-Filho. *et al.,* 2013). The genomes, growth rates, and pigments analysed in the present study hint at several possible roles performed by the novel

Cyanobacteria in turfs. The genomes of CCMR0081$^T$ and CCMR0082 are among the largest cyanobacterial genomes. Unicellular cyanobacteria such as *Prochlorococcus* have reduced genomes (1 Mb), while filamentous cyanobacteria such as *Scytonema* sp. have large genomes (10 Mb) (Dagan *et al.*, 2013; Dufresne *et al.*, 2008). Prokaryotes with larger genomes tend to have higher phenotypic and physiologic plasticity and wide niche occupancy (Gevers *et al.*, 2005; Han *et al.*, 2013; Huntley *et al.*, 2011; Schneiker *et al.*, 2007). Cyanobacterial taxonomy is a challenge and is actually under an overhauling being the phylum re-assessment out of the scope of the present study. However, our phylogenomic reconstruction and genomic taxonomy analysis allowed us to place the new genomes within a new genus, demonstrating the genomic breadth within the cyanobacterium phylum.

## (Photo)-adaptation and nutrient related-genes

A surprising observation was that CCMR0081$^T$ and CCMR0082 were able to acclimate to far-red light by altering greatly the synthesis of PE and APC (Figure 1A-B). Occurrence for acclimative responses has been found in Cyanobacteria, such as Complementary Chromatic Acclimation (CCA) (Grossman 2003), and Far-Red Light Photoacclimation (FaRLiP) (e.g. *Leptolyngbya* JSC-1, Gan *et al*., 2014).

Through far-red light, CCMR0081$^T$ gains enhanced absorbance for APC (630 nm), while CCRM0082 raise their absorbance for PE (567 nm). However, specifically genes involved in grown under far-red light (Gan *et al*., 2014; Gan *et al*., 2015), including synthesis of Chls *f* and *d* were not found in CCMR0081$^T$ and CCMR0082, suggesting the operation of different paths. The ability to grown under far-red light presenting those characteristics may indicate a key strategy for these strains in the South Atlantic Ocean (SAO).

Regulation of the expression of phycobilisome-pigmented proteins (rod linkers of the light-harvesting antennae proteins phycoerythrin and phycocyanin) in the novel strains may be mediated by *cpe* genes. Chromatic acclimation is present in many

Cyanobacteria species and modulates them to the spectral distribution of ambient light. Variations in colour (red and blue-green) have been observed in filaments of these novel Cyanobacteria in the turfs of the Abrolhos reefs (Walter *et al.,* 2016) and in cultures (Figure S1). Environments with variable ratios of green and red light, such as reef systems, will favour cyanobacterial species that can perform chromatic acclimation (Postius *et al.,* 2001). This potential ecophysiological feature of the novel Cyanobacteria may represent an important advantage to occupy the benthic compartment of reef systems. In addition, during nutrient (N) limitation, Cyanobacteria may modify the photosynthetic apparatus through the machinery that coordinates the phycobilisome degradation (Dolganov and Grossman 1999). The phycobiliprotein : chlorophyll *a* ratios measured in our isolates are higher than ratios reported for other marine and freshwater Cyanobacteria, indicating that the novel turf-forming cyanobacteria have important N reserves in the form of phycoerythrin (Bennett and Bogorad 1973; Kana and Glibert 1987; Wojtasiewicz and Stoń-Egiert 2016). In addition, cultures of the novel strains change colour from red to green in late stationary growth phase, indicating their potential for rapid degradation of phycobilisomes under limiting nutrient conditions. The genes *nbl*A and *nbl*B, which are involved in the process of chromatic acclimation, were detected in the genomes of the two novel isolates. Both novel strains (CCMR0081$^T$ and CCMR0082) may access an alternative nitrogen source through degradation of phycobilisomes via increased activity of NblA, in addition to the nitrogen obtained by nitrogen fixation.

Phosphorus is also a crucial nutrient for Cyanobacteria growth. The genomes of the two novel strains (CCMR0081$^T$ and CCMR0082) have a higher abundance of the alkaline phosphatases gene (N = 26 - 34) than other cyanobacterial genomes. This gene encodes a key enzyme in the Pho regulon involved in phosphorus acquisition that releases phosphate (Pi) from complex organophosphate compounds (such as phosphoesters and phosphonates in the pool of dissolved organic phosphate, DOP) in a phosphate environmental limiting condition (Singh *et al.*, 2006). The Pho regulon is

also part of a complex network involved in bacterial virulence (e.g. biofilm formation and quorum-sensing circuits) and stress response (Ren *et al.,*, 2004). This ability to use organic phosphorus may grant competitive advantages to these novel cyanobacteria strains within the turf microhabitat, where organic matter is abundant and access to dissolved inorganic phosphorus from the water column may be limited (Mueller *et al.*, 2016).

The iron requirement for photosynthesis in Cyanobacteria is exceptionally large. Only recently have the mechanisms of iron uptake in Cyanobacteria been elucidated (Jiang *et al.*, 2015). The siderophore-mediated iron uptake mechanism described in non-photosynthetic bacteria has not been identified in Cyanobacteria (Hopkinson and Morel 2009; Ito and Butler 2005; Mirus *et al.*, 2009; Stevanovic *et al.*, 2012). Some Cyanobacteria can use iron bound to the siderophores of other organisms (Kranzler *et al.*, 2011). Iron ABC transporters, the iron-uptake regulation protein FUR and the ExbB-ExbD complexes are involved in the activation of different classes of outer membrane transporters for inorganic iron uptake (Jiang *et al.*, 2015). Several copies of ExbB (3-4), ExbD (3-5) and a Ferric Iron ABC transporter permease (4-6) were found in the two novel genomes in the present study, whereas siderophores genes were found in the Abrolhos turf metagenomes (Walter *et al.,* 2016), indicating that a possible efficient consortium was assembled in these turfs for iron capture and use (Walter *et al.,* 2016).

## Secondary metabolites

The *Acrophormium turfae* strains CCMR0081$^{T}$ and CCMR0082 have multiple predicted genes coding for secondary metabolites, ranking them among the highest within the Cyanobacteria (Shih *et al.,* 2013). According to a previous study, the phylogenetic clade encompassing PCC7375 (major subclade C) has a minor fraction of secondary metabolite genes (Shih *et al.,* 2013). CCMR0081$^{T}$ had the second highest number of PKS, NRPS and hybrid clusters (Table S3), the first being *Fischerella* sp. PCC9339 from clade B. While most Cyanobacteria has an average of five NRPS/PKS

clusters per genome (approximately 5% of cyanobacterial genomes) (Shih *et al.*, 2013), CCMR0081$^{T}$ and CCMR0082 have an average of 15 clusters.

## Ecological relevance of *Acrophormium turfae*

Analysis of metagenomes representing a wide area of the Abrolhos reefs and two different seasons (Walter et al., 2016) unveiled a vast spatial distribution, persistence, and high relative biomass contribution of *A. turfae* in turfs across this important coral reef ecosystem in the SAO. This reef-wide occurrence of *A. turfae* along a 50 Km cross-shelf gradient highlights the ability of this cyanobacterium in colonizing the benthic substrate over a vast range of environmental physical and chemical properties e.g. irradiance, temperature and nutrient levels. Such relatively large physiological plasticity is in agreement with the rich genomic repertoire described herein for the species, that includes potential for chromatic acclimation, nitrogen fixation, and high affinity for iron and phosphorus. Filamentous cyanobacteria-rich turf coverage is increasing in Abrolhos as well as in other reef systems worldwide, outcompeting corals (Knowlton and Jackson 2008). Direct, positive effects of physical and chemical environmental factors (e.g. temperature and nutrients) on the Cyanobacteria growth are often evoked as the underlying mechanism for this observed expansion of turfs, combined with indirect factors that reduce coral coverage e.g. infectious diseases that open niches for turf and macroalgae colonization (Richardson *et al.*, 2009). Both *A. turfae* strains described here had relatively low growth rates (doubling times up to 9 d) at light and temperature levels representative of their natural environment. Thus, it reasonable to assume that *A. turfae* widespread occurrence in the environment is due to other characteristics than simply growth potential. On the other hand, the strikingly high abundance and diversity of secondary metabolite genes unveiled in this cyanobacterium is likely to be the key to its success in Abrolhos reefs. Secondary metabolite genes are involved in important ecological processes in Cyanobacteria, such as signalling pathways, resource competition through allelopathy,

feeding deterrence, and UV protection. Moreover, these metabolites include various bioactive compounds known to negatively influence coral health (Barott et al., 2012; Charpy et al., 2012; Gregg et al., 2013; Wild et al., 2014). Persistence and abundance of a species in its environment is ultimately determined by the balance between gain and loss factors. Herbivorous fish and small invertebrates are common members of reef systems representing potential loss factor for microbial mats like cyanobacteria-rich turfs. Thus, any increased grazing deterrence capability due to high production of secondary metabolites may confer the necessary competitive advantage for the establishment and spread of these microbial assemblages.

## Conclusions

This study is a first attempt to improve our understanding of the function of turf-forming Cyanobacteria. An unexpected high abundance and diversity of potential enzymes coding for secondary metabolite biosynthesis, and an uncommon acclimative response to far-red light (enhanced absorption in PE and APC nm) are the most distinguishable features of the novel *A. turfae*. Based on its unique genomic features and basic physiological traits, we hypothesize that *A. turfae* is likely a key species for turfs success in the reef system due to its high potential to produce secondary metabolites that might help to prevent the whole assemblage from being heavily impacted by grazers and by negatively affecting competing benthic organisms. The high number of photosynthesis-related genes and the high content of phycobiliproteins hint at the potential of these Cyanobacteria to thrive under light limitation, a possible scenario in reef systems. The presence of a diverse secondary metabolite genetic repertoire may allow these Cyanobacteria to inhibit the growth of competing benthic organisms by means of toxin production. Further, our study demonstrates that novel Cyanobacteria isolates have the potential to retrieve nitrogen from both $N_2$ fixation and phycobilisome degradation.

## ACKNOWLEDGMENT

The authors acknowledge National Counsel of Technological and Scientific Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES), and Rio de Janeiro Research Foundation (FAPERJ) for their financial support for this study. JMW thank CNPq (207751-2014/5) and CAPES for the PhD scholarships. This paper is part of the D. Sc. requirements of JMW at the Biodiversity and Evolutionary Biology Graduate Program of the Federal University of Rio de Janeiro (UFRJ). The sequence data selected from the Geba database (Accession number 2509601039) and used in this paper was produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community.

## Field Study Permissions

The following information was supplied relating to field study approvals: sampling permit SISBIO no. 27147-2 issue by the Ministry of Environment Institute Chico Mendes (ICMBio).

## REFERENCES

Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Auch AF, von Jan M, Klenk HP, Göker M. (2010). Digital DNA-DNA Hybridization for Microbial Species Delineation by Means of Genome-to-Genome Sequence Comparison. *Stand in Gen Scien* **2**(1):117–34.

Auch AF, Klenk HP, Göker M. (2010). Standard Operating Procedure for Calculating Genome-to-Genome Distances Based on High-Scoring Segment Pairs. *Stand in Gen Scien* **2**(1):142–48.

Bankevich A, Nirk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**(5):455–77.

Barott KL, Rodriguez-Mueller B, Youle M, Marhaver KL, Vermeij MJ, Smith JE, *et al.* (2012). Microbial to Reef Scale Interactions between the Reef-Building Coral Montastraea Annularis and Benthic Algae. *Proc R Soc B: Biol Sci* **279**:1655–64.

Bennett A, Bogorad L. (1973). Complementary Chromatic Adaption in a Filamentous Blue-Green Alga. *J Cell Biol* **58**:419–35.

Bonfield JK, Whitwham A. (2010). Gap5-Editing the Billion Fragment Sequence Assembly. *Bioinformatics* **26**(14):1699–1703.

Bryant DA, Guglielmi G, de Marsac NT, Castets AM, Cohen-Bazire G. (1979). The Structure of Cyanobacterial Phycobilisomes: A Model. *Archi Microb* **123**:113–27.

Buchfink B, Xie C, Huson DH. (2014). Fast and Sensitive Protein Alignment Using DIAMOND. *Nat Methods* **12**(1):59–60.

Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ. (1997). Trichodesmium, a

Globally Significant Marine Cyanobacterium. *Science* **276**(5316):1221–29.

Charpy L, Casareto BE, Langlade MJ, Suzuki Y. (2012). Cyanobacteria in Coral Reef Ecosystems: A Review. *J Mar Bio* **2012**:1–9.

Cheung MY, Liang S, Lee J. (2013). Toxin-Producing Cyanobacteria in Freshwater: A Review of the Problems, Impact on Drinking Water Safety, and Efforts for Protecting Public Health. *J Microbiol* **51**(1):1–10.

Coil D, Jospin G, Darling AE. (2014). A5-Miseq : An Updated Pipeline to Assemble Microbial Genomes from Illumina MiSeq Data. *Bioinformatics* **2**:5–8.

Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, *et al.* (2013). Genomes of Stigonematalean Cyanobacteria (Subsection V) and the Evolution of Oxygenic Photosynthesis from Prokaryotes to Plastids. *Genome Biol Evol* **5**(1):31–44.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, *et al.* (2009). Community-Wide Analysis of Microbial Genome Sequence Signatures. *Genome Biol* **10**(8):R85.

Dolganov N, Grossman AR. (1999). A Polypeptide with Similarity to Phycocyanin - Subunit Phycocyanobilin Lyase Involved in Degradation of Phycobilisomes. *J Bacteriol* **181**(2):610–17.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, *et al.* (2008). Unraveling the Genomic Mosaic of a Ubiquitous Genus of Marine Cyanobacteria. *Genome Biol* **9**(5):R90.

Dupraz C, Reid P, Braissant O, Decho AW, Norman RS, Visscher PT. (2009). Processes of Carbonate Precipitation in Modern Microbial Mats. *Earth-Sci Rev* **96**(3):141–62.

Gan F, Zhang S, Rockwell NC, Martin SS, Lagarias JC, Bryant DA (2014) Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science*. 345:1312–1317. doi:10.1126/science.1256963

Gan F, Zhang S, Bryant DA (2015) Occurrence of Far-Red Light Photoacclimation (FaRLiP) in Diverse Cyanobacteria. *Life*. 5: 4-24. doi:10.3390/life5010004

Garcia-Pichel F, Belnap J, Neuer S, Schanz F. (2003). Estimates of Global Cyanobacterial Biomass and Its Distribution. *Algol Stud* **109**(1):213–27.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, *et al.* (2005). Re-Evaluating Prokaryotic Species. *Nat Rev Microbiol* **3**(1740–1526):733–39.

Grant JR, Arantes AS, Stothard P. (2012). Comparing Thousands of Circular Genomes Using the CGView Comparison Tool. *BMC Genomics* **13**(1):202.

Gregg AK, Hatay M, Haas AF, Robinett NL, Barott K, Vermeij MJA, *et al.* (2013). Biological Oxygen Demand Optode Analysis of Coral Reef-Associated Microbial Communities Exposed to Algal Exudates. *PeerJ* **1**:e107. doi: 10.7717/peerj.107.

Grissa I, Vergnaud G, Pourcel C. (2007). The CRISPRdb Database and Tools to Display CRISPRs and to Generate Dictionaries of Spacers and Repeats. *BMC Bioinformatics* **8**:172.

Grossman AR (2003) A molecular understanding of complementary chromatic adaptation. *Photosynth. Res*. 2003, 76, 207–215, doi:0.1007/1-4020-3324-9_85.

Guillard RRL, Ryther JH. (1962). Studies of Marine Planktonic Diatoms: I. Cyclotella Nana Hustedt, and Detonula Confervacea (Cleve) Gran. *Can J Microbiol* **8**(1140):229–39.

Guillard RL. (1975). Culture of Phytoplankton for Feeding Marine Inverterbrates. p. 26–60 in Culture of Marine Invertebrate Animals, edited by Smith WL and Chanley MH. New York, NY: Plenum Press.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **29**(8):1072–75.

Hopkinson B, Brian M., Morel FMM. (2009). The Role of Siderophores in Iron Acquisition by Photosynthetic Marine Microorganisms. *BioMetals* **22**(4):659–69.

Ito Y, Butler A. (2005). Structure of Synechobactins, New Siderophores of the Marine Cyanobacterium Synechococcus Sp. PCC 7002. *Limnol Oceanogr* **50**(6):1918–23.

Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. (2012). Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science* **335**(6068):587–90.

Jiang HB, Lou WJ, Ke WT, Song WY, Price NM, Qiu BS. (2015). New Insights into Iron Acquisition by Cyanobacteria: An Essential Role for ExbB-ExbD Complex in Inorganic Iron Uptake. *ISME J* **9**:1–13.

Kana TM, Glibert PM. (1987). Effect of Irradiances up to 2000 μE M-2 S-1 on Marine Synechococcus WH7803. Growth, Pigmentation, and Cell Composition. *Deep Sea Research Part I: Oceanogr Res Pap* **34**(4A):479–95.

Karlin S, Mrázek J, Campbell AM. (1997). Compositional Biases of Bacterial Genomes and Evolutionary Implications. *J Bacteriol* **179**(12):3899–3913.

Knowlton N, Jackson JBC. (2008). Shifting Baselines, Local Impacts, and Global Change on Coral Reefs. *PLoS Biol* **6**(2):e54.

Konstantinidis KT & Tiedje JM. (2005). Genomic Insights That Advance the Species Definition for Prokaryotes. *Proc Natl Acad Sci USA* **102**(7):2567–72.

Kranzler C, Lis H, Shaked Y, Keren N. (2011). The Role of Reduction in Iron Uptake Processes in a Unicellular, Planktonic Cyanobacterium. *Environ Microbiol* **13**(11):2990–99.

Langmead B, Salzberg SL. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods* **9**(4):357–59.

Luque I, Zabulon G, Contreras A, Houmard J. (2001). Convergence of Two Global Transcriptional Regulators on Nitrogen Induction of the Stress-Acclimation Gene nblA in the Cyanobacterium Synechococcus Sp. PCC 7942. *Mol Microbiol* **41**(4):937–47.

Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. (2013). Genome Sequence-Based

Species Delimitation with Confidence Intervals and Improved Distance Functions. *BMC bioinformatics* **14**:60.

Micallef ML, D'Agostino PM, Al-Sinawi B, Neilan BA, Moffitt MC. (2014). Exploring Cyanobacterial Genomes for Natural Product Biosynthesis Pathways. *Mar Genomics* **21**:1–12.

Mirus O, Strauss S, Nicolaisen K, von Haeseler A, & Schleiff E. (2009). TonB-Dependent Transporters and Their Occurrence in Cyanobacteria. *BMC Biol* **7**(1):68.

Mueller B, den Haan J, Visser PM, Vermeij MJA, van Duyl FC. (2016). Effect of Light and Nutrient Availability on the Release of Dissolved Organic Carbon (DOC) by Caribbean Turf Algae. *Sci Rep* **6**:23248.

Muramatsu M, Hihara Y. (2012). Acclimation to High-Light Conditions in Cyanobacteria: From Gene Expression to Physiological Responses. *J Plant Res* **125**(1):11–39.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res* **25**(7):1043–55.

Paul R, Jinkerson RE, Buss K, Steel J, Mohr R, Hess WR, *et al.* (2014). Draft Genome Sequence of the Filamentous Cyanobacterium Leptolyngbya Sp. Strain Heron Island J, Exhibiting Chromatic, Acclimation. *Genome Announc* **2**(1):e01166-13.

Richardson LL, Miller AW, Broderick E, Kaczmarsky L, Gantar M, Stanic D, *et al.* (2009). Sulfide, Microcystin, and the Etiology of Black Band Disease. *Dis Aquat Org* **87**(1–2):79–90.

Schmieder R, Edwards R. (2011). Quality Control and Preprocessing of Metagenomic Datasets. *Bioinformatics* **27**(6):863–64.

Seemann T. (2014). Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **30**(14):2068–69.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, *et al.* (2013). Improving the Coverage of the Cyanobacterial Phylum Using Diversity-Driven Genome Sequencing. *Proc Natl Acad Sci USA* **110**(3):1053–58.

Singh SK, Singh SS, Pandey VD, Mishra AK. (2006). Factors Modulating Alkaline Phosphatase Activity in the Diazotrophic Rice-Field Cyanobacterium, Anabaena Oryzae. *World J Microbiol Biotechnol* **22**(9):927–35.

Stevanovic M, Hahn A, Nicolaisen K, Mirus O, Schleiff E. (2012). The Components of the Putative Iron Transport System in the Cyanobacterium Anabaena Sp. PCC 7120. *Environ Microbiol* **14**(7):1655–70.

Sweet MJ, Bythell JC, Nugues MM. (2013). Algae as Reservoirs for Coral Pathogens. *PLoS ONE* **8**(7):e69717.

Tenório MMB, Le Borgne R, Rodier M, Neveux J. (2005). The Impact of Terrigeneous Inputs on the Bay of Ouinné (New Caledonia) Phytoplankton Communities: A Spectrofluorometric and Microscopic Approach. Estuarine, Coastal and Shelf

*Science* **64**:531–45.

Tritt A, Eisen JA, Facciotti MT, Darling AE. (2012). An Integrated Pipeline for de Novo Assembly of Microbial Genomes. *PLoS ONE* **7**(9).

Wadhams GH, Armitage JP. (2004). Making Sense of It All: Bacterial Chemotaxis. *Nat Rev Mol Cell Biol* **5**(12):1024–37.

Walter JM, Tschoeke D, Meirelles PM, de Oliveira L, Leomil L, Tenório M, *et al.* (2016). Taxonomic and Functional Metagenomic Signature of Turfs in the Abrolhos Reef System (Brazil). *PLoS ONE* **11**(8):e0161168.

Watanabe M, Semchonok DA, Webber-Birungi MT, Ehira E, Kondo K, Narikawa R, *et al.,* (2014). Attachment of Phycobilisomes in an Antenna-Photosystem I Supercomplex of Cyanobacteria. *Proc Natl Acad Sci USA* **111**(7):2512–17.

Waterbury JB, Watson SW, Guillard RRL, Brand LE. (1979). Widespread Occurance of a Unicellular, Marine, Planktonic, Cyanobacterium. *Nature* **277**:293–94.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, *et al.* (2015). antiSMASH 3.0--a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res* **43**:237–43.

Welker M, von Döhren W. (2006). Cyanobacterial Peptides - Nature's Own Combinatorial Biosynthesis. *FEMS Microbiol Rev* **30**(4):530–63.

Wild C, Jantzen C, Kremb SG. (2014). Turf Algae-Mediated Coral Damage in Coastal Reefs of Belize, Central America. *PeerJ* **2**:e571.

Wojtasiewicz B, Stoń-Egiert J. (2016). Bio-Optical Characterization of Selected Cyanobacteria Strains Present in Marine and Freshwater Ecosystems. *J Appl Phycol* 1–16.

Wu M, Eisen JA. (2008). A Simple, Fast, and Accurate Method of Phylogenomic Inference. *Genome Biology* **9**(10):R151.

Zerbino DR, Birney E. (2008). Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs. *Genome Res* **18**(5):821–29.

## Supplementary Information

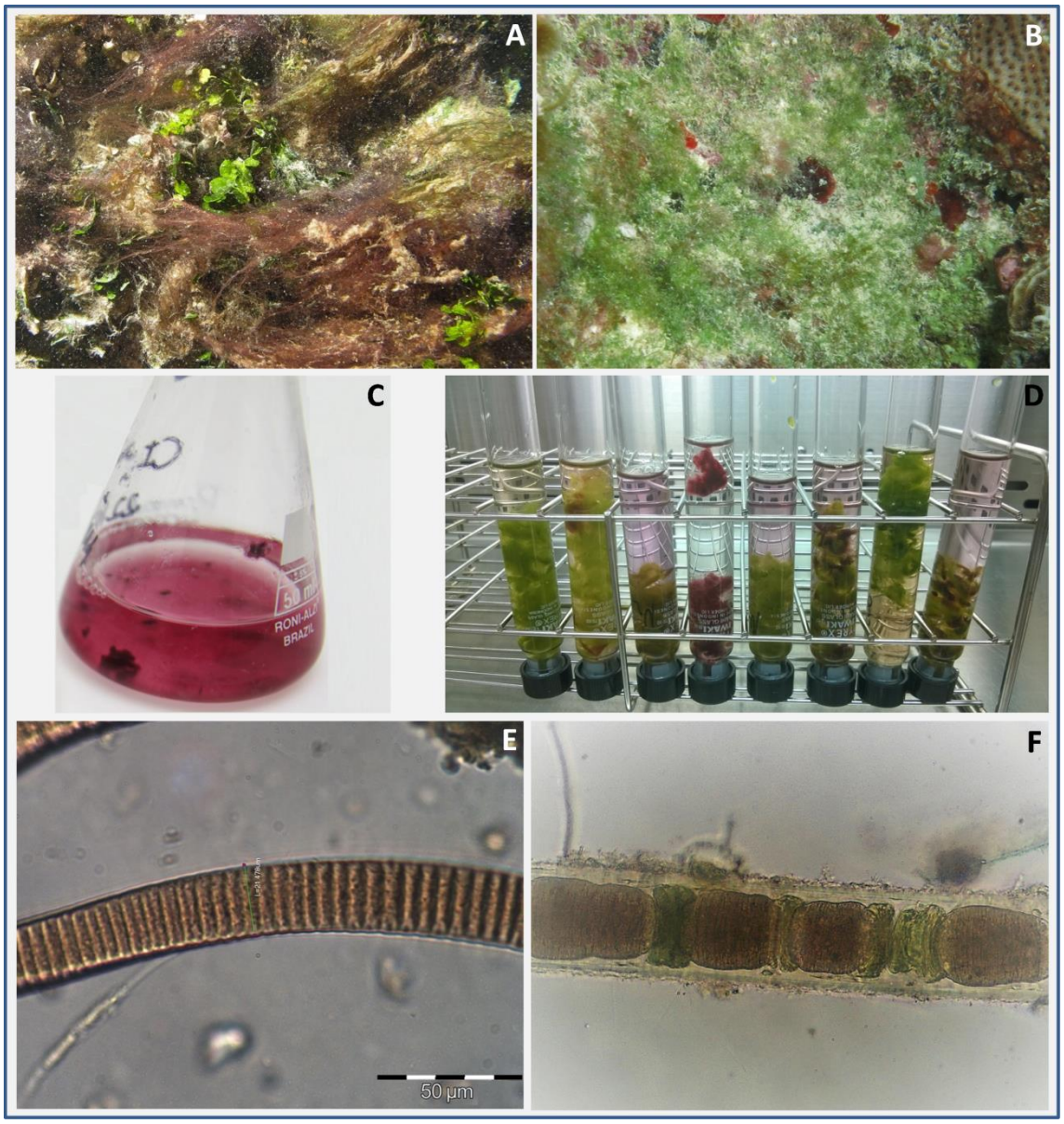



**Figure S1.** Overview of turfs and *Acrophormium turfae* culture isolated from them. (A-B) Macroscopic view of a typical cyanobacteria-dominated turf mat growing over corals in Abrolhos reefs. Turfs presenting variations in color (brick red and blue-green). The red color results from phycoerythrin-rich filamentous cyanobacteria. (C-D) Example of the cyanobacterial strain growing in liquid f/2 medium, forming tufts of filamentous mats contained phycoerythrin, where in D shown culture in different stages of pigment loss, showing variations in colors. (E-F) Filaments of Cyanobacteria observed under microscope, where in F shown detail of cells variation color from red to blue-green.

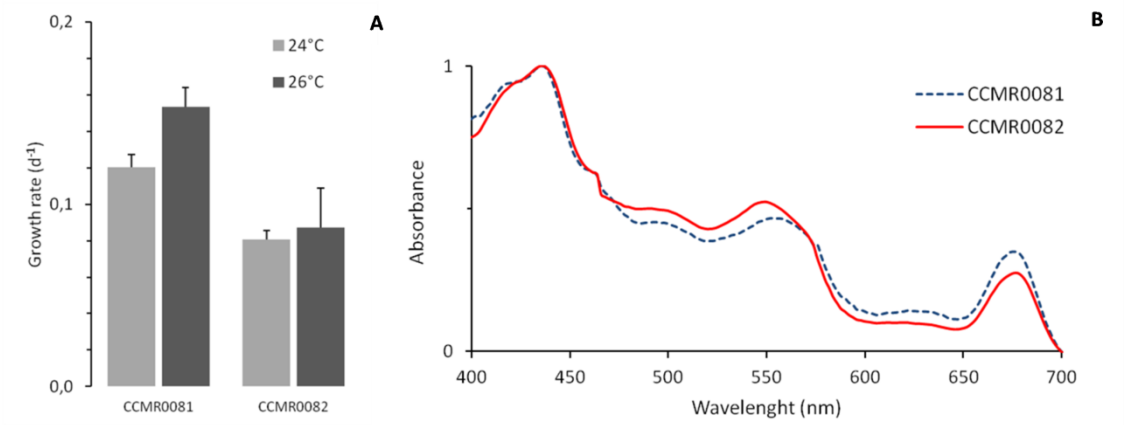



**Figure S2.** Growth rates and optical characteristics of strains CCMR0081$^T$ and CCMR0082. (A) Specific growth rates (μ) of strains CCMR0081$^T$ and CCMR0082 cultivate at 24°C and 26°C in F ½ culture medium at 50 μE m-2 s-1 with a 16:8 light:dark cycle. Bars are the average of 4 replicates. Whiskers: 1 SD. (B) Optical characteristics of strains CCMR0081$^T$ and CCMR0082. *In vivo* absorption spectra from 400 to 700 nm of both strains grown under the same irradiance (50 μE $m^{-2}$ $s^{-1}$; white light; 16:8 light:dark cycle) and temperature (24°C). Cells were harvested from a 7-day old culture. The spectra were normalized at the blue absorption peak of chlorophyll *a*. Both strains showed strong absorbance at the PE region (550-552 nm).

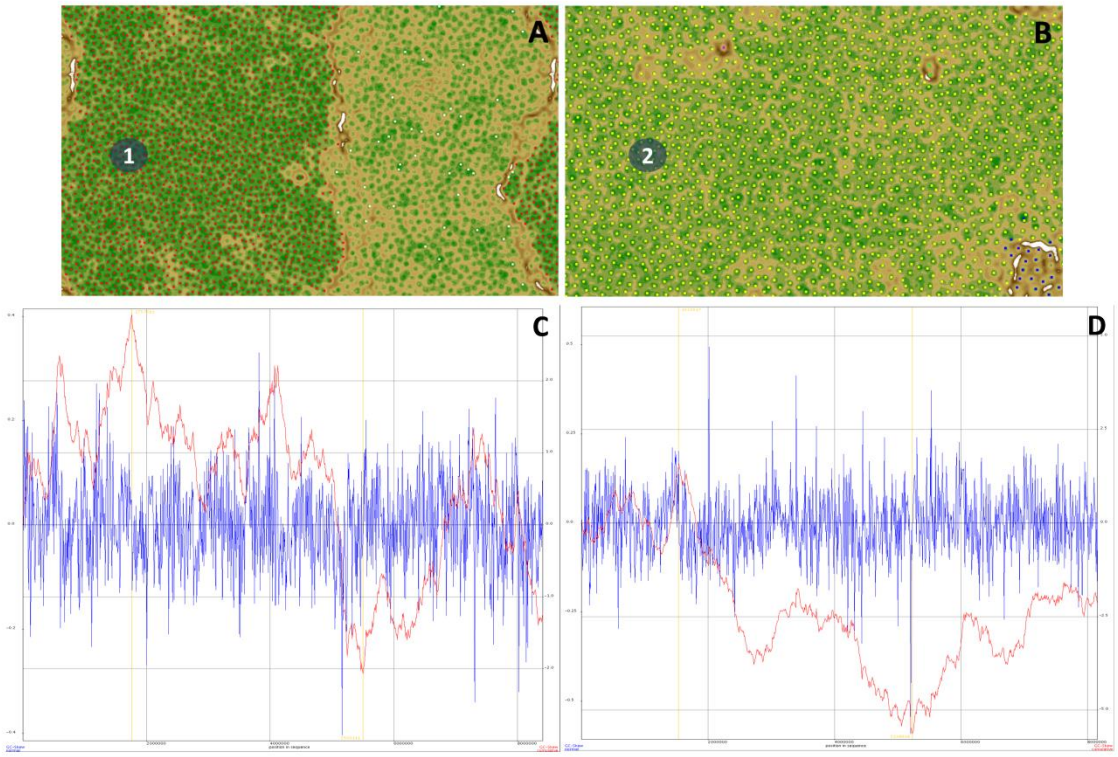



**Figure S3.** Binning and G+C skew. (A-B) ESOM binning of the CCRM0081 and CCRM0082 isolates bins. Sequence fragments were assigned to particular genomic bins on the basis of tetranucleotide frequency. Maps are showed in tiled mode,

revealing complete bins. Colors represent the differences in tetranucleotide frequency profiles between nodes of the ESOM matrix. Same color presenting same bins. Only one grid is displayed to each isolate, avoiding redundancy of data points. Respective bins are labeled with numbers. Panel (A) shows a map revealing the CCRM0081 bins (highlighted with number 1, red), and (B) shows a map revealing the CCRM0082 bins (highlighted with number 2, yellow). (C-D) Base frequencies deviations of G=C for both genomes, showing the origin and the terminus in each circular chromosomes (http://genskew.csb.univie.ac.at/).

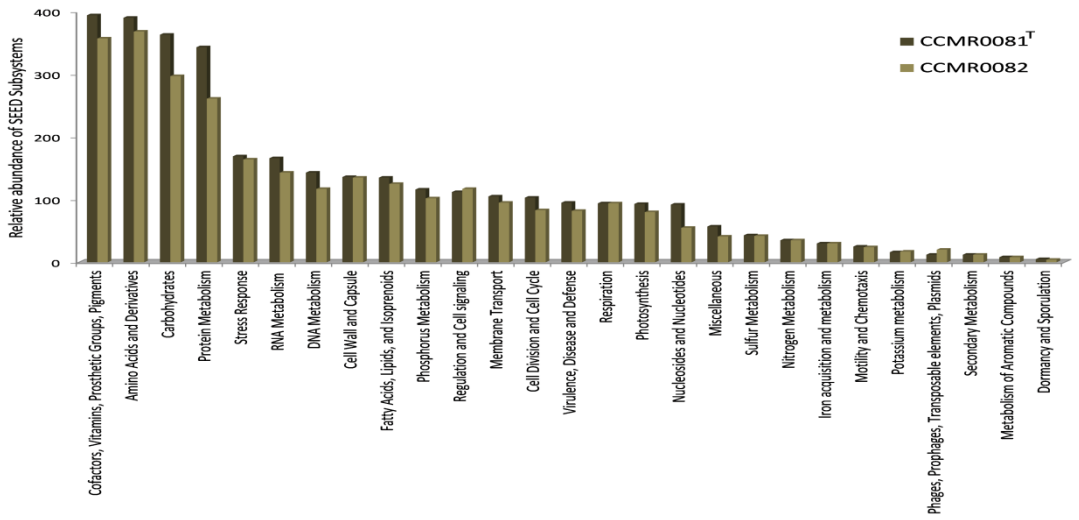



**Figure S4.** Distribution of *Acrophormium turfae* strains genomic sequences categorized into functional roles (level 1 – SEED subsystems).

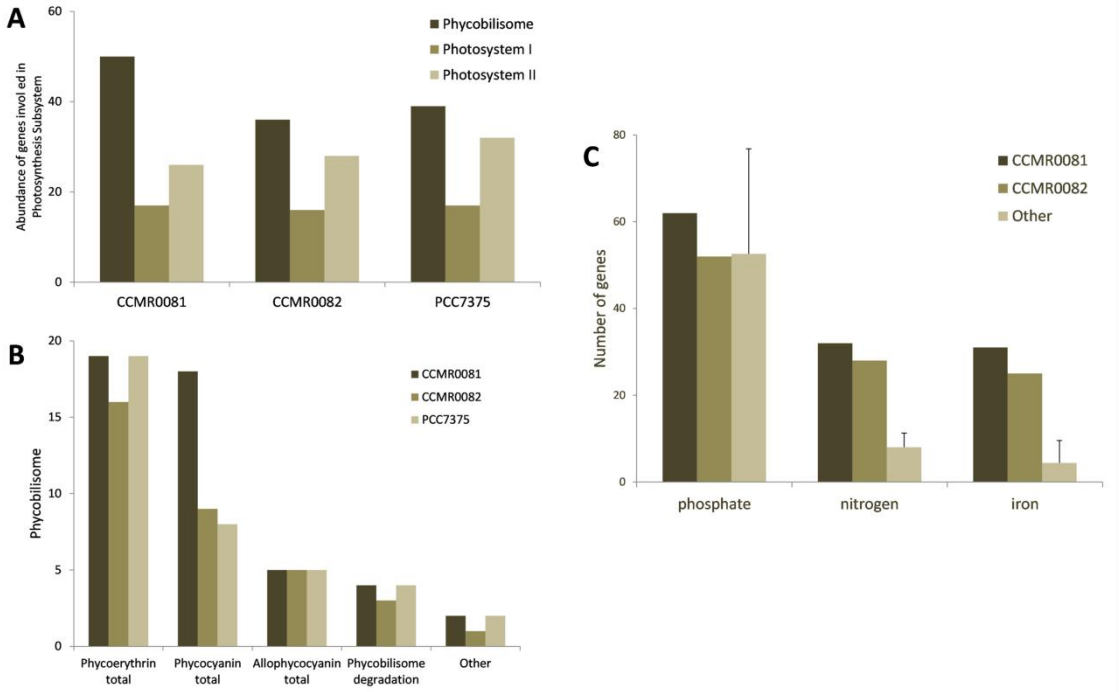


**Figure S5.** Distribution of genes involved in photosynthesis apparatus. (A) Distribution of genes involved in light-harvesting machineries. Photosystem II major genes: *psb*A, *psb*27, *psb*O, *psb*I, *psb*H, *psb*C, *psb*D, *psb*P, *psb*E, *psb*Z, *psb*W, *psb*F, *psb*J, *psb*B, *psb*K, *psb*N, *psb*V, *psb*U (N = 23 for CCRM0081$^T$; N = 24 for CCRM0082; and N = 28 for PCC7375). CCRM0082 and PCC7375 also hold *psb*L and *psb*X genes, whereas CCRM0081$^T$ carry on the *psb*T gene. Photosystem I major genes: *psa*E, *psa*K, *psa*K1, *psa*D, *psa*J, *psa*L, *psa*B, *psa*C, *psa*F, *psa*A (N = 13 for CCRM0081$^T$; N = 12 for CCRM0082; and N = 13 for PCC7375). (B) Abundance of genes involved in phycobilisome structure formation in CCRM0081$^T$, CCRM0082 and PCC7375. Plots shown the presence of genes involved in alpha and beta chains formation of phycocyanin, phycocyanobilin lyase, phycoerythrin, and allophycocyanin, further those involved in rod-core linker proteins of phycocyanin, phycoerythrin (CpeR and CpeS homologs), and in core-membrane linker polypeptide of phycobilisome. Also, they hold genes involved in phycobilisome *in vivo* degradation process (*nbl*A, *nbl*B). (C) Abundance of Photosystem I and II and Phycobilisome related-genes presents in the Photosynthesis subsystem in *Acrophormium turfae* strains.

**Table S1.** Prediction of RNAs in cyanobacterial genomes. Prokka tool (using Prodigal) was used to annotate genes.

| Species | Strain ID | RNAs (total) | tRNAs | rRNAs | Other RNA genes | CDS |
|---|---|---|---|---|---|---|
| *Acrophormium turfae* | CCMR0081 | 66 | 60 | 6 | 0 | 7276 |
| *Acrophormium turfae* | CCMR0082 | 61 | 58 | 3 | 0 | 7713 |
| *Acrophormium splendidus* | PCC 7375 * | 82 | 63 | 9 | 10 | 8366 |

* Formerly known as Leptolyngbya sp. PCC 7375. Reference: Shih *et al*. (2013).

**Table S2.** The 30 more abundant proteins of the core genome of CCRM0081$^{T}$, CCRM0082, and PCC7375 strains.

| Protein | Function | SS | Abundance |
|---|---|---|---|
| High-affnity carbon uptake protein Hat | $CO_2$ fixation | Carbohydrates | 126 |
| Glycosyl transferase | Biosynthesis of galactoglycans and related lipopolysacharides | Carbohydrates | 103 |
| Circadian input kinase A | Cyanobacterial circadian clock | Cell Division and Cell Cycle | 87 |
| Adenylate cyclase (EC 4.6.1.1) | cAMP signaling in bacteria | Regulation and Cell signaling | 77 |
| Membrane proteins related to metalloendopeptidases | Putative protein export | Clustering-based subsystems | 44 |
| Ton_and_Tol_transport_systems | Ton and Tol transport systems | Membrane Transport | 43 |
| Alkaline phosphatase (EC 3.1.3.1) | Phosphate uptake | Phosphate Metabolism | 39 |
| diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) | Bacterial hemoglobins | Stress Response | 37 |
| Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) | Cell wall of Mycobacteria (mycolic acid synthesis) | Cell Wall and Capsule | 29 |
| Permease of the drug/metabolite transporter (DMT) superfamily | RNA processing and modification (Queuosine-Archaeosine biosynthesis) | RNA Metabolism | 28 |
| Ferredoxin | Inorganic sulfur assimilation | Sulfur Metabolism | 23 |
| Chemotaxis regulator (CheY) | Bacterial chemotaxis | Virulence, Disease and Defense | 21 |
| Serine phosphatase RsbU | SigmaB stress responce regulation | Stress Response | 19 |
| 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) | Cell wall of Mycobacteria (mycolic acid synthesis) | Cell Wall and Capsule | 19 |
| Cysteine desulfurase (EC 2.8.1.7) | RNA processing and modification | RNA Metabolism | 18 |
| N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) | Involved in recycling and identification of peptidoglycan (cell wall and capsule) | Regulation and Cell signaling | 18 |
| Aspartate aminotransferase (EC 2.6.1.1) | Ammonia assimilation | Amino Acids and Derivatives | 16 |

| | | | |
|---|---|---|---|
| DnaJ-class molecular chaperone CbpA | Protein folding | Protein Metabolism | 13 |
| Type_IV_pilus\|\|Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) | Protein and nucleoprotein secretion system, Type IV pilus | Membrane Transport | 12 |
| Alkyl hydroperoxide reductase subunit C-like protein | Oxidative stress | Stress Response | 12 |
| Exonuclease SbcC | bacterial DNA repair | DNA Metabolism | 12 |
| transcriptional regulator (Crp/Fnr family) | Oxidative stress | Stress Response | 12 |
| soluble [2Fe-2S] ferredoxin | Soluble cytochromes and functionally related electron carriers | Respiration | 12 |
| Universal stress protein family | Universal stress protein family | Stress Response | 12 |
| Chaperone protein DnaK | Heat shock dnaK gene cluster extended | Stress Response | 12 |
| Adenylate kinase (EC 2.7.4.3) | Phosphotransferases with a phosphate group as acceptor | Nucleosides and Nucleotides | 12 |
| Phycocyanin alpha phycocyanobilin lyase (NblB) | Light-harvesting complexes (Phycobilisome) | Photosynthesis | 12 |
| Two-component system response regulator (virulence regulator) | Regulation of virulence | Regulation and Cell signaling | 12 |
| GTPase (FtsY) | Protein biosynthesis | Protein Metabolism | 12 |
| Phosphoglycerate mutase (EC 5.4.2.1) | Central carbohydrate metabolism | Carbohydrates | 11 |
| ADP-ribose pyrophosphatase (EC 3.6.1.13) | NAD and NADP cofactor biosynthesis | Cofactors, Vitamins, Prosthetic Groups, Pigments | 10 |
| Thioredoxin | Selenoproteins related | Protein Metabolism | 9 |
| No Hit | NA | NA | 5747 |

**Table S3.** Distribution of the nonribosomal encoded peptide and polyketide biosynthetic pathways in Cyanobacteria. Number and occurrence of the nonribosomal peptide and polyketide gene clusters within each genome. Data was retrieved from Shih *et al*. (2013) using selected cyanobacterial genomes. Corresponding name and details of each strain was maintained according Shih *et al*. (2013).

| Strain | Strain ID | Genome size (Mb) | NCBI Project ID | PKS | Hybrid NRPS/PKS | NRPS | Total |
|---|---|---|---|---|---|---|---|
| *Acaryochloris* sp. | CCMEE 5410 | 7.88 | 16707 | 0 | 0 | 0 | 0 |
| *Acaryochloris marina* | MBIC11017$^{T}$ | 8.36 | 12997 | 0 | 1 | 0 | 1 |
| *Chamaesiphon minutus* | PCC 6605 | 6.76 | 158825 | 1 | 2 | 2 | 5 |
| *Crocosphaera watsonii* | WH 0003 | 5.89 | 61839 | 2 | 1 | 6 | 9 |
| *Crocosphaera watsonii* | WH 8501 | 6.24 | 10651 | 1 | 1 | 7 | 9 |
| *Cyanobacterium aponinum* | PCC 10605$^{T}$ | 4.18 | 158691 | 0 | 0 | 0 | 0 |
| *Cyanobacterium stanieri* | PCC 7202 $^{T}$ | 3.16 | 39697 | 0 | 1 | 0 | 1 |
| *Cyanobium gracile* | PCC 6307 $^{T}$ | 3.34 | 158695 | 2 | 0 | 0 | 2 |
| *Cyanobium* sp. | PCC 7001 | 2.83 | 19301 | 1 | 0 | 0 | 1 |
| *Cyanothece* sp. | ATCC 51142 | 5.46 | 20319 | 1 | 1 | 1 | 3 |
| *Cyanothece* sp. | ATCC 51472 | 5.46 | 59973 | 1 | 1 | 1 | 3 |
| *Cyanothece* sp. | CCY 0110 | 5.88 | 18951 | 1 | 1 | 2 | 4 |
| *Cyanothece* sp. | PCC 7424 | 6.55 | 20479 | 2 | 1 | 3 | 6 |
| *Cyanothece* sp. | PCC 7425 | 5.79 | 28337 | 1 | 0 | 0 | 1 |
| *Cyanothece* sp. | PCC 7822 | 7.84 | 28535 | 1 | 2 | 1 | 4 |
| *Cyanothece* sp. | PCC 8801 | 4.79 | 20503 | 0 | 1 | 2 | 3 |
| *Cyanothece* sp. | PCC 8802 | 4.8 | 28339 | 0 | 1 | 2 | 3 |
| *Dactylococcopsis salina* | PCC 8305 | 3.78 | 158703 | 0 | 0 | 1 | 1 |
| *Geminocystis herdmanii* | PCC 6308 $^{T}$ | 4.26 | 62511 | 1 | 0 | 0 | 1 |
| *Gloeobacter violaceus* | PCC 7421 $^{T}$ | 4.66 | 9606 | 3 | 0 | 0 | 3 |
| *Gloeocapsa* sp. | PCC 73106 | 4.03 | 159497 | 5 | 1 | 1 | 7 |
| *Gloeocapsa* sp. | PCC 7428 | 5.88 | 158831 | 2 | 2 | 2 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Halothece* sp. | PCC 7418 | 4.18 | 40817 | 0 | 0 | 0 | 0 |
| *Microcystis aeruginosa* | NIES-843 | 5.84 | 27835 | 1 | 3 | 1 | 5 |
| *Microcystis aeruginosa* | PCC 7806 | 5.2 | 15702 | 3 | 2 | 1 | 6 |
| *Prochlorococcus marinus* | AS9601 | 1.67 | 13548 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9202 | 1.69 | 19343 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9211 | 1.69 | 13551 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9215 | 1.74 | 18633 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9301 | 1.64 | 15746 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9303 | 2.68 | 13496 | 1 | 0 | 1 | 2 |
| *Prochlorococcus marinus* | MIT9312 | 1.71 | 13910 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | MIT9313 | 2.41 | 220 | 1 | 0 | 0 | 1 |
| *Prochlorococcus marinus* | MIT9515 | 1.7 | 13617 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | NATL1A | 1.86 | 15660 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus* | NATL2A | 1.84 | 13911 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus*, subsp. *marinus* | CCMP1375$^{T}$ | 1.75 | 419 | 0 | 0 | 0 | 0 |
| *Prochlorococcus marinus*, subsp. *marinus* | CCMP1986 | 1.66 | 213 | 0 | 0 | 0 | 0 |
| *Prochloron didemni* (metagenome) | P1 | 6.2 | 13452 | 1 | 1 | 0 | 2 |
| *Synechococcus elongatus* | PCC 6301 | 2.7 | 13282 | 0 | 0 | 0 | 0 |
| *Synechococcus elongatus* | PCC 7942 | 2.74 | 10645 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | BL 107 | 2.28 | 13559 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | CB0101 | 2.69 | 46501 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | CB0205 | 2.43 | 46503 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | CC9311 | 2.61 | 12530 | 2 | 0 | 0 | 2 |
| *Synechococcus* sp. | CC9605 | 2.51 | 13643 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | CC9902 | 2.23 | 13655 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | JA-2-3B | 3.05 | 16252 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | JA-3-3Ab | 2.93 | 16251 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | PCC 6312 | 3.72 | 158717 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Synechococcus* sp. | PCC 7002 | 3.41 | 28247 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | PCC 7335 | 5.97 | 19377 | 0 | 2 | 1 | 3 |
| *Synechococcus* sp. | PCC 7336 | 5.14 | 158719 | 0 | 1 | 1 | 2 |
| *Synechococcus* sp. | PCC 7502 | 3.58 | 159509 | 0 | 0 | 1 | 1 |
| *Synechococcus* sp. | RCC307 | 2.22 | 13654 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | RS9916 | 2.66 | 13557 | 0 | 0 | 0 | 0 |
| *Synechococcus* sp. | RS9917 | 2.58 | 13555 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | WH 5701 | 3.04 | 13554 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | WH 7803 | 2.37 | 13642 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | WH 7805 | 2.62 | 13553 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | WH 8016 | 2.71 | 61805 | 2 | 0 | 0 | 2 |
| *Synechococcus* sp. | WH 8102 | 2.43 | 230 | 1 | 0 | 0 | 1 |
| *Synechococcus* sp. | WH 8109 | 2.12 | 37911 | 0 | 0 | 0 | 0 |
| *Synechocystis* sp. | PCC 6803 | 3.95 | 60 | 0 | 0 | 0 | 0 |
| *Synechocystis* sp. | PCC 7509 | 4.77 | 159501 | 0 | 1 | 2 | 3 |
| *Thermosynechococcus elongatus* | BP-1 | 2.59 | 308 | 0 | 0 | 0 | 0 |
| Unidentified cyanobacterium (symbiont) | UCYN-A | 1.44 | 30917 | 0 | 0 | 0 | 0 |
| *Chroococcidiopsis* sp. | PCC 6712 | 5.7 | 158687 | 6 | 2 | 0 | 8 |
| *Chroococcidiopsis thermalis* | PCC 7203 | 6.69 | 38119 | 0 | 2 | 3 | 5 |
| *Pleurocapsa* sp. | PCC 7319 | 7.39 | 158813 | 5 | 3 | 3 | 11 |
| *Pleurocapsa* sp. | PCC 7327 | 4.99 | 158829 | 5 | 0 | 0 | 5 |
| *Stanieria cyanosphaera* | PCC 7437 | 5.55 | 158877 | 3 | 2 | 0 | 5 |
| *Xenococcus* sp. | PCC 7305 | 5.93 | 159499 | 4 | 2 | 2 | 8 |
| *Arthrospira maxima* | CS-328 | 6 | 29085 | 0 | 0 | 0 | 0 |
| *Arthrospira platensis* | NIES-39 | 6.79 | 42161 | 0 | 0 | 0 | 0 |
| *Arthrospira platensis* | Paraca | 5 | 34793 | 0 | 0 | 0 | 0 |
| *Arthrospira* sp. | PCC 8005 | 6.15 | 40633 | 0 | 0 | 0 | 0 |
| *Coleofasciculus chthonoplastes* | PCC 7420 | 8.68 | 19325 | 1 | 1 | 1 | 3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Crinalium epipsammum* | PCC 9333 | 5.62 | 158835 | 0 | 1 | 3 | 4 |
| *Geitlerinema* sp. | PCC 7105 | 6.15 | 158727 | 1 | 1 | 1 | 3 |
| *Geitlerinema* sp. | PCC 7407 | 4.68 | 158833 | 0 | 0 | 0 | 0 |
| *Leptolyngbya boryana* | PCC 6306 | 7.26 | 158729 | 0 | 1 | 3 | 4 |
| *Leptolyngbya* sp. | PCC 6406 | 5.61 | 159511 | 1 | 1 | 0 | 2 |
| *Leptolyngbya* sp. | PCC 7375 | 9.42 | 43137 | 2 | 1 | 2 | 5 |
| *Leptolyngbya* sp. | PCC 7376 | 5.13 | 43487 | 1 | 0 | 0 | 1 |
| *Lyngbya* sp. | CCY 9616 | 7.04 | 13409 | 0 | 0 | 0 | 0 |
| *Microcoleus* sp. | PCC 7113 | 7.97 | 158839 | 0 | 1 | 2 | 3 |
| *Microcoleus vaginatus* | FGP-2 | 6.7 | 47601 | 0 | 0 | 3 | 3 |
| *Moorea producta* | 3L$^{T}$ | 8.48 | 60895 | 3 | 7 | 4 | 14 |
| *Nodosilinea nodulosa* | PCC 7104 | 6.89 | 62311 | 0 | 0 | 0 | 0 |
| *Oscillatoria acuminata* | PCC 6304 | 7.8 | 158709 | 0 | 2 | 1 | 3 |
| *Oscillatoria formosa* | PCC 6407 | 6.89 | 158733 | 2 | 4 | 3 | 9 |
| *Oscillatoria nigro-viridis* | PCC 7112 | 8.27 | 158711 | 0 | 1 | 1 | 2 |
| *Oscillatoria* sp. | PCC 10802 | 8.59 | 158815 | 0 | 0 | 5 | 5 |
| *Oscillatoria* sp. | PCC 6506 | 6.68 | 49445 | 2 | 4 | 2 | 8 |
| *Prochlorothrix hollandica* | PCC 9006$^{T}$ | 5.65 | 158811 | 0 | 0 | 2 | 2 |
| *Pseudanabaena* sp. | PCC 6802 | 5.62 | 158731 | 0 | 0 | 0 | 0 |
| *Pseudanabaena* sp. | PCC 7367 | 4.89 | 158713 | 0 | 0 | 0 | 0 |
| *Pseudanabaena* sp. | PCC 7429 | 5.48 | 158837 | 1 | 1 | 0 | 2 |
| *Spirulina major* | PCC 6313 | 5.05 | 158715 | 0 | 1 | 0 | 1 |
| *Spirulina subsalsa* | PCC 9445 | 5.32 | 158827 | 0 | 0 | 0 | 0 |
| *Trichodesmium erythraeum* | IMS101 | 7.75 | 318 | 0 | 1 | 1 | 2 |
| *Anabaena cylindrica* | PCC 7122 | 7.06 | 43355 | 2 | 2 | 3 | 7 |
| *Anabaena* sp. | PCC 7108 | 5.89 | 158737 | 3 | 3 | 1 | 7 |
| *Anabaena variabilis* | ATCC 29413 | 7.11 | 10642 | 1 | 5 | 2 | 8 |
| *Calothrix* sp. | PCC 6303 | 6.96 | 158041 | 2 | 1 | 1 | 4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Calothrix* sp. | PCC 7103 | 11.58 | 159495 | 3 | 2 | 3 | 8 |
| *Calothrix* sp. | PCC 7507 | 7.02 | 158683 | 1 | 1 | 1 | 3 |
| *Cylindrospermopsis raciborskii* | CS-505 | 3.88 | 40109 | 2 | 1 | 1 | 4 |
| *Cylindrospermum stagnale* | PCC 7417 | 7.61 | 158809 | 8 | 3 | 4 | 15 |
| *Microchaete* sp. | PCC 7126 | 5.74 | 158817 | 2 | 2 | 1 | 5 |
| *Nodularia spumigena* | CCY 9414 | 5.32 | 13447 | 3 | 2 | 4 | 9 |
| *Nostoc azollae* (endosymbiont) | 0708 | 5.49 | 30807 | 2 | 0 | 1 | 3 |
| *Nostoc punctiforme* | PCC 73102 | 9.06 | 216 | 4 | 7 | 1 | 12 |
| *Nostoc* sp. | PCC 7107 | 6.33 | 158705 | 1 | 1 | 0 | 2 |
| *Nostoc* sp. | PCC 7120 | 7.21 | 244 | 1 | 3 | 0 | 4 |
| *Nostoc* sp. | PCC 7524 | 6.72 | 158707 | 3 | 1 | 0 | 4 |
| *Raphidiopsis brookii* | D9 | 3.19 | 40111 | 2 | 0 | 1 | 3 |
| *Rivularia* sp. | PCC 7116 | 8.73 | 63147 | 2 | 1 | 0 | 3 |
| *Tolypothrix* sp. | PCC 9009 | 8.18 | 63425 | 1 | 2 | 2 | 5 |
| *Fischerella* sp. | JSC-11 | 5.38 | 61093 | 3 | 2 | 1 | 6 |
| *Fischerella* sp. | PCC 9339 | 8.4 | 159505 | 10 | 7 | 5 | 22 |
| *Fischerella* sp. | PCC 9431 | 7.14 | 158821 | 4 | 4 | 0 | 8 |
| *Fischerella* sp. | PCC 9605 | 8.2 | 158819 | 2 | 1 | 2 | 5 |
| *Mastigocladopsis repens* | PCC 10914 | 6.31 | 158735 | 0 | 3 | 2 | 5 |
| Unidentified cyanobacterium | PCC 7702 | 4.9 | 158823 | 2 | 1 | 1 | 4 |
| *Acrophormium turfae* | CCMR0081$^{T}$ | 8.41 | - | 7 | 8 | 3 | 18 |
| *Acrophormium turfae* | CCMR0082 | 8.16 | - | 6 | 6 | 1 | 13 |

# Chapter IV

# MERGING ECOGENOMICS AND TAXONOMY OF CYANOBACTERIA

*Manuscript under revision, Frontiers in Microbiology Journal*

Juline M. Walter[1,2], Felipe H. Coutinho[1,2], Bas E. Dutilh[2,4], Jean Swings[5], Fabiano L. Thompson[1,3*], Cristiane C. Thompson[1*]

[1]Laboratory of Microbiology, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[2]Radboud University Medical Centre, Radboud Institute for Molecular Life Sciences, Centre for Molecular and Biomolecular Informatics (CMBI), Nijmegen, The Netherlands.

[3]Center of Technology - CT2, SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

[4]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands.

[5]Laboratory of Microbiology, Ghent University, Gent, Belgium.

*Corresponding Authors: Av. Carlos Chagas Filho 373. Sala 102, Bloco A, CCS/IB/BIOMAR, Lab. de Microbiologia, Cidade Universitária, Rio de Janeiro, RJ, Brasil, CEP 21941-599, T/F: +552139386567, E-mails: fabianothompson1@gmail.com (FLT); thompsoncristiane@gmail.com (CCT).

**Short/ Running Title:** Cyanobacterial Diversity

**Keywords:** Microbial Ecology / Ecological Niches / Charting Biodiversity / Genome-based Microbial Taxonomy / Metagenome / High-Throughput Sequencing Technology.

## Abstract

Cyanobacteria are major contributors to global biogeochemical cycles. The genetic diversity among Cyanobacteria enables them to thrive across many habitats, although only a few studies have analyzed the association of phylogenomic clades to specific environmental niches. In this study, we adopted an ecogenomics strategy with the aim to delineate ecological niches preferences of Cyanobacteria and link them to the genomic taxonomy of these bacteria. First, an appropriate phylogenomic framework was established using a set of genomic taxonomy signatures (multilocus gene sequence tree, in silico Genome-to-Genome hybridization, GGH/GGD; Average Amino acid Identity, AAI; Average Nucleic acid Identity, ANI) to analyze 100 publicly available cyanobacterial genomes. Next, the relative abundances of these genomes were determined throughout diverse global marine and freshwater ecosystems. This allowed the distinction of three major ecological groups of Cyanobacteria (i. Low Temperature; ii. Low Temperature Copiotroph; and iii. High Temperature Oligotroph) that were linked to the genomic taxonomy. Finally, we determined a high coherence between the phylogenomic approach and the ecogenomic groups.

## Introduction

Earth is home to nearly one trillion (1012) microbial species that have evolved over approximately 4 billion years (Locey and Lennon 2016). Cyanobacteria emerged approximately 3 billion years ago, ushering Earth's transition from anoxygenic to oxygenic conditions through photosynthesis (Shih et al. 2013). Throughout their evolution, Cyanobacteria became one of the most diverse and widely distributed Prokaryotes, occupying many niches within terrestrial, planktonic, and benthic habitats. Their long history evolved in a broad heterogeneity comprising unicellular and multicellular, photosynthetic and non-photosynthetic (i.e. Melainabacteria) (Di Rienzi et al. 2013; Soo et al. 2014), free-living, symbiotic, toxic and predatory organisms, with genomes sizes ranging from 1 to 10 Mb (Shih et al. 2013). Here we consider Cyanobacteria phylum as consisted only of oxygenic phototrophs.

The Cyanobacteria form a challenging group for the microbiologists. Their traditional taxonomy based on morphologic traits does not reflect the results of phylogenetic analyses (Gugger and Hoffmann 2004; Hugenholtz et al. 2016; Schirrmeister et al. 2011; Boone and Castenholz 2001; Rippka et al. 1979). The predominance of morphology assembled unrelated Cyanobacteria into polyphyletic species and genera and higher taxonomic categories which require revisions in the future (Komárek et al. 2014). The 16S rRNA gene sequences were useful in charting and characterizing microbial communities (Kózlov et al. 2016) but this molecule lack sensitivity for evolutionary changes that occur in ecological dynamics, where microbial diversity is organized by physicochemical parameters (Choudoir et al. 2012; Becraft et al. 2015). Hence, the processes that shape cyanobacterial communities over space and time are less known. A recent study proposed that there should be 170 genera of Cyanobacteria based on 16S rRNA sequences only (Kózlov et al. 2016). Farrant et al. (2016) delineated 121 Prochlorococcus and 15 Synechococcus ecologically significant taxonomic units (ESTUs) in the global ocean using single-copy petB sequences

(encoding cytochrome b6) and environmental cues. Another hurdle in the study of this group of bacteria is the cultivation in the laboratory.

High Throughput Sequencing (HTS) have revolutionized the practice of microbial systematics, providing an informative, reproducible, and portable tool to delineate species, reconstruct their evolutionary history, and infer ecogenomic features (Hugenholtz et al. 2016; Gevers et al. 2005; Konstantinidis and Tiedje 2005; Gribaldo and Brochier-Armanet 2012; Sutcliffe et al. 2013; Garrity and Oren 2012; Shih et al. 2013). This approach allows both cultured (Al-Saari et al. 2015; Appolinario et al. 2016) and uncultured microorganisms (Iverson et al. 2012; Hugerth et al. 2015; Brown et al. 2015) to be studied. A genomic taxonomy approach has successfully been applied to elucidate the taxonomic structure of the cyanobacterial genera Prochlorococcus and Synechococcus (Coutinho et al. 2015; Coutinho et al. 2016; Thompson et al. 2013). The availability of whole-genomes opened the doors for an in-depth knowledge in microbial diversity and ecology, where the entire genomic pool may be applied to understanding the forces that govern community structure. The use of ecogenomic analyses postulates a reliable and scalable approach to delineate species and genera in order to reconstruct their evolution and to draw a global picture of possible ecological determinants (Garrity 2016; Hugenholtz et al. 2016; Thompson et al. 2015; Di Rienzi et al. 2013; Soo et al. 2014; Anantharaman et al. 2016; Spang et al. 2015; Hug et al. 2016). Our hypothesis is that a phylogenomic framework will reflect the ecologic groups found in nature.

To test this hypothesis, we first established a phylogenomic framework, using genomic signatures (i.e. multilocus gene sequence tree, average amino acid identity, and Genome-to-Genome distance), with the circumscription of species and genera. We then classified the genomes in three major groups according to their ecological traits as inferred through metagenomics and environmental metadata. Finally, we correlated the three disclosed ecogenomic groups (i. Low Temperature; ii. Low Temperature Copiotroph; and iii. High Temperature Oligotroph) with the circumscribed species and

genera. We observed that the taxonomic delineation of species and genera is coherent with the ecogenomic groups.

## Material and Methods

### Genome selection

A total of 100 complete genomes publicly available as of January 2016 were retrieved from RefSeq (NCBI Reference Sequence Database), GenBank and GEBA (Genomic Encyclopedia of Bacteria and Archaea) databases for analysis (Table 1). Completeness of all genomes selected was assessed with CheckM (Parks et al. 2015), and the genomes that were at least 90% complete and assembled in < 500 contigs were used for further analysis. All genomes used in this paper are listed in Table 1.

**Table 1. Details of all cyanobacterial genomes included in this study.**

Ecological and molecular features were indicated, such as environment sampling, as well as number of contigs, genome size, GC % content, completeness score, and carboxysome type. The following classifications for detailed for comparison: NCBI (order and family), numeral identification and genera according to Kózlov et al. (2016), and identification based in curated database CyanoType v.1 (Ramos et al. 2017). Overwritten T indicates type strain. [The table was splited in two for better visualization].

**Part 1:**

| NCBI Order Classification [§] | NCBI Family Classification [§] | ID Classification (Kózlov et al., 2016) [¢] | Genus Classification (Kózlov et al., 2016) [¢] | CyanoType v.1 [ᵠ] | Bacterial Strain [1] | Strain [1] | NCBI or JGI Reference Sequence |
|---|---|---|---|---|---|---|---|
| Nostocales | Nostocaceae | Family2;Genus133 | NA | *Anabaena cylindrica* | *Anabaena cylindrica* | PCC 7122 [T] | NC_019771.1 |
| Nostocales | Nostocaceae | Family2;Genus128 | NA | *Anabaena sp.* | *Anabaena sp.* | PCC 7108 | NZ_AJWF00000000.1 |
| Oscillatoriales | x | NA | NA | *Arthrospira platensis* | *Arthrospira platensis* | C1 [2] | NZ_CM001632.1 |
| Oscillatoriales | x | Family1;Genus90 | *Arthrospira* | *Arthrospira platensis* | *Arthrospira platensis* | NIES-39 | NC_016640.1 |
| Oscillatoriales | x | NA | NA | *Arthrospira platensis* | *Arthrospira platensis* | Paraca | NZ_ACSK00000000.3 |
| Oscillatoriales | x | NA | NA | *Arthrospira sp.* | *Arthrospira sp.* | PCC 8005 | NZ_FO818640.1 |
| Nostocales | Rivulariaceae | NA | NA | *Calotrix sp.* | *Calotrix sp.* | PCC 7103 | NZ_ALVJ00000000.1 |
| Chroococcales | x | NA | NA | *Chamaesiphon minutus* | *Chamaesiphon minutus* | PCC 6605 | NC_019697.1 |
| Pleurocapsales | x | Family1;Genus111 | *Chroococcidiopsis* | *Chroococcidiopsis thermalis* | *Chroococcidiopsis thermalis* | PCC 7203 [T] | NC_019695.1 |
| Oscillatoriales | Phormidiaceae | Family1;Genus100 | *Coleofasciculus* | *Coleofasciculus chthonoplastes* | *Coleofasciculus chthonoplastes* | PCC 7420 [3 T] | NZ_ABRS00000000.1 |
| Oscillatoriales | x | NA | NA | *Crinalium epipsammum* | *Crinalium epipsammum* | PCC 9333 | NC_019753.1 |
| Chroococcales | x | NA | NA | filamentous cyanobacterium | *Cyanobacterium* | ESFC-1 | NZ_ARCP00000000.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Chroococcales | x | NA | NA | unidentified Oscillatoriales | *Cyanobacterium* | JSC-12 | NZ_CM001633.1 |
| Chroococcales | x | Family1;Genus76 | *Cyanobacterium* | *Cyanobacterium stanieri* | *Cyanobacterium stanieri* | PCC 7202 [T] | CP003940.1 |
| Nostocales | Nostocaceae | Family2;Genus122 | *Cylindrospermum* | *Cylindrospermum stagnale* | *Cylindrospermum stagnale* | PCC 7417 [T] | NC_019757.1 |
| Chroococcales | x | Family1;Genus57 | *Dactylococcopsis* | *Dactylococcopsis salina* | *Dactylococcopsis salina* | PCC 8305 [T] | NC_019780.1 |
| Stigonematales | x | NA | NA | *Fischerella sp.* | *Fischerella sp.* | JSC-11 | NZ_AGIZ00000000.1 |
| Stigonematales | x | NA | NA | *Fischerella sp.* | *Fischerella sp.* | PCC 9339 | NZ_ALVS00000000.1 |
| Stigonematales | x | NA | NA | *Fischerella sp.* | *Fischerella sp.* | PCC 9431 | ALVX00000000.1 |
| Stigonematales | x | NA | NA | *Fischerella sp.* | *Fischerella sp.* | PCC 9605 | NZ_ALVT00000000.1 |
| Oscillatoriales | x | Family1;Genus59 | *Lyngbya* | *Geitlerinema sp.* | *Geitlerinema sp.* | PCC 7105 | NZ_ANFQ00000000.1 |
| Oscillatoriales | x | Family1;Genus23 | NA | *Geitlerinema sp.* | *Geitlerinema sp.* | PCC 7407 | NC_019703.1 |
| Chroococcales | x | Family1;Genus76 | *Cyanobacterium* | *Geminocystis herdmanii* | *Geminocystis herdmanii* | PCC 6308 [T] | NZ_ALVO00000000.1 |
| Chroococcales | x | Family1;Genus114 | *Chroogloeocystis* | *Gloeocapsa sp.* | *Gloeocapsa sp.* | PCC 7428 | NC_019745.1 |
| Chroococcales | x | Family1;Genus78 | *Gleocapsa* | *Gloeocapsa sp.* | *Gloeocapsa sp.* | PCC 73106 | NZ_ALVY00000000.1 |
| Chroococcales | x | Family1;Genus57 | *Euhalothece* | *Halothece sp.* | *Halothece sp.* | PCC 7418 | NC_019779.1 |
| Oscillatoriales | x | NA | NA | *Leptolyngbya boryana* | *Leptolyngbya boryana* | PCC 6306 [T] | NZ_ALVM00000000.1 |
| Oscillatoriales | x | NA | NA | *Nodosilinea nodulosa* PCC7104 | *Leptolyngbya sp.* | PCC 7104 [4] | NZ_ALVP00000000.1 |
| Oscillatoriales | x | Family1;Genus14 | *Leptolyngbya* | *Leptolyngbya ectocarpi PCC7375* | *Leptolyngbya sp.* | PCC 7375 | NZ_ALVN00000000.1 |
| Oscillatoriales | x | Family1;Genus65 | *Oscillatoria* | *Leptolyngbya sp.* | *Leptolyngbya sp.* | PCC 7376 | NC_019683.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Oscillatoriales | x | NA | NA | *Leptolyngbya sp.* | *Leptolyngbya sp.* | PCC 6406 | NZ_ALVV00000000.2 |
| Oscillatoriales | x | NA | NA | NA | *Lyngbya aestuarii* | BL-J | NZ_AUZM00000000.1 |
| Oscillatoriales | x | NA | NA | *Lyngbya confervoides* | *Lyngbya confervoides* | BDU | NZ_JTHE00000000.1 |
| Oscillatoriales | x | NA | NA | *Lyngbya sp.* | *Lyngbya sp.* | PCC 8106 [5] | NZ_AAVU00000000.1 |
| Oscillatoriales | x | NA | NA | *Microcoleus sp.* | *Microcoleus sp.* | PCC 7113 | NC_019738.1 |
| Oscillatoriales | x | NA | NA | *Microcoleus vaginatus* | *Microcoleus vaginatus* | FGP-2 | NZ_AFJC00000000.1 |
| x | x | NA | NA | *Moorea producens* | *Moorea producens* | 3L [6 T] | NZ_AEPQ00000000.1 |
| Nostocales | Nostocaceae | Family2;Genus120 | NA | *Nostoc sp.* | *Nostoc sp.* | PCC 7107 | NC_019676.1 |
| Nostocales | Nostocaceae | NA | NA | *Nostoc sp.* | *Nostoc sp.* | PCC 7524 | NC_019684.1 |
| Oscillatoriales | x | NA | NA | *Oscillatoria acuminata* | *Oscillatoria acuminata* | PCC 6304 [T] | NC_019693.1 |
| Oscillatoriales | x | Family1;Genus94 | *Microcoleus* | *Oscillatoria nigro-viridis* | *Oscillatoria nigroviridis* | PCC 7112 | NC_019729.1 |
| Oscillatoriales | x | NA | NA | *Oscillatoria princeps* | *Oscillatoria sp.* | PCC 10802 | NZ_ANKO00000000.1 |
| Oscillatoriales | x | NA | NA | *Kamptonema* sp. | *Oscillatoria sp.* | PCC 6506 | NZ_CACA00000000.1 |
| Oscillatoriales | x | NA | NA | *Kamptonema formosum* | *Oscillatoria sp.* | PCC 6407 [7] | NZ_ALVI00000000.1 |
| x | x | Family1;Genus1 | *Synechococcus* | NA | *Parasynechococcus africanus* | CC9605 [T] • | NC_007516 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus chillensis* | CC9902 [T] • | NC_007513 |
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus marearabicus* | WH8109 [T] • | ACNY00000000.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus marenigrum* | WH8102 [T] • | NC_005070.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus nordiatlanticus* | BL107 $^{T}$ • | NZ_DS022298.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus benguelii* | CC9311 $^{T}$ • | NC_008319.1 |
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus equatorialis* | RS9917 $^{T}$ • | NZ_CH724158.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus gyrus* | RS9916 $^{T}$ • | NZ_DS022299.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus pacificus* | WH7803 $^{T}$ • | NC_009481 |
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus subtropicalis* | WH7805 $^{T}$ • | NZ_CH724168.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus sudipacificus* | WH8016 $^{T}$ • | AGIK00000000.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus antarcticus* | WH5701 $^{T}$ • | NZ_CH724159–NZ_CH724167 |
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus indicus* | CB0205 • | NZ_ADXM00000000.1 |
| x | x | NA | NA | *Synechococcus sp.* | *Parasynechococcus sudiatlanticus* | CB0101 $^{T}$ • | NZ_ADXL00000000.1 |
| x | x | Family1;Genus1 | *Synechococcus* | *Synechococcus sp.* | *Parasynechococcus mediterranei* | RCC307 $^{T}$ • | NC_009482.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix agardhii* | *Planktothrix agardhii* | NIVA-CYA 126/8 | NZ_CM002803.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix agardhii* | *Planktothrix agardhii* | NIVA-CYA 15 | NZ_AVFS00000000.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix agardhii* | *Planktothrix agardhii* | NIVA-CYA 56/3 | NZ_AVFY00000000.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix mougeotii* | *Planktothrix mougeotii* | NIVA-CYA 405 | NZ_AVFU00000000.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Oscillatoriales | x | NA | NA | *Planktothrix prolifica* | *Planktothrix prolifica* | NIVA-CYA 406 | NZ_AVFV00000000.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix prolifica* | *Planktothrix prolifica* | NIVA-CYA 540 | NZ_AVFX00000000.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix rubescens* | *Planktothrix prolifica* | NIVA-CYA 98 | NZ_AVFZ00000000.1 |
| Oscillatoriales | x | NA | NA | *Planktothrix rubescens* | *Planktothrix rubescens* | NIVA-CYA 407 | NZ_AVFW00000000.1 |
| Pleurocapsales | x | Family1;Genus71 | *Pleurocapsa* | *Pleurocapsa sp.* | *Pleurocapsa sp.* | PCC 7319 | NC_019689.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus chisholmii* | AS9601 $^{T}$ ° | NC_008816.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus marinus* | CCMP1986 | NC_005072.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus neptunius* | MIT9312 $^{T}$ ° | NC_007577.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | NA | NA | *Prochlorococcus marinus* | *Prochlorococcus nereus* | MIT9202 $^{T}$ ° | NZ_ACDW00000000.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus nereus* | MIT9215 ° | NC_009840.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus ponticus* | MIT9301 $^{T}$ ° | NC_009091.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus tetisii* | MIT9515 $^{T}$ ° | NC_008817.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus proteus* | NATL1A ° | NC_008819.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | NA | NA | *Prochlorococcus marinus* | *Prochlorococcus proteus* | NATL2A $^{T}$ ° | NC_007335.2 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus marinus* | CCMP1375 $^{T}$ | NC_005042.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Prochlorococcus* | *Prochlorococcus marinus* | *Prochlorococcus ceticus* | MIT9211 [T] | NC_009976.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Synechococcus* | *Prochlorococcus marinus* | *Prochlorococcus swingsii* | MIT9303 ° | NC_008820.1 |
| Prochlorales / Prochlorophytes | Prochlorococcaceae | Family1;Genus1 | *Synechococcus* | *Prochlorococcus marinus* | *Prochlorococcus swingsii* | MIT9313 [T] ° | NC_005071.1 |
| Oscillatoriales | x | NA | NA | *Pseudanabaena biceps* | *Pseudanabaena biceps* | PCC 7429 | NZ_ALWB00000000.1 |
| Oscillatoriales | x | Family3;Genus166 | *Pseudanabaena* | *Pseudanabaena sp.* | *Pseudanabaena sp.* | PCC 7367 | NC_019701.1 |
| Oscillatoriales | x | Family3-1;Genus169 | *Phormidium* | *Pseudanabaena sp.* | *Pseudoanabaena sp.* | PCC 6802 | ALVK00000000.1 |
| Nostocales | Rivulariaceae | Family2;Genus150 | *Rivularia* | *Rivularia sp.* | *Rivularia sp.* | PCC 7116 | NC_019678.1 |
| Oscillatoriales | x | NA | NA | *Spirulina subsalsa* | *Spirulina subsalsa* | PCC 9445 | NZ_ALVR00000000.1 |
| Pleurocapsales | x | Family1;Genus75 | *Stanieria* | *Stanieria cyanosphaera* | *Stanieria cyanosphaera* | PCC 7437 [T] | NC_019748.1 |
| Chroococcales | x | Family1;Genus3 | *Synechococcus* | *Synechococcus elongatus* | *Synechococcus elongatus* | PCC 6301 [T] | NC_006576.1 |
| Chroococcales | x | Family1;Genus3 | *Synechococcus* | *Synechococcus elongatus* | *Synechococcus elongatus* | PCC 7942 | NC_007604.1 |
| Chroococcales | x | NA | NA | NA | *Synechococcus springii* | JA23Ba213 [T] • | NC_007776 |
| Chroococcales | x | NA | NA | NA | *Synechococcus yellowstonii* | JA33Ab [T] • | NC_007775.1 |
| Chroococcales | x | Family1;Genus44 | *Thermosynechococcus* | *Synechococcus* sp. | *Synechococcus californii* | PCC 6312 [T] • | NC_019680.1 |
| Chroococcales | x | Family1;Genus65 | *Synechococcus* | *Synechococcus* sp. | *Synechococcus euryhalinus* | PCC 7002 [T] • | NC_010475.1 |
| Chroococcales | x | NA | NA | *Synechococcus* sp. | *Synechococcus* | PCC 7335 [T] • | ABRV00000000.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | *mexicanus* | | |
| Chroococcales | x | NA | NA | *Synechococcus* sp. | *Synechococcus berkleyi* | PCC 7336 [T] • | ALWC00000000.1 |
| Chroococcales | x | NA | NA | *Synechococcus* sp. | *Synechococcus bogii* | PCC 7502 [T] • | CP003594.1 |
| Chroococcales | x | NA | NA | *Synechocystis sp.* | *Synechocystis sp.* | PCC 7509 | ALVU00000000.2 |
| Oscillatoriales | x | Family1;Genus96 | *Trichodesmium* | *Trichodesmium erythraeum* | *Trichodesmium erythraeum* | IMS101 [T] | NC_008312.1 |
| Pleurocapsales | x | Family1;Genus75 | *Pleurocapsa* | *Xenococcus sp.* | *Xenococcus sp.* | PCC 7305 | NZ_ALVZ00000000.1 |
| Gloeobacterales | x | NA | NA | *Gloeobacter violaceum* | *Gloeobacter violaceum* | PCC 7421 [8] | NC_005125.1 |

## Part 2:

| Bacterial Strain [1] | Strain [1] | New Genus Proposal | New Species Proposal | Habitat | Type Source / Place | # Contigs [Δ] | Lenght (Mbp) [Δ] | % mol GC [Δ] | # CDS | Completeness * | Carboxysome |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Anabaena cylindrica* | PCC 7122 [T] | | | Freshwater | Cambridge, UK | 7 | 7.06 | 38.79 | 6182 | 99.44 | β |
| *Anabaena sp.* | PCC 7108 | | *A. mossi* | Marine (coastal) | Intertidal zone, Moss Beach, CA, USA | 3 | 5.9 | 38.78 | 5169 | 99.63 | β |
| *Arthrospira platensis* | C1 [2] | | *A. sesilensis* | Freshwater | Alkaline salt lakes | 63 | 6 | 44.69 | 4852 | 99.71 | β |
| *Arthrospira platensis* | NIES-39 | | | Freshwater | Alkaline salt lakes | 1 | 6.78 | 44.27 | 6676 | 99.13 | β |
| *Arthrospira platensis* | Paraca | | | Freshwater | Alkaline salt lakes | 239 | 6.49 | 44.31 | 5436 | 99.34 | β |
| *Arthrospira sp.* | PCC 8005 | | *A. nitrilium* | Unknown | Unknown | 119 | 6.27 | 44.7 | 5171 | 99.93 | β |
| *Calotrix sp.* | PCC 7103 | | *C. wisconsii* | Freshwater | Crawford Co, Wisconsin, USA | 12 | 11.58 | 38.55 | 9371 | 99.39 | β |
| *Chamaesiphon minutus* | PCC 6605 | | | Freshwater | Berkeley, CA, USA | 1 | 6.28 | 45.73 | 5956 | 99.48 | β |
| *Chroococcidiopsis thermalis* | PCC 7203 [T] | | | Soil | Greifswald, Germany | 3 | 6.68 | 44.47 | 5618 | 99.63 | β |
| *Coleofasciculus chthonoplastes* | PCC 7420 [3] [T] | | | Marine (coastal) | Salt marsh in Woods Hole, Massachusetts, USA | 142 | 8.65 | 45.43 | 7100 | 99.37 | β |
| *Crinalium epipsammum* | PCC 9333 | | | NA | NA | 1 | 5.31 | 40.16 | 5002 | 99.48 | β |
| *Cyanobacterium* | ESFC-1 | *Cyclospexia* | *C. valenium* | Marine (coastal) | Extremophylic mat communities, Elkhorn Slough estuary, CA, USA | 52 | 5.62 | 46.51 | 4857 | 99.59 | β |
| *Cyanobacterium* | JSC-12 | *Tapinonema* | *T. colecalium* | Freshwater | NA | 20 | 5.52 | 47.49 | 5024 | 99.29 | β |
| *Cyanobacterium stanieri* | PCC 7202 [T] | *Geminocystis* | *G. stanieri* | Freshwater | Thermal springs, alkaline pod | 1 | 3.16 | 38.66 | 2886 | 99.52 | β |
| *Cylindrospermum stagnale* | PCC 7417 [T] | *Anabaena* | *A. stagnale* | Soil | Stockholm, Sweden | 4 | 7.61 | 42.2 | 6127 | 99.78 | β |
| *Dactylococcopsis salina* | PCC 8305 [T] | | | Freshwater | Solar Lake, Israel | 1 | 3.78 | 42.44 | 3412 | 99.55 | β |
| *Fischerella sp.* | JSC-11 | | *F. sesquitii* | NA | NA | 34 | 5.38 | 41.05 | 4627 | 99.76 | β |
| *Fischerella sp.* | PCC 9339 | | *F. hapalii* | NA | NA | 13 | 8 | 40.16 | 6720 | 99.76 | β |
| *Fischerella sp.* | PCC 9431 | | *F. welwii* | NA | NA | 8 | 7.16 | 40.19 | 6104 | 99.76 | β |
| *Fischerella sp.* | PCC 9605 | | *F. peptidasii* | Soil | Limestone, Jerucham, Har Rahama, Israel | 12 | 8.08 | 42.61 | 7060 | 100 | β |
| *Geitlerinema sp.* | PCC 7105 | | *G. catellasis* | NA | USA | 8 | 6.15 | 51.59 | 4735 | 93.75 | β |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Geitlerinema sp.* | PCC 7407 | *Pseudogeitlerinema* | *P. shalloid* | Unknow | Unknow | 1 | 4.68 | 58.46 | 3727 | 99.87 | β |
| *Geminocystis herdmanii* | PCC 6308 [T] | | | Freshwater | Lake near Madison, Wisconsin, USA | 1 | 4.26 | 34.28 | 3887 | 99.78 | β |
| *Gloeocapsa sp.* | PCC 7428 | *Rotundosa* | *R. thermolimnetic* | Thermal - Freshwater | Moderate hot spring | 1 | 5.43 | 43.27 | 5254 | 99.78 | β |
| *Gloeocapsa sp.* | PCC 73106 | | *G. sphagnus* | Freshwater | Sphagnum bog, Switzerland | 228 | 4.025 | 41.11 | 3704 | 98.84 | β |
| *Halothece sp.* | PCC 7418 | *Dactylococcopsis* | *D. halotolerans* | Freshwater | Solar Lake, Israel | 1 | 4.18 | 42.92 | 3663 | 99.48 | β |
| *Leptolyngbya boryana* | PCC 6306 [T] | | | Freshwater | Lake near Madison, Wisconsin, USA | 5 | 7.26 | 47.02 | 6827 | 99.41 | β |
| *Leptolyngbya sp.* | PCC 7104 [4] | *Allonema* | *A. longislandicus* | Marine (coastal) | Rock at shoreline, Montauk Point, Long Island, NY, USA | 2 | 6.89 | 57.69 | 6414 | 99.18 | β |
| *Leptolyngbya sp.* | PCC 7375 | *Acrophormium* | *A. splendidus* | Marine (coastal) | Plankton, Woods Hole, Massachusetts, USA | 5 | 9.42 | 47.62 | 8366 | 99.73 | β |
| *Leptolyngbya sp.* | PCC 7376 | *Symphothece* | *S. fragile* | Marine (coastal) | Limestone, Crystal Cave, Bermuda | 1 | 5.12 | 43.87 | 4601 | 99.42 | β |
| *Leptolyngbya sp.* | PCC 6406 | *Euryforis* | *E. eilemai* | Freshwater | California, USA | 3 | 5.77 | 55.18 | 5156 | 98.64 | β |
| *Lyngbya aestuarii* | BL-J | | | NA | NA | 432 | 6.87 | 41.16 | 5597 | 99.74 | β |
| *Lyngbya confervoides* | BDU | *Euryforis* | *E. confervoides* | NA | NA | 298 | 8.79 | 55.63 | 8370 | 99.34 | β |
| *Lyngbya sp.* | PCC 8106 [5] | | *L. limosa* | Marine (coastal) | NA | 110 | 7.03 | 41.11 | 5854 | 99.3 | β |
| *Microcoleus sp.* | PCC 7113 | | *M. franciscanus* | Soil | San Francisco, California, USA | 1 | 7.47 | 46.21 | 6734 | 99.56 | β |
| *Microcoleus vaginatus* | FGP-2 | *Toxinema* | *T. vaginatus* | Soil | Canyonlands National Park, UT, USA | 40 | 6.69 | 46.04 | 5519 | 99.67 | β |
| *Moorea producens* | 3L [6 T] | | | NA | NA | 287 | 8.38 | 43.68 | 6979 | 98.56 | β |
| *Nostoc sp.* | PCC 7107 | *Anabaena* | *A. reyesii* | Freshwater | Point Reyes Peninsula, California, USA | 1 | 6.32 | 40.36 | 5200 | 99.26 | β |
| *Nostoc sp.* | PCC 7524 | *Anabaena* | *A. nostocii* | Freshwater | Hot spring, Amparai District, Maha Oya, Sri Lanka | 3 | 6.71 | 41.53 | 5326 | 99.33 | β |
| *Oscillatoria acuminata* | PCC 6304 [T] | | | Soil | NA | 1 | 7.68 | 47.6 | 6004 | 99.71 | β |
| *Oscillatoria nigroviridis* | PCC 7112 | *Toxinema* | *T. nigroviridis* | Soil | USA | 1 | 7.47 | 45.87 | 6925 | 99.78 | β |
| *Oscillatoria sp.* | PCC 10802 | *Somacatellium* | *S. hydroxylic* | NA | NA | 9 | 8.59 | 54.1 | 7012 | 100 | β |
| *Oscillatoria sp.* | PCC 6506 | *Toxinema* | *T. oscillati* | NA | NA | 377 | 6.67 | 43.4 | 6007 | 99.12 | β |
| *Oscillatoria sp.* | PCC 6407 [7] | *Toxinema* | *T. oscillati* | Freshwater | NA | 12 | 6.89 | 43.43 | 5693 | 99.56 | β |
| *Parasynechococcus africanus* | CC9605 [T] • | | | Marine | California current, Pacific, oligotrophic, 51 m | 1 | 2.51 | 59.2 | 2583 | 99.73 | α |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Parasynechococcus chillensis* | CC9902 $^{T}$ • | | | Marine | California current, Pacific, oligotrophic, 5 m | 1 | 2.23 | 54.2 | 2289 | 99.46 | α |
| *Parasynechococcus marearabicus* | WH8109 $^{T}$ • | | | Marine | Sargasso Sea | 1 | 2.12 | 60.1 | 2661 | 99.32 | α |
| *Parasynechococcus marenigrum* | WH8102 $^{T}$ • | | | Marine | Sargasso Sea | 1 | 2.43 | 59.4 | 2461 | 99.46 | α |
| *Parasynechococcus nordiatlanticus* | BL107 $^{T}$ • | | | Marine | Blanes Bay, Mediterranean Sea, 1,800 m | 1 | 2.29 | 54.2 | 2322 | 99.46 | α |
| *Parasynechococcus benguelii* | CC9311 $^{T}$ • | *Pseudosynechococcus* | *P. benguelii* | Marine | California current, Pacific, coastal, 95 m | 1 | 2.61 | 52.4 | 2627 | 99.73 | α |
| *Parasynechococcus equatorialis* | RS9917 $^{T}$ • | *Pseudosynechococcus* | *P. equatorialis* | Marine | Gulf of Aqaba, Red Sea, 10 m | 1 | 2.58 | 64.4 | 2575 | 99.46 | α |
| *Parasynechococcus gyrus* | RS9916 $^{T}$ • | *Pseudosynechococcus* | *P. gyrus* | Marine | Gulf of Aqaba, Red Sea, 10 m | 1 | 2.66 | 59.8 | 2603 | 99.73 | α |
| *Parasynechococcus pacificus* | WH7803 $^{T}$ • | *Pseudosynechococcus* | *P. pacificus* | Marine | Sargasso Sea, 25 m | 1 | 2.37 | 60.2 | 2439 | 99.18 | α |
| *Parasynechococcus subtropicalis* | WH7805 $^{T}$ • | *Pseudosynechococcus* | *P. subtropicalis* | Marine | Sargasso Sea | 3 | 2.63 | 57.6 | 2595 | 99.73 | α |
| *Parasynechococcus sudipacificus* | WH8016 $^{T}$ • | *Pseudosynechococcus* | *P. sudipacificus* | Marine | Woods Hole, MA, USA | 16 | 2.69 | 54.1 | 2990 | 99.18 | α |
| *Parasynechococcus antarcticus* | WH5701 $^{T}$ • | *Regnicoccus* | *R. antarcticus* | Marine | Long Island Sound, Connecticut, USA | 116 | 3.28 | 65.4 | 2917 | 99.46 | α |
| *Parasynechococcus indicus* | CB0205 • | *Magnicoccus* | *M. indicus* | Marine | Chesapeake Bay, Baltimore, Maryland, USA | 78 | 2.43 | 63 | 2473 | 99.18 | α |
| *Parasynechococcus sudiatlanticus* | CB0101 $^{T}$ • | *Magnicoccus* | *M. sudiatlanticus* | Marine | Chesapeake Bay, Baltimore, Maryland, USA | 94 | 2.69 | 64.2 | 2757 | 99.73 | α |
| *Parasynechococcus mediterranei* | RCC307 $^{T}$ • | *Inmanicoccus* | *I. mediterranei* | Marine | Mediterranean Sea, 15 m | 1 | 2.22 | 60.8 | 2348 | 99.64 | α |
| *Planktothrix agardhii* | NIVA-CYA 126/8 | | *P. stereotis* | Freshwater | NA | 13 | 5.04 | 39.57 | 4188 | 100 | β |
| *Planktothrix agardhii* | NIVA-CYA 15 | | | Freshwater | NA | 238 | 5.38 | 39.48 | 4606 | 100 | β |
| *Planktothrix agardhii* | NIVA-CYA 56/3 | | | Freshwater | NA | 185 | 5.48 | 39.48 | 4674 | 99.78 | β |
| *Planktothrix mougeotii* | NIVA-CYA 405 | | *P. agardhii* | Freshwater | NA | 240 | 5.46 | 39.47 | 4697 | 99.56 | β |
| *Planktothrix prolifica* | NIVA-CYA 406 | | *P. agardhii* | Freshwater | NA | 375 | 5.62 | 39.51 | 4873 | 100 | β |
| *Planktothrix prolifica* | NIVA-CYA 540 | | *P. agardhii* | Freshwater | NA | 157 | 5.5 | 39.48 | 4710 | 99.78 | β |
| *Planktothrix prolifica* | NIVA-CYA 98 | | *P. agardhii* | Freshwater | NA | 346 | 5.61 | 39.52 | 4862 | 99.78 | β |
| *Planktothrix rubescens* | NIVA-CYA 407 | | *P. agardhii* | Freshwater | NA | 219 | 5.39 | 39.46 | 4658 | 100 | β |
| *Pleurocapsa sp.* | PCC 7319 | | *P. penascus* | Marine (coastal) | Arizona Station, Gulf of California, Puerto Penasco, | 10 | 7.38 | 38.74 | 4516 | 99.56 | β |

| | | | | | Mexico | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Prochlorococcus chisholmii* | AS9601 [T] ° | *Eurycolium* | *E. chisholmii* | Marine | Arabian Sea, 50 m | 1 | 1.66 | 31.32 | 1769 | 99.64 | α |
| *Prochlorococcus marinus* | CCMP1986 | *Eurycolium* | *E. marinus* | Marine | Mediterranean Sea, 5 m | 1 | 1.65 | 30.8 | 1777 | 99.46 | α |
| *Prochlorococcus neptunius* | MIT9312 [T] ° | *Eurycolium* | *E. neptunius* | Marine | Gulf Stream, 135 m | 1 | 1.7 | 31.21 | 1815 | 99.73 | α |
| *Prochlorococcus nereus* | MIT9202 [T] ° | *Eurycolium* | *E. nereus* | Marine | South Pacific, 79 m | 1 | 1.69 | 31.1 | 1795 | 98.78 | α |
| *Prochlorococcus nereus* | MIT9215 ° | *Eurycolium* | *E. nereus* | Marine | Equatorial Pacific, surface | 1 | 1.73 | 31.15 | 1840 | 99.73 | α |
| *Prochlorococcus ponticus* | MIT9301 [T] ° | *Eurycolium* | *E. ponticus* | Marine | Sargasso Sea, 90 m | 1 | 1.64 | 31.34 | 1774 | 99.46 | α |
| *Prochlorococcus tetisii* | MIT9515 [T] ° | *Eurycolium* | *E. tetisii* | Marine | Equatorial Pacific, 15 m | 1 | 1.7 | 30.79 | 1784 | 100 | α |
| *Prochlorococcus proteus* | NATL1A ° | *Prolificoccus* | *P. proteus* | Marine | Northern Atlantic, 30 m | 1 | 1.86 | 34.98 | 2204 | 99.73 | α |
| *Prochlorococcus proteus* | NATL2A [T] ° | *Prolificoccus* | *P. proteus* | Marine | Northern Atlantic, 10 m | 1 | 1.84 | 35.12 | 1930 | 99.45 | α |
| *Prochlorococcus marinus* | CCMP1375 [T] | | | Marine | Sargasso Sea, 120 m | 1 | 1.75 | 36.44 | 1883 | 100 | α |
| *Prochlorococcus ceticus* | MIT9211 [T] | *Expansiococcus* | *E. ceticus* | Marine | Equatorial Pacific, 83 m | 1 | 1.68 | 38.01 | 1748 | 99.73 | α |
| *Prochlorococcus swingsii* | MIT9303 ° | *Parasynechococcus* | *P. swingsii* | Marine | Sargasso Sea, 100 m | 1 | 2.68 | 50.01 | 2504 | 100 | α |
| *Prochlorococcus swingsii* | MIT9313 [T] ° | *Parasynechococcus* | *P. swingsii* | Marine | Gulf Stream, 135 m | 1 | 2.41 | 50.74 | 2339 | 99.46 | α |
| *Pseudanabaena biceps* | PCC 7429 | | | Freshwater | NA | 464 | 5.47 | 43.18 | 4774 | 99.29 | β |
| *Pseudanabaena sp.* | PCC 7367 | *Leptocystis* | *L. gracile* | Marine (coastal) | Intertidal zone, Mexico | 1 | 4.55 | 46.31 | 3960 | 98.23 | β |
| *Pseudoanabaena sp.* | PCC 6802 | *Paraleptovivax* | *P. allomegium* | Freshwater | California, USA | 6 | 5.62 | 47.83 | 5363 | 99.76 | β |
| *Rivularia sp.* | PCC 7116 | | *R. bajacalifornii* | Marine (coastal) | La Paz, Baja California Sur, Mexico | 3 | 8.72 | 37.53 | 6612 | 99.78 | β |
| *Spirulina subsalsa* | PCC 9445 | *Paraspirulina* | *P. subsalsa* | NA | NA | 10 | 5.32 | 47.39 | 4580 | 99.56 | β |
| *Stanieria cyanosphaera* | PCC 7437 [T] | | | Freshwater | Havana, Cuba | 6 | 5.54 | 36.22 | 4895 | 99.56 | β |
| *Synechococcus elongatus* | PCC 6301 [T] | | | Freshwater | NA | 1 | 2.7 | 55.5 | 2576 | 99.73 | β |
| *Synechococcus elongatus* | PCC 7942 | | | Freshwater | NA | 2 | 2.74 | 55.46 | 2655 | 100 | β |
| *Synechococcus springii* | JA23Ba213 [T] • | *Leptococcus* | *L. springii* | Thermal - Freshwater | Octopus Spring, Yellowstone Park, USA | 1 | 3.05 | 58.5 | 3064 | 100 | β |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Synechococcus yellowstonii* | JA33Ab [T] • | *Leptococcus* | *L. yellowstonii* | Thermal - Freshwater | Octopus Spring, Yellowstone Park, USA | 1 | 2.93 | 60.2 | 3036 | 100 | β |
| *Synechococcus californii* | PCC 6312 [T] • | *Stenotopis* | *S. californii* | Freshwater | California, USA | 2 | 3.72 | 48.49 | 3795 | 99.29 | β |
| *Synechococcus euryhalinus* | PCC 7002 [T] • | *Symphothece* | *S. euryhalinus* | Unknown | Unknown | 7 | 3.41 | 49.16 | 3121 | 100 | β |
| *Synechococcus mexicanus* | PCC 7335 [T] • | *Coccusdissimilis* | *C. mexicanus* | Marine (coastal) | Snail shell, intertidal zone, Puerto Penasco, Mexico | 11 | 5.97 | 48.2 | 5702 | 98.91 | β |
| *Synechococcus berkleyi* | PCC 7336 [T] • | *Eurycoccus* | *E. berkleyi* | Marine (coastal) | Sea Water Tank, Berkeley University, CA, USA | 1 | 5.07 | 53.7 | 5093 | 100 | β |
| *Synechococcus bogii* | PCC 7502 [T] • | *Leptovivax* | *L. bogii* | Sphagnum bog (peat bog) | NA | 3 | 3.58 | 40.6 | 3703 | 99.76 | β |
| *Synechocystis sp.* | PCC 7509 | *Doliumcoccus* | *D. switzii* | Soil | Rock scraping, Switzerland | 4 | 4.9 | 41.67 | 4859 | 99.67 | β |
| *Trichodesmium erythraeum* | IMS101 [T] | | | Marine (coastal) | NA | 1 | 7.75 | 34.14 | 4358 | 99.71 | β |
| *Xenococcus sp.* | PCC 7305 | | *X. lajollai* | Marine (coastal) | Aquarium, La Jolla, CA, USA | 234 | 5.92 | 39.68 | 4992 | 99.78 | β |
| *Gloeobacter violaceum* | PCC 7421 [8] | | | Soil | Calcareous (chalky) rock, Switzerland | 1 | 4.66 | 62 | 4511 | 99.15 | β |

**Add informations:**

[§] Taxonomic scheme (orders and families) according to NCBI Taxonomy Browser (August 31, 2009). Source: *http://www.bacterio.net/*

¢ Classification proposed by Kózlov et al. (2016), based on 16S rRNA sequences.
⍉ Classification proposed by Ramos et al. (2017), based on 16S rRNA sequences. Database: CyanoType version 1 (*http://lege.ciimar.up.pt/cyanotype*).

[1] Cyanobacterial genomes used in Komarek et al. (2014) paper and available at public database in January 2016 were retrieved for this study.

Type strains or type species are indicated with overwritten T at the end of the name.

[2] *Arthrospira platensis* is also called *Spirulina platensis*.

[3] *Coleofasciculus chthonoplastes* PCC 7420 is also called *Microcoleus chthonoplastes* PCC 7420.

[4] *Leptolyngbya* sp. PCC 7104 is also called *Nodosilinea nodulosa* PCC 7104.

[5] *Lyngbya aestuarii* PCC8106 is also called *L. aestuarii* CCY9616, and even the former name *Oscillatoria limosa* PCC8106.

[6] *Moorea producens* 3L is also called *Moorea producta* 3L.

[7] *Oscillatoria* sp. PCC 6407 is also called *Kamptonema formosum* PCC 6407, and even *O. formosa* PCC 6407.

[8] Outgroup used in the phylogenetic analysis.

• New taxonomic identification proposed by Coutinho et al. 2016.

° New taxonomic classification proposed by Thompson et al. 2013.

[Δ] Number of contigs, total length and GC content values were obtained using QUAST tool.

* Values using CheckM tool.

## Annotation and genomic taxonomy

All genomes were annotated using Prokka version 1.11 (Seemann 2014), with default settings. Genomic similarity between the 100 strains was determined by Average Amino Acid Identity (AAI) (Konstantinidis and Tiedje 2005), Average Nucleotide Identity (ANI), and in silico DNA-DNA Hybridization (DDH) or Genome-to-Genome Distance (GGD) (Auch et al. 2010; Auch et al. 2010; Meier-Kolthoff et al. 2013). GGD were calculated using the Genome-to-Genome Distance Calculator tool, version 2.1 (Meier-Kolthoff et al. 2013) (http://ggdc.dsmz.de/). Genomes were considered to belong to the same species if they displayed AAI and ANI ranging from 95 to 100% and 70 to 100% GGD. To verify the taxonomic similarity, the genomic content values were plotted in heatmaps, which were generated in R (R Development Core Team 2011), using the heatmap.2 {gplots} package.

## Phylogenetic analysis

To establish the phylogenetic structure of the phylum Cyanobacteria, phylogenetic trees were constructed using the 16S rRNA gene sequences and the concatenated alignments of a set of conserved genes, most of which encode ribosomal proteins (Table 2).

**Table 2. Conserved marker genes used in MLST phylogenetic reconstruction.**

| Gene Name | Gene Product |
|---|---|
| *dnaG* | DNA primase |
| *rplB* | 50S ribosomal protein L2 |
| *rplM* | 50S ribosomal protein S13 |
| *rpsC* | 30S ribosomal protein S3 |
| *frr* | ribosome recycling factor |
| *rplC* | 50S ribosomal protein L3 |
| *rplN* | 50S ribosomal protein L14 |
| *rpsE* | 30S ribosomal protein S5 |
| *nusA* | transcription termination protein (64,000 mol. wts.) |
| *rplD* | 50S ribosomal protein L4 |
| *rplP* | 50S ribosomal protein L16 |
| *rpsI* | 30S ribosomal protein S9 |

| | |
|---|---|
| *pgk* | phosphoglycerate kinase |
| *rplE* | 50S ribosomal protein L5 |
| *rplT* | 50S ribosomal protein L20 |
| *rpsS* | 30S ribosomal protein S19 |
| *pyrG* | CTP synthase |
| *rplF* | 50S ribosomal protein L6 |
| *rpoB* | β subunit of bacterial RNA polymerase |
| *smpB* | RNA-binding protein |
| *rplA* | 50S ribosomal protein L1 |
| *rplL* | 50S ribosomal protein L12 |
| *rpsB* | 30S ribosomal protein S2 |
| *tsf* | elongation factor Ts |

**Ribosomal RNA sequences.** The small subunit ribosomal RNA (16S rRNA) sequences from all cyanobacterial strains for which whole genome sequence data are publically available (exception see below, thus N = 97), as well as 16S rRNA gene sequences from additional type-strains available (N = 14) were all analyzed. The sequences were retrieved from the ARB SILVA database (Pruesse et al. 2007; Quast et al. 2013). Whenever sequences were not available, they were retrieved directly from the genomes using RNammer 1.2 Server (Lagesen et al. 2007). Sequences were aligned through MUSCLE v. 3.8 (Edgar 2004), with default settings, and Gblocks 0.91b (Castresana 2000; Talavera and Castresana 2007) was used for alignment curation. Using MEGA 6 (Tamura et al. 2013), best-fitting nucleic acid substitution models were calculated through the MLModelTest feature. Models were ranked based on their Bayesian Information Criterion (BIC) scores as described by Tamura et al. (2013). The model with the lowest BIC score was selected and used for further phylogenetic analysis. The phylogenetic inference was obtained using the Maximum Likelihood method based on the Kimura 2 parameter method with the Gamma distributed rate variation (K2+G) as the nucleotide substitution model, which was estimated from the data. The support branches of tree topology were checked by 1,000 bootstrap replicates. The 16S rRNA gene alignments were used to estimate the degree of genetic distance between strains through the Tajima-Nei method.

*Gloeobacter violaceus* PCC 7421 was set as the outgroup in both trees. Trees were visualized with FigTree, version 1.4.2 (Rambaut 2015). Due to incomplete or partial sequences, Synechococcus sp. CB0101 was omitted from these analyses. Planktothrix mougeotii NIVA-CYA 405 as well as Planktothrix prolifica NIVA-CYA 540 were not included in the phylogenetic analyses because 16S rRNA sequences are not currently available for these strains (and not retrievable from their genomes).

The type-strains of each taxa were included in the 16S phylogenetic tree to confirm the phylogenetic relatedness of the cyanobacterial genomes. Designations of type strain and type species were not available for *Chaemaesiphon minutus* PCC6605, *Pleurocapsa* sp. PCC7319, *Rivularia* sp. PCC7116, *Synechocystis* sp. PCC7509, *Trichodesmium erythraeum* IMS01, *Xenococcus* sp. PCC7305, cyanobacterium ESFC-1, and cyanobacterium JSC-12. *Geitlerinema* sp. PCC7105 is the reference strain for marine species of *Geitlerinema*, and PCC73106 is the reference strain for *Gloeocapsa* (Bhattacharya, 2012).

**Conserved marker genes: multilocus sequence analysis (MLSA).** A phylogenetic tree was generated using 31 conserved gene sequences previously validated as phylogenetic markers for (cyano)bacteria (Shih et al. 2013; Wu and Eisen 2008). The sequences of these proteins were mined using the *AutoMated Phylogenomic infeRence Application* - AMPHORA2 tool (Wu and Scott 2012), through default settings for the Bacteria option, and with a cut-off value of $1.e^{-10}$. Individual alignments were performed for each of the 31 gene sets through MUSCLE v. 3.8 with default settings (Edgar 2004). All alignments were then concatenated. Only OTUs with complete sequences for all 31 genes were used in the phylogenetic analysis. A Maximum Likelihood tree was constructed using RaxML v. 7 (Stamatakis 2006) and the Dayhoff+G likelihood model. One thousand bootstrap replications were calculated to evaluate the relative support of the branches. Trees were visualized with FigTree, version 1.4.2 (Rambaut 2015).

## Abundance of cyanobacterial genomes across aquatic environments and ecological correlations

Marine and freshwater metagenomes were retrieved to determine the abundance of 100 cyanobacterial genomes across the Earth. A set of 191 marine metagenomes from the Tara Ocean project were retrieved for analysis along with their associated metadata (Sunagawa et al. 2015). Sample-associated environmental data were inferred across multiple depths at global scale of Tara's metagenomics sampling: i) surface water layer (5 m, s.d. = 0); and ii) subsurface layer, including deep chlorophyll maximum zone (71 m, s.d. = 41 m) and mesopelagic zone (600 m, s.d. = 220 m) (Sunagawa et al. 2015). Eight freshwater metagenomes were retrieved for analysis from the Caatinga biome microbial community project along with their associated metadata (Lopes et al. 2016).

Metagenome reads were mapped to a database containing the 100 analyzed cyanobacterial genomes through Bowtie2 (Langmead and Salzberg 2012) using -very-sensitive-local and -a options. Abundance of genomes across samples was calculated based on the number of mapped reads as described by Iverson et al. (2012). Metagenomes were compared based on the relative abundances of the 100 analyzed genomes within them using non-metric multidimensional scaling (NMDS).

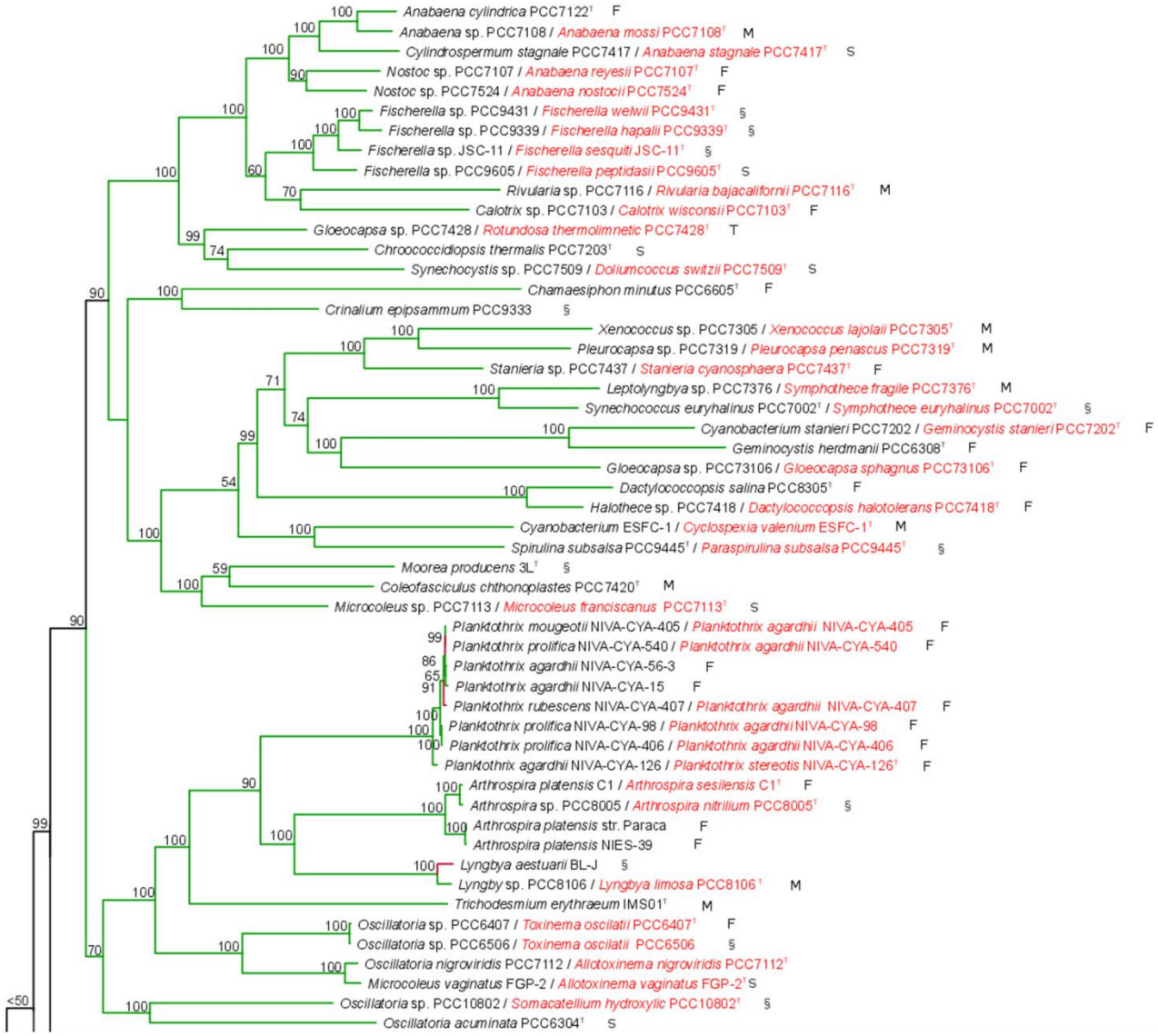
Next, Spearman correlation coefficients (R, or Spearman's rho) were calculated for the abundance of each genome and the levels of measured environmental parameters across samples. Next, a dissimilarity matrix of Manhattan distances was calculated based on the Spearman correlation values of every genome. All correlations were used by this analyzes regardless of the corrected p-value, as non-significant correlations are still ecologically informative as they indicate weak associations between microorganisms and environmental parameters. Finally, this dissimilarity matrix was used as input for hierarchical clustering using the complete linkage method

within the *hclust*() function in R. The resulting dendogram was visually inspected to define groups (i.e. ecogenomic groups) of organisms with similar correlation patterns which were named based on the main correlated feature.

## Results

### Phylogenomic framework reconstruction

The MLSA tree (Figure 1) based on conserved genes sequences gave a higher resolution than the 16S rRNA phylogenetic analysis (Figure S1 and Table S1). The 16S rRNA topology showed incongruences on the taxonomy of Cyanobacteria. In total, 59 branches corresponding to genera were recognized.


Anabaena cylindrica PCC7122T F
Anabaena sp. PCC7108 / Anabaena mossi PCC7108T M
Cylindrospermum stagnale PCC7417 / Anabaena stagnale PCC7417T S
Nostoc sp. PCC7107 / Anabaena reyesii PCC7107T F
Nostoc sp. PCC7524 / Anabaena nostocii PCC7524T F
Fischerella sp. PCC9431 / Fischerella welwii PCC9431T §
Fischerella sp. PCC9339 / Fischerella hapalii PCC9339T §
Fischerella sp. JSC-11 / Fischerella sesquiti JSC-11T §
Fischerella sp. PCC9605 / Fischerella peptidasii PCC9605T S
Rivularia sp. PCC7116 / Rivularia bajacalifornii PCC7116T M
Calotrix sp. PCC7103 / Calotrix wisconsii PCC7103T F
Gloeocapsa sp. PCC7428 / Rotundosa thermolimnetic PCC7428T T
Chroococcidiopsis thermalis PCC7203T S
Synechocystis sp. PCC7509 / Doliumcoccus switzii PCC7509T S
Chamaesiphon minutus PCC6605T F
Crinalium epipsammum PCC9333 §
Xenococcus sp. PCC7305 / Xenococcus lajolaii PCC7305T M
Pleurocapsa sp. PCC7319 / Pleurocapsa penascus PCC7319T M
Stanieria sp. PCC7437 / Stanieria cyanosphaera PCC7437T F
Leptolyngbya sp. PCC7376 / Symphothece fragile PCC7376T M
Synechococcus euryhalinus PCC7002T / Symphothece euryhalinus PCC7002T §
Cyanobacterium stanieri PCC7202 / Geminocystis stanieri PCC7202T F
Geminocystis herdmanii PCC6308T F
Gloeocapsa sp. PCC73106 / Gloeocapsa sphagnus PCC73106T F
Dactylococcopsis salina PCC8305T F
Halothece sp. PCC7418 / Dactylococcopsis halotolerans PCC7418T F
Cyanobacterium ESFC-1 / Cyclospexia valenium ESFC-1T M
Spirulina subsalsa PCC9445T / Paraspirulina subsalsa PCC9445T §
Moorea producens 3LT §
Coleofasciculus chthonoplastes PCC7420T M
Microcoleus sp. PCC7113 / Microcoleus franciscanus PCC7113T S
Planktothrix mougeotii NIVA-CYA-405 / Planktothrix agardhii NIVA-CYA-405 F
Planktothrix prolifica NIVA-CYA-540 / Planktothrix agardhii NIVA-CYA-540 F
Planktothrix agardhii NIVA-CYA-56-3 F
Planktothrix agardhii NIVA-CYA-15 F
Planktothrix rubescens NIVA-CYA-407 / Planktothrix agardhii NIVA-CYA-407 F
Planktothrix prolifica NIVA-CYA-98 / Planktothrix agardhii NIVA-CYA-98 F
Planktothrix prolifica NIVA-CYA-406 / Planktothrix agardhii NIVA-CYA-406 F
Planktothrix agardhii NIVA-CYA-126 / Planktothrix stereotis NIVA-CYA-126T F
Arthrospira platensis C1 / Arthrospira sesilensis C1T F
Arthrospira sp. PCC8005 / Arthrospira nitrilium PCC8005T §
Arthrospira platensis str. Paraca F
Arthrospira platensis NIES-39 F
Lyngbya aestuarii BL-J §
Lyngby sp. PCC8106 / Lyngbya limosa PCC8106T M
Trichodesmium erythraeum IMS01T M
Oscillatoria sp. PCC6407 / Toxinema oscilatii PCC6407T F
Oscillatoria sp. PCC6506 / Toxinema oscilatii PCC6506 §
Oscillatoria nigroviridis PCC7112 / Allotoxinema nigroviridis PCC7112T
Microcoleus vaginatus FGP-2 / Allotoxinema vaginatus FGP-2T S
Oscillatoria sp. PCC10802 / Somacatellium hydroxylic PCC10802T §
Oscillatoria acuminata PCC6304T S

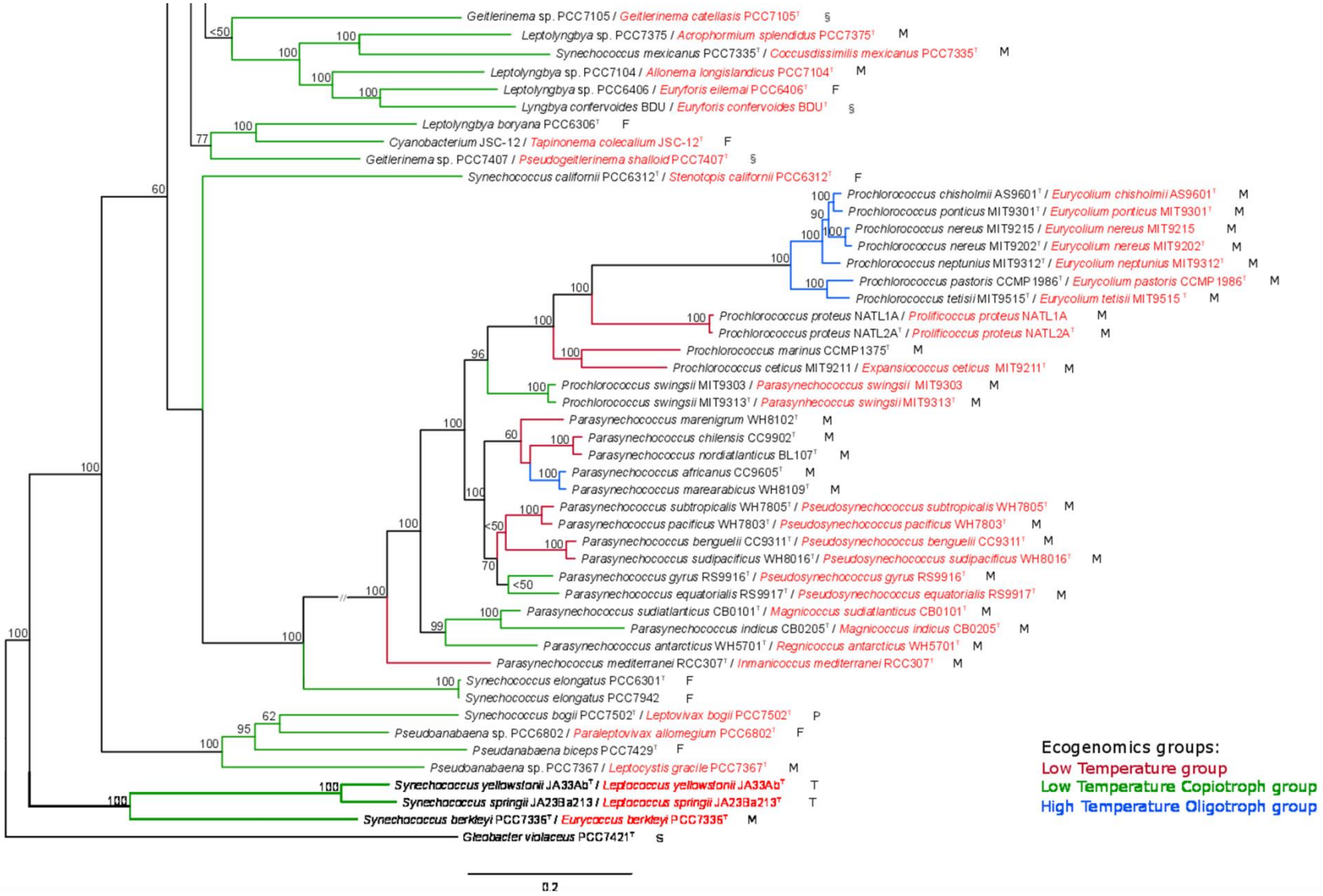

Geitlerinema sp. PCC7105 / Geitlerinema catellasis PCC7105T §
Leptolyngbya sp. PCC7375 / Acrophormium splendidus PCC7375T M
Synechococcus mexicanus PCC7335T / Coccusdissimilis mexicanus PCC7335T M
Leptolyngbya sp. PCC7104 / Allonema longislandicus PCC7104T M
Leptolyngbya sp. PCC6406 / Euryforis eilemai PCC6406T F
Lyngbya confervoides BDU / Euryforis confervoides BDUT §
Leptolyngbya boryana PCC6306T F
Cyanobacterium JSC-12 / Tapinonema colecalium JSC-12T F
Geitlerinema sp. PCC7407 / Pseudogeitlerinema shalloid PCC7407T §
Synechococcus californii PCC6312T / Stenotopis californii PCC6312T F
Prochlorococcus chisholmii AS9601T / Eurycolium chisholmii AS9601T M
Prochlorococcus ponticus MIT9301T / Eurycolium ponticus MIT9301T M
Prochlorococcus nereus MIT9215 / Eurycolium nereus MIT9215 M
Prochlorococcus nereus MIT9202T / Eurycolium nereus MIT9202T M
Prochlorococcus neptunius MIT9312T / Eurycolium neptunius MIT9312T M
Prochlorococcus pastoris CCMP1986T / Eurycolium pastoris CCMP1986T M
Prochlorococcus tetisii MIT9515T / Eurycolium tetisii MIT9515T M
Prochlorococcus proteus NATL1A / Prolificoccus proteus NATL1A M
Prochlorococcus proteus NATL2AT / Prolificoccus proteus NATL2AT M
Prochlorococcus marinus CCMP1375T M
Prochlorococcus ceticus MIT9211 / Expansiococcus ceticus MIT9211T M
Prochlorococcus swingsii MIT9303 / Parasynechococcus swingsii MIT9303 M
Prochlorococcus swingsii MIT9313T / Parasynhecoccus swingsii MIT9313T M
Parasynechococcus marenigrum WH8102T M
Parasynechococcus chilensis CC9902T M
Parasynechococcus nordiatlanticus BL107T M
Parasynechococcus africanus CC9605T M
Parasynechococcus marearabicus WH8109T M
Parasynechococcus subtropicalis WH7805T / Pseudosynechococcus subtropicalis WH7805T M
Parasynechococcus pacificus WH7803T / Pseudosynechococcus pacificus WH7803T M
Parasynechococcus benguelii CC9311T / Pseudosynechococcus benguelii CC9311T M
Parasynechococcus sudipacificus WH8016T / Pseudosynechococcus sudipacificus WH8016T M
Parasynechococcus gyrus RS9916T / Pseudosynechococcus gyrus RS9916T M
Parasynechococcus equatorialis RS9917T / Pseudosynechococcus equatorialis RS9917T M
Parasynechococcus sudiatlanticus CB0101T / Magnicoccus sudiatlanticus CB0101T M
Parasynechococcus indicus CB0205T / Magnicoccus indicus CB0205T M
Parasynechococcus antarcticus WH5701T / Regnicoccus antarcticus WH5701T M
Parasynechococcus mediterranei RCC307T / Inmanicoccus mediterranei RCC307T M
Synechococcus elongatus PCC6301T F
Synechococcus elongatus PCC7942 F
Synechococcus bogii PCC7502T / Leptovivax bogii PCC7502T P
Pseudobanabaena sp. PCC6802 / Paraleptovivax allomegium PCC6802T F
Pseudanabaena biceps PCC7429T F
Pseudbanabaena sp. PCC7367 / Leptocystis gracile PCC7367T M
Synechococcus yellowstonii JA33AbT / Leptococcus yellowstonii JA33AbT T
Synechococcus springii JA23Ba213 / Leptococcus springii JA23Ba213T T
Synechococcus berkleyi PCC7335T / Eurycoccus berkleyi PCC7335T M
Gleobacter violaceus PCC7421T S
0.2
Ecogenomics groups:
Low Temperature group
Low Temperature Copiotroph group
High Temperature Oligotroph group

**Figure 1. Multilocus sequence analysis (MLSA) phylogenetic tree of the Cyanobacteria phylum with the proposed new names.**

Tree construction was performed using 100 genomes. The numbers at the nodes indicate bootstrap values as percentages greater than 50 %. Bootstrap tests were conducted with 1,000 replicates. The unit of measure for the scale bars is the number of nucleotide substitutions per site. The *Gloeobacter violaceus* PCC 7421 sequence was designated as outgroup. Capital letters indicate environmental source: F, freshwater; M, marine; S, soil; T, thermal; and §, other. New names are highlighted in red. Overwritten T indicates type strain. Ecogenomic groups are depicted in different colors as indicated in the legend: Low Temperature group; Low Temperature Copiotroph group; and High Temperature Oligotroph group.

## Genomic diversity of Cyanobacteria

In total, we found 59 branches corresponding to genera based on the AAI, ANI and GGD analyses (Figure 2). The genus and species cut-off delimitation were ≥ 70% and ≥ 95% AAI similarity respectively. Thirty-one new genera and 87 species (of which 32 new spp.) were circumscribed. From a total of 100 genomes used in this study, 70 were previously classified to the species level, whereas the remaining 30 had incomplete taxonomic classification (i.e. only sp. or unclassified). In total, 16 genera (from a total of 44) and 38 species (from a total of 69) were taxonomically reclassified and/or re-named. Thus, we found that 73 of all analyzed genomes required reassignment at one or more ranks to reconcile existing taxonomic classifications with our new genomic taxonomy (Figures 2 and S1).

Meanwhile, our results showed that *Anabaena* and *Nostoc* should be merged into a single genus. This case has been debated for some time, and similar results were obtained in other studies (Zehr et al. 1997; Tamas et al. 2000). Over the next section, we highlight four specific cases to exemplify cyanobacterial taxonomic issues that were resolved through our genome-driven approach (see Figure S2). These cases illustrate how the use of genomic taxonomy in Cyanobacteria provides relevant information (supplementary material).


**Figure 2. Heatmap based on similarity matrix of AAI between 100 genomes.**

The intraspecies limit is assumed as ≥ 95 %, whereas genera delimitation is assumed as ≥ 70 % (dashed lines) AAI. Hierarchical clustering was performed based on Manhattan distances. The proposed new names were adopted in this figure.

**Case I.** *Oscillatoria* group. Analysis of the five genomes of *Oscillatoria* distinguished three genera, based on the genomic signatures (i.e., GGDH, AAI, 16S, and MLSA): i. *Oscillatoria acuminata* PCC 6304 type strain formed a separate group; ii. *Oscillatoria* PCC 10802 formed a separate divergent group, corresponding to a new genus named *Somacatellium* (*S. hydroxylic* PCC 10802); and iii. *Oscillatoria* strains PCC 7112 and *Microcoleus vaginatus* FGP-2 formed a new genus named *Allotoxinema* (*A. nigroviridis* PCC7112 and A*. vaginatus* FGP-2); and iv. *Oscillatoria* strains PCC 6407 and PCC 6506 formed a new genus named *Toxinema* (*T. oscillati* PCC 6407 and *T. oscillati* PCC 6506).

**Case II.** *Leptolyngbya* group. The five *Leptolyngbya* strains were polyphyletic, forming different phylogenetic branches. Thus, i. *Leptolyngbya boryana* PCC 6306 type strain forms a separate group with cyanobacterium JSC-12, while the rest of the *Leptolyngbya* strains cluster apart; ii. strain PCC 7376 forms a group along with *Synechococcus euryhalinus* PCC 7002, representing a new genus named *Symphothece* (*S. fragile* PCC 7376 and *S. euryhalinus* PCC 7002); iii. strain PCC 7375 forms a new genus named *Acrophormium* (*A. splendidus* PCC 7375); iv. strain PCC 7104 forms a new genus named *Allonema* (*A. longislandicus* PCC 7104); and v. strains PCC 6406 and BDU (*Lyngbya confervoides*) form a new genus named *Euryforis* (*E. eilemai* PCC 6406 and *E. confervoides* BDU).

**Case III.** *Arthrospira* group. Examination of the five *Arthrospira* strains indicated that i. *A. platensis* C1 should be considered a new species, named *A. sesilensis*; ii. strain 8005 belongs to a new species, named *A. nitrilium*; and iii. the type strain of *Arthrospira platensis* (PCC 7345) formed a tight cluster with NIES-39 and Paraca.

**Case IV.** *Synechococcus* group. The nine *Synechococcus* strains split in i. *S. elongatus* PCC 6301 type strain forms a separate group with *S. elongatus* PCC 7942; ii. strain PCC 6312 forms a new genus named *Stenotopis* (*S. californii* PCC 6312); iii.

strain PCC 7335 belongs to a new genus named *Coccusdissimilis* (*C. mexicanus* PCC 7335); iv. strains JA23Ba213 and JA33Ab formed a new genus named *Leptococcus* (*L. springii* JA23Ba213 and *L. yellostonii* JA33Ab); v. strain PCC 7336 formed a new genus named *Eurycoccus* (*E. berkleyi* PCC 7336); vi. strain PCC 7502 belonged to a new genus named *Leptovivax* (*L. bogii* PCC 7502); and vii. strain PCC 7002 formed a tight group with strain PCC 7376 (*Leptolyngbya* sp.), representing a new genus named *Symphothece* (*S. euryhalinus* PCC 7002 and *S. fragile* PCC 7376).

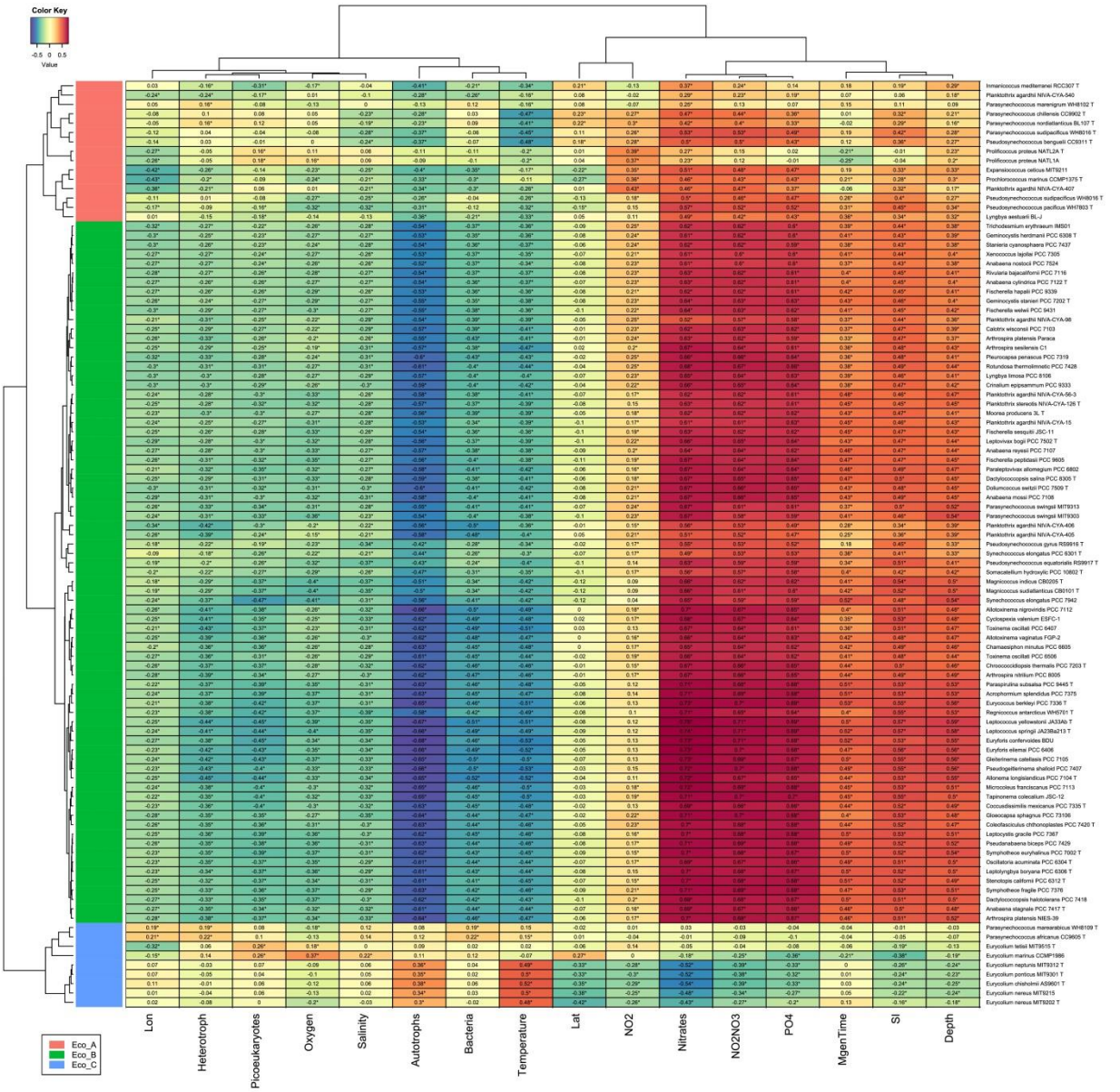



**Figure 3. Environmental correlations profile used to define ecogenomic groups.**

Heatmap displays Spearman correlation scores between the abundance of cyanobacterial genomes and measured environmental parameters at Tara Ocean sampling sites. Correlations that showed q corrected p-value < 0.05 are marked with stars. Variables were grouped through the complete linkage clustering method using Manhattan distances as input. The proposed new names were adopted in this figure.

## Charting ecological groups of Cyanobacteria

Our phylogenomic analysis was complemented by an ecological characterization of the analyzed strains. Correlating the relative genome abundances with environmental parameters measured at *Tara* Oceans samples (Sunagawa et al. 2015) revealed associations between Cyanobacteria and physical, chemical and biological variables of their habitats (Figure 3). The ecogenomic analysis clustered genomes based on their profiles of correlations to environmental parameters. Three major ecogenomic groups were found: a) Low Temperature; b) Low Temperature Copiotroph; and c) High Temperature Oligotroph (Figure 4A-C). Closely related species of the same genus show tight associations with environmental parameters, grouped to the same ecogenomic group (with few exceptions, i.e. *Planktothrix* and *Parasynechococcus*).

Members of the Low Temperature group were characterized by positive correlations with the concentration of nitrogen and phosphorus sources; weak positive correlations with minimum generation time, silicate and depth; and by negative correlations with temperature, microbial cell abundance, oxygen availability, and salinity. Meanwhile, members of the Low Temperature Copiotroph group were characterized by strong positive correlations with the concentration of nitrogen and phosphorus; positive correlations (stronger than those presented by Low Temperature group) with minimum generation time, silicate and depth; and by negative correlations (also stronger than those presented by Low Temperature group) with temperature, microbial cell abundance (in particular with autotroph cell density), oxygen availability, and salinity. Finally, members of High Temperature Oligotroph group were characterized by negative correlations with the concentration of nitrogen and phosphorus and positive correlations with temperature and autotroph cell abundance.

As suggested by correlation analyses (Figure 3), NMDS revealed the Low Temperature Copiotroph group to be more abundant in cold and eutrophic waters, while the High Temperature Oligotroph group exhibited the opposite pattern and was

more abundant in warm and oligotrophic environments (Figures 4D-E). In turn, Low Temperature was more abundant at intermediate conditions between these polar opposites and was shown to be more abundant in samples with higher cell densities and $NO_2$ concentrations.

We also investigated the abundance of the ecogenomic groups in freshwater environments. Unfortunately, there is no currently available large-scale dataset of freshwater metagenomes with associated metadata comparable to the *Tara* Oceans dataset. Unable to define freshwater ecogenomic groups we chose to extrapolate the classification obtained from the analyses of the marine dataset. In freshwater metagenomes, the Low Temperature Copiotroph was the dominant group in all the analyzed samples (Figure S3A). NMDS of freshwater samples suggested that Low Temperature group displayed a preference for higher pH and DOC, nitrite and total nitrogen concentrations whereas the High Temperature Oligotroph group has a preference for habitats with higher concentrations of POC, phosphorus, ammonia and nitrate (Figures S3B-C).

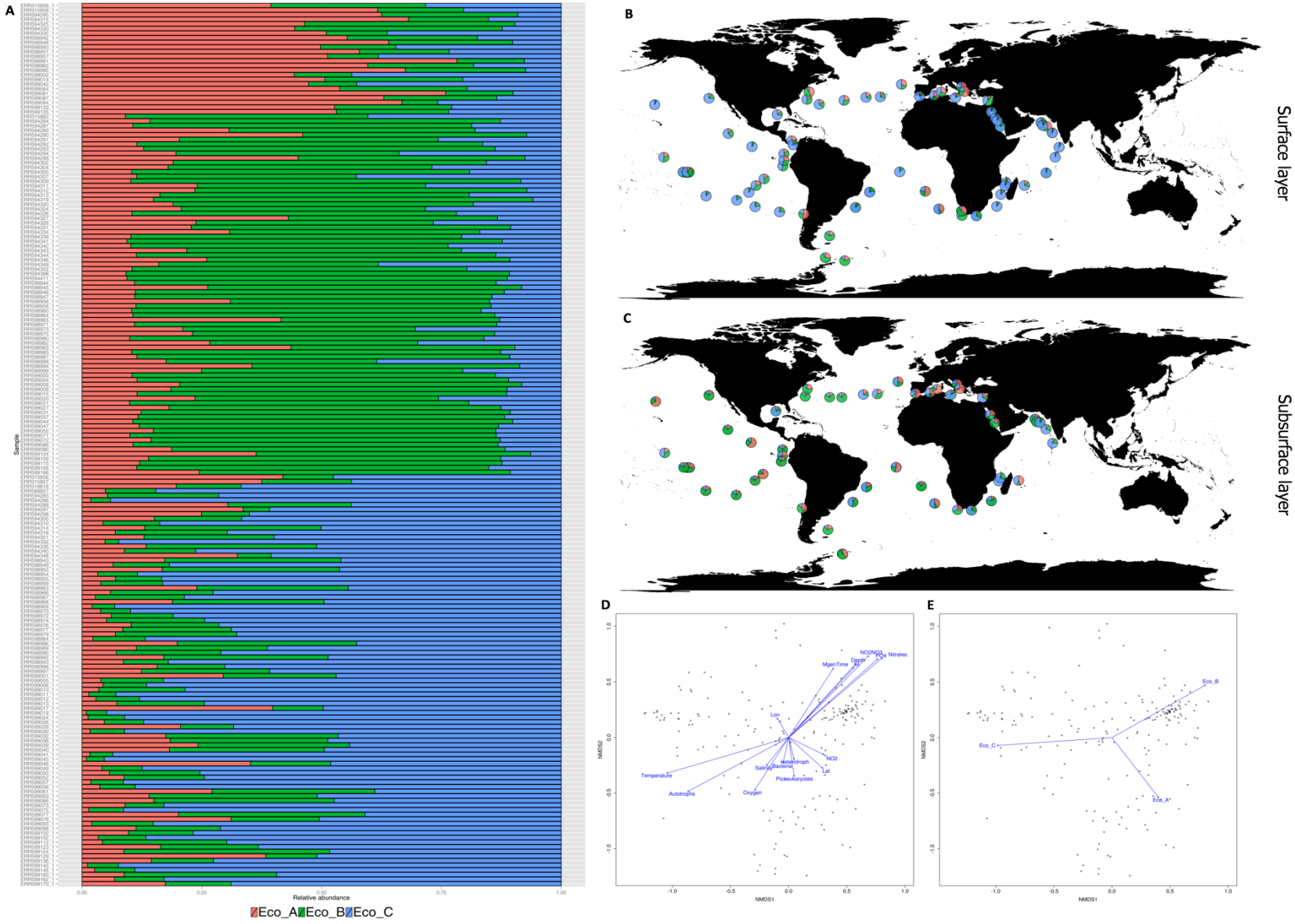

A
Sample
Relative abundance
0.00
0.25
0.50
0.75
1.00
Eco_A
Eco_B
Eco_C
B
C
Surface layer
Subsurface layer
D
E
NMDS1
NMDS2
MeanTime
Depth
NO2NO3
PO4
Nitrates
Lon
Lat
NO2
Heterotroph
Bacteria
Salinity
Picoeukaryotes
Temperature
Autotrophs
Oxygen
Eco_B
Eco_C
Eco_A

**Figure 4. Ecogenomic of Cyanobacteria across the Earth.**

(A) Abundance and distribution of ecogenomic clusters across global marine metagenomes. Relative abundance of A- Low Temperature group; B -Low Temperature Copiotroph group; and C- High Temperature Oligotroph group at the global scale. (B) Distribution of the dominant ecogenomic groups along the *Tara* Ocean transect sampling from surface layer (5 m). (C) Distribution of the dominant ecogenomic groups along the Tara Ocean transect sampling from subsurface layer (> 5 m). (D) Non-metric multidimensional scaling (NMDS) analysis of the marine metagenomes and environmental parameters. Ordination plot of physicochemical parameters. Dots indicate the metagenomes samples. Distances were calculated based on the Bray-Curtis Method. NMDS stress value = 0.15. (E) Non-metric multidimensional scaling (NMDS) analysis of the marine metagenomes and environmental parameters. Ordination plot of ecogenomic clusters. Dots indicate the metagenomes samples. Distances were calculated based on the Bray-Curtis Method. NMDS stress value = 0.15.

## Discussion

The use of high-throughput sequence technologies and environmental surveys have allowed studies that link phylogenomics and ecogenomics of Cyanobacteria. High-throughput genome sequence technologies (e.g. Illumina) are causing a revolution in microbial diversity studies. Through the process of binning of contigs or scaffolds derived from the same strain, complete genomes can be reconstructed. Recent studies have obtained dozens of new metagenome-assembled genomes from complex environmental samples (Almstrand et al. 2016; Brown et al. 2015; Haroon et al. 2016; Hugerth et al. 2015; Pinto et al. 2016). The abundance of these genomes across different environments can now be inferred from metagenomics, including their metabolic and ecological potential. It is clear that a new system is required to allow for precise taxonomic identification of these new genomes.

## WGS as the basic unit for Cyanobacteria genomic taxonomy (CGT)

Comparative genomics studies allows for identification of sequence groups with high genotypic similarity based on variation in protein coding genes distributed across the genomes. Analyses of environmental metagenomes and microbiomes have shown that microbial communities consist of genotypic clusters of closely related organisms (Farrant et al. 2016). These groups display cohesive environmental associations and

dynamics that differentiate them from other groups co-existing in the same samples. In light of new concepts, restlessness is mounting with the inability to define the microbial species itself. Evolution studies on closely related bacteria show rapid and highly variable gene fluxes in evolving microbial genomes, suggesting that extensive gene loss and horizontal gene transfer leading to innovation are the dominant evolutionary processes (Batut et al. 2014; Puigbò et al. 2014). CGT will solve the problem of the frequent observation that even closely related genomes can have high gene content variation that gives phenotypic variation. CGT is completely adjusting to the genomics era, addressing the needs of its users in microbial ecology and clinical microbiology, in a new paradigm of open access (Beiko 2015). CGT will provide a predictive operational framework for reliable automated and openly available identification and classification (Thompson et al. 2015).

## Proposals for cyanobacterial taxonomy

A main gap exists and is growing each day between the formal taxonomy of Cyanobacteria and the forest of acronyms and numbers in the different databases. Indeed, the nameless OTUs, strains, isolates and WGS sequences (Beiko 2015; Kózlov et al. 2016) form the great majority of data in private and public databases. There is a need to re-examine the Cyanobacteria prokaryote species, taking into account all recently developed concepts, e.g. the gene flow unit, OTU, ETUs and CTU in the context of a pragmatic genome-based taxonomic scheme. The type can be a culture, DNA or a WGS. The CGT system should maintain all of the existing information, integrating it with new data on DNA, genomes, isolates/strains, cultured and uncultured, “Candidatus” cases and reconstructed genomes from metagenomes (Brown et al. 2015; Hugerth et al. 2015). The international initiatives of GEBA is currently working on determining the WGS of all type strains of known microbial species to shorten this gap (more than eleven thousand genomes).

A modern taxonomy should be based on WGS. The enormous amount of unique gene sequences (e.g. 16S rRNA gene) databases, which contain many misleading, should be compared to the available genomes-based phylogeny. Studies focusing on one specifically taxa/group cannot disregarded the phylogenetic analysis for the whole major taxa. It will avoid the inclusion of the previously erroneous taxa on the analysis. Further, the anxiety to give a new name should be reconsidered. On the literature is easily to found proposes of new taxa, even that the phylogenetic relationship was not firmly established (e.g. Rajaniemi et al. 2005).

## Ecogenomics and the delineation of the ecological niches of Cyanobacteria

Correlation analysis allowed us to characterize how the abundance of the analyzed genomes is associated with environmental parameters at both marine and freshwater habitats. These associations shed light on ecological interactions taking place within aquatic habitats that are responsible for delineating the ecological niches of Cyanobacteria. Our results shown that taxonomic affiliation and niche occupancy are linked, i.e. closely related species of the same genus often shared correlation patterns, and consequently were assigned to the same ecogenomic group.

Identifying the specific features responsible for the defining niche occupancy among these organisms depends on extensive experimental data focusing on both physiological and morphological features, which is outside of our scope. Nevertheless, we speculate that a some features are likely playing a role in this process:

1) Transcriptional patterns: The way in which Cyanobacteria regulate gene expression in response to changing environmental conditions is likely to play a role in defining which habitats are better suitable for growth of different species.

2) Nutrient uptake and utilization: Throughout the aquatic environment a myriad of gradients of nutrient abundance are formed (Stocker 2012). The capacity of Cyanobacteria to uptake and utilize sources of limiting nutrients (e.g. P, N and Fe) is associated with the ecological niches they occupy (Coutinho et al. 2016; Thompson et

al. 2013; Farrant et al. 2016). Considering that several significant associations were detected between the abundance of the analyzed genomes and phosphorus and nitrogen sources, we assume that their diversity and efficiency of their nutrient transporters plays a major role in defining their affiliation to the proposed ecogenomic groups.

3) Photosynthetic machinery and efficiency: Cyanobacteria are remarkably diverse when considering their photosynthetic physiology. Species differ regard their ideal light intensities and preferred wavelengths which affects their photosynthetic efficiency (Moore et al., 1998, Ting et al., 2002). They also can be differentiated regarding their Carboxysomes, sub-cellular structures where carbon fixation takes place (Yeates et al., 2008). To our knowledge, no study has consistently compared the photosynthetic yields of all the strains analyzed here, therefore we cannot determine if the proposed ecogenomic groups differ regarding this parameter. Nevertheless, their distinctions regarding their requirements for efficient photosynthesis is likely linked to their patterns of niche occupancy.

## Ecogenomics, global changes and cyanobacterial communities

Over the past two centuries, human development has affected aquatic ecosystems due to nutrient over-enrichment (eutrophication), hydrologic alterations, global warming and ocean acidification. Temperature is one of the most important factors determining the taxonomic composition of marine microbial communities (Sunagawa et al. 2015). Our data shows that temperature is central for regulating the composition and functioning of cyanobacterial communities. Global warming can affect growth rates and bloom potentials of many taxa within this phylum (Flombaum et al. 2013; Fu et al. 2007; Pittera et al. 2014; Paerl and Huisman 2008). Niche based models predict an increase in the absolute levels of organisms formerly classified as *Prochlorococcus* and *Synechococcus* due to global warming (Flombaum et al. 2013). Consequently the functioning of the biogeochemical cycles in which these organisms

are involved will also be affected (Fu et al. 2007). Nevertheless, much less is known regarding how global warming could affect communities of Cyanobacteria aside from these two groups of organisms.

The ecogenomic groups identified and their associations with environmental parameters shed light into the potential changes that communities of Cyanobacteria will undergo following global climate changes. Our results indicate that an increase in temperatures will lead to decreases in the relative abundances of Low Temperature and Low Temperature Copiotroph groups, while that of High Temperature Oligotroph group increases, especially those of species *Eurycolium neptunis*, *E. ponticus*, *E. chisholmi* and *E. nereus*. One major impact of this alteration is a possible effect on the degree of nitrogen fixation mediated by Cyanobacteria, as none of the species assigned to the Low Temperature group are known to fix nitrogen (Latysheva et al. 2012). In fact, our data shows that higher temperatures are associated with lower relative abundances of nitrogen fixating Cyanobacteria of the genera *Trichodesmium* and *Anabaena* (Zehr 2011). This result is in agreement with recent laboratory and field study (Hong et al. 2017). Rising temperatures might shift cyanobacterial community composition towards a state were diazotrophs are relatively less abundant. Because nitrogen is often a limiting nutrient to marine primary productivity (Moore et al. 2013; Tyrrel 1999), alterations to oceanic levels of nitrogen fixation could affect not only non-diazotrophic Cyanobacteria but also heterotrophic microbes as well as the higher tropic levels that are sustained by microorganisms.

Furthermore, our findings suggest that changes in temperature can affect the contributions of Cyanobacteria to the global carbon pump. For example, four out of the five strongest positive correlations with temperature among High Temperature Oligotroph group involve the high-light adapted members of the *Eurycolium* genus (i.e. strains MIT9312, MIT9301, MIT9215, MIT9202 and AS9601). These are high-light adapted strains that display lower photosynthetic efficiency than their low-light adapted counterparts (Moore et al. 1998; Moore et al. 1999). Our results suggest that relative

abundance of high-light adapted strains would increase with rising temperatures. In turn, these changes could affect the efficiency of carbon fixation in the ocean, a change that could also be influenced by the alterations in nitrogen fixation mentioned above.

## Conclusions

The present study proposes a first attempt towards merging taxonomy and ecogenomics, offering a new opportunity for the development of Cyanobacteria studies. Our results show that closely related genomes often shared niches and were assigned to the same ecogenomic group. End-users of Cyanobacteria taxonomy may benefit from a more reproducible and portable taxonomic scheme. Future studies are needed to determine the evolutionary and physiological basis for niche occupancy of known and newly discovered Cyanobacteria, as well as linking their niche to other important ecological variables not analyzed here such as phage susceptibility, light utilization strategies, horizontal gene transfer, and inter-species interactions.

## Acknowledgments

The authors thankfully acknowledge National Counsel of Technological and Scientific Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES), and Rio de Janeiro Research Foundation (FAPERJ) for their financial support for this study. This paper is part of the D. Sc. requirements of JMW at the Biodiversity and Evolutionary Biology Graduate Program of the Federal University of Rio de Janeiro (UFRJ), and was developed within the Science Without Borders Program (CNPq, 207751/2014-5). We also state no conflict of interest or competing commercial interests in relation to the work.

## Sequence Data

All publicly available sequence data used in this paper was retrieved from the RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) and GenBank, as part of the International Nucleotide Sequence Database Collaboration, and also from the GEBA database, produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community.

## References

Almstrand, R., Pinto, A.J., Figueroa, L.A., Sharp, J.O. (2016). Draft genome sequence of a novel Desulfobacteraceae member from a sulfate-reducing bioreactor

metagenome. *Genome Announc*. 4(1): e01540-15. doi:10.1128/genomeA.01540-15.

Auch, A.F., von Jan, M., Klenk, H.-P., Göker, M. (2010a). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences*. 2(1):117–134. doi: 10.4056/sigs.531120.

Auch, A.F., Klenk, H.-P., Göker, M. (2010b). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Standards in Genomic Sciences*. 2(1):142–148. doi: 10.4056/sigs.541628

Batut, B., Knibbe, C., Marais, G., Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nature Reviews Microb*. 12:841–850.

Becraft, E. D., Wood, J. M., Rusch, D. B., Kühl, M., Jensen, S. I., Bryant, D. A., et al. (2015). The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of Synechococcus inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Front. Microbiol*. 6:590. doi: 10.3389/fmicb.2015.00590

Beiko, R.G. (2015). Microbial malaise: How can we classify the microbiome? Trends in Microbiology. 23(11):671–679.

Biller, S.J., Berube, P.M., Lindell, D., Chisholm, S.W. (2015). Prochlorococcus: the structure and function of collective diversity. *Nature Reviews Microb*. 13:13-27.

Boone, D.R., and Castenholz, R.W. (2001). Bergey's Manual of Systematic Bacteriology Volume One : The Archaea and the Deeply Branching and Phototrophic Bacteria 2nd ed. G. Garrity, D. R. Boone, and R. W. Castenholz, eds., New York, NY: Springer-Verlag New York.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. 523(7559):208–211. doi:10.1038/nature14486

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17(4):540–552.

Choudoir, M.L., Campbell, A.N., Buckley, D.H. (2012). Grappling with Proteus: population-level approaches to understanding microbial diversity. *Front Microb*. 3:336. doi: 10.3389/fmicb.2012.00336

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P., Swings, J.(2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiology Reviews*. 29(2):147–167.

Coutinho, F., Tschoeke, D.A., Thompson, F., Thompson, C. (2016). Comparative genomics of Synechococcus and proposal of the new genus Parasynechococcus. *PeerJ*. 4:e1522. doi:10.7717/peerj.1522

Coutinho, F.H., Meirelles, P.M., Moreira, A.P.B., Paranhos, R.P., Dutilh, B.E., Thompson, F.L. (2015). Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. *PeerJ*. 3:e1008. doi:10.7717/peerj.1008

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., et al. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife*. 2:e01102. doi: 10.7554/eLife.01102

Edgar, R.C.(2004). MUSCLE: Multiple sequence alignment with high accuracy and

high throughput. *Nucleic Acids Research*. 32(5):1792–1797.

Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., et al. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci USA*. 113(24):E3365-74. doi: 10.1073/pnas.1524865113

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., et al. (2013). Present and future global distributions of the marine cyanobacteria Prochlorococcus and Synechococcus. *Proc Natl Acad Sci USA*. 110(24):9824–9829.

Fu, F.-X., Warner, M.E., Zhang, Y., Feng, Y., Hutchins, D.A. (2007). Effects of increasedtemperature and $CO_2$ on photosynthesis, growth, and elemental ratios inmarineSynechococcus and Prochlorococcus. *J Phycol*. 43:485–496.

Garrity, G.M., and Oren, A. (2012). Response to Gribaldo and Brochier-Armanet: time for order in microbial systematics. *Trends Microbiol*. 20(8):353-4. doi: 10.1016/j.tim.2012.05.004

Garrity, G.M. (2016). A new genomics-driven taxonomy of Bacteria and Archaea: are we there yet? *J. Clin. Microb*. 54:1956-1963. doi:10.1128/JCM.00200-16

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., et al. (2005). Re-evaluating prokaryotic species. *Nat.Rev.Microbiol.*, 3:733–739.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*. 14:81–91. doi: 10.1099/ijs.0.64483-0

Gugger, M.F., and Hoffmann, L. (2004). Polyphyly of true branching cyanobacteria (Stigonematales). *International Journal of Systematic and Evolutionary Microbiology*. 54(2):349–357.

Haroon, M.F., Thompson, L.R., Parks, D.H., Hugenholstz, P., Stingl, U. (2016). Data Descriptor : a catalogue of 136 microbial draft genomes from Red Sea metagenomes. *Scientific Data.* 3:160050. doi: 10.1038/sdata.2016.50

Hong, H., Shen, R., Zhang, F., Wen, Z., Chang, S., Lin, W., et al. (2017). The complex effects of ocean acidification on the prominent $N_2$-fixing cyanobacterium Trichodesmium. *Science*. eaal2981. doi: 10.1126/science.aal2981

Hugenholtz, P., Skarshewski, A., Parks, D.H. (2016). Genome-based microbial taxonomy coming of age. *Cold Spring Harbor Perspectives in Biology*. 8(6):a018085. doi: 10.1101/cshperspect.a018085

Hugerth, L.W., Larsson, J., Alneberg, J., Lingh, M., Legrand, C., Pinhassi, J., et al. (2015). Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biology*. 16:279. doi:10.1186/s13059-015-0834-7

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science*. 335:587–590. doi:10.1126/science.1212665

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., Chisholm, S.W. (2006). Partitioning among Prochlorococcus ecotypes along environmental gradients. *Science*. 311:1737–1740.

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science*. 344:416–420. doi:10.1126/science.1248575

Komárek, J., Kastovsky, J., Mares, J., Johansen, J.R. (2014). Taxonomic classification

of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, 86: 295-335.

Konstantinidis, K.T., and Tiedje, J.M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Nat. Acad. Scienc.USA*. 102(7):2567–2572.

Konstantinidis, K.T., and Tiedje, J.M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187(18):6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005.

Kózlov, A.M., Zhang, J., Yilmaz, P., Glöckner, F.O., Stamatakis, A. (2016). Phylogeny-aware Identification and Correction of Taxonomically Mislabeled Sequences. *bioRxiv*. 2(13):1–12. doi:10.1093/nar/gkw396

Lagesen, K., Hallin, P., Rodland, mE.A., Staerfeldt, H.-H., Rognes, T., Ussery, D. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*. 35(9):3100–3108.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4):357–359.

Locey, K.J., and Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*. 113(21):5970–5975. doi: 10.1073/pnas.1521291113

Lopes, F.A.C., Catão, E.C.P., Santana, R.H., Cabral, A.S., Paranhos, R., Rangel, T.P., et al. (2016). Microbial community profile and water quality in a protected area of the Caatinga biome. *Plos One*. 11(2):e0148296.doi:10.1371/journal.pone.0148296

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M.(2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14:60.

Moore, C. M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nat. Geosci*. 6:701–710.doi:10.1038/ngeo1765

Moore, L.R., Rocap, G., and Chisholm, S.W. (1998). Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. *Nature*. 393:464–467.

Moore, L.R., and Chisholm, S.W. (1999). Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr*. 44:628–638.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*. 25(7):1043–1055.

Paerl, H.W. and Huisman, J. (2008). Blooms Like It Hot. *Science*. 320(5872):57-58. doi: 10.1126/science.1155398

Pinto, A.J., Sharp, J.O., Yoder, M.J., Almstrand, R. (2016). Draft genome sequences of two novel Acidimicrobiaceae members from an acid mine drainage biofilm metagenome. *Genome Announc*. 4(1):e01563-15. doi:10.1128/genomeA.01563-15

Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., Six, C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine Synechococcus. *The ISME J*. 8:1221-1236.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*.

35(21):7188–7196.doi:10.1093/nar/gkm864
Puigbò, P., Lobkovsky, A.E., Kristensen, D.M., Wolf, Y.I., Koonin, E.V. (2014). Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biology*. 12(1):66.
Qin, Q.L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *Journal of Bacteriology*. 196(12):2210-2215.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*. 41:D590–D596.doi:10.1093/nar/gks1219
Rambaut, A.(2015). FigTree v1.4.2: Tree figure drawing tool. Available at: http://tree.bio.ed.ac.uk/software/figtree/
Ramos, V., Morais, J., Vasconcelos, V.M. (2017). A curated database of cyanobacterial strains relevant for modern taxonomy and phylogenetic studies. *Scientific Data*. 4:170054. doi:10.1038/sdata.2017.54
Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R.Y. (1979). Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. *J. General Microb.* 111(1):1–61.
Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*. 424(6952):1042–1047.
Sarma, T.A. (2012). *Handbook of Cyanobacteria*, Boca Raton, FL: CRC Press.
Scanlan, D.J. (2003). Physiological diversity and niche adaptation in marine Synechococcus. *Advances in Microbial Physiology*. 47:1–64.
Scanlan, D.J., and West, N.J. (2002). Molecular ecology of the marine cyanobacteria genera Prochlorococcus and Synechococcus. *FEMS Microbiology Ecology*. 40:1–12.
Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev*. 73:249–299.
Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Nat. Acad. Scienc. USA*. 110(3):1053–1058.
Schirrmeister, B.E., Anisimova, M., Antonelli, A., Baghri, H.C. (2011). Evolution of cyanobacterial morphotypes: Taxa required for improved phylogenomic approaches. Communicative and Integrative Biology. 4(4):424–427. doi: 10.4161/cib.4.4.16183
Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., et al. (2014). An expanded genomic representation of the phylum Cyanobacteria. *Genome Biol. Evol*. 6(5):1031–1045.doi:10.1093/gbe/evu073
Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21):2688–2690.
Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science*. 348(6237):1261359-1–1261359-9. doi: 10.1126/science.1261359
Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing

divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*. 56(4):564–77.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*. 30(12): 2725–2729.

Thompson, C.C., Silva, G.G., Vieira, N.M., Edwards, R., Vicente, A.C., Thompson, F.L. (2013). Genomic Taxonomy of the Genus Prochlorococcus. *Microbial Ecology*, 66(4):752-762. doi: 10.1007/s00248-013-0270-8

Thompson, C.C., Chimetto, L., Edwards, R.A., Swings, J., Stackebrandt, E., Thompson, F.L. (2013). Microbial genomic taxonomy. *BMC Genomics*. 14:913.

Thompson, C.C., Amaral, G.R., Campeão, M., Edwards, R.A., Polz, M.F., Dutilh, B.E., et al. (2015). Microbial taxonomy in the post-genomic era: Rebuilding from scratch? *Archives of Microbiology*. 197(3):359–370.

Tyrrell, T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature*. 400:525–531.

Wu, M., and Eisen, J. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol*. 9(10):R151.

Wu, M., and Scott, A.J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics*. 28(7):1033–1034.

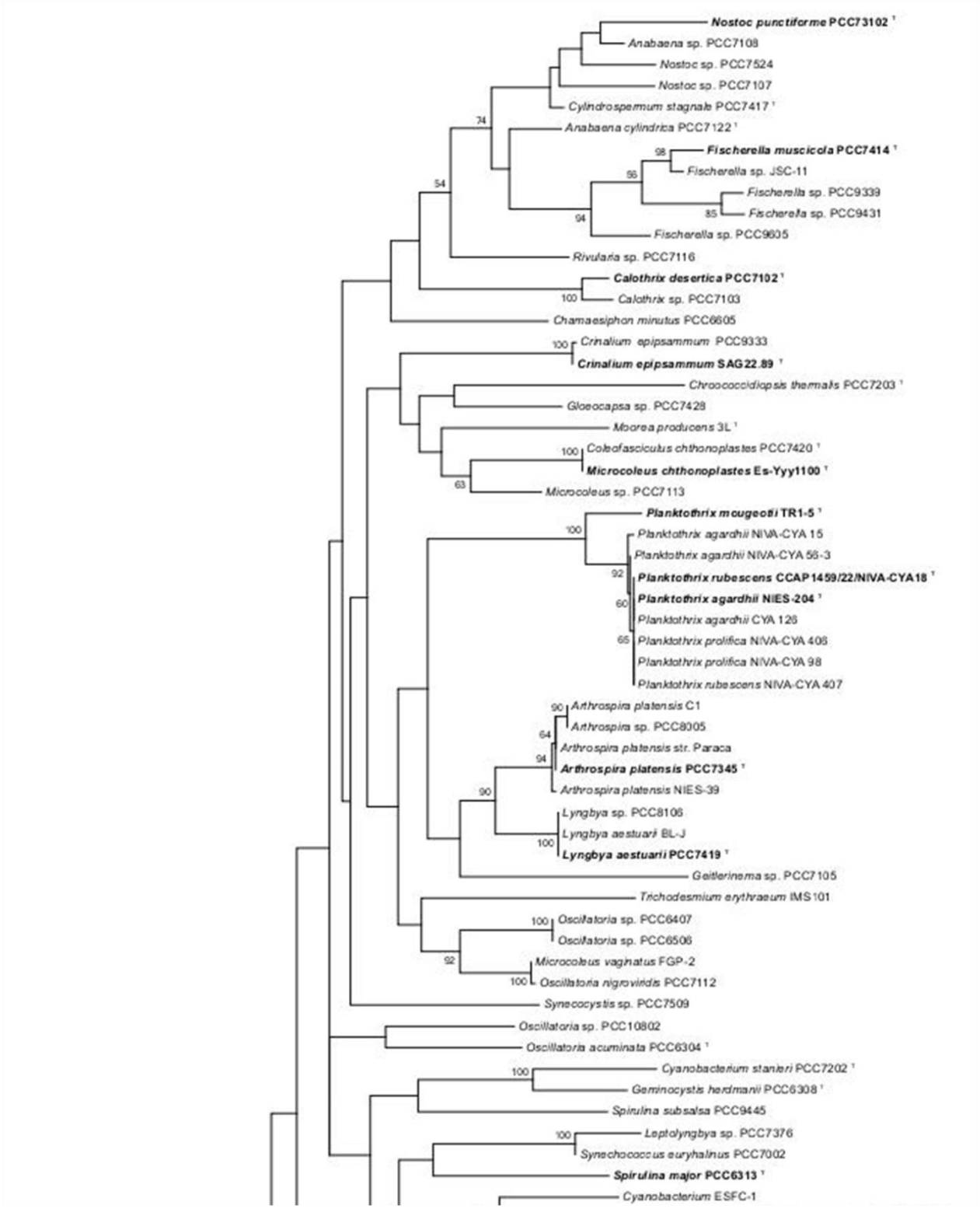
Yarza, P., Richter, mM., Peplies, J., Euzeby, J., Amann, R., Schleifer, K.-H., et al. (2008). The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Systematic and Applied Microbiology*. 31(4):241–250.

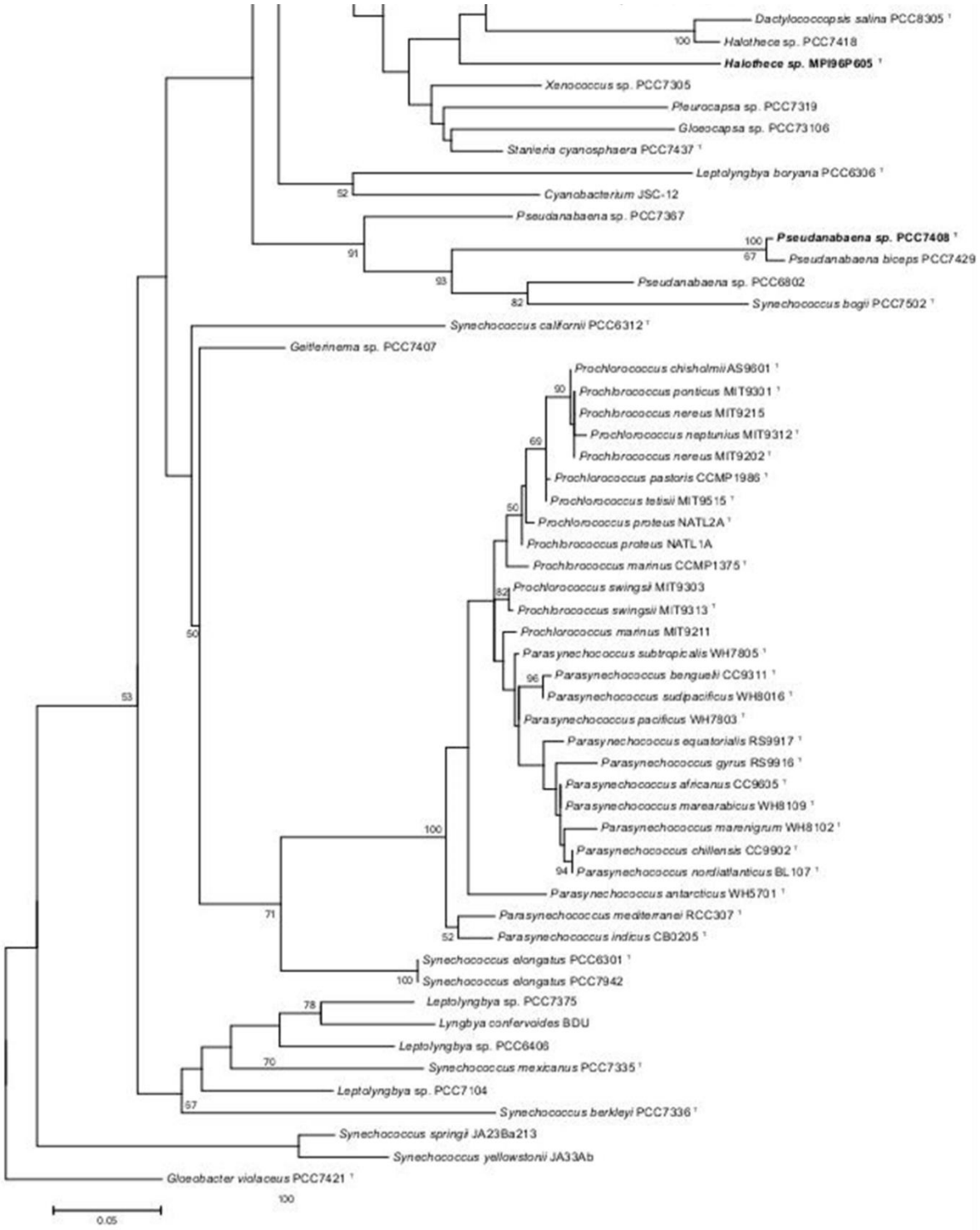
Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* 19:162–173.

## Author Contributions

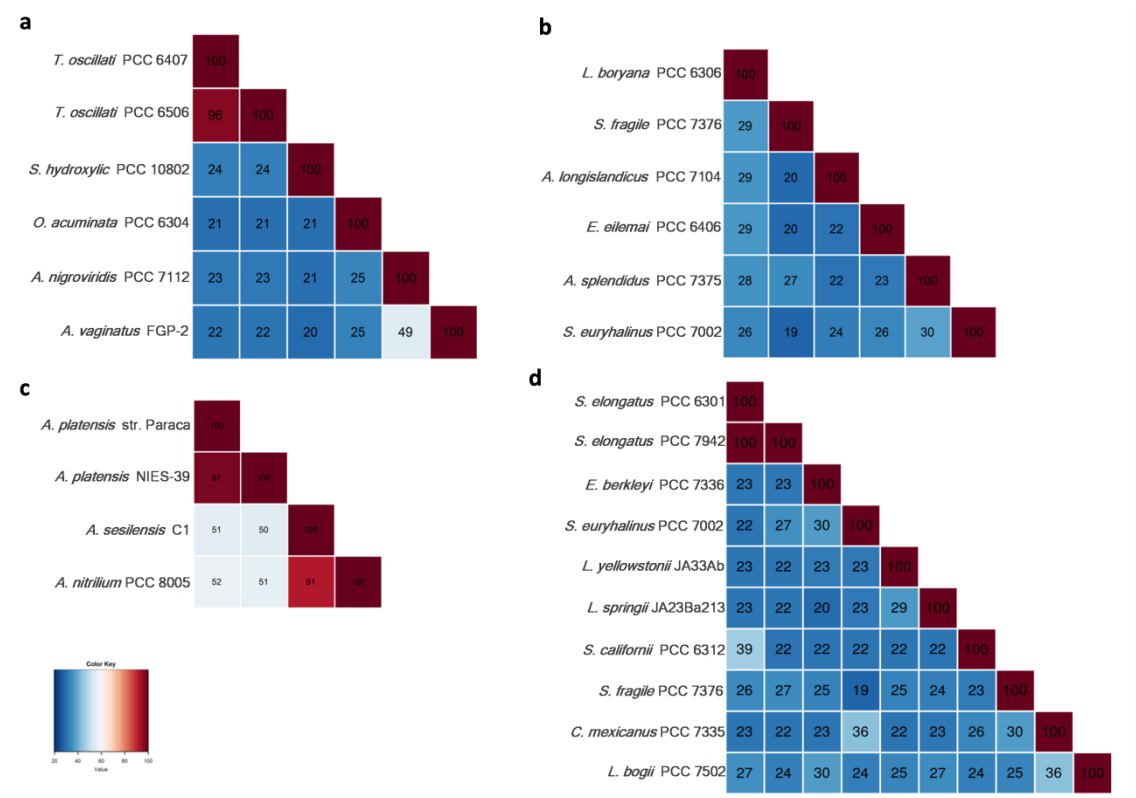
All authors contributed to the writing of the manuscript. JW, FC, BD, JS, FT and CT designed and planned the study. JW and FC performed the bioinformatics analyses. JW and FC analyzed the results and compiled the data. All authors approved the final version of the manuscript.

## Supplementary Information

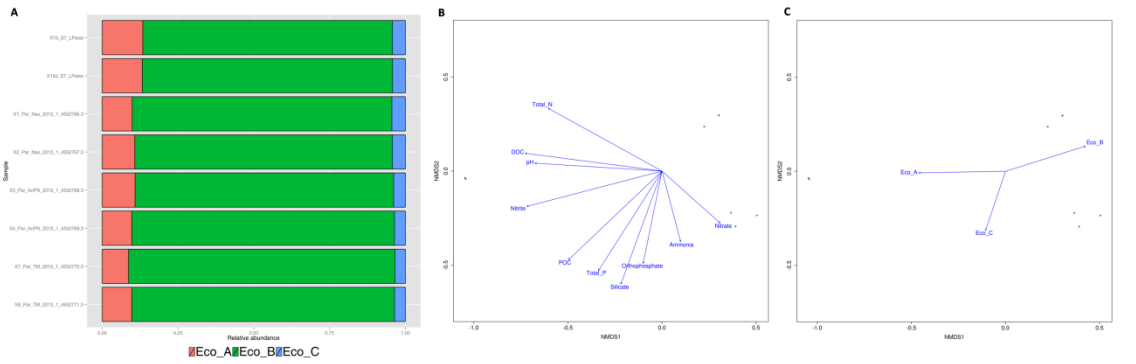





**Supplementary Figure S1. Ribosomal phylogenetic reconstruction of the Cyanobacteria phylum.** Tree was constructed through ML using the Kimura 2-parameter method, and GTR+G substitution model. Tree was inferred from 110 16S rRNA gene sequences (~ 1,400 bp). The species cut-off was 98.8 % similarity (Thompson et al., 2014). The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. Nodes supported with a bootstrap of ≥ 50 % are indicated. Overwritten T indicates type strain. Bold names indicate the included type strains. The unit of measure for the scale bars is the number of nucleotide substitutions per site. *Gloeobacter violaceus* PCC 7421 sequence was designated as outgroup.




**Supplementary Figure S2. Heatmaps based on GGD metrics of especific cases.** (A) Heatmap of GGD values between *Oscillatoria* group (case I); (B) Heatmap of GGD values between *Leptolyngbya* group (case II); (C) Heatmap of GGD values between *Arthrospira* group (case III); and (D) Heatmap of GGD values between *Synechococcus* group (case IV). The intraspecies limit is assumed as ≥ 70 % GGD. The proposed new names were adopted in this figure.




**Supplementary Figure S3. Abundance and distribution of ecogenomic clusters across freshwater metagenomes.** (A) Relative abundance of ecogenomic clusters in Caatinga biome (metagenomes, N = 8). (B) Non-metric multidimensional scaling (NMDS) analysis of the freshwater metagenomes and environmental parameters. Ordination plot of physicochemical parameters. Dots indicate the metagenomes samples. Distances were calculated based on the Bray-Curtis Method. NMDS stress value = 0.15. (C) Non-metric multidimensional scaling (NMDS) analysis of the freshwater metagenomes and environmental parameters Ordination plot of

ecogenomic clusters. Dots indicate the metagenomes samples. Distances were calculated based on the Bray-Curtis Method. NMDS stress value = 0.15.

**Supplementary Table S1. Estimates of genome relatedness of cyanobacterium strains.** Values at the matrix indicates the intergenomic distances (i.e. evolutionary divergence between sequences). The numbers of base substitutions per site between sequences are shown. Analyses were conducted accordingly Tamura et al. (2004) method. The analysis involved 110 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 759 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. [*Online Material*]

**Supplementary Material. Formal description of new genera and species:**

**Formal description of new genera and species**

**Description of *Pseudogeitlerinema* gen. nov.**
*Pseudogeitlerinema* (Pseudo.gei.tle.ri.ne.ma. Gr. adj. *pseud*, false*; geitlerinema*, other cyanobacteria genus [cf. genus *Geitlerinema*]; M. L. fem. n. *Pseudogeitlerinema*).

Description of *Pseudogeitlerinema shalloid* (formerly known as *Geitlerinema* sp.) (shal.loid. Gr. *-oid*, resembling, having the appearance of; *shalloid* referring to the appearance of very shallow adjacent cells): This species is characterized by β-carboxysome. Type strain is PCC 7407. The genome of this strain contains 4,68 Mbp (GC = 58,46%) harboring 3,727 coding DNA sequences.

**Description of *Somacatellium* gen. nov.**
*Somacatellium* (Som.a.ca.te.llium. Gr. n. *soma* (*somatiko*), body; L. fem. n. *catella*, a small chain; *Somacatellium* referring to a small chain filamentous body).

*Somacatellium hydroxylic* (formerly known as *Oscillatoria* sp.) (hy.drox.y.lic. Gr. *-ic (ikos)*, relating to or having some characteristic of; *hydroxylic* referring to the hydroxyl-proline and hydroxyl aspartic acid compounds the new cinnamycin variant found in this strain): This species is characterized by β-carboxysome. Type strain is PCC 10802. The genome of this strain contains 8,59 Mbp (GC = 54,1%) harboring 7,012 coding DNA sequences.

**Description of *Tapinonema* gen. nov.**
*Tapinonema (ta.pi.no.nema. Gr. tapino small*, modest, weak; Gr. n. *nema*, thread; *Tapinonema*, small filament).

Description of *Tapinonema colecalium* (formerly known as cyanobacterium) (col.e.ca.lium. L. *cole*, inhabit; Gr. *calo*, beautiful; *–ium*, quality or relationship; *colecalium*, beautiful inhabitant): This species is characterized by β-carboxysome. Type strain is JSC-12, isolated from freshwater habitat. The genome of this strain contains 5,52 Mbp (GC = 47,49%) harboring 5,024 coding DNA sequences.

**Description of *Toxinema* gen. nov.**
*Toxinema* (tox.i.nema. N.L. n. *toxi* toxins related to the strains; Gr. neut. n. *nema* thread, filament; N.L. neut. n. *Toxinema* a filament with release toxins).

Description of *Toxinema nigroviridis* comb. nov. (formerly known as *Oscillatoria nigroviridis*) (nig.ro.vi.ri.dis. Gr. adj. *nigro*, distinctus; Gr. adj. *viridis*, campus, a green field; *Oscillatoria* is other genus of cyanobacteria [cf. genus *Oscillatoria*]): This species

is characterized by β-carboxysome. Type strain is PCC7112, isolated from a soil habitat in USA. The genome of this strain contains 7,47 Mbp (GC = 45,87%) harboring 6,925 coding DNA sequences.

Description of *Toxinema vaginatus* comb. nov. (formerly known as *Microcoleus vaginatus*) [cf. genus species *Microcoleus vaginatus*]: This species is characterized by β-carboxysome. Type strain is FGP-2, isolated from soil habitat in Canyonlands National Park, UT, USA. The genome of this strain contains 6,69 Mbp (GC = 46,04%) harboring 5,519 coding DNA sequences.

Description of *Toxinema oscillati* (formerly known as *Oscillatoria* sp.) (os.ci.lla.ti. Gr. o*scillare*, to swing, *Oscillatoria* is other genus of cyanobacteria [cf. genus *Oscillatoria*]): This species is characterized by β-carboxysome. Type strain is PCC 6407, isolated from freshwater habitat. The genome of this strain contains 6,89 Mbp (GC = 43,43%) harboring 5,693 coding DNA sequences.

**Description of *Stenotopis* gen. nov.**
*Stenotopis* (Ste.no.to.pis. Gr. adj. *stenos*, narrow; Gr. n. *topos*, place; *Stenotopis* referring a living within a narrow range of places).

Description of *Stenotopis californii* comb. nov. (formerly known as *Synechococcus californii*) (Ste.no.to.pis. Gr. adj. *stenos*, narrow; Gr. n. *topos*, place; *Stenotopis* referring a living within a narrow range of places): This species is characterized by β-carboxysome. Type strain is PCC 6312, isolated from freshwater habitat in California, USA. The genome of this strain contains 3.72 Mbp (GC = 48.49%) harboring 3,795 coding DNA sequences.

**Description of *Euryforis* gen. nov.**
*Euryforis* (Eur.y.fo.rys. Gr. masc. n. *eury*, broad; N.L. masc. n. *foris*, outside; N.L. masc. n. *Euryforis* a *habitat* widespread genus).

Description of *Euryforis confervoides* comb. nov. (formerly known as *Lyngbya confervoides*) [cf. species *L. confervoides*]: This species is characterized by β-carboxysome. Type strain is BDU, isolated from a NA habitat in NA. The genome of this strain contains 8,79 Mbp (GC = 55,63%) harboring 8,370 coding DNA sequences.

Description of *Euryforis eilemai* (formerly known as *Leptolyngbya* sp.) (ei.le.mai. N.L. gen. n. *eilemai*, referring to the presence of sheath): This species is characterized by β-carboxysome. Type strain is PCC 6406, isolated from freshwater habitat in California, USA. The genome of this strain contains 5,77 Mbp (GC = 55,18%) harboring 5,156 coding DNA sequences.

**Description of *Allonema* gen. nov.**
*Allonema* (All.o.nema. Gr. adj. *allos*, other; Gr. n. *nema*, thread; M. L. fem. n. *Allonema*, other filament).

*Allonema longislandicus* (formerly known as *Leptolyngbya* sp.) (long.is.lan.di.cus. referring to the habitat collected, Long Island, NY, USA): This species is characterized by β-carboxysome. Type strain is PCC 7104, isolated from marine habitat rockshore, Montauk Point, Long Island, NY, USA. The genome of this strain contains 6,89 Mbp (GC = 57,69%) harboring 6,414 coding DNA sequences.

**Description of *Coccusdissimilis* gen. nov.**
*Coccusdissimilis* (Co.ccus.di.ssi.mi.lis. Gr. n. *kokkos*, grain or kernel; L. adj. *dissimilis*, different; *Coccusdissimilis*, a different coccos).

Description of *Coccusdissimilis mexicanus* comb. nov. (formerly known as *Synechococcus mexicanus*) (coccusdissimulis. Gr. n. *kokkos*, grain or kernel; L. adj. *dissimilis*, different; *Coccusdissimilis*, a different coccos): This species is characterized by β-carboxysome. Type strain is PCC 7335, isolated from marine habitat in Snail shell, intertidal zone, Puerto Penasco, Mexico. The genome of this strain contains 5.97 Mbp (GC = 48.20%) harboring 5,702 coding DNA sequences.

**Description of *Cyclospexia* gen. nov.**
*Cyclospexia* (Cy.clos.pex.ia. Gr. n. kyklos, cycle; Gr. *pexia* (adj. *-pectic*), fixation; n. *Cyclospexia* involved in N2 fixation cycle).

*Cyclospexia valenium* (formerly known as cyanobacterium) (val.e.nium. L. *valens*, strong; L. suff. *-icus -a -um*, suffix used with the sense of pertaining to; *valenium* referring to an organism restricted to few types of environmental conditions, such as extremophylic habitat): This species is characterized by β-carboxysome. Type strain is ESFC-1, isolated from marine habitat in extremophylic mat communities, Elkhorn Slough estuary, CA, USA. The genome of this strain contains 5.62 Mbp (GC = 46.51%) harboring 4,857 coding DNA sequences.

**Description of *Paraspirulina* gen. nov.**
*Paraspirulina* (Par.as.pi.ru.li.na. Gr. adj. *para*, to bring forth, to bear, alongside; *Spirulina*, other genus of cyanobacteria [cf. genus *Spirulina*]; M.L. fem. n. *Paraspirulina*).

Description of *Paraspirulina subsalsa* comb. nov. (formerly known as *Spirulina subsalsa*) [c.f. *S. subsalsa*]: This species is characterized by β-carboxysome. Type strain is PCC 9445. The genome of this strain contains 5,32 Mbp (GC = 47,39%) harboring 4,580 coding DNA sequences.

**Description of *Symphothece* gen. nov.**
*Symphothece* (Sym.pho.the.ce. Gr. *sympho*, to grow together; Gr. fem. n. *theke*, case, envelope; *Symphothece*, cells-box that growing together).

Description of *Symphothece euryhalinus* comb. nov. (formerly known as *Synechococcus euryhalinus*) [c.f. *S. euryhalinus*]: This species is characterized by β-carboxysome. Type strain is PCC 7002. The genome of this strain contains 3.41 Mbp (GC = 49.16%) harboring 3,121 coding DNA sequences.

Description of *Symphothece fragile* (formerly known as *Leptolyngbya* sp.) (fra.gi.le. *fragile*, sensitive, delicate): This species is characterized by β-carboxysome. Type strain is PCC 7376, isolated from marine habitat in Limestone, Crystal Cave, Bermuda. The genome of this strain contains 5.12 Mbp (GC = 43.87%) harboring 4,601 coding DNA sequences.

***Description of Leptococcus gen. nov.***
*Leptococcus* (Lept.o.co.ccus. Gr. adj. *leptos,* delicate or thin; N.L. masc. n. coccus *[*from Gr. Masc. n. *kokkos*, grain, seed, kernel]; N.L masc. n. *Leptococcus*, a delicate coccos).

Description of *Leptococcus springii* com. nov. (formerly known as *Synechococcus springii*) [c.f. *S. springii*]: This species is characterized by β-carboxysome. Type strain is JA23Ba213, isolated from freshwater habitat in Octopus Spring, Yellowstone Park, USA. The genome of this strain contains 3.05 Mbp (GC = 58.50%) harboring 3,064 coding DNA sequences.

Description of *Leptococcus yellowstonii* comb. nov. (formerly known as *Synechococcus yellowstonii*) [c.f. *S. yellowstonii*]: This species is characterized by β-carboxysome. Type strain is JA33Ab, isolated from freshwater habitat in Octopus Spring, Yellowstone Park, USA. The genome of this strain contains 2.93 Mbp (GC = 60.20%) harboring 3,036 coding DNA sequences.

**Description of *Eurycoccus* gen. nov.**
*Eurycoccus* (Eur.y.co.ccus. Gr. adj. *eury*, wide, broad; N.L. masc. n. *coccus [*from Gr. Masc. n. *kokkos*, grain, seed, kernel]; N.L masc. n. *Eurycoccus*, a wide coccus).

Description of *Eurycoccus berkleyi* comb. nov. (formerly known as *Synechococcus berkleyi*) [c.f. *S. berkleyi*]: This species is characterized by β-carboxysome. Type strain is PCC 7336, isolated from marine habitat in Sea Water Tank, Berkeley University, CA, USA. The genome of this strain contains 5.07 Mbp (GC = 53.70%) harboring 5,093 coding DNA sequences.

**Description of *Paraleptovivax* gen. nov.**
*Paraleptovivax* (Pa.ra.lep.to.vi.vax. *para*, to bring forth, to bear, alongside; Gr. adj. *leptos,* delicate or thin; L. adj. *vivax,* long-lived, tenacious of life; *Leptovivax*, delicate long-lived).

Description of *Paraleptovivax allomegium* (formerly known as *Pseudoanabaena* sp.) (all.o.me.gium. Gr. adj. *allos*, other; Gr. nom. neut. adj. m*ega*, big; Gr. *–ium*, quality or relationship): This species is characterized by β-carboxysome. Type strain is PCC 6802, isolated from freshwater habitat in California, USA. The genome of this strain contains 5.62 Mbp (GC = 47.83%) harboring 5,363 coding DNA sequences.

**Description of *Leptovivax* gen. nov.**
*Leptovivax* (Lept.o.vi.vax. Gr. adj. *leptos,* delicate or thin; L. adj. *vivax,* long-lived, tenacious of life; *Leptovivax*, delicate long-lived).

Description of *Leptovivax bogii* (formerly known as *Synechococcus bogii*) [c.f. *S. bogii*]: This species is characterized by β-carboxysome. Type strain is PCC 7502, isolated from a sphagnum bog habitat. The genome of this strain contains 3.58 Mbp (GC = 40.60%) harboring 3,703 coding DNA sequences.

**Description of *Calotaxis* gen. nov.**
*Calotaxis* (Ca.lo.ta.xis. Gr. *calo*, beautiful; Gr. *taxis*, arrangement; *Calotaxis*, beautiful arrangement of the filament).

Description of *Calotaxis gracile* (formerly known as *Pseudoanabaena* sp.) (gra.ci.le, L. neut. adj. *gracile*, slender): This species is characterized by β-carboxysome. Type strain is PCC 7367, isolated from marine habitat in intertidal zone, Mexico. The genome of this strain contains 4.55 Mbp (GC = 46.31%) harboring 3,960 coding DNA sequences.

**Description of *Acrophormium* gen. nov.**
*Acrophormium* (A.cro.phor.mium. Gr. *acro*, high, top; Gr. *phorm*, basket, mat; *-um*, suffix used with the sense of pertaining to; *Acrophormium*, living on the top of the mat).

Description of *Acrophormium splendidus* (formerly known as *Leptolyngbya* sp.) (splen.di.dus. L. adj. *splendidus*, brilliant): This species is characterized by β-carboxysome. Type strain is PCC 7375, isolated from marine habitat in Woods Hole,

Massachusetts, USA. The genome of this strain contains 9,42 Mbp (GC = 47,62%) harboring 8,366 coding DNA sequences.

**Description of *Eurycolium* gen. nov.**
*Eurycolium* (Eur.y.co.lium. Gr. adj. *eury*, wide, broad; L. *cole*, inhabit; Gr. *–ium*, quality or relationship, *Eurycolium* referring to the spread inhabiting trait in marine habitats).

Description of *Eurycolium pastoris* (formerly known as *Prochlorococcus pastoris*) [c.f. *P. pastoris*]: This species is characterized by α-carboxysome. Type strain is CCMP 1986 (or MED4), isolated from marine habitat. The genome of this strain contains 1,65 Mbp (GC = 30,8%) harboring 1,777 coding DNA sequences.

Description of *Eurycolium tetisii* (formerly known as *Prochlorococcus tetisii*) [c.f. *P. tetisii*]: This species is characterized by α-carboxysome. Type strain is MIT 9515, isolated from marine habitat. The genome of this strain contains 1,7 Mbp (GC = 30,79%) harboring 1,784 coding DNA sequences.

Description of *Eurycolium neptunius* (formerly known as *Prochlorococcus neptunius*) [c.f. *P. neptunius*]: This species is characterized by α-carboxysome. Type strain is MIT 9312, isolated from marine habitat. The genome of this strain contains 1,7 Mbp (GC = 31,21%) harboring 1,815 coding DNA sequences.

Description of *Eurycolium ponticus* (formerly known as *Prochlorococcus ponticus*) [c.f. *P. ponticus*]: This species is characterized by α-carboxysome. Type strain is MIT 9301, isolated from marine habitat. The genome of this strain contains 1,64 Mbp (GC = 31,34%) harboring 1,774 coding DNA sequences.

Description of *Eurycolium nereus* (formerly known as *Prochlorococcus nereus*) [c.f. *P. nereus*]: This species is characterized by α-carboxysome. Type strain is MIT 9202, isolated from marine habitat. The genome of this strain contains 1,69 Mbp (GC = 31,1%) harboring 1,795 coding DNA sequences.

Description of *Eurycolium chisholmii* (formerly known as *Prochlorococcus chisholmii*) [c.f. *P. chisholmii*]: This species is characterized by α-carboxysome. Type strain is AS9601, isolated from marine habitat. The genome of this strain contains 1,66 Mbp (GC = 31,32%) harboring 1,769 coding DNA sequences.

**Description of *Prolificoccus* gen. nov.**
*Prolificoccus* (Pro.li.fi.co.ccus. L. *prolificus*, productive, abundant, numerous; *Prolificoccus*, referring to an abundant coccus).

Description of *Prolificoccus proteus* comb. nov. (formerly known as *Prochlorococcus proteus*) [c.f. *P. proteus*]: This species is characterized by α-carboxysome. Type strain is NATL2A, isolated from a marine habitat in Northern Atlantic (10 m depth). The genome of this strain contains 1,84 Mbp (GC = 35,12%) harboring 1,930 coding DNA sequences.

**Description of *Expansiococcus* gen. nov.**
*Expansiococcus* (Ex.pan.sio.co.ccus. L. *expansio*, increase, expansion; *Expansiococcus*, referring to an organism that occupy large extensions).

Description of *Expansiococcus swingsii* comb. nov. (formerly known as *Prochlorococcus swingsii*) [c.f. *P. swingsii*]: This species is characterized by α-carboxysome. Type strain is MIT 9313, isolated from a marine habitat in Gulf Stream

(135 m depth). The genome of this strain contains 2,41 Mbp (GC = 50,74%) harboring 2,339 coding DNA sequences.

**Formal description of new species (within known genera)**

***Arthrospira genus***

Description of *Arthrospira nitrilium* (formerly known as *Arthrospira* sp.) (ni.tri.lium. N.L. gen. n. *nitrilium*, referring to the capacity to utilize nitriles (R-C≡N) as the sole source of nitrogen): this species is characterized by β-carboxysome and the phycobilisome pigmentation of this strain has not been characterized. Type strain is PCC 8005, isolated from a freshwater habitat in alkaline salt lakes. The genome of this strain contains 6,27 Mbp (GC = 44,7%) harboring 5,171 coding DNA sequences.

Description of *Arthrospira sesilensis* (formerly known as *Arthrospira platensis*) (ses.i.len.sis. N.L. gen. n. *sesilensis*, referring to the lack of gliding): this species is characterized by β-carboxysome and the phycobilisome pigmentation of this strain has not been characterized. Type strain is PCC 9438 (or C1), isolated from a freshwater habitat in alkaline salt lakes. The genome of this strain contains 6 Mbp (GC = 44,69%) harboring 4,852 coding DNA sequences.

***Geitlerinema genus***

Description of *Geitlerinema catellasis* (formerly known as *Geitlerinema* sp.) (cat.e.lla.sis. L. fem. n. *catella*, referring to the filamentous small chain): This species is characterized by α-carboxysome. Type strain is PCC 7105, isolated in USA. The genome of this strain contains 6,15 Mbp (GC = 51,59%) harboring 4,735 coding DNA sequences.

***Coleofasciculus genus***

Description of *Coleofasciculus microcolis* (formerly known as *Microcoleus* sp.) (mi.cro.co' lis. Gr. adj. *mikros* small (or thin); Gr. n. *koleos* sheath; M.L. masc. n. *microcolis* referring to a small or thin sheath): This species is characterized by β-carboxysome. Type strain is PCC 7113, isolated from a soil habitat in San Francisco, California, USA. The genome of this strain contains 7,47 Mbp (GC = 46,21%) harboring 6,734 coding DNA sequences.

***Chroococcidiopsis genus***

Description of *Chroococcidiopsis thermolimnetic* (formerly known as *Gloeocapsa* sp.) (ther.mo.lim.ne.tic. Gr. adj. *thermos*, hot ; N.L. fem. adj. *limnetica* (from Gr. n. *limnê*), lake; N.L. gen. n. *thermolimnetic*, referring to the habitat, freshwater thermal springs): This species is characterized by β-carboxysome. Type strain is PCC 7428, isolated from freshwater habitat in moderate hot spring. The genome of this strain contains 5,43 Mbp (GC = 43,27%) harboring 5,254 coding DNA sequences.

Description of *Chroococcidiopsis synechocystis* (formerly known as *Synechocystis* sp.) (syn.e.cho.cys.tis, *synechocystis* referring to *Synechocystis* genus): This species is characterized by β-carboxysome. Type strain is PCC 7509, isolated from soil habitat in Switzerland. The genome of this strain contains 4,9 Mbp (GC = 41,67%) harboring 4,859 coding DNA sequences.

***Lyngbya genus***

Description of *Lyngbya hydrogenis* (formerly known as *Lyngbya aestuarii*) (hydr.o.ge.nis. *hydrogenis*, referring to the powerful hydrogen production of this strain): This species is characterized by β-carboxysome. Type strain is BL-J. The genome of this strain contains 6,87 Mbp (GC = 41,16%) harboring 5,597 coding DNA sequences.

***Anabaena genus***

Description of *Anabaena mossii* (formerly known as *Anabaena* sp.) (mos.si. *mossi*, referring to the isolation area, intertidal zone, Moss Beach, CA, USA): This species is characterized by β-carboxysome. Type strain is PCC 7108, isolated from marine habitat in intertidal zone, Moss Beach, CA, USA. The genome of this strain contains 5,9 Mbp (GC = 38,78%) harboring 5,169 coding DNA sequences.

Description of *Anabaena stagnale* (formerly known as *Cylindrospermum stagnale*) [c.f. *C. stagnale*]: This species is characterized by β-carboxysome. Type strain is PCC 7417, isolated from soil habitat in Stockholm, Sweden. The genome of this strain contains 7,61 Mbp (GC = 42,2%) harboring 6,127 coding DNA sequences.

Description of *Anabaena nostocii* (formerly known as *Nostoc* sp.) [c.f. genus *Nostoc*]: This species is characterized by β-carboxysome. Type strain is PCC 7524, isolated from freshwater habitat in hot spring, Amparai District, Maha Oya, Sri Lanka. The genome of this strain contains 7,61 Mbp (GC = 42,2%) harboring 5,326 coding DNA sequences.

Description of *Anabaena reyesii* (formerly known as *Nostoc* sp.) (rey.es.si. *reyessi*, referring to the isolation area, Point Reyes Peninsula, CA, USA): This species is characterized by β-carboxysome. Type strain is PCC 7107, isolated from freshwater habitat in Point Reyes Peninsula, CA, USA. The genome of this strain contains 6,32 Mbp (GC = 40,36%) harboring 5,200 coding DNA sequences.

Description of *Anabaena welwii* (formerly known as *Fischerella* sp.) (wel.wii. *welwii*, referring to the presence of welwitindolinone gene cluster, a hapalindole-type family of natural products): This species is characterized by β-carboxysome. Type strain is PCC 9431. The genome of this strain contains 7,16 Mbp (GC = 40,19%) harboring 6,104 coding DNA sequences.

Description of *Anabaena hapalii* (formerly known as *Fischerella* sp.) (hap.a.lii. *hapalii*, referring to the presence of hapalindole gene cluster, a hapalindole-type family of natural products): This species is characterized by β-carboxysome. Type strain is PCC 9339. The genome of this strain contains 8 Mbp (GC = 40,16%) harboring 6,720 coding DNA sequences.

Description of *Anabaena sesquitii* (formerly known as *Fischerella* sp.) (ses.qui.tii. *sesquitii*, referring to the presence of sesquiterpene gene cluster, encoding three proteins): This species is characterized by β-carboxysome. Type strain is JSC-11. The genome of this strain contains 5,38 Mbp (GC = 41,05%) harboring 4,627 coding DNA sequences.

Description of *Anabaena peptidasii* (formerly known as *Fischerella* sp.) (pep.ti.da.sii. *peptidasii*, referring to the presence of genes encoding peptidases, such as M16, characteristic of the group II bacteriocin gene clusters): This species is characterized by β-carboxysome. Type strain is PCC 9605, isolated from soil (limestone) habitat in Jerucham, Har Rahama, Israel. The genome of this strain contains 8,08 Mbp (GC = 42,61%) harboring 7,060 coding DNA sequences.

***Xenococcus genus***

Description of *Xenococcus lajollai* (formerly known as *Xenococcus* sp.) (la.jo.llai. *lajollai*, referring to the isolation area, La Jolla Aquarium, CA, USA): This species is characterized by β-carboxysome. Type strain is PCC 7305, isolated from marine

habitat in Aquarium, La Jolla, CA, USA. The genome of this strain contains 5,92 Mbp (GC = 39,68%) harboring 4,992 coding DNA sequences.

***Pleurocapsa* genus**
Description of *Pleurocapsa penascus* (formerly known as *Pleurocapsa* sp.) (pe.nas.cus. *penascus*, referring to the isolation area, Puerto Penasco): This species is characterized by β-carboxysome. Type strain is PCC 7319, isolated from marine habitat in Arizona Station, Gulf of California, Puerto Penasco, Mexico. The genome of this strain contains 7,38 Mbp (GC = 38,74%) harboring 4,516 coding DNA sequences.

***Microcoleus* genus**
Description of *Microcoleus franciscanus* (formerly known as *Microcoleus* sp.) (fran.cis.ca.nus. *franciscanus*, referring to the isolation area, San Francisco): This species is characterized by β-carboxysome. Type strain is PCC 7113, isolated from soil habitat in San Francisco, California, USA. The genome of this strain contains 7,47 Mbp (GC = 46,21%) harboring 6,734 coding DNA sequences.

***Dactylococcopsis genus***
Description of *Dactylococcopsis halotolerans* (formerly known as *Halothece* sp.) (ha.lo.to.le.rans*. Gr*. n. *hals*, *halos*, salt, sea; L. pres. part. *tolerans*, tolerating; M. L. part. adj. *Halotolerans* salt-tolerating): This species is characterized by β-carboxysome. Type strain is PCC 7418, isolated from freshwater habitat in Solar Lake, Israel. The genome of this strain contains 4,18 Mbp (GC = 42,92%) harboring 3,663 coding DNA sequences.

***Geminocystis genus***
Description of *Geminocystis stanieri* comb. nov. (formerly known as *Cyanobacterium stanieri*) [c.f. *C. stanieri*]: This species is characterized by β-carboxysome. Type strain is PCC 7202, isolated from freshwater habitat in thermal spring, alkaline pod. The genome of this strain contains 3,16 Mbp (GC = 38,66%) harboring 2,886 coding DNA sequences.

***Planktothrix genus***
Description of *Planktothrix stereotis* comb. nov. (formerly known as *Planktothrix agardhii*) (ster.eo.tis. Gr. *stereos*, solid; *stereotis*, referring to a solid and straight filament): This species is characterized by β-carboxysome. Type strain is NIVA-CYA 126, isolated from freshwater habitat. The genome of this strain contains 5,04 Mbp (GC = 39,57%) harboring 4,188 coding DNA sequences.

**Abbreviations**

adj. adjective; *Gr*. *Greek*; L. Latin; *n*. noun; neut. neutro; c.f. confer, conferre; v. verb; CDS protein-coding sequence; NA "Not Available" or "Not Applicable" or "Not Announced".

# Chapter V

# CONCLUDING REMARKS AND PERSPECTIVES

This thesis contributes information of direct value to our ultimate understanding of the modern coral reefs. Here, our blueprint of study area was the Abrolhos reefs. We bring knowlegment regarding to the abundance and functional roles of turf assemblages present in these reefs (Chapter II), and we extended our understanding of how the genomic content of an expressive bacteria living in – the cyanobacterial species named *Acrophormium turfae* (strains CCRM0081$^T$ and CCRM0082) – may allow these organisms enhance the turf persistence (Chapter III). The phylogenomic reconstruction for the newly strains raises issues about the current taxonomy for the whole Cyanobacteria phylum, which leading us to propose an ordered, coherent and practical frame for Cyanobacteria classification. It was supported by a broad identification of ecological associations within aquatic habitats that are responsible for delineating the ecological niches of Cyanobacteria (Chapter IV).

First, we shed light on the homogeneous functional core of turfs across the Abrolhos Bank, and their differentiated microbiome (when compared to corals, rhodoliths, and seawater), which holds diverse microbial guilds acting in synergism. It makes turf an energetic benthic component of Abrolhos. Oxygenic photosynthesis, AANP, sulfur cycle (S oxidation, and DMSP consumption), and nitrogen metabolism ($N_2$ fixation, ammonia assimilation, dissimilatory nitrate and nitrite ammonification) showed a conserved functional trait of the turf microbiomes (Chapter II).

Second, isolation, culture maintenance, sequencing, and genomic and physiological properties analysis of two cyanobacterial strains were the first attempt to improve our understanding on the functioning of turf-forming cyanobacteria from Abrolhos. Both isolates, named *Acrophormium turfae* CCRM0081$^T$ and CCRM0082

showed an unexpected frequency and diversity of secondary metabolites, high presence of phycobilisomes genes, and a surprising acclimative response to far-red light (enhanced absorption in PE and APC nm) (Chapter III). CCRM0081$^{T}$ had the second highest number of NRPS/PKS clusters. Findings from this chapter hints that *A. turfae* is likely a key species for turfs success in the reef system, where the potential to produce secondary metabolites might help to prevent the whole assemblage from being heavily impacted by grazers and by negatively affecting competing benthic organisms; the high number of photosynthesis-related genes and the high content of phycobiliproteins might help these Cyanobacteria to thrive under light limitation, a possible scenario in reef systems; and the acclimative response to far-red light raise PE and APC absorption, enhancing these organisms to grown efficiently with limitin conditions.

Finally, great effort was sued extending the phylogenomic analysis for the whole Cyanobacteria phylum, through an ecogenomic approach. Then, in the fourth chapter was shown a high coherence between the phylogenomic and the ecological delineation results applied to 100 cyanobacterial genomes, i.e. our results demonstrated that taxonomic affiliation and niche occupancy are linked, where closely related species of the same genus often shared correlation patterns, and consequently belong to the same ecogenomic group. It allows us to propose a new taxonomic framework for the Cyanobacteria phylum, intending to facilitate ecological studies in terms of evolutionary-lineage coherence of that group. This main and solid way to classify the established and the new incomings of strains is based on their genomes. Their accuracy and practical analysis matters because once we have a quickly growing number of described taxa, we maintains large gaps in our current knowledge of cyanobacterial biodiversity and distribution.

The knowledge obtained in this thesis constitute a valuable platform for future scientific progress, since filled some gaps concerning the proper definition and

ecological roles of one expressive puzzle of the Abrolhos coral reefs ecosystem – turf – helping forthcoming researches understand their persistence across the time, for example. Moreover, we propose a stepping stones towards a genome-based Cyanobacteria classification. Furthermore, studies including other isolates of the same collection of CCRM0081$^{T}$ and CCRM0082 strains will add more information to build a microbial benthic landscape of Abrolhos Bank. Finally, transcriptome sequencing and analysis of these Cyanobacteria in controlled conditions will answer some issues related to the control and regulation of photosynthetic machinaries, and the related persistance and success of turfs.

"There is a word for what Darwin and the rest of us have felt when in the presence of the reef: 'awe.' Confronted with the reef, awe is the most appropriate response. It is probably in our nature. It is also, apparently, in our nature to destroy that which we hold in awe." -- from The Enchanted Braid