UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The diversity of polyketides in the South Atlantic and Southern Ocean

By

Oluwatayo Ayotunde Makinde

18372008

Supervisor: Prof. Thulani Peter Makhalanyane

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor **(PhD) Genetics**

In the

Faculty of Natural & Agricultural Sciences

Department of Biochemistry, Genetics and Microbiology

University of Pretoria

September 2023

Declaration

I, Oluwatayo Ayotunde Makinde, declare that the thesis submitted to the Department of Biochemistry, Genetics and Microbiology at the University of Pretoria, Hatfield Campus, for the degree *Philosophiae Doctor* (PhD) in Genetics has never been submitted at the university or at any other institution elsewhere. This is my own work and all the sources used or quoted have been indicated and acknowledged.

_____

Oluwatayo Ayotunde Makinde

_____

Date

Plagiarism Declaration

Full names of student: Oluwatayo Ayotunde Makinde

Student number: 18372008

Topic of work: Diversity of polyketides in South Atlantic Ocean and Southern Ocean

Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this **thesis** is my own original work. Where other people's work has been used (either from a printed source, internet, or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

_____

SIGNATURE

Dedication

This research thesis is dedicated to God Almighty for his infinite mercy over my life during the course of this programme, and also to my late father Pa S.J Ola Makinde of blessed memory.

Acknowledgements

My profound gratitude goes to my brilliant supervisor Professor Thulani Makhalanyane and his crucial role in making this thesis happen. His constant support, challenging mindset and understanding have contributed significantly to this work. To the same extent, I am very grateful for the support from Doctor Oliver Bezuidt who was always willing to listen to my analysis rants and review my scientific texts and presentations. I cannot but thank Doctor Surendra Vikram for the time spent teaching me bioinformatics in the developmental stage of this thesis. Furthermore, I would like to thank all members of my thesis jury for accepting to be part of my final defense committee and their useful suggestions.

Special thanks to my colleagues at the Microbiome@UP and Centre for Microbial Ecology and Genomics (CMEG), all of whom have contributed to the person I am today particularly Percy, Marvellous, Smakalegen, Tidimalo, Mubanga, Mable, Choaro, Diego and Tiffiny for their assistance. I also thank Ms. Prudy Seepy for her supportive technical, secretarial, and administrative work.

I am thankful to the South African National Antarctic Programme (SANAP) for supporting field logistics and Adekunle Ajasin University Akungba-Akoko (AAUA), Nigeria, my home institution, for financial and career support during my studies.

I am immensely grateful to my family especially my siblings and my mother, Mrs. Felicia Mamisola Makinde. You have contributed so much to helping me get here in life. I'm sorry I missed so many of your celebrations during my studies. Thank you all for your love, encouragement, prayers and support.

To the big one, Okadazim Niella Makinde, my beautiful wife, words cannot describe my deep appreciation for your tender love, patience, prayers, intellectual support, commitment and sacrifices. You are the best!

Finally, I am supremely grateful to God Almighty for his ever faithfulness. My Rock and present help, thank you.

Abstract

Nature has long provided valuable pharmaceuticals, with many effective antibiotics originating from soil and plants. Yet, rising antibiotic resistance and repeated rediscovery of known compounds have redirected natural product drug discovery toward the ocean. Previous studies have shown that the extreme and oligotrophic environment substantially produce novel compounds. The significant pharmaceutical value of polyketides has sparked extensive research into their biosynthesis. The bulk of these studies have focused on the diversity of polyketides in an oligotrophic ocean, limited in iron and nitrogen, interpreting and understanding their role to the microbial community, by harnessing the available metagenomic tools. The knowledge deficit is especially true of South Atlantic Ocean (SAO) and Southern Ocean (SO), given that they are largely under sampled and pristine respectively. To reduce this knowledge deficit, we established a broad scale analysis of the SAO euphotic zone. We thereafter assessed the diversity of polyketide in the euphotic zone of SAO. We further assessed the role viruses play in the propagation of polyketides.

The analysis was carried out using the contextualize approach designed due to the continuous increase of metadata in databases. We further assessed the efficacy of the approach. Our analysis revealed there is high diversity of polyketides in the euphotic zone of SAO, with the majority encoding arylpolyene pathway suggesting its role in photo-protection. Most of the identified polyketides had low similarity to known BGCs, suggesting substantial functional novelty. The result also suggested virus may be playing a role in the propagation of the abundant polyketide type.

To elucidate the biosynthetic potential across the ocean depths, samples were taken from the epipelagic, mesopelagic and bathypelagic zones of the SO, in two seasons (Spring and Winter). Using metagenome-assembled genome approach, a total of 21 BGC types were recovered with

at least 8 different types of BGC in each zone. The highest diversity was observed in the bathypelagic zones of the two seasons. We observed terpene and PUFA to be the most abundant in the environment and are thought to play important roles in the ecosystem function. Interestingly, we discovered SAR324 as the most biosynthetically diverse microbe in SO.

Taken together, our findings substantially broaden insights regarding the diversity of polyketides in SAO and SO communities. We reveal how the continuous increase in metadata in databases can be harnessed for prospecting. These analyses demonstrate that the bathypelagic zone of SO harbours more diverse BGC types. The insights generated from the study provide an important basis for prospective researchers about the SAO and SO's biosynthetic potentials and ability to produce different specialized metabolites in different zones. In addition, the MAGs generated will fill some phylogenetic gaps in the available genome collection, making them useful for inferring details about microbial phylogenetic relationships.

Table of contents

Appendices

List of tables

List of figures

List of Abbreviations

ACP	Acyl carrier protein
AMR	Antimicrobial Resistance
ANI	Average Nucleotide Identity
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
AR	Antibiotic resistance
ARGs	Antibiotic resistance genes
AT	Acyltransferase
BGCs	Biosynthetic Gene Clusters
CAT	Contig annotation tool
CO_2	Carbon dioxide
DHs	Dehydratases
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ERs	Enoylreductases
ESI-MS	Electrospray Ionization Mass Spectrometry
FASs	Fatty acid synthases
GCF	Gene cluster family
GC-MS	Gas Chromatography Mass Spectrometry

GTDB-Tk	Genome Taxonomy Database Toolkit
HgIE-KS	Heterocyst specific glycolipids
HGT	Horizontal gene transfer
HMM	Hidden Markov Model
HNLC	High-nutrient low chlorophyll
HPLC	High Performance Liquid Chromatography
KS	Ketosynthases
KSα	alpha ketosynthase
KSβ	beta ketosynthase
KRs	Ketoreductases
MAGs	Metagenomic assembled genomes
MGEs	Mobile genetic elements
MIBiG	Minimum Information about a biosynthetic gene cluster
ML	Machine learning
MS	Mass Spectrometry
MTs	Methyltransferases
NaCl	Sodium chloride
NaPDoS	Natural Product Domain Seeker
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing

NMR	Nuclear Magnetic Resonance
NP	Natural product
NRPS	Non-ribosomal peptide synthetases
OMZ	Oxygen minimum zone
ORFs	Open reading frames
PCR	Polymerase chain reaction
PES	Polyethersulfone
PKs	Polyketides
PKS	Polyketide synthases
PKS I/T1PKS	Type I Polyketide synthases
PKS II/T2PKS	Type II Polyketide synthases
PKS III/T3PKS	Type III Polyketide synthases
PUFA	Polyunsaturated fatty acid
RiPPs	Ribosomally synthesized and post-translationally modified peptides
rRNA	Ribosomal ribonucleic acid
SAGs	Single amplified genomes
SAO	South Atlantic Ocean
SCFAs	Short chain fatty acids
SCG	Single-cell genomics
SDS	Sodium Dodecyl Sulfate

SO	Southern ocean
T1PKS-NRPS	Type I Polyketide synthase and non-ribosomal peptide synthase hybrid
TAE	Tris acetate EDTA
TAR	Transformation-associated recombination
TEs	Thioesterases
Tris-HCl	Tris (hydroxymethyl) aminomethane
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
UV-B	Ultraviolet B

Chapter One: Literature review

Metagenomic tools for unravelling novel polyketides

1.1 Introduction

Infectious diseases have shaped scientific discovery, the history of the world, and social evolution. Alexander Fleming's discovery of Penicillin, in 1928, is single-handedly one of the major discoveries of the 20th century and has vastly reduced the number of deaths from infectious diseases (Borchert, 2017, Zoffmann *et al.*, 2019). This discovery ushered in the 'golden age of antibiotics' during which numerous molecules used for treating various ailments were discovered (Metsä-Ketelä *et al.*, 2002). Ironically, these efforts were so successful, and the combination of declining net revenues resulted in the shift towards more profitable ventures. Pharmaceutical companies have nevertheless since diverted their attention to the discovery of drugs that focused on the treatment of lifelong illnesses (Von Bubnoff, 2006, Lewis, 2012). The regulatory challenges and costs associated with the production of new drugs (Wang *et al.*, 2020) has also led to a focus on treating chronic diseases through so called 'blockbuster drugs' (Luepke *et al.*, 2017). These drugs include recent high revenue generators such as Lipitor® that is used to lower bad cholesterol and reduce the risk of stroke, heart attack and other heart and blood vessel problem. Such drug has been more lucrative compared to drugs for treating bacterial infections (Hammel *et al.*, 2022).

Approximately half of the currently available antibiotic classes were discovered during the 'golden age of antibiotics'(Palazzotto *et al.*, 2019) [Figure 1]. Since then, resistance to antibiotics has increased, primarily due to increased selective pressure on pathogens and as a result of the overuse or misuse of these drugs (Davies & Davies, 2010). Consequently, an increasing number of pathogens have developed resistance to the currently available antibiotics (Golkar *et al.*, 2014).

For some infections, this has resulted in a situation akin to the "pre-antibiotic era" when common infections were life threatening (Ventola, 2015, Schalk, 2018). It has been predicted that, by 2050, antibiotic resistance will result in the deaths of over 10 million people if the current situation continues unabated (O'Neill, 2014, UN, 2019). This has led to the World Health Organisation (WHO) declaring antibiotic resistance and the resultant health implications a crisis, which requires urgent action and intervention (WHO, 2014). Recently, there has been a substantial increase in antibiotic discovery efforts from several leading pharmaceutical companies and researchers (Adu-Oppong *et al.*, 2017, Presterl *et al.*, 2019).

The traditional methods used for discovering novel antibiotics include the once-off fermentation-based screening of pure isolates (Hata *et al.*, 1971, Ikeda *et al.*, 1983, Medema & Fischbach, 2015, Katz & Baltz, 2016). This approach was highly valuable during the golden age of antibiotics and led to the development of several routinely used antibiotics (Katz & Baltz, 2016).

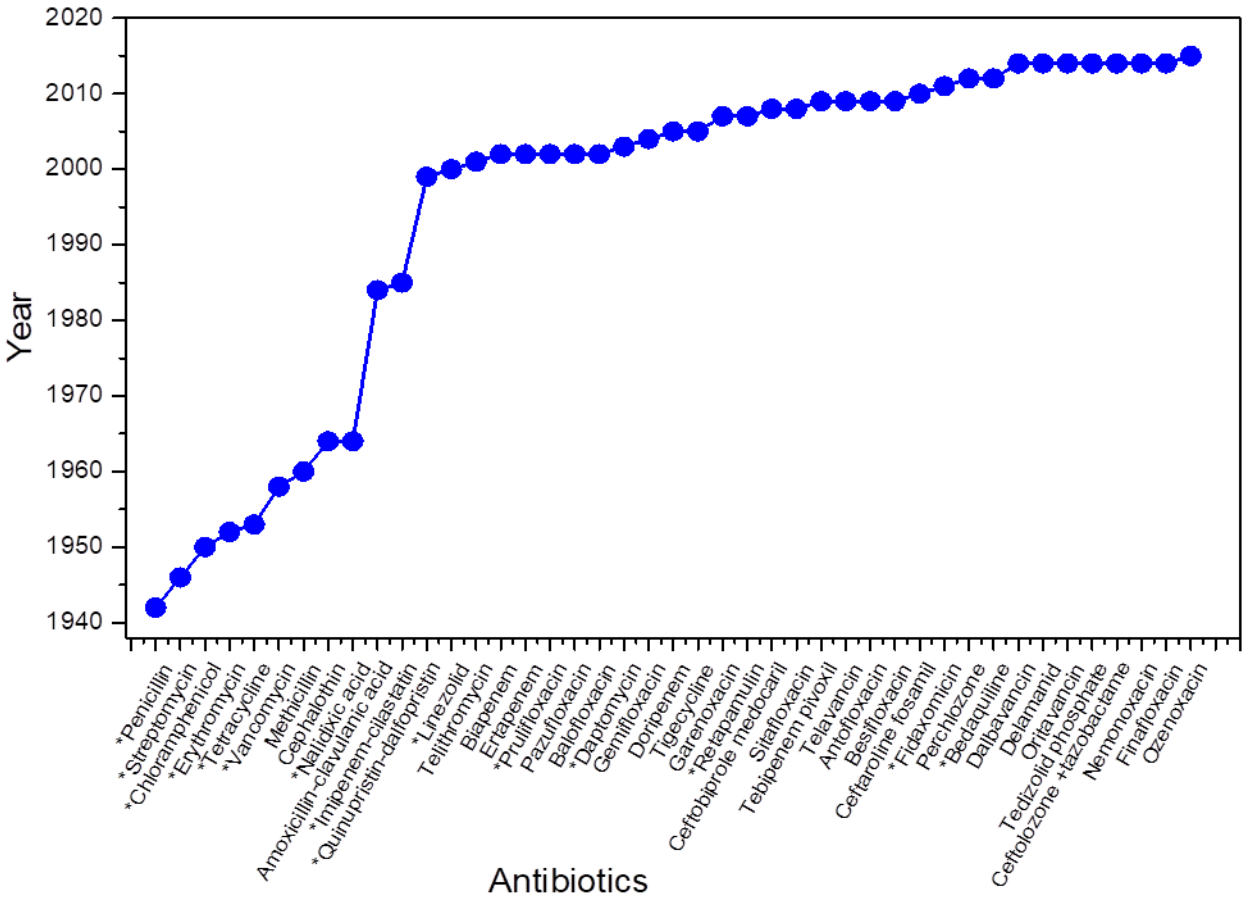

Figure 1.1: A timeline of antibiotic discovery. Between the 1930s and mid-1960, there was a rapid increase in the rate of novel antibiotic discovery. This golden age of antibiotics was followed by a period of low biodiscovery (only two targets between mid-1960 and 1999). Since 2000, there has been resurgence in the discovery of new antibiotics, although most are related to known classes of antibiotics. The asterisked antibiotics represent a new class. Antibiotic information compiled from Marston et al., 2016; Butler et al., 2016; Stuart et al., 2020. This image was created with Origin.

However, owing to a comparatively slow advance in our ability to reproduce microbial culturing conditions, for the >99% of organisms that are difficult to culture, this approach is arguably not as viable as it was at the outset (Hirsch *et al.*, 2010, Pham & Kim, 2012). Indeed, there is strong

evidence that culture-based approaches may fail to cultivate ecologically rare microorganisms with complex metabolisms (Vanwonterghem *et al.*, 2016, Nayfach *et al.*, 2021). Concomitantly, this approach also fails to detect cryptic or low-level expression of novel compounds produced in tandem with previously produced compounds (Foulston, 2019). However, two advances have provided access to so called uncultivatable organisms. The first of these advances is the emergence of culture-independent isolation technologies, which have expanded the catalogue of known microorganisms and resulted in the increased enumeration of various candidatus phyla (Hirsch *et al.*, 2010). The second is the rapid advancements linked to next generation sequencing (NGS) approaches, with the emergence of several platforms capable of providing inexpensive, long and accurate sequencing reads at comparatively low costs (Wang *et al.*, 2015, Chiu & Miller, 2016). Harnessing and integrating these advances to streamline the study of genomic diversity and identification of new compounds from natural products is key in the search for promising novel antibiotic compounds. This is because natural products are an unrivaled source of bioactive compounds (Demain, 2014, Newman & Cragg, 2016, Scherlach & Hertweck, 2020). The two most important groups of natural products are the polyketides and non-ribosomal peptides, because of their role in pharmaceutical and agricultural applications (Rego *et al.*, 2020).

This chapter aims to provide a general overview and synthesis of recent advances in studies on polyketides (PKs). Taking advantage of recent developments, we suggest an approach for the discovery of novel PKs from complex environments, using Ketosynthases (KS) as a probe via state-of-the-art bioinformatics tools. We also highlight how these tools may be complemented with the expression of biosynthetic gene clusters to determine function. We further discuss how the development of artificial intelligence tools, such as GECCO, may aid in the discovery of new

antibiotics. Finally, we discuss fruitful avenues for improving research on novel natural compounds.

1.2 Structure of polyketides

Approximately 70% of commercially developed anti-infectives are derived from natural products (Cragg & Newman, 2013, Newman & Cragg, 2016, Schneider, 2021). Polyketides are one of the most important classes of natural products (Rego *et al.*, 2020). These polyketides are complex compounds, encoded by polyketide synthase (PKS) genes (Hertweck, 2009, Klaus & Grininger, 2018). Architecturally, polyketide synthases are gene clusters organized in operons. The complexity of PKs is due to the sequential occurrence of various PKS catalytic domains (Yadav *et al.*, 2009, Shelest *et al.*, 2015, Selvin *et al.*, 2016). Molecular analysis has shown that PKs are structurally and functionally diverse and include antibacterial (erythromycin A), antifungal (amphotericin B), immunosuppressing (rapamycin), anticancer (epothilone B), anti-inflammatory (Flavonoids), antiviral (balticolid), anti-cholesterol (lovastatin), insecticide (spinosyn), and chemopreventive (resveratrol) functions amongst others (Gomes *et al.*, 2013, Risdian *et al.*, 2019, Singh et al., 2019) [Figure 1.2].

PKs are mainly produced by bacteria (Risdian *et al.*, 2019) for communication and defence (Flórez *et al.*, 2018, Rodríguez-Hernández *et al.*, 2019). Of these bacteria, members of the genus Streptomycetes are the most important and produce up to 75% of all antibiotics used in clinical practice (Janardhan *et al.*, 2014, Kang & Kim, 2021). The metabolic versatility of bacteria probably explains why they produce most known PKs. As a result, there is a need to search for novel PKs in uncharacterized bacteria, particularly those from complex extreme environments.

However, several studies have shown that other organisms including fungi (e.g., lovastatin), protists (e.g., maitotoxin-1), insects (e.g., stegobinone), mollusks (e.g., elysione), and plants (e.g., emodin) also produce PKs (Risdian *et al.*, 2019).

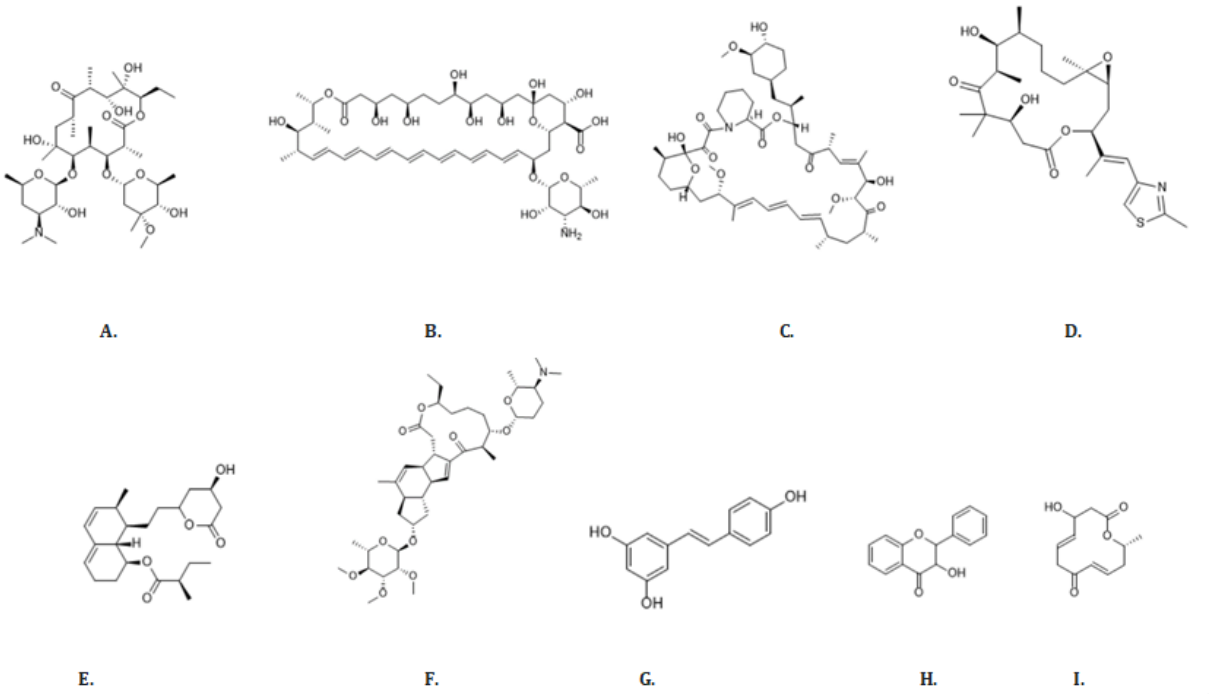

Figure 1.2: Structurally distinct polyketide compounds. A. an antibacterial compound (Erythromycin A). **B.** an antifungal compound (Amphotericin B). **C.** an immunosuppresor (Rapamycin). **D.** an anticancer compound (Epothilone). **E.** an anti-cholesterol compound (Lovastatin). **F.** an insecticide (Spinosyn). **G.** a chemopreventive compound (Resveratrol). **H.** a flavonoid (Flavanonol) and **I.** an antiviral compound (Balticolid).

PKS are multifunctional and multidomain mega-enzymes (Singh et al., 2019). These enzymes catalyse the assembly of the PKs skeleton, from coenzyme-A fatty esters (Gomes *et al.*, 2013), a function similar to fatty acid synthases (FAS) (Foerstner *et al.*, 2008, Shelest *et al.*, 2015). However, PKS are functionally different from FAS in terms of their capacity to use a wide range of structural components for the formation of various chain lengths (Hertweck, 2009). PKS are classified into three distinct types (I, II, and III) based principally on the architecture, mode of action and arrangement of their catalytic domains (Foerstner *et al.*, 2008, Hertweck, 2009, Beedessee *et al.*, 2019, Singh et al., 2019).

Type I PKS (PKS I) are mostly multi-modular enzymes with catalytic domains arranged in modules for selection, installation, and tailoring of each molecule (Risdian *et al.*, 2019). Each module catalyzes a single round of polyketide chain extension in a non-iterative way (Risdian *et al.*, 2019). However, PKS I in fungi mostly act iteratively (Chooi & Tang, 2012). Nonetheless, all PKS I have three core domains, the acyltransferase (AT), ketosynthase (KS), and acyl carrier protein (ACP) domains, which are the basic catalytic domains required by a module for polyketide chain elongation. The AT domain is responsible for the selection of the initial substrate (acyl unit) and transfers same to the ACP (Foerstner *et al.*, 2008, Shelest *et al.*, 2015). The ACP, contains the phosphopantetheinyl arm, and is responsible for the movement and activation of the acyl unit attached to the KS active site (Kurnia *et al.*, 2017, Singh et al., 2019). The KS domain catalyses the decarboxylative condensation reaction of the acyl units (Kornfuehrer & Eustáquio, 2019). Apart from these core domains, each module may contain one or more additional processing domains that will increase the complexity (Foerstner *et al.*, 2008, Kornfuehrer & Eustáquio, 2019, Singh et al., 2019) [Figure 1.3][Table 1.1]. The PKS I group is further differentiated based on three traits including whether it may be partially reducing (lacking KR, DH or ER optional reductive

domains), highly reducing (having all the optional reductive domains) and non-reducing (lacking all the optional reductive domains) (Fatema *et al.*, 2018). In addition, PKS I may be categorized as either cis-AT or trans-AT, depending on whether each module has its own AT domain (cis) or whether it possesses a distinct AT domain which serves all modules (Hertweck, 2009).

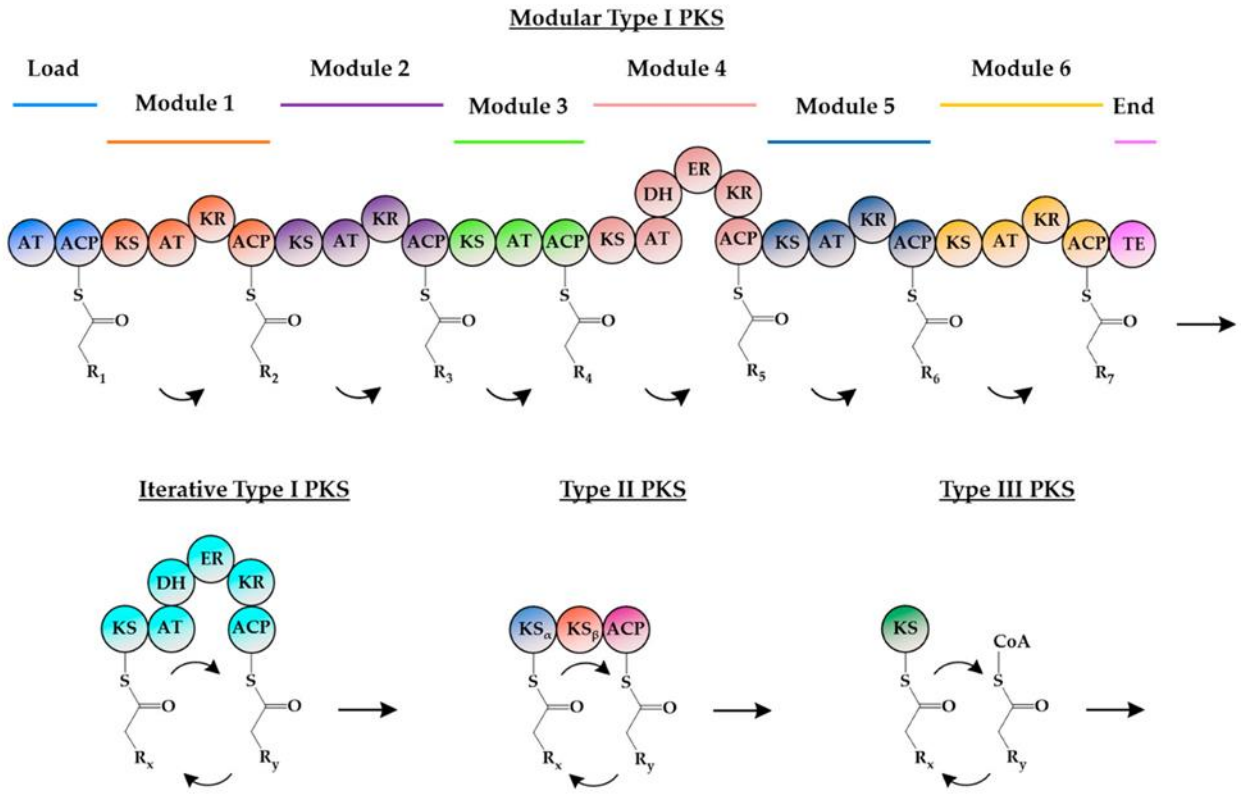

Figure 1.3: Domain organization of the different PKSs. Putative domains are represented by circles. In modular type I PKSs, functional domains are grouped into multiple modules, where each module handles a single decarboxylative condensation step in polyketide formation. Conversely, for iterative type I PKSs, the functional domains are consolidated into a single module, and each domain is reused multiple times during polyketide synthesis. Type II PKSs consist of separable multi-enzyme complexes, where each protein carries an independent catalytic domain

used iteratively in polyketide formation. Type III PKS reactions are also iterative but do not require an ACP for the attachment of the growing polyketide chain.

Key: AT: Acyltransferase; ACP: Acyl carrier protein; KS: Ketosynthase; KR: Ketoreductase; DH: Dehydratase; ER: Enoyl reductase; TE: Thioesterase; CoA: Coenzyme A. Image extracted from Lim et al., 2016.

On the other hand, type II PKSs **(**PKS II) are multifunctional enzyme-complexes (Bogdanova *et al.*, 2022). These complexes are responsible for the biosynthesis of PKs in an exclusively iterative manner (Singh et al., 2019). PKS II complexes consist of both mono- and bi-functional enzymes, which collectively synthesize PKs. These enzymes harbour key components, similar to PKS I, which are regarded as the minimal domains, sometimes referred to as "minimal PKS", associated with PKS II (Risdian *et al.*, 2019). These components include three enzymatic subunits: the alpha ketosynthase (KSα), beta ketosynthase (KSβ), and ACP (Shimizu *et al.*, 2017, Singh et al., 2019). KSα and KSβ form a heterodimer, and ACP serves as an anchor for prospective PKs by forming a thioester bond with the initial substrate (acyl CoA) (Singh et al., 2019). The KSα unit subsequently catalyses the condensation of the ACP-bound substrate (Risdian *et al.*, 2019). The KSβ, also regarded as the chain length factor, works in cooperation with the KSα to produce the poly β-keto ester or β-diketone depending on the substrates (Risdian *et al.*, 2019, Singh et al., 2019, Villebro *et al.*, 2019).

Table 1.1: Processing domains. Different domains that may be included in a single module with the activity they perform.

Domains	Function	Article
Ketoreductases (KRs)	Catalyzes the hydroxyl (OH) group formation	Baerga-Ortiz et al., 2006, Bonnett et al., 2013, Xie et al., 2016
Dehydratases (DHs)	Catalyzes the double bond formation following removal of water	Akey et al., 2010, Gay et al., 2013, Fiers et al., 2016, Barajas et al., 2019
Enoylreductases (ERs)	Catalyzes the conversion of double bond to single bond	Khare et al., 2015
Methyltransferases (MTs)	Catalyzes the modification of acyl units after condensation	Xiang, 2020
Thioesterases (TEs)	Catalyzes the release of the finished polyketide product	Zhou et al., 2015, Galea et al., 2017.

Other enzymes, apart from this "minimal PKS", which further transform the poly β-keto chain into an aromatic compound include cyclases and aromatases (Zhang *et al.*, 2017). This characteristic contributes to the unique ability of PKS II to generate aromatic PKs. In addition to these, there are several enzymes involved in tailoring the PKs aside from the KR, DH, or ER mentioned earlier. These enzymes include the methyltransferase (MT) and glycosyltransferase (GT) that catalyse the

introduction of methyl group or sugar to a polyketide intermediate (Skiba *et al.*, 2016, Hofeditz *et al.*, 2018, Villebro *et al.*, 2019).

PKS III have the simplest structure of all the PKS enzymes (Shimizu *et al.*, 2017). These enzymes are also referred to as chalcone synthases, naringenin-chalcone synthases, and stilbene synthases (Chen *et al.*, 2011, Gomes *et al.*, 2013, Singh et al., 2019) based on their mechanism of action and end-products. These synthases are similar to PKS II, in that they act iteratively only, but differ because they do not use ACP for substrate anchoring during PKs production (Risdian *et al.*, 2019). Instead, these enzymes use free CoA thioesters (Gomes *et al.*, 2013, Singh et al., 2019). PKS III is also different from PKS I and PKS II because it is a homodimeric enzyme, with only a single KS domain performing the functions of both core PKS I and minimal PKS II domains (Shimizu *et al.*, 2017). To synthesize PKs, free CoA or its derivative binds to the active site of the KS domain. The KS domain then catalyses the condensation reaction, with several extender substrates, to generate linear PK intermediates. These intermediates will later undergo intramolecular cyclization producing the final PK product (Singh et al., 2019).

1.3 Structure of ketosynthases and their potential as a probe

Ketosynthases (KS), also referred to as ketoacyl synthases, 3-oxoacyl synthases, and β-ketoacyl synthases (Chen *et al.*, 2011), exist as a single enzyme with a core component for type II fatty acid and polyketide synthases and are also found as a domain in mega-enzymes, such as the type I fatty acid synthases (FAS) and PKSs (Smith & Tsai, 2007, Chen *et al.*, 2011). KS domains are members of the thiolase superfamily, known to perform the Claisen condensation reaction for generating carbon to carbon bonds (C-C). Members of this superfamily form part of the three core PKS I

domains (Beedessee *et al.*, 2019), the "minimal PKS" of PKS II, and PKS III (Yu *et al.*, 2012, Shimizu *et al.*, 2017, Villebro *et al.*, 2019). Generally, in natural products chemistry, phylogenetic markers have been used to classify organisms based on their taxonomy and secondary metabolite production (Engene *et al.*, 2011, Ziemert & Jensen, 2012). However, phylogenetic analysis is inadequate for PKS classification due, in part, to their large and complex sequences (Nguyen *et al.*, 2008, Ridley *et al.*, 2008, Ziemert & Jensen, 2012). It has been proposed that an alternative approach to exploring the biosynthetic diversity of a complex metagenome may include the amplification and phylogenetic analysis of a conserved biosynthetic domain (Thompson *et al.*, 2017, Libis *et al.*, 2019). Phylogenetic analyses of specific domains in different modules of PKS assembly lines have been used to predict substrate incorporation (AT) (Bravo-Rodriguez *et al.*, 2015), stereochemistry of the hydroxyl group produced (KR) (Baerga-Ortiz *et al.*, 2006), and the final structure of the product, whether linear or cyclic (TE) (Zhou *et al.*, 2015). However, the KS domain has been the most informative (highly conserved) in terms of predicting the overall pathways (Jenke-Kodama & Dittmann, 2009, Ziemert & Jensen, 2012). Therefore, finding KS-encoding sequences provides a valid basis for detecting and isolating gene clusters encoding the biosynthesis of novel pharmacologically relevant PKs (Kurnia *et al.*, 2017). Studies have detected the presence of PKs in the environment, using KS degenerate primers specific for Actinomycetes (Selvin *et al.*, 2016, Arteaga *et al.*, 2017, Wei *et al.*, 2018). Public repositories such as UniprotKB (Boutet *et al.*, 2016) and NaPDoS (Ziemert *et al.*, 2012) include several characterized KS genes. These developments, combined with the improvements in NGS, have led to the generation of different environmental data (Shokralla *et al.*, 2012, Hassan *et al.*, 2020). Based on these advances, there is an urgent need to harness these data for PKs prospecting.

For instance, recent studies used Hidden Markov Models (HMMs) to evaluate the abundance and novelty of genes involved in biogeochemical cycling and energy metabolisms (Anantharaman *et al.*, 2016, Qiu *et al.*, 2020, Levy-Booth *et al.*, 2021, Trilla-Prieto *et al.*, 2021). Indeed, HMMs have been demonstrated to be more accurate than many other alignment-based approaches (Eddy, 2011). Since domains are units that can be used as building blocks for proteins due to their evolutionary or structural conservation (Moore *et al.*, 2008, Klasberg *et al.*, 2016), and protein domains are typically described by probabilistic models, specifically HMMs (Klasberg *et al.*, 2016). Consequently, we propose the use of HMM models based on publicly available characterized KS genes to prospect for PKs, instead of having to use degenerate primers.

1.4 The discovery of novel PKs with the use of HMMs

Here, we briefly discuss how HMM can be utilized to discover novel PKs. Initially, we describe how metagenomic samples can be collected from the environment. In addition, the contextualized approach proposed in this study together with the method of accessing the clusters of the novel KS. Finally, point out other methods that can be coupled with the approach for better detection and characterization.

1.4.1 Construction of metagenomes from complex environments

Sample collection from complex environments, such as deep-sea hydrothermal vents, and DNA extraction procedures for metagenomic analysis is vital for bioprospecting (Sekurova *et al.*, 2019, Davies-Bolorunduro *et al.*, 2021). It is crucial that samples are analysed or flash frozen immediately after collection to avoid alteration in microbial community profiles (Quince *et al.*, 2017). Sufficient sample quantities must be collected from environmental sites with low biomass,

such as deep-sea hydrothermal vents, to obtain adequate DNA for downstream analysis. DNA extraction protocols must be carefully considered to select, optimize, and maximize yields (Farraj *et al.*, 2020). Several protocols for obtaining high quality DNA for metagenomic sequencing have recently been reported (Josefsen *et al.*, 2015, Méndez-García *et al.*, 2018). Recently, Méndez-García et al. (2018) and Su et al. (2022) reviewed useful techniques for sample collection, preservation, and transportation, as well as DNA extraction from complex samples, before DNA sequencing.

The Illumina short read sequencing platform is currently the most widely used platforms available. This is due to relative cost effectiveness and high-throughput provision of high quality sequence data (Van Dijk *et al.*, 2014, Almeida & De Martinis, 2019). This sequencing technology, its quality control system, and sequence assembly have been recently reviewed by Almeida and De Martinis (2019). Generally, short read technology typically begins with DNA fragmentation, DNA end-repair, adapter ligation, surface attachment, and in-situ amplification (Hu *et al.*, 2021). This requires reassembling sequencing data over long stretches of DNA, which presents challenges with structural variations or low-complexity regions. However, considering the complexity of PK, the Pacific Bioscience sequencing platform which provides long reads may have huge technical advantage in capturing very long nucleotide sequences from a DNA molecule (i.e. more informative in terms of coverage)(Weirather *et al.*, 2017). The raw read sequences either from the constructed metagenomes or the online sequence data may be quality filtered and trimmed using programs like Trimmomatic (Bolger *et al.*, 2014) and Cutadapt (Martin, 2011). The quality filtered sequences may be assembled into contigs using MetaSPAdes (Nurk *et al.*, 2017), Megahit (Li *et al.*, 2016) and IDBA (Peng *et al.*, 2012). In order to predict ORFs or putative protein coding regions from the contigs, programs such as Prodigal are used (Hyatt *et al.*, 2010). The ORFs must be

examined such that only sequences with full reading frame within contigs will be selected to avoid truncated genes making up the metagenome query-ready.

1.4.2 Retrieval of KS protein sequences

Characterized KS domain protein sequences of PKS may be retrieved in FASTA format from online resources, such as the UniprotKB (Boutet *et al.*, 2016) and NaPDos (Ziemert *et al.*, 2012). Consequently, we recommend grouping sequences based on 95% similarity using CD-HIT (Chen *et al.*, 2016) and representative sequences from the grouping may subsequently be aligned using existing multiple sequence alignment programs, such as MUSCLE (Edgar, 2004) or MAFFT (Katoh *et al.*, 2002, Katoh *et al.*, 2019). These may then be edited using an alignment trimming tool such as trimAl (Capella-Gutiérrez *et al.*, 2009), and the edited alignments may thereafter be used to generate profile HMMs (Hidden Markov Models) for each group using HMMER (Eddy, 2011). Translated protein sequences predicted from metagenomes or genomes of interest may therefore get queried against a set of HMM profiles generated. We argue that the use of HMMs represents a more sensitive approach compared to running homolog BLASTN or BLASTP (Altschul *et al.*, 1990, Camacho *et al.*, 2009) searches using all the related sequences from the databases as a direct query.

1.4.3 Screening metagenomes for KS encoding genes

First, protein sequences predicted from metagenomes may be queried against putative profile HMMs using the HMMER hmmsearch approach (Eddy, 2011). For increased sensitivity and to avoid false positives predictions, stringent E-values i.e 1e-15 for both sequence domain reporting must be set (Finn *et al.*, 2011). The evolutionary relationships of the candidate KS genes identified in the metagenomes may be calculated to determine phylogenetic and functional novelty based on compounds produced by existing KS genes. KS genes that are distantly related to the already characterised or existing KS genes are regarded as putatively novel KS genes.

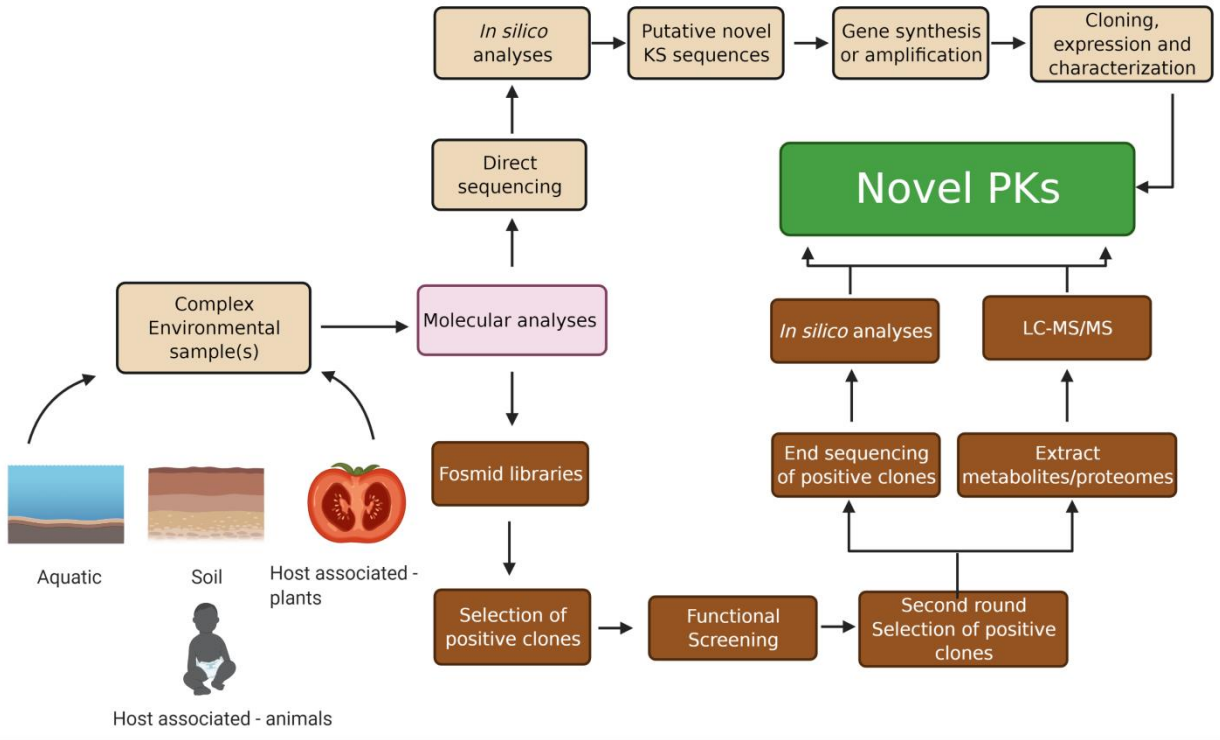

Figure 1.4: The conceptual framework of the pipeline. Metagenomic analyses will be used to survey and characterize total community DNA from natural environments including aquatic, soil

and host associated (animal and plant) microbiomes. Two approaches may be following initial molecular analysis (DNA extraction and quality processing). The functional screening approach may be used to elucidate the composition, diversity, structure, and functions of novel PKs in a complex environment. While the direct sequencing approach may be used to provide an overview of putative novel targets. Used correctly, both approaches may reveal novel PKs from complex environments. This image was created with BioRender.com

1.4.4 Accessing the clusters of PKSs associated with novel KS genes

There is a need to have access to the co-localized and associated genes (clusters) of the phylogenetically distinct KS genes, to fully characterize and confirm their novelty. Hence, contigs in possession of KS genes are parsed from the metagenome for cluster identification using programs such as ClusterFinder (Cimermancic et al., 2014), DeepBGC (Hannigan et al., 2019), antiSMASH (Blin *et al.*, 2021). AntiSMASH resource possesses a database of experimentally characterized biosynthetic gene clusters which may be used as reference to further characterize and define the novelty of queried candidate clusters. A suite of parameters can also be enabled such as “KnownClusterBlast” (searches queried clusters against MIBiG repository), “ClusterBlast” (searches queried clusters against a comprehensive gene cluster database), “SubClusterBlast” (searches queried clusters against a database containing operons involved in the biosynthesis of common metabolites building blocks) which better characterizes and defines the novelty of these clusters.

1.5 Functional characterization of the novel clusters

Candidate novel gene clusters with sequence and domain features that are distinct from those in antiSMASH may then be synthesized directly (especially when the environmental DNA is not available) and cloned for expression in a related host for compound characterization. On the other hand, when the environmental DNA is available, this access may be provided by the construction of functional metagenomic libraries. PKS clusters may be as small as 18 kb or as large as 84kb, for clusters responsible for the production of bikaverin and aflatoxin in *Fusarium fujikuromi* and *Aspergillus parasiticus*, respectively (Noar & Daub, 2016). However, using fosmid libraries combine the ability to accommodate large inserts, such as complete gene clusters, with relative ease of construction compared to other types of libraries (Martínez & Osburne, 2013). Furthermore, larger insert-derived libraries, like the bacterial artificial chromosome, usually yield metagenomic libraries with clones that are smaller than Fosmid-sized clones by two to three orders of magnitude (Kim *et al.*, 2010). Alternatively, if a larger cluster is involved, the transformation-associated recombination (TAR) cloning method in yeasts can be used to merge several overlapping clones from different Fosmid libraries together so as to capture the larger cluster (Kim *et al.*, 2010, Libis *et al.*, 2019). However, there is a major challenge in natural product discovery, which is usually the transcriptionally silent nature of biosynthetic gene clusters (BGC), hence, there is often the need to trigger some BGC's expression (Zhang *et al.*, 2017). This may be done by screening the clones generated from Fosmid libraries against similar or putative predicted compounds, based on bioinformatic results obtained via KS analysis. The positive clones that are resistant to the compounds can therefore be extracted and sequenced for adequate characterization of the BGC and the PKs involved. Various other methods can be used to awaken the silent nature

of the BGC such as ribosome engineering, regulatory network optimization and chemical elicitors (Choi *et al.*, 2018). Finally, PKs can be extracted using the polyol extraction method for structural elucidation via LC-MS/MS (Beedessee *et al.*, 2019). The activity of the crude PK extracts can also be assayed at various concentrations against multidrug resistant-pathogens (Fatema *et al.*, 2018). The mechanism of action of the PKs can also be determined via phenotypic analyses as previously described (Zoffmann *et al.*, 2019).

.

1.6 Using metagenome-assembled genomes (MAGs) and single amplified genomes (SAGs) to uncover novel biosynthetic gene clusters

The assembly of high-throughput, short-read metagenomic data into contigs and binning into draft genomes results in metagenome assembled genomes (MAGs) (Tyson *et al.*, 2004, Albertsen *et al.*, 2013). Similarly, single-cell genomics (SCG) which allows for the sequencing of individual cells or a genome region in an uncultured cell, provides a valuable opportunity to measure different molecules with ultimate resolution, such as DNA, RNA, protein, and chromatin (Stepanauskas, 2012, Hu *et al.*, 2018). The recovery of both MAGs and single amplified genomes (SAGs) has transformed microbial research. These techniques have become critical for studying microbial communities in various environments, removing the limitation in the previously uncultured population of microorganisms. MAGs (Parks *et al.*, 2017, Anantharaman *et al.*, 2018, Ward *et al.*, 2018) and SAGs (Marcy *et al.*, 2007, Stepanauskas, 2012, Rinke *et al.*, 2013, Hu *et al.*, 2018) have been used in several studies to disentangle the diversity, metabolic potential, and evolutionary history of environmental microbes. The problem of rediscovery of known compounds by the

traditional method of prospecting for antibiotics is heavily hinged on the slow advancement or inability to culture the greater proportion of microorganisms in nature. As a result, MAGs and SAGs may provide us with access to the biosynthetic gene clusters (BGCs) that encode the pathway required to produce secondary metabolites from non-culturable organisms. For instance, Kogawa et al. (2021) localized chemical features of a specific bacterium in an uncultivated microbiome using single-cell sequencing. With this, they were able to identify the bacteria in the uncultured lineage as a producer of aurantoside (an antifungal natural product) from a chemically and microbially complex sponge (Kogawa *et al.*, 2021). Hence, we propose combining this approach with other bioinformatics computational tools such as BIG-SCAPE (Navarro-Muñoz *et al.*, 2020) and inferring phylogenies will facilitate the identification of BGCs from non-culturable organisms and the elimination of BGCs of known compounds. In addition, the potential ecological contributions of the organisms in their specific environments can also be inferred.

1.7 Metabolomics and metaproteomics as complementary resources

Metabolomic analyses involve the characterization and identification of small weight molecules (metabolites) which represent the final products of biological processes (Muthubharathi *et al.*, 2021). Such products may be either in fluids of a single cell or a microbial community (microbiome) (Weckwerth, 2007, Abid *et al.*, 2018, Muthubharathi *et al.*, 2021) at a given time. With the measurement of these molecules, it is possible to deduce changes that will occur within a biological system in response to both genetic and environmental factors (Buescher *et al.*, 2012, Tautenhahn *et al.*, 2012, Zhang *et al.*, 2014). For instance, Tobia et al. (2017) used high-resolution mass spectrometry analyses to reveal a huge chemical diversity in two taxa (*Photorhabdus* and *Xenorhabdus*) that appeared very similar at the DNA sequence level. They

went further to uncover a large number of previously unidentified metabolite classes including the xefoampeptides and tilivalline using molecular network reconstruction (Tobias *et al.*, 2017). Since metagenomics affords us the opportunity to access the relative gene abundance of a particular microbial community (i.e genetic makeup) and subsequent detection of a novel KS, metabolomics provide direct correlations between genotype and phenotype in biological systems (Cuperlovic-Culf *et al.*, 2013, Stuart *et al.*, 2020). As a result, acquiring and analyzing paired datasets which include MS/MS data from culture extracts and genome sequences from their producers may facilitate improved bacterial NP discovery programs. We propose combining the two approaches whether on a community level (as meta-omics), or on a cultured organism (as in the case of conventional method of drug discovery via activity assay of a particular organism), or on the clones of the synthesized clusters (as in the case at 4.4 above). This will reveal the extent of novel PKs and/or molecules which may be linked with the identified novel KS. There is some evidence that direct quantification of molecules in biochemical samples, using Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS) coupled with chromatography (Wishart, 2008, Stuart *et al.*, 2020), may facilitate the identification of novel PKs compounds. Although NMR and MS are part of the major techniques used for metabolomics, other platforms including Gas Chromatography Mass Spectrometry (GC-MS), Electrospray Ionization Mass Spectrometry (ESI-MS), High Performance Liquid Chromatography (HPLC), Ultra Performance Liquid Chromatography (UPLC), Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Imaging (MALDI-MSI) may be employed to study metabolomes (Peng, 2000, Zhang et al., 2012, Boufridi and Quinn, 2016, Stuart et al., 2020, He et al., 2022). Metabolomics may also be applied at various stages of drug discovery, ranging from the detection phase to the clinical trial stage (Zhang *et al.*, 2020). This technique has been used to optimise media ingredients/nutrients for

improved functional expression (Bose *et al.*, 2017, Kurtböke, 2017). However, there are still limitations to this approach which has been reviewed recently by Yang et al. (2019) and Zhang et al. (2020). At the crux of these limitations is the high number of unknown metabolites, which leads to difficulties in the mapping (Yang *et al.*, 2019, Zhang *et al.*, 2020). The reduced ability to map these metabolites substantially limits the biological interpretations which may be inferred from specific targets (Zhang *et al.*, 2013).

Metaproteomics or proteomics is also considered an effective way of identifying gene products of a community or organisms or clones (Rath *et al.*, 2011). This is related to measuring protein productivity and expression as a physiological state monitor (Zhang *et al.*, 2014). These measurements are accomplished by using mass spectrometry for quantitative characterization of proteins, and subsequently predicting their abundance and function (Srivastava *et al.*, 2022). Consequently, metaproteomics has made it possible to investigate the bacterial response to stress condition and starvation (Abid *et al.*, 2018), which is especially applicable when triggering silent BGCs. Furthermore, metaproteomics combines the immediacy of traditional screening (activity-based screening) with the culture-independent "meta-omic" approach (Sukul *et al.*, 2017). This makes metaproteomics a vital tool in natural product drug discovery. For instance, Proteomic analysis were used to uncover the regulatory mechanisms controlling antibiotic production at various stages of culture growth (Gallo *et al.*, 2016), resulting in the identification of all types of natural products and their associated gene clusters. In addition, the duo of metabolomics and proteomics was also used by Gubbens et al. (2014) to elucidate gene clusters of natural products in *Bacillus* and *Streptomyces*.

Hence, in an ideal scenario where funds are unlimited, combining metabolomics and metaproteomics with metagenomics may be a powerful tool in disentangling the PKs of interest.

These approaches are crucial for unravelling the complexity of the regulatory mechanisms controlling the expression of BGCs of interest and the subsequent discovery of novel natural products or molecules (Nieselt *et al.*, 2010, Covington *et al.*, 2017, Palazzotto & Weber, 2018) [Figure 1.4]. These approaches are important for harnessing the currently untapped biodiversity in natural ecosystems. However, there are currently large unexplored datasets, which may also provide information on uncharacterized BGCs. With the development of computational approaches, these datasets may be explored to characterize novel natural products.

1.8 The application of machine learning as a tool for bioprospecting

Machine learning (ML) is a combination of statistical approaches classified as artificial intelligence. These approaches rely on initial training, using available known datasets, and then generating predictions based on various algorithms (Beam & Kohane, 2018, Shouval *et al.*, 2021). The resultant autonomous model is trained using a subset of validated datasets (which is typically around 60% of the overall validated data) (Guo *et al.*, 2021, Marya *et al.*, 2021). These data may vary in size but are usually constructed to represent the variability of the complete validated dataset. The recognition that biological systems vary greatly in complexity is an important criterion for applying these approaches.

ML-based algorithms use both supervised and unsupervised learning approaches (Vamathevan *et al.*, 2019). The supervised method (regression and classifier methods) is responsible for training models using known input and output data (El Mrabet *et al.*, 2021). These data are then used to predict future values of any given data (El Mrabet *et al.*, 2021). The unsupervised method is responsible for the exploratory purposes of hidden information by using clustering patterns in a meaningful way for data analysis (Aouedi *et al.*, 2021). The combination of these approaches is useful in drug discovery and may yield substantial insights regarding novel molecules produced

by microorganisms (Fang *et al.*, 2016, Lo *et al.*, 2018, Masalha *et al.*, 2018, Li *et al.*, 2019, Rayan *et al.*, 2019, Zoffmann *et al.*, 2019).

For example, Zoffmann et al. (2019) used a ML approach to capture the complexity of bacterial phenotypic fingerprints of a test compound to infer its novel mode of action (Zoffmann *et al.*, 2019). ClusterFinder (Cimermancic *et al.*, 2014) and DeepBGC (Hannigan *et al.*, 2019) form part of the widely used ML tools that have been developed to discover novel and previously undetected BGCs (Prihoda *et al.*, 2021). To demonstrate the efficiency of these approaches, Cimermancic et al. (2014) identified BGCs in an extensive extant microbial sequence data using the ClusterFinder tool. The authors further generate a network analysis of the predicted BGCs and revealing a large gene cluster family, dominated by uncharacterized sequences. Experimental characterization of the most prominent family, consisting of two subfamilies of hundreds of BGCs revealed their products to be arylpolyene (Cimermancic *et al.*, 2014). Liu et al. (2021) also utilized DeepBGC to identify specialized metabolite biosynthetic gene clusters in *Streptomyces scabiei* a causative agent of common scab disease. They were able to detect and annotate new molecules with diverse bioactivities that were not previously known to be produced by *S. scabiei* (Liu *et al.*, 2021). The results were further supported by untargeted liquid chromatography-coupled tandem mass spectrometry (LC-MS2) (Liu *et al.*, 2021). These studies provide clear evidence that the predictive nature of these algorithms yields substantial insights regarding unknown bioactive compounds.

Another recent example that uses an unsupervised approach to detect BGCs of unknown architecture is GECCO (Carroll *et al.*, 2021). GECCO employs conditional random fields, which is a high-precision, scalable method for identifying novel BGCs. Its predictions rely on protein domains with both known and novel associations to secondary metabolism, which has also indicated some efficiency in the detection and expansion of novel BGCs (Carroll *et al.*, 2021).

These studies (and others) show that the application of ML in drug discovery may revolutionize the field by improving the success rate. Since the success rate of drug discovery and development is comparatively low across multiple facets of therapeutics (Wong *et al.*, 2019), and pipelines for drug discovery are usually complex and time-consuming (Vamathevan *et al.*, 2019) these approaches offer exceptional potential. We propose the incorporation of ML algorithms and software at various stages of drug discovery and development pipeline. This incorporation may ultimately result in improvements to the final outputs by providing novel compounds. However, since the predictive power of ML approaches depends on the availability of large volumes of high quality data (Koscielny *et al.*, 2017), there is a need to generate such data for both positive and negative annotations (Adadi *et al.*, 2019, Vamathevan *et al.*, 2019). Such data are crucial for the development of ML based drug discovery especially for programs which use HMMs including ClusterFinder.

1.9 Conclusion and recommendation

The discovery of novel biosynthetic pathways, and their related natural products of pharmaceutical relevance from the environment, has been the subject of extensive research (Newman & Cragg, 2020, Kogawa *et al.*, 2021, Paoli *et al.*, 2022, Waschulin *et al.*, 2022). This is particularly true for targets derived from complex and extreme environments with underexplored diversity (Mahajan & Balachandran, 2017, Hug *et al.*, 2018). Previous studies suggest that these environments are likely to harbour microbial communities with unique adaptative capabilities, and collectively remain enormous untapped resources (Rego *et al.*, 2020, Rego *et al.*, 2021, Paoli *et al.*, 2022, Waschulin *et al.*, 2022). To facilitate the discovery of novel compounds, we recommend the construction of HMM profiles as a viable tool for exploring these uncharacterized resources. We also suggest that KSs sequences, available in public domains, be used as these provide high

confidence in the detection of novel and phylogenetically distinct KS genes. In addition, these HMMs will provide more robust *in silico* data of PKS clusters prior to laboratory-based phenotypic characterization. This approach circumvents the traditional targeted method (using degenerate primers), which are biased to specific groups of organisms. In addition, the identification of novel KSs with high activity stability and substrate promiscuity will also foster new PKs derivatives through genetic engineering of the existing chimeric PKS in synthetic biology, by improving their activities.

1.10 Problem statement

The effects of rising antibiotic resistance are clearly evidenced by the loss of lives annually across the world (Boolchandani *et al.*, 2019, Cassini *et al.*, 2019, Su *et al.*, 2019). This provides clear motivation for the discovery of new antimicrobials. However, only one new class of antibiotic belonging to the narrow-spectrum daptomycin has been discovered and implemented clinically in the last 50 years (Lewis, 2012). It is clear that searching for novel antibiotics, using traditional approaches, such as screening towards well-known producers, will result in the rediscovery of known compounds (Von Bubnoff, 2006, Li *et al.*, 2019). Relative to soil environments, which are highly explored, marine environments provide essential sources of novel natural products, as marine organisms have shown exceptional biological, biochemical, and biosynthetic potential (Jeewon *et al.*, 2019, Paoli *et al.*, 2022). One of the methods used to uncover potential BGC is through computational approaches such as demonstrated by antiSMASH, and these rely on structural similarity-based approaches limiting the detection of novel BGCs. However, this study implements the use of HMMs as an approach to search for potentially novel polyketide contigs in a metagenome and how it may enhance the detection of diverse groups of drug classes.

1.11 Hypothesis

This thesis explores questions related to the diversity of polyketides (PKs) in oligotrophic oceans (South Atlantic Ocean (SAO) and Southern Ocean (SO)), these regions represent pristine marine microbiomes. Recent studies have shown that microbial diversity and functionality increases with depth. This has been attributed to increased specialization to niche environments and use of alternative pathways. We hypothesize that these environments will harbor novel polyketides using an HMM based approach and other metagenomics approach, and that there would be a proportionate increase in the diversity and novelty of BGCs with depth.

1.12 Aims and objectives

1.12.1 Aims

Several studies have focused on the diversity of PKs. Most of these studies have focused on the diversity of PKs in the microbiome of sponge and sediments of the marine environment, and little work has been done on the free-living organisms in the ocean column. Consequently, this thesis aims to elucidate the diversity of polyketides across oligotrophic ocean stratification by exploring the following questions:

i. Are there unique polyketides in the euphotic zone of an oligotrophic ocean
ii. Do viruses play a role in the diversity of polyketides
iii. Do the distribution of secondary metabolite gene vary across the ocean stratification

1.12.2 Objectives

The specific objectives of this study include:

1.12.2.1 Polyketides diversity analysis of the euphotic zone of SAO

i. Building and validation of HMM profiles for KS.

ii. Using the profile to characterize polyketides in the euphotic zone of SAO, and determining the extent of their novelty through gene cluster family (GCF) network formation.

iii. Checking if viruses are potential vectors for the dissemination of polyketides in these environments

1.12.2.2 Analysis of the biosynthetic potentials of the Southern Ocean

i. Using metagenome-assembled genome to assess taxonomy of the uncultured microbial population of the SO

ii. Characterizing the secondary metabolite gene clusters of the SO

iii. Comparing the biosynthetic potentials of three major layers of the ocean over two seasons.

Chapter Two

Metagenomic analysis reveals diverse and phylogenetically distinct polyketides in euphotic zones of the South Atlantic Ocean

2.0 Abstract

The South Atlantic Ocean (SAO) is nitrogen-limited due to little or no deposition of iron (Fe) rich dust. The resultant oligotrophic habitats harbour a wide array of complex, and novel, secondary metabolites. One such group includes polyketides (PKs), a biotechnologically important group of enzymes with several pharmacological properties. However, little is known regarding their diversity in the euphotic regions including those in the SAO. Here, we used an *in silico* approach to investigate the diversity of polyketides in the SAO. Ketosynthase-based HMM profiles were designed to query 24 metagenomes, and their associated biosynthetic clusters were determined using the antiSMASH tool. We demonstrate that the SAO possesses a diverse range of phylogenetically distinct and potentially novel PKs/KSs. A fraction of these were determined to relate to Arylpolyene, T1PKS and HgIE-KS associated with *Proteobacteriota* and *Bacteroidota* phyla. We further indicate that viruses may contribute towards the evolution and dissemination of BGCs in the SAO**.**

2.1 Introduction

Marine ecosystems are regarded as a source of structurally diverse secondary metabolites (Tortorella *et al.*, 2018, Paoli *et al.*, 2022). Polyketides (PKs) constitute some of the most important secondary metabolites and have been pursued for their complex biological properties (O'Hagan,

1991, Hertweck *et al.*, 2007). The chemical structures of these secondary metabolites possess properties that make them ideal for application in several pharmaceutical and agricultural industries (Barajas *et al.*, 2017). In medicine, these metabolites have been found to produce compounds with antibacterial (Igarashi *et al.*, 2011, Ola *et al.*, 2018), antifungal (Khan *et al.*, 2011, Gao *et al.*, 2020) and antitumor (Li *et al.*, 2018, Zabala *et al.*, 2020) activities. Apart from these activities, PKs contribute towards biogeochemical cycling of nutrients including nitrogen in the ocean (Saito & Awai, 2020), suggesting that they may have broad roles in microbial primary production (Saito & Awai, 2020). In the ocean, the evolution and diversity of polyketides have been shown to be influenced by oligotrophic conditions, which results in the distribution of biosynthetic operons through horizontal gene transfer (HGT) (Jenke-Kodama *et al.*, 2005, Jenke-Kodama & Dittmann, 2009, Romero *et al.*, 2011, Nivina *et al.*, 2019). However, comparatively few studies have explored the extent of diversity and phylogeny of these compounds in marine ecosystems.

Previous studies have explored readily cultivable organisms, recovered from soil, for natural products (Lavrova *et al.*, 1972, Campbell *et al.*, 1983, Ehling-Schulz *et al.*, 1997). However, more recent studies suggest that using such approaches to explore compounds produced by culturable microorganisms may result in the rediscovery of known molecules (Amos *et al.*, 2015, Charlop-Powers *et al.*, 2016, Lemetre *et al.*, 2017, Borsetto *et al.*, 2019, Sharrar *et al.*, 2020, Zhang *et al.*, 2021). Therefore, to circumvent such problems recent studies have focused on uncultivable organisms, which inhabit pristine environments such as the ocean (Khalifa *et al.*, 2019, Subramani & Sipkema, 2019, Blasiak *et al.*, 2020). Some of these studies have explored the capacity of marine environments to produce novel biotechnologically useful polyketides based on the application of high throughput technologies (Paoli *et al.*, 2022). These are mainly conducted on microbiota

acquired from different oceanic regions such as the South Pacific (Rust *et al.*, 2020), Western Pacific (Wei *et al.*, 2018) and Caribbean sea (Tuttle *et al.*, 2019) associated with the benthic or sediments (Wei *et al.*, 2018, Tuttle *et al.*, 2019, Bech *et al.*, 2020) and sponge microbiome (Trindade-Silva *et al.*, 2013, Ueoka *et al.*, 2015, Rust *et al.*, 2020, Kaluzhnaya & Itskovich, 2022). Yet, comparatively few studies have focused on the discovery of polyketides in the South Atlantic Ocean (Trindade-Silva *et al.*, 2013).

The South Atlantic Ocean (SAO) is an oligotrophic ocean, known to be limited in nitrogen due to low or no deposition of iron rich dust (Howarth & Marino, 2006, Moore *et al.*, 2013, Conway *et al.*, 2018). Iron is a significant micronutrient required for nitrogen fixation (Han *et al.*, 2006, Moore *et al.*, 2009, Sohm *et al.*, 2011) . The oligotrophic nature is likely to be more in the euphotic zone where nitrogen fixation is critical for regulating the marine nitrogen budget (Shiozaki *et al.*, 2014, Zehr & Capone, 2020), by providing a significant source of nitrogen for new production which supports this ecosystem (Shiozaki *et al.*, 2014). Since the evolution and diversity of polyketides have been shown to be influenced by two major factors: namely genetic (via gene duplication, recombination, and Horizontal Gene Transfer (HGT)) and environmental factors (Jenke-Kodama *et al.*, 2005, Jenke-Kodama & Dittmann, 2009, Romero *et al.*, 2011, Nivina *et al.*, 2019), conditions such as nutrient deficiency, environmental competitiveness amongst many others are part of the driving forces for the evolution of biosynthetic genes (Giordano *et al.*, 2015). Microorganisms in the euphotic zone of SAO are therefore presumed to harbour a wide range of complex and novel PKs in response to the nutrient stress.

Here, we explore the diversity of polyketides in euphotic zones of the SAO using an HMM based approach. Using this approach, we demonstrate that the SAO possesses a diverse range of phylogenetically distinct and potentially novel KSs. A fraction of these were determined to relate

to Arylpolyene, T1PKS and HgIE-KS associated with *Proteobacteriota* and *Bacteroidota* phyla. We further indicate that viruses may contribute towards the evolution and dissemination of BGCs in the SAO**.**

2.2 Materials and methods

2.2.1 Sample collection and molecular analysis

Sampling was performed aboard the *RV SA Agulhas II* during spring in September 2018 from eight oceanographic stations (Figure 2.1 & Table S1). In total, 45 liters of water was collected per station (5 m depth) and filtered through 0.2 µm Polyethersulfone (PES) filter membranes (Merck, RSA) using the approaches detailed previously. The membranes were stored at -80^{0}C until DNA extraction was done (Makinde *et al.*, 2022).

DNA extraction and library preparation was performed as previously described by (Hirai *et al.*, 2017). Briefly, membrane filter pieces were incubated in 400 µl of DNA extraction buffer (400 mM Tris-HCl pH8, 60 mM EDTA pH8, 150 mM NaCl and 1 % SDS) for 10 mins at 60°C. 120 µl of 3 M potassium acetate buffer was used to precipitate organic matter followed by incubation on ice for five minutes and centrifugation at 15000 rpms at 4°C for one minute. DNA retrieved from the supernatant was subsequently purified through a two-step process by initially adding 800 µl of solution C4 and 500 µl of solution C5 from the DNAeasy Power soil kit (Qiagen, Hilden, Germany) following manufacturers specifications. Nucleic acid concentrations were determined using the Qubit 4 Fluorimeter (Thermofisher, Yokosuka, Japan). High quality DNA was used to construct libraries using the KAPA Hyper Prep Kit (KAPA biosystems, Cape Town, RSA) as

detailed in the manufacturer's protocol. Samples were sequenced using an Illumina HiSeq2000 instrument at Macrogen next-generation sequencing service.

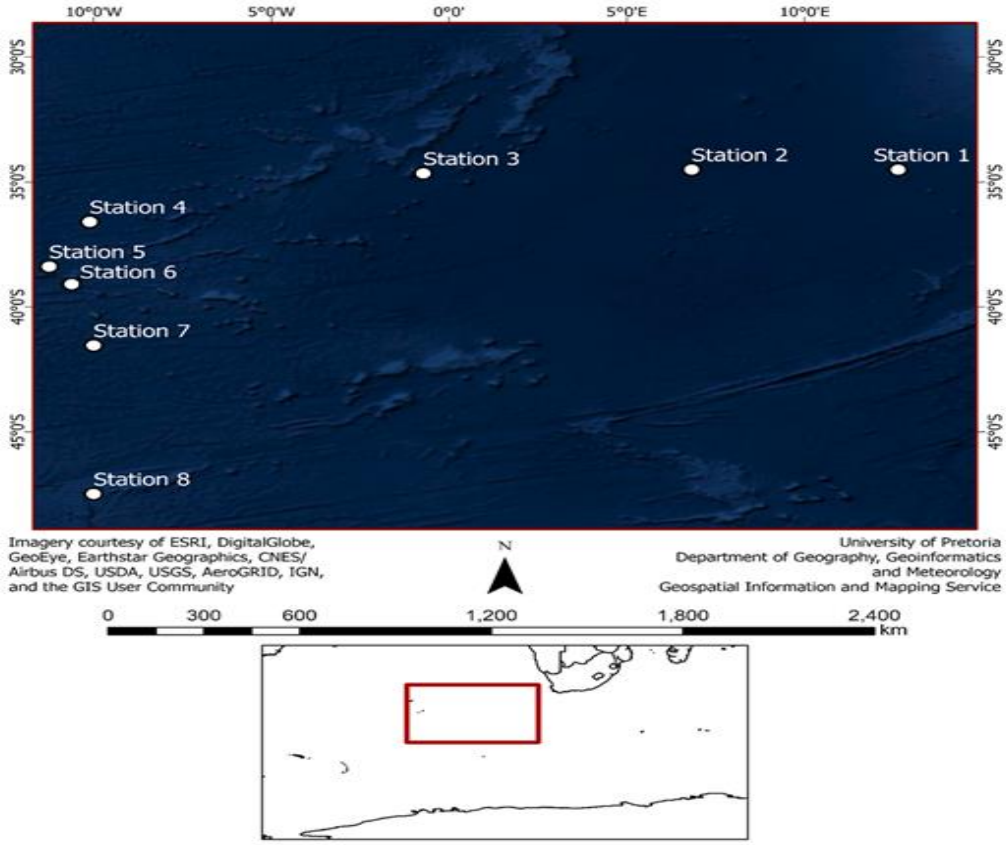

Figure 2.1: The eight sampling stations in the South Atlantic Ocean. Sampling stations are denoted as white bullets and the outcrop shows the entire sampling site relative to South Africa and the South Atlantic Ocean (courtesy of Lauren Pijper, University of Pretoria, Department of Geography, Geoinformatics and Meteorology).

2.2.2 Shotgun metagenomic data analysis and processing

Raw metagenomic data were quality filtered using Trimmomatic v0.36 (Bolger *et al.*, 2014). Filtered metagenomes were assembled with MEGAHIT v1.0.3 with default parameters (Li *et al.*,

2016). Each assembly was quality controlled to a minimum contig length of 500 bp and resulting contigs were used to predict proteins using Prodigal v2.6.3 with the -p meta parameter implemented (Hyatt *et al.*, 2010). Next, we determined the microbial taxonomy of each metagenome using singleM v0.12.1 (https://github.com/wwood/singlem) on the raw reads with default parameters. The resultant taxonomic classifications at phylum level were visualized using R ggplot2 package.

2.2.3 Screening for ketosynthases using an HMM approach

To construct HMM profiles, we retrieved 194 and 458 Ketosynthase domain (KS) sequences of polyketide synthase (PKS) from UniprotKB (March 2020) (Boutet *et al.*, 2016) and NaPDoS (February 2020) (Ziemert *et al.*, 2012), respectively. The sequences were merged and clustered at 40% protein similarity (-c 0.4 -n 2 -d 40) using CD-HIT (Chen *et al.*, 2016). This resulted in 29 clusters and, from these, a total of 21 with >=2 sequences were aligned using ClustalW2 using default parameters (Larkin *et al.*, 2007). The resultant alignments were used to build Hidden Markov Model (HMM) profiles (Eddy, 2011) using hmmbuild (HMMER 3.1b2). The profiles were searched against complete protein sequences from all 24 metagenomes using hmmsearch (HMMER 3.1b2) with --incE 1e-10 --incdomE 1e-10 parameters. To provide an ecological context, we retrieved the 12 Tara Ocean surface co-assembled metagenomes acquired from 12 oceanic regions of a study by Delmont et al. (2018) for comparison. Protein sequences were predicted from these data, as detailed previously, and were subject to the HMM analysis.

2.2.4 Diversity of polyketides KS

For downstream phylogenetic analysis, all the KS sequences with >= 200bp from both the SAO and Tara metagenomes, were clustered using CD-HIT (Chen *et al.*, 2016). The criteria for cluster

analysis were based on 98% similarity, over an alignment of 80% of the shortest sequence (-c 0.98 -n 5 -d 0 -aS 0.8). The resultant KS representative sequences were further merged with the NaPDoS and UniprotKB KSs and aligned using MAFFT (Katoh *et al.*, 2019) with –auto option. The resultant alignments were trimmed using trimAl (Capella-Gutiérrez *et al.*, 2009) using the noallgaps option. The alignments were used to reconstruct maximum likelihood trees using Fasttree based on default parameters (Price *et al.*, 2009).

BlastP was used to determine the functional potential of the KSs associated with the SAO and TARA. Initially, KS protein sequences from the NaPDoS and UniprotKB databases were used to create a diamond database. Then, SAO and TARA ocean KSs were queried against this database using blastp (Buchfink *et al.*, 2015) with --query-cover 80 -k1 option.

2.2.5 HMM efficacy

To assess whether our HMM approach predicts KSs associated with known BGCs, available in public repositories. Contigs associated with the KS domain were used to predict for the presence of biosynthetic gene clusters (BGCs) using antiSMASH version 6.0 pipeline (Blin *et al.*, 2021). We used the strict detection mode with all the analysis options selected (KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREFinder, Cluster Pfam analysis, Pfam-based GO term annotation, TIGRFam analysis) in antiSMASH. From these analyses, contigs which were not predicted to possess BGCs were considered novel PKs encoding for unique pathways. Furthermore, contigs that were predicted to possess BGCs but lacked the presence of known KS genes were also regarded as novel PKs. These two groups were not

considered for BGC downstream analysis. However, those that were predicted to possess BGC with known KSs were regarded as polyketides (PKs) for downstream analysis.

2.2.6 The distribution of polyketides gene clusters across the SAO sampling stations

To determine the distribution of polyketide gene clusters associated with KS hits in each metagenome across sampling stations, protein sequences which shared similarity with the KS profiles were analysed individually, clustered using CD-HIT (Chen *et al.*, 2016). The threshold was set at 98% sequence similarity over an alignment of 80% of the shortest sequence (-c 0.98 -n 5 -d 0 -aS 0.8) to reduce redundancy. Representative sequences were retrieved from the clusters, and antiSMASH was used to determine BGCs from the contigs that contained these representative sequences. The taxonomic provenience of the identified polyketide clusters was determined using the Contig Annotation Tool (CAT) (von Meijenfeldt *et al.*, 2019). R studio was used to visualise the distribution of the PKs cluster across sampling stations and the polyketide clusters' taxonomic provenience using the ggplot2 package (Allaire, 2012, Gómez-Rubio, 2017).

2.2.7 The diversity of BGCs in SAO, TARA and MiBIG

To reduce the knowledge deficit regarding the diversity and distribution of BGCs in SAO and TARA, we searched for their presence in metagenomes. Gene cluster frequently results in fragmented assemblies from short reads and therefore, contigs with lengths >= 10kb were searched for the presence of BGCs with Antismash version 6 pipeline using the default parameters (Blin *et al.*, 2021). The BGCs predicted in all datasets were further compared with a set of experimentally verified BGCs in the MIBiG version 2.0 database. We used the BiG-SCAPE pipeline using a

threshold of 0.7 (Navarro-Muñoz *et al.*, 2020) to verify BGCs. The resultant networks were visualized using Cytoscape version 3.8.2 (Su *et al.*, 2014).

2.2.8 Viral contribution to the propagation of polyketides

Viral contigs were identified using VirSorter2 as previously described (--include-groups all --min-length 5000 --min-score 0.5) (Guo *et al.*, 2021), checkV (v0.7.0) (Nayfach *et al.*, 2021) and some manual curation as described by (Guo *et al.*, 2021). All viral contigs that did not meet these criteria (viral_gene >0 OR score >=0.95 OR hallmark >2 OR viral_gene =0 AND host_gene =0), were discarded. Gene prediction was performed on the viral contigs with Prodigal (Hyatt *et al.*, 2010). These were further subjected to KS HMM screening as explained above. Contigs found to be in possession of non-redundant KS genes were subsequently analyzed using antiSMASH with all the analysis option (Blin *et al.*, 2021). For the taxonomic classification; vconTACT2 was used to classify viral contigs, and predicted open reading frames were queried against the NCBI Bacterial and Archaeal Viral RefSeq V85 with ICTV and NCBI taxonomy used in vconTACT2 (Jang *et al.*, 2019). To determine the origin of the viral KS we merge the 3 viral KS identified into clusters by antiSMASH with the non-redundant bacterial KS, aligned using MAFFT (Katoh *et al.*, 2019) with –auto option, trimmed using trimAl (Capella-Gutiérrez *et al.*, 2009) with noallgaps, and build the tree using Fasttree (Price *et al.*, 2009).

2.3 Results and Discussion

2.3.1 Taxonomic diversity and the detection of PKs using KS domains

Marine microorganisms are a valuable untapped source of diverse natural products (Tortorella *et al.*, 2018, Loges *et al.*, 2020). One of the most important classes of these natural products are polyketides (Della Sala *et al.*, 2014, Rego *et al.*, 2020). Despite their importance, little is known

about their diversity in SAO. To address this knowledge deficit, we assessed taxonomy and distribution of the microbial taxonomic composition in the SAO euphotic zones. This indicated that the SAO euphotic zone consisted of highly diverse groups of bacterial (*Proteobacteriota, Bacteroidota, Actinomycetota, Marinimicrobia)* and archaeal (*Crenarchaeota, Thermoplasmatota, Halobacterota)* phyla (Figure 2.2). These observed diverse taxonomic range is in agreement with other previous reports that have also looked at the SAO's euphotic zones using metagenomics approaches (Coutinho *et al.*, 2021, Ferreira *et al.*, 2022). Protein sequences predicted in metagenomic assemblies associated with these data were further assessed for the presence of potentially novel polyketides (PKs) using KS HMM profiles. This resulted in the detection of a total of 578 non-redundant putative KS sequences associated with 562 contigs with lengths ranging from 750 to 863838bp. Taxonomic assignments indicated that most of the contigs which possessed these KSs were overly associated with *Proteobacteriota* (52%), *Bacteroidota* (15%) and unassigned taxa (22%) phyla (Figure S1 &Table S2). Since Actinomycetes are the well-known sources of most antimicrobials today (Mast & Stegmann, 2019). To compare the KSs predicted in our data ecologically, we acquired the Tara oceans metagenomic data from Delmont et al. (2018) to help understand the distribution of PKs across the different oceanic regions and their microbial drivers. Protein sequences from the TARA metagenomes were also analysed following the same HMM based approach used to mine for the KS genes in our data. This resulted in the detection of 1654 non-redundant KS genes associated with 1554 contigs. Taxonomic analysis of the contigs indicated that these were also overly associated with *Proteobacteriota* (53%), *Bacteroidota* (11%) and unassigned taxa (16%) phyla (Figure S2). Similar to previous studies conducted by (Gavriilidou *et al.*, 2021, Rego *et al.*, 2021), our results further showed that marine *Proteobacteriota* possess a high abundance of genes that code for polyketide clusters

(Timmermans *et al.*, 2017, Buijs *et al.*, 2019). This is contrary to the studies that utilized KS degenerate primers, which indicated that *Actinomycetota* possess disproportionate numbers of PKS genes (Borsetto *et al.*, 2019, Rego *et al.*, 2020). Although, these observations may be due to the differences in the studied environments (as the studied environment is soil as against the ocean) or the bias in their approach (using primers originally used to amplify the domains of *Actinomycetota*).

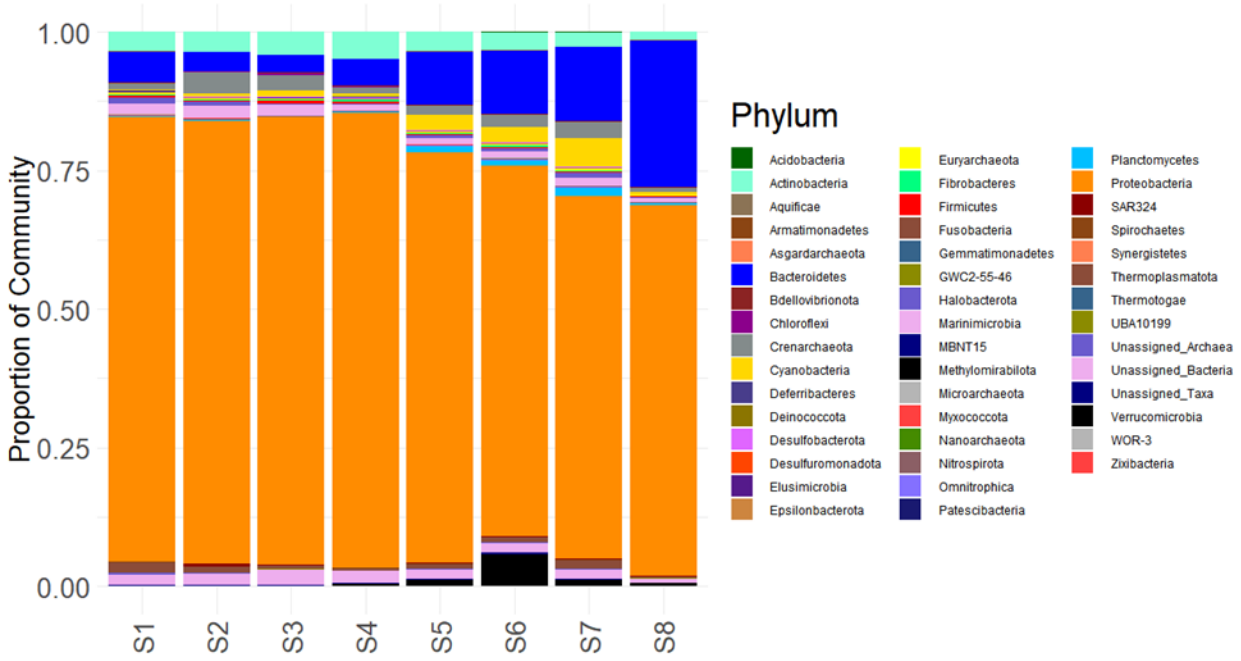

Figure 2.2: The taxonomy of the community composition at phylum level. The microbial taxonomic profile of the community is similar across all the sampling stations except for a few changes that occurred in some stations where the relative abundance of the *Marinimicrobia* and *Crenarchaeota* were increased while *the Bacteroidota, Cyanobacteria, Planctomycetes* and *Verrucomicrobia* decreased as seen in station 2 and station 3.

To determine their diversity, putative non redundant KSs from the SAO and TARA were merged and aligned with 652 KSs from NaPDoS and UniprotKB. A maximum likelihood phylogenetic tree reconstructed from these alignments indicated a great disparity between marine associated

KSs and those from both NaPDoS and UniprotKB KS (Figure 2.3). This maybe as a result of the origin, as most of the NaPDoS and UniprotKB KSs are from soil origin and mostly from *Streptomyces*. Before now, Streptomycetes has been vigorously explored and designated as the most important genus, producing up to 75% of all antibiotics used in clinical practice (Janardhan *et al.*, 2014, Kang & Kim, 2021) leading to having the highest number of KSs in the databases. In addition, many KS drivers could not be assigned to taxa, and a higher number of those that were assigned at phylum level could not be assigned at a lower level (Table S2). This might also suggest why the KSs are unique (Figure 2.3), as previously unexplored genera for natural product research are being reported as a source of new metabolites in the marine ecosystem on a regular basis (Tiwari & Gupta, 2012, Schorn *et al.*, 2016, Waschulin *et al.*, 2022). We also observed that KSs from SAO and TARA shared similar clusters, suggesting that niche specificity may select for highly similar groups of polyketides.

To assess the potential functions or compounds associated with both the SAO and TARA KSs that is similar to NaPDoS and UniprotKB own, Diamond blastp were conducted against NaPDoS and UniprotKB KSs. The results indicated that only 7 (of the 578) and 26 (of the 1654) KSs from SAO and TARA were predicted to likely produce similar compounds with known antitumor, antibiotics and antifungal compounds (Table 2.1). These results suggest the lack of sequence similarity shared between KSs from both SAO and TARA with those from NaPDoS and UniprotKB. A phylogenetic analysis of the 7 SAO, 26 TARA and their subsequent hits from NaPDoS and UniprotKB revealed that almost all of the SAO's KS form a distinct clade along with Tara ocean KS despite the fact that they got hit with the NaPDoS and UniprotKB natural product sequences, except for 11_k141_569222_2 that cluster with EpoC_Q9L8C8_H (Figure S3). This further indicates that the detection of KSs using HMM-based approaches contribute towards identification of novel

polyketides, and highlights that marine environment may potentially possess highly diverse natural products relative to terrestrial ecosystems (Kong *et al.*, 2010, Mayer *et al.*, 2013, Mayer *et al.*, 2019).

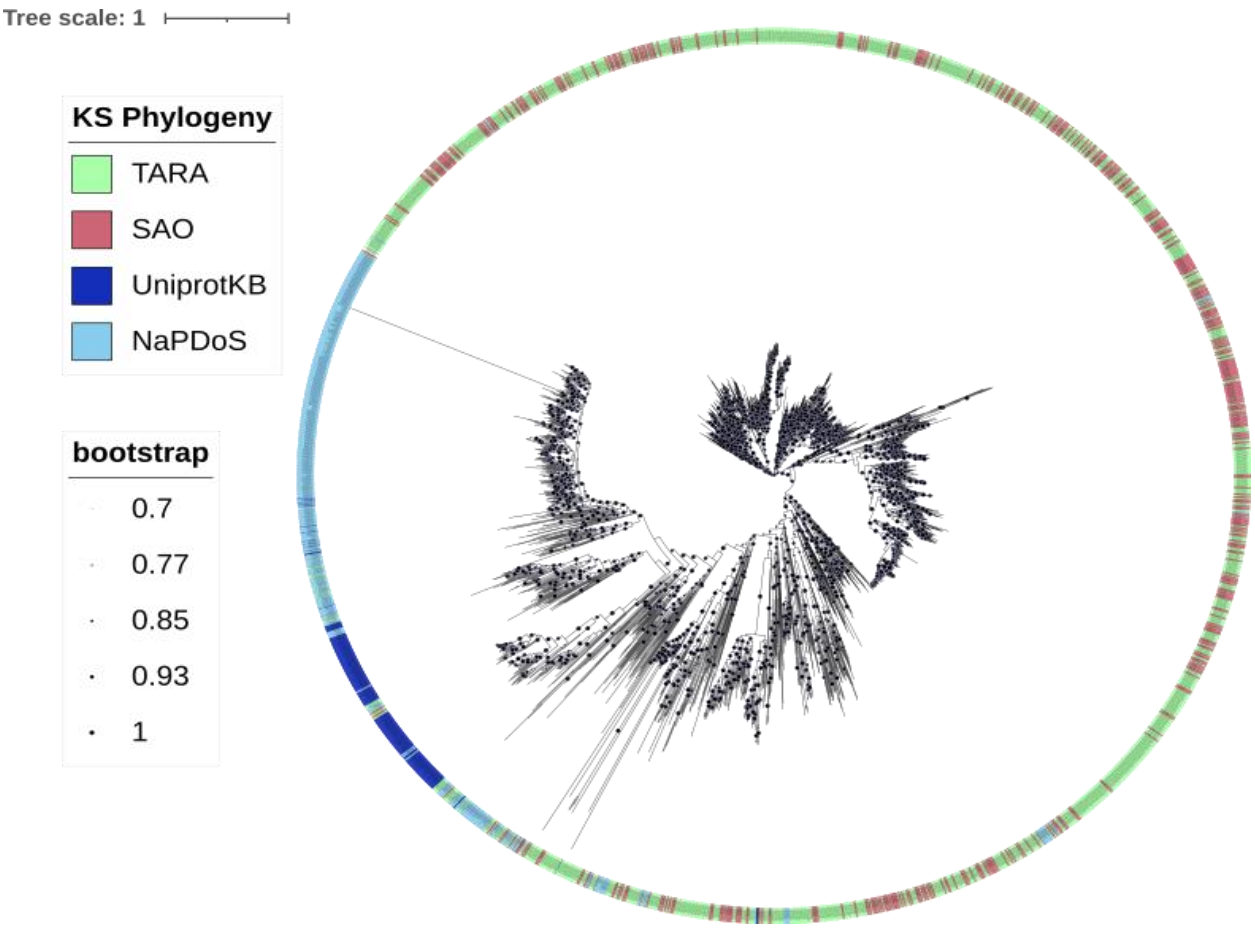

Figure 2.3: Phylogenetic tree of ketosynthase (KS) from the SAO (This study), Tara Ocean with NaPDoS and UniprotKB database KSs. There is a great disparity between the SAO, Tara and both NaPDoS and UniprotKB KSs. Most of the KSs from this study aligned to that of the Tara Ocean.

Table 2.1: Showing the KSs from both SAO and TARA that are similar to the known KSs from NaPDoS and UniprotKB with their functions

SAO/TARA	Napdos/UniprotKB hits ID	Compound name	Activity	Type of PKS	organisms	Compound family
1_k141_623915_1	KirAII_CAN89632_5T	kirromycin	antibiotics (inhibit protein synthesis)	Trans	Streptomyces collinus	complex linear polyketide peptide-bonded to sugar-like moiety
6_k141_159118_7	AknB_AF257324_KSa	aclacinomycin	antitumor	PKS2	Streptomyces galilaeus	
6_k141_111942_3	AlnM_ACI88862_KSb	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
6_k141_279575_6	AlnM_ACI88862_KSb	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
10_k141_548346_4	sp|N4WHE3|14-439	T-toxin	mycotoxin	PKS1	Bipolaris maydis	Reducing polyketide synthase
11_k141_569222_2	EpoC_Q9L8C8_H	epothilone	antitumor	hybrids	Sorangium cellulosum	Epothilone B is a 16-membered polyketide macrolacton e with a methylthiazole group connected to the macrocycle by an olefinic bond
23_k141_453723_2	mycos_Q9R9J1_T	mycosubtilin	antifungal agent	hybrid Trans	Bacillus subtilis	cyclic lipopeptides
TARA_RED-k99_598896_2	AknB_AF257324_KSa	aclacinomycin	antitumor	PKS2	Streptomyces galilaeus	
TARA_RED-k99_598896_6	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_RED-k99_6832670_5	sp|A0A0C6E0I7|9-444	betaenones	phytotoxins	PKS1	Phoma betae	Highly reducing polyketide synthase
TARA_RED-k99_5933435_40	actinorh_NP_629237_KSa	actinorhodin	antibiotic	PKS2	Streptomyces coelicolor	
TARA_RED-k99_8689661_24	SaqB_ACP19354_KSb	saquayamycin	anticancer and antibacterial activities	PKS2	Micromonospo ra sp	aromatic polyketide
TARA_RED-k99_3365992_8	AlnM_ACI88862_KSb	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_PSW-k99_2151472_21	sp|S0EEY3|12-443	fusarin C	mycotoxin	hybrid	Fusarium fujikuroi	
TARA_PSW-k99_58599_15	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_PSW-k99_448684_21	AlnM_ACI88862_KSb	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_PSW-k99_6686297_43	AlnM_ACI88862_KSb	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_PSE-k99_2493361_110	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide

TARA_PSE-k99_5662366_18	sp|B1GVX7|8-441	botcinic acid & its botcinin derivatives	antifungal activities	modular	Botrytis cinerea	Reducing polyketide synthase
TARA_PSE-k99_14366286_14	sp|B1GVX7|8-441	botcinic acid & its botcinin derivatives	antifungal activities	modular	Botrytis cinerea	Reducing polyketide synthase
TARA_PSE-k99_27046995_12	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_PSE-k99_27365180_20	sp|A0A0C6E0I7|9-444	betaenones	phytotoxins	PKS1	Phoma betae	Highly reducing polyketide synthase
TARA_PSE-k99_13046137_36	AknB_AF257324_KSa	aclacinomycin	antitumor	PKS2	Streptomyces galilaeus	
TARA_PSE-k99_25439984_63	AknB_AF257324_KSa	aclacinomycin	antitumor	PKS2	Streptomyces galilaeus	
TARA_PON-k99_1231741_3	TylGI_O33954_2mod	tylosin	antibiotics	modular	Streptomyces fradiae	macrolide antibiotic and bacteriostatic feed additive
TARA_PON-k99_5958106_3	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide
TARA_MED-k99_1829769_8	sp|A0A0C6E0I7|9-444	betaenones	phytotoxins	PKS1	Phoma betae	Highly reducing polyketide synthase
TARA_ION-k99_9870663_4	AknB_AF257324_KSa	aclacinomycin	antitumor	PKS2	Streptomyces galilaeus	
TARA_ASW-k99_1596744_18	sp|A0A3G9GQ29|35-459	chrodrimanin B	insecticidal activity	PKS1	Penicillium verruculosum	Highly reducing polyketide synthase
TARA_ASE-k99_5870610_8	actinorh_NP_629237_KSa	actinorhodin	antibiotic	PKS2	Streptomyces coelicolor	
TARA_ANW-k99_1566735_9	EpoD_Q9L8C7_2mod	epothilone	anticancer/ antitumor	modular	Sorangium cellulosum	Epothilone B is a 16-membered polyketide macrolacton e with a methylthiazole group connected to the macrocycle by an olefinic bond
TARA_ANE-k99_2156692_6	sp|S0EEY3|12-443	fusarin C	mycotoxin	hybrid	Fusarium fujikuroi	
TARA_ANE-k99_1507929_11	AlnL_ACI88861_KSa	alnumycin	antibiotic and cytostatic activities	PKS2	Streptomyces sp	aromatic polyketide

2.3.2 HMM efficacy

Based on the novelty discover with the KSs, we further explore the potential novelty of our approach by assessing whether the 562 contigs predicted to possess KS domains in the SAO data will be associated with known BGC that are available in the repositories using the antiSMASH pipeline. Of these, 560 contigs are above the 1000bp threshold that antiSMASH can identify into BGC. However, only 79 contigs with lengths 1487 to 53492 bp were predicted to possess BGCs (Table S3), highlighting the effectiveness of our approach in identifying novel polyketide structures as antiSMASH's detection algorithm is tailored towards known biosynthetic architecture. Of the 79 identified contigs, only 12 contigs were determined to possess non-truncated biosynthetic gene clusters (i.e. that did not fall on contig edge). However, of the total predicted BGCs, only 77 contigs were identified as polyketide clusters associated with pathways that relate to Arylpolyene, T1PKS, T1PKS-NRPS, heterocyst specific glycolipids, T3PKS, resorcinol and ladderane, respectively (Table S3). Despite the relationship of the 77 identified clusters with known BGC's pathways, the highest similarity recorded was 66% suggesting they might still be coding for a novel compounds (Table S3). Aside arylpolyene that has been designated the most abundant biosynthetic gene cluster family when categorizing BGCs from sequenced genomes spanning the prokaryotic tree of life (Cimermancic *et al.*, 2014), we identified more of T1PKS-related polyketide pathway clusters in our data. Previous study by Trindade-Silva et al. (2013) on Sponge microbiome (*Arenosclera brasiliensis*), Endemic to the Southern Atlantic Ocean (Trindade-Silva *et al.*, 2013) also detected a great diversity of T1PKS, suggesting that SAO might be rich in T1PKS pathway.

This was followed by assessing the diversity and potential novelty of BGCs associated with 1554 contigs predicted to possess KS domains in the TARA data. Of these, only 217 contigs were

identified into BGCs by antiSMASH. This analysis indicated that similar to our SAO findings, KS domain associated contigs from the TARA possessed T1PKS, Arylpolyene, T1PKS-NRPS, heterocyst specific glycolipids, T3PKS, resorcinol, ladderane, betalactone and hserlactone BGCs associated with polyketide clusters (Table S4). Similar to our study, most of their percentage similarity to known BGC pathway was lower than 66% except for the 2 that has 100% similarity to known ones. However, only 46 contigs were determined to possess non-truncated biosynthetic gene clusters.

2.3.3 Distribution of polyketides gene clusters across the SAO sampling stations

To assess the distribution of the polyketide clusters across the stations, individual metagenome's representative KS sequences and subsequently their contigs were retrieved. This resulted into the identification of 1255 contigs that were associated with KS domains of PKs. Of these, only 152 contigs were identified into their clusters with arylpolyene pathway occurring in high abundance in almost all the stations (Figure 2.4 & Table S5). However, only 33 contigs were determined to be non-truncated biosynthetic gene clusters (BGC) (i.e. that did not fall on contig edge). The low number of complete BGC retrieved; highlight the deficiency of using short reads to recover complete BGCs / polyketides.

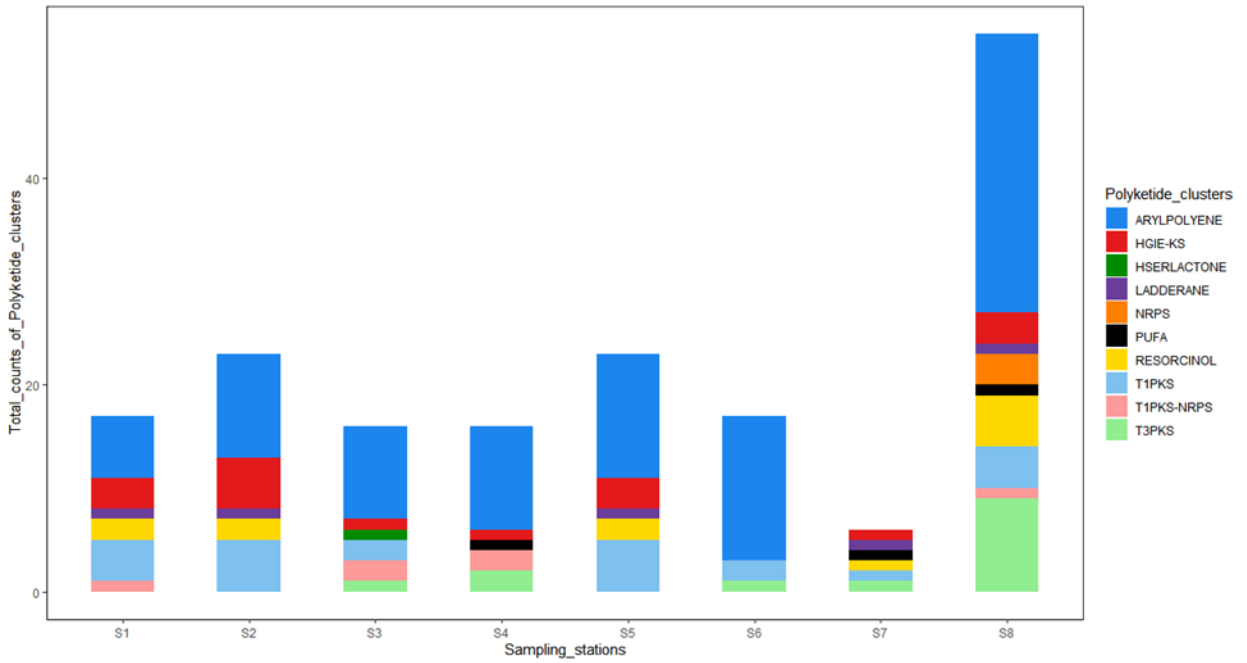

Figure 2.4: Distribution of the polyketides pathway across the sampling stations. The distribution of each type of polyketides identified as well as their total occurrence per station was shown. A total of 152 polyketide biosynthetic gene clusters were identified across the sampling station of the SAO, although only 33 were non-truncated biosynthetic gene clusters according to the antiSMASH. All hybrid clusters were separated into component parts to better visualize the variety of categories present except for the NRPS-PKS hybrid clusters (Table S5).

2.3.4 The diversity of BGCs in SAO, TARA and MiBIG

Because it has been demonstrated that BGCs frequently result in fragmented assemblies from short reads (Meleshko *et al.*, 2019). We then assess the diversity and distribution of BGCs in SAO and TARA in contigs lengths with >= 10kb. Candidate BGCs from both SAO and Tara were compared to those acquired from the MIBiG database by constructing Gene cluster families (GCF). Grouping gene clusters into larger families makes it easier to identify and prioritize potential new classes of BGCs and their associated products (Medema & Fischbach, 2015). Hence, BGCs from large sequencing datasets are increasingly being compared using GCF similarity networks (Cimermancic *et al.*, 2014, Doroghazi *et al.*, 2014, Ziemert *et al.*, 2014, Navarro-Muñoz *et al.*,

2020). The BGCs sequence similarity was first performed using the recommended distance matrix cutoff of 0.3 by BiG-SCAPE, which allows the prediction of BGCs with same compound into GCF. However, using this cutoff the SAO euphotic BGCs did not form GCFs with any of the BGCs from the MIBiG database, suggesting that these may encode for novel molecules (data not shown). However, since lower cutoffs are more appropriate for grouping BGCs that produce identical compounds while higher cutoffs provide a broader perspective on related families of natural products (Navarro-Muñoz *et al.*, 2020). We therefore increased the cutoff threshold to 0.7 so as to accommodate the relationship to the experimentally validated ones in MIBiG database (Figure 2.5). Despite this, the relatedness of the SAO euphotic BGCs to the experimentally validated ones is still low which suggests potential novel encoded secondary metabolites. Interestingly, there is an overlap of some PKS-other, T1PKS and BGC-others from SAO BGCs relative to Tara BGCs which further suggest probable niche specific PKs.

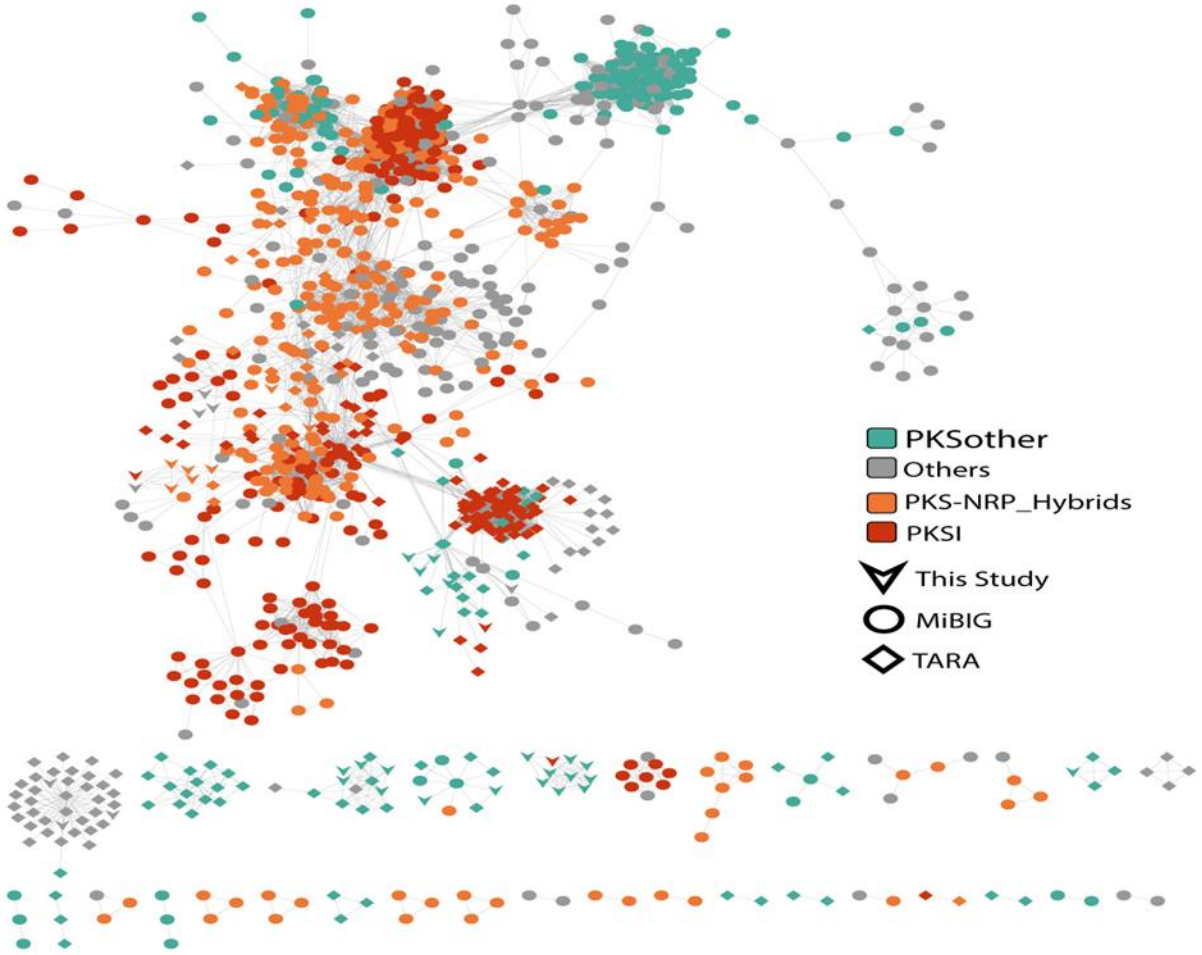

Figure 2.5: BGC similarity network of SAO, Tara and MiBiG BGCs. The network is based on 0.7 cutoff thresholds in order to accommodate relationship with MiBiG reference BGCs. Nonetheless, the relationship of the SAO BGC to the experimentally validated ones is still low. (PKSother includes T3PKS, HgIE-KS and their combination with T1PKS while Others includes PUFA, Arylpolyene, Ladderane or any of their combination).

2.3.5 Viral contribution to the propagation of polyketides

To determine the role played in the distribution of PKSs in the ocean by virus, we searched for the presence of viruses in our metagenomes using virsorter. From this analysis we detected a total of 7176 viral contigs and of these, only 58 contigs were predicted to possess KS genes. This initial

result suggested that polyketides presence is rare on the viral contigs since 0.81% of the total identified viral contigs harbors KS genes (Figure 2.6). However, subsequent determination of the BGCs of the non-redundant KS viral contigs using antiSMASH only detected 5.2% clusters (3 of 58 contigs) from the total non-redundant KS viral contigs (Figure 2.6). These detected clusters all belong to arylpolyene pathway. The presence of only arylpolyene clusters on the viral contigs could be as a result of their small cluster size, unlike the sizes of the multi-domain mega-synthases of T1PKS, as multi-gene assemblies are also probably constrained by the efficiency of the viral DNA packing process, which translates to their stability (Hatfull & Hendrix, 2011). BGCs are associated with mobile genetic elements (Ruzzini & Clardy, 2016), and viruses, the most abundant biological agents on earth, are a major component of mobile genetic elements in bacteria. Considering the relatively high abundance of arylpolyene clusters across the different sampling sites, we posit that virus likely play a significant role towards their transmission between bacterial cells. In order to ascertain further whether the viral KS are of bacteria origin, we built a viral and bacterial KS phylogeny using the 3 viral KS that was identified into cluster. We found out that the viral KS formed a cluster with the bacterial KS which further substantiate our hypothesis (Figure 2.7). Studies have also shown that the viruses can encode different compounds that are potentially valuable for their hosts such as toxins and virulence factors (Boyd, 2012, Jamet *et al.*, 2017, Castillo *et al.*, 2018). As a result, we can also hypothesize that these viral BGCs improve their host's fitness against ROS in this environment. Surprisingly, all of the viral contigs harboring BGCs belong to an order of unclassified virus.

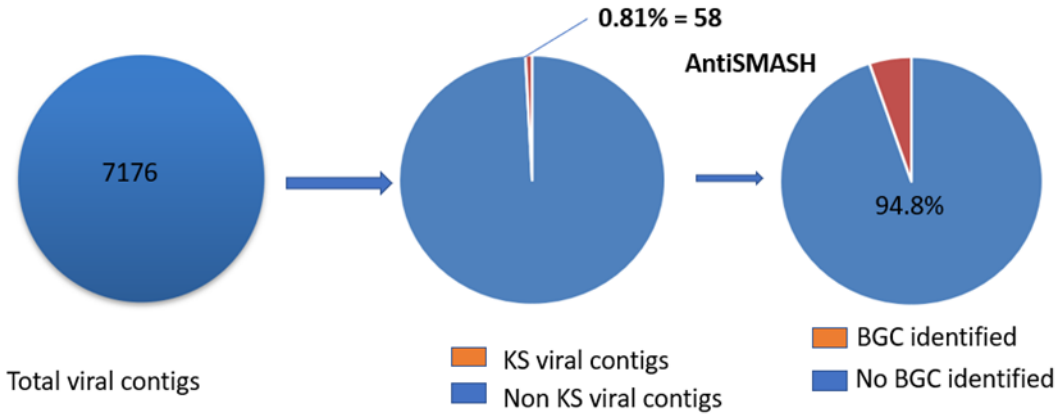

Figure 2.6: The proportion of the total viral contig identified in the metagenomic data together with the proportion that harbor polyketides is shown. Out of the total of 7176 viral contigs identified, 58 were found to contain KS genes. Only 5.2 percent (3 of 58) of the non-redundant KS viral contigs BGCs were identified using antiSMASH, and they were all identified as arylpolyene.

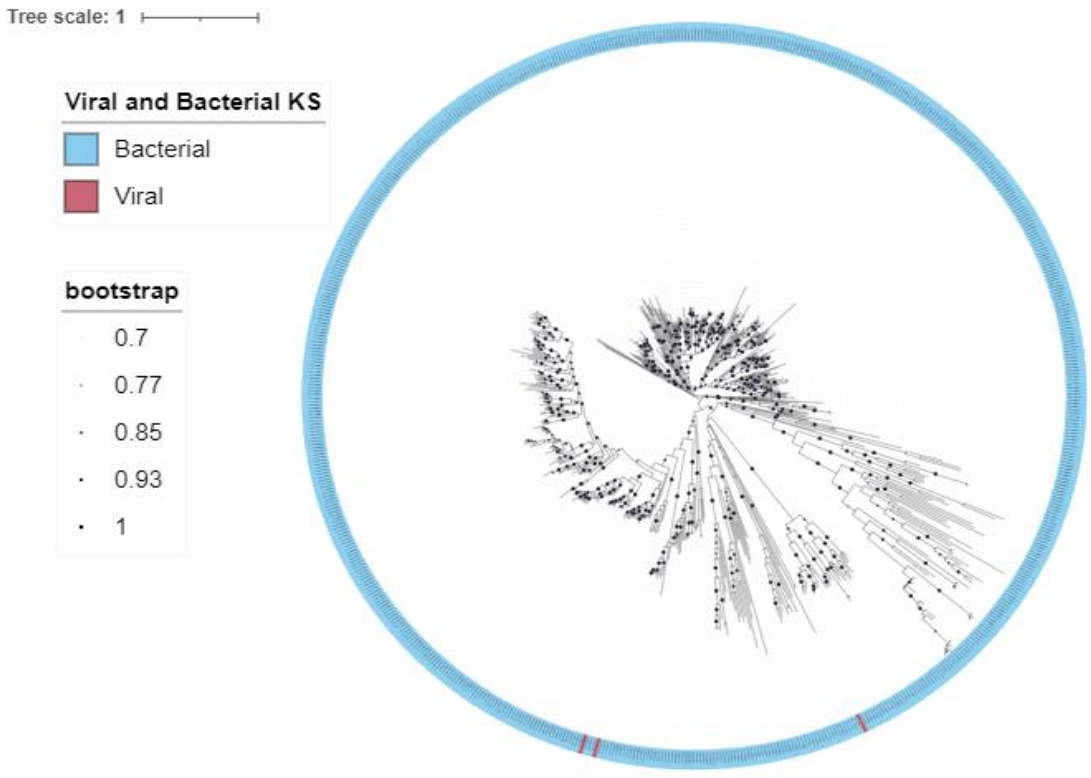

Figure 2.7: Phylogenetic tree of the viral KS and bacterial KS. The viral KS clustered with the bacterial KS.

2.4 Conclusion

The use of the contextualize approach in this study has revealed an overview of the biosynthetic diversity and subsequent identification of the potential novel biosynthetic pathway from a metagenomic data which antiSMASH could not detect. This has provided a new framework for future survey of biosynthetic potential of a microbiome prior to a major biosynthetic project in this era of continuous increase of metadata in public databases. The study confirms and expands our knowledge of the biosynthetic potential of the SAO euphotic zone, as large percentage of the identified polyketide clusters had no or low similarity to known biosynthetic gene clusters. This is indicative of an extensive novel metabolites being encoded in the euphotic zone of the SAO, which might lead to the discovery of novel compounds. This data is an interesting addition to the already known literature on polyketides of the marine ecosystem. Furthermore, our approach can serve as a good starting point for bio-prospecting research because it focuses efforts on identifying biosynthetically rich environments and candidate BGCs for heterologous expression. Future studies, which will integrate the different water column of an ocean that is limited in iron rich dust deposition, will be required to accurately elucidate the dynamics associated with the polyketide biosynthetic pathways, and also the exploration of these novel clusters for compounds production and activity.

Chapter Three

Genome centric analysis of biosynthetic gene clusters from the Southern Ocean

Abstract

Natural product-based drug discovery, primarily from terrestrial environments, has resulted in the continuous 'rediscovery' of known compounds. Consequently, the search for novel compounds has recently shifted to marine environments, which represent untapped and uncharacterized sources of metabolites. The Southern Ocean (SO) is limited in trace elements including iron. This limitation presents an ecological stressor which may result in microbial adaptation and evolution. However, little is known regarding the biosynthetic potential of the SO, and the variation in secondary metabolite production by oceanographic provinces and seasons remains unknown. Here, we used genome centric metagenomics to assess the biosynthetic potential of three SO zones (epipelagic, mesopelagic, and bathypelagic) over two seasons. We recovered a total of 109 medium to high-quality MAGs out of which biosynthetic gene clusters (BGCs) were detected in 54. In total, 21 BGC types were recovered with at least seven different types in each zone. The highest BGC diversity was observed in samples from bathypelagic zones. Terpene and PUFA BGCs were abundant in this environment and have been suggested to play important roles in ecosystem function. Interestingly, our analysis showed that SAR324 were the most biosynthetically diverse microbes in the SO. Taken together, our findings provide insights regarding the genetic potential of biosynthetic gene clusters in the SO for different specialized metabolites in different zones.

3.1 Introduction

Secondary metabolites represent unparalleled sources of bioactive compounds, which have been used in medicine and agriculture (Newman and Cragg, 2016, Scherlach and Hertweck, 2020). These metabolites are auxiliary compounds produced by microorganisms and serve several beneficial roles including as agents in nutrient acquisition, cell to cell communication and bacterial defense (Hibbing et al., 2010, Davies, 2013, Newman and Cragg, 2020). Secondary metabolites are encoded by biosynthetic gene clusters (BGCs) (Lee et al., 2021). These BGCs are groups of two or more proximally located genes in the genome (Medema et al., 2015). There are several different structural classes of BGCs including polyketide synthases (PKS) (Kealey et al., 2021), non-ribosomal peptide synthetases (NRPS) (Martínez-Núñez and Rodríguez-Escamilla, 2020), ribosomally synthesized and post-translationally modified peptides (RiPPs) (Hug et al., 2020), terpenes (Muchlinski et al., 2019), betalactone (Suciu et al., 2021), bacteriocins (Zhang et al., 2022). Of these diverse classes, PKS and NRPS are the two largest and have been targeted for most natural product discovery due to their capacity to produce compounds with wide ranging activities (Wang et al., 2014, Dror et al., 2023).

The discovery of these natural products has largely been performed using conventional methods, based on screening cultivable microbes in the laboratory (Hata et al., 1971, Ikeda et al., 1983, Medema and Fischbach, 2015, Katz and Baltz, 2016). These approaches mostly rely on the isolation of soil actinomycetes as prominent sources for the discovery of natural compounds (Kim et al., 2021). Although these cultivable actinomycetes possess diverse types of antimicrobial compounds per genome due to their large genomes (Hamedi and Mohammadipanah, 2015, Siro et al., 2022), the functional and biosynthetic potentials of uncultivated microbiomes from other ecosystems remain underrepresented (Mori et al., 2018, Paoli et al., 2022). Recent advances in the

field of metagenomics have enabled the reconstruction of metagenome-assembled genomes (MAGs) of complex microbiomes (Albertsen et al., 2013, Liu et al., 2021, Arikawa et al., 2021, Duncan et al., 2022). This approach has facilitated the studies of biosynthetic potential of both cultivable and uncultivable organisms, directly from the environment (Wu et al., 2022). Exploring diverse types of biosynthetic genes using a taxon centric approach helps to tie novel compounds to known taxa and their ecological contributions in the environment (Chen et al., 2020). Following this approach, recent studies have shown that, despite the fact that the ocean covers more than 70% of the earth's surface (DeLong, 2007, Fauville et al., 2019), natural products encoded by microorganisms in this environment remain largely underexplored (Paoli et al., 2022). Rego et al. (2021) further indicated that this is particularly true for marine polar regions, highlighting that only a handful natural products have been derived from these due to logistic challenges. There is an urgent need to discover natural products encoded by BGCs from geographically diverse marine environments (Khalifa et al., 2019, Nweze et al., 2020).

The Southern Ocean (SO) is a unique and ecologically crucial system responsible for the global uptake of 40% of anthropogenic carbon dioxide (CO_2) from the atmosphere (Sabine et al., 2004, Landschützer et al., 2015, Frölicher et al., 2015). This ocean is also the largest high-nutrient low chlorophyll (HNLC) region on earth and is characterized by a lack of limiting nutrients such as iron (Queguiner, 2013, Hunt et al., 2021). These physico-chemical properties have been shown to influence metabolic mechanisms of SO microbial communities leading to adaptation and evolution (Debeljak et al., 2019, Boyd, 2019, Fourquez et al., 2020). However, despite such insights, we lack information regarding biosynthetic potential associated with the diverse microbial communities in the SO. Furthermore, the interactions of organisms with their physical environment may result in differences in the biotic diversity (Silknetter et al., 2020), trophic structure (Kaur and Dutta, 2020,

Erktan et al., 2020), energy (Abrahms et al., 2018) and the materials that flow between them (Odum and Barrett, 1971, Sharma and Sharma, 2012, Jeronen, 2019).

The physical, chemical, and biological properties provide a basis for distinguishing oceanic water bodies (Sherman et al., 2005, Spalding et al., 2007, Longhurst, 2010, Proud et al., 2017, Reygondeau et al., 2017). Based on these criteria, the ocean is divided into epipelagic, mesopelagic, bathypelagic, abyssopelagic and hadal zones (Netburn, 2018, Xue et al., 2020). However, the extent to which these zones affect the distribution of specialized secondary metabolites remains unclear (Hay, 1996). In addition, little is known regarding the diversity of bacterial and archaeal species which harbour these secondary metabolites. Previous studies suggest an increase in both the taxonomy and functional richness of the ocean with depth (Sunagawa et al., 2015). In addition, studies have also shown that seasonality affects microbial community composition (Salmaso et al., 2018, Bolaños et al., 2021). However, no study has investigated the effects of these variations on the composition of BGCs. Based on these observations, we therefore hypothesize that microbiota in the SO may possess novel BGCs, and that there would be a proportionate increase in the diversity of BGC with depth.

Secondary metabolic gene clusters were screened from metagenome assembled genomes produced in a recent study by Castillo et al (in review). We used the antiSMASH and NaPDoS pipelines following established methods (Ziemert et al., 2012, Blin et al., 2021). We evaluated the potential of SO microbial communities for future research on potential novel drugs, focusing on three major layers (epipelagic, mesopelagic, and bathypelagic zones) over two seasons. A total of 166 secondary metabolite clusters were identified from 54 MAGs recovered from our samples. This resulted in the detection of 21 BGC types with at least 7 different types occurring in each zone. Of these, a total of 23 PKS and 1 NRPS clusters were recovered. Overall, we found no differences in

the diversity of BGC due to seasonal variations. The distribution of broad BGCs types was mostly consistent in the bathypelagic zone across the two seasons. Overall, we found high relative abundances of several gene clusters and these were observed to be more common in epipelagic and mesopelagic zones.

3.2 Materials and Methods

3.2.1 Sample collection

Sampling was carried out aboard the *RV SA Agulhas II* in June (winter) and October (spring) 2019 from the Southern Ocean at a single station (-51.40167°S latitude, 0.00115°E and -51.4°S latitude, 0.001°E longitude in winter and spring, respectively). At each depth (200m, 500m, and 2600m), 5 liters of water were collected using a CTD rosette multisampler and filtered through 0.2 m Polyethersulfone (PES) filter membranes (Merck, RSA) resulting in a total of six samples. The membranes were kept at -80°C until DNA was extracted.

3.2.2 DNA extraction and sequencing

DNA extraction was carried out as previously described by (Hirai et al., 2017). Briefly, filter membranes were cut and incubated for 10 minutes at 60°C, in a DNA extraction buffer (400 mM Tris-HCl pH8, 60 mM EDTA pH8, 150 mM NaCl, and 1% SDS). To precipitate organic matter, 120 µl of 3 M potassium acetate buffer was used, followed by five minutes incubation on ice and one minute centrifugation at 15000 rpms at 4°C. The DNA extracted from the supernatant was then purified in two steps by first adding 800 µl of solution C4 and 500 µl of solution C5 from the DNAeasy Power soil kit (Qiagen, Hilden, Germany). The concentrations of nucleic acids were determined using the Qubit 4 Fluorimeter (Invitrogen by Thermo fisher scientific, Singapore) in

accordance with the manufacturer's instructions. High-quality DNA of each sample was sequenced using an Illumina NovaSeq S4 2x150 instrument.

3.2.3 Quality control, assembly and MAGs construction

FastQC v0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used for quality control. Raw reads and low-quality bases were filtered using Trimmomatic v0.36 (Bolger et al., 2014). The resultant high quality reads were assembled using metaSPAdes v3.13.0 (Nurk et al., 2017). To determine the percentage of assembled reads, quality-filtered sequences were mapped back to the contigs using BBMap Aligner (BBMap) (minid=0.90 maxindel=3) (Bushnell, 2014). The alignments were 'sorted' and 'indexed' using SAMtools v1.9, as previously described (Li et al., 2009). CONCOCT v1.1.0 (Alneberg et al., 2014), MetaBAT2 v2.12.1 (Kang et al., 2019), and MaxBin2 v2.2.6 (Wu et al., 2016) were then used to construct the MAGs from the resulting contigs. CheckM v1.0.18 was used to evaluate the completeness and contamination value of each MAG (Parks et al., 2015). DasTool v1.1.2 (Sieber et al., 2018) was used to dereplicate the medium to high-quality MAGs (>50% completeness and <10% contamination to >90% completeness and <5% contamination) from each metagenome.

3.2.4 Community composition, taxonomic assignment and phylogeny

Taxonomic composition was determined using singleM v0.12.1 (https://github.com/wwood/singlem) on the raw reads with default parameters. The GTDB-tk pipeline v1.6.0 release 89 was used to assign taxonomy to each of the medium to high-quality MAG (Chaumeil et al., 2020). GTOTree v1.4.7 (-G 0.2) was used to construct a phylogenetic tree

using all MAGs (Lee, 2019). The phylogenetic tree was then visualized and annotated using iTOL version 6.4.1 (Letunic and Bork, 2021).

3.2.5 Detection of BGCs in MAGs

All the medium to high-quality MAGs were subjected to BGC detection using antiSMASH (Blin et al., 2021). For this, all analysis options were enabled (KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREFinder, Cluster Pfam analysis, Pfam-based GO term annotation, TIGRFam analysis). Using python, a heatmap representing MAGs by phylum and the number of BGCs detected was created.

3.2.6 Comparison of BGC abundance between three SO Ocean zones

To determine the relative abundances of BGCs in individual metagenomes, BBMap was used to recruit BGCs from the respective quality trimmed reads as detailed previously (Bushnell, 2014). The relative abundance of each BGC, per sample, was estimated from the resulting bam files and converted into a table using a custom wrapper script from BamM (https://github.com/ecogenomics/BamM). Coverage values were calculated and represented as relative abundance using the "tpmean" algorithm, which is normalized for the size of each metagenome in bases and the length of each BGC as previously described (Gregory et al., 2019, Gazitúa et al., 2021). The resultant data were used to assess the distribution and abundance of each BGC type among the three sampling zones and plotted using ggplot2 in R v4.0.3 (Wickham and Wickham, 2007, Team, 2013).

3.2.7 Analysis of the KS Domains from PKS Clusters

KS domains predicted in PKSs linked to each MAG, were used for comparisons. These were compared with sequences in the NaPDoS database using maximum likelihood phylogenetic trees.

The analyses were conducted using the NaPDoS web server (Ziemert et al., 2012). iToL was used to visualize and annotate the trees (Letunic and Bork, 2021).

3.3 Results and discussion

3.3.1 Community composition of Southern Ocean

To assess the overall taxonomic distribution of the microbial communities in our data, a total of six metagenomes (12 raw reads) were analyzed. We used single-copy marker genes of 14 ribosomal proteins for the analysis and the data revealed that the dominant microbial lineages were similar both in winter and spring. For instance, we observed a decrease in the abundance of *Actinomycetota*, with concomitant increase in the abundance of *Proteobacteriota* in the winter epipelagic and mesopelagic zones (Figure 3.1). On the other hand, the proportion of unclassified bacterial sequences in the community composition was relatively high, across all depths (Figure 3.1). This suggests that there is still a large proportion of bacterial novelty in the SO. This result is consistent with previous studies, where a large number of unclassified bacterial sequences have also been reported in the SO (Phoma et al., 2018, Phoma and Makhalanyane, 2021).

Metagenome assembled genomes were used to explore potentially novel taxa in the SO. This approach has been shown to increase the likelihood of identifying complete BGCs (Rego et al., 2021). Binning of six metagenomes, from epipelagic, mesopelagic and bathypelagic zones yielded a total of 109 high and medium quality (Table S6) MAGs (89 bacterial and 20 archaeal) (Bowers et al., 2017). Of these, 56 MAGs were recovered from winter while 53 were recovered from spring (Table S6). Taxonomic assignments of these using GTDB-tk indicated that these represented a total of 11 phyla (9 bacteria and 2 archaea), which include *Proteobacteriota* (28.4%), SAR324

(10.1%), *Verrucomicrobiota* (9.1%), *Chloroflexota* (9.1%), *Actinobacteriota* (8.2%), *Marinisomatota* (5.5%) and *Thermoplasmatota* (17.4%) from bacteria and archaea, respectively (Figure 3.2). Similar taxonomic dominance was also reported from the study conducted by Milici et al. (2017) on particle-associated and free-living microorganisms from SO, which corroborated our result. Pairwise comparison of the MAGs against GTDB-tk reference genomes using ANI analysis indicates that most genomes may represent distinct species (61 Bacterial and 6 Archaeal) (Table S7). In addition to the candidatus phyla, one unique MAG assigned to TA06 phylum was uncovered with 57.26% completeness and 0.75% contamination (Table S6). As a result, most of the MAGs may fill some phylogenetic gaps in the available genome collection, making them useful for inferring details about microbial phylogenetic relationships.

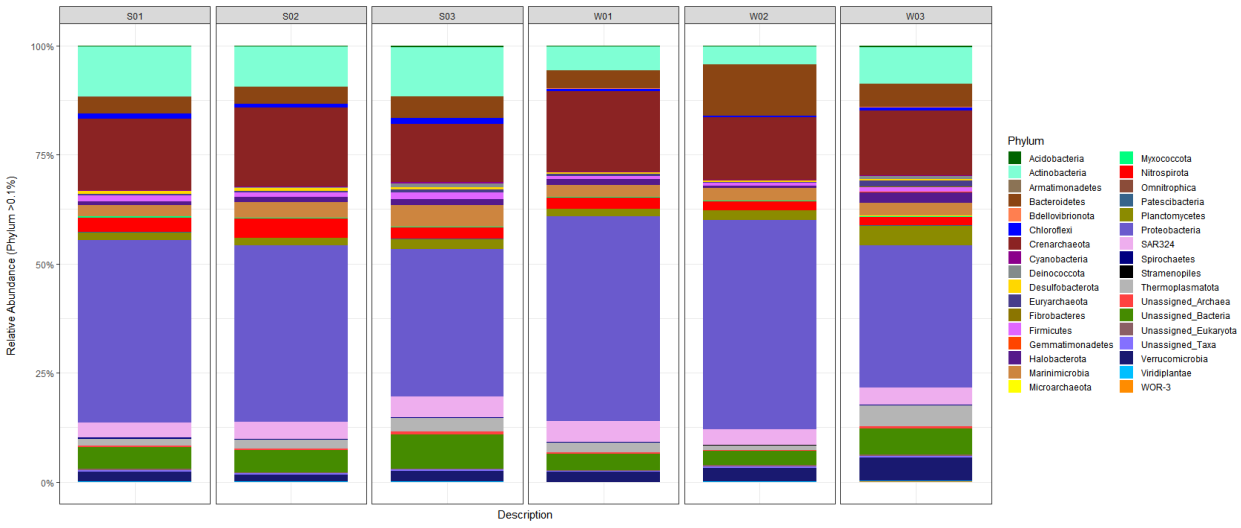

Figure 3.1: A snapshot of the community composition of Southern Ocean showed similar patterns of relative abundance in both spring and winter. S01, S02 and S03 represent 200m, 500m and 2600m respectively in spring while W01, W02 and W03 represent 200m, 500m and 2690m respectively in winter.

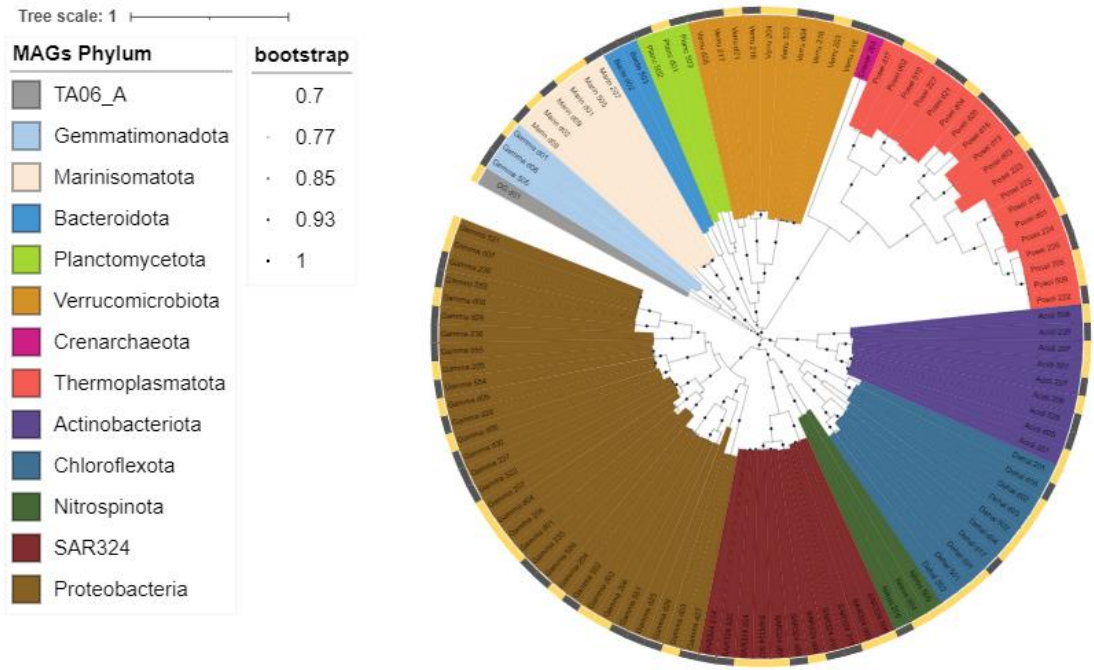

Figure 3.2: Maximum likelihood phylogenetic tree of microorganisms recovered from the Southern Ocean. In total, 109 medium to high-quality metagenome assembled genomes were classified into phyla using color-coded taxonomies. Grey and yellow outer ring represent winter and spring MAGs, respectively. *Proteobacteriota* and *Thermoplasmatota* accounted for the large proportion of the medium to high-quality MAGs.

3.3.2 Terpenes and HgIE-KS are overrepresented amongst SO microbial communities seasonally

To explore the diversity of BGCs possessed by microbiota in the SO, we detected a total of 166 biosynthetic gene clusters in 54 MAGs (40 bacterial and 14 archaeal) (Figure 3.3 & Table S8). However, despite the recovery of nearly complete genomes, the majority of the identified BGCs remain on contig edges (i.e., incomplete BGCs) likely due to the use of short read sequencing technology for this study (Van Goethem et al., 2021). Although short reads have limitations in the assembly of complete BGCs (Libis et al., 2019), these still provide the basis for exploring partial genomic fragments which may be associated with potentially novel BGC types (Robinson et al., 2021). We identified an overrepresentation of BGCs associated with Terpenes and HgIE-KSs from the MAGs. These Terpenes and HgIE-KSs were associated with 8 (comprising 7 bacterial and 1 archaeal) and 5 (4 bacterial and 1 archaeal) different phyla, respectively. The majority of these BGCs were detected in SAR324, *Verrucomicrobiota*, *Thermoplasmotota* and *Chloroflexota*. Overall differential abundance estimates suggest that these phyla harbor 45, 32, 22 and 21 BGCs, respectively. This proportion of high number of BGCs in underexplored phyla is consistent with previous reports by Borsetto et al (2019), Sharrar et al (2020) and Geller-McGrath et al (2023) where phyla such as *Verrucomicrobiota*, *Chloroflexota* and *Thermoplasmatota* were observed to encode BGCs. Similarly, higher terpenes abundance has been reported by Rego et al (2021) and Paoli et al (2022) in Artic and global ocean microbiome, respectively.

Interestingly, we observed a genome of the candidatus phylum TA06 (Huang et al., 2019) (Table S6),), which had T3PKS, Terpene, Ladderane, and Sactipeptide clusters. To the best of our knowledge, no secondary metabolisms for this phylum have been reported; this suggests that TA06

may encode diverse BGCs that remain to be explored. Overall, based on similarity to known sequences, our analyses suggest that the majority of BGC identified may be novel. The similarity values ranged between 0 to 28%, with the exception of a BGC affiliated with a *Bacteriodota* MAG, which had a 58% similarity to known arylpolyene (Table S8).

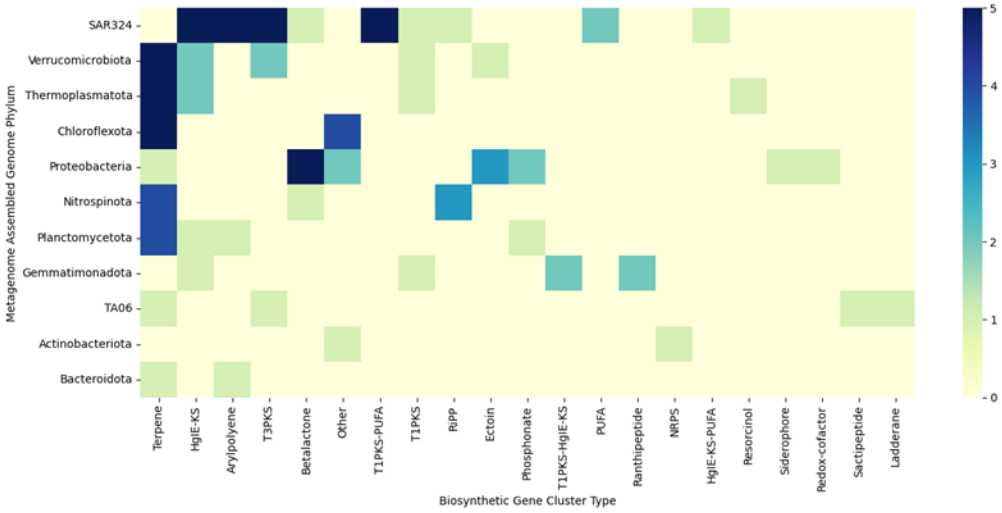

Figure 3.3: BGCs detected by antiSMASH in 11 different assigned phyla (9 bacteria and 2 archaea). MAGs are arranged top to bottom by the highest to the lowest number of BGCs been harbored while gene cluster types were arranged from left to right by the abundance of BGC type. The abundance was adjusted to the scale of 5 for better view.

To assess the ecological distribution of BGCs between two seasons, read recruitment of the BGCs identified in the metagenome assembled genomes to the trimmed reads was performed (Methods

3.2.6). This recruitment also allowed us to assess BGCs in the SO metagenomes that were not assembled into the metagenome assembled genomes. Biogeography may have a large impact on organism distribution (Hill et al., 2017), and various depths in the ocean may represent distinct biogeographic regions. Given the distinct niche, it is reasonable to predict that these niches may select for distinct microbial communities and, in turn, harbor concomitantly unique biosynthetic potential. As a result, the abundance and diversity of the biosynthetic potential of the Southern Ocean was investigated across three zones/depths over two seasons. (Figure 3.4).

In spring, the abundance of BGC decreases with depth, with the epipelagic zone having the highest abundance, whereas during the winter season, the abundance of BGC was at the highest in the mesopelagic zone (Figure 3.4). These two prolific zones (epipelagic in spring and mesopelagic in winter) were significantly abundant in HgIE-KS, terpene and PUFA or PUFA-T1PKS.

We also observed that terpene, T3PKS, Other-BGC and arylpolyene BGCs were present throughout the three zones/depths and also in the two seasons. Terpene gene clusters are responsible for the production of a diverse range of functionally different secondary metabolites including carotenoids and geosmin, which provide protection against reactive oxygen species and the removal of excess metabolites when microorganisms are stressed respectively (Watson, 2003, Pattanaik and Lindberg, 2015). Furthermore, because of their polyene structural system, arylpolyene is also similar to carotenoids (Schöner et al., 2016) which further suggest to be playing similar role. Terpene may also be used for communication purposes by bacteria, both within themselves and even with fungi (Schmidt et al., 2017). The presence of Other-BGC refers to the presence of clusters that contain secondary metabolite-related proteins but that does not fit into any class known (https://docs.antismash.secondarymetabolites.org/glossary/#current-types). This

may suggest the ability of the BGCs having an interesting and possibly novel function that may be unique to the environment.

Surprisingly, the bathypelagic zone of the two seasons had the lowest abundance of BGC but the greatest diversity of BGC types (Figure 3.4). This is similar to previous studies that have shown increase in taxonomy and functional diversity with depth (Sunagawa et al., 2015, Walsh et al., 2016). The observed differences in the abundance of BGCs between winter and spring could be linked to changes in microbial community structure which in turn is influenced by change in season (Bandekar et al., 2016, Kumar et al., 2019). For example in spring high abundance of microbial communities in the epipelagic zone have been shown to be influenced by high productivity as opposed to winter conditions (Bolaños et al., 2021).

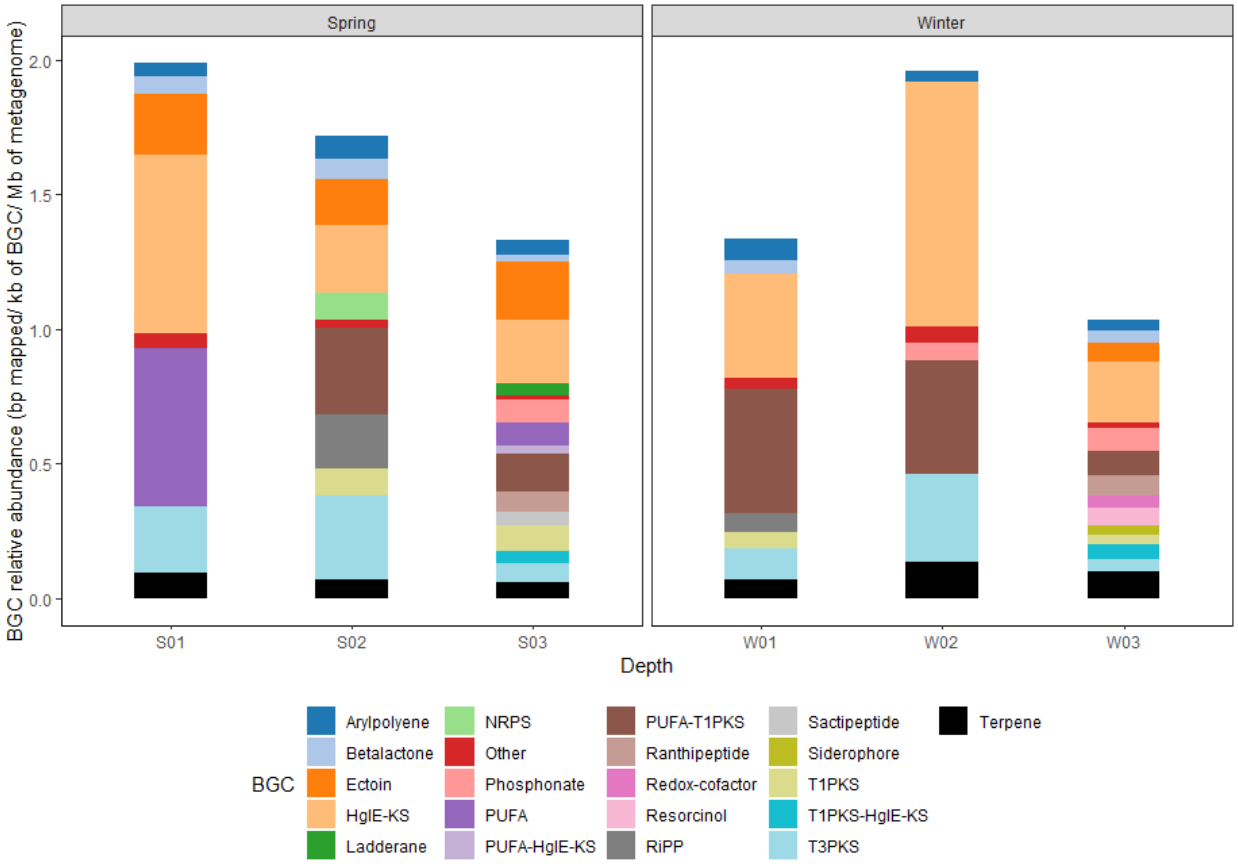

Figure 3.4: Relative abundance of BGCs in each metagenome (coverage values normalized by metagenome size and BGC length) from different zones in two seasons. S01, S02 and S03 represent 200m, 500m and 2600m respectively in spring while W01, W02 and W03 represent 200m, 500m and 2690m respectively in winter.

3.3.3 SAR324 phyla encodes for disproportionately diverse PKS

SAR324 is a monophyletic group of marine Deltaproteobacteria that has recently been reclassified into its own phylum (Waite et al., 2020). These are found across all ocean depths (Wright et al., 1997, Pham et al., 2008) which is also demonstrated in this study (Table S6). Studies have been made to elucidate the potential metabolic capabilities of this group of organisms (Swan et al., 2011,

Sheik et al., 2014, Cao et al., 2016) but little is known about their biosynthetic potential. We found several biosynthetic gene clusters in SAR324 including T1PKS, T3PKS, Arylpolyene, PUFA, RiPP, HgIE-KS and betalactone (Figure 3.3). In contrast to a previous study where a single small PKS pathway was observed from a SAR324 genome (Baltz, 2017). This highlights the high biosynthetic potential of this new phylum aside the common metabolic traits (such as the ability to fix inorganic carbon, ability to oxidize sulfur among others) that have been reported before. Interestingly, PUFA and PUFA related metabolites (T1PKS-PUFA and HgIE-KS-PUFA) were exclusively discovered in SAR324. This may suggest PUFA to be one of the surviving mechanisms utilized by SAR324 for their occurrence across all ocean depths as PUFAs have long been thought to be effective modulators of membrane fluidity (Yoshida et al., 2016). Several studies have also reported PUFA's cold adaptation role in polar bacteria and algae (Alcaíno et al., 2015, Schulze et al., 2019, Morales-Sánchez et al., 2020). Surprisingly, despite the fact that SAR324 has the highest number of BGCs in this environment, the most abundant BGC type (terpene) was not found in them.

We observed a total of 20 KS domain sequences from the T1PKS clusters of this study, most of which belong to the SAR324 phylum (Table S9). Although the criterion for determining functional gene novelty is still being debated, sequences with less than 80 percent hits may indicate that the domain of interest may contribute to natural products that have yet to be identified, according to NaPDoS (Ziemert et al., 2012). Several of the SAR324, *Gemmatimonadota*, *Verrucomicrobiota*, and *Thermoplasmatota* KS domain sequences identified in this study, however, aligned with poly-unsaturated fatty acid (PUFA) compounds from the Azotobacter and Shewanella genus with less than 50% identity (Figure 3.5 & Table S9). Firstly, this suggests that these polyketides may be

novel. The observation of different BGCs in SAR324 genomes suggest that they may play important ecological roles.

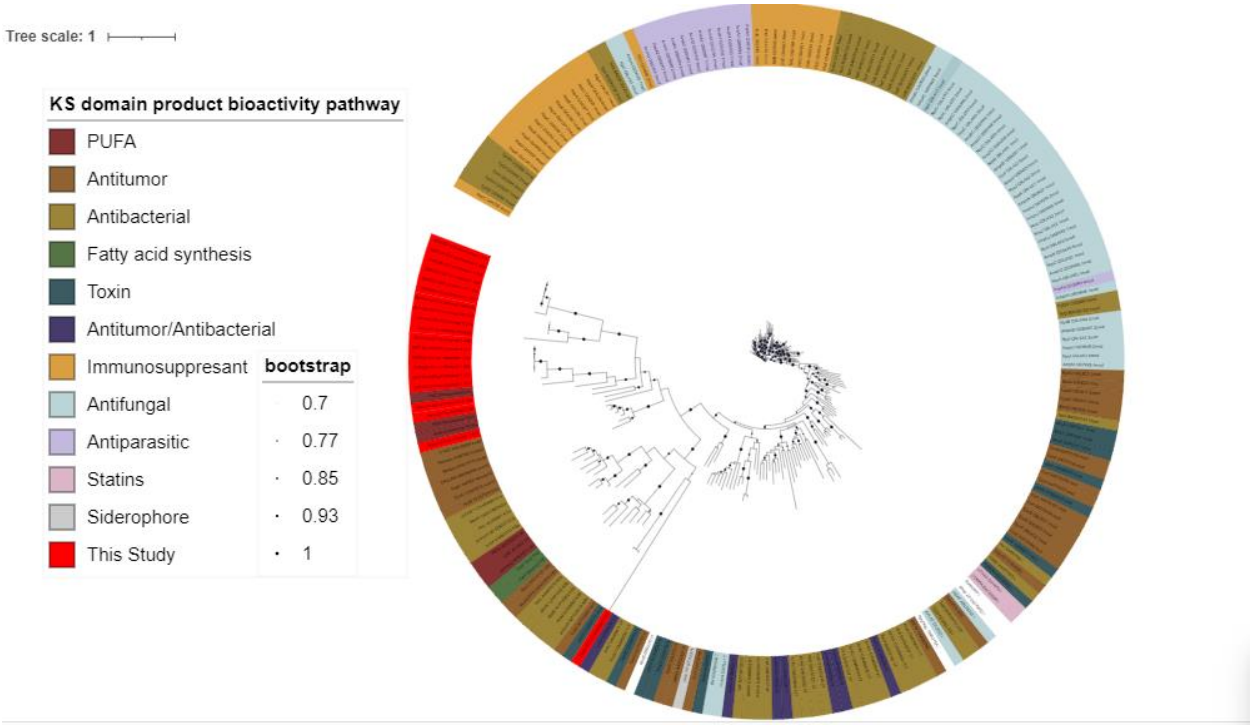

Figure 3.5: Maximum likelihood phylogenetic tree of ketosynthase domains based on the NaPDoS domain database. The bioactivity of natural products is color coded and represented on the outer ring. Several of the KS domains identified in this study were closely related to PUFA compounds. *Thermoplasmatota's* KS domain was found to be highly aligned with antitumor compounds.

Secondly, the results suggest the importance of PUFA in this environment as these compounds are well known for their adaptability in photo-protection of microorganisms against UV radiation (Nishida et al., 2006, Chen et al., 2020). In addition, these compounds play a role in the regulation of cell membrane fluidity and function (Khozin-Goldberg et al., 2011, Bellou et al., 2014, Bellou et al., 2016, Yoshida et al., 2016, Kothri et al., 2020). This may explain why microorganisms inhabiting the Southern Ocean harbor more PUFA related compounds, as the Southern Ocean is

one of the coldest water body on the planet (Livermore et al., 2005, Chen et al., 2019), with water temperatures ranging between +1.5 and -1.9 ^{0}C at the most northerly and southerly latitudes, respectively (Littlepage, 1965, Núñez-Pons et al., 2018), combined with the large ozone holes that typically form over the Southern Hemisphere in the spring, thereby increasing the amount of UV-B that will reach the marine environments (Karentz and Bosch, 2001, Núñez-Pons et al., 2018).

The discovery of several PUFA-related BGCs in SAR324 may also point to this phylum's contribution to the marine food web (Nichols, 2003). This is because in the past marine organisms had been observed to lack the ability to produce PUFA denovo even though it was an important source of nutrients in the deep ocean (Kanazawa et al., 1979, Intriago and Jones, 1993). PUFA has also been shown to have other chemical properties for application in skin health, cardiovascular and other inflammatory diseases (Ziboh et al., 2000, Kowal-Bielecka et al., 2001, Luigi Capella, 2011, Kendall and Nicolaou, 2013, Pérez-Sánchez et al., 2018). This suggests that the SO's SAR324 natural products may be a repertoire of novel compounds that may found pharmaceutical application in this regard.

3.4 Conclusion

Microbial communities in the Southern Ocean are subject to extreme environmental conditions including oligotrophy and drastic shifts in climate (Hindell et al., 2020). These conditions may result in the selection of microorganisms resulting in communities with adaptations to these extreme conditions. The data from this study suggests that microbial communities in these environments may harbour a suit of genetic mechanisms including those linked to the production of natural productions. We report the abundance and variety of BGCs recovered from Southern Ocean microorganisms using a metagenome-assembled genome approach, with the potential for some novel BGCs encoding for new natural products in the microbial dark matter. We found a

total of 166 BGCs that grouped into 21 BGC types. These BGCs included T1PKS, T3PKS, Arylpolyene, PUFA, RiPP, HgIE-KS, Terpene, betalactone and they were observed in MAGs that belong to SAR324, *Verrucomicrobiota*, *Thermoplasmotota* and *Chloroflexota*. Amongst understudied microbial lineages such as the SAR324, we observed that these genomes coded for a disproportionate number of BGCs suggesting that understudied microbial lineages may encode more BGCs that are yet to be discovered. This insight provides novel area for potential drug discovery.

In general, the study has improved our understanding of microbial adaptations in the Southern Ocean and provides a framework for future research by mapping secondary metabolism pathways of Southern Ocean microorganisms, as well as their potential to encode for new natural products of interest and the zone with the most outstanding biosynthetic potential.

Chapter Four

Thesis synthesis

4.1 Potential impact of the findings

Natural products have held a pivotal role in drug discovery, serving as the foundation for many early medicines. These compounds have proven to be a valuable resource for developing new drugs to combat numerous life-threatening diseases where effective treatments are either unavailable or still in development (Chopra and Dhingra, 2021). Unique and pristine environments, which include the oceans, have been predicted to harbour novel natural products (Liu et al., 2019, Taufa et al., 2021). These novel compounds, including polyketides, have been the major targets based on their complex biological properties such antibacterial, antifungal and antitumor (O'Hagan, 1991, Hertweck *et al.*, 2007, Rego *et al.*, 2020). The development of simple approaches to prospect for novel polyketides in ocean microbiomes and other biosynthetic gene cluster rich environments is important to discover new gene clusters. These clusters may encode distinct chemical structures different from the existing ones, leading to improved ways in abating the antimicrobial resistance threat.

The data generated in this thesis advances current insights regarding polyketide diversity and abundance in SAO and SO. Furthermore, it fills the knowledge deficit related to the diversity of BGCs encoded amongst microbial lineages. For instance, SAR324 was identified in our research as the microbial lineage with the highest biosynthetic gene diversity in the SO. This suggested that there is an enormous biosynthetic potential encoded in this microbial lineage contrary to what has been reported by Baltz (2017). In addition, we have shown that the SAO and the SO harbour unprecedented levels of functional diversity despite the substantial environmental constraints. Firstly, our study revealed that the continuous increase in metagenomic data in databases can be

harnessed to bio-prospect for BGCs from unique and pristine environments through the application of metagenomic tools. The application of these tools could subsequently lead to the identification of potentially novel natural products from target environments. For instance, in this study the diversity of polyketides in the euphotic zone of South Atlantic Ocean was dominated by arylpolyene, T1PKS and HgIE-KS pathway. The novelty of these biosynthetic gene clusters was demonstrated by lack of similarity to known biosynthetic gene clusters. However, this could also mean that there is lack of adequate reference sequences in current databases to make accurate BGC annotations from pristine microbiomes such as the SAO and SO. Based on our result, we further speculate that viruses played a role towards the propagation of polyketides in the euphotic zone. On the other hand in SO, our results showed that bathypelagic zone microbial lineages had low abundance but disproportionately diverse BGCs in contrast to epipelagic and mesopelagic zones. This suggests the influence of depth in the distribution of microorganisms that encode BGCs which still remains to be explored. Altogether, these results suggest that the South Atlantic Ocean and Southern Ocean microbial communities harbour a novel chemical ecology which plays a key role in the ecosystem function and could also be harnessed for pharmaceutical purposes.

4.2 Caveat

The project relied on the acquisition, analysis, and direct interpretation of DNA-based sequence information for BGCs without any functional characterization. Although several studies have also utilized DNA-based sequence analysis to describe the composition and diversity of BGCs in ecological studies (Charlop-Powers et al., 2016, Lemetre et al., 2017, Borsetto et al., 2019, Rego et al., 2020, Sharrar et al., 2020). This is usually hugely reliant on the assembly of high-quality DNA sequences (Bezuidt et al., 2016, Nurk et al., 2022, Makinde et al., 2022). One of the limitations experienced in this study was the sequencing depth of the SAO and SO metagenomes

which was approximately around ten million reads per sample respectively. This sequencing depth technically introduces a bias towards abundant microbial communities (Sims et al., 2014). In addition, metagenomic assemblies were constructed using MEGAHIT and MetaSPAdes for the SAO and SO respectively. These assemblers currently do not scale well on metagenomes when assembling contigs to identify BGCs. Currently, there exist an assembler called BioysntheticSPAdes which improves the prediction of BGCs from metagenomes (Meleshko et al., 2019). Lastly, antiSMASH was used as the genome mining tool to identify BGCs which is also limited in its predictive capability. To circumvent these caveats a few future prospects are described below.

4.3 Future prospect

Future prospects will be aimed at generating metagenomes by combining both short read and long read sequencing technologies to identify BGCs from the SAO and SO. For instance, Van Goethem et al. (2021) showed that, combining the two approaches overcomes the limitation of incomplete BGCs and biases introduced towards communities that naturally occur in high abundances. Briefly, co-assemblies of both long-read and short-read sequences were used to elucidate the secondary metabolism of biocrust metagenomes. The results provided both deeper insight and recovery of higher complete BGCs than individual approaches when used separately (Van Goethem et al., 2021). In addition to using better sequencing platforms, another future prospect will be to utilize non-traditional computational strategies. A number of BGC genome mining tools exist which include antiSMASH, RiPPER, PRIMS 3 and RODEO (Skinnider et al., 2017, Tietz et al., 2017, Santos-Aberturas et al., 2019, Blin et al., 2021). However, these tools rely on rule-based algorithms and heuristic approaches that compare genes or proteins against annotated BGCs

(Gupta Vinod et al., 2022). There is a great need to use artificial intelligence and machine learning strategies to identify BGCs in marine metagenomes. In a recent study, machine learning was used to detect BGCs with antimicrobial properties from gut microbiomes, where at least 83% of the genes showed promising activity via experimental assays (Ma et al., 2022).

Furthermore, while bioinformatics approaches primarily help in the identification of potential novel BGCs, an effort will be made to generate relevant functional datasets such as metaproteomics, metabolomics and the physico-chemical parameters of the microbiomes. Integrating this information with metagenomics will be beneficial in consolidating the functional hypothesis and further demystify the ecological contribution of the involved unique microorganisms. A proper validation of these BGCs through heterologous expression in a microbial host and subsequent experimental assays will also be important. For instance, fosmid metagenomic libraries could be constructed from high quality DNA and pooled for Illumina sequencing as previously described (Negri et al., 2022). Another approach will be to directly clone the fosmid library into a vector and perform activity assays such as bacterial cell viability assays against known antibiotic resistant strains (Wang et al., 2021, Moretta et al., 2020). *E.coli* has been the most successfully used microbial host for natural product biosynthesis (Yang et al., 2020). Terpenoids such as β-Carotene (Yang and Guo, 2014), Crocin-5 (Wang et al., 2019), Armophadiene (Shukal et al., 2019) and polyketides such as Erythromycin A (Zhang et al., 2010), 6-Methylsalicylic acid (Yang et al., 2018), Olivetolic (Tan et al., 2018) have all been produced in *E.coli*. The biosynthesis of the identified natural products will afford further determination of properties such as the mechanism of action and spectrum via phenotypic analyses (Zoffmann et al., 2019). These features are particularly important in the study of drugs and natural products.

Chapter Five

Reference

ABID, F., ZAHID, M. A., ABEDIN, Z. U., NIZAMI, S. B., ABID, M. J., KAZMI, S. Z. H., KHAN, S. U., HASAN, H., ALI, M. & GUL, A. 2018. Omics approaches in marine biotechnology: The treasure of ocean for human betterments. *Omics Technologies and Bio-Engineering*. Elsevier.

ABRAHMS, B., SCALES, K. L., HAZEN, E. L., BOGRAD, S. J., SCHICK, R. S., ROBINSON, P. W. & COSTA, D. P. 2018. Mesoscale activity facilitates energy gain in a top predator. *Proceedings of the Royal Society B*, 285**,** 20181101.

ADADI, A., ADADI, S. & BERRADA, M. 2019. Gastroenterology meets machine learning: Status quo and quo vadis. Advances in bioinformatics, 2019.

ADU-OPPONG, B., GASPARRINI, A. J. & DANTAS, G. 2017. Genomic and functional techniques to mine the microbiome for novel antimicrobials and antimicrobial resistance genes. *Annals of the New York Academy of Sciences*, 1388**,** 42-58.

AKEY, D. L., RAZELUN, J. R., TEHRANISA, J., SHERMAN, D. H., GERWICK, W. H. & SMITH, J. L. 2010. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure*, 18**,** 94-105.

ALBERTSEN, M., HUGENHOLTZ, P., SKARSHEWSKI, A., NIELSEN, K. L., TYSON, G. W. & NIELSEN, P. H. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, 31**,** 533-538.

ALCAÍNO, J., CIFUENTES, V. & BAEZA, M. 2015. Physiological adaptations of yeasts living in cold environments and their potential applications. *World Journal of Microbiology and Biotechnology*, 31**,** 1467-1473.

ALLAIRE, J. 2012. RStudio: integrated development environment for R. Boston, MA, 770, 165-171.

ALMEIDA, O. G. & DE MARTINIS, E. C. 2019. Bioinformatics tools to assess metagenomic data for applied microbiology. *Applied Microbiology and Biotechnology*, 103**,** 69-82.

ALNEBERG, J., BJARNASON, B. S., DE BRUIJN, I., SCHIRMER, M., QUICK, J., IJAZ, U. Z., LAHTI, L., LOMAN, N. J., ANDERSSON, A. F. & QUINCE, C. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11**,** 1144-1146.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215**,** 403-410.

AMOS, G. C., BORSETTO, C., LASKARIS, P., KRSEK, M., BERRY, A. E., NEWSHAM, K. K., CALVO-BADO, L., PEARCE, D. A., VALLIN, C. & WELLINGTON, E. M. 2015. Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. *PloS one*, 10**,** e0138327.

ANANTHARAMAN, K., BROWN, C. T., HUG, L. A., SHARON, I., CASTELLE, C. J., PROBST, A. J., THOMAS, B. C., SINGH, A., WILKINS, M. J. & KARAOZ, U. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature communications, 7, 1-11.

ANANTHARAMAN, K., HAUSMANN, B., JUNGBLUTH, S. P., KANTOR, R. S., LAVY, A., WARREN, L. A., RAPPÉ, M. S., PESTER, M., LOY, A. & THOMAS, B. C. 2018.

Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *The ISME Journal,* 12**,** 1715-1728.

AOUEDI, O., PIAMRAT, K., HAMMA, S. & PERERA, J. 2021. Network Traffic Analysis using Machine Learning: an unsupervised approach to understand and slice your network. *Annals of Telecommunications***,** 1-13.

ARIKAWA, K., IDE, K., KOGAWA, M., SAEKI, T., YODA, T., ENDOH, T., MATSUHASHI, A., TAKEYAMA, H. & HOSOKAWA, M. 2021. Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics. *Microbiome,* 9**,** 1-16.

ARTEAGA, R. D. C., ICHIWAKI, S., MASSINI, K., GARRIDO, L. M., ROSERO, E. M. B. & PADILLA, G. 2017. Screening of polyketide genes from Brazilian Atlantic Forest soil. *Universitas Scientiarum,* 22**,** 87-96.

BAERGA-ORTIZ, A., POPOVIC, B., SISKOS, A. P., O'HARE, H. M., SPITELLER, D., WILLIAMS, M. G., CAMPILLO, N., SPENCER, J. B. & LEADLAY, P. F. 2006. Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. *Chemistry & Biology,* 13**,** 277-285.

BALTZ, R. H. 2017. Gifted microbes for genome mining and natural product discovery. Journal of industrial microbiology & biotechnology, 44, 573-588.

BANDEKAR, M., RAMAIAH, N., JAIN, A. & MURTI MEENA, R. 2016. Distinctly different bacterial communities in surface and oxygen minimum layers in the Arabian Sea. *Biogeosciences Discussions***,** 1-36.

BARAJAS, J. F., BLAKE-HEDGES, J. M., BAILEY, C. B., CURRAN, S. & KEASLING, J. D. 2017. Engineered polyketides: synergy between protein and host level engineering. *Synthetic and Systems Biotechnology,* 2**,** 147-166.

BARAJAS, J. F., MCANDREW, R. P., THOMPSON, M. G., BACKMAN, T. W., PANG, B., DE ROND, T., PEREIRA, J. H., BENITES, V. T., MARTÍN, H. G. & BAIDOO, E. E. 2019. Structural insights into dehydratase substrate selection for the borrelidin and fluvirucin polyketide synthases. *Journal of Industrial Microbiology & Biotechnology,* 46**,** 1225-1235.

BEAM, A. L. & KOHANE, I. S. 2018. Big data and machine learning in health care. *Jama,* 319**,** 1317-1318.

BECH, P. K., LYSDAL, K. L., GRAM, L., BENTZON-TILIA, M. & STRUBE, M. L. 2020. Marine Sediments Hold an Untapped Potential for Novel Taxonomic and Bioactive Bacterial Diversity. *Msystems,* 5**,** e00782-20.

BEEDESSEE, G., HISATA, K., ROY, M. C., VAN DOLAH, F. M., SATOH, N. & SHOGUCHI, E. 2019. Diversified secondary metabolite biosynthesis gene repertoire revealed in symbiotic dinoflagellates. *Scientific Reports,* 9**,** 1204.

BELLOU, S., BAESHEN, M. N., ELAZZAZY, A. M., AGGELI, D., SAYEGH, F. & AGGELIS, G. 2014. Microalgal lipids biochemistry and biotechnological perspectives. *Biotechnology Advances,* 32**,** 1476-1493.

BELLOU, S., TRIANTAPHYLLIDOU, I.-E., AGGELI, D., ELAZZAZY, A. M., BAESHEN, M. N. & AGGELIS, G. 2016. Microbial oils as food additives: recent approaches for improving microbial oil production and its polyunsaturated fatty acid content. *Current Opinion in Biotechnology,* 37**,** 24-35.

BEZUIDT, O. K., GOMRI, M. A., PIERNEEF, R., VAN GOETHEM, M. W., KHARROUB, K., COWAN, D. A. & MAKHALANYANE, T. P. 2016. Draft genome sequence of

Thermoactinomyces sp. strain AS95 isolated from a Sebkha in Thamelaht, Algeria. *Standards in Genomic Sciences,* 11**,** 1-6.

BIEBER, B., NÜSKE, J., RITZAU, M. & GRÄFE, U. 1998. Alnumycin a new naphthoquinone antibiotic produced by an endophytic Streptomyces sp. *The Journal of Antibiotics,* 51**,** 381-382.

BLASIAK, R., WYNBERG, R., GRORUD-COLVERT, K., THAMBISETTY, S., BANDARRA, N. M., CANARIO, A. V., DA SILVA, J., DUARTE, C. M., JASPARS, M. & ROGERS, A. 2020. The ocean genome and future prospects for conservation and equity. Nature Sustainability, 3, 588-596.

BLIN, K., SHAW, S., KLOOSTERMAN, A. M., CHARLOP-POWERS, Z., VAN WEZEL, G. P., MEDEMA, M. H. & WEBER, T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research,* 49**,** W29-W35.

BOEUF, D., EPPLEY, J. M., MENDE, D. R., MALMSTROM, R. R., WOYKE, T. & DELONG, E. F. 2021. Meta-pangenomics Reveals Depth-dependent Shifts in Metabolic Potential for the Ubiquitous Marine Bacterial SAR324 Lineage.

BOGDANOVA, A., REFELD, A., PRAZDNOVA, E., CHIKINDAS, M., OLSHEVSKAYA, A., MALTSEVA, T. & CHEGGE, V. Polyketides Are Agents for Probiotics Antagonizing Pathogenic Microbiota. XIV International Scientific Conference "INTERAGROMASH 2021", 2022. Springer, 60-68.

BOLAÑOS, L. M., CHOI, C. J., WORDEN, A. Z., BAETGE, N., CARLSON, C. A. & GIOVANNONI, S. 2021. Seasonality of the microbial community composition in the north atlantic. Frontiers in Marine Science, 8, 624164.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120.

BONNETT, S. A., WHICHER, J. R., PAPIREDDY, K., FLOROVA, G., SMITH, J. L. & REYNOLDS, K. A. 2013. Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chemistry & Biology,* 20**,** 772-783.

BOOLCHANDANI, M., D'SOUZA, A. W. & DANTAS, G. 2019. Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics***,** 1.

BORCHERT, E. 2017. The biotechnological potential of deep sea sponges and their associated microbiome.

BORSETTO, C., AMOS, G. C., DA ROCHA, U. N., MITCHELL, A. L., FINN, R. D., LAIDI, R. F., VALLIN, C., PEARCE, D. A., NEWSHAM, K. K. & WELLINGTON, E. M. 2019. Microbial community drivers of PK/NRP gene diversity in selected global soils. *Microbiome,* 7**,** 1-11.

BOSE, U., ORTORI, C. A., SARMAD, S., BARRETT, D. A., HEWAVITHARANA, A. K., HODSON, M. P., FUERST, J. A. & SHAW, P. N. 2017. Production of N-acyl homoserine lactones by the sponge-associated marine actinobacteria Salinispora arenicola and Salinispora pacifica. *FEMS Microbiology Letters,* 364.

BOUFRIDI, A. & QUINN, R. J. 2016. Turning metabolomics into drug discovery. *Journal of the Brazilian Chemical Society,* 27**,** 1334-1338.

BOUTET, E., LIEBERHERR, D., TOGNOLLI, M., SCHNEIDER, M., BANSAL, P., BRIDGE, A. J., POUX, S., BOUGUELERET, L. & XENARIOS, I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Plant Bioinformatics.* Springer.

BOWERS, R. M., KYRPIDES, N. C., STEPANAUSKAS, R., HARMON-SMITH, M., DOUD, D., REDDY, T. B. K., SCHULZ, F., JARETT, J., RIVERS, A. R., ELOE-FADROSH, E. A., TRINGE, S. G., IVANOVA, N. N., COPELAND, A., CLUM, A., BECRAFT, E. D., MALMSTROM, R. R., BIRREN, B., PODAR, M., BORK, P., WEINSTOCK, G. M., GARRITY, G. M., DODSWORTH, J. A., YOOSEPH, S., SUTTON, G., GLÖCKNER, F. O., GILBERT, J. A., NELSON, W. C., HALLAM, S. J., JUNGBLUTH, S. P., ETTEMA, T. J. G., TIGHE, S., KONSTANTINIDIS, K. T., LIU, W. T., BAKER, B. J., RATTEI, T., EISEN, J. A., HEDLUND, B., MCMAHON, K. D., FIERER, N., KNIGHT, R., FINN, R., COCHRANE, G., KARSCH-MIZRACHI, I., TYSON, G. W., RINKE, C., LAPIDUS, A., MEYER, F., YILMAZ, P., PARKS, D. H., EREN, A. M., SCHRIML, L., BANFIELD, J. F., HUGENHOLTZ, P. & WOYKE, T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology,* 35**,** 725-731.

BOYD, E. F. 2012. Bacteriophage-encoded bacterial virulence factors and phage–pathogenicity island interactions. *Advances in Virus Research,* 82**,** 91-118.

BOYD, P. W. 2019. Physiology and iron modulate diverse responses of diatoms to a warming Southern Ocean. *Nature Climate Change,* 9**,** 148.

BRAVO-RODRIGUEZ, K., KLOPRIES, S., KOOPMANS, K. R., SUNDERMANN, U., YAHIAOUI, S., ARENS, J., KUSHNIR, S., SCHULZ, F. & SANCHEZ-GARCIA, E. 2015. Substrate flexibility of a mutated acyltransferase domain and implications for polyketide biosynthesis. *Chemistry & Biology,* 22**,** 1425-1430.

BRETAGNON, M., PAULMIER, A., GARÇON, V., DEWITTE, B., ILLIG, S., LEBLOND, N., COPPOLA, L., CAMPOS, F., VELAZCO, F. & PANAGIOTOPOULOS, C. 2018. Modulation of the vertical particle transfer efficiency in the oxygen minimum zone off Peru. *Biogeosciences,* 15**,** 5093-5111.

BROWN, M. V. & DONACHIE, S. P. 2007. Evidence for tropical endemicity in the Deltaproteobacteria Marine Group B/SAR324 bacterioplankton clade. *Aquatic Microbial Ecology,* 46**,** 107-115.

BUCHFINK, B., XIE, C. & HUSON, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods,* 12**,** 59-60.

BUESCHER, J. M., LIEBERMEISTER, W., JULES, M., UHR, M., MUNTEL, J., BOTELLA, E., HESSLING, B., KLEIJN, R. J., LE CHAT, L. & LECOINTE, F. 2012. Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science,* 335**,** 1099-1103.

BUIJS, Y., BECH, P. K., VAZQUEZ-ALBACETE, D., BENTZON-TILIA, M., SONNENSCHEIN, E. C., GRAM, L. & ZHANG, S.-D. 2019. Marine Proteobacteria as a source of natural products: advances in molecular tools and strategies. *Natural Product Reports,* 36**,** 1333-1350.

BUSHNELL, B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

BUTLER, M. S., BLASKOVICH, M. A. & COOPER, M. A. 2017. Antibiotics in the clinical pipeline at the end of 2015. *The Journal of Antibiotics,* 70**,** 3-24.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

CAMPBELL, W., FISHER, M., STAPLEY, E., ALBERS-SCHONBERG, G. & JACOB, T. 1983. Ivermectin: a potent new antiparasitic agent. *Science,* 221**,** 823-828.

CAO, H., DONG, C., BOUGOUFFA, S., LI, J., ZHANG, W., SHAO, Z., BAJIC, V. B. & QIAN, P.-Y. 2016. Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge. *Scientific Reports,* 6**,** 1-9.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J. M. & GABALDÓN, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics,* 25**,** 1972-1973.

CARROLL, L. M., LARRALDE, M., FLECK, J. S., PONNUDURAI, R., MILANESE, A., BARAZZONE, E. C. & ZELLER, G. 2021. Accurate de novo identification of biosynthetic gene clusters with GECCO. *bioRxiv*.

CASSINI, A., HÖGBERG, L. D., PLACHOURAS, D., QUATTROCCHI, A., HOXHA, A., SIMONSEN, G. S., COLOMB-COTINAT, M., KRETZSCHMAR, M. E., DEVLEESSCHAUWER, B. & CECCHINI, M. 2019. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases,* 19**,** 56-66.

CASTILLO, D., KAUFFMAN, K., HUSSAIN, F., KALATZIS, P., RØRBO, N., POLZ, M. F. & MIDDELBOE, M. 2018. Widespread distribution of prophage-encoded virulence factors in marine Vibrio communities. *Scientific Reports,* 8**,** 1-9.

CHANG, Z., SITACHITTA, N., ROSSI, J. V., ROBERTS, M. A., FLATT, P. M., JIA, J., SHERMAN, D. H. & GERWICK, W. H. 2004. Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium Lyngbya m ajuscula. *Journal of Natural Products,* 67**,** 1356-1367.

CHARLOP-POWERS, Z., PREGITZER, C. C., LEMETRE, C., TERNEI, M. A., MANIKO, J., HOVER, B. M., CALLE, P. Y., MCGUIRE, K. L., GARBARINO, J. & FORGIONE, H. M. 2016. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proceedings of The National Academy of Sciences,* 113**,** 14811-14816.

CHATOW, L., NUDEL, A., NESHER, I., HAYO HEMO, D., ROZENBERG, P., VOROPAEV, H., WINKLER, I., LEVY, R., KEREM, Z. & YANIV, Z. 2021. In vitro evaluation of the activity of terpenes and cannabidiol against human coronavirus E229. *Life,* 11**,** 290.

CHAUMEIL, P.-A., MUSSIG, A. J., HUGENHOLTZ, P. & PARKS, D. H. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Oxford University Press.

CHEN, L., LU, Y., LI, W., REN, Y., YU, M., JIANG, S., FU, Y., WANG, J., PENG, S. & BILYK, K. T. 2019. The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. *GigaScience,* 8**,** giz016.

CHEN, Q., WAN, Y., LEI, Y., ZOBEL, J. & VERSPOOR, K. Evaluation of CD-HIT for constructing non-redundant databases. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016. IEEE, 703-706.

CHEN, R., WONG, H. L., KINDLER, G. S., MACLEOD, F. I., BENAUD, N., FERRARI, B. C. & BURNS, B. P. 2020. Discovery of an abundance of biosynthetic gene clusters in shark bay microbial mats. *Frontiers in Microbiology,* 11**,** 1950.

CHEN, T. C., DA FONSECA, C. O. & SCHÖNTHAL, A. H. 2015. Preclinical development and clinical use of perillyl alcohol for chemoprevention and cancer therapy. *American Journal of Cancer Research,* 5**,** 1580.

CHEN, Y., KELLY, E. E., MASLUK, R. P., NELSON, C. L., CANTU, D. C. & REILLY, P. J. 2011. Structural classification and properties of ketoacyl synthases. *Protein Science,* 20**,** 1659-1667.

CHIU, C. & MILLER, S. 2016. Next-generation sequencing. *Molecular Microbiology: Diagnostic Principles and Practice***,** 68-79.

CHOPRA, B. & DHINGRA, A. K. 2021. Natural products: A lead for drug discovery and development. *Phytotherapy Research,* 35**,** 4660-4702.

CHOI, S.-S., KATSUYAMA, Y., BAI, L., DENG, Z., OHNISHI, Y. & KIM, E.-S. 2018. Genome engineering for microbial natural product discovery. *Current Opinion in Microbiology,* 45**,** 53-60.

CHOOI, Y.-H. & TANG, Y. 2012. Navigating the fungal polyketide chemical space: from genes to molecules. *The Journal of Organic Chemistry,* 77**,** 9933-9953.

CIMERMANCIC, P., MEDEMA, M. H., CLAESEN, J., KURITA, K., BROWN, L. C. W., MAVROMMATIS, K., PATI, A., GODFREY, P. A., KOEHRSEN, M. & CLARDY, J. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell,* 158**,** 412-421.

CONWAY, T. M., PALTER, J. B. & DE SOUZA, G. F. 2018. Gulf Stream rings as a source of iron to the North Atlantic subtropical gyre. Nature Geoscience, 11, 594-598.

COUTINHO, F., VON MEIJENFELDT, F., WALTER, J., HARO-MORENO, J., LOPÉZ-PÉREZ, M., VAN VERK, M., THOMPSON, C., COSENZA, C., APPOLINARIO, L. & PARANHOS, R. 2021. Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome. *Science of The Total Environment,* 765**,** 142758.

COVINGTON, B. C., MCLEAN, J. A. & BACHMANN, B. O. 2017. Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Natural Product Reports,* 34**,** 6-24.

CRAGG, G. M. & NEWMAN, D. J. 2013. Natural products: a continuing source of novel drug leads. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1830**,** 3670-3695.

CUPERLOVIC-CULF, M., CULF, A. S., MORIN JR, P. & TOUAIBIA, M. 2013. Application of metabolomics in drug discovery, development and theranostics. *Current Metabolomics,* 1**,** 41-57.

DAVIES, J. 2013. Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics,* 66**,** 361-364.

DAVIES, J. & DAVIES, D. 2010. Origins and evolution of antibiotic resistance. *Microbiology & Molecular Biology Review,* 74**,** 417-433.

DAVIES-BOLORUNDURO, O., OSUOLALE, O., SAIBU, S., ADELEYE, I. & AMINAH, N. 2021. Bioprospecting marine actinomycetes for antileishmanial drugs: current perspectives and future prospects. Heliyon, 7, e07710.

DEBELJAK, P., TOULZA, E., BEIER, S., BLAIN, S. & OBERNOSTERER, I. 2019. Microbial iron metabolism as revealed by gene expression profiles in contrasted Southern Ocean regimes. *Environmental microbiology,* 21**,** 2360-2374.

DELLA SALA, G., HOCHMUTH, T., TETA, R., COSTANTINO, V. & MANGONI, A. 2014. Polyketide synthases in the microbiome of the marine sponge Plakortis halichondrioides: a metagenomic update. *Marine Drugs,* 12**,** 5425-5440.

DELMONT, T. O., QUINCE, C., SHAIBER, A., ESEN, Ö. C., LEE, S. T., RAPPÉ, M. S., MCLELLAN, S. L., LÜCKER, S. & EREN, A. M. 2018. Nitrogen-fixing populations of

Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology,* 3**,** 804-813.

DELONG, E. F. 2007. Modern microbial seascapes. Forward. *Nature Reviews Microbiology,* 5**,** 755-7.

DELONG, E. F., PRESTON, C. M., MINCER, T., RICH, V., HALLAM, S. J., FRIGAARD, N.-U., MARTINEZ, A., SULLIVAN, M. B., EDWARDS, R. & BRITO, B. R. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science,* 311**,** 496-503.

DEMAIN, A. L. 2014. Importance of microbial natural products and the need to revitalize their discovery. *Journal of Industrial Microbiology and Biotechnology,* 41**,** 185-201.

DICK, G. J., ANANTHARAMAN, K., BAKER, B. J., LI, M., REED, D. C. & SHEIK, C. S. 2013. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology,* 4**,** 124.

DOROGHAZI, J. R., ALBRIGHT, J. C., GOERING, A. W., JU, K.-S., HAINES, R. R., TCHALUKOV, K. A., LABEDA, D. P., KELLEHER, N. L. & METCALF, W. W. 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology,* 10**,** 963-968.

DROR, B., JURKEVITCH, E. & CYTRYN, E. 2023. Global-scale analysis reveals distinct patterns of non-ribosomal peptide and polyketide synthase encoding genes in root and soil bacterial communities. *Soil Ecology Letters,* 5**,** 38-45.

DU, X., LIU, H., YUAN, L., WANG, Y., MA, Y., WANG, R., CHEN, X., LOSIEWICZ, M. D., GUO, H. & ZHANG, H. 2019. The diversity of cyanobacterial toxins on structural characterization, distribution and identification: A systematic review. *Toxins,* 11**,** 530.

DUNCAN, A., BARRY, K., DAUM, C., ELOE-FADROSH, E., ROUX, S., SCHMIDT, K., TRINGE, S. G., VALENTIN, K. U., VARGHESE, N. & SALAMOV, A. 2022. Metagenome-assembled genomes of phytoplankton microbiomes from the Arctic and Atlantic Oceans. *Microbiome,* 10**,** 67.

EDDY, S. R. 2011. Accelerated profile HMM searches. *PLoS Computational Biology,* 7**,** e1002195.

EDGAR, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.(2004) Nucleic Acids Research 32: 1792–1797. pmid: 15034147 View Article PubMed. NCBI.

EDWARDS, D. J., MARQUEZ, B. L., NOGLE, L. M., MCPHAIL, K., GOEGER, D. E., ROBERTS, M. A. & GERWICK, W. H. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chemistry & Biology,* 11**,** 817-833.

EHLING-SCHULZ, M., BILGER, W. & SCHERER, S. 1997. UV-B-induced synthesis of photoprotective pigments and extracellular polysaccharides in the terrestrial cyanobacterium Nostoc commune. *Journal of Bacteriology,* 179**,** 1940-1945.

EL MRABET, M. A., EL MAKKAOUI, K. & FAIZE, A. Supervised Machine Learning: A Survey. 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet), 2021. IEEE, 1-10.

ENGENE, N., CHOI, H., ESQUENAZI, E., BYRUM, T., VILLA, F. A., CAO, Z., MURRAY, T. F., DORRESTEIN, P. C., GERWICK, L. & GERWICK, W. H. 2011. Phylogeny-guided isolation of ethyl tumonoate A from the marine cyanobacterium cf. Oscillatoria margaritifera. *Journal of Natural Products,* 74**,** 1737-1743.

ERKTAN, A., OR, D. & SCHEU, S. 2020. The physical structure of soil: determinant and consequence of trophic interactions. *Soil Biology and Biochemistry,* 148**,** 107876.

FANG, J., PANG, X., YAN, R., LIAN, W., LI, C., WANG, Q., LIU, A.-L. & DU, G.-H. 2016. Discovery of neuroprotective compounds by machine learning approaches. *RSC Advances,* 6**,** 9857-9871.

FARRAJ, S. A., EL-KAFRAWY, S. A., KUMOSANI, T. A., YOUSEF, J. M. & AZHAR, E. I. 2020. Evaluation of Extraction Methods for Clinical Metagenomic Assay. Microorganisms, 8, 1128.

FATEMA, U., BROBERG, A., JENSEN, D. F., KARLSSON, M. & DUBEY, M. 2018. Functional analysis of polyketide synthase genes in the biocontrol fungus Clonostachys rosea. *Scientific Reports,* 8**,** 15009.

FAUVILLE, G., STRANG, C., CANNADY, M. A. & CHEN, Y.-F. 2019. Development of the International Ocean Literacy Survey: measuring knowledge across the world. *Environmental Education Research,* 25**,** 238-263.

FERREIRA, J. C. N., BERGO, N. M., TURA, P. M., CHUQUI, M. G., BRANDINI, F. P., JOVANE, L. & PELLIZARI, V. H. 2022. Abundance and microbial diversity from surface to deep water layers over the Rio Grande Rise, South Atlantic. *Progress in Oceanography,* 201**,** 102736.

FIERS, W. D., DODGE, G. J., SHERMAN, D. H., SMITH, J. L. & ALDRICH, C. C. 2016. Vinylogous dehydration by a polyketide dehydratase domain in curacin biosynthesis. *Journal of The American Chemical Society,* 138**,** 16024-16036.

FINN, R. D., CLEMENTS, J. & EDDY, S. R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research,* 39**,** W29-W37.

FLÓREZ, L. V., SCHERLACH, K., MILLER, I. J., RODRIGUES, A., KWAN, J. C., HERTWECK, C. & KALTENPOTH, M. 2018. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in Lagria villosa beetles. *Nature Communications,* 9**,** 1-10.

FOERSTNER, K. U., DOERKS, T., CREEVEY, C. J., DOERKS, A. & BORK, P. 2008. A computational screen for type I polyketide synthases in metagenomics shotgun data. *PLoS One,* 3**,** e3515.

FOULSTON, L. 2019. Genome mining and prospects for antibiotic discovery. *Current Opinion in Microbiology,* 51**,** 1-8.

FOURQUEZ, M., BRESSAC, M., DEPPELER, S. L., ELLWOOD, M., OBERNOSTERER, I., TRULL, T. W. & BOYD, P. W. 2020. Microbial competition in the subpolar southern ocean: an Fe–C Co-limitation experiment. *Frontiers in Marine Science,* 6**,** 776.

FREITAS, S., HATOSY, S., FUHRMAN, J. A., HUSE, S. M., WELCH, D. B. M., SOGIN, M. L. & MARTINY, A. C. 2012. Global distribution and diversity of marine Verrucomicrobia. *The ISME Journal,* 6**,** 1499-1505.

FRÖLICHER, T. L., SARMIENTO, J. L., PAYNTER, D. J., DUNNE, J. P., KRASTING, J. P. & WINTON, M. 2015. Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate,* 28**,** 862-886.

FU, J., CHEN, L., YANG, S., LI, Y., JIN, L., HE, X., HE, L., AO, X., LIU, S. & LIU, A. 2021. Metagenome and analysis of metabolic potential of the microbial community in pit mud used for Chinese strong-flavor liquor production. *Food Research International,* 143**,** 110294.

FUHRMAN, J. A., CRAM, J. A. & NEEDHAM, D. M. 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology, 13, 133-146.

GALEA, C. A., ROBERTS, K. D., ZHU, Y., THOMPSON, P. E., LI, J. & VELKOV, T. 2017. Functional characterization of the unique terminal thioesterase domain from polymyxin synthetase. *Biochemistry,* 56**,** 657-668.

GALLO, G., RENZONE, G., PALAZZOTTO, E., MONCIARDINI, P., ARENA, S., FADDETTA, T., GIARDINA, A., ALDUINA, R., WEBER, T. & SANGIORGI, F. 2016. Elucidating the molecular physiology of lantibiotic NAI-107 production in Microbispora ATCC-PTA-5024. *BMC Genomics,* 17**,** 42.

GAO, Y., WANG, L., KALSCHEUER, R., LIU, Z. & PROKSCH, P. 2020. Antifungal polyketide derivatives from the endophytic fungus Aplosporella javeedii. *Bioorganic & Medicinal Chemistry,* 28**,** 115456.

GAVRIILIDOU, A., KAUTSAR, S. A., ZABURANNYI, N., KRUG, D., MÜLLER, R., MEDEMA, M. H. & ZIEMERT, N. 2021. A global survey of specialized metabolic diversity encoded in bacterial genomes. *bioRxiv*.

GAY, D., YOU, Y.-O., KEATINGE-CLAY, A. & CANE, D. E. 2013. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry,* 52**,** 8916-8928.

GAZITÚA, M. C., VIK, D. R., ROUX, S., GREGORY, A. C., BOLDUC, B., WIDNER, B., MULHOLLAND, M. R., HALLAM, S. J., ULLOA, O. & SULLIVAN, M. B. 2021. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *The ISME Journal,* 15**,** 981-998.

GHIGLIONE, J.-F., GALAND, P. E., POMMIER, T., PEDRÓS-ALIÓ, C., MAAS, E. W., BAKKER, K., BERTILSON, S., KIRCHMAN, D. L., LOVEJOY, C. & YAGER, P. L. 2012. Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences,* 109**,** 17633-17638.

GILBERT, J. A., FIELD, D., SWIFT, P., THOMAS, S., CUMMINGS, D., TEMPERTON, B., WEYNBERG, K., HUSE, S., HUGHES, M. & JOINT, I. 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic'study of seasonal and diel temporal variation. *PloS one,* 5**,** e15545.

GIORDANO, D., COPPOLA, D., RUSSO, R., DENARO, R., GIULIANO, L., LAURO, F. M., DI PRISCO, G. & VERDE, C. 2015. Marine microbial secondary metabolites: pathways, evolution and physiological roles. *Advances in Microbial Physiology,* 66**,** 357-428.

GOLKAR, Z., BAGASRA, O. & PACE, D. G. 2014. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *The Journal of Infection in Developing Countries,* 8**,** 129-136.

GOMES, E. S., SCHUCH, V. & LEMOS, E. G. D. M. 2013. Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Brazilian Journal of Microbiology,* 44**,** 1007-1034.

GÓMEZ-RUBIO, V. 2017. ggplot2-elegant graphics for data analysis. Journal of Statistical Software, 77, 1-3.

GREGORY, A. C., ZAYED, A. A., CONCEIÇÃO-NETO, N., TEMPERTON, B., BOLDUC, B., ALBERTI, A., ARDYNA, M., ARKHIPOVA, K., CARMICHAEL, M. & CRUAUD, C. 2019. Marine DNA viral macro-and microdiversity from pole to pole. *Cell,* 177**,** 1109-1123. e14.

GUBBENS, J., ZHU, H., GIRARD, G., SONG, L., FLOREA, B. I., ASTON, P., ICHINOSE, K., FILIPPOV, D. V., CHOI, Y. H. & OVERKLEEFT, H. S. 2014. Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products. *Chemistry & Biology,* 21**,** 707-718.

GUO, D.-L., CHEN, J.-F., TAN, L., JIN, M.-Y., JU, F., CAO, Z.-X., DENG, F., WANG, L.-N., GU, Y.-C. & DENG, Y. 2019. Terpene glycosides from Sanguisorba officinalis and their anti-inflammatory effects. *Molecules,* 24**,** 2906.

GUO, J., BOLDUC, B., ZAYED, A. A., VARSANI, A., DOMINGUEZ-HUERTA, G., DELMONT, T. O., PRATAMA, A. A., GAZITÚA, M. C., VIK, D. & SULLIVAN, M. B. 2021a. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome,* 9**,** 1-13.

GUO, J., LI, Y., JESSELL, M. W., GIRAUD, J., LI, C., WU, L., LI, F. & LIU, S. 2021b. 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods. *Computers & Geosciences,* 149**,** 104701.

HAMEDI, J. & MOHAMMADIPANAH, F. 2015. Biotechnological application and taxonomical distribution of plant growth promoting actinobacteria. *Journal of industrial microbiology and biotechnology,* 42**,** 157-171.

HAMMEL, E., WEBSTER, T. F., GURNEY, R. & HEIGER-BERNAYS, W. 2022. Implications of PFAS definitions using fluorinated pharmaceuticals. *iScience***,** 104020.

HAN, Q., MOORE, J. K. & ZENDER, C. Constraining Oceanic dust deposition using surface ocean dissolved Al. AGU Fall Meeting Abstracts, 2006. OS32B-06.

HANNIGAN, G. D., PRIHODA, D., PALICKA, A., SOUKUP, J., KLEMPIR, O., RAMPULA, L., DURCAK, J., WURST, M., KOTOWSKI, J. & CHANG, D. 2019. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Research,* 47**,** e110-e110.

HASSAN, F. M., SALIH, W. Y. & AL-HAIDERI, H. H. 2020. Next-Generation Sequencing Technologies for Environmental DNA as an Efficient Bio Indicator for Bacterial Biodiversity in Tigris River, Iraq. Systematic Reviews in Pharmacy, 11, 1107-1114.

HATA, T., OMURA, S., IWAI, Y., NAKAGAWA, A., OTANI, M., ITO, S. & MATSUYA, T. 1971. A new antibiotic, kinamycin: Fermentation, isolation, purification and properties. *The Journal of Antibiotics,* 24**,** 353-359.

HATFULL, G. F. & HENDRIX, R. W. 2011. Bacteriophages and their genomes. *Current Opinion in Virology,* 1**,** 298-303.

HAY, M. E. 1996. Marine chemical ecology: what's known and what's next? *Journal of Experimental Marine Biology and Ecology,* 200**,** 103-134.

HE, M. J., PU, W., WANG, X., ZHANG, W., TANG, D. & DAI, Y. 2022. Comparing DESI-MSI and MALDI-MSI mediated spatial metabolomics and their applications in cancer studies. *Frontiers in Oncology,* 12**,** 891018.

HERTWECK, C. 2009. The biosynthetic logic of polyketide diversity. *Angewandte Chemie International Edition,* 48**,** 4688-4716.

HERTWECK, C., LUZHETSKYY, A., REBETS, Y. & BECHTHOLD, A. 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports,* 24**,** 162-190.

HIBBING, M. E., FUQUA, C., PARSEK, M. R. & PETERSON, S. B. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology,* 8**,** 15-25.

HILL, P., HEBERLIG, G. W. & BODDY, C. N. 2017. Sampling terrestrial environments for bacterial polyketides. *Molecules,* 22**,** 707.

HINDELL, M. A., REISINGER, R. R., ROPERT-COUDERT, Y., HÜCKSTÄDT, L. A., TRATHAN, P. N., BORNEMANN, H., CHARRASSIN, J.-B., CHOWN, S. L., COSTA, D. P. & DANIS, B. 2020. Tracking of marine predators to protect Southern Ocean ecosystems. *Nature,* 580**,** 87-92.

HIRAI, M., NISHI, S., TSUDA, M., SUNAMURA, M., TAKAKI, Y. & NUNOURA, T. 2017. Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. *Microbes and Environments***,** ME17132.

HIRSCH, P. R., MAUCHLINE, T. H. & CLARK, I. M. 2010. Culture-independent molecular techniques for soil microbial ecology. *Soil Biology and Biochemistry,* 42**,** 878-887.

HOFEDITZ, T., UNSIN, C., WIESE, J., IMHOFF, J., WOHLLEBEN, W., GROND, S. & WEBER, T. 2018. Lysoquinone-TH1, a new polyphenolic tridecaketide produced by expressing the lysolipin minimal PKS II in Streptomyces albus. *Antibiotics,* 7**,** 53.

HOWARTH, R. W. & MARINO, R. 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: evolving views over three decades. Limnology and Oceanography, 51, 364-376.

HU, Y., AN, Q., SHEU, K., TREJO, B., FAN, S. & GUO, Y. 2018. Single cell multi-omics technology: methodology and application. *Frontiers in Cell and Developmental Biology,* 6**,** 28.

HU, T., CHITNIS, N., MONOS, D. & DINH, A. 2021. Next-generation sequencing technologies: An overview. Human Immunology, 82, 801-811.

HUANG, I.-S. & ZIMBA, P. V. 2019. Cyanobacterial bioactive metabolites—A review of their chemistry and biology. *Harmful Algae,* 83**,** 42-94.

HUANG, J.-M., BAKER, B. J., LI, J.-T. & WANG, Y. 2019. New microbial lineages capable of carbon fixation and nutrient cycling in deep-sea sediments of the northern South China Sea. *Applied and environmental microbiology,* 85**,** e00523-19.

HUG, J. J., BADER, C. D., REMŠKAR, M., CIRNSKI, K. & MÜLLER, R. 2018. Concepts and methods to access novel antibiotics from actinomycetes. Antibiotics, 7, 44.

HUG, J. J., DASTBAZ, J., ADAM, S., REVERMANN, O., KOEHNKE, J., KRUG, D. & MÜLLER, R. 2020. Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from Myxococcus xanthus. *ACS chemical biology,* 15**,** 2221-2231.

HUNT, B. P., ESPINASSE, B., PAKHOMOV, E. A., CHEREL, Y., COTTÉ, C., DELEGRANGE, A. & HENSCHKE, N. 2021. Pelagic food web structure in high nutrient low chlorophyll (HNLC) and naturally iron fertilized waters in the Kerguelen Islands region, Southern Ocean. *Journal of Marine Systems,* 224**,** 103625.

HYATT, D., CHEN, G.-L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics,* 11**,** 119.

IGARASHI, Y., OGURA, H., FURIHATA, K., OKU, N., INDANANDA, C. & THAMCHAIPENET, A. 2011. Maklamicin, an antibacterial polyketide from an endophytic Micromonospora sp. *Journal of Natural Products,* 74**,** 670-674.

IKEDA, Y., IDEMOTO, H., HIRAYAMA, F., YAMAMOTO, K., IWAO, K., ASAO, T. & MUNAKATA, T. 1983. Safracins, New Antitumor Antibiotics I. Producing Organism, Fermentation and Isolation. *The Journal of Antibiotics,* 36**,** 1279-1283.

INTRIAGO, P. & JONES, D. 1993. Bacteria as food for Artemia. *Aquaculture,* 113**,** 115-127.

JAMET, A., TOUCHON, M., RIBEIRO-GONÇALVES, B., CARRIÇO, J. A., CHARBIT, A., NASSIF, X., RAMIREZ, M. & ROCHA, E. P. 2017. A widespread family of polymorphic toxins encoded by temperate phages. *BMC Biology,* 15**,** 1-12.

JANARDHAN, A., KUMAR, A. P., VISWANATH, B., SAIGOPAL, D. & NARASIMHA, G. 2014. Production of bioactive compounds by actinomycetes and their antioxidant properties. *Biotechnology Research International,* 2014.

JANG, H. B., BOLDUC, B., ZABLOCKI, O., KUHN, J. H., ROUX, S., ADRIAENSSENS, E. M., BRISTER, J. R., KROPINSKI, A. M., KRUPOVIC, M. & LAVIGNE, R. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology,* 37**,** 632-639.

JEEWON, R., LUCKHUN, A. B., BHOYROO, V., SADEER, N. B., MAHOMOODALLY, M. F., RAMPADARATH, S., PUCHOOA, D., SARMA, V. V., DURAIRAJAN, S. S. K. & HYDE, K. D. 2019. Pharmaceutical Potential of Marine Fungal Endophytes. *Endophytes and Secondary Metabolites***,** 1-23.

JENKE-KODAMA, H. & DITTMANN, E. 2009. Evolution of metabolic diversity: insights from microbial polyketide synthases. *Phytochemistry,* 70**,** 1858-1866.

JENKE-KODAMA, H., SANDMANN, A., MÜLLER, R. & DITTMANN, E. 2005. Evolutionary implications of bacterial polyketide synthases. *Molecular biology and evolution,* 22**,** 2027-2039.

JERONEN, E. 2019. Ecology and Ecosystem: Sustainability. *Encyclopedia of Sustainable Management; Springer Nature: Cham Switzerland*.

JOSEFSEN, M. H., ANDERSEN, S. C., CHRISTENSEN, J. & HOORFAR, J. 2015. Microbial food safety: Potential of DNA extraction methods for use in diagnostic metagenomics. *Journal of Microbiological Methods,* 114**,** 30-34.

KALUZHNAYA, O. & ITSKOVICH, V. 2022. Features of Diversity of Polyketide Synthase Genes in the Community of Freshwater Sponge Baikalospongia fungiformis. *Russian Journal of Genetics,* 58**,** 336-346.

KANAZAWA, A., TESHIMA, S. & ONO, K. 1979. Relationship between essential fatty acid requirements of aquatic animals and the capacity for bioconversion of linolenic acid to highly unsaturated fatty acids. *Comparative biochemistry and physiology. B, Comparative Biochemistry,* 63**,** 295-298.

KANG, D. D., LI, F., KIRTON, E., THOMAS, A., EGAN, R., AN, H. & WANG, Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ,* 7**,** e7359.

KANG, H.-S. & KIM, E.-S. 2021. Recent advances in heterologous expression of natural product biosynthetic gene clusters in Streptomyces hosts. *Current Opinion in Biotechnology,* 69**,** 118-127.

KARENTZ, D. & BOSCH, I. 2001. Influence of ozone-related increases in ultraviolet radiation on Antarctic marine organisms. *American Zoologist,* 41**,** 3-16.

KARL, D. M., KNAUER, G. A. & MARTIN, J. H. 1988. Downward flux of particulate organic matter in the ocean: a particle decomposition paradox. *Nature,* 332**,** 438-441.

KATOH, K., MISAWA, K., KUMA, K. I. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research,* 30**,** 3059-3066.

KATOH, K., ROZEWICKI, J. & YAMADA, K. D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics,* 20**,** 1160-1166.

KATZ, L. & BALTZ, R. H. 2016. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology,* 43**,** 155-176.

KAUR, T. & DUTTA, P. S. 2020. Persistence and stability of interacting species in response to climate warming: the role of trophic structure. *Theoretical Ecology,* 13**,** 333-348.

KEALEY, J. T., CRAIG, J. P. & BARR, P. J. 2021. Identification of a lichen depside polyketide synthase gene by heterologous expression in Saccharomyces cerevisiae. *Metabolic Engineering Communications,* 13**,** e00172.

KENDALL, A. C. & NICOLAOU, A. 2013. Bioactive lipid mediators in skin inflammation and immunity. *Progress in Lipid Research,* 52**,** 141-164.

KHALID, S. & KELLER, N. P. 2021. Chemical signals driving bacterial–fungal interactions. *Environmental Microbiology,* 23**,** 1334-1347.

KHALIFA, S. A., ELIAS, N., FARAG, M. A., CHEN, L., SAEED, A., HEGAZY, M.-E. F., MOUSTAFA, M. S., ABD EL-WAHED, A., AL-MOUSAWI, S. M. & MUSHARRAF, S. G. 2019. Marine natural products: A source of novel anticancer drugs. Marine drugs, 17, 491.

KHAN, N., RAWLINGS, B. & CAFFREY, P. 2011. A labile point in mutant amphotericin polyketide synthases. *Biotechnology Letters,* 33**,** 1121-1126.

KHARE, D., HALE, W. A., TRIPATHI, A., GU, L., SHERMAN, D. H., GERWICK, W. H., HÅKANSSON, K. & SMITH, J. L. 2015. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase. *Structure,* 23**,** 2213-2223.

KHOZIN-GOLDBERG, I., ISKANDAROV, U. & COHEN, Z. 2011. LC-PUFA from photosynthetic microalgae: occurrence, biosynthesis, and prospects in biotechnology. *Applied Microbiology and Biotechnology,* 91**,** 905-915.

KIM, J. H., FENG, Z., BAUER, J. D., KALLIFIDAS, D., CALLE, P. Y. & BRADY, S. F. 2010. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers,* 93**,** 833-844.

KIM, J. H., LEE, N., HWANG, S., KIM, W., LEE, Y., CHO, S., PALSSON, B. O. & CHO, B.-K. 2021. Discovery of novel secondary metabolites encoded in actinomycete genomes through coculture. Journal of Industrial Microbiology and Biotechnology, 48, kuaa001.

KLASBERG, S., BITARD-FEILDEL, T. & MALLET, L. 2016. Computational identification of novel genes: current and future perspectives. Bioinformatics and Biology insights, 10, BBI. S39950.

KLAUS, M. & GRININGER, M. 2018. Engineering strategies for rational polyketide synthase design. *Natural Product Reports,* 35**,** 1070-1081.

KOGAWA, M., MIYAOKA, R., HEMMERLING, F., ANDO, M., YURA, K., IDE, K., NISHIKAWA, Y., HOSOKAWA, M., ISE, Y. & CAHN, J. K. 2021. Single-cell metabolite detection and genomics reveals uncultivated talented producer. *bioRxiv*.

KONG, D.-X., JIANG, Y.-Y. & ZHANG, H.-Y. 2010. Marine natural products as sources of novel scaffolds: Achievement and concern. Elsevier Current Trends.

KORNFUEHRER, T. & EUSTÁQUIO, A. S. 2019. Diversification of polyketide structures via synthase engineering. *MedChemComm*.

KOSCIELNY, G., AN, P., CARVALHO-SILVA, D., CHAM, J. A., FUMIS, L., GASPARYAN, R., HASAN, S., KARAMANIS, N., MAGUIRE, M. & PAPA, E. 2017. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Research,* 45**,** D985-D994.

KOTHRI, M., MAVROMMATI, M., ELAZZAZY, A. M., BAESHEN, M. N., MOUSSA, T. A. & AGGELIS, G. 2020. Microbial sources of polyunsaturated fatty acids (PUFAs) and the prospect of organic residues and wastes as growth media for PUFA-producing microorganisms. *FEMS Microbiology Letters,* 367**,** fnaa028.

KOWAL-BIELECKA, O., DISTLER, O., NEIDHART, M., KÜNZLER, P., RETHAGE, J., NAWRATH, M., CAROSSINO, A., PAP, T., MÜLLER-LADNER, U. & MICHEL, B. A. 2001. Evidence of 5-lipoxygenase overexpression in the skin of patients with systemic sclerosis: a newly identified pathway to skin inflammation in systemic sclerosis. *Arthritis & Rheumatism,* 44**,** 1865-1875.

KUMAR, R., MISHRA, A. & JHA, B. 2019. Bacterial community structure and functional diversity in subsurface seawater from the western coastal ecosystem of the Arabian Sea, India. *Gene,* 701**,** 55-64.

KURNIA, N. M., URIA, A. R., KUSNADI, Y., DINAWATI, L., ZILDA, D. S., HADI, T. A., SETYAHADI, S. & FELIX, F. 2017. Metagenomic Survey of Potential Symbiotic Bacteria and Polyketide Synthase Genes in an Indonesian Marine Sponge. *HAYATI Journal of Biosciences,* 24**,** 6-15.

KURTBÖKE, İ. 2017. Revisiting biodiscovery from microbial sources in the light of molecular advances. *Microbiology Australia,* 38**,** 58-61.

LANDSCHÜTZER, P., GRUBER, N., HAUMANN, F. A., RÖDENBECK, C., BAKKER, D. C., VAN HEUVEN, S., HOPPEMA, M., METZL, N., SWEENEY, C. & TAKAHASHI, T. 2015. The reinvigoration of the Southern Ocean carbon sink. *Science,* 349**,** 1221-1224.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A. & LOPEZ, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LASRAM, S., ZEMNI, H., HAMDI, Z., CHENENAOUI, S., HOUISSA, H., TOUNSI, M. S. & GHORBEL, A. 2019. Antifungal and antiaflatoxinogenic activities of Carum carvi L., Coriandrum sativum L. seed essential oils and their major terpene component against Aspergillus flavus. *Industrial Crops and Products,* 134**,** 11-18.

LAVROVA, N., PREOBRAZHENSKAIA, T. & SVESHNIKOVA, M. 1972. Isolation of soil actinomycetes on selective media with rubomycin. *Antibiotiki,* 17**,** 965-970.

LEE, M. D. 2019. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics,* 35**,** 4162-4164.

LEE, N., HWANG, S., KIM, W., LEE, Y., KIM, J. H., CHO, S., KIM, H. U., YOON, Y. J., OH, M.-K. & PALSSON, B. O. 2021. Systems and synthetic biology to elucidate secondary metabolite biosynthetic gene clusters encoded in Streptomyces genomes. *Natural Product Reports,* 38**,** 1330-1361.

LEMETRE, C., MANIKO, J., CHARLOP-POWERS, Z., SPARROW, B., LOWE, A. J. & BRADY, S. F. 2017. Bacterial natural product biosynthetic domain composition in soil correlates with changes in latitude on a continent-wide scale. *Proceedings of the National Academy of Sciences,* 114**,** 11615-11620.

LETUNIC, I. & BORK, P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research,* 49**,** W293-W296.

LEVY-BOOTH, D. J., HASHIMI, A., ROCCOR, R., LIU, L.-Y., RENNECKAR, S., ELTIS, L. D. & MOHN, W. W. 2021. Genomics and metatranscriptomics of biogeochemical cycling and degradation of lignin-derived aromatic compounds in thermal swamp sediment. The ISME journal, 15, 879-893.

LEWIS, K. 2012. Antibiotics: Recover the lost art of drug discovery. *Nature,* 485**,** 439.

LI, D., LUO, R., LIU, C.-M., LEUNG, C.-M., TING, H.-F., SADAKANE, K., YAMASHITA, H. & LAM, T.-W. 2016. MEGAHIT v1. 0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods,* 102**,** 3-11.

LI, J. S., BARBER, C. C. & ZHANG, W. 2019a. Natural products from anaerobes. *Journal of Industrial Microbiology & Biotechnology,* 46**,** 375-383.

LI, L., KOH, C. C., REKER, D., BROWN, J., WANG, H., LEE, N. K., LIOW, H.-H., DAI, H., FAN, H.-M. & CHEN, L. 2019b. Predicting protein-ligand interactions based on bow-pharmacological space and Bayesian additive regression trees. *Scientific Reports,* 9**,** 1-12.

LI, W., MA, Z., CHEN, L. & YIN, W.-B. 2018. Synthesis and production of the antitumor polyketide aurovertins and structurally related compounds. *Applied Microbiology and Biotechnology,* 102**,** 6373-6381.

LIBIS, V., ANTONOVSKY, N., ZHANG, M., SHANG, Z., MONTIEL, D., MANIKO, J., TERNEI, M. A., CALLE, P. Y., LEMETRE, C. & OWEN, J. G. 2019. Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences. *Nature Communications,* 10**,** 1-9.

LITTLEPAGE, J. L. 1965. Oceanographic investigations in McMurdo sound, Antarctica. *Biology of The Antarctic Seas II,* 5**,** 1-37.

LIU, J., NOTHIAS, L.-F., DORRESTEIN, P. C., TAHLAN, K. & BIGNELL, D. R. 2021. Genomic and metabolomic analysis of the potato common scab pathogen Streptomyces Scabiei. *ACS Omega,* 6**,** 11474-11487.

LIU, L., WANG, Y., YANG, Y., WANG, D., CHENG, S. H., ZHENG, C. & ZHANG, T. 2021. Charting the complexity of the activated sludge microbiome through a hybrid sequencing strategy. *Microbiome,* 9**,** 1-15.

LIVERMORE, R., NANKIVELL, A., EAGLES, G. & MORRIS, P. 2005. Paleogene opening of Drake passage. *Earth and Planetary Science Letters,* 236**,** 459-470.

LLOYD, K. G., STEEN, A. D., LADAU, J., YIN, J. & CROSBY, L. 2018. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *MSystems,* 3**,** e00055-18.

LO, Y.-C., RENSI, S. E., TORNG, W. & ALTMAN, R. B. 2018. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today,* 23**,** 1538-1546.

LOGES, L. A., SILVA, D. B., PAULINO, G. V., LANDELL, M. F. & MACEDO, A. J. 2020. Polyketides from marine-derived Aspergillus welwitschiae inhibit Staphylococcus aureus virulence factors and potentiate vancomycin antibacterial activity in vivo. *Microbial Pathogenesis,* 143**,** 104066.

LONGHURST, A. R. 2010. *Ecological Geography of The Sea*, Elsevier.

LUEPKE, K. H., SUDA, K. J., BOUCHER, H., RUSSO, R. L., BONNEY, M. W., HUNT, T. D. & MOHR III, J. F. 2017. Past, present, and future of antibacterial economics: increasing bacterial resistance, limited antibiotic pipeline, and societal implications. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy,* 37**,** 71-84.

LUIGI CAPELLA, G. 2011. Strategies for Leukotriene Modulation in Dermatology: Even More Visionary Perspectives? An Update. *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Inflammatory and Anti-Allergy Agents)*, 10**,** 407-417.

MAHAJAN, G. B. & BALACHANDRAN, L. 2017. Sources of antibiotics: Hot springs. Biochemical pharmacology, 134, 35-41.

MAKINDE, O. A., LUNGA, P. M., BEZUIDT, O. K. & MAKHALANYANE, T. P. 2022. Four Lentisphaerae Family Metagenome-Assembled Genomes from the South Atlantic Ocean. Microbiology Resource Announcements, 11, e00496-22.

MALVE, H. 2016. Exploring the ocean for new drug developments: Marine pharmacology. *Journal of Pharmacy & Bioallied Sciences,* 8**,** 83.

MARCY, Y., OUVERNEY, C., BIK, E. M., LÖSEKANN, T., IVANOVA, N., MARTIN, H. G., SZETO, E., PLATT, D., HUGENHOLTZ, P. & RELMAN, D. A. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences,* 104**,** 11889-11894.

MARSTON, H. D., DIXON, D. M., KNISELY, J. M., PALMORE, T. N. & FAUCI, A. S. 2016. Antimicrobial resistance. *Jama,* 316**,** 1193-1204.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal,* 17**,** 10-12.

MARTÍNEZ, A. & OSBURNE, M. S. 2013. Preparation of fosmid libraries and functional metagenomic analysis of microbial community DNA. *Methods in Enzymology*. Elsevier.

MARTÍNEZ-NÚÑEZ, M. A. & RODRÍGUEZ-ESCAMILLA, Z. 2020. Mining the Yucatan Coastal Microbiome for the Identification of Non-Ribosomal Peptides Synthetase (NRPS) Genes. *Toxins,* 12**,** 349.

MARYA, N. B., POWERS, P. D., CHARI, S. T., GLEESON, F. C., LEGGETT, C. L., DAYYEH, B. K. A., CHANDRASEKHARA, V., IYER, P. G., MAJUMDER, S. & PEARSON, R. K. 2021. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut,* 70**,** 1335-1344.

MASALHA, M., RAYAN, M., ADAWI, A., ABDALLAH, Z. & RAYAN, A. 2018. Capturing antibacterial natural products with in silico techniques. *Molecular Medicine Reports,* 18**,** 763-770.

MAST, Y. & STEGMANN, E. 2019. Actinomycetes: The antibiotics producers. Multidisciplinary Digital Publishing Institute.

MAYER, A., RODRÍGUEZ, A. D., TAGLIALATELA-SCAFATI, O. & FUSETANI, N. 2013. Marine pharmacology in 2009–2011: Marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, and antiviral activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of action. *Marine Drugs,* 11**,** 2510-2573.

MAYER, A. M., GUERRERO, A. J., RODRÍGUEZ, A. D., TAGLIALATELA-SCAFATI, O., NAKAMURA, F. & FUSETANI, N. 2019. Marine pharmacology in 2014–2015: Marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, antiviral, and anthelmintic activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of action. *Marine Drugs,* 18**,** 5.

MEDEMA, M. H. & FISCHBACH, M. A. 2015. Computational approaches to natural product discovery. *Nature Chemical Biology,* 11**,** 639.

MEDEMA, M. H., KOTTMANN, R., YILMAZ, P., CUMMINGS, M., BIGGINS, J. B., BLIN, K., DE BRUIJN, I., CHOOI, Y. H., CLAESEN, J. & COATES, R. C. 2015. Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology,* 11**,** 625.

MELESHKO, D., MOHIMANI, H., TRACANNA, V., HAJIRASOULIHA, I., MEDEMA, M. H., KOROBEYNIKOV, A. & PEVZNER, P. A. 2019. BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. *Genome Research,* 29**,** 1352-1362.

MÉNDEZ-GARCÍA, C., BARGIELA, R., MARTÍNEZ-MARTÍNEZ, M. & FERRER, M. 2018. Metagenomic Protocols and Strategies. *Metagenomics.* Elsevier.

METSÄ-KETELÄ, M., HALO, L., MUNUKKA, E., HAKALA, J., MÄNTSÄLÄ, P. & YLIHONKO, K. 2002. Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various Streptomyces species. *Applied and Environmental Microbiology,* 68**,** 4472-4479.

MILICI, M., VITAL, M., TOMASCH, J., BADEWIEN, T. H., GIEBEL, H. A., PLUMEIER, I., WANG, H., PIEPER, D. H., WAGNER-DÖBLER, I. & SIMON, M. 2017. Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: evidence of dispersal limitation in the Bransfield Strait. *Limnology and Oceanography,* 62**,** 1080-1095.

MOORE, C. M., MILLS, M. M., ACHTERBERG, E. P., GEIDER, R. J., LAROCHE, J., LUCAS, M. I., MCDONAGH, E. L., PAN, X., POULTON, A. J. & RIJKENBERG, M. J. 2009. Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geoscience,* 2**,** 867-871.

MOORE, A. D., BJÖRKLUND, Å. K., EKMAN, D., BORNBERG-BAUER, E. & ELOFSSON, A. 2008. Arrangements in the modular evolution of proteins. Trends in biochemical sciences, 33, 444-451.

MOORE, C., MILLS, M., ARRIGO, K., BERMAN-FRANK, I., BOPP, L., BOYD, P., GALBRAITH, E., GEIDER, R., GUIEU, C. & JACCARD, S. 2013. Processes and patterns of oceanic nutrient limitation. Nature geoscience, 6, 701-710.

MORALES-SÁNCHEZ, D., SCHULZE, P. S., KIRON, V. & WIJFFELS, R. H. 2020. Production of carbohydrates, lipids and polyunsaturated fatty acids (PUFA) by the polar marine microalga Chlamydomonas malina RCC2488. *Algal Research,* 50**,** 102016.

MORI, T., CAHN, J. K., WILSON, M. C., MEODED, R. A., WIEBACH, V., MARTINEZ, A. F. C., HELFRICH, E. J., ALBERSMEIER, A., WIBBERG, D. & DÄTWYLER, S. 2018. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated Entotheonella sponge symbionts. *Proceedings of the National Academy of Sciences,* 115**,** 1718-1723.

MUCHLINSKI, A., CHEN, X., LOVELL, J. T., KÖLLNER, T. G., PELOT, K. A., ZERBE, P., RUGGIERO, M., CALLAWAY III, L., LALIBERTE, S. & CHEN, F. 2019. Biosynthesis and emission of stress-induced volatile terpenes in roots and leaves of switchgrass (Panicum virgatum L.). *Frontiers in plant science,* 10**,** 1144.

MUTHUBHARATHI, B. C., GOWRIPRIYA, T. & BALAMURUGAN, K. 2021. Metabolomics: small molecules that matter more. *Molecular Omics,* 17**,** 210-229.

NAUGHTON, L. M., ROMANO, S., O'GARA, F. & DOBSON, A. D. 2017. Identification of secondary metabolite gene clusters in the Pseudovibrio genus reveals encouraging

biosynthetic potential toward the production of novel bioactive compounds. *Frontiers in Microbiology,* 8**,** 1494.

NAVARRO-MUÑOZ, J. C., SELEM-MOJICA, N., MULLOWNEY, M. W., KAUTSAR, S. A., TRYON, J. H., PARKINSON, E. I., DE LOS SANTOS, E. L., YEONG, M., CRUZ-MORALES, P. & ABUBUCKER, S. 2020. A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology,* 16**,** 60-68.

NAYFACH, S., CAMARGO, A. P., SCHULZ, F., ELOE-FADROSH, E., ROUX, S. & KYRPIDES, N. C. 2021a. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology,* 39**,** 578-585.

NAYFACH, S., ROUX, S., SESHADRI, R., UDWARY, D., VARGHESE, N., SCHULZ, F., WU, D., PAEZ-ESPINO, D., CHEN, I.-M. & HUNTEMANN, M. 2021b. A genomic catalog of Earth's microbiomes. *Nature Biotechnology,* 39**,** 499-509.

NEWMAN, D. J. & CRAGG, G. M. 2016. Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products,* 79**,** 629-661.

NEWMAN, D. J. & CRAGG, G. M. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products,* 83**,** 770-803.

NGUYEN, T., ISHIDA, K., JENKE-KODAMA, H., DITTMANN, E., GURGUI, C., HOCHMUTH, T., TAUDIEN, S., PLATZER, M., HERTWECK, C. & PIEL, J. 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature Biotechnology,* 26**,** 225.

NICHOLS, D. S. 2003. Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiology Letters,* 219**,** 1-7.

NIESELT, K., BATTKE, F., HERBIG, A., BRUHEIM, P., WENTZEL, A., JAKOBSEN, Ø. M., SLETTA, H., ALAM, M. T., MERLO, M. E. & MOORE, J. 2010. The dynamic architecture of the metabolic switch in Streptomyces coelicolor. *BMC Genomics,* 11**,** 10.

NISHIDA, T., ORIKASA, Y., WATANABE, K. & OKUYAMA, H. 2006. The cell membrane-shielding function of eicosapentaenoic acid for Escherichia coli against exogenously added hydrogen peroxide. *FEBS Letters,* 580**,** 6690-6694.

NIVINA, A., YUET, K. P., HSU, J. & KHOSLA, C. 2019. Evolution and Diversity of Assembly-Line Polyketide Synthases: Focus Review. *Chemical Reviews,* 119**,** 12524-12547.

NOAR, R. D. & DAUB, M. E. 2016. Bioinformatics prediction of polyketide synthase gene clusters from Mycosphaerella fijiensis. *PLoS One,* 11**,** e0158471.

NÚÑEZ-PONS, L., AVILA, C., ROMANO, G., VERDE, C. & GIORDANO, D. 2018. UV-protective compounds in marine organisms from the Southern Ocean. *Marine Drugs,* 16**,** 336.

NURK, S., KOREN, S., RHIE, A., RAUTIAINEN, M., BZIKADZE, A. V., MIKHEENKO, A., VOLLGER, M. R., ALTEMOSE, N., URALSKY, L. & GERSHMAN, A. 2022. The complete sequence of a human genome. *Science,* 376**,** 44-53.

NURK, S., MELESHKO, D., KOROBEYNIKOV, A. & PEVZNER, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Research,* 27**,** 824-834.

NWEZE, J. A., MBAOJI, F. N., HUANG, G., LI, Y., YANG, L., ZHANG, Y., HUANG, S., PAN, L. & YANG, D. 2020. Antibiotics development and the potentials of marine-derived compounds to stem the tide of multidrug-resistant pathogenic bacteria, fungi, and protozoa. *Marine Drugs,* 18**,** 145.

O'HAGAN, D. 1991. *The polyketide metabolites,* Ellis Horwood Limited.

O'NEILL, J. 2014. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. *Review Antimicrobial Resistance*. http://amr-review.org/Publications.

ODUM, E. P. & BARRETT, G. W. 1971. *Fundamentals of ecology*, Saunders Philadelphia.

OJA, T., NIIRANEN, L., SANDALOVA, T., KLIKA, K. D., NIEMI, J., MÄNTSÄLÄ, P., SCHNEIDER, G. & METSÄ-KETELÄ, M. 2013. Structural basis for C-ribosylation in the alnumycin A biosynthetic pathway. *Proceedings of the National Academy of Sciences*, 110**,** 1291-1296.

OJA, T., PALMU, K., LEHMUSSOLA, H., LEPPÄRANTA, O., HÄNNIKÄINEN, K., NIEMI, J., MÄNTSÄLÄ, P. & METSÄ-KETELÄ, M. 2008. Characterization of the alnumycin gene cluster reveals unusual gene products for pyran ring formation and dioxan biosynthesis. *Chemistry & Biology*, 15**,** 1046-1057.

OLA, A. R., TAWO, B. D., BELLI, H. L. L., PROKSCH, P., TOMMY, D. & HAKIM, E. H. 2018. A new antibacterial polyketide from the endophytic fungi Aspergillus fumigatiaffinis. *Natural Product Communications*, 13**,** 1934578X1801301202.

PALAZZOTTO, E., TONG, Y., LEE, S. Y. & WEBER, T. 2019. Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. *Biotechnology Advances*.

PALAZZOTTO, E. & WEBER, T. 2018. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Current Opinion in Microbiology*, 45**,** 109-116.

PAOLI, L., RUSCHEWEYH, H.-J., FORNERIS, C. C., HUBRICH, F., KAUTSAR, S., BHUSHAN, A., LOTTI, A., CLAYSSEN, Q., SALAZAR, G., MILANESE, A., CARLSTRÖM, C. I., PAPADOPOULOU, C., GEHRIG, D., KARASIKOV, M., MUSTAFA, H., LARRALDE, M., CARROLL, L. M., SÁNCHEZ, P., ZAYED, A. A., CRONIN, D. R., ACINAS, S. G., BORK, P., BOWLER, C., DELMONT, T. O., GASOL, J. M., GOSSERT, A. D., KAHLES, A., SULLIVAN, M. B., WINCKER, P., ZELLER, G., ROBINSON, S. L., PIEL, J. & SUNAGAWA, S. 2022. Biosynthetic potential of the global ocean microbiome. Nature, 607, 111-118.

PARKS, D. H., CHUVOCHINA, M., CHAUMEIL, P.-A., RINKE, C., MUSSIG, A. J. & HUGENHOLTZ, P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, 38**,** 1079-1086.

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**,** 1043-1055.

PARKS, D. H., RINKE, C., CHUVOCHINA, M., CHAUMEIL, P.-A., WOODCROFT, B. J., EVANS, P. N., HUGENHOLTZ, P. & TYSON, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2**,** 1533-1542.

PATTANAIK, B. & LINDBERG, P. 2015. Terpenoids and their biosynthesis in cyanobacteria. *Life*, 5**,** 269-293.

PENG, S. X. 2000. Hyphenated HPLC-NMR and its applications in drug discovery. *Biomedical Chromatography*, 14**,** 430-441.

PENG, Y., LEUNG, H. C., YIU, S.-M. & CHIN, F. Y. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28**,** 1420-1428.

PÉREZ-SÁNCHEZ, A., BARRAJÓN-CATALÁN, E., HERRANZ-LÓPEZ, M. & MICOL, V. 2018. Nutraceuticals for skin care: A comprehensive review of human clinical studies. *Nutrients,* 10**,** 403.

PHAM, V. D., KONSTANTINIDIS, K. T., PALDEN, T. & DELONG, E. F. 2008. Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environmental Microbiology,* 10**,** 2313-2330.

PHAM, V. H. & KIM, J. 2012. Cultivation of unculturable soil bacteria. *Trends in Biotechnology,* 30**,** 475-484.

PHOMA, B. S. & MAKHALANYANE, T. P. 2021. Depth-dependent variables shape community structure and functionality in the Prince Edward Islands. *Microbial Ecology,* 81**,** 396-409.

PHOMA, S., VIKRAM, S., JANSSON, J. K., ANSORGE, I. J., COWAN, D. A., VAN DE PEER, Y. & MAKHALANYANE, T. P. 2018. Agulhas Current properties shape microbial community diversity and potential functionality. *Scientific Reports,* 8**,** 1-12.

PRESTERL, E., DIAB-EL SCHAHAWI, M., LUSIGNANI, L. S., PAULA, H. & REILLY, J. S. 2019. Antimicrobial Agents (Antibiotics). *Basic Microbiology and Infection Control for Midwives.* Springer.

PRICE, M. N., DEHAL, P. S. & ARKIN, A. P. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution,* 26**,** 1641-1650.

PRIHODA, D., MARITZ, J. M., KLEMPIR, O., DZAMBA, D., WOELK, C. H., HAZUDA, D. J., BITTON, D. A. & HANNIGAN, G. D. 2021. The application potential of machine learning and genomics for understanding natural product diversity, chemistry, and therapeutic translatability. *Natural Product Reports,* 38**,** 1100-1108.

PROUD, R., COX, M. J. & BRIERLEY, A. S. 2017. Biogeography of the global ocean's mesopelagic zone. *Current Biology,* 27**,** 113-119.

QUEGUINER, B. 2013. Iron fertilization and the structure of planktonic communities in high nutrient regions of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography,* 90**,** 43-54.

QIU, F., ZENG, J., WANG, J., HUANG, J. P., ZHOU, W., YANG, C., LAN, X., CHEN, M., HUANG, S. X. & KAI, G. 2020. Functional genomics analysis reveals two novel genes required for littorine biosynthesis. New Phytologist, 225, 1906-1914.

QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN, N. J. & SEGATA, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology,* 35**,** 833.

RATH, C. M., JANTO, B., EARL, J., AHMED, A., HU, F. Z., HILLER, L., DAHLGREN, M., KREFT, R., YU, F. & WOLFF, J. J. 2011. Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chemical Biology,* 6**,** 1244-1256.

RAYAN, M., ABDALLAH, Z., ABU-LAFI, S., MASALHA, M. & RAYAN, A. 2019. Indexing natural products for their antifungal activity by filters-based approach: Disclosure of discriminative properties. *Current Computer-Aided Drug Design,* 15**,** 235-242.

REGO, A., FERNANDEZ-GUERRA, A., DUARTE, P., ASSMY, P., LEÃO, P. N. & MAGALHÃES, C. 2021. Secondary metabolite biosynthetic diversity in Arctic Ocean metagenomes. *Microbial Genomics,* 7.

REGO, A., SOUSA, A. G., SANTOS, J. P., PASCOAL, F., CANÁRIO, J., LEÃO, P. N. & MAGALHÃES, C. 2020. Diversity of Bacterial Biosynthetic Genes in Maritime Antarctica. *Microorganisms,* 8**,** 279.

REYGONDEAU, G., GUIEU, C., BENEDETTI, F., IRISSON, J.-O., AYATA, S.-D., GASPARINI, S. & KOUBBI, P. 2017. Biogeochemical regions of the Mediterranean Sea: An objective multidimensional and multivariate environmental approach. *Progress in oceanography,* 151**,** 138-148.

RIDLEY, C. P., LEE, H. Y. & KHOSLA, C. 2008. Evolution of polyketide synthases in bacteria. *Proceedings of the National Academy of Sciences,* 105**,** 4595-4600.

RINKE, C., SCHWIENTEK, P., SCZYRBA, A., IVANOVA, N. N., ANDERSON, I. J., CHENG, J.-F., DARLING, A., MALFATTI, S., SWAN, B. K. & GIES, E. A. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature,* 499**,** 431-437.

RISDIAN, C., MOZEF, T. & WINK, J. 2019. Biosynthesis of Polyketides in Streptomyces. *Microorganisms,* 7**,** 124.

ROBINSON, S. L., PIEL, J. & SUNAGAWA, S. 2021. A roadmap for metagenomic enzyme discovery. *Natural Product Reports,* 38**,** 1994-2023.

RODRÍGUEZ-HERNÁNDEZ, D., MELO, W. G., MENEGATTI, C., LOURENZON, V. B., DO NASCIMENTO, F. S. & PUPO, M. T. 2019. Actinobacteria associated with stingless bees biosynthesize bioactive polyketides against bacterial pathogens. *New Journal of Chemistry,* 43**,** 10109-10117.

ROMERO, D., TRAXLER, M., LÓPEZ, D. & KOLTER, R. 2011. Antibiotics as signal molecules. *Chemical Reviews,* 111, 5492–5505.

RUST, M., HELFRICH, E. J., FREEMAN, M. F., NANUDORN, P., FIELD, C. M., RÜCKERT, C., KÜNDIG, T., PAGE, M. J., WEBB, V. L. & KALINOWSKI, J. 2020. A multiproducer microbiome generates chemical diversity in the marine sponge Mycale hentscheli. *Proceedings of the National Academy of Sciences,* 117**,** 9508-9518.

RUXTON, C., REED, S. C., SIMPSON, M. & MILLINGTON, K. 2004. The health benefits of omega-3 polyunsaturated fatty acids: a review of the evidence. *Journal of Human Nutrition and Dietetics,* 17**,** 449-459.

RUZZINI, A. C. & CLARDY, J. 2016. Gene flow and molecular innovation in bacteria. *Current Biology,* 26**,** R859-R864.

SABINE, C. L., FEELY, R. A., GRUBER, N., KEY, R. M., LEE, K., BULLISTER, J. L., WANNINKHOF, R., WONG, C., WALLACE, D. W. & TILBROOK, B. 2004. The oceanic sink for anthropogenic CO2. *Science,* 305**,** 367-371.

SAITO, A. Y., RODRIGUEZ, A. A. M., VEGA, D. S. M., SUSSMANN, R. A., KIMURA, E. A. & KATZIN, A. M. 2016. Antimalarial activity of the terpene nerolidol. *International Journal of Antimicrobial Agents,* 48**,** 641-646.

SAITO, T. & AWAI, K. 2020. A polyketide synthase HglEA, but not HglE2, synthesizes heterocyst specific glycolipids in Anabaena sp. PCC 7120. *The Journal of General and Applied Microbiology***,** 2019.11. 004.

SALMASO, N., ALBANESE, D., CAPELLI, C., BOSCAINI, A., PINDO, M. & DONATI, C. 2018. Diversity and cyclical seasonal transitions in the bacterial community in a large and deep perialpine lake. Microbial ecology, 76, 125-143.

SCHALK, I. 2018. Siderophore-antibiotic conjugates: exploiting iron uptake to deliver drugs into bacteria. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases,* 24**,** 801.

SCHERLACH, K. & HERTWECK, C. 2020. Chemical mediators at the bacterial-fungal interface. *Annual Review of Microbiology,* 74**,** 267-290.

SCHMIDT, R., JAGER, V. D., ZÜHLKE, D., WOLFF, C., BERNHARDT, J., CANKAR, K., BEEKWILDER, J., IJCKEN, W. V., SLEUTELS, F. & BOER, W. D. 2017. Fungal volatile compounds induce production of the secondary metabolite Sodorifen in Serratia plymuthica PRI-2C. *Scientific Reports,* 7**,** 1-14.

SCHMIDT, S., KILDGAARD, S., GUO, H., BEEMELMANNS, C. & POULSEN, M. 2022. The chemical ecology of the fungus-farming termite symbiosis. *Natural Product Reports,* 39**,** 231-248.

SCHNEIDER, Y. K. 2021. Bacterial Natural Product Drug Discovery for New Antibiotics: Strategies for Tackling the Problem of Antibiotic Resistance by Efficient Bioprospecting. *Antibiotics,* 10**,** 842.

SCHÖNER, T. A., GASSEL, S., OSAWA, A., TOBIAS, N. J., OKUNO, Y., SAKAKIBARA, Y., SHINDO, K., SANDMANN, G. & BODE, H. B. 2016. Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *ChemBioChem,* 17**,** 247-253.

SCHORN, M. A., ALANJARY, M. M., AGUINALDO, K., KOROBEYNIKOV, A., PODELL, S., PATIN, N., LINCECUM, T., JENSEN, P. R., ZIEMERT, N. & MOORE, B. S. 2016. Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology,* 162**,** 2075.

SCHULZE, P. S., HULATT, C. J., MORALES-SÁNCHEZ, D., WIJFFELS, R. H. & KIRON, V. 2019. Fatty acids and proteins from marine cold adapted microalgae for biotechnology. *Algal Research,* 42**,** 101604.

SEKUROVA, O. N., SCHNEIDER, O. & ZOTCHEV, S. B. 2019. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. Microbial biotechnology, 12, 828-844.

SELVIN, J., SATHIYANARAYANAN, G., LIPTON, A. N., AL-DHABI, N. A., VALAN ARASU, M. & KIRAN, G. S. 2016. Ketide synthase (KS) domain prediction and analysis of iterative type II PKS gene in marine sponge-associated actinobacteria producing biosurfactants and antimicrobial agents. *Frontiers in Microbiology,* 7**,** 63.

SHARMA, P. & SHARMA, P. 2012. *Ecology and environment,* Rastogi Publications.

SHARRAR, A. M., CRITS-CHRISTOPH, A., MÉHEUST, R., DIAMOND, S., STARR, E. P. & BANFIELD, J. F. 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *Mbio,* 11.

SHEIK, C. S., JAIN, S. & DICK, G. J. 2014. Metabolic flexibility of enigmatic SAR 324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology,* 16**,** 304-317.

SHELEST, E., HEIMERL, N., FICHTNER, M. & SASSO, S. 2015. Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. *BMC Genomics,* 16**,** 1015.

SHERMAN, K., SISSENWINE, M., CHRISTENSEN, V., DUDA, A., HEMPEL, G., IBE, C., LEVIN, S., LLUCH-BELDA, D., MATISHOV, G. & MCGLADE, J. 2005. A global movement toward an ecosystem approach to management of marine resources. *Marine Ecology Progress Series,* 300**,** 275-279.

SHIMIZU, Y., OGATA, H. & GOTO, S. 2017a. Discriminating the reaction types of plant type III polyketide synthases. *Bioinformatics,* 33**,** 1937-1943.

SHIMIZU, Y., OGATA, H. & GOTO, S. 2017b. Type III polyketide synthases: functional classification and phylogenomics. *ChemBioChem,* 18**,** 50-65.

SHIOZAKI, T., IJICHI, M., KODAMA, T., TAKEDA, S. & FURUYA, K. 2014. Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. Global Biogeochemical Cycles, 28, 1096-1110.

SHOKRALLA, S., SPALL, J. L., GIBSON, J. F. & HAJIBABAEI, M. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular ecology, 21, 1794-1805.

SHOUVAL, R., FEIN, J. A., SAVANI, B., MOHTY, M. & NAGLER, A. 2021. Machine learning and artificial intelligence in haematology. *British Journal of Haematology,* 192**,** 239-250.

SIEBER, C. M., PROBST, A. J., SHARRAR, A., THOMAS, B. C., HESS, M., TRINGE, S. G. & BANFIELD, J. F. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology,* 3**,** 836-843.

SIENIAWSKA, E., SWATKO-OSSOR, M., SAWICKI, R., SKALICKA-WOŹNIAK, K. & GINALSKA, G. 2017. Natural terpenes influence the activity of antibiotics against isolated Mycobacterium tuberculosis. *Medical Principles and Practice,* 26**,** 108-112.

SILKNETTER, S., CREED, R. P., BROWN, B. L., FRIMPONG, E. A., SKELTON, J. & PEOPLES, B. K. 2020. Positive biotic interactions in freshwaters: A review and research directive. *Freshwater Biology,* 65**,** 811-832.

SINGH ET AL. 2019. *Polyketide and Its Derivatives*.

SIRO, G., PIPITE, A., CHRISTI, K., SRINIVASAN, S. & SUBRAMANI, R. 2022. Marine actinomycetes associated with stony corals: A potential hotspot for specialized metabolites. *Microorganisms,* 10**,** 1349.

SKIBA, M. A., SIKKEMA, A. P., FIERS, W. D., GERWICK, W. H., SHERMAN, D. H., ALDRICH, C. C. & SMITH, J. L. 2016. Domain organization and active site architecture of a polyketide synthase C-methyltransferase. ACS chemical biology, 11, 3319-3327.

SMITH, S. & TSAI, S.-C. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports,* 24**,** 1041-1072.

SOHM, J. A., HILTON, J. A., NOBLE, A. E., ZEHR, J. P., SAITO, M. A. & WEBB, E. A. 2011. Nitrogen fixation in the South Atlantic Gyre and the Benguela upwelling system. *Geophysical Research Letters,* 38.

SPALDING, M. D., FOX, H. E., ALLEN, G. R., DAVIDSON, N., FERDAÑA, Z. A., FINLAYSON, M., HALPERN, B. S., JORGE, M. A., LOMBANA, A. & LOURIE, S. A. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience,* 57**,** 573-583.

SRIVASTAVA, N., SARETHY, I. P., JEEVANANDAM, J. & DANQUAH, M. 2022. Emerging strategies for microbial screening of novel chemotherapeutics. *Journal of Molecular Structure***,** 132419.

STEPANAUSKAS, R. 2012. Single cell genomics: an individual look at microbes. *Current Opinion in Microbiology,* 15**,** 613-620.

STUART, K. A., WELSH, K., WALKER, M. C. & EDRADA-EBEL, R. 2020. Metabolomic tools used in marine natural product drug discovery. *Expert Opinion on Drug Discovery,* 15**,** 499-522.

SU, M., SATOLA, S. W. & READ, T. D. 2019. Genome-based prediction of bacterial antibiotic resistance. *Journal of Clinical Microbiology,* 57**,** e01405-18.

SU, G., MORRIS, J. H., DEMCHAK, B. & BADER, G. D. 2014. Biological network exploration with Cytoscape 3. Current protocols in bioinformatics, 47, 8.13. 1-8.13. 24.

SUBRAMANI, R. & SIPKEMA, D. 2019. Marine rare actinomycetes: a promising source of structurally diverse and unique novel natural products. Marine drugs, 17, 249.

SUCIU, R. M., LUVAGA, I. K., HAZEEN, A., WEERASOORIYA, C., RICHARDSON, S. K., FIRESTONE, A. J., SHANNON, K., HOWELL, A. R. & CRAVATT, B. F. 2021. Chemical proteomic analysis of palmostatin beta-lactone analogs that affect N-Ras palmitoylation. *Bioorganic & medicinal chemistry letters*, 53**,** 128414.

SUKUL, P., SCHÄKERMANN, S., BANDOW, J. E., KUSNEZOWA, A., NOWROUSIAN, M. & LEICHERT, L. I. 2017. Simple discovery of bacterial biocatalysts from environmental samples through functional metaproteomics. *Microbiome*, 5**,** 1-5.

SUNAGAWA, S., COELHO, L. P., CHAFFRON, S., KULTIMA, J. R., LABADIE, K., SALAZAR, G., DJAHANSCHIRI, B., ZELLER, G., MENDE, D. R. & ALBERTI, A. 2015. Structure and function of the global ocean microbiome. *Science*, 348**,** 1261359.

SWAN, B. K., MARTINEZ-GARCIA, M., PRESTON, C. M., SCZYRBA, A., WOYKE, T., LAMY, D., REINTHALER, T., POULTON, N. J., MASLAND, E. D. P. & GOMEZ, M. L. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science*, 333**,** 1296-1300.

TAUTENHAHN, R., CHO, K., URITBOONTHAI, W., ZHU, Z., PATTI, G. J. & SIUZDAK, G. 2012. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature Biotechnology*, 30**,** 826-828.

TEAM, R. C. 2013. R: A language and environment for statistical computing.

THOMPSON, L. R., SANDERS, J. G., MCDONALD, D., AMIR, A., LADAU, J., LOCEY, K. J., PRILL, R. J., TRIPATHI, A., GIBBONS, S. M. & ACKERMANN, G. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551**,** 457.

TIMMERMANS, M. L., PAUDEL, Y. P. & ROSS, A. C. 2017. Investigating the biosynthesis of natural products from marine proteobacteria: a survey of molecules and strategies. *Marine Drugs*, 15**,** 235.

TIWARI, K. & GUPTA, R. K. 2012. Rare actinomycetes: a potential storehouse for novel antibiotics. *Critical Reviews in Biotechnology*, 32**,** 108-132.

TOBIAS, N. J., WOLFF, H., DJAHANSCHIRI, B., GRUNDMANN, F., KRONENWERTH, M., SHI, Y.-M., SIMONYI, S., GRÜN, P., SHAPIRO-ILAN, D. & PIDOT, S. J. 2017. Natural product diversity associated with the nematode symbionts Photorhabdus and Xenorhabdus. *Nature Microbiology*, 2**,** 1676-1685.

TORRES-MARTÍNEZ, R., GARCÍA-RODRÍGUEZ, Y. M., RÍOS-CHÁVEZ, P., SAAVEDRA-MOLINA, A., LÓPEZ-MEZA, J. E., OCHOA-ZARZOSA, A. & GARCIGLIA, R. S. 2017. Antioxidant activity of the essential oil and its major terpenes of Satureja macrostema (Moc. and Sessé ex Benth.) Briq. *Pharmacognosy Magazine*, 13**,** S875.

TORTORELLA, E., TEDESCO, P., PALMA ESPOSITO, F., JANUARY, G. G., FANI, R., JASPARS, M. & DE PASCALE, D. 2018. Antibiotics from deep-sea microorganisms: current discoveries and perspectives. *Marine Drugs*, 16**,** 355.

TRILLA-PRIETO, N. R., VILA-COSTA, M., CASAS, G., JIMÉNEZ, B. & DACHS, J. 2021. Dissolved Black Carbon and Semivolatile Aromatic Hydrocarbons in the Ocean: Two Entangled Biogeochemical Cycles? Environmental Science & Technology Letters, 8, 918-923.

TRINDADE-SILVA, A. E., RUA, C. P., ANDRADE, B. G., VICENTE, A. C. P., SILVA, G. G., BERLINCK, R. G. & THOMPSON, F. L. 2013. Polyketide synthase gene diversity within the microbiome of the sponge Arenosclera brasiliensis, endemic to the Southern Atlantic Ocean. *Applied and Environmental Microbiology,* 79**,** 1598-1605.

TUTTLE, R. N., DEMKO, A. M., PATIN, N. V., KAPONO, C. A., DONIA, M. S., DORRESTEIN, P. & JENSEN, P. R. 2019. Detection of natural products and their producers in ocean sediments. *Applied and Environmental Microbiology,* 85**,** e02830-18.

TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. & BANFIELD, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature,* 428**,** 37-43.

UEOKA, R., URIA, A. R., REITER, S., MORI, T., KARBAUM, P., PETERS, E. E., HELFRICH, E. J., MORINAKA, B. I., GUGGER, M. & TAKEYAMA, H. 2015. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nature Chemical Biology,* 11**,** 705-712.

UNITED NATION NEWS. 2019. *Global health agencies sound alarm on drug-resistant infections; new recommendations to reduce 'staggering number' of future deaths* https://news.un.org/en/story/2019/04/1037471.

VAMATHEVAN, J., CLARK, D., CZODROWSKI, P., DUNHAM, I., FERRAN, E., LEE, G., LI, B., MADABHUSHI, A., SHAH, P. & SPITZER, M. 2019. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery***,** 1.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics,* 30**,** 418-426.

VAN GOETHEM, M. W., OSBORN, A. R., BOWEN, B. P., ANDEER, P. F., SWENSON, T. L., CLUM, A., RILEY, R., HE, G., KORIABINE, M. & SANDOR, L. 2021. Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. *Communications biology,* 4**,** 1302.

VANWONTERGHEM, I., JENSEN, P. D., RABAEY, K. & TYSON, G. W. 2016. Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. *Environmental Microbiology,* 18**,** 3144-3158.

VENTOLA, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics,* 40**,** 277.

VILLEBRO, R., SHAW, S., BLIN, K. & WEBER, T. 2019. Sequence-based classification of type II polyketide synthase biosynthetic gene clusters for antiSMASH. *Journal of Industrial Microbiology & Biotechnology,* 46**,** 469-475.

VON BUBNOFF, A. 2006. Seeking new antibiotics in nature's backyard. *Cell,* 127**,** 867-869.

VON MEIJENFELDT, F. B., ARKHIPOVA, K., CAMBUY, D. D., COUTINHO, F. H. & DUTILH, B. E. 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology,* 20**,** 1-14.

WAITE, D. W., CHUVOCHINA, M., PELIKAN, C., PARKS, D. H., YILMAZ, P., WAGNER, M., LOY, A., NAGANUMA, T., NAKAI, R. & WHITMAN, W. B. 2020. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *International Journal of Systematic and Evolutionary Microbiology,* 70**,** 5972-6016.

WANG, C.-C., ZHAO, Y. & CHEN, X. 2020. Drug-pathway association prediction: from experimental results to computational models. *Briefings in Bioinformatics*.

WANG, H., FEWER, D. P., HOLM, L., ROUHIAINEN, L. & SIVONEN, K. 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences,* 111**,** 9259-9264.

WANG, Y., YANG, Q. & WANG, Z. 2015. The evolution of nanopore sequencing. *Frontiers in Genetics,* 5**,** 449.

WARD, L. M., HEMP, J., SHIH, P. M., MCGLYNN, S. E. & FISCHER, W. W. 2018. Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer. *Frontiers in Microbiology,* 9**,** 260.

WASCHULIN, V., BORSETTO, C., JAMES, R., NEWSHAM, K. K., DONADIO, S., CORRE, C. & WELLINGTON, E. 2022. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *The ISME Journal,* 16**,** 101-111.

WATSON, S. B. 2003. Cyanobacterial and eukaryotic algal odour compounds: signals or by-products? A review of their biological activity. *Phycologia,* 42**,** 332-350.

WECKWERTH, W. 2007. *Metabolomics: methods and protocols,* Springer Science & Business Media.

WEI, Y., ZHANG, L., ZHOU, Z. & YAN, X. 2018. Diversity of gene clusters for polyketide and nonribosomal peptide biosynthesis revealed by metagenomic analysis of the yellow sea sediment. *Frontiers in Microbiology,* 9**,** 295.

WEIRATHER, J. L., DE CESARE, M., WANG, Y., PIAZZA, P., SEBASTIANO, V., WANG, X.-J., BUCK, D. & AU, K. F. 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research,* 6.

WICKHAM, H. & WICKHAM, M. H. 2007. The ggplot package. *URL: https://cran. r-project. org/web/packages/ggplot2/index. html*.

WISHART, D. S. 2008. Applications of metabolomics in drug discovery and development. *Drugs in R & D,* 9**,** 307-322.

WONG, C. H., SIAH, K. W. & LO, A. W. 2019. Estimation of clinical trial success rates and related parameters. *Biostatistics,* 20**,** 273-286.

WORLD HEALTH ORGANIZATION. 2014. *Antimicrobial resistance: global report on surveillance* https://www.who.int/mediacentre/news/releases/2014/amr-report/en/.

WRIGHT, T. D., VERGIN, K. L., BOYD, P. W. & GIOVANNONI, S. J. 1997. A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Applied and Environmental Microbiology,* 63**,** 1441-1448.

WU, C., YIN, Y., ZHU, L., ZHANG, Y. & LI, Y.-Z. 2022. Metagenomic sequencing-driven multidisciplinary approaches to shed light on the untapped microbial natural products. *Drug Discovery Today,* 27**,** 730-742.

WU, Y.-W., SIMMONS, B. A. & SINGER, S. W. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics,* 32**,** 605-607.

XIANG, X. 2020. Structural and biochemical studies of erythromycin-resistance methyltransferases: insights into structure-based drug design.

XIE, X., GARG, A., KEATINGE-CLAY, A. T., KHOSLA, C. & CANE, D. E. 2016. Epimerase and reductase activities of polyketide synthase ketoreductase domains utilize the same conserved tyrosine and serine residues. *Biochemistry,* 55**,** 1179-1186.

XUE, C.-X., LIU, J., LEA-SMITH, D. J., ROWLEY, G., LIN, H., ZHENG, Y., ZHU, X.-Y., LIANG, J., AHMAD, W. & TODD, J. D. 2020. Insights into the vertical stratification of microbial ecological roles across the deepest seawater column on Earth. *Microorganisms,* 8**,** 1309.

YADAV, G., GOKHALE, R. S. & MOHANTY, D. 2009. Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS Computational Biology,* 5**,** e1000351.

YANG, Q., ZHANG, A.-H., MIAO, J.-H., SUN, H., HAN, Y., YAN, G.-L., WU, F.-F. & WANG, X.-J. 2019. Metabolomics biotechnology, applications, and future trends: a systematic review. RSC advances, 9, 37245-37257.

YOSHIDA, K., HASHIMOTO, M., HORI, R., ADACHI, T., OKUYAMA, H., ORIKASA, Y., NAGAMINE, T., SHIMIZU, S., UENO, A. & MORITA, N. 2016. Bacterial long-chain polyunsaturated fatty acids: their biosynthetic genes, functions, and practical use. *Marine Drugs,* 14**,** 94.

YU, D., XU, F., ZENG, J. & ZHAN, J. 2012. Type III polyketide synthases in natural product biosynthesis. *IUBMB life,* 64**,** 285-295.

ZABALA, D., SONG, L., DASHTI, Y., CHALLIS, G. L., SALAS, J. A. & MÉNDEZ, C. 2020. Heterologous reconstitution of the biosynthesis pathway for 4-demethyl-premithramycinone, the aglycon of antitumor polyketide mithramycin. *Microbial Cell Factories,* 19**,** 1-12.

ZEHR, J. P. & CAPONE, D. G. 2020. Changing perspectives in marine nitrogen fixation. Science, 368, eaay9514.

ZHANG, A., SUN, H., WANG, P., HAN, Y. & WANG, X. 2012. Modern analytical techniques in metabolomics analysis. *Analyst,* 137**,** 293-300.

ZHANG, A., SUN, H. & WANG, X. 2014. Potential Applications and Development of Cell Metabolomics in Natural Products. *Journal of Drug Metabolism and Toxicology,* 5**,** 2.

ZHANG, A., SUN, H., XU, H., QIU, S. & WANG, X. 2013. Cell metabolomics. *Omics: a Journal of Integrative Biology,* 17**,** 495-501.

ZHANG, H., LIU, Y., WANG, X., HU, R., XU, G., MAO, C., INAM, M., WANG, Y., JIANG, X. & KONG, L. 2021. Gene sequence diversity of the nonribosomal peptide and polyketide natural products in Changbaishan soil correlates with changes in landscape belts. *Ecological Indicators,* 133**,** 108160.

ZHANG, M. M., WONG, F. T., WANG, Y., LUO, S., LIM, Y. H., HENG, E., YEO, W. L., COBB, R. E., ENGHIAD, B. & ANG, E. L. 2017a. CRISPR–Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nature Chemical Biology,* 13**,** 607.

ZHANG, Z., PAN, H.-X. & TANG, G.-L. 2017b. New insights into bacterial type II polyketide biosynthesis. *F1000Research,* 6.

ZHANG, X.-W., LI, Q.-H. & DOU, J.-J. 2020. Mass spectrometry-based metabolomics in health and medical science: A systematic review. RSC advances, 10, 3092-3104.

ZHANG, T., ZHANG, Y., LI, L., JIANG, X., CHEN, Z., ZHAO, F. & YI, Y. 2022. Biosynthesis and production of class II bacteriocins of food-associated lactic acid bacteria. *Fermentation,* 8**,** 217.

ZHOU, Y., PREDIGER, P., DIAS, L. C., MURPHY, A. C. & LEADLAY, P. F. 2015. Macrodiolide formation by the thioesterase of a modular polyketide synthase. *Angewandte Chemie International Edition,* 54**,** 5232-5235.

ZIBOH, V. A., MILLER, C. C. & CHO, Y. 2000. Metabolism of polyunsaturated fatty acids by skin epidermal enzymes: generation of antiinflammatory and antiproliferative metabolites. *The American Journal of Clinical Nutrition,* 71**,** 361s-366s.

ZIEMERT, N. & JENSEN, P. R. 2012. Phylogenetic approaches to natural product structure prediction. *Methods in Enzymology.* Elsevier.

ZIEMERT, N., LECHNER, A., WIETZ, M., MILLÁN-AGUIÑAGA, N., CHAVARRIA, K. L. & JENSEN, P. R. 2014. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. *Proceedings of the National Academy of Sciences,* 111**,** E1130-E1139.

ZIEMERT, N., PODELL, S., PENN, K., BADGER, J. H., ALLEN, E. & JENSEN, P. R. 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS one,* 7**,** e34064.

ZOFFMANN, S., VERCRUYSSE, M., BENMANSOUR, F., MAUNZ, A., WOLF, L., MARTI, R. B., HECKEL, T., DING, H., TRUONG, H. H. & PRUMMER, M. 2019. Machine learning-powered antibiotics phenotypic drug discovery. *Scientific Reports,* 9**,** 5013.

Appendices

Appendix 1

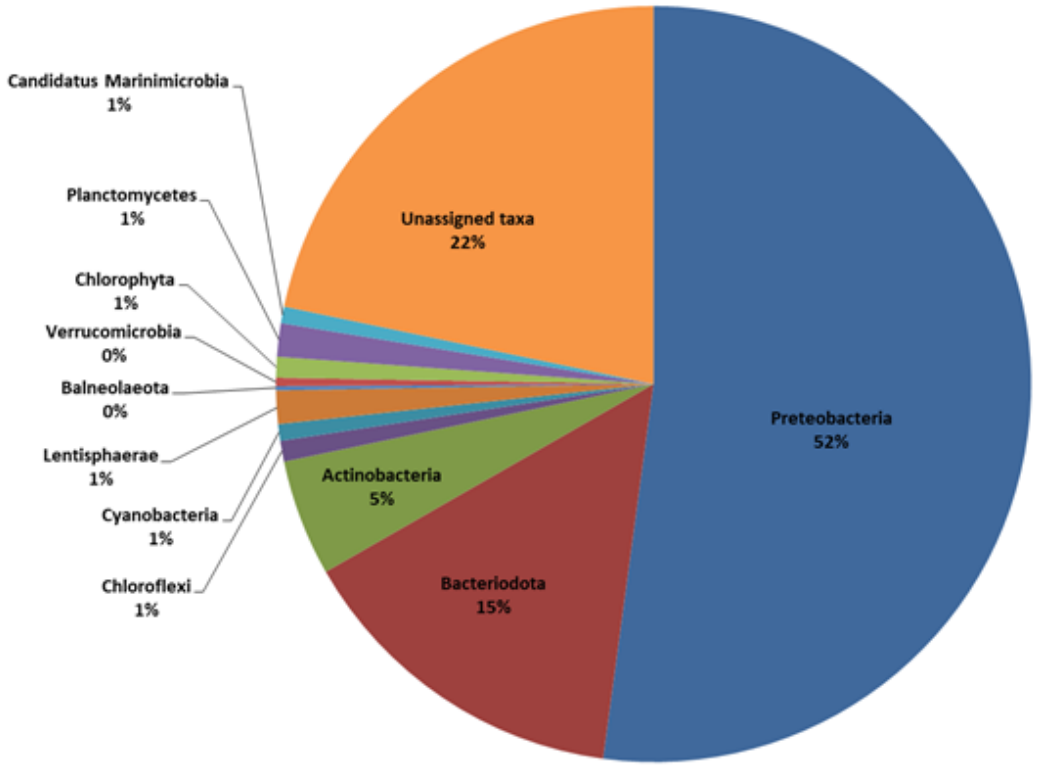

Figure S1: The taxonomy provenience of the contigs harboring the non-redundant KSs from SAO.

Appendix 2

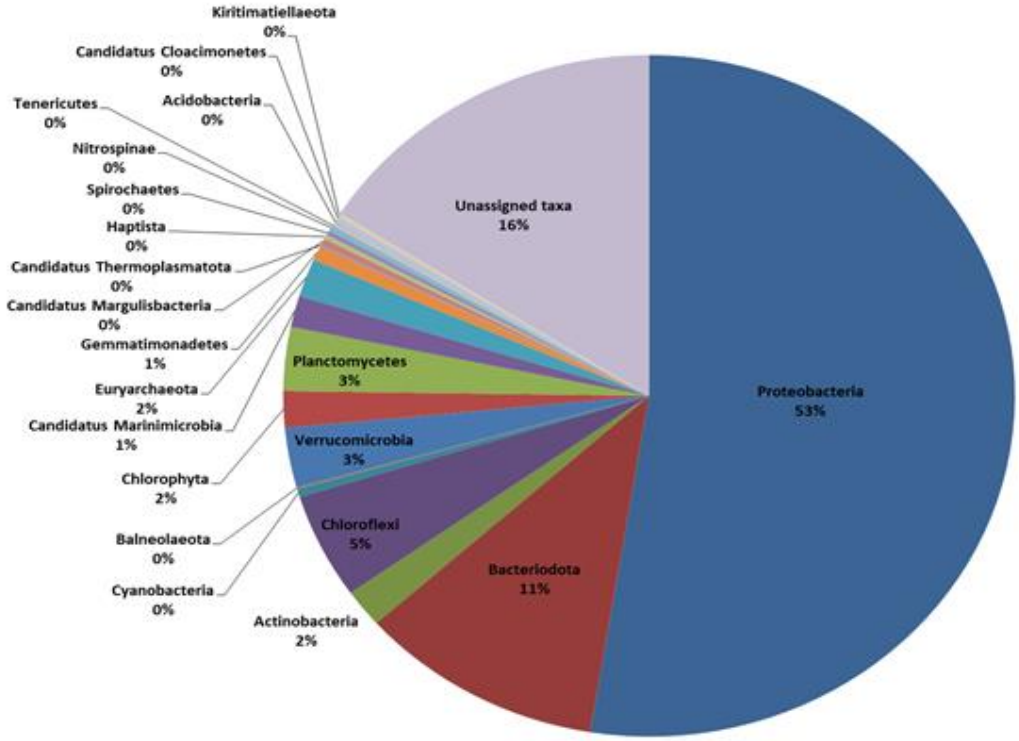

Figure S2: The taxonomy provenience of the contigs harboring the non-redundant KSs from TARA.

Appendix 3

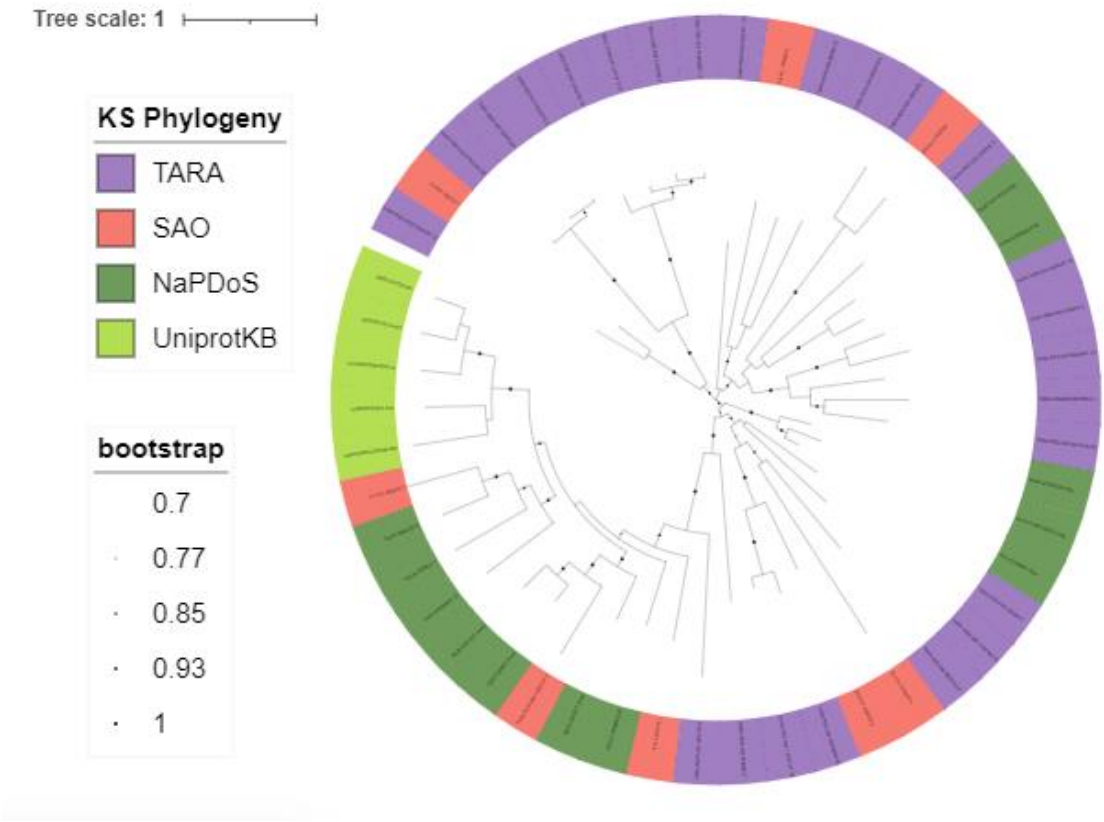

Figure S3: Phylogenetic tree of the 7 SAO (This study), 26 Tara Ocean with their NaPDoS and UniprotKB database KS hits. Most of the KSs from this study aligned to that of the Tara Ocean despite the fact that got hit with the NaPDoS and UniprotKB KSs

Appendix 4

Table S1: SAO euphotic zones coordinate

Sampling number stattions	Latitude	Longitude	Air temperature	Humidity	Salinity Temperature	Air pressure/bar	Surface salinity/PSU
1	34°30.265'S	12°38.087'E	14.8	57%	16.71	1.025	35.46
2	34°30.086'S	06°49.622'E	13.9	68%	15.92	1.022	35.46
3	34°39.249'S	00°43.394'W	15.0	87%	16.33	1.006	35.57
4	36°35.591'S	10°06.384'W	10.0	87%	13.99	1.009	35.13
5	38°22.947'S	11°14.942'W	8.8	74%	12.82	1.011	34.96
6	39°05.226'S	10°36.910'W	9.4	66%	12.63	1.015	34.95
7	41°32.438'S	09°59.916'W	10.2	90%	10.75	1.029	34.67
8	47°28.906'S	09°59.970'W	4.6	99%	3.98	1.000	33.78

Appendix 5

Table S2: Showing the taxonomy provenience of the contigs harboring the non-redundant KSs from SAO

# contig	superkingdom	phylum	class	order	family	genus	species
10_k141_134455	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium TMED226: 1.00
10_k141_168623	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support
10_k141_186322	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.80
10_k141_221256	Bacteria: 0.96	Proteobacteria: 0.96	Gammaproteobacteria: 0.96	Pseudomonadales: 0.96	Moraxellaceae: 0.96	Psychrobacter: 0.93	no support
10_k141_23602	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.70
10_k141_295956	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
10_k141_300809	no support	no support	no support	no support	no support	no support	no support
10_k141_435931	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium: 0.60
10_k141_447286	Bacteria: 0.91	Actinobacteria: 0.91	Actinomycetia: 0.91	no support	no support	no support	no support
10_k141_484731	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
10_k141_509418	Bacteria: 1.00	Chloroflexi: 1.00	no support	no support	no support	no support	no support
10_k141_548346	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
10_k141_56722	Bacteria: 1.00	Actinobacteria: 0.99	Actinomycetia: 0.99	no support	no support	no support	no support
10_k141_569952	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.95	Pseudomonadales: 0.95	Moraxellaceae: 0.95	Psychrobacter: 0.95	no support
10_k141_569960	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.72	Oceanospirillales: 0.72	Kangiellaceae: 0.72	Kangiella: 0.72	no support
10_k141_580458	Bacteria: 0.91	no support	no support	no support	no support	no support	no support
10_k141_605615	Bacteria: 1.00	Proteobacteria: 0.97	Gammaproteobacteria: 0.97	no support	no support	no support	no support
10_k141_661505	Bacteria: 0.92	Cyanobacteria: 0.92	NA	Synechococcales: 0.92	Prochloraceae: 0.92	Prochlorococcus: 0.92	no support
10_k141_679757	Bacteria: 1.00	Proteobacteria: 0.88	Gammaproteobacteria: 0.88	no support	no support	no support	no support
10_k141_849482	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support
10_k141_862166	Bacteria: 0.68	Actinobacteria: 0.68	no support	no support	no support	no support	no support

10_k141_873368	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
10_k141_941237	Bacteria: 1.00	Bacteroidetes: 0.91	Flavobacteriia: 0.91	Flavobacteriales: 0.91	Flavobacteriaceae: 0.91	no support	no support
10_k141_994982	Bacteria: 0.86	Lentisphaerae: 0.54	Lentisphaeria: 0.53	Lentisphaerales: 0.51	Lentisphaeraceae: 0.51	no support	no support
11_k141_1036746	Bacteria: 1.00	Proteobacteria: 0.74	Gammaproteobacteria: 0.74	no support	no support	no support	no support
11_k141_1088515	Bacteria: 0.93	Proteobacteria: 0.72	no support	no support	no support	no support	no support
11_k141_1098531	Bacteria: 0.51	no support	no support	no support	no support	no support	no support
11_k141_128779	Bacteria: 1.00	Chloroflexi: 1.00	no support	no support	no support	no support	no support
11_k141_203925	Bacteria: 0.99	Proteobacteria: 0.90	Gammaproteobacteria: 0.83	no support	no support	no support	no support
11_k141_236961	Bacteria: 0.98	Proteobacteria: 0.95	Gammaproteobacteria: 0.88	Vibrionales: 0.60	no support	no support	no support
11_k141_254968	no support	no support	no support	no support	no support	no support	no support
11_k141_272457	Bacteria: 0.85	no support	no support	no support	no support	no support	no support
11_k141_280191	Bacteria: 0.97	Proteobacteria: 0.95	Gammaproteobacteria: 0.94	Vibrionales: 0.57	no support	no support	no support
11_k141_288184	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
11_k141_389331	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Vibrionales: 0.79	Vibrionaceae: 0.79	Photobacterium: 0.79	no support
11_k141_414433	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
11_k141_444075	Bacteria: 1.00	Bacteroidetes: 0.87	no support	no support	no support	no support	no support
11_k141_474531	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
11_k141_499329	Bacteria: 1.00	Actinobacteria: 0.93	Actinomycetia: 0.55	NA	NA	NA	Actinobacteria bacterium: 0.53
11_k141_524149	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	no support	no support	no support
11_k141_538133	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
11_k141_569222	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Vibrionales: 1.00	Vibrionaceae: 1.00	Vibrio: 1.00	no support
11_k141_641980	Bacteria: 1.00	Bacteroidetes: 0.92	Flavobacteriia: 0.92	Flavobacteriales: 0.92	Flavobacteriaceae: 0.92	NA	Flavobacteriaceae bacterium: 0.92
11_k141_770495	no support	no support	no support	no support	no support	no support	no support
11_k141_857945	Bacteria: 1.00	Proteobacteria: 0.75	Gammaproteobacteria: 0.75	no support	no support	no support	no support
11_k141_86364	Bacteria: 0.89	no support	no support	no support	no support	no support	no support
11_k141_942231	Bacteria: 1.00	Proteobacteria: 0.77	no support	no support	no support	no support	no support
11_k141_956206	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support
12_k141_1026498	Bacteria: 0.78	Proteobacteria: 0.78	no support	no support	no support	no support	no support
12_k141_1074131	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
12_k141_236173	Bacteria: 1.00	Proteobacteria: 0.87	Gammaproteobacteria: 0.87	no support	no support	no support	no support

12_k141_27542	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
12_k141_334417	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
12_k141_347618	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
12_k141_654680	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
12_k141_667193	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
12_k141_718223	Bacteria: 1.00	Balneolaeota: 0.57	Balneolia: 0.57	Balneolales: 0.57	Balneolaceae: 0.57	Balneola: 0.57	no support
12_k141_938416	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
12_k141_95577	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium: 0.81
12_k141_973732	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 1.00	Rhodobiaceae: 1.00	Amorphus: 1.00	Amorphus coralli*: 1.00
13_k141_105401	Bacteria: 0.81	Bacteroidetes: 0.64	Flavobacteriia: 0.64	Flavobacteriales: 0.64	Flavobacteriaceae: 0.64	no support	no support
13_k141_173175	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
13_k141_177830	Bacteria: 1.00	Proteobacteria: 0.86	Alphaproteobacteria: 0.86	Pelagibacterales: 0.86	NA	NA	Pelagibacterales bacterium SAG-MED32: 0.86
13_k141_179656	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	NA	NA	NA	SAR116 cluster bacterium: 1.00
13_k141_201867	Bacteria: 0.91	Proteobacteria: 0.91	Alphaproteobacteria: 0.91	no support	no support	no support	no support
13_k141_22851	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	NA	Rhodobacteraceae bacterium: 1.00
13_k141_254135	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
13_k141_27070	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 0.66	no support	no support
13_k141_353863	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	NA	Rhodobacteraceae bacterium: 1.00
13_k141_372264	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
13_k141_438144	Bacteria: 1.00	Bacteroidetes: 1.00	no support	no support	no support	no support	no support
13_k141_445123	no support	no support	no support	no support	no support	no support	no support
13_k141_458188	no support	no support	no support	no support	no support	no support	no support
13_k141_491816	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
13_k141_508260	Bacteria: 0.98	Proteobacteria: 0.98	Alphaproteobacteria: 0.98	Rhodobacterales: 0.98	Rhodobacteraceae: 0.97	NA	Rhodobacteraceae bacterium: 0.97
13_k141_553218	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
13_k141_555845	Bacteria: 0.80	Proteobacteria: 0.52	no support	no support	no support	no support	no support
13_k141_569378	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
13_k141_570367	Bacteria: 1.00	Verrucomicrobia: 0.90	Verrucomicrobiae: 0.90	Verrucomicrobiales: 0.90	no support	no support	no support
13_k141_637990	Eukaryota: 1.00	Chlorophyta: 1.00	Mamiellophyceae: 1.00	Mamiellales: 1.00	Bathycoccaceae: 1.00	Bathycoccus: 1.00	Bathycoccus prasinos*: 1.00

13_k141_699125	Bacteria: 1.00	Actinobacteria: 0.89	no support	no support	no support	no support	no support
13_k141_746611	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium TMED61: 1.00
13_k141_75535	Bacteria: 1.00	Proteobacteria: 0.94	Alphaproteobacteria: 0.94	Rhodobacterales: 0.94	Rhodobacteraceae: 0.71	no support	no support
13_k141_864914	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	NA	NA	NA	Alphaproteobacteria bacterium: 1.00
13_k141_865803	Bacteria: 1.00	Planctomycetes: 1.00	no support	no support	no support	no support	no support
13_k141_888939	Bacteria: 0.98	Proteobacteria: 0.98	Alphaproteobacteria: 0.98	Rhodobacterales: 0.98	Rhodobacteraceae: 0.96	NA	Rhodobacteraceae bacterium: 0.96
13_k141_915898	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
13_k141_925562	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
13_k141_939321	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
13_k141_941699	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 0.73	no support	no support
13_k141_953875	Bacteria: 0.93	NA	NA	NA	NA	NA	bacterium: 0.93
13_k141_955160	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 0.78	no support	no support	no support
13_k141_980930	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.76	no support	no support	no support	no support
13_k141_98537	Bacteria: 1.00	Actinobacteria: 1.00	no support	no support	no support	no support	no support
14_k141_1012099	Bacteria: 0.60	Bacteroidetes: 0.55	no support	no support	no support	no support	no support
14_k141_1014186	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
14_k141_1029109	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
14_k141_1033523	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 0.95	Rhodobacteraceae: 0.88	NA	Rhodobacteraceae bacterium: 0.81
14_k141_1036163	NA	NA	NA	NA	NA	NA	uncultured organism MedDCM-OCT-S04-C478: 0.52
14_k141_1074864	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.71
14_k141_1088418	Bacteria: 1.00	Lentisphaerae: 0.81	Lentisphaeria: 0.68	Lentisphaerales: 0.61	Lentisphaeraceae: 0.61	no support	no support
14_k141_1188494	Bacteria: 1.00	Bacteroidetes: 0.94	Flavobacteriia: 0.94	Flavobacteriales: 0.94	no support	no support	no support
14_k141_129047	Bacteria: 1.00	Lentisphaerae: 1.00	Lentisphaeria: 1.00	Lentisphaerales: 1.00	Lentisphaeraceae: 1.00	NA	Lentisphaeraceae bacterium*: 0.74
14_k141_140905	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
14_k141_143323	Bacteria: 1.00	Bacteroidetes: 0.93	Flavobacteriia: 0.93	Flavobacteriales: 0.93	NA	NA	Flavobacteriales bacterium: 0.93
14_k141_149681	Eukaryota: 1.00	Chlorophyta: 1.00	Mamiellophyceae: 1.00	Mamiellales: 1.00	Mamiellaceae: 1.00	Micromonas: 1.00	no support
14_k141_15179	Bacteria: 1.00	Bacteroidetes: 0.94	Flavobacteriia: 0.94	Flavobacteriales: 0.94	Flavobacteriaceae: 0.94	NA	Flavobacteriaceae bacterium: 0.61
14_k141_223467	Bacteria: 0.76	no support	no support	no support	no support	no support	no support
14_k141_298901	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.70

14_k141_324473	Bacteria: 0.85	Proteobacteria: 0.85	Alphaproteobacteria: 0.85	Pelagibacterales: 0.85	NA	NA	Pelagibacterales bacterium: 0.85
14_k141_356488	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
14_k141_405531	Bacteria: 1.00	Planctomycetes: 0.90	no support	no support	no support	no support	no support
14_k141_438417	Bacteria: 1.00	Proteobacteria: 0.99	Alphaproteobacteria: 0.99	Rhodobacterales: 0.98	Rhodobacteraceae: 0.97	Sulfitobacter: 0.87	no support
14_k141_494019	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 0.81	Rhodobacteraceae: 0.81	no support	no support
14_k141_615262	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.76
14_k141_626152	Bacteria: 0.94	Proteobacteria: 0.94	Alphaproteobacteria: 0.94	Rhodobacterales: 0.94	no support	no support	no support
14_k141_635947	Bacteria: 1.00	Planctomycetes: 1.00	no support	no support	no support	no support	no support
14_k141_649119	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
14_k141_652364	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
14_k141_678859	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
14_k141_753018	Bacteria: 1.00	Verrucomicrobia: 1.00	NA	NA	NA	NA	Verrucomicrobia bacterium: 1.00
14_k141_776388	Bacteria: 1.00	Proteobacteria: 0.90	Alphaproteobacteria: 0.90	no support	no support	no support	no support
14_k141_795239	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodospirillales: 1.00	no support	no support	no support
14_k141_897973	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.87	Oceanospirillales: 0.87	Alcanivoracaceae: 0.87	Alcanivorax: 0.87	Alcanivorax sp.: 0.87
14_k141_963741	Bacteria: 0.97	Proteobacteria: 0.97	Gammaproteobacteria: 0.97	NA	NA	NA	Gammaproteobacteria bacterium: 0.78
15_k141_102478	Bacteria: 1.00	Bacteroidetes: 0.96	Flavobacteriia: 0.70	Flavobacteriales: 0.70	Flavobacteriaceae: 0.70	no support	no support
15_k141_106375	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
15_k141_146662	Bacteria: 1.00	Proteobacteria: 0.90	no support	no support	no support	no support	no support
15_k141_272442	Bacteria: 1.00	Bacteroidetes: 1.00	no support	no support	no support	no support	no support
15_k141_526901	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
15_k141_609318	Bacteria: 0.95	Actinobacteria: 0.84	Nitriliruptoria: 0.73	Euzebyales: 0.73	Euzebyaceae: 0.73	Euzebya: 0.73	no support
15_k141_864794	Bacteria: 1.00	Bacteroidetes: 0.87	Flavobacteriia: 0.87	Flavobacteriales: 0.87	no support	no support	no support
15_k141_87354	Eukaryota: 1.00	Chlorophyta: 1.00	Mamiellophyceae: 1.00	Mamiellales: 1.00	Mamiellaceae: 1.00	Micromonas: 1.00	Micromonas commoda: 0.58
16_k141_1095244	Bacteria: 0.93	Lentisphaerae: 0.71	Lentisphaeria: 0.66	Lentisphaerales: 0.66	Lentisphaeraceae: 0.66	no support	no support
16_k141_163115	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	NA	NA	NA	Alphaproteobacteria bacterium: 1.00
16_k141_187429	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
16_k141_295939	no support	no support	no support	no support	no support	no support	no support
16_k141_311071	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support

16_k141_319369	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium TMED61: 0.84
16_k141_339408	Bacteria: 1.00	Bacteroidetes: 0.86	no support	no support	no support	no support	no support
16_k141_392261	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
16_k141_414130	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	no support	no support
16_k141_506975	Bacteria: 1.00	Bacteroidetes: 0.76	Flavobacteriia: 0.76	Flavobacteriales: 0.76	no support	no support	no support
16_k141_552797	no support	no support	no support	no support	no support	no support	no support
16_k141_567513	Bacteria: 0.68	no support	no support	no support	no support	no support	no support
16_k141_580000	Eukaryota: 1.00	NA	Pelagophyceae: 0.51	Pelagomonadales: 0.51	NA	Aureococcus: 0.51	Aureococcus anophagefferens*: 0.51
16_k141_697129	Bacteria: 0.54	no support	no support	no support	no support	no support	no support
16_k141_698950	no support	no support	no support	no support	no support	no support	no support
16_k141_73806	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
16_k141_752405	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
16_k141_805129	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
16_k141_814350	Eukaryota: 1.00	NA	Pelagophyceae: 1.00	Pelagomonadales: 1.00	NA	Aureococcus: 1.00	Aureococcus anophagefferens*: 1.00
16_k141_884553	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
16_k141_91757	Bacteria: 1.00	Lentisphaerae: 0.76	Lentisphaeria: 0.70	Lentisphaerales: 0.58	Lentisphaeraceae: 0.58	no support	no support
16_k141_920012	Bacteria: 0.90	no support	no support	no support	no support	no support	no support
16_k141_982964	Bacteria: 0.63	no support	no support	no support	no support	no support	no support
17_k141_239172	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
17_k141_271030	Bacteria: 1.00	Lentisphaerae: 0.85	Lentisphaeria: 0.83	Lentisphaerales: 0.71	Lentisphaeraceae: 0.71	NA	Lentisphaeraceae bacterium*: 0.60
17_k141_336292	Bacteria: 1.00	Lentisphaerae: 0.83	Lentisphaeria: 0.66	Lentisphaerales: 0.64	Lentisphaeraceae: 0.64	no support	no support
17_k141_358034	Bacteria: 0.90	no support	no support	no support	no support	no support	no support
17_k141_389870	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
17_k141_434876	no support	no support	no support	no support	no support	no support	no support
17_k141_451521	Bacteria: 1.00	Actinobacteria: 0.80	no support	no support	no support	no support	no support
17_k141_462954	Bacteria: 1.00	Proteobacteria: 0.56	Gammaproteobacteria: 0.56	NA	NA	Candidatus Thioglobus: 0.56	Candidatus Thioglobus sp. NP1: 0.56
17_k141_540426	Bacteria: 1.00	Bacteroidetes: 0.92	Cytophagia: 0.92	no support	no support	no support	no support
17_k141_560283	Bacteria: 0.85	Proteobacteria: 0.85	Betaproteobacteria: 0.82	no support	no support	no support	no support
17_k141_565457	Bacteria: 0.92	no support	no support	no support	no support	no support	no support

17_k141_599502	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
17_k141_639123	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
17_k141_738967	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.64	no support	no support	no support	no support
17_k141_798470	Bacteria: 0.58	no support	no support	no support	no support	no support	no support
18_k141_133495	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
18_k141_186503	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
18_k141_252393	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	NA	Rhodobacteraceae bacterium: 1.00
18_k141_353367	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
18_k141_437839	Eukaryota: 1.00	NA	Pelagophyceae: 1.00	Pelagomonadales: 1.00	NA	Aureococcus: 1.00	Aureococcus anophagefferens*: 1.00
18_k141_554657	Bacteria: 0.81	Bacteroidetes: 0.74	no support	no support	no support	no support	no support
18_k141_673355	Bacteria: 1.00	Bacteroidetes: 0.91	no support	no support	no support	no support	no support
18_k141_700482	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
18_k141_80123	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
18_k141_865812	Bacteria: 1.00	Bacteroidetes: 0.88	no support	no support	no support	no support	no support
18_k141_898049	Eukaryota: 1.00	Chlorophyta: 1.00	Mamiellophyceae: 1.00	Mamiellales: 1.00	Bathycoccaceae: 1.00	Bathycoccus: 1.00	Bathycoccus prasinos*: 1.00
19_k141_179943	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.51
19_k141_288541	Bacteria: 1.00	Planctomycetes: 1.00	no support	no support	no support	no support	no support
19_k141_298365	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
19_k141_385876	Bacteria: 1.00	Planctomycetes: 1.00	no support	no support	no support	no support	no support
19_k141_392029	Bacteria: 0.93	Proteobacteria: 0.93	Alphaproteobacteria: 0.93	Pelagibacterales: 0.93	Pelagibacteraceae: 0.93	no support	no support
19_k141_399029	Bacteria: 1.00	Rhodothermaeota: 0.63	no support	no support	no support	no support	no support
19_k141_495976	Bacteria: 0.72	no support	no support	no support	no support	no support	no support
19_k141_524369	Bacteria: 1.00	Proteobacteria: 0.74	Alphaproteobacteria: 0.74	Rhodobacterales: 0.74	no support	no support	no support
19_k141_57710	Eukaryota: 1.00	NA	Pelagophyceae: 1.00	Pelagomonadales: 1.00	NA	Aureococcus: 1.00	Aureococcus anophagefferens*: 1.00
19_k141_674417	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
19_k141_757971	Bacteria: 1.00	Bacteroidetes: 0.85	Flavobacteriia: 0.85	Flavobacteriales: 0.85	no support	no support	no support
19_k141_860717	Bacteria: 1.00	Planctomycetes: 1.00	no support	no support	no support	no support	no support
19_k141_916416	Bacteria: 0.92	no support	no support	no support	no support	no support	no support
1_k141_110569	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_160480	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support

1_k141_178908	Bacteria: 0.90	Actinobacteria: 0.73	Actinomycetia: 0.73	Micrococcales: 0.73	Microbacteriaceae: 0.73	Microbacterium: 0.73	no support
1_k141_188434	Bacteria: 1.00	Proteobacteria: 0.56	no support	no support	no support	no support	no support
1_k141_188667	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
1_k141_202180	Bacteria: 0.89	Actinobacteria: 0.65	no support	no support	no support	no support	no support
1_k141_208982	Bacteria: 0.99	no support	no support	no support	no support	no support	no support
1_k141_210006	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_226966	Bacteria: 0.89	no support	no support	no support	no support	no support	no support
1_k141_253685	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_306129	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
1_k141_318982	Bacteria: 0.83	Actinobacteria: 0.68	no support	no support	no support	no support	no support
1_k141_328873	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_369165	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.71
1_k141_39786	Bacteria: 1.00	Proteobacteria: 0.94	Gammaproteobacteria: 0.87	no support	no support	no support	no support
1_k141_401067	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
1_k141_415862	Bacteria: 1.00	Proteobacteria: 0.81	Gammaproteobacteria: 0.81	no support	no support	no support	no support
1_k141_440756	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 0.68	no support	no support	no support
1_k141_443738	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_4489	Bacteria: 1.00	Bacteroidetes: 0.99	Flavobacteriia: 0.83	Flavobacteriales: 0.81	Flavobacteriaceae: 0.81	NA	Flavobacteriaceae bacterium TMED208: 0.54
1_k141_455033	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_460960	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_481169	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Pelagibacterales: 1.00	Pelagibacteraceae: 1.00	no support	no support
1_k141_497682	Bacteria: 1.00	Proteobacteria: 0.98	Gammaproteobacteria: 0.98	NA	NA	NA	Gammaproteobacteria bacterium TMED186: 0.76
1_k141_498695	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_49987	Bacteria: 0.95	Bacteroidetes: 0.82	no support	no support	no support	no support	no support
1_k141_516519	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_55460	Bacteria: 1.00	Proteobacteria: 0.77	Gammaproteobacteria: 0.77	NA	NA	NA	SAR86 cluster bacterium: 0.64
1_k141_557398	Bacteria: 1.00	Actinobacteria: 1.00	Acidimicrobiia: 0.70	NA	NA	NA	Acidimicrobiia bacterium: 0.70
1_k141_563953	Bacteria: 1.00	Actinobacteria: 0.99	Actinomycetia: 0.99	Micrococcales: 0.98	Microbacteriaceae: 0.98	Microbacterium: 0.98	no support
1_k141_588384	Bacteria: 0.94	Proteobacteria: 0.94	no support	no support	no support	no support	no support

1_k141_623915	Bacteria: 1.00	Proteobacteria: 1.00	Betaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_646012	Bacteria: 1.00	Rhodothermaeota: 0.87	NA	NA	NA	NA	Rhodothermaeota bacterium MED-G16: 0.55
1_k141_665714	Bacteria: 1.00	Proteobacteria: 0.69	Alphaproteobacteria: 0.69	no support	no support	no support	no support
1_k141_681347	Bacteria: 0.97	no support	no support	no support	no support	no support	no support
1_k141_690695	Bacteria: 0.95	Bacteroidetes: 0.86	no support	no support	no support	no support	no support
1_k141_709904	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_712452	Bacteria: 0.87	Actinobacteria: 0.87	no support	no support	no support	no support	no support
1_k141_714245	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_720556	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
1_k141_726662	Bacteria: 1.00	Actinobacteria: 0.90	no support	no support	no support	no support	no support
1_k141_73404	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_746314	Bacteria: 0.98	Proteobacteria: 0.92	Alphaproteobacteria: 0.92	Sphingomonadales: 0.92	Sphingomonadaceae: 0.91	Sphingomonas: 0.90	no support
1_k141_755953	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.57	no support	no support	no support	no support
1_k141_761992	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Pelagibacterales: 0.94	no support	no support	no support
1_k141_774024	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_783086	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_785748	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_83086	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
1_k141_843551	Bacteria: 0.88	Proteobacteria: 0.88	no support	no support	no support	no support	no support
1_k141_845589	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_8665	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_873093	Bacteria: 0.88	Bacteroidetes: 0.77	no support	no support	no support	no support	no support
1_k141_878817	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
1_k141_878846	Bacteria: 1.00	Proteobacteria: 0.96	Gammaproteobacteria: 0.65	NA	NA	NA	SAR86 cluster bacterium: 0.65
1_k141_895014	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
1_k141_954557	Bacteria: 0.87	no support	no support	no support	no support	no support	no support
1_k141_97664	no support	no support	no support	no support	no support	no support	no support
20_k141_173528	Bacteria: 0.71	Cyanobacteria: 0.71	NA	Synechococcales: 0.71	no support	no support	no support
20_k141_330210	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
20_k141_549731	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.91

20_k141_607934	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.68
20_k141_633824	Bacteria: 1.00	Bacteroidetes: 0.94	Flavobacteriia: 0.94	Flavobacteriales: 0.94	no support	no support	no support
20_k141_754708	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
20_k141_887530	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
21_k141_290075	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
21_k141_356839	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
21_k141_493080	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
21_k141_53160	Eukaryota: 1.00	Chlorophyta: 1.00	Mamiellophyceae: 1.00	Mamiellales: 1.00	Bathycoccaceae: 1.00	Bathycoccus: 1.00	Bathycoccus prasinos*: 1.00
21_k141_554987	Bacteria: 0.93	Cyanobacteria: 0.93	NA	Synechococcales: 0.93	Prochloraceae: 0.93	Prochlorococcus: 0.93	no support
21_k141_566531	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
21_k141_610497	no support	no support	no support	no support	no support	no support	no support
21_k141_627732	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Legionellales: 1.00	NA	NA	Legionellales bacterium: 1.00
21_k141_818310	Bacteria: 1.00	Bacteroidetes: 0.87	no support	no support	no support	no support	no support
21_k141_875102	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium TMED226: 0.82
22_k141_114585	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_12412	Bacteria: 1.00	Proteobacteria: 0.62	Alphaproteobacteria: 0.57	no support	no support	no support	no support
22_k141_126198	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_144797	Bacteria: 1.00	Actinobacteria: 0.92	Actinomycetia: 0.59	no support	no support	no support	no support
22_k141_157481	Bacteria: 0.54	no support	no support	no support	no support	no support	no support
22_k141_161416	Bacteria: 1.00	Actinobacteria: 1.00	Nitriliruptoria: 1.00	Euzebyales: 1.00	Euzebyaceae: 1.00	Euzebya: 1.00	no support
22_k141_172111	Bacteria: 1.00	Proteobacteria: 0.90	Gammaproteobacteria: 0.90	Cellvibrionales: 0.57	Porticoccaceae: 0.57	no support	no support
22_k141_175488	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_21204	Bacteria: 0.96	Proteobacteria: 0.96	Alphaproteobacteria: 0.96	Rhodobacterales: 0.96	Rhodobacteraceae: 0.96	no support	no support
22_k141_218930	Bacteria: 1.00	Bacteroidetes: 0.55	no support	no support	no support	no support	no support
22_k141_262217	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
22_k141_280657	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.79	no support	no support	no support	no support
22_k141_280758	Bacteria: 1.00	Proteobacteria: 0.95	Alphaproteobacteria: 0.94	Rhodobacterales: 0.94	Rhodobacteraceae: 0.94	NA	Rhodobacteraceae bacterium: 0.86
22_k141_286225	Bacteria: 1.00	Bacteroidetes: 0.91	no support	no support	no support	no support	no support
22_k141_291662	Bacteria: 1.00	Bacteroidetes: 0.92	Flavobacteriia: 0.92	Flavobacteriales: 0.92	no support	no support	no support
22_k141_291818	Bacteria: 1.00	Bacteroidetes: 0.61	no support	no support	no support	no support	no support

22_k141_296647	Bacteria: 1.00	Bacteroidetes: 0.92	no support	no support	no support	no support	no support
22_k141_301027	Bacteria: 1.00	Bacteroidetes: 0.77	Flavobacteriia: 0.77	Flavobacteriales: 0.77	no support	no support	no support
22_k141_311775	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Cellvibrionales: 0.89	Porticoccaceae: 0.83	no support	no support
22_k141_319695	Bacteria: 0.94	Bacteroidetes: 0.94	no support	no support	no support	no support	no support
22_k141_332080	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.99	Cellvibrionales: 0.99	Porticoccaceae: 0.56	no support	no support
22_k141_336485	Bacteria: 1.00	Bacteroidetes: 0.54	no support	no support	no support	no support	no support
22_k141_342631	Bacteria: 0.96	Bacteroidetes: 0.96	Flavobacteriia: 0.96	Flavobacteriales: 0.96	no support	no support	no support
22_k141_354190	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	NA	NA	Flavobacteriales bacterium: 1.00
22_k141_356700	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_403570	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
22_k141_422854	Bacteria: 1.00	Bacteroidetes: 0.71	no support	no support	no support	no support	no support
22_k141_44723	Bacteria: 0.98	Proteobacteria: 0.81	Alphaproteobacteria: 0.78	Rhodobacterales: 0.78	Rhodobacteraceae: 0.71	Planktomarina: 0.57	no support
22_k141_467055	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	NA	Flavobacteriaceae bacterium: 1.00
22_k141_470688	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 0.88	Flavobacteriales: 0.88	NA	NA	Flavobacteriales bacterium: 0.88
22_k141_477067	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	no support	no support
22_k141_490279	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_495890	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_510147	Bacteria: 0.99	Bacteroidetes: 0.99	Flavobacteriia: 0.53	Flavobacteriales: 0.53	NA	NA	Flavobacteriales bacterium: 0.53
22_k141_513461	no support	no support	no support	no support	no support	no support	no support
22_k141_515583	Bacteria: 0.85	no support	no support	no support	no support	no support	no support
22_k141_518137	Bacteria: 0.92	no support	no support	no support	no support	no support	no support
22_k141_518376	Bacteria: 1.00	Bacteroidetes: 0.95	Flavobacteriia: 0.95	Flavobacteriales: 0.95	no support	no support	no support
22_k141_537266	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_54015	Bacteria: 1.00	Bacteroidetes: 1.00	no support	no support	no support	no support	no support
22_k141_559517	Bacteria: 1.00	Bacteroidetes: 0.94	NA	NA	NA	NA	Bacteroidetes bacterium: 0.80
22_k141_564034	Bacteria: 1.00	Bacteroidetes: 0.65	Flavobacteriia: 0.61	Flavobacteriales: 0.61	no support	no support	no support
22_k141_572781	Bacteria: 1.00	Proteobacteria: 0.80	Alphaproteobacteria: 0.80	Rhodobacterales: 0.80	Rhodobacteraceae: 0.80	NA	Rhodobacteraceae bacterium: 0.80
22_k141_585394	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	NA	Rhodobacteraceae bacterium: 1.00
22_k141_588206	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_605048	Bacteria: 0.88	no support	no support	no support	no support	no support	no support

22_k141_610153	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.89
22_k141_628785	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_630754	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_638071	Bacteria: 1.00	Planctomycetes: 1.00	Phycisphaerae: 0.62	Phycisphaerales: 0.62	NA	NA	Phycisphaerales bacterium*: 0.62
22_k141_64415	Bacteria: 1.00	Proteobacteria: 0.92	Gammaproteobacteria: 0.92	no support	no support	no support	no support
22_k141_656692	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_65869	Bacteria: 0.82	no support	no support	no support	no support	no support	no support
22_k141_662416	Bacteria: 1.00	Bacteroidetes: 0.77	Flavobacteriia: 0.73	Flavobacteriales: 0.73	Cryomorphaceae: 0.50	NA	Cryomorphaceae bacterium: 0.50
22_k141_670722	Bacteria: 1.00	Bacteroidetes: 0.69	Flavobacteriia: 0.69	Flavobacteriales: 0.69	no support	no support	no support
22_k141_676685	Bacteria: 1.00	Bacteroidetes: 0.70	Flavobacteriia: 0.70	Flavobacteriales: 0.70	no support	no support	no support
22_k141_6928	Bacteria: 1.00	Bacteroidetes: 0.86	Flavobacteriia: 0.86	Flavobacteriales: 0.86	no support	no support	no support
22_k141_697813	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_716741	Bacteria: 1.00	Proteobacteria: 0.80	Gammaproteobacteria: 0.80	no support	no support	no support	no support
22_k141_734013	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodospirillales: 1.00	Rhodospirillaceae: 1.00	no support	no support
22_k141_744365	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Cellvibrionales: 1.00	Porticoccaceae: 0.81	no support	no support
22_k141_765910	Bacteria: 1.00	Bacteroidetes: 0.94	Flavobacteriia: 0.94	Flavobacteriales: 0.94	no support	no support	no support
22_k141_807710	Bacteria: 1.00	Bacteroidetes: 0.73	Flavobacteriia: 0.64	Flavobacteriales: 0.64	NA	NA	Flavobacteriales bacterium: 0.52
22_k141_80919	Bacteria: 1.00	Bacteroidetes: 0.53	Flavobacteriia: 0.53	Flavobacteriales: 0.53	no support	no support	no support
22_k141_82869	Bacteria: 1.00	Bacteroidetes: 0.53	no support	no support	no support	no support	no support
22_k141_84523	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_86807	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
22_k141_89967	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
22_k141_95830	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
23_k141_10153	Bacteria: 1.00	Proteobacteria: 0.84	Gammaproteobacteria: 0.84	no support	no support	no support	no support
23_k141_110579	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	no support	no support	no support
23_k141_112504	Bacteria: 0.92	Proteobacteria: 0.92	Gammaproteobacteria: 0.92	no support	no support	no support	no support
23_k141_148276	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 1.00	Oceanospirillaceae: 1.00	NA	Oceanospirillaceae bacterium: 1.00
23_k141_148952	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 1.00	Pseudoalteromonas: 1.00	no support
23_k141_158766	Bacteria: 1.00	Proteobacteria: 0.54	Betaproteobacteria: 0.54	no support	no support	no support	no support

23_k141_179779	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
23_k141_2370	Bacteria: 1.00	Bacteroidetes: 0.51	no support	no support	no support	no support	no support
23_k141_247444	Bacteria: 0.67	Cyanobacteria: 0.67	NA	Synechococcales: 0.67	no support	no support	no support
23_k141_28421	Bacteria: 0.93	no support	no support	no support	no support	no support	no support
23_k141_293657	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
23_k141_337458	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	SAR86 cluster bacterium: 0.61
23_k141_359393	Bacteria: 1.00	Bacteroidetes: 0.76	Flavobacteriia: 0.76	Flavobacteriales: 0.76	no support	no support	no support
23_k141_374042	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 0.86	Flavobacteriales: 0.86	Crocinitomicaceae: 0.83	no support	no support
23_k141_379353	Bacteria: 1.00	Bacteroidetes: 0.61	Flavobacteriia: 0.61	Flavobacteriales: 0.61	no support	no support	no support
23_k141_399777	Bacteria: 1.00	Bacteroidetes: 0.72	Flavobacteriia: 0.72	Flavobacteriales: 0.72	no support	no support	no support
23_k141_399967	Bacteria: 1.00	Proteobacteria: 0.93	Gammaproteobacteria: 0.93	NA	NA	NA	Gammaproteobacteria bacterium: 0.62
23_k141_406909	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
23_k141_453723	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 1.00	no support	no support	no support
23_k141_457731	Bacteria: 0.96	no support	no support	no support	no support	no support	no support
23_k141_478074	Bacteria: 1.00	Bacteroidetes: 0.83	Flavobacteriia: 0.83	Flavobacteriales: 0.83	no support	no support	no support
23_k141_504782	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium: 0.55
23_k141_537934	Bacteria: 1.00	Proteobacteria: 0.85	Gammaproteobacteria: 0.85	no support	no support	no support	no support
23_k141_538063	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
23_k141_549371	Bacteria: 1.00	Bacteroidetes: 0.81	Flavobacteriia: 0.81	no support	no support	no support	no support
23_k141_574308	Bacteria: 1.00	Actinobacteria: 1.00	Actinomycetia: 1.00	NA	NA	NA	Actinobacteria bacterium: 1.00
23_k141_584539	Bacteria: 1.00	Bacteroidetes: 1.00	no support	no support	no support	no support	no support
23_k141_589697	Bacteria: 1.00	Bacteroidetes: 0.57	Flavobacteriia: 0.57	Flavobacteriales: 0.57	no support	no support	no support
23_k141_605682	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 0.80	Rhodobacteraceae: 0.80	NA	Rhodobacteraceae bacterium: 0.80
23_k141_624169	Bacteria: 1.00	Bacteroidetes: 0.86	Flavobacteriia: 0.86	Flavobacteriales: 0.86	Flavobacteriaceae: 0.85	no support	no support
23_k141_647662	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
23_k141_683534	Bacteria: 1.00	Bacteroidetes: 0.75	Flavobacteriia: 0.71	Flavobacteriales: 0.71	no support	no support	no support
23_k141_716104	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 1.00	Pseudoalteromonas: 1.00	Pseudoalteromonas denitrificans: 0.88
23_k141_730969	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodospirillales: 1.00	Rhodospirillaceae: 1.00	no support	no support
23_k141_745178	Bacteria: 0.93	Planctomycetes: 0.87	NA	NA	NA	NA	Planctomycetes bacterium: 0.87
23_k141_753057	Bacteria: 1.00	Proteobacteria: 0.89	Gammaproteobacteria: 0.89	Cellvibrionales: 0.59	Porticoccaceae: 0.59	no support	no support

23_k141_794766	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Cellvibrionales: 0.96	Porticoccaceae: 0.96	Porticoccus: 0.82	Porticoccus sp.*: 0.82
23_k141_815283	Bacteria: 1.00	Bacteroidetes: 0.67	Flavobacteriia: 0.67	Flavobacteriales: 0.56	NA	NA	Flavobacteriales bacterium: 0.56
24_k141_160455	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.98	Rhodobacterales: 0.98	no support	no support	no support
24_k141_171662	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_191308	Bacteria: 1.00	Proteobacteria: 0.99	Gammaproteobacteria: 0.98	Alteromonadales: 0.95	Pseudoalteromonadaceae: 0.91	Pseudoalteromonas: 0.91	no support
24_k141_192790	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Cellvibrionales: 0.71	Porticoccaceae: 0.71	no support	no support
24_k141_208898	Bacteria: 0.99	Proteobacteria: 0.99	Gammaproteobacteria: 0.99	Alteromonadales: 0.94	Pseudoalteromonadaceae: 0.89	Pseudoalteromonas: 0.89	no support
24_k141_212920	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.76	Oceanospirillales: 0.76	no support	no support	no support
24_k141_223133	Bacteria: 0.96	Proteobacteria: 0.91	Alphaproteobacteria: 0.79	Rhodobacterales: 0.79	NA	NA	Rhodobacterales bacterium HTCC2255: 0.72
24_k141_225047	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 0.89	Pseudoalteromonadaceae: 0.74	Pseudoalteromonas: 0.74	no support
24_k141_229082	Bacteria: 0.87	no support	no support	no support	no support	no support	no support
24_k141_274140	Bacteria: 0.99	Bacteroidetes: 0.88	Flavobacteriia: 0.81	Flavobacteriales: 0.79	Flavobacteriaceae: 0.79	no support	no support
24_k141_286195	Bacteria: 1.00	Bacteroidetes: 0.83	Flavobacteriia: 0.83	Flavobacteriales: 0.83	Flavobacteriaceae: 0.61	no support	no support
24_k141_328670	no support	no support	no support	no support	no support	no support	no support
24_k141_345032	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Crocinitomicaceae: 1.00	Fluviicola: 1.00	Fluviicola sp.: 1.00
24_k141_357622	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 1.00	Rhodobacteraceae: 1.00	NA	Rhodobacteraceae bacterium: 0.94
24_k141_363920	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
24_k141_417363	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support
24_k141_421138	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_428569	Bacteria: 0.96	Bacteroidetes: 0.96	Flavobacteriia: 0.57	Flavobacteriales: 0.57	no support	no support	no support
24_k141_454983	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_46276	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 0.98	no support	no support
24_k141_482584	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
24_k141_494523	Bacteria: 1.00	Bacteroidetes: 0.90	Flavobacteriia: 0.88	Flavobacteriales: 0.88	Flavobacteriaceae: 0.87	no support	no support
24_k141_508814	Bacteria: 1.00	Proteobacteria: 0.99	Gammaproteobacteria: 0.98	Alteromonadales: 0.86	Pseudoalteromonadaceae: 0.76	Pseudoalteromonas: 0.76	no support
24_k141_560063	Bacteria: 1.00	Bacteroidetes: 0.74	Flavobacteriia: 0.74	no support	no support	no support	no support
24_k141_576453	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
24_k141_621418	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
24_k141_625175	Bacteria: 0.93	Proteobacteria: 0.93	Alphaproteobacteria: 0.93	Rhodobacterales: 0.93	Rhodobacteraceae: 0.93	NA	Rhodobacteraceae bacterium: 0.93

24_k141_657725	Bacteria: 0.94	Proteobacteria: 0.94	Alphaproteobacteria: 0.94	Rhodobacterales: 0.89	Rhodobacteraceae: 0.89	NA	Rhodobacteraceae bacterium: 0.73
24_k141_681529	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_695390	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_708757	Bacteria: 1.00	Bacteroidetes: 0.60	Flavobacteriia: 0.60	Flavobacteriales: 0.60	NA	NA	Flavobacteriales bacterium: 0.60
24_k141_715393	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
24_k141_81428	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.98	Rhodobacterales: 0.91	no support	no support	no support
24_k141_89444	Bacteria: 1.00	Lentisphaerae: 0.93	Lentisphaeria: 0.93	Lentisphaerales: 0.93	Lentisphaeraceae: 0.93	NA	Lentisphaeraceae bacterium*: 0.64
2_k141_107175	Bacteria: 1.00	Proteobacteria: 0.95	Betaproteobacteria: 0.95	no support	no support	no support	no support
2_k141_231850	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
2_k141_258309	Bacteria: 0.93	Proteobacteria: 0.76	Gammaproteobacteria: 0.57	no support	no support	no support	no support
2_k141_2996	Bacteria: 1.00	Proteobacteria: 0.59	Gammaproteobacteria: 0.59	no support	no support	no support	no support
2_k141_353594	Bacteria: 0.92	Proteobacteria: 0.75	no support	no support	no support	no support	no support
2_k141_408943	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
2_k141_410077	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
2_k141_410135	Bacteria: 0.91	Actinobacteria: 0.91	no support	no support	no support	no support	no support
2_k141_501545	Bacteria: 0.92	Proteobacteria: 0.70	Gammaproteobacteria: 0.55	no support	no support	no support	no support
2_k141_568663	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
2_k141_600930	Bacteria: 0.63	Bacteroidetes: 0.58	Flavobacteriia: 0.58	Flavobacteriales: 0.58	NA	NA	Flavobacteriales bacterium: 0.58
2_k141_6427	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
2_k141_66582	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
2_k141_66913	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.64	no support	no support	no support	no support
2_k141_975062	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
3_k141_143843	Bacteria: 1.00	Proteobacteria: 0.81	Alphaproteobacteria: 0.79	Hyphomonadales: 0.79	Hyphomonadaceae: 0.79	Hyphomonas: 0.79	no support
3_k141_159802	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
3_k141_198040	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Alteromonadaceae: 1.00	Marinobacter: 1.00	no support
3_k141_250370	Bacteria: 1.00	Candidatus Marinimicrobia: 0.90	NA	NA	NA	NA	Candidatus Marinimicrobia bacterium: 0.90
3_k141_259586	NA	NA	NA	NA	NA	NA	uncultured organism MedDCM-OCT-S04-C478: 0.64
3_k141_292961	Bacteria: 1.00	Candidatus Marinimicrobia: 0.72	NA	NA	NA	NA	Candidatus Marinimicrobia bacterium: 0.72
3_k141_302629	Bacteria: 1.00	Actinobacteria: 0.89	Actinomycetia: 0.89	Micrococcales: 0.89	Microbacteriaceae: 0.89	Leifsonia: 0.89	Leifsonia sp.: 0.89

3_k141_429847	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 0.94	Myxococcales: 0.94	no support	no support	no support
3_k141_45958	Bacteria: 0.92	Proteobacteria: 0.80	Gammaproteobacteria: 0.80	Alteromonadales: 0.74	Alteromonadaceae: 0.74	Marinobacter: 0.74	no support
3_k141_467259	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	Sandaracinaceae: 0.50	NA	Sandaracinaceae bacterium*: 0.50
3_k141_486053	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.74
3_k141_503823	Bacteria: 1.00	Proteobacteria: 0.77	Gammaproteobacteria: 0.77	no support	no support	no support	no support
3_k141_593222	Bacteria: 0.92	Proteobacteria: 0.74	Gammaproteobacteria: 0.74	Alteromonadales: 0.72	Alteromonadaceae: 0.72	Marinobacter: 0.72	no support
3_k141_633769	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodospirillales: 1.00	no support	no support	no support
3_k141_6476	Bacteria: 1.00	Actinobacteria: 0.92	no support	no support	no support	no support	no support
3_k141_663669	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
3_k141_66437	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
3_k141_687471	Bacteria: 1.00	Bacteroidetes: 0.86	Flavobacteriia: 0.86	Flavobacteriales: 0.86	no support	no support	no support
3_k141_714338	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.96	Hyphomicrobiales: 0.96	Aurantimonadaceae: 0.77	no support	no support
3_k141_731125	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
3_k141_784722	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
3_k141_789508	Bacteria: 1.00	Proteobacteria: 0.97	Alphaproteobacteria: 0.96	Hyphomonadales: 0.92	Hyphomonadaceae: 0.92	Hyphomonas: 0.84	no support
3_k141_830730	Bacteria: 1.00	Proteobacteria: 0.73	Alphaproteobacteria: 0.70	Sphingomonadales: 0.67	no support	no support	no support
3_k141_836045	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 0.85	no support	no support	no support
3_k141_836609	Bacteria: 1.00	Proteobacteria: 0.99	Alphaproteobacteria: 0.99	Hyphomicrobiales: 0.94	Aurantimonadaceae: 0.94	no support	no support
3_k141_892661	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
3_k141_936940	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
3_k141_938916	Bacteria: 0.89	Actinobacteria: 0.78	no support	no support	no support	no support	no support
4_k141_147618	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 0.83	Alcanivoracaceae: 0.83	Alcanivorax: 0.83	no support
4_k141_161800	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.90
4_k141_253932	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
4_k141_268362	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.79	NA	NA	NA	Gammaproteobacteria bacterium: 0.79
4_k141_386327	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
4_k141_495312	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
4_k141_513089	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 0.90	no support	no support	no support
4_k141_60506	Bacteria: 0.89	Proteobacteria: 0.89	Alphaproteobacteria: 0.89	no support	no support	no support	no support

4_k141_70679	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
4_k141_711263	Bacteria: 1.00	Proteobacteria: 0.95	Gammaproteobacteria: 0.95	Oceanospirillales: 0.95	Oleiphilaceae: 0.95	Oleiphilus: 0.95	no support
4_k141_855023	Bacteria: 1.00	Actinobacteria: 0.75	no support	no support	no support	no support	no support
4_k141_862362	Bacteria: 1.00	Proteobacteria: 1.00	NA	NA	NA	NA	Proteobacteria bacterium: 1.00
4_k141_866455	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
4_k141_89201	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
4_k141_950971	Bacteria: 0.87	Proteobacteria: 0.81	Gammaproteobacteria: 0.81	Alteromonadales: 0.65	Alteromonadaceae: 0.65	Marinobacter: 0.65	no support
4_k141_981843	Bacteria: 0.96	Proteobacteria: 0.96	Alphaproteobacteria: 0.96	Hyphomicrobiales: 0.96	no support	no support	no support
5_k141_1041368	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
5_k141_1052317	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
5_k141_1054409	Bacteria: 1.00	Proteobacteria: 0.96	Gammaproteobacteria: 0.96	no support	no support	no support	no support
5_k141_1082158	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Vibrionales: 1.00	no support	no support	no support
5_k141_1185582	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Pseudomonadales: 1.00	Pseudomonadaceae: 0.91	Pseudomonas: 0.91	no support
5_k141_1226140	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 1.00	Rhodobiaceae: 1.00	NA	Rhodobiaceae bacterium*: 1.00
5_k141_141154	Bacteria: 1.00	Candidatus Marinimicrobia: 0.92	NA	NA	NA	NA	Candidatus Marinimicrobia bacterium: 0.92
5_k141_156088	Bacteria: 1.00	Rhodothermaeota: 0.65	no support	no support	no support	no support	no support
5_k141_164137	Bacteria: 0.89	Proteobacteria: 0.89	Alphaproteobacteria: 0.89	Maricaulales: 0.89	Maricaulaceae: 0.89	no support	no support
5_k141_181738	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
5_k141_251289	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomonadales: 1.00	Hyphomonadaceae*: 1.00	no support	no support
5_k141_25794	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Maricaulales: 0.93	Maricaulaceae: 0.93	Maricaulis: 0.82	no support
5_k141_409288	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
5_k141_442065	Bacteria: 1.00	Proteobacteria: 0.95	Gammaproteobacteria: 0.95	NA	NA	NA	Gammaproteobacteria bacterium: 0.95
5_k141_443192	Bacteria: 1.00	Proteobacteria: 0.93	Gammaproteobacteria: 0.93	no support	no support	no support	no support
5_k141_503021	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
5_k141_68513	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Pseudomonadales: 1.00	Pseudomonadaceae: 1.00	Pseudomonas: 0.65	no support
5_k141_780267	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 0.98	Flavobacteriales: 0.98	Flavobacteriaceae: 0.87	NA	Flavobacteriaceae bacterium: 0.87
5_k141_817973	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodospirillales: 1.00	no support	no support	no support
5_k141_833048	Bacteria: 1.00	Chloroflexi: 1.00	no support	no support	no support	no support	no support
5_k141_876822	Bacteria: 0.95	no support	no support	no support	no support	no support	no support
5_k141_918191	Bacteria: 1.00	Actinobacteria: 1.00	Actinomycetia: 1.00	Propionibacteriales: 0.81	Nocardioidaceae: 0.81	Pimelobacter: 0.72	Pimelobacter sp.*: 0.72

5_k141_924327	Bacteria: 1.00	NA	NA	NA	NA	NA	bacterium TMED46: 0.70
5_k141_928548	Bacteria: 0.72	Proteobacteria: 0.72	Gammaproteobacteria: 0.62	Vibrionales: 0.62	no support	no support	no support
5_k141_963588	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.96	NA	NA	NA	Gammaproteobacteria bacterium TMED257: 0.96
5_k141_990402	Bacteria: 0.96	Actinobacteria: 0.90	Actinomycetia: 0.90	no support	no support	no support	no support
6_k141_104421	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.64	no support	no support	no support	no support
6_k141_111942	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
6_k141_159118	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
6_k141_193528	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	no support	no support	no support
6_k141_196731	Bacteria: 0.96	Proteobacteria: 0.96	Alphaproteobacteria: 0.96	Rhodobacterales: 0.90	Rhodobacteraceae: 0.90	Ahrensia: 0.90	Ahrensia sp.: 0.90
6_k141_238268	Bacteria: 0.95	Proteobacteria: 0.95	Alphaproteobacteria: 0.81	Rhodobacterales: 0.55	Rhodobacteraceae: 0.55	Ahrensia: 0.55	Ahrensia sp.: 0.55
6_k141_266046	Bacteria: 1.00	Proteobacteria: 1.00	Deltaproteobacteria: 1.00	Myxococcales: 1.00	Sandaracinaceae: 0.70	NA	Sandaracinaceae bacterium*: 0.70
6_k141_279575	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.95	Rhodobacterales: 0.93	Rhodobacteraceae: 0.93	Ahrensia: 0.93	Ahrensia sp.: 0.93
6_k141_546	no support	no support	no support	no support	no support	no support	no support
6_k141_68143	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 0.99	Rhodospirillales: 0.90	no support	no support	no support
6_k141_8233	Bacteria: 1.00	Actinobacteria: 1.00	Nitriliruptoria: 1.00	Euzebyales: 1.00	Euzebyaceae: 1.00	Euzebya: 1.00	no support
6_k141_8774	Bacteria: 1.00	Proteobacteria: 0.96	Alphaproteobacteria: 0.96	Hyphomonadales: 0.94	Hyphomonadaceae*: 0.94	no support	no support
6_k141_97482	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomonadales: 1.00	Hyphomonadaceae*: 1.00	no support	no support
7_k141_141434	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
7_k141_155479	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 0.98	Alcanivoracaceae: 0.98	Alcanivorax: 0.69	no support
7_k141_203753	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Vibrionales: 1.00	no support	no support	no support
7_k141_234494	Bacteria: 1.00	Proteobacteria: 0.98	Gammaproteobacteria: 0.98	Oceanospirillales: 0.81	Alcanivoracaceae: 0.81	Alcanivorax: 0.65	no support
7_k141_243606	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
7_k141_291656	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 0.99	Alcanivoracaceae: 0.99	Alcanivorax: 0.74	no support
7_k141_334343	Bacteria: 0.99	Proteobacteria: 0.99	Alphaproteobacteria: 0.98	Rhodobacterales: 0.94	Rhodobacteraceae: 0.91	Sulfitobacter: 0.79	no support
7_k141_417033	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
7_k141_490004	Bacteria: 1.00	Proteobacteria: 0.88	Alphaproteobacteria: 0.88	Pelagibacterales: 0.88	NA	NA	Pelagibacterales bacterium SAG-MED32: 0.88
7_k141_496110	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Hyphomicrobiales: 1.00	Rhodobiaceae: 1.00	NA	Rhodobiaceae bacterium*: 1.00
7_k141_509836	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Oceanospirillales: 1.00	Alcanivoracaceae: 1.00	Alcanivorax: 0.61	no support
7_k141_548834	Bacteria: 1.00	Proteobacteria: 0.92	Gammaproteobacteria: 0.92	Oceanospirillales: 0.92	Alcanivoracaceae: 0.92	Alcanivorax: 0.84	no support

7_k141_586997	Bacteria: 1.00	Proteobacteria: 0.98	Alphaproteobacteria: 0.98	Rhodobacterales: 0.96	Rhodobacteraceae: 0.96	Sulfitobacter: 0.84	no support
7_k141_591226	Bacteria: 0.99	Proteobacteria: 0.99	Gammaproteobacteria: 0.99	Oceanospirillales: 0.92	Alcanivoracaceae: 0.92	Alcanivorax: 0.88	no support
7_k141_627629	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.94	Alteromonadales: 0.84	Alteromonadaceae: 0.84	Marinobacter: 0.84	no support
7_k141_745429	Bacteria: 1.00	Proteobacteria: 0.98	Alphaproteobacteria: 0.98	Rhodobacterales: 0.96	Rhodobacteraceae: 0.95	Sulfitobacter: 0.83	no support
7_k141_84381	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	Rhodobacterales: 0.92	Rhodobacteraceae: 0.91	Sulfitobacter: 0.76	no support
9_k141_1021009	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_1039522	Bacteria: 1.00	Proteobacteria: 0.94	Gammaproteobacteria: 0.94	Alteromonadales: 0.61	no support	no support	no support
9_k141_190102	Bacteria: 1.00	Candidatus Marinimicrobia: 0.70	NA	NA	NA	NA	Candidatus Marinimicrobia bacterium: 0.70
9_k141_208497	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.98	Oceanospirillales: 0.95	Alcanivoracaceae: 0.95	Alcanivorax: 0.72	no support
9_k141_284466	Bacteria: 1.00	no support	no support	no support	no support	no support	no support
9_k141_308801	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 0.91	Legionellales: 0.78	NA	NA	Legionellales bacterium: 0.78
9_k141_328917	Bacteria: 1.00	Chloroflexi: 1.00	NA	NA	NA	NA	Chloroflexi bacterium: 0.58
9_k141_352615	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 1.00	Pseudoalteromonas: 1.00	no support
9_k141_360001	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 0.70	Pseudoalteromonas: 0.61	no support
9_k141_371773	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.66
9_k141_379141	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
9_k141_425602	Bacteria: 1.00	Proteobacteria: 1.00	no support	no support	no support	no support	no support
9_k141_461938	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 0.93
9_k141_473766	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 0.74	no support	no support
9_k141_510210	Bacteria: 0.68	Actinobacteria: 0.68	no support	no support	no support	no support	no support
9_k141_535753	Bacteria: 1.00	Chloroflexi: 1.00	no support	no support	no support	no support	no support
9_k141_579011	Bacteria: 1.00	Proteobacteria: 0.62	no support	no support	no support	no support	no support
9_k141_580773	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_639864	Bacteria: 1.00	Actinobacteria: 0.99	Actinomycetia: 0.99	Propionibacteriales: 0.96	Nocardioidaceae: 0.96	Nocardioides: 0.96	no support
9_k141_685811	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_716922	Bacteria: 1.00	Proteobacteria: 1.00	Alphaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_726320	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	NA	NA	NA	Gammaproteobacteria bacterium: 1.00
9_k141_733915	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	Pseudoalteromonadaceae: 1.00	no support	no support
9_k141_749334	Bacteria: 1.00	Proteobacteria: 0.84	Gammaproteobacteria: 0.84	no support	no support	no support	no support

9_k141_796666	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_804910	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 0.57	Pseudoalteromonadaceae: 0.57	Pseudoalteromonas: 0.57	no support
9_k141_823203	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support
9_k141_82633	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	Tenacibaculum: 0.63	Tenacibaculum soleae: 0.63
9_k141_833100	Bacteria: 1.00	Proteobacteria: 0.89	Gammaproteobacteria: 0.89	no support	no support	no support	no support
9_k141_84122	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_847541	Bacteria: 1.00	Proteobacteria: 0.91	Gammaproteobacteria: 0.91	no support	no support	no support	no support
9_k141_868470	Bacteria: 0.91	no support	no support	no support	no support	no support	no support
9_k141_870089	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	no support	no support	no support	no support
9_k141_919001	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Legionellales: 0.95	NA	NA	Legionellales bacterium: 0.95
9_k141_921415	Bacteria: 1.00	Proteobacteria: 1.00	Gammaproteobacteria: 1.00	Alteromonadales: 1.00	no support	no support	no support
9_k141_983411	Bacteria: 1.00	Bacteroidetes: 1.00	Flavobacteriia: 1.00	Flavobacteriales: 1.00	Flavobacteriaceae: 1.00	no support	no support

Appendix 6

Table S3: Showing known BGC pathway identified from the SAO non-redundant KS contigs and their percentage similarity

Contigs/region	BGC type	From location	To location	Most similar known cluster	Type	Similarity
10_k141_580458/Region 7.1	arylpolyene	6,400	49,981	APE Vf	Other	30%
10_k141_295956/Region 8.1	arylpolyene	1	1,814			
10_k141_849482/Region 13.1	arylpolyene	1	1,955	flexirubin	Polyketide	5%
10_k141_168623/Region 14.1	arylpolyene	1	3,184	flexirubin	Polyketide	11%
10_k141_569960/Region 19.1	arylpolyene	1	3,951	APE Vf	Other	15%
11_k141_956206/Region 25.1	arylpolyene	1	8,578	flexirubin	Polyketide	25%
11_k141_569222/Region 31.1	NRPS,T1PKS	1	36,118			
11_k141_203925/Region 32.1	PUFA,hglE-KS	60,376	106,022	eicosapentaenoic acid	Other	66%
11_k141_389331/Region 45.1	T1PKS,NRPS	1	33,699			
13_k141_105401/Region 72.1	arylpolyene	1	3,024	flexirubin	Polyketide	13%
14_k141_1012099/Region 95.1	arylpolyene	1	4,671	flexirubin	Polyketide	8%
14_k141_223467/Region 104.1	arylpolyene	1	6,269	flexirubin	Polyketide	16%
14_k141_897973/Region 107.1	arylpolyene	1	2,186	APE Ec	Other	15%
14_k141_753018/Region 114.1	ladderane	1	2,582			
14_k141_1029109/Region 116.1	arylpolyene	1	3,314	APE Vf	Other	15%
15_k141_526901/Region 128.1	arylpolyene	1	3,163			
16_k141_697129/Region 134.1	T1PKS	1	14,790			
16_k141_920012/Region 135.1	arylpolyene	12,294	55,858	APE Vf	Other	40%
16_k141_392261/Region 136.1	arylpolyene	1	2,080			

16_k141_982964/Region 142.1	arylpolyene	1	8,748	APE Vf	Other	15%
17_k141_565457/Region 155.1	arylpolyene	187,459	231,040	APE Vf	Other	30%
17_k141_358034/Region 156.1	arylpolyene	7,209	50,772	APE Vf	Other	35%
17_k141_798470/Region 159.1	arylpolyene	1	3,575	APE Ec	Other	15%
18_k141_554657/Region 176.1	arylpolyene	1	3,204			
1_k141_73404/Region 194.1	arylpolyene	1	4,034			
1_k141_873093/Region 202.1	terpene	166,809	187,642	carotenoid	Terpene	42%
1_k141_873093/Region 202.2	T1PKS,NRPS	681,163	731,905			
1_k141_328873/Region 230.1	arylpolyene	1	1,769	APE Vf	Other	10%
1_k141_845589/Region 250.1	T1PKS	1	9,986			
22_k141_605048/Region 269.1	arylpolyene,resorcinol	162,351	208,418	flexirubin	Polyketide	13%
22_k141_336485/Region 276.1	arylpolyene,resorcinol	1	8,641	flexirubin	Polyketide	16%
22_k141_65869/Region 284.1	arylpolyene	371,576	417,912	flexirubin	Other	36%
22_k141_422854/Region 288.1	arylpolyene	1	4,367	flexirubin	Polyketide	8%
22_k141_470688/Region 311.1	arylpolyene	1	2,476	flexirubin	Polyketide	5%
22_k141_54015/Region 312.1	NRPS,T1PKS	1	12,344			
22_k141_559517/Region 325.1	NRPS	1	11,289			
22_k141_80919/Region 332.1	T3PKS	1	8,831			
22_k141_354190/Region 335.1	arylpolyene	1	2,460			
23_k141_716104/Region 348.1	arylpolyene	1	3,595	APE Vf	Other	15%
23_k141_549371/Region 349.1	T3PKS	1	2,036			
23_k141_453723/Region 359.1	arylpolyene	1	1,487			
24_k141_681529/Region 375.1	arylpolyene	1	3,368	APE Vf	Other	10%
24_k141_225047/Region 378.1	arylpolyene	2,427	24,274	APE Vf	Other	10%
24_k141_208898/Region 380.1	hglE-KS,T1PKS,PUFA	1	37,171	eicosapentaenoic acid-like compound	Other	18%
24_k141_560063/Region 386.1	T3PKS	1	2,541			
24_k141_708757/Region 387.1	arylpolyene	1	2,853	flexirubin	Polyketide	8%
24_k141_212920/Region 391.1	arylpolyene	1	2,484	APE Vf	Other	10%
24_k141_46276/Region 392.1	arylpolyene,resorcinol	1	23,289	flexirubin	Polyketide	44%
24_k141_345032/Region 394.1	arylpolyene	1	3,352	flexirubin	Polyketide	13%

24_k141_695390/Region 398.1	arylpolyene	1	10,359	APE Vf	Other	25%
24_k141_417363/Region 399.1	arylpolyene	1	8,227	flexirubin	Polyketide	25%
24_k141_328670/Region 400.1	arylpolyene	1	8,979	APE Vf	Other	15%
2_k141_501545/Region 412.1	resorcinol	73,959	115,107			
2_k141_353594/Region 413.1	hglE-KS,T1PKS	74,570	128,062	paulomycin	Other	3%
3_k141_633769/Region 427.1	arylpolyene	1	9,764			
3_k141_467259/Region 430.1	arylpolyene	1	4,134			
3_k141_663669/Region 434.1	arylpolyene	1	11,616			
3_k141_789508/Region 442.1	T1PKS,hglE-KS	213,391	264,848			
3_k141_892661/Region 449.1	ladderane	1	2,597			
5_k141_251289/Region 472.1	T1PKS	1	9,903			
5_k141_68513/Region 476.1	arylpolyene	1	2,041	APE Vf	Other	10%
5_k141_1185582/Region 486.1	arylpolyene	1	3,093			
6_k141_193528/Region 497.1	arylpolyene	1	3,876			
6_k141_159118/Region 499.1	arylpolyene	1	6,746	aryl polyenes	Other	22%
6_k141_266046/Region 500.1	hglE-KS	1	8,499			
6_k141_68143/Region 501.1	ectoine	89,601	99,999	ectoine	Other	80%
6_k141_97482/Region 503.1	hglE-KS,T1PKS	1	17,833			
6_k141_111942/Region 506.1	ladderane	1	6,760			
7_k141_627629/Region 513.1	T1PKS,NRPS	1	26,776			
7_k141_84381/Region 519.1	T1PKS	17,712	64,164			
7_k141_203753/Region 521.1	arylpolyene	1	1,901	APE Vf	Other	10%
7_k141_334343/Region 522.1	hserlactone	49,579	70,301			
9_k141_82633/Region 526.1	arylpolyene	1	4,367			
9_k141_84122/Region 529.1	arylpolyene	1	2,425	APE Vf	Other	10%
9_k141_733915/Region 540.1	arylpolyene	1	4,667	APE Vf	Other	15%
9_k141_868470/Region 544.1	T1PKS,NRPS	1	15,886			
9_k141_473766/Region 549.1	arylpolyene	1	2,051	APE Vf	Other	10%
9_k141_983411/Region 557.1	arylpolyene	1	17,309	flexirubin	Polyketide	33%
9_k141_823203/Region 558.1	T3PKS	1	7,838			

Appendix 7

Table S4: Showing known BGC pathway identified from the TARA non-redundant KS contigs and their percentage similarity

Contigs/region	BGC type	From location	To location	Most similar known cluster	Type	Similarity
TARA_ANE-k99_2692542/Region 22.1	arylpolyene	1	26,262	flexirubin	Polyketide	44%
TARA_ANE-k99_2888596/Region 26.1	NRPS	1	26,980			
TARA_ANE-k99_3002501/Region 27.1	T1PKS	1	22,133			
TARA_ANE-k99_3031332/Region 28.1	T1PKS	1	10,359			
TARA_ANE-k99_3872539/Region 35.1	arylpolyene	1	12,229			
TARA_ANE-k99_3964636/Region 37.1	T1PKS	1,931	39,463			
TARA_ANE-k99_4257134/Region 40.1	T1PKS	2,810	38,546			
TARA_ANE-k99_6730098/Region 59.1	arylpolyene	1	20,706			
TARA_ANE-k99_9620988/Region 80.1	T1PKS	1	12,545			
TARA_ANE-k99_11352568/Region 97.1	T1PKS	1	10,274			
TARA_ASE-k99_973679/Region 111.1	hglE-KS,T1PKS	1	10,522			
TARA_ASE-k99_3319788/Region 123.1	NRPS,T1PKS	1	28,095			
TARA_ASE-k99_3539774/Region 125.1	arylpolyene	1	28,171	APE Vf	Other	25%
TARA_ASE-k99_4740099/Region 129.1	PUFA,T1PKS	1	10,518			
TARA_ASE-k99_6939433/Region 140.1	hglE-KS	1	10,751			
TARA_ASW-k99_771653/Region 145.1	T1PKS	1	12,766			
TARA_ASW-k99_1259523/Region 148.1	terpene	8,457	29,605			
TARA_ASW-k99_1596744/Region 151.1	ladderane	1	22,194			
TARA_ASW-k99_1853689Region 153.1	T1PKS	1	25,277			

TARA_ASW-k99_2069705/Region 157.1	T1PKS	1	25,955	berninamycin A	RiPP	26%
TARA_ASW-k99_2429447/Region 161.1	T3PKS	1	27,216			
TARA_ASW-k99_2555030/Region 162.1	arylpolyene	1	27,585			
TARA_ASW-k99_4841156/Region 180.1	arylpolyene	1,046	42,329			
TARA_ASW-k99_5140954/Region 183.1	T1PKS	1	16,672			
TARA_ASW-k99_5827606/Region 187.1	NRPS	1	11,047			
TARA_ASW-k99_6277563/Region 190.1	T1PKS	1	10,748			
TARA_ASW-k99_6302083/Region 191.1	T3PKS	1	19,574			
TARA_ASW-k99_7931438/Region 200.1	T1PKS	20,313	61,757			
TARA_ASW-k99_9390589/Region 216.1	hglE-KS	1	22,629			
TARA_ION-k99_2637854/Region 231.1	terpene	86,893	108,583			
TARA_ION-k99_2637854/Region 231.2	redox-cofactor	119,596	141,803	azinomycin B	NRP + Polyketide	4%
TARA_ION-k99_4336073/Region 239.1	RiPP-like	73,949	84,737			
TARA_ION-k99_4336073/Region 239.2	terpene	270,808	292,486			
TARA_ION-k99_4336073/Region 239.3	redox-cofactor	302,157	324,364	azinomycin B	NRP + Polyketide	4%
TARA_ION-k99_4828864/Region 244.1	T1PKS	1	28,582			
TARA_ION-k99_7062432/Region 261.1	arylpolyene	1	14,390			
TARA_ION-k99_8207243/Region 265.1	T1PKS	9,731	50,555			
TARA_ION-k99_8530172/Region 267.1	T1PKS	1	15,090			
TARA_ION-k99_8926338/Region 268.1	betalactone	3,627	27,955			
TARA_ION-k99_11168747/Region 278.1	NRPS,T1PKS	1	70,319	oxazolepoxidomycin A	Polyketide	13%
TARA_ION-k99_12523780/Region 284.1	hglE-KS	1	23,027			
TARA_ION-k99_14455400/Region 289.1	T1PKS	1	11,187			
TARA_IOS-k99_1812130/Region 301.1	T1PKS	1	11,145			
TARA_IOS-k99_3481211/Region 311.1	hglE-KS	1	28,651			
TARA_IOS-k99_5040029/Region 325.1	RiPP-like	29,169	40,065			
TARA_IOS-k99_5585544/Region 331.1	T1PKS	1	21,832			
TARA_IOS-k99_7245556/Region 343.1	arylpolyene	176	35,541			
TARA_IOS-k99_9531409/Region 354.1	T1PKS,NRPS	1	19,050			
TARA_IOS-k99_11608981/Region 366.1	T1PKS	1	11,032			
TARA_IOS-k99_12211857/Region 369.1	T1PKS	1	11,231			

TARA_IOS-k99_13206630/Region 372.1	arylpolyene	1	13,823			
TARA_MED-k99_829504/Region 404.1	T1PKS	1	12,785			
TARA_MED-k99_1413773/Region 421.1	arylpolyene	63,351	104,526	lactonamycin	Polyketide	3%
TARA_MED-k99_3702765/Region 482.1	arylpolyene	1	10,301	APE Vf	Other	10%
TARA_MED-k99_4407384/Region 493.1	hserlactone	36,722	57,456			
TARA_MED-k99_4997507/Region 505.1	arylpolyene	1	26,464			
TARA_MED-k99_5022155/Region 508.1	arylpolyene	1	11,550			
TARA_MED-k99_5205461/Region 511.1	RiPP-like	76,533	87,387			
TARA_MED-k99_5299240/Region 515.1	arylpolyene	1	17,684			
TARA_MED-k99_5879508/Region 525.1	T1PKS	1	33,433			
TARA_MED-k99_6432729/Region 533.1	T1PKS	38,099	64,987			
TARA_MED-k99_6609820/Region 536.1	hglE-KS	1	23,583			
TARA_MED-k99_7993022/Region 559.1	T1PKS	1	33,921			
TARA_MED-k99_9189128/Region 583.1	terpene	150,907	171,839			
TARA_MED-k99_9456448/Region 592.1	NRPS,T1PKS	1	15,666			
TARA_PON-k99_213996/Region 598.1	hglE-KS	1	10,050			
TARA_PON-k99_1011198/Region 601.1	NRPS	1	17,478			
TARA_PON-k99_1175079/Region 603.1	hglE-KS	9,261	45,539			
TARA_PON-k99_1231741/Region 604.1	NRPS	1	18,509	mycobactin	NRP + Polyketide	60%
TARA_PON-k99_1254868/Region 605.1	hglE-KS	1	27,700			
TARA_PON-k99_1570427/Region 609.1	arylpolyene	4,925	45,977			
TARA_PON-k99_2437080/Region 619.1	NRPS,T1PKS	1	15,286			
TARA_PON-k99_4118086/Region 632.1	T1PKS,NRPS	1	39,156			
TARA_PON-k99_4235523/Region 633.1	T1PKS	1	24,469			
TARA_PON-k99_4270088/Region 634.1	T1PKS	1	14,505			
TARA_PON-k99_5387792/Region 647.1	T1PKS	1	15,080			
TARA_PON-k99_5593556/Region 649.1	arylpolyene	32,011	73,213			
TARA_PON-k99_5958106Region 653.1	arylpolyene,ladderane	1	22,722			
TARA_PON-k99_6842487/Region 660.1	arylpolyene	1	16,610			
TARA_PON-k99_7218854/Region 665.1	T1PKS	1	18,114	glycopeptidolipid	NRP	16%
TARA_PON-k99_7763728/Region 673.1	T1PKS	1	37,613			

TARA_PON-k99_8804150/Region 683.1	T1PKS	1	10,311	1-heptadecene	Polyketide:Modular type I	100%
TARA_PON-k99_9329632/Region 689.1	T1PKS	1	10,155			
TARA_PON-k99_9956518/Region 693.1	arylpolyene	1	34,284			
TARA_PON-k99_10334859/Region 697.1	RiPP-like	28,923	39,804			
TARA_PON-k99_11831479/Region 710.1	T1PKS,NRPS	1	62,748	chondrochloren A	NRP + Polyketide:Modular type I	22%
TARA_PON-k99_12794088/Region 718.1	terpene	7,532	28,335			
TARA_PON-k99_13894771/Region 725.1	arylpolyene	1	13,082	aryl polyenes	Other	22%
TARA_PON-k99_13954325/Region 726.1	T1PKS	1	18,322			
TARA_PON-k99_15561696/Region 741.1	T1PKS,hglE-KS	1	10,188	TP-1161	RiPP:Thiopeptide	12%
TARA_PON-k99_15758939/Region 743.1	T1PKS	1	22,183			
TARA_PSE-k99_775059/Region 753.1	hglE-KS	1	10,575			
TARA_PSE-k99_2030147/Region 765.1	NRPS	1	34,704	didemnin B / nordidemnin B / didemnin X / didemnin Y	Polyketide + NRP:Cyclic depsipeptide	45%
TARA_PSE-k99_2180720/Region 770.1	terpene	90,920	113,718			
TARA_PSE-k99_2180720/Region 770.2	proteusin	125,119	145,490			
TARA_PSE-k99_2493361/Region 775.1	T1PKS	105,550	153,184			
TARA_PSE-k99_2550615/Region 776.1	arylpolyene	1	11,246	APE Vf	Other	30%
TARA_PSE-k99_2558412/Region 778.1	T1PKS	7,343	53,744			
TARA_PSE-k99_6484808/Region 814.1	arylpolyene	1	21,719			
TARA_PSE-k99_9211891/Region 827.1	T1PKS	1	33,484			
TARA_PSE-k99_9266729/Region 828.1	arylpolyene	1	17,393	aryl polyenes	Other	16%
TARA_PSE-k99_9404822/Region 830.1	ladderane	1	13,789			
TARA_PSE-k99_10001996/Region 837.1	T1PKS	1	27,029			
TARA_PSE-k99_11445862/Region 851.1	arylpolyene	1	10,735			
TARA_PSE-k99_11687429/Region 852.1	T1PKS	1	32,146	vazabitide A	NRP	4%
TARA_PSE-k99_11941962/Region 855.1	T3PKS	34,625	75,677	ET-743	NRP:Beta-lactam	8%
TARA_PSE-k99_13046137/Region 864.1	arylpolyene	13,524	54,234	APE Vf	Other	50%
TARA_PSE-k99_13059475/Region 865.1	T1PKS	49,169	96,719	s56-p1	NRP	3%

TARA_PSE-k99_13248480/Region 869.1	NRPS	1	10,758			
TARA_PSE-k99_13335927/Region 871.1	arylpolyene	1	29,093	APE Vf	Other	40%
TARA_PSE-k99_14423699/Region 878.1	terpene	120,431	141,267	carotenoid	Terpene	42%
TARA_PSE-k99_14677836/Region 882.1	proteusin	1	12,587			
TARA_PSE-k99_14786034/Region 884.1	butyrolactone	7,402	20,827			
TARA_PSE-k99_16004998/Region 895.1	arylpolyene,resorcinol	1	38,120	flexirubin	Polyketide	8%
TARA_PSE-k99_16767245/Region 897.1	T1PKS	1	11,881			
TARA_PSE-k99_17406297/Region 904.1	arylpolyene	1	14,169	APE Ec	Other	15%
TARA_PSE-k99_18921703/Region 922.1	NRPS	1	11,723	didemnin B / nordidemnin B / didemnin X / didemnin Y	Polyketide + NRP:Cyclic depsipeptide	36%
TARA_PSE-k99_19871325/Region 934.1	arylpolyene	1	10,561	flexirubin	Polyketide	11%
TARA_PSE-k99_22430428/Region 955.1	NRPS,T1PKS	1	49,749	puwainaphycin F / minutissamide A	NRP	33%
TARA_PSE-k99_22441888/Region 956.1	NRPS	1,382	44,736	thalassospiramide A	NRP:Lipopeptide + Polyketide:Trans-AT type I	100%
TARA_PSE-k99_22651754/Region 957.1	NRPS,T1PKS	1	16,993	micacocidin	NRP:NRP siderophore + Polyketide:Modular type I + Polyketide:Iterative type I	20%
TARA_PSE-k99_22688052/Region 959.1	T1PKS	1	15,183			
TARA_PSE-k99_22709537/Region 960.1	T1PKS	1	20,351			
TARA_PSE-k99_24883498/Region 971.1	NRPS,T1PKS	1	15,348			
TARA_PSE-k99_25439984/Region 980.1	arylpolyene	47,604	91,157	APE Vf	Other	50%
TARA_PSE-k99_25631123/Region 984.1	arylpolyene	1	28,676	flexirubin	Polyketide	19%
TARA_PSE-k99_26181230/Region 990.1	terpene	16,324	32,186			
TARA_PSE-k99_26535095/Region 993.1	T1PKS	1	21,434			
TARA_PSE-k99_27046995/Region 995.1	arylpolyene	1	21,038	APE Ec	Other	10%
TARA_PSE-k99_28122088/Region 1003.1	T1PKS	1	30,344			
TARA_PSW-k99_58599/Region 1005.1	ladderane	1	38,932	TP-1161	RiPP:Thiopeptide	16%
TARA_PSW-k99_236712/Region 1008.1	T3PKS	273,400	314,494			
TARA_PSW-k99_236712/Region 1008.2	terpene	500,632	521,468	carotenoid	Terpene	57%

TARA_PSW-k99_934618/Region 1013.1	arylpolyene	1	11,869			
TARA_PSW-k99_937050/Region 1014.1	hglE-KS,T1PKS	16,802	69,586	eicoseicosapentaenoic acid	Other	15%
TARA_PSW-k99_2012911/Region 1027.1	NRPS,T1PKS	9,559	61,230			
TARA_PSW-k99_2415730/Region 1033.1	T1PKS	1	46,107			
TARA_PSW-k99_2714889/Region 1036.1	hserlactone	27,485	47,952			
TARA_PSW-k99_3002637/Region 1040.1	T1PKS,NRPS	1	46,731			
TARA_PSW-k99_3826268/Region 1054.1	T1PKS	1,041	33,110	funisamine	Polyketide	5%
TARA_PSW-k99_3907640/Region 1055.1	terpene	8,240	29,079	carotenoid	Terpene	28%
TARA_PSW-k99_3927251/Region 1057.1	T1PKS	11,745	52,824			
TARA_PSW-k99_4475253/Region 1060.1	T1PKS	1	43,645			
TARA_PSW-k99_5115786/Region 1071.1	arylpolyene	1	14,993			
TARA_PSW-k99_5373462/Region 1077.1	T1PKS	1	30,218			
TARA_PSW-k99_5716256/Region 1087.1	T1PKS	1	12,218			
TARA_PSW-k99_5945439/Region 1090.1	arylpolyene	1	19,978			
TARA_PSW-k99_6084385/Region 1092.1	betalactone	12,901	40,420	fengycin	NRP	13%
TARA_PSW-k99_6638563/Region 1102.1	terpene	4,785	25,789			
TARA_PSW-k99_6887639/Region 1113.1	hglE-KS,T1PKS	1	35,886	eicoseicosapentaenoic acid	Other	21%
TARA_PSW-k99_6997347/Region 1114.1	arylpolyene	1	15,874			
TARA_PSW-k99_7461075/Region 1121.1	NRPS	224	46,397	bacillaene	Polyketide + NRP	28%
TARA_PSW-k99_7461075/Region 1121.2	NRPS,T1PKS	55,123	99,761			
TARA_PSW-k99_7488810/Region 1122.1	terpene	49,071	66,284			
TARA_PSW-k99_7524880/Region 1123.1	T1PKS	1	31,670	phenalamide A2	NRP + Polyketide	25%
TARA_PSW-k99_7785607/Region 1127.1	arylpolyene	1	23,816			
TARA_PSW-k99_7966335/Region 1130.1	T1PKS	1	30,931			
TARA_PSW-k99_8083719/Region 1133.1	arylpolyene,RiPP-like	1	28,165			
TARA_PSW-k99_8235592/Region 1135.1	terpene	1	16,935			
TARA_PSW-k99_8428153/Region 1139.1	T1PKS	57,245	104,117			
TARA_PSW-k99_8942982/Region 1145.1	arylpolyene	117,114	158,352			
TARA_PSW-k99_9218994/Region 1149.1	hglE-KS,T1PKS	1	34,438			

TARA_PSW-k99_9502407/Region 1155.1	hserlactone,T1PKS,butyrolactone	16,576	75,506	thailanstatin A	NRP + Polyketide	10%
TARA_PSW-k99_9520575/Region 1156.1	T1PKS	1	31,142			
TARA_PSW-k99_9885307/Region 1161.1	RiPP-like	123,220	134,071			
TARA_PSW-k99_10292531/Region 1168.1	NRPS	37,211	63,407	micacocidin	NRP:NRP siderophore + Polyketide:Modular type I + Polyketide:Iterative type I	20%
TARA_PSW-k99_11124378/Region 1176.1	NRPS,T1PKS	1	32,096			
TARA_PSW-k99_11132354/Region 1177.1	T1PKS	26,669	55,796			
TARA_RED-k99_165427/Region 1179.1	T1PKS	1	44,666			
TARA_RED-k99_598896/Region 1186.1	arylpolyene	1	12,070			
TARA_RED-k99_1972275/Region 1222.1	T1PKS	1	14,296			
TARA_RED-k99_2043078/Region 1224.1	T1PKS	496,208	542,477			
TARA_RED-k99_2052333/Region 1225.1	arylpolyene	1	23,209			
TARA_RED-k99_2210738/Region 1227.1	terpene	83,356	104,408			
TARA_RED-k99_2529078/Region 1238.1	terpene	20,674	41,504			
TARA_RED-k99_2796487/Region 1246.1	hglE-KS	1	25,265			
TARA_RED-k99_3927568/Region 1270.1	lassopeptide	1,120	17,726			
TARA_RED-k99_3927568/Region 1270.2	lassopeptide	58,412	80,792			
TARA_RED-k99_3930265/Region 1271.1	T1PKS	1	32,947	oviedomycin	Polyketide:Type II	8%
TARA_RED-k99_4022104/Region 1272.1	RiPP-like	15,157	26,029			
TARA_RED-k99_4941388/Region 1293.1	hserlactone	32,380	52,994			
TARA_RED-k99_5017231/Region 1297.1	T1PKS	1	25,563	eicoseicosapentaenoic acid	Other	10%
TARA_RED-k99_5778951/Region 1308.1	T1PKS	1	16,817			
TARA_RED-k99_7992211/Region 1360.1	T1PKS	1	28,918	guadinomine	NRP + Polyketide	7%
TARA_RED-k99_8017624/Region 1362.1	T1PKS	1	19,939			
TARA_RED-k99_8701261/Region 1374.1	terpene	22,497	44,248			
TARA_RED-k99_9355052/Region 1383.1	arylpolyene	1	37,543			
TARA_ANW-k99_31938/Region 1392.1	T1PKS	1	11,964			
TARA_ANW-k99_426342/Region 1394.1	RiPP-like	1	7,387			

TARA_ANW-k99_1018361/Region 1399.1	hserlactone	1	18,955			
TARA_ANW-k99_2040536/Region 1411.1	NRPS	1	27,340			
TARA_ANW-k99_2931189/Region 1419.1	terpene,hglE-KS	1	17,323			
TARA_ANW-k99_3409679/Region 1423.1	T1PKS	1	15,796			
TARA_ANW-k99_3484098/Region 1424.1	arylpolyene	1	20,154			
TARA_ANW-k99_4094640/Region 1432.1	T3PKS	1	11,081			
TARA_ANW-k99_6319736/Region 1459.1	NRPS,T1PKS	1	46,187			
TARA_ANW-k99_6695559/Region 1463.1	T1PKS	1	11,561			
TARA_ANW-k99_6760279/Region 1464.1	arylpolyene	1	15,451			
TARA_ANW-k99_6797346/Region 1465.1	T1PKS	2,692	35,844			
TARA_ANW-k99_7068302/Region 1469.1	arylpolyene	16,001	41,743			
TARA_ANW-k99_7171645/Region 1470.1	arylpolyene	1	15,144			
TARA_ANW-k99_7836192/Region 1478.1	arylpolyene	1	49,520	APE Vf	Other	25%
TARA_ANW-k99_8234935/Region 1481.1	arylpolyene,resorcinol	1	28,112	flexirubin	Polyketide	66%
TARA_ANW-k99_9622640/Region 1494.1	ladderane,arylpolyene	1	13,696			
TARA_ANW-k99_9709422/Region 1497.1	T1PKS	1	12,580	eicoseicosapentaenoic acid	Other	21%
TARA_ANW-k99_10064631/Region 1501.1	T1PKS	1	13,120			
TARA_ANW-k99_11223962/Region 1508.1	T1PKS	1	25,756			
TARA_ANW-k99_11302352/Region 1510.1	T1PKS	1	13,861			
TARA_SOC-k99_17348/Region 1511.1	arylpolyene	1	12,814			

TARA_SOC-k99_48115/Region 1512.1	T3PKS	1	16,812			
TARA_SOC-k99_237486/Region 1515.1	arylpolyene	1	29,040			
TARA_SOC-k99_893914/Region 1524.1	arylpolyene	1	20,646	flexirubin	Polyketide	41%
TARA_SOC-k99_1216356/Region 1528.1	T3PKS	1	29,224			
TARA_SOC-k99_1784331/Region 1533.1	arylpolyene	1	13,234	flexirubin	Polyketide	16%
TARA_SOC-k99_1983783/Region 1536.1	terpene	1	14,013	carotenoid	Terpene	28%
TARA_SOC-k99_2535372/Region 1545.1	arylpolyene,T3PKS	1	20,896			
TARA_SOC-k99_3385775/Region 1553.1	arylpolyene	1	20,337			

Appendix 8

Table S5: Showing the distribution of known BGC pathway identified from the SAO non-redundant KS contigs and their percentage similarity

Contigs/region	BGC type	From location	To location	Most similar known cluster	Type	Similarity
1_k141_73404/Region 2.1	arylpolyene	1	4,034			
1_k141_873093/Region 10.1	terpene	166,809	187,642	carotenoid	Terpene	42%
1_k141_873093/Region 10.2	T1PKS,NRPS	681,163	731,905			
1_k141_328873/Region 39.1	arylpolyene	1	1,769	APE Vf	Other	10%
1_k141_845589/Region 60.1	T1PKS	1	9,986			
2_k141_501545/Region 67.1	resorcinol	73,959	115,107			
2_k141_353594/Region 69.1	hglE-KS,T1PKS	74,570	128,062	paulomycin	Other	3%
3_k141_96687/Region 95.1	T1PKS,hglE-KS	25,309	78,801	paulomycin	Other	3%
3_k141_702016/Region 97.1	resorcinol	73,959	115,107			
3_k141_633769/Region 107.1	arylpolyene	1	9,764			
3_k141_467259/Region 114.1	arylpolyene	1	4,134			
3_k141_808400/Region 122.1	arylpolyene	1	3,808			
3_k141_663669/Region 125.1	arylpolyene	1	11,616			
3_k141_789508/Region 136.1	T1PKS,hglE-KS	213,391	264,848			
3_k141_892661/Region 144.1	ladderane	1	2,597			
4_k141_912315/Region 168.1	arylpolyene	1	3,998			
5_k141_251289/Region 185.1	T1PKS	1	9,903			
5_k141_652315/Region 189.1	resorcinol	46,182	87,330			
5_k141_68513/Region 194.1	arylpolyene	1	2,041	APE Vf	Other	10%
5_k141_688086/Region 205.1	T1PKS,hglE-KS	26,083	79,575	paulomycin	Other	3%

5_k141_1185582/Region 212.1	arylpolyene	1	3,093			
5_k141_78374/Region 214.1	arylpolyene	1	9,904			
5_k141_610370/Region 223.1	T1PKS,hglE-KS	259,544	311,001			
5_k141_920741/Region 226.1	arylpolyene	1	4,645			
6_k141_50111/Region 228.1	arylpolyene	1	14,956			
6_k141_235002/Region 232.1	resorcinol	74,124	115,272			
6_k141_193528/Region 236.1	arylpolyene	1	3,876			
6_k141_265014/Region 237.1	arylpolyene	1	2,301			
6_k141_159118/Region 239.1	arylpolyene	1	6,746	aryl polyenes	Other	22%
6_k141_59400/Region 241.1	T1PKS,hglE-KS	26,083	79,575	paulomycin	Other	3%
6_k141_266046/Region 242.1	hglE-KS	1	8,499			
6_k141_68143/Region 244.1	ectoine	89,601	99,999	ectoine	Other	80%
6_k141_97482/Region 247.1	T1PKS,hglE-KS	1	17,833			
6_k141_111942/Region 250.1	ladderane	1	6,760			
6_k141_225096/Region 251.1	arylpolyene	1	37,391	APE Vf	Other	10%
7_k141_627629/Region 261.1	T1PKS,NRPS	1	26,776			
7_k141_84381/Region 271.1	T1PKS	17,712	64,164			
7_k141_203753/Region 276.1	arylpolyene	1	1,901	APE Vf	Other	10%
7_k141_334343/Region 277.1	hserlactone	49,579	70,301			
8_k141_535735/Region 282.1	arylpolyene	1	1,809			
8_k141_101970/Region 285.1	arylpolyene	1	7,498			
8_k141_471464/Region 287.1	arylpolyene	1	2,204	APE Vf	Other	20%
8_k141_281465/Region 298.1	terpene	4,286	25,290			
9_k141_82633/Region 321.1	arylpolyene	1	4,367			
9_k141_84122/Region 324.1	arylpolyene	1	2,425	APE Vf	Other	10%
9_k141_733915/Region 344.1	arylpolyene	1	4,667	APE Vf	Other	15%
9_k141_868470/Region 348.1	NRPS,T1PKS	1	15,886			
9_k141_473766/Region 355.1	arylpolyene	1	2,051	APE Vf	Other	10%
9_k141_22135/Region 361.1	T1PKS	1	11,467			
9_k141_983411/Region 367.1	arylpolyene	1	17,309	flexirubin	Polyketide	33%

9_k141_823203/Region 368.1	T3PKS	1	7,838			
10_k141_680081/Region 370.1	arylpolyene	1	5,570	flexirubin	Polyketide	19%
10_k141_580458/Region 379.1	arylpolyene	6,400	49,981	APE Vf	Other	30%
10_k141_451084/Region 383.1	arylpolyene	1	3,710			
10_k141_295956/Region 384.1	arylpolyene	1	1,814			
10_k141_849482/Region 397.1	arylpolyene	1	1,955	flexirubin	Polyketide	5%
10_k141_168623/Region 399.1	arylpolyene	1	3,184	flexirubin	Polyketide	11%
10_k141_330940/Region 406.1	arylpolyene	1	5,763	flexirubin	Polyketide	13%
10_k141_904164/Region 407.1	T3PKS	1	10,754			
10_k141_569960/Region 411.1	arylpolyene	1	3,951	APE Vf	Other	15%
11_k141_956206/Region 419.1	arylpolyene	1	8,578	flexirubin	Polyketide	25%
11_k141_569222/Region 431.1	NRPS,T1PKS	1	36,118			
11_k141_203925/Region 432.1	PUFA,hglE-KS	60,376	106,022	eicosapentaenoic acid	Other	66%
11_k141_389331/Region 460.1	T1PKS,NRPS	1	33,699			
11_k141_813671/Region 464.1	T3PKS	1	4,477			
12_k141_1096211/Region 484.1	NRPS	100,883	132,197	chejuenolide A / chejuenolide B	Polyketide	15%
13_k141_100709/Region 509.1	hglE-KS,T1PKS	24,394	75,851			
13_k141_105401/Region 528.1	arylpolyene	1	3,024	flexirubin	Polyketide	13%
13_k141_839269/Region 540.1	arylpolyene	1	1,181			
13_k141_86176/Region 541.1	arylpolyene	1	4,338			
13_k141_211967/Region 544.1	T1PKS,hglE-KS	26,083	79,575	paulomycin	Other	3%
13_k141_742488/Region 551.1	resorcinol	73,959	115,107			
13_k141_39623/Region 553.1	T1PKS	1	20,047			
14_k141_399450/Region 578.1	T1PKS	31,333	77,902			
14_k141_1012099/Region 580.1	arylpolyene	1	4,671	flexirubin	Polyketide	8%

14_k141_771963/Region 589.1	resorcinol	73,932	115,080			
14_k141_223467/Region 595.1	arylpolyene	1	6,269	flexirubin	Polyketide	16%
14_k141_897973/Region 600.1	arylpolyene	1	2,186	APE Ec	Other	15%
14_k141_873357/Region 613.1	T1PKS,hglE-KS	26,083	79,575	paulomycin	Other	3%
14_k141_753018/Region 618.1	ladderane	1	2,582			
14_k141_1029109/Region 620.1	arylpolyene	1	3,314	APE Vf	Other	15%
14_k141_272023/Region 637.1	arylpolyene	1	2,552	flexirubin	Polyketide	11%
15_k141_759704/Region 654.1	arylpolyene	1	1,939	flexirubin	Polyketide	8%
15_k141_305974/Region 655.1	arylpolyene	1	1,619	flexirubin	Polyketide	8%
15_k141_526901/Region 662.1	arylpolyene	1	3,163			
15_k141_528765/Region 663.1	arylpolyene	1	4,033	flexirubin	Polyketide	13%
16_k141_697129/Region 679.1	T1PKS	1	14,790			
16_k141_502314/Region 680.1	arylpolyene	35,558	82,131	flexirubin	Polyketide	55%
16_k141_920012/Region 681.1	arylpolyene	12,294	55,858	APE Vf	Other	40%
16_k141_392261/Region 682.1	arylpolyene	1	2,080			
16_k141_284832/Region 691.1	T3PKS	74,197	115,300			
16_k141_982964/Region 703.1	arylpolyene	1	8,748	APE Vf	Other	15%
16_k141_44289/Region 712.1	arylpolyene	1	8,813	APE Vf	Other	10%
16_k141_294334/Region 713.1	arylpolyene	1	3,738	flexirubin	Polyketide	13%
16_k141_387750/Region 735.1	arylpolyene	1	33,881	APE Vf	Other	30%
17_k141_565457/Region 745.1	arylpolyene	187,459	231,040	APE Vf	Other	30%
17_k141_358034/Region 747.1	arylpolyene	7,209	50,772	APE Vf	Other	35%
17_k141_798470/Region 759.1	arylpolyene	1	3,575	APE Ec	Other	15%
17_k141_537429/Region 781.1	arylpolyene	1	4,004	APE Vf	Other	10%

17_k141_665540/Region 782.1	arylpolyene	1	2,456	flexirubin	Polyketide	13%
18_k141_875853/Region 818.1	T1PKS	17,595	64,164			
18_k141_554657/Region 823.1	arylpolyene	1	3,204			
18_k141_168703/Region 845.1	arylpolyene	1	17,174	APE Vf	Other	30%
19_k141_209518/Region 893.1	ladderane	1	2,663			
21_k141_98863/Region 928.1	T3PKS	1	8,723			
21_k141_327557/Region 958.1	resorcinol	73,959	115,107			
21_k141_957892/Region 962.1	T1PKS,hglE-KS	26,083	79,575	paulomycin	Other	3%
22_k141_605048/Region 977.1	arylpolyene,resorcinol	162,351	208,418	flexirubin	Polyketide	13%
22_k141_336485/Region 988.1	resorcinol,arylpolyene	1	8,641	flexirubin	Polyketide	16%
22_k141_275063/Region 996.1	resorcinol	3,074	27,493			
22_k141_65869/Region 1001.1	arylpolyene	371,576	417,912	flexirubin	Other	36%
22_k141_422854/Region 1006.1	arylpolyene	1	4,367	flexirubin	Polyketide	8%
22_k141_28569/Region 1011.1	arylpolyene	1	6,773	APE Vf	Other	15%
22_k141_697585/Region 1020.1	arylpolyene	1	1,387	flexirubin	Polyketide	5%
22_k141_470688/Region 1036.1	arylpolyene	1	2,476	flexirubin	Polyketide	5%
22_k141_54015/Region 1037.1	NRPS,T1PKS	1	12,344			
22_k141_825735/Region 1038.1	ladderane	1	4,998			
22_k141_54573/Region 1041.1	ectoine	87,239	97,637	ectoine	Other	80%
22_k141_222669/Region 1043.1	arylpolyene	1	5,589			
22_k141_392020/Region 1053.1	T1PKS	19,237	47,400			
22_k141_412564/Region 1057.1	arylpolyene	1	3,610			
22_k141_559517/Region 1058.1	NRPS	1	11,289			

22_k141_457442/Region 1064.1	arylpolyene	1	4,127			
22_k141_686181/Region 1065.1	arylpolyene	1	37,391	APE Vf	Other	10%
22_k141_80919/Region 1068.1	T3PKS	1	8,831			
22_k141_354190/Region 1071.1	arylpolyene	1	2,460			
23_k141_716104/Region 1095.1	arylpolyene	1	3,595	APE Vf	Other	15%
23_k141_549371/Region 1096.1	T3PKS	1	2,036			
23_k141_694466/Region 1097.1	arylpolyene	75,771	122,344	flexirubin	Polyketide	55%
23_k141_71496/Region 1111.1	NRPS	1	12,215			
23_k141_453723/Region 1121.1	arylpolyene	1	1,487			
23_k141_98389/Region 1128.1	T3PKS	1	9,743			
23_k141_602719/Region 1132.1	T3PKS	498,211	530,245			
23_k141_499467/Region 1138.1	arylpolyene	1	28,639	APE Vf	Other	10%
23_k141_187763/Region 1148.1	T3PKS	1	7,623			
24_k141_386724/Region 1156.1	T1PKS	1	7,190			
24_k141_681529/Region 1157.1	arylpolyene	1	3,368	APE Vf	Other	10%
24_k141_350974/Region 1161.1	arylpolyene	1	34,398	flexirubin	Polyketide	55%
24_k141_111922/Region 1166.1	T3PKS	1	3,867			
24_k141_462903/Region 1173.1	T3PKS	1	17,999			
24_k141_225047/Region 1177.1	arylpolyene	2,427	24,274	APE Vf	Other	10%

24_k141_614420/Region 1183.1	resorcinol	73,959	115,107			
24_k141_208898/Region 1184.1	hglE-KS,T1PKS,PUFA	1	37,171	eicosapentaenoic acid-like compound	Other	18%
24_k141_653222/Region 1197.1	arylpolyene	28,003	65,238	APE Vf	Other	10%
24_k141_560063/Region 1201.1	T3PKS	1	2,541			
24_k141_708757/Region 1203.1	arylpolyene	1	2,853	flexirubin	Polyketide	8%
24_k141_212920/Region 1216.1	arylpolyene	1	2,484	APE Vf	Other	10%
24_k141_46276/Region 1217.1	arylpolyene,resorcinol	1	23,289	flexirubin	Polyketide	44%
24_k141_325214/Region 1218.1	NRPS	1	12,113			
24_k141_230241/Region 1219.1	T1PKS,hglE-KS	26,083	69,203			
24_k141_345032/Region 1221.1	arylpolyene	1	3,352	flexirubin	Polyketide	13%
24_k141_471897/Region 1222.1	hglE-KS	82,396	106,393			
24_k141_695390/Region 1228.1	arylpolyene	1	10,359	APE Vf	Other	25%
24_k141_417363/Region 1231.1	arylpolyene	1	8,227	flexirubin	Polyketide	25%
24_k141_328670/Region 1234.1	arylpolyene	1	8,979	APE Vf	Other	15%
24_k141_639534/Region 1238.1	T3PKS	134,622	175,725			

Appendix 9

Table S6: Showing the medium to high quality MAGs with their respective taxonomic classification, completeness and contamination

MAG ID	Classification	Completeness	Contamination
200m S6 Spring			
SAR324_204	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	88.65	0.84
Posei_205	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	83.96	0.4
Verru_203	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	85.83	4.77
Acidi_207	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	85.31	3.85
Dehal_201	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA1328;s__	87.36	3.56
Gamma_204	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	58.79	6.04
Dehal_202	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__	67.54	5.17
Verru_204	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	63.31	6.67
Acidi_208	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	83.1	2.99
Gamma_205	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	50.81	0

Gamma_206	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__	77.3	6.4
Gamma_207	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	84.76	3.66
Gamma_208	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__UBA11791;s__	86.91	6.33
200m Winter S39			
Posei_222	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	83.16	0
Nitros_205	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__LS-NOB;s__LS-NOB sp003545625	81.09	2.23
Posei_223	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-P;s__	56.13	3.3
Posei_224	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O2;s__	58.51	1.6
Posei_225	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	70.11	1.84
SAR324_214	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	90.07	3.69
Verru_216	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	97.75	0.81
Verru_217	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__;s__	76.12	2.99
SAR324_215	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	61.81	1.73
Acidi_227	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	89.32	3.85
Posei_226	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__	52.4	4.8
Verru_218	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	74.13	5.45
Posei_227	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	72	6.89
Gamma_235	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__GCA-2726415;s__	75.44	3.05
Gamma_236	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__GCA-002728725;s__	77.59	2.83
Gamma_237	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	96.65	1.83
Gamma_238	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	60.8	2.17
Acidi_228	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__TK06;g__UBA2110;s__UBA2110 sp002331465	72.98	5.57
Marin_207	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	90.11	0
Crenar_203	d__Archaea;p__Crenarchaeota;c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopelagicus;s__	66.83	8.74
500m Spring S7			
Dehal_501	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__	66.97	4.36

Nitros_504	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__LS-NOB;s__LS-NOB sp003545625	92.47	7.34
SAR324_506	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	83.94	2.52
Verru_503	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	83.56	2.4
Nitros_505	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__SCGCAAA288-L16;s__	56.21	7.76
SAR324_507	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	91.26	0.84
Acidi_507	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	94.87	6.41
Posei_509	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	82.36	0
Marin_505	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	90.11	1.1
Gamma_521	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	59.88	2.01
Posei_510	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	69.73	0.29
Acidi_508	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__TK06;g__UBA2110;s__UBA2110 sp002331465	93.16	2.14
Gamma_522	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	82.62	1.83
Dehal_502	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA826;g__UBA11996;s__	66.49	1.98
500m Winter			
S40			
SAR324_520	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	87.99	1.03
Planc_502	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__UBA1268;g__UBA1268;s__	55.15	1.25
Verru_516	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	79.94	8.61
Gemma_505	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__GCA-2718595;s__	63.49	2.81
Gamma_551	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__	91.11	0.68
Gamma_552	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	63.07	1.7
Gamma_553	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	55.27	5.69
Acidi_528	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	83.57	2.99
Bacte_503	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-121220-bin8;s__	92.42	2.94
Gamma_554	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	67.85	0
Gamma_555	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__GCA-002728725;s__	57.77	7.92
Gamma_556	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__GCA-2726415;s__	58.02	0.64
Planc_503	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__Pirellulaceae;g__Mariniblastus;s__	77.1	2.35

2600m Spring S8			
Posei_d01	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	86.4	0
DG_d01	d__Bacteria;p__TA06_A;c__DG-26_A;o__;f__;g__;s__	57.26	0.75
Gemma_d01	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__UBA1138;s__UBA1138 sp003447875	60.84	3.3
Gamma_d01	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__UBA11791;s__	53.28	1.32
Posei_d02	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp003193815	70.92	1.64
Planc_d01	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__Pirellulaceae;g__;s__	72.74	3.53
Gamma_d02	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	84.71	2.89
SAR324_d01	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	90.42	0
Posei_d03	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__UBA60;s__UBA60 sp002503395	68.63	2.53
SAR324_d02	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	56.2	5.93
Verru_d01	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	65.21	8.91
SAR324_d03	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__	85.17	0
Dehal_d01	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__UBA9611 sp002746355	86.56	2.97
Gamma_d03	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__AEGEAN-183;s__	65.76	0.74
Gamma_d04	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__	94.78	1.83
Dehal_d02	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA9455;s__UBA9455 sp002313245	94.55	0
Dehal_d03	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__UBA1123;s__UBA1123 sp002313895	86.33	4.95
Posei_d04	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	84.4	0.8
Gamma_d05	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	93.91	0
Acidi_d01	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	93.16	2.14
Gamma_d06	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__;s__	98.23	0
Gamma_d07	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	58.35	0.78
Dehal_d04	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA11650;s__UBA11650 sp002401285	85.52	9.16
Marin_d01	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	86.81	1.1
Gamma_d08	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	54.78	0.35
Marin_d02	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA8229;g__UBA8229;s__UBA8229 sp003535775	52.38	3.36
2690m Winter S37			

Gamma_d25	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__UBA9145 sp002731775	93.95	3.46
Gamma_d26	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Oleiphilaceae;g__Marinobacter;s__Marinobacter salarius	91.88	0.55
Posei_d16	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__;s__	65.01	1.14
Posei_d17	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp003193815	75.73	3.77
SAR324_d09	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	87.02	0
Posei_d18	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	85.6	0
Posei_d19	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__UBA60;s__UBA60 sp002503395	61.62	4.57
Dehal_d16	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA9455;s__UBA9455 sp002313245	65.75	0
Gemma_d06	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__UBA1138;s__UBA1138 sp003447875	91.75	4.4
Posei_d20	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-L1;s__	77.33	0.8
Gamma_d27	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__AEGEAN-183;s__	64.87	0.49
Bacte_d02	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__GCA-2700405;s__	67.36	1.23
Acidi_d05	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	95.44	2.14
Gamma_d28	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	59.42	1.06
Posei_d21	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	81.2	0.8
Gamma_d29	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	91.92	0.66
Verru_d04	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	73.14	0
Gamma_d30	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__;s__	95.92	0.17
Verru_d05	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__;s__	89.58	4.25
Marin_d08	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__SCGC-AAA003-L08;f__;g__;s__	80.47	1.1
Marin_d09	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	86.81	0
SAR324_d10	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__	87.9	0
Dehal_d17	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__UBA9611 sp002746355	85.64	0.99

Appendix 10

Table S7: Showing the MAGs and their reference genomes ANI values

MAGs_ID	classification	fastani_reference	fastani_reference_radius	fastani_ani	fastani_af	closest_placement_reference	closest_placement_ani	closest_placement_af	aa_percent
200m spring S6									
SAR324_204	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	97.43	0.82	GCA_000213335.2	97.73	0.49	90.18
Verru_203	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	N/A	N/A	N/A	N/A	GCA_002299785.1	91.03	0.61	76.65
Acidi_207	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	GCA_002699725.1	95.0	97.26	0.89	GCA_002699725.1	97.26	0.89	74.31
Dehal_201	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA1328;s__	N/A	N/A	N/A	N/A	GCA_002501045.1	77.78	0.29	79.62
Gamma_204	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	56.71
Dehal_202	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	60.71
Verru_204	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	48.61
Acidi_208	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	N/A	N/A	N/A	N/A	GCA_002717365.1	78.98	0.36	76.15

Gamma_205	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	N/A	N/A	N/A	N/A	GCA_002698045.1	95.93	0.59	51.77
Gamma_206	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	71.13
Gamma_207	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	GCA_002456995.1	95.0	97.65	0.76	GCA_002456995.1	97.65	0.76	79.82
Gamma_208	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__UBA11791;s__	N/A	N/A	N/A	N/A	GCA_002714255.1	93.61	0.61	79.88
Posei_205	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	GCA_002495735.1	95.0	99.11	0.87	GCA_002495735.1	99.11	0.87	78.71
200m winter S39									
Acidi_227	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	GCA_002699725.1	95.0	97.15	0.86	GCA_002699725.1	97.15	0.86	79.38
Acidi_228	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__TK06;g__UBA2110;s__UBA2110 sp002331465	GCA_002331465.1	95.0	96.82	0.81	GCA_002331465.1	96.82	0.81	71.85
Marin_207	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	GCA_002715035.1	95.0	98.79	0.87	GCA_002715035.1	98.79	0.87	89.48
Gamma_237	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	GCA_002456995.1	95.0	97.57	0.82	GCA_002456995.1	97.57	0.82	95.93
SAR324_214	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	97.74	0.75	GCA_002082305.1	97.74	0.75	89.01
Nitros_205	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__LS-NOB;s__LS-NOB sp003545625	GCA_003545625.1	95.0	97.93	0.82	GCA_003545625.1	97.93	0.82	72.4
Verru_217	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	62.2
Verru_216	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	N/A	N/A	N/A	N/A	GCA_002299785.1	91.28	0.73	86.49
Verru_218	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	54.6
Gamma_236	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__GCA-002728725;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	98.1
Gamma_235	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__GCA-2726415;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	73.99
Gamma_238	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	84.05
SAR324_215	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	62.26
Posei_222	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	GCA_002495735.1	95.0	99.36	0.88	GCA_002495735.1	99.36	0.88	81.25
Posei_225	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	GCA_002498525.1	95.0	97.67	0.99	GCA_002498525.1	97.67	0.99	64.62
Posei_226	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	40.57
Posei_224	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O2;s__	N/A	N/A	N/A	N/A	GCA_002499785.1	89.0	0.82	55.0
Posei_223	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-P;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	44.63

Posei_227	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	70.69
Crenar_203	d__Archaea;p__Crenarchaeota;c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopelagicus;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	55.82
500m spring S7									
Acidi_508	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__TK06;g__UBA2110;s__UBA2110 sp002331465	GCA_002331465.1	95.0	96.98	0.82	GCA_002331465.1	96.98	0.82	87.84
Gamma_522	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002456995	GCA_002456995.1	95.0	97.45	0.79	GCA_002456995.1	97.45	0.79	83.21
Acidi_507	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002699725	GCA_002699725.1	95.0	97.15	0.86	GCA_002699725.1	97.15	0.86	89.46
Marin_505	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	GCA_002715035.1	95.0	98.64	0.88	GCA_002715035.1	98.64	0.88	89.29
Nitros_504	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__LS-NOB;s__LS-NOB sp003545625	GCA_003545625.1	95.0	98.71	0.74	GCA_003545625.1	98.71	0.74	86.07
SAR324_507	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	97.92	0.79	GCA_002082305.1	97.92	0.79	92.72
Dehal_502	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA826;g__UBA11996;s__	N/A	N/A	N/A	N/A	GCA_002708495.1	95.95	0.57	56.85
Dehal_501	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__	N/A	N/A	N/A	N/A	GCA_002730485.1	79.31	0.57	60.65
Verru_503	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	N/A	N/A	N/A	N/A	GCA_002299785.1	90.79	0.58	68.77
Gamma_521	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	78.33
SAR324_506	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	N/A	N/A	N/A	N/A	GCA_002707655.1	98.2	0.58	81.35
Nitros_505	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__SCGCAAA288-L16;s__	N/A	N/A	N/A	N/A	GCA_000372225.1	93.7	0.25	47.58
Posei_509	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__Thalassarchaeum;s__Thalassarchaeum sp002495735	GCA_002495735.1	95.0	99.38	0.89	GCA_002495735.1	99.38	0.89	80.43
Posei_510	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	65.53
500m winter S40									
SAR324_520	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	97.13	0.82	GCA_000213335.2	97.16	0.42	87.1
Acidi_528	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	N/A	N/A	N/A	N/A	GCA_002717365.1	78.63	0.39	74.62
Verru_516	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	N/A	N/A	N/A	N/A	GCA_002299785.1	90.79	0.61	74.54
Planc_503	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__Pirellulaceae;g__Mariniblastus;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	64.11
Planc_502	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__UBA1268;g__UBA1268;s__	N/A	N/A	N/A	N/A	GCA_002862165.1	87.7	0.74	49.68
Bacte_503	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-121220-bin8;s__	N/A	N/A	N/A	N/A	GCA_002700465.1	77.83	0.39	83.49

Gemma_505	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__GCA-2718595;s__	N/A	N/A	N/A	N/A	GCA_002718595.1	78.83	0.28	59.5
Gamma_552	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	63.43
Gamma_551	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	97.22
Gamma_554	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	N/A	N/A	N/A	N/A	GCA_002698045.1	96.08	0.54	59.56
Gamma_555	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__GCA-002728725;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	62.54
Gamma_556	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__GCA-2726415;s__	N/A	N/A	N/A	N/A	GCA_002726415.1	77.53	0.18	53.35
Gamma_553	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	73.79
2600m spring S8									
Dehal_d01	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__UBA9611 sp002746355	GCA_002746355.1	95.0	99.25	0.94	GCA_002746355.1	99.25	0.94	80.28
Dehal_d02	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA9455;s__UBA9455 sp002313245	GCA_002313245.1	95.0	99.4	0.86	GCA_002313245.1	99.4	0.86	87.38
Gamma_d04	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp003447825	GCA_003447825.1	95.0	99.41	0.89	GCA_003447825.1	99.41	0.89	92.94
Dehal_d03	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__UBA1123;s__UBA1123 sp002313895	GCA_002313895.1	95.0	99.52	0.89	GCA_002313895.1	99.52	0.89	76.33
Marin_d01	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	GCA_002715035.1	95.0	96.78	0.82	GCA_002715035.1	96.78	0.82	90.18
SAR324_d01	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	99.16	0.88	GCA_002082305.1	99.16	0.88	92.32
Dehal_d04	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA11650;s__UBA11650 sp002401285	GCA_002401285.1	95.0	98.7	0.68	GCA_002401285.1	98.7	0.68	70.04
Gemma_d01	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__UBA1138;s__UBA1138 sp003447875	GCA_003447875.1	95.0	98.79	0.91	GCA_003447875.1	98.79	0.91	53.85
Marin_d02	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA8229;g__UBA8229;s__UBA8229 sp003535775	GCA_003535775.1	95.0	98.68	0.68	GCA_003535775.1	98.68	0.68	57.18
Acidi_d01	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	N/A	N/A	N/A	N/A	GCA_002717365.1	79.26	0.43	88.17
Verru_d01	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	51.01
Planc_d01	d__Bacteria;p__Planctomycetota;c__Planctomycetes;o__Pirellulales;f__Pirellulaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	63.83
DG_d01	d__Bacteria;p__TA06_A;c__DG-26_A;o__;f__;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	51.29
Gamma_d02	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	80.28
Gamma_d05	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	N/A	N/A	N/A	N/A	GCA_002698045.1	94.21	0.82	92.34
Gamma_d06	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	97.08
Gamma_d01	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA10353;f__LS-SOB;g__UBA11791;s__	N/A	N/A	N/A	N/A	GCA_002714255.1	97.85	0.45	52.9
Gamma_d03	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__AEGEAN-183;s__	N/A	N/A	N/A	N/A	GCA_001628005.1	79.76	0.67	84.4

Gamma_d07	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	59.56
Gamma_d08	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	69.27
SAR324_d03	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	85.95
SAR324_d02	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__	N/A	N/A	N/A	N/A	GCA_002707655.1	N/A	N/A	44.84
Posei_d01	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	GCA_002498525.1	95.0	98.4	0.98	GCA_002498525.1	98.4	0.98	83.7
Posei_d02	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp003193815	GCA_003193815.1	95.0	95.39	0.67	GCA_003193815.1	95.39	0.67	62.33
Posei_d03	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__UBA60;s__UBA60 sp002503395	GCA_002503395.1	95.0	96.97	0.93	GCA_002503395.1	96.97	0.93	63.41
Posei_d04	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	N/A	N/A	N/A	N/A	GCA_002506485.1	91.89	0.73	81.07
2690m winter S37									
Dehal_d16	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__Bin127;g__UBA9455;s__UBA9455 sp002313245	GCA_002313245.1	95.0	98.91	0.87	GCA_002313245.1	98.91	0.87	55.52
SAR324_d09	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002082305	GCA_002082305.1	95.0	99.1	0.89	GCA_002082305.1	99.1	0.89	89.15
Gamma_d25	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__UBA9145 sp002731775	GCA_002731775.1	95.0	98.6	0.79	GCA_002731775.1	98.6	0.79	97.28
Gamma_d26	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Oleiphilaceae;g__Marinobacter;s__Marinobacter salarius	GCF_000831005.1	95.0	98.67	0.95	GCF_000831005.1	98.67	0.95	93.89
Marin_d09	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__TCS55;s__TCS55 sp002715035	GCA_002715035.1	95.0	96.74	0.8	GCA_002715035.1	96.74	0.8	88.65
Dehal_d17	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__UBA9611 sp002746355	GCA_002746355.1	95.0	99.05	0.94	GCA_002746355.1	99.05	0.94	73.85
Gemma_d06	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__UBA1138;s__UBA1138 sp003447875	GCA_003447875.1	95.0	98.94	0.93	GCA_003447875.1	98.94	0.93	87.98
Acidi_d05	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales;f__MedAcidi-G1;g__UBA9410;s__	N/A	N/A	N/A	N/A	GCA_002717365.1	79.16	0.44	89.17
Verru_d05	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	77.18
Verru_d04	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__	N/A	N/A	N/A	N/A	GCA_002299785.1	90.69	0.62	65.81
Bacte_d02	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__GCA-2700405;s__	N/A	N/A	N/A	N/A	GCA_002700405.1	81.92	0.58	90.62
Marin_d08	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__SCGC-AAA003-L08;f__;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	73.45
Gamma_d29	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__Thioglobus;s__	N/A	N/A	N/A	N/A	GCA_002698045.1	94.15	0.81	91.31
Gamma_d30	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiomicrospirales;f__Thioglobaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	91.87
Gamma_d27	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__AEGEAN-183;s__	N/A	N/A	N/A	N/A	GCA_001628005.1	79.77	0.63	83.33
Gamma_d28	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	84.4

SAR324_d10	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	88.21
Posei_d19	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__UBA60;s__UBA60 sp002503395	GCA_002503395.1	95.0	96.83	0.97	GCA_002503395.1	96.83	0.97	50.02
Posei_d18	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassoarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002498525	GCA_002498525.1	95.0	98.37	0.97	GCA_002498525.1	98.37	0.97	83.72
Posei_d17	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp003193815	GCA_003193815.1	95.0	96.02	0.69	GCA_003193815.1	96.02	0.69	64.97
Posei_d20	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-L1;s__	N/A	N/A	N/A	N/A	GCA_002499015.1	81.87	0.73	76.13
Posei_d16	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	N/A	48.67
Posei_d21	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__	N/A	N/A	N/A	N/A	GCA_002506485.1	91.84	0.74	79.61

Appendix 11

Table S8: The Southern Ocean MAGs and their BGC abundance

MAG_ID	Phylum	Region	Type	From	To	Most similar known cluster		Similarity
200m Spring								
Dehal_201	Chloroflexota	Region 340.1	terpene	1	1,567			
Dehal_202	Chloroflexota	Region 17.1	other	1	6,957			
		Region 61.1	terpene	1	4,852	hopene	Terpene	15%
		Region 157.1	terpene	1	3,441			
		Region 608.1	terpene	1	1,557			
Gamma_204	Proteobacteria	Region 6.1	other	1	7,050			
Gamma_208	Proteobacteria	Region 216.1	betalactone	1	4,465			
Posei_205	Thermoplasmatota	Region 14.1	terpene	9,997	24,461			

SAR324_204	SAR324	Region 158.1	arylpolyene	1	5,621	APE Vf	Other	10%
		Region 258.1	PUFA	1	3,654			
		Region 292.1	T3PKS	1	3,239			
		Region 298.1	hglE-KS	1	3,136			
Verru_203	Verrucomicrobiota	Region 62.1	terpene	1	8,278			
		Region 134.1	terpene	1	5,632			
Verru_204	Verrucomicrobiota	Region 17.1	terpene	1	4,364			
		Region 183.1	terpene	1	2,334			
		Region 578.1	terpene	1	1,385			
		Region 799.1	ectoine	1	1,170			
		Region 861.1	terpene	1	1,124			
200m winter								
Acidi_228	Actinobacteriota	Region 127.1	other	1	6,379			
Gamma_237	Proteobacteria	Region 63.1	betalactone	1	15,291			
Nitros_205	Nitrospinota	Region 92.1	terpene	1	4,139			
		Region 188.1	RiPP-like	1	3,134			
		Region 458.1	terpene	1	1,884			
Posei_222	Thermoplasmatota	Region 34.1	terpene	1	14,411			
Posei_224	Thermoplasmatota	Region 5.1	terpene	1	10,670			
Posei_225	Thermoplasmatota	Region 99.1	terpene	1	4,222			
Posei_226	Thermoplasmatota	Region 225.1	terpene	1	1,768			
Posei_227	Thermoplasmatota	Region 3.1	terpene	4,092	21,629			
SAR324_214	SAR324	Region 3.1	T3PKS	9,389	32,131			
		Region 114.1	T1PKS,PUFA	1	7,587			
		Region 185.1	arylpolyene	1	5,177	APE Vf	Other	15%
		Region 265.1	hglE-KS	1	3,863			
		Region 329.1	RiPP-like	1	3,150			
SAR324_215	SAR324	Region 35.1	arylpolyene	1	4,064	APE Vf	Other	15%
		Region 556.1	hglE-KS	1	1,517			
Verru_216	Verrucomicrobiota	Region 28.1	terpene	1	11,941			

		Region 39.1	terpene	1,890	16,773			
Verru_217	Verrucomicrobiota	Region 45.1	T3PKS	1	7,338			
		Region 134.1	terpene	1	5,423			
		Region 442.1	T1PKS	1	3,114			
		Region 506.1	terpene	1	2,896			
		Region 694.1	hglE-KS	1	2,365	heterocyst glycolipids	Other	28%
		Region 792.1	hglE-KS	1	2,128			
Verru_218	Verrucomicrobiota	Region 144.1	terpene	1	2,755			
		Region 152.1	terpene	1	2,737			
		Region 561.1	terpene	1	1,465			
500m Spring								
Acidi_507	Actinobacteriota	Region 272.1	NRPS	1	1,140			
Dehal_501	Chloroflexota	Region 24.1	other	1	5,408			
		Region 136.1	terpene	1	3,404			
		Region 457.1	terpene	1	1,899			
		Region 685.1	terpene	1	1,456			
Gamma_522	Proteobacteria	Region 104.1	ectoine	1	5,765			
Nitros_504	Nitrospinota	Region 55.1	terpene	1	9,723			
		Region 242.1	terpene	1	3,942			
		Region 303.1	RiPP-like	1	3,145			
Nitros_505	Nitrospinota	Region 5.1	betalactone	1	15,228			
		Region 56.1	RiPP-like	1	6,026			
Posei_509	Thermoplasmatota	Region 20.1	terpene	10,374	27,024			
Posei_510	Thermoplasmatota	Region 28.1	terpene	1	7,879			
SAR324_506	SAR324	Region 159.1	hglE-KS	1	4,614			
		Region 243.1	arylpolyene	1	3,718			
		Region 299.1	betalactone	1	3,240			
		Region 423.1	hglE-KS	1	2,557			
		Region 440.1	T1PKS	1	2,501			
		Region 851.1	arylpolyene	1	1,265			

SAR324_507	SAR324	Region 104.1	PUFA,T1PKS	1	7,588			
		Region 164.1	arylpolyene	1	5,373	APE Vf	Other	15%
		Region 276.1	hglE-KS	1	3,543			
		Region 298.1	T3PKS	1	3,271			
Verru_503	Verrucomicrobiota	Region 111.1	terpene	1	7,031			
		Region 240.1	terpene	1	4,833			
500m winter								
Bacte_503	Bacteroidota	Region 11.1	arylpolyene	1	34,972	flexirubin	Polyketide	58%
Gamma_552	Proteobacteria	Region 65.1	other	1	7,408			
Planc_502	Planctomycetota	Region 295.1	terpene	1	2,873			
		Region 414.1	hglE-KS	1	2,353			
		Region 432.1	terpene	1	2,285			
Planc_503	Planctomycetota	Region 74.1	terpene	1	9,041			
		Region 495.1	phosphonate	1	3,811			
		Region 597.1	arylpolyene	1	3,226			
SAR324_520	SAR324	Region 53.1	arylpolyene	1	8,032	APE Vf	Other	15%
		Region 82.1	T1PKS,PUFA	1	6,661			
		Region 230.1	arylpolyene	1	3,706			
		Region 276.1	T3PKS	1	3,382			
		Region 822.1	hglE-KS	1	1,082			
Verru_516	Verrucomicrobiota	Region 442.1	terpene	1	2,593			
		Region 604.1	terpene	1	2,013			
		Region 1187.1	terpene	1	1,025			
2600m spring								
Dehal_d01	Chloroflexota	Region 14.1	other	1	35,614			
		Region 19.1	terpene	19,396	33,252			
		Region 39.1	terpene	9,158	22,060			
Dehal_d02	Chloroflexota	Region 11.1	terpene	13,202	34,672	heme D1	Other	17%
Dehal_d03	Chloroflexota	Region 38.1	terpene	2,965	17,572			
		Region 193.1	terpene	1	6,795			
Dehal_d04	Chloroflexota	Region 20.1	terpene	1	13,760			

		Region 114.1	terpene	1	6,941			
DG_d01	TA06	Region 219.1	sactipeptide	1	3,202			
		Region 339.1	terpene	1	2,789			
		Region 430.1	ladderane	1	2,570			
		Region 487.1	T3PKS	1	2,417			
Gamma_d01	Proteobacteria	Region 313.1	ectoine	1	3,316			
Gamma_d02	Proteobacteria	Region 35.1	betalactone	1	14,478			
Gamma_d04	Proteobacteria	Region 38.1	betalactone	6,366	27,948			
Gamma_d08	Proteobacteria	Region 14.1	phosphonate	1	29,327			
Gemma_d01	Gemmatimonadota	Region 229.1	T1PKS,hglE-KS	1	4,105			
		Region 259.1	hglE-KS	1	3,835	heterocyst glycolipids	Other	28%
		Region 472.1	ranthipeptide	1	2,723			
		Region 538.1	T1PKS	1	2,509			
Planc_d01	Planctomycetota	Region 88.1	terpene	1	5,443			
Posei_d01	Thermoplasmatota	Region 17.1	terpene	4,692	25,367			
Posei_d02	Thermoplasmatota	Region 194.1	terpene	1	1,924			
Posei_d03	Thermoplasmatota	Region 42.1	terpene	1	6,630			
Posei_d04	Thermoplasmatota	Region 6.1	terpene	1	13,655			
SAR324_d01	SAR324	Region 6.1	T3PKS	1	22,743			
		Region 41.1	PUFA,T1PKS	1	15,217			
		Region 200.1	hglE-KS	1	3,944			
SAR324_d02	SAR324	Region 347.1	hglE-KS	1	1,909			
		Region 360.1	PUFA	1	1,896			
		Region 469.1	hglE-KS	1	1,696			
		Region 476.1	arylpolyene	1	1,688	aryl polyenes	Other	16%
		Region 871.1	arylpolyene	1	1,231			
		Region 1048.1	arylpolyene	1	1,099			
SAR324_d03	SAR324	Region 65.1	T3PKS	1	15,028			
		Region 71.1	arylpolyene	1	13,956	APE Vf	Other	20%

		Region 87.1	hglE-KS,PUFA	1	11,989	heterocyst glycolipids	Other	28%
Verru_d01	Verrucomicrobiota	Region 153.1	terpene	1	2,693			
		Region 293.1	terpene	1	2,114			
		Region 332.1	terpene	1	2,004			
		Region 1273.1	terpene	1	1,029			
2960m winter								
Bacte_d02	Bacteroidota	Region 10.1	terpene	1	11,223			
Dehal_d16	Chloroflexota	Region 133.1	terpene	1	4,557			
Dehal_d17	Chloroflexota	Region 15.1	other	1	17,248			
		Region 21.1	terpene	1	15,476			
		Region 151.1	terpene	1	6,979			
Gamma_d25	Proteobacteria	Region 216.1	terpene	1	3,817			
Gamma_d26	Proteobacteria	Region 14.1	ectoine	1	9,007			
		Region 16.1	betalactone	1	21,551			
		Region 69.1	siderophore	1	13,192	xanthoferrin	Other	28%
		Region 86.1	redox-cofactor	1	12,068	lankacidin C	NRP + Polyketide	13%
		Region 238.1	betalactone	1	5,310	fengycin	NRP	13%
		Region 465.1	betalactone	1	2,527			
Gamma_d28	Proteobacteria	Region 5.1	phosphonate	17,896	48,300			
Gemma_d06	Gemmatimonadota	Region 206.1	hglE-KS,T1PKS	1	6,323			
		Region 229.1	ranthipeptide	1	5,626			
Posei_d16	Thermoplasmatota	Region 283.1	terpene	1	1,903			
		Region 387.1	hglE-KS	1	1,551			
Posei_d17	Thermoplasmatota	Region 143.1	terpene	1	2,314			
Posei_d18	Thermoplasmatota	Region 26.1	terpene	1	20,648			
Posei_d19	Thermoplasmatota	Region 156.1	terpene	1	2,801			
		Region 424.1	terpene	1	1,589			
		Region 614.1	hglE-KS	1	1,230			
Posei_d20	Thermoplasmatota	Region 5.1	terpene	27,172	42,624			

		Region 33.1	T1PKS	1	16,443			
		Region 64.1	resorcinol	1	9,188			
SAR324_d09	SAR324	Region 3.1	T3PKS	10,842	33,584			
		Region 42.1	PUFA,T1PKS	1	15,056			
		Region 210.1	hglE-KS	1	3,881			
SAR324_d10	SAR324	Region 24.1	T3PKS	1	22,125			
		Region 70.1	arylpolyene	1	14,071	APE Vf	Other	25%
		Region 76.1	PUFA,T1PKS	1	13,513			
		Region 140.1	hglE-KS	1	6,699			
Verru_d04	Verrucomicrobiota	Region 40.1	terpene	1	16,291			
		Region 215.1	terpene	1	3,631			
Verru_d05	Verrucomicrobiota	Region 60.1	T3PKS	1	17,132			
		Region 68.1	terpene	1	16,075			
		Region 77.1	terpene	1	15,804			

Appendix 12

Table S9: The SO KS domain similarity to NaPDoS KS domain

Query id	Database match id	percent identity	align length	e-value	pathway product	domain class

SAR324_507_s1	PfaC_Shewanella_PUFA	50	455	2.00E-123	polyunsaturated fatty acid	PUFA
SAR324_507_s2	PfaA_Shewanella_PUFA	28	459	6.00E-46	polyunsaturated fatty acid	PUFA
Gemma_d01_s2	PfaC_Shewanella_PUFA	48	274	5.00E-62	polyunsaturated fatty acid	PUFA
Gemma_d01_s2	PfaA_Shewanella_PUFA	36	293	1.00E-40	polyunsaturated fatty acid	PUFA
SAR324_d01_s1	PfaC_Shewanella_PUFA	50	462	5.00E-126	polyunsaturated fatty acid	PUFA
SAR324_d01_s2	PfaA_Shewanella_PUFA	28	459	5.00E-46	polyunsaturated fatty acid	PUFA
SAR324_214_w1	PfaC_Shewanella_PUFA	50	455	2.00E-123	polyunsaturated fatty acid	PUFA
SAR324_214_w2	PfaA_Shewanella_PUFA	28	459	8.00E-46	polyunsaturated fatty acid	PUFA
Verru_217_w	PfaA_Shewanella_PUFA	33	288	5.00E-38	polyunsaturated fatty acid	PUFA
Verru_217_w	PfaA_Shewanella_PUFA	47	60	1.00E-12	polyunsaturated fatty acid	PUFA
SAR_324_520_w1	PfaC_Shewanella_PUFA	52	413	7.00E-117	polyunsaturated fatty acid	PUFA
SAR324_520_w2	PfaA_Shewanella_PUFA	28	459	2.00E-44	polyunsaturated fatty acid	PUFA
SAR324_d09_w1	PfaC_Shewanella_PUFA	50	462	5.00E-126	polyunsaturated fatty acid	PUFA
SAR324_d09_w2	PfaA_Shewanella_PUFA	28	459	3.00E-46	polyunsaturated fatty acid	PUFA
Gemma_d06_w1	PfaC_Shewanella_PUFA	41	375	2.00E-65	polyunsaturated fatty acid	PUFA
Gemma_d06_w2	PfaA_Shewanella_PUFA	36	293	1.00E-40	polyunsaturated fatty acid	PUFA
Posei_d20_w1	ArsA_Azotobacter_PUFA	41	461	4.00E-76	alkylresorcinol	PUFA
Posei_d20_w2	AlnL_ACI88861_KSa	26	434	2.00E-16	alnumycin	typeII
SAR324_d10_w1	PfaC_Shewanella_PUFA	49	462	1.00E-118	polyunsaturated fatty acid	PUFA

SAR324_d10_w2	PfaA_Shewanella_PUFA	28	459	5.00E-45	polyunsaturated fatty acid	PUFA