UNIVERSITY OF SAO PAULO

SCHOOL OF PHARMACEUTICAL SCIENCES OF RIBEIRAO PRETO

Bioinformatics and biogeography to mine natural products in metagenomes

Bioinformática e biogeografia para buscar produtos naturais em metagenomas

ULYSSES AMANCIO DE FRIAS

Ribeirão Preto

2017

UNIVERSITY OF SAO PAULO

SCHOOL OF PHARMACEUTICAL SCIENCES OF RIBEIRAO PRETO

Bioinformatics and biogeography to mine natural products in metagenomes

Bioinformática e biogeografia para buscar produtos naturais em metagenomas

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas da Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP para obtenção do Título de Doutor em Ciências

Área de Concentração: Produtos Naturais e Sintéticos.

Dissertation submitted to the Pharmaceutical Sciences Graduate Program in partial fulfillment of the requirements for the degree of Doctor in Phylosophy.

Field of Study: Natural and synthetic products

PhD candidate: Ulysses Amancio de Frias

Advisor: Profa. Dra. Monica Tallarico Pupo

Revised version of the dissertation submitted to the Pharmaceutical Sciences Graduate Program at 12/14/2017. The original version is available at the School of Pharmaceutical Sciences of Ribeirão Preto/USP.

Ribeirão Preto

2017

De Frias, Ulysses Amancio.

Bioinformatics and biogeography to mine natural products in metagenomes. Ribeirão Preto, 2017.

123 p. : il. ; 30cm.

Dissertation submitted to School of Pharmaceutical Sciences of Ribeirão Preto / USP – Field of Study: Natural and synthetic products.

Advisor: Pupo, Mônica Tallarico.

1. Metagenômica. 2. Produtos naturais. 3. Bioinformática.

DISSERTATION APPROVAL SHEET

Ulysses Amancio de Frias
Bioinformatics and biogeography to mine natural products in metagenomes

Dissertation submitted to the Pharmaceutical Sciences Graduate Program in partial fulfillment of the requirements for the degree of Doctor in Philosophy.

Field of Study: Natural and synthetic products

Advisor: Profa. Dra. Mônica Tallarico Pupo

Date Approved:

Committee

Prof. Dr. _____

Institution: _____ Signature: _____

Prof. Dr. _____

Institution: _____ Signature: _____

Prof. Dr. _____

Institution: _____ Signature: _____

Prof. Dr. _____

Institution: _____ Signature: _____

Prof. Dr. _____

Institution: _____ Signature: _____

This thesis is dedicated to

My family, friends, and colleagues.

In loving memory of José Amancio and Alcebiades Frias

Acknowledgments

I am exceptionally grateful to Dr. Monica Tallarico Pupo for her perseverance, leadership, honestly and fairness. I am extremely pleased for you provide the best environment to conduct my Ph.D. I am very grateful.

I would like to acknowledge Dr. Sean F. Brady for the time I spent in his lab, for his encouragement and the opportunities he has granted to me at Rockefeller University.

I would like to offer special thanks to Dr. Sergio Akira Uyemura, who opened his laboratory for this thesis. You have remarkably contributed to this research.

I must thank Dr. Denise Oliveira Guimaraes. Her thesis inspired me many years ago an initial knowledge of the Natural Products research.

I must acknowledge the past members of the LQMo Fernanda Chagas, Mauricio Caraballo, Eduardo Afonso, Camila Paludo, Raphael Conti, Larissa Varella, Marília Almeida, Mayra Tuiche, Adriana Lopes, Gisele Messiano, Ana Ligia Oliveira, Rita Pessotti, Isabela Innocentini, Bárbara Matos, and the present members Carla Menegatti, Gabriela Toninato, Vitor Lourenzon, Marina Cusinato, Marina Geraldi, Humberto Ortega, Weilan Paixão, Taise Fukuda, Camira Fiori, Diego Rodriguez. I acknowledge all of you for making the laboratory an exciting experience. I would like to give particular appreciate to Claudia Macedo for her extraordinary assistance - "The Clau."

I must thank Dr. Zachary Charlop-Powers for all guidance and the attention during the time that I spent at Rockefeller University. I want to acknowledge all Brady lab members. I also thank Paula Calle for her support.

I would like to thank the School of Pharmaceutical Sciences of Ribeirao Preto - University of Sao Paulo and all professors and employees. A special acknowledge to Eleni and Rafael for all support across these years.

I gratefully acknowledge the Sao Paulo Research Foundation (FAPESP) for the financial support during my research in Brazil (2011/50869-8, 2012/16139-5) as well as in United States (2014/22877-4). I would like to thank the CEPID-CIBFAR (2013/07600-3), the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq).

Of course I am extremely grateful to my wife and family, who have supported me through everything in life and academia.

I am exceptionally grateful to all of you.

“I have not failed. I've just found 10,000 ways that won't work”

Thomas A. Edison

“Every carbon atom in every living thing on the planet was produced in the heart of a dying star”

Brian Cox

“Science is a way of thinking much more than it is a body of knowledge”

Carl Sagan

RESUMO

FRIAS, U. A. **Bioinformática e biogeografia para buscar produtos naturais em metagenomas.** 2017. 123 p. Tese de doutorado, Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2017.

Os produtos naturais microbianos (NP) tem demonstrado ser inestimáveis pontos de partida na descoberta e desenvolvimento de medicamentos aprovados pelo FDA. A abordagem tradicional para a identificação de produtos naturais microbianos exige a cultura em laboratório. Infelizmente, os métodos convencionais baseados nesta metodologia foram desestimulados devido a altas taxas de redescoberta de moléculas. Os métodos independentes de cultura que se baseiam no sequenciamento do metagenoma microbiano sugerem a ocorrência de um enorme reservatório inexplorado de *clusters* biossintéticos de produtos naturais (BGCs) no meio ambiente. Neste trabalho utilizamos uma metodologia baseada em PCR e *barcoding amplicon*-sequencing para buscar importantes famílias de produtos naturais como peptídeos não ribossomais (NRP), ácido 3-amino-5-hidroxibenzóico (AHBA), dímeros de triptofano (TD), policetídeos, aminoglicosídeos e outros. Para isto desenvolvemos um *script* chamado SecMetPrimer que nos permitiu bioinformaticamente desenhar conjuntos de *primers* contendo um gradiente de degenerâncias. No total, desenhamos 165 conjuntos de *primers.* Os *amplicons* foram obtidos por PCR padrão, tendo sido concatenados *barcodes* específicos por amostra e sequenciados através de Illumina MiSeq. Para validar, utilizamos eDNA (*environmental* DNA) de bibliotecas metagenômicas, totalizando 223 milhões de clones. Através das análises bioinformáticas, as curvas de rarefação foram calculadas e a diversidade para cada família foi determinada. Foi realizada uma re-amplificação dos domínios de adenilação de peptídeo não ribossomal e domínios de cetosintase de policetídeos utilizando eDNA isolado de 25 amostras diferentes coletadas em Mata Atlântica, Cerrado e ambiente marinho. Nossos dados indicaram a correlação entre distância geográfica e o tipo ecológico dos biomas. Deste modo, foi possível assim atribuir genes relacionados à *clusters* biossintéticos que codificam importantes produtos naturais à informações taxonômicas e metabólicas. Deste modo identificamos os melhores *hotspots* para busca de diversidade biossintética dentre as amostras analisadas.

Palavras-chave: Produtos naturais, Metagenômica, Bioinformática

ABSTRACT

FRIAS, U. A. **Bioinformatics and biogeography to mine natural products in metagenomes.** 2017. 123 p. Doctoral Dissertation. School of Pharmaceutical Sciences of Ribeirão Preto – University of Sao Paulo, Ribeirão Preto, 2017.

Microbial natural products (NP) have proven to be invaluable starting points in the discovery and development of many drugs approved by FDA. The traditional approach to identify microbial natural products requires the culturing in the laboratory. Unfortunately, conventional culture-based methods have been deemphasized due to high rediscovery rates. Culture-independent methods applying microbial (meta)genome sequencing suggest the occurrence of an enormous untapped reservoir of natural-product-encoding biosynthetic gene clusters (BGCs) in the environment. Here we have used a PCR-based approach and barcoding amplicon-sequencing derived from important families of microbial natural products such as nonribosomal peptides (NRP), polyketides (PK), 3-amino-5-hydroxybenzoic acid-containing NPs (AHBA), tryptophan dimmers (TD), aminoglycosides, phosphono-containing NPs and others. We have written an internal script called SecMetPrimer that allowed us to bioinformatically design sets of primers containing a range of degeneracy to amplify these genes. At the total, we designed 165 different sets of primers. The amplicons were obtained by standard PCR containing double-barcoded-target primers and sequenced by Illumina MiSeq platform. The validation process was conducted using eDNA from metagenomic libraries containing a 223 millions of clones. The rarefaction and diversity analyses were assigned, and the best-hit primer for each family was chosen. We have re-amplified the nonribosomal peptide adenylation domains and polyketide ketosynthase domains, using as substrate environmental DNA isolated from 25 different samples collected in Atlantic Forest, Cerrado and marine environment. Our data indicate a correlation between geographic distance and biome-type, and the biosynthetic diversity found in these environments. Thus, by assigning reads to known BGCs against taxonomic and metabolic profiles, we have identified the hotspots of relevant biosynthetic diversity among the analyzed samples.

Keywords: Natural Products, Metagenomics, Bioinformatics

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

2-DOI	2-Deoxy streptamine scyllo-inosose synthase
A	Adenine
ACP	Acyl carrier protein
AD	Adenylation
afa	Aligned FASTA
AHBA	3-amino-5-hydroxy benzoic acid
Anti-Smash	Antibiotic and secondary metabolite analysis shell
AT	Acyltransferase
ATP	Adenosine triphosphate
B	Not A
BGC	Biosynthetic gene cluster
Co	Condesation
C	Cytosine
CAPES	Coordination for the Improvement of Higher Education Personnel
CNPq	National Council for Scientific and Technological Development
CPA	Chromopyrrolic acid
CTAB	Cetyl trimethylammonium bromide
D	Not C
DFD	Drugs from dirt
DNA	Deoxyribonucleic acid
EC	*Escherichia coli*
eDNA	Environmental DNA
EDTA	Ethylenediamine tetraacetic acid
FAPESP	São Paulo Research Foundation
FDA	Food and drug admnistration
G	Guanine
GPA	Glycopeptide antibiotic
H	Not G
ID	Identification
K	Keto
kb	kilo base
Km	Kilometers
KS	Ketosynthase
KSα	alpha Ketoacyl synthase
KSβ	beta Ketoacyl synthase
LPA	Lipopeptide antibiotic
LQMo	Laboratório de química de Micro-organismos
M	Amino
MG-RAST	Metagenomics rapid annotation using subsystem technology
MiBig	Minimum information about a biosynthetic gene cluster
mM	milli Molar
MO	Microorganism
MSA	Multiple sequence alignment

N	all Nucleotides
NJ	Neibor-Joininng
NMDS	Non-metric multidimentional scaling
NP	Natural products
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
OTU	Operational taxonomy units
PCA	Principal component analysis
PCR	Polymerase chain reaction
PK	Polyketide
PKS	Polyketide synthase
R	Purine
R	Environment for statistical computing and graphics
RNA	Ribonucleic acid
S	Strong
SDS	Sodium dodecyl sulfate
T	Thymine
TAR	Trans-activation response element
TD	Tryptophan dimer
U	Uracil
µL	microliter
µM	micromolar
UV	Ultraviolet
V	Not T
W	Weak
Y	Pyrimidine
Z	Zero

TABLE OF CONTENTS

INTRODUCTION

1 Introduction

1.1 The Importance of Natural Products

During most of the historic records, small molecule natural products have been a source of innovative therapeutics agents against a broad spectrum of diseases. Humans have used plant extracts for therapeutic purpose for millennia (Handelsman *et al.*, 1998; Ji *et al.*, 2009). Since the early 1800s, the identification of natural products has risen a brand-new era of medicine, and drugs started to be isolated from plants, and microorganisms (Newman e Cragg, 2016). The term natural product literally proposes any naturally occurring molecule but is commonly considered to mean a secondary metabolite, a molecule that has apparently no fundamental function in the metabolism (Williams *et al.*, 1989).

It is believed that molecules exist to enhance the reasonableness of an organism's survival by attracting or repulsing other organisms. The initial reports on natural products include strychnine, morphine, atropine, quinine, and colchicine (**Figure 1**) (Hosztafi, 1997; Cordell *et al.*, 2001; Corson e Crews, 2007; Zenk e Juenger, 2007; Kaiser, 2008; Cragg e Newman, 2014). In 1826, E. Merck revealed the first business natural product, morphine (Sertuerner, 1817) and this was succeeded by the first semisynthetic, pure drug based on a natural product, aspirin, by Bayer in 1899 (**Figure 1**) (Sneader, 2000). The next breakthrough was the identification of penicillin from the filamentous fungus *Penicillium notatum* by Alexander Fleming in 1929. This unquestionably led in the "Golden Age of Antibiotics", the period from the 1940s to the 1970s (Li e Vederas, 2009; Newman e Cragg, 2016).

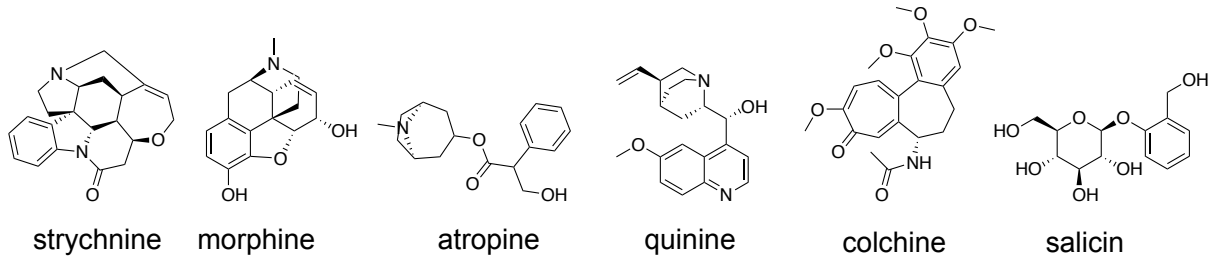


Figure 1: Natural Products breakthrough isolated from plants

Pharmaceutical companies started extensive studies of microbes as sources of novel antibiotics and led to the identification of a host of different antibacterial compounds, including the tetracyclines (e.g., doxycycline), cephalosporins, aminoglycosides (e.g., streptomycin), lipopeptides (e.g., daptomycin), glycopeptides (e.g., vancomycin), and macrocyclic compounds such as erythromycin (Newman *et al.*, 2000; Cragg e Newman, 2001). Microbial natural products have also been the source of different anti-infective drugs having antifungal (e.g., amphotericin, nystatin) and antiparasitic, (e.g., ivermectin, fumagillin) activities (**Figure 2**) (Newman *et al.*, 2000; Cragg e Newman, 2001; Newman, David J. e Cragg, Gordon M., 2012).

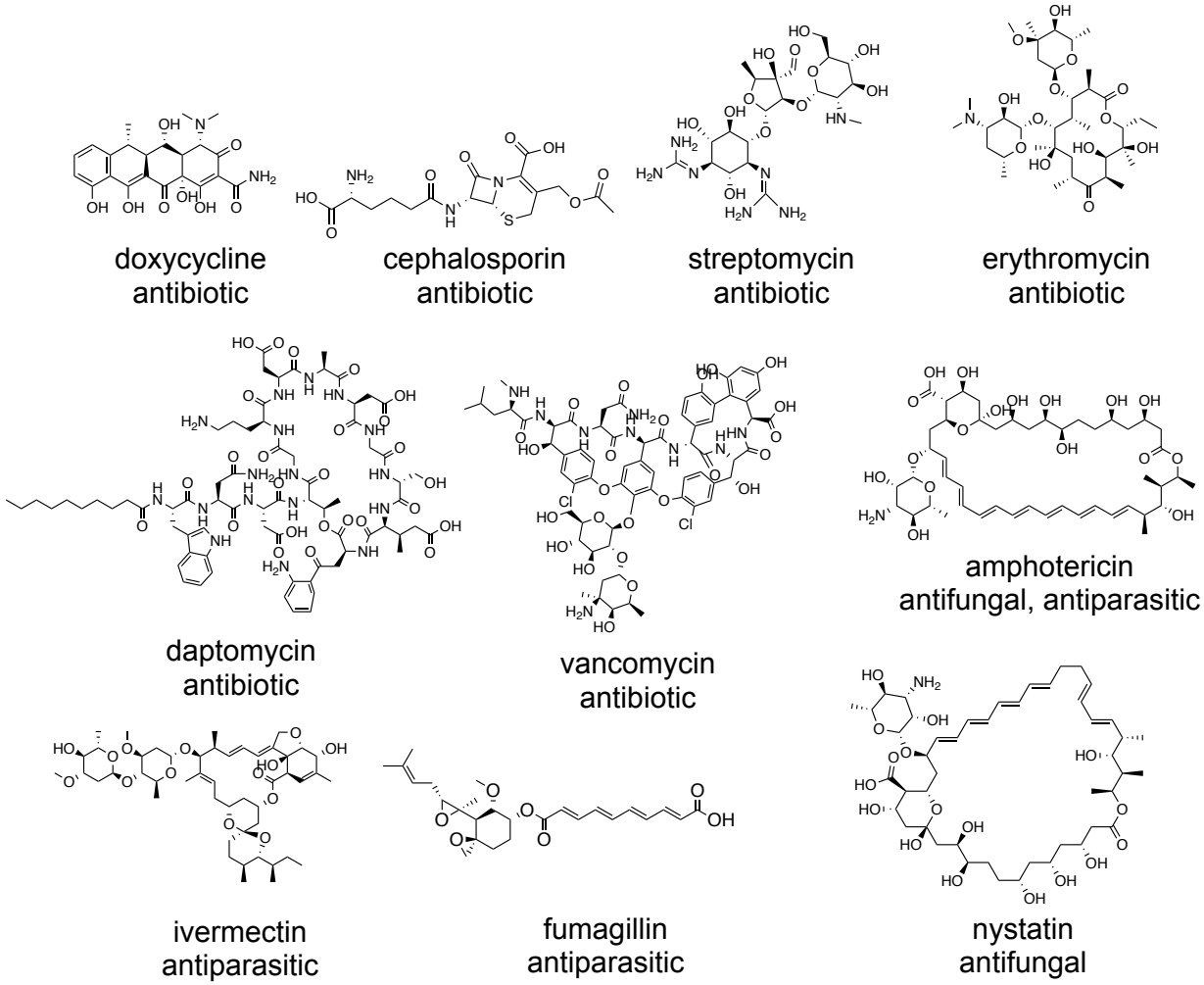
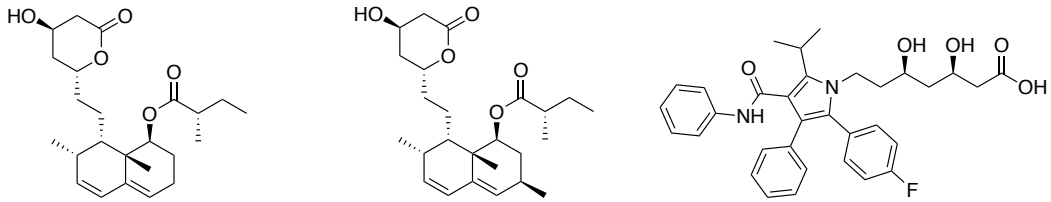


Figure 2: Natural Products as therapeutic agents

In this way, microbes have given the template for the development of anticholesterolemic drugs, such as statins (Cragg e Newman, 2014). The determination of the mevastatin (compactin) and lovastatin as inhibitors of HMG-CoA reductase led to the progress of synthetic statin analogs, such as atorvastatin (**Figure 3**).



mevastatin (compactin) lovastatin atorvastatin

Figure 3: Natural and synthetic anticholesteromic drugs

The antitumor antibiotics families of anthracyclines (e.g., doxorubicin), enediynes (e.g., calicheamicin), ansamycins (e.g., geldanamycin), peptolides (e.g., dactinomycin), epothilones (e.g., ixabepilone), mitosanes (e.g., mitomycin C), are microbial natural products that represent meaningful role in the drug discovery (**Figure 4**) (Cragg *et al.*, 2009; Cragg *et al.*, 2012; Newman, D. J. e Cragg, G. M., 2012; Newman e Cragg, 2016). The Type I polyketide rapamycin are another relevant class of immunosuppressive drugs that complement the performance of another immunosuppressant microbial natural product, cyclosporine (Cragg e Newman, 2001).

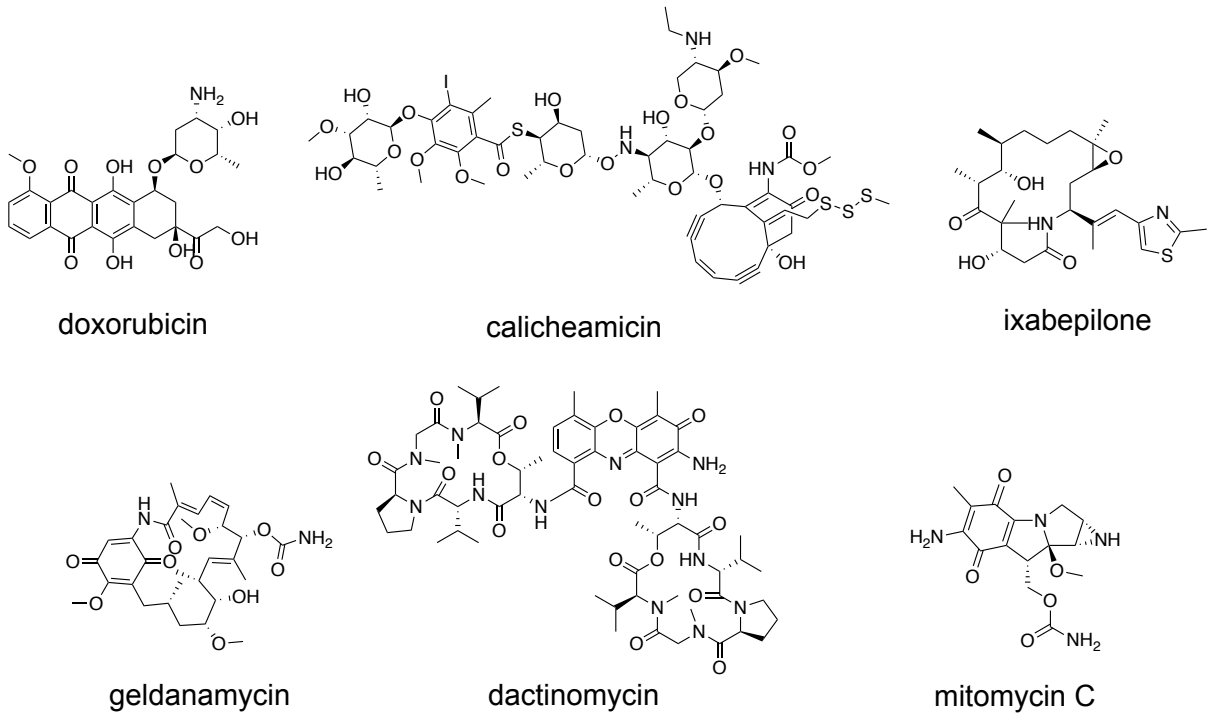


Figure 4: Therapeutic natural products isolated from bacteria

Thus, it is calculated that approximately 70% of all small molecule drugs have the origin in natural products (**Figure 5**) (Newman e Cragg, 2016). Therefore, natural products have been, and remain to be, an essential source of innovative drugs and pharmaceutical leads.

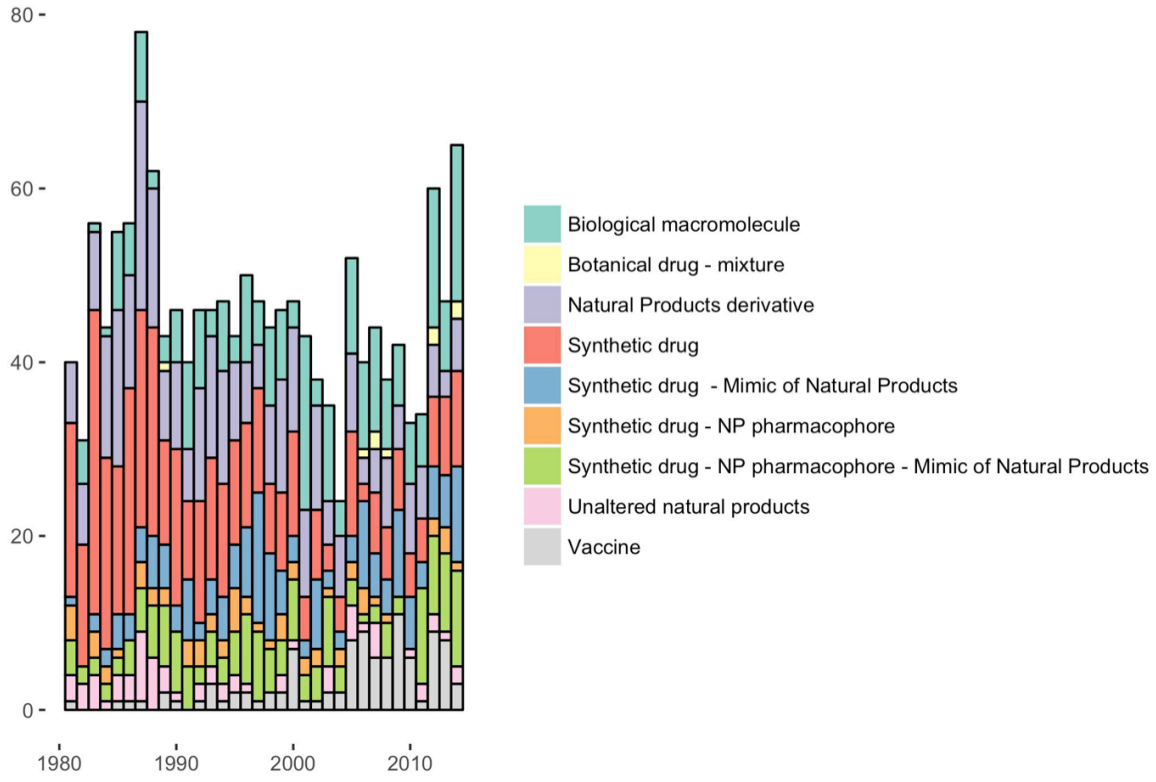


Figure 5: Distribution of 1562 drugs approved by FDA the during 1980 - 2014
This graphic represents the correlation of the amount of drugs approved by FDA (x-axis) by year (y-axis). The colors in legend represent drugs class. This data is available in reference - Newman and Cragg (2016).

Natural products isolated from cultivated microorganisms have given the large dimension of the most significant pharmacophores discovered to date (Newman e Cragg, 2016). Their bioactivity is associated, to a considerable amount, to selective pressures that crowned in the evolution of structural and chemical features of natural products to perform their function efficiently in biological contexts (Maplestone *et al.*, 1992). Presumably, in the natural microbial ecosystem the metabolites have been revealed to perform crucial functions in processes as self-defense, nutrient scavenging, and virulence (Omura *et al.*, 2001; Wolfgang *et al.*, 2003; Nougayrede *et al.*, 2006; Wyatt *et al.*, 2010). Soil-dwelling bacteria, particularly actinomycetes, have been an overflowing reservoir of bioactive natural products. Admittedly, a predicted two-thirds of all clinically valuable antibiotics were isolated from one bacterial genus singly, *Streptomyces* (Kieser et al. 2000). A single gram of soil is predicted to harbor

up 10,000 unique bacterial species. Consequently, soil-bacteria appear to represent an encouraging reservoir of new natural products with broad therapeutic potential (Torsvik et al. 1990; Rappe and Giovannoni 2003; Charlop-Powers, Owen*, et al.*, 2014; Charlop-Powers *et al.*, 2015). There are hundreds of thousands of different natural products reported in literature. The CRC Dictionary of Natural Products (DNP) lists 190,939 records (DNP, Chapman and Hall, 2005). Regrettably, there is no rigidly established design for classifying natural products. Their endless heterogeneity in structure, function, and biosynthesis is exceptionally vast to enable them to fit into a few simple categories.

1.2 Metagenomics and uncultured microorganisms

Mapping biologically significant chemical space is of highest interest for drug discovery and development. Nevertheless, the discovery of new natural product antimicrobials has decreased over the past decades (Fox, 2006; Banin *et al.*, 2017). Multiple factors are providing this decay. The isolation of small molecules frequently relies on the decade-old methods and techniques (**Figure 6**). The conventional protocols employed to isolate and identify natural products from microorganisms usually apply extraction of bacterial cultures using organic solvents, and isolation of the molecules by activity-guided fractionation (**Figure 6**). However, the conventional culture-based model appears to be one of the main obstacles to new natural product discovery. Notwithstanding decades of historical productivity, pharmaceutical companies have deemphasized natural product discovery applications due to the rising of rediscovery rates of known metabolites produced by cultured bacteria - a rate that now outstrips 99% (Newman e Cragg, 2016). The challenges associated

with culturing environmental bacteria limit traditional culture-dependent methods to be applied to a small fraction of the bacterial species in nature. The microbial 16S rRNA sequencing efforts showed that just a tiny fraction of the microbes had been cultured through standard microbiological techniques (Torsvik *et al.*, 1990; Kaeberlein *et al.*, 2002; Zengler *et al.*, 2002; Gans *et al.*, 2005; Tringe *et al.*, 2005). In the bacterial domain, this fraction is represented by less than 1% of all species in the environment.

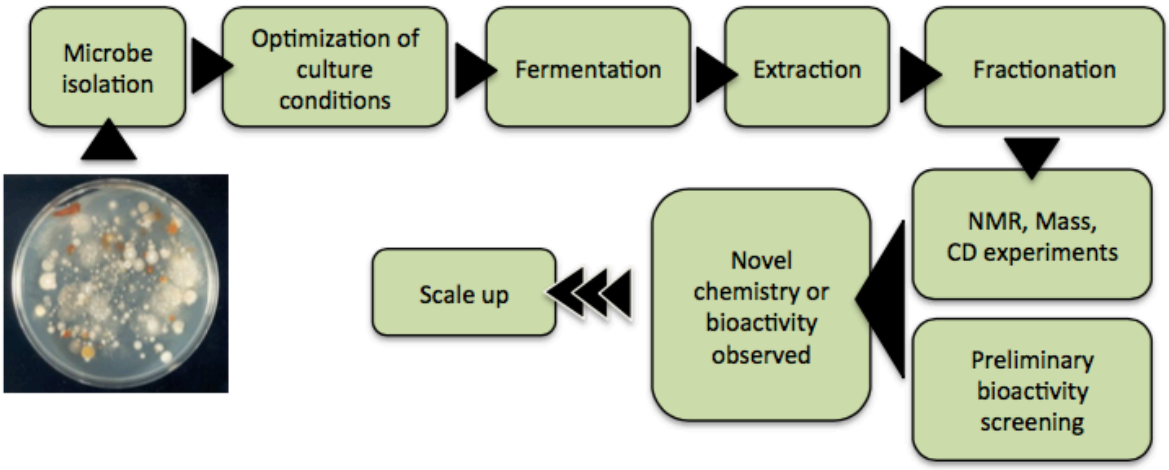


Figure 6: Traditional culture-based methods in microbial natural products isolation
Arrows guide the pipeline for isolation and characterization of natural products.
Methods could change across laboratories, however the general idea is maintained.

These likewise investigations comprehend that more than 80 dominant bacterial divisions exist, but cultured isolates less than half (Keller e Zengler, 2004; Schloss e Handelsman, 2004; Desantis *et al.*, 2006). New methods suggest that uncultured bacteria are apparent the biggest outstanding pool of genetic and chemical diversity on the planet. Hence, it is evident that the large preponderance of natural products molecules biosynthesized by microorganisms in the biosphere are exceeding the reach of the common culture-dependent model (Charlop-Powers *et al.*, 2015).

There are multiple strategies employing both culture-dependent and culture-independent methods, which are presently being extended to access this untapped reservoir of chemical diversity. Typically, culture-independent approaches require the cloning of the genetic material from an environmental sample (eDNA - environmental DNA), which carries the genes encoding small molecules (**Figure 7**). This genetic heterogeneity of group of the bacteria across given environmental sample has been termed the "metagenome" (Handelsman *et al.*, 1998). While genomics is the investigation of an individual organism's DNA, metagenomics transcends the single genome. Catching advantage of the progress in DNA-sequencing technology, this approach provides the investigation of whole populations of microorganisms simultaneously, circumventing the requirement for isolation and culture (Handelsman *et al.*, 1998).

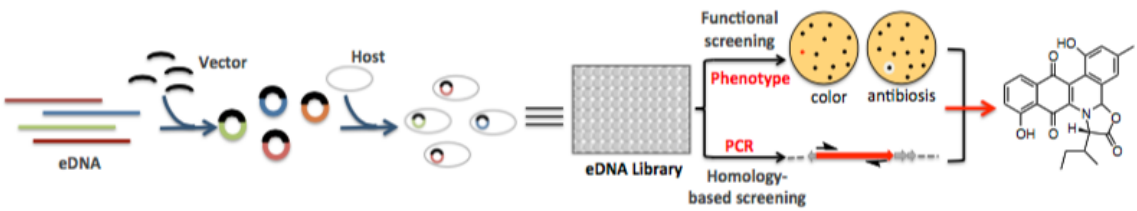


Figure 7: Metagenomic library construction and screening
The general idea is connect the eDNA with vectors using ligase and put inside a host (in general *E. coli*). The library is construct and screening by phenotypic screenings.

It consequently appears possible that we can investigate the biosynthetic potential of uncultured bacteria using metagenomics and obtains new ways to identify unknown bioactive metabolites for drug discovery applications. Metagenomics emerged fast as an option to standard microbial screening for natural product discovery. Investigating natural products biosynthetic gene clusters (BGCs) applying metagenomic approach requires eDNA (environmental DNA) cloning and

sequencing improvement techniques. The DNA obtained from an environmental sample, before-mentioned as soil and water, contains vast amounts of genetic material, including the secondary metabolite biosynthetic gene clusters.

DNA extracted from an environmental sample, such as soil and water, contains large quantities of genetic material, including secondary metabolite biosynthetic gene clusters. This eDNA is purified and cloned into a vector that provides the high-efficiency transformation of a tractable library host, typically in *Escherichia coli*. To perform this, eDNA can be ligated into a cosmid vector, packaged into lambda phage and transfected into *E. coli* (Brady, 2007). This cloning approach can capture up to 50 kb pieces of eDNA in a single clone. Considering that genes responsible for the biosynthesis of natural products are usually clustered collectively, it is possible to capture a complete biosynthetic gene cluster or a large part of it in a single cosmid (Handelsman *et al.*, 1998).

Some remarkable works have showed how the metagenomic approaches have been applied to microbial communities. As examples from soils/sediments (Rondon *et al.*, 2000; Brady *et al.*, 2002; Voget *et al.*, 2003), rumen gut (Brulc *et al.*, 2009), planktonic marine microbial associations (Beja *et al.*, 2000; Breitbart *et al.*, 2002), deep sea microbiome (Sogin *et al.*, 2006), acid mine site (Tyson *et al.*, 2004), arctic sediments and the Sargasso sea (Venter *et al.*, 2004).

The basis of all metagenomic strategies is the isolation and subsequent examination of DNA extracted directly from naturally occurring microbial communities, which bypasses the challenges associated with culturing environmental bacteria (Handelsman *et al.*, 1998). Metagenomics is expressly fascinating to natural product investigation because a high fraction of the gene clusters are under 100 kb.

Thus, become possible to capture the BGCs on a small number of eDNA clones and by heterologous expression obtain molecules (Handelsman *et al.*, 1998).

1.3 Functional Metagenomics

In functional metagenomic approach, eDNA libraries are analyzed in simple high throughput assays intended to recognize clones that have phenotypes correlated with the biosynthesis of some small molecules, such as pigment production, antibiosis, or altered colony morphology. One of the manageable approaches employed to detect eDNA clones that might provide small molecule antibiotics has been the screening of libraries hosted in *E. coli* for clones that generate inhibition zones against pathogenic microbes in top agar overlay assays.

Initial attempts to mine metagenomes looking for natural product began with the construction of eDNA cosmid libraries in *E. coli*. The expression of a new molecule was accompanied by a visual or chromatographic screening of these libraries looking for phenotypes commonly associated with natural product expression (e.g., color, antibiosis, HPLC peak). These simple screening produced some interesting metabolites as violacein, turbomycin A, palmitoylputrescine, a long-chain fatty acid enol ester, long-chain *N*-acyl tryptophan, long-chain *N*-acyl arginine (**Figure 8**) (Rondon *et al.*, 1999; Rondon *et al.*, 2000; Brady *et al.*, 2001; Courtois *et al.*, 2001; Macneil *et al.*, 2001; Brady *et al.*, 2002; 2004; Wang *et al.*, 2010; Feng *et al.*, 2011). However, they were less productive than initially expected.

Screening a complex metagenome library for natural product biosynthesis poses a severe challenge yielded the small fraction of clones that are expected to receive biosynthetic genes of interest. An approach to avoid this is the selectively enriching of metagenomic libraries for secondary metabolite biosynthetic machinery.

This has been illustrated by the functional screening of eDNA clones for phosphopantetheine transferase (PPTase) genes using a PPTase deficient *E. coli* strain (Charlop-Powers *et al.*, 2013). PPTases are responsible for activating nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) modules by post-translational attachment of phosphopantetheine (PPT). They are required for the function of NRPS and PKS gene clusters, including those which produce the secondary metabolite iron-chelators, siderophores, required for bacterial growth under iron-limiting conditions (Matzanke *et al.*, 1984). This method conduct to the isolation of vibrioferrin and enterobactin from metagenomics libraries (**Figure 8**).

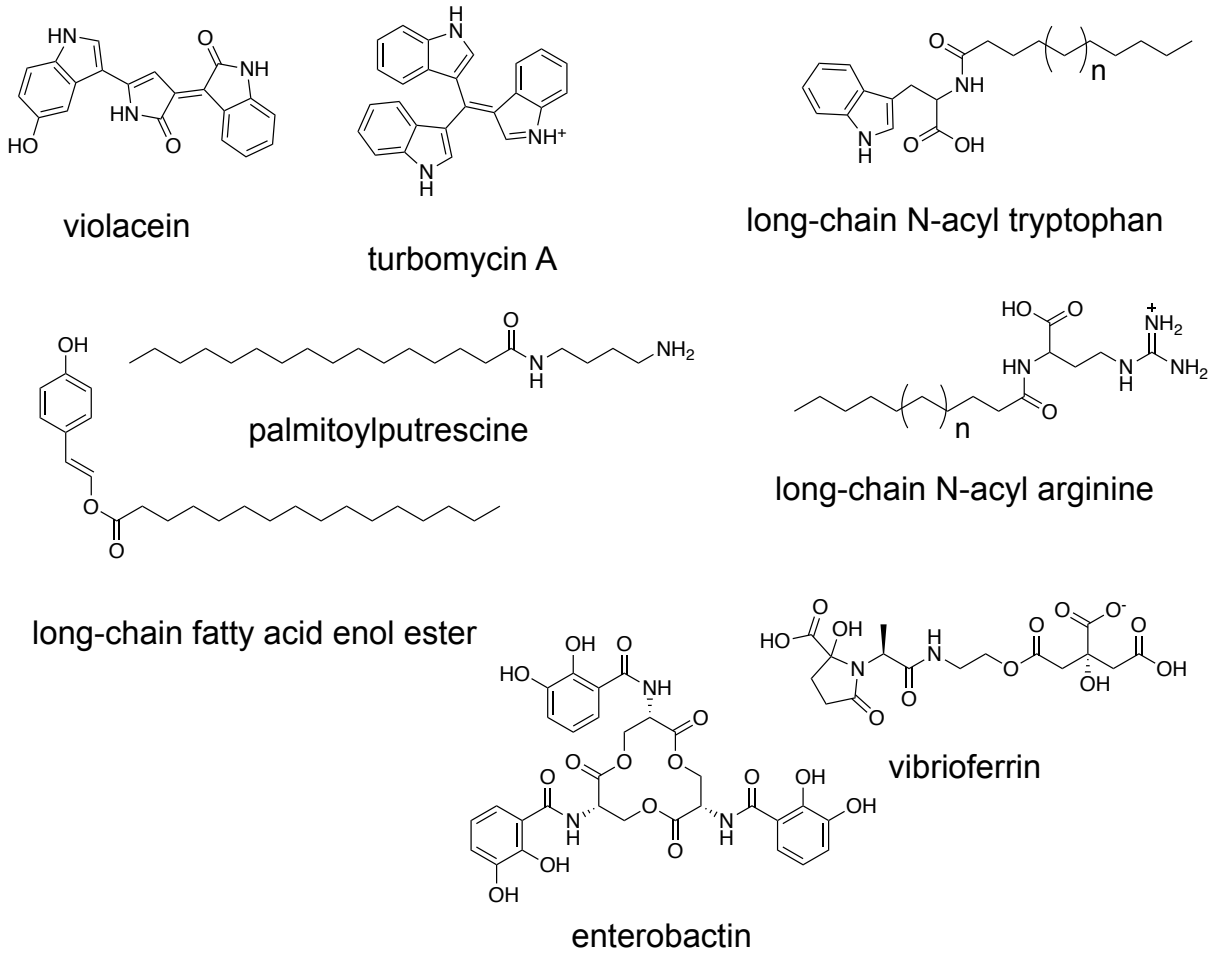


Figure 8: Natural products identified by functional metagenomic screening
This molecules were obtained from different libraries and screening using different host

This approach has led to the identification of novel longchain *N*-acylated amines, new isonitrile indole antibiotic (Brady e Clardy, 2000; 2005). Antibiotics have also been detected by monitoring pigmented eDNA clones, and also through the direct screening of fermentation extracts from randomly picked clones (Wang *et al.*, 2000; Brady *et al.*, 2001; Macneil *et al.*, 2001; Gillespie *et al.*, 2002; Lim *et al.*, 2005).

Examples of compounds with bioactivity recognized from these models of investigations include the antibiotic pigments violacein and indigo recovered from soil libraries (Chang *et al.*, 2015). The functional metagenomics has also been used to distinguish clones that produce proteins with potential anti-infective activities (Schipper *et al.*, 2009). Although almost all small molecules by functional metagenomic studies have been identified in *E. coli*, it is expected that the majority of the biosynthetic diversity existing in an environmental sample is not functionally accessible employing the same heterologous host. To investigate this difficulty, a computational approach of promoters and ribosomal binding sites used by a taxonomically diverse group of sequenced bacteria detected that at most, 40% of the enzymatic features present within a conventional metagenomic sample could be reached employing *E. coli* as a heterologous host (Gabor *et al.*, 2004). Consequently, libraries originally created in *E. coli* were later alternated into different hosts including *Streptomyces lividans*, *Ralstonia metallodurans*, *Rhizobium leguminosraum*, *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, and *Pseudomonas putida* (Wang *et al.*, 2000; Martinez *et al.*, 2004; Li *et al.*, 2005; Craig *et al.*, 2009; Craig *et al.*, 2010).

One primary advantage of a functional screening is that no a priori understanding of the biosynthetic enzymes is required to discover novel metabolites. In functional screenings, a secondary metabolite is immediately linked to its

biosynthetic genes providing the impartial discovery of novel biosynthetic sequences that have not been previously described. Although many novel metabolites have been revealed applying functional screenings, the expression of complete BGCs demands the coordinated composition of multiple proteins in a clone supporting laboratory culture conditions (Bentley *et al.*, 2002; Craig *et al.*, 2010; Garcia *et al.*, 2011; Milshteyn *et al.*, 2014). The biosynthetic gene clusters isolated from a metagenomic sample are of distinct phylogenetic origin. Thus, the plausibility that a biosynthetic gene cluster attends all of the requirements is very low. Consequently, the hit rates for functional screenings of metagenomic libraries are commonly around 0.01% (Courtois *et al.*, 2003; Williamson *et al.*, 2005; Guan *et al.*, 2007; Brady *et al.*, 2009). For these restrictions, functional screenings of metagenomic libraries have been planned so that they can be efficiently developed on a large number of clones (Katz *et al.*, 2016).

Although functional metagenomic screening is a powerful method to access to the chemical diversity encoded in uncultured ecosystems, the approach still needs optimization. As explained, metagenomic cloning relies fundamentally on cosmid-based vectors. However, multiple canonical biosynthetic gene clusters are too large (>50 kb) to be carried on single cosmids and are trimmed. This special barrier performs a significant restriction of using functional assays to identify new natural products from cosmid-based eDNA libraries. Apart from this, the multifariousness present in environmental samples performs the adoption of an optimal heterologous screening host considerably challenging (Torsvik e Ovreas, 2002). Heterologous expression hosts are essentially limited in their capacity to functionally process foreign DNA. The extended use of phylogenetically diverse heterologous expression hosts will continue to benefit functional screening purposes. Searching vector-host

sets that support the screening of metagenomic libraries in phylogenetically diverse microorganisms has the potential to increase the chemical diversity within functional metagenomic approaches (Katz *et al.*, 2016).

1.4 Homology-based screening for Natural Products Discovery

The expression-independent or homology-based screening of metagenomes is based on PCR amplification of conserved natural product biosynthetic gene sequences to identify and recover gene clusters from eDNA libraries. This approach enables targeted recovery of precise biosynthetic pathways from the genomes of all bacterial species present within a metagenome.

Homology-based screens rely on the similarity of the unknown, eDNA-derived secondary metabolite biosynthetic pathways to sequenced clusters of known metabolites (Milshteyn *et al.*, 2014). In these studies, the eDNA libraries are investigated to distinguish clones that contain conserved sequences correlated with a specific group of secondary metabolite biosynthetic gene. Bioinformatics analyses of the sequences obtained by homology screening provide ways to eliminate sequences associated with known gene clusters. Consequently, homology-based screening associated with phylogenetic analyses of sequences obtained in the eDNA libraries permit the discovery of biosynthetic gene clusters encoding for metabolites related to rare families of bioactive natural products (Kang e Brady, 2014; Milshteyn *et al.*, 2014).

Homology-based screening utilizes degenerate oligonucleotides to amplify homologs sequence via the polymerase chain reaction (PCR) (Seow *et al.*, 1997). Degenerate PCR primers are designed from beforehand sequenced conserved

regions of a biosynthetic gene of interest. The screening of metagenomic libraries with these degenerate primers permits the identification and recovery of uncommon clones with homologous genes from the eDNA library (King *et al.*, 2009). Those models of screening can be applied to detect new derivatives of known metabolites by targeting pathway-specific BGCs. These primers can additionally be designed to identify novel structures by targeting conserved universal biosynthetic genes observed in a class of compounds. As illustration, degenerate primers for amplifying conserved regions of minimal polyketide (PK) ketoacylsynthase alfa (KSα) gene are used to amplify KSα sequences obtained in the Texas eDNA library (King *et al.*, 2009). KSα genes eDNA-derived were applied as probes to identify and recover the cosmid clones containing type II PKS biosynthetic gene clusters from the metagenomic library (**Figure 9**).

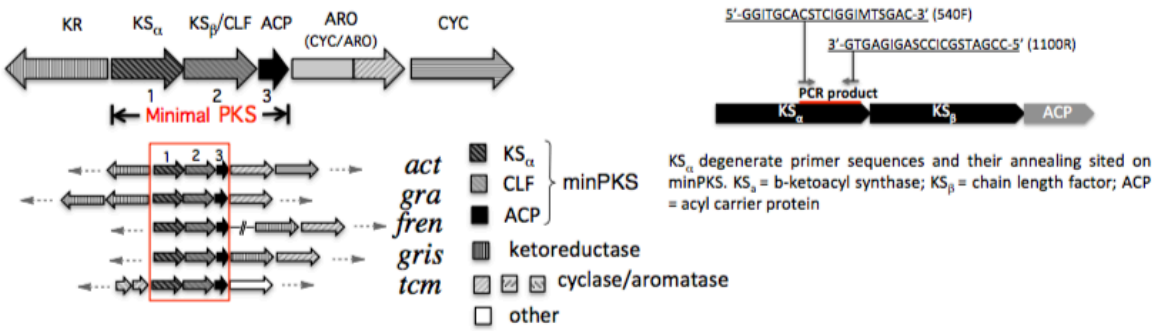


Figure 9: PCR-based screening for novel natural products
The type II PKS here described show an example of natural products gene cluster and how the PCR approach could be used to design primers targeting conserved genes.

The homology-based metagenomics screening has been proving to be an attractive approach to access the structural diversity encoded by uncultured bacteria. It has been successfully used to detect novel structures by targeting conserved broad biosynthetic genes observed in a class of molecules. One of the first groups of molecules targeted applying this approach was the iterative (type II, aromatic)

polyketides. Type II polyketides use a small number of biosynthetic enzymes to produce a distinct array of scaffolds, which are transformed by various downstream reactions (Zhang *et al.*, 2017). Although the biosynthetic gene clusters that encode the biosynthesis of this class of molecules can fundamentally differ in gene content, they all are encoded by a profoundly conserved minimal PKS formed of three proteins: ketosynthase alpha (KSα); ketosynthase beta/chain length factor (KSβ); and acyl carrier protein (ACP)(Robbins *et al.*, 2016). This “minimal PKS”, which is responsible for the iterative condensation of malonyl-CoA into a polyketide chain, maintains a conserved system and is consequently an excellent target to identify molecules using homology-based metagenomics (Hertweck, 2009).

As mentioned, BGCs of interest can be obtained from the particular pool of metagenomic libraries, sequenced, and analyzed in heterologous expression studies (Milshteyn *et al.*, 2014; Katz *et al.*, 2016). This concept was illustrated using a multi-million membered metagenomic library created using eDNA isolated from Arizona desert soil (Owen *et al.*, 2013). The screening utilized degenerate NRPS- and KS-targeting primers to identified tags concerning BGCs to encode congeners of a class of bioactive molecules. One BGC predicted a glycopeptide-like antibiotic and was obtained by heterologous expression using *Streptomyces toyocaensis*:ΔStaL as host. This provided the isolation of three novel glycopeptide congeners.

Strong promoters, such as the *erm*E* erythromycin resistance encoding promoter, can also be located in close of positive regulatory elements, or individual biosynthetic genes to activate the pathways (Laureti *et al.*, 2011; Kallifidas *et al.*, 2012; Luo *et al.*, 2013). Thus, resulting in the production of the secondary metabolites. This method has been used to produce 6-epialteramides, candicidins, antimycins (Luo *et al.*, 2013) and the 51-membered macrolide, stambomycin (Laureti

et al., 2011) and tetarimycin A (Kallifidas *et al.*, 2012) in a heterologous host. Some investigations have shown complete refactoring of biosynthetic pathways as a way of obtaining metabolites from cryptic cluster families, resulting in the heterologous expression of spectinabilin (Shao *et al.*, 2013) and several new polycyclic tetramate macrolactams (Luo *et al.*, 2013). Recently, yeast homologous recombination was used to production of the indolotryptoline and the antiproliferative agents, lazarimides A and B (Montiel *et al.*, 2015).

Using Trans-activation response element (TAR) and shuttled into *Streptomyces albus* for heterologous expression led the identification of seven novel epoxyketone protease inhibitors natural products: clarepoxins A-E and landepoxins A and B (Owen *et al.*, 2015). Some metagenomic discovery efforts were focused on the screening of eDNA libraries searching for Tryptophan dimmers (TD) BGCs using degenerate primers based on an alignment of chromopyrrolic acid synthase (CPAS) genes, which encode the enzymes involved in the dimerization of oxytryptophan (Chang e Brady, 2013; Chang *et al.*, 2013). These investigations produced the identification of some TD BGCs and the characterization of both known the antitumor molecule BE-54017 (Chang e Brady, 2011), and novel TDs (e.g., erdasporine (Chang, 2011), borregomycins (Chang e Brady, 2013). Some clusters were heterologously expressed to produce hydroxysporine: a pyrrolinone indolocarbazole core containing TD that had been described as a synthetic compound but never seen in nature. The second cluster, which was correlated with a phylogenetically novel CPAS sequence tag, encoded the new TD, reductasporine, which includes a new pyrrolinium indolocarbazole core (Nakano e Omura, 2009; Chang *et al.*, 2013). Thus, these investigation supports the premise that “outlier” sequence tags have a high potential of being associated with functionally novel gene clusters (Katz *et al.*, 2016).

This approach has led to the identification of a number of natural products with new or rarely ring systems such as arimetamycin A, tetarimycin A, arixanthomycin A, calixanthomycin A; UT-X26, AZ154 (King *et al.*, 2009; Feng *et al.*, 2011; Kallifidas *et al.*, 2012; Kang e Brady, 2013; 2014), further illustrating the utility of targeted screens in conjunction with a metagenomic natural product-mining pipeline to discover diverse new secondary metabolites (**Figure 10**).

The application of sequence‑based metagenomic natural product discovery allowed to use these metagenomic analyses to other environments. As example, the symbiotic bacteria found associated with organisms like marine sponges and tunicates have pointed to the characterization of diverse metabolites, such as the antitumor polyketides, bryostatins (Davidson e Haygood, 1999; Davidson *et al.*, 2001; Hildebrand *et al.*, 2004; Trindade *et al.*, 2015), onnamide (Hori *et al.*, 1993; Wilson *et al.*, 2014), and polytheonamides (Hamada *et al.*, 2010; Wilson *et al.*, 2014). In the case of the cytotoxic natural product calyculin A, it took a targeted sequence-based metagenomic method to identify the symbiotic organism and associated BGCs responsible for its production almost three decades after its initial discovery from extracts of the marine sponge *Discodermia calyx* (Kato *et al.*, 1986).

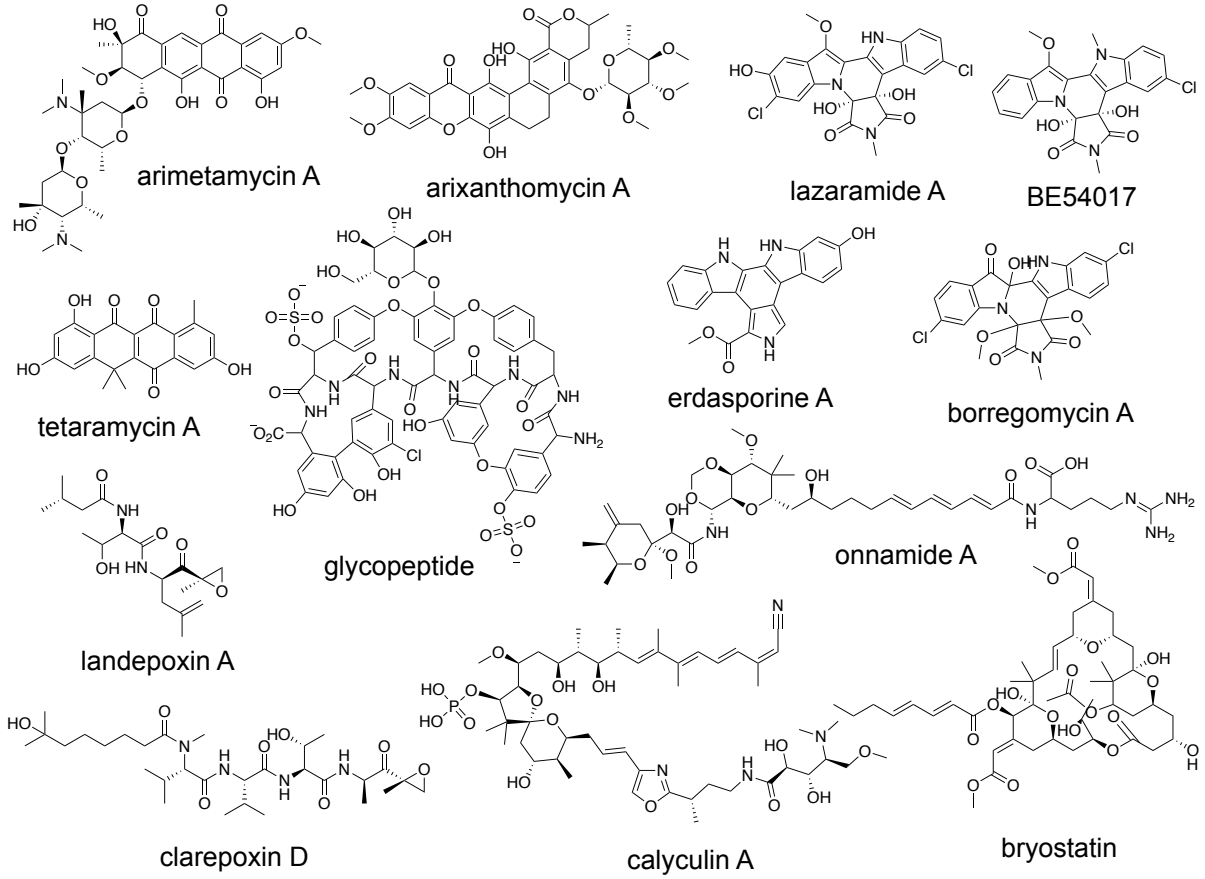


Figure 10: Some natural products molecules from homology screenings
The examples described were obtained from different metagenomic libraries. The homology were also detected using PCR-based approach

The analysis of eDNA libraries and metagenomic sequencing data for families of known BGCs is likely to be a valuable approach for recognizing new structural analogs of multiple bacterially derived antimicrobials, conceivably granting ready access to molecules with a refined spectrum of activity (**Figure 11**). Kang and Brady identified a collection of unique bioactive pentangular polyphenols, calixanthomycin A, arenimycins C-D, arixanthomycins A-C (Kang e Brady, 2013; 2014) through homology-based screening. Applying related strategy, the antibiotic fasamycins were isolated from soil eDNA and found to inhibit FabF of type II fatty acid biosynthesis (Feng *et al.*, 2011; Feng *et al.*, 2012). Expression of a pathway encoding the biosynthesis of a new erdacin derivative was identified with the unknown pentacyclic skeleton **(Figure 11)**. This confirmed the hypothesis that uncultured bacteria contains

a chemical diversity different from that detected in cultivated bacteria (King *et al.*, 2009).

A set of fluostatins and the methicillin-resistant *Staphylococcus aureus* (MRSA)-active antibiotic tetarimycin A have also been isolated from soil libraries (Feng *et al.*, 2010; Kallifidas *et al.*, 2012) (**Figure 11**). In extension to isolating type II polyketides, homology-based metagenomic screening has been fortunately used to detect novel tryptophan dimer analogs from eDNA libraries (Chang e Brady, 2011; Chang *et al.*, 2013; Chang e Brady, 2014; Chang *et al.*, 2015).

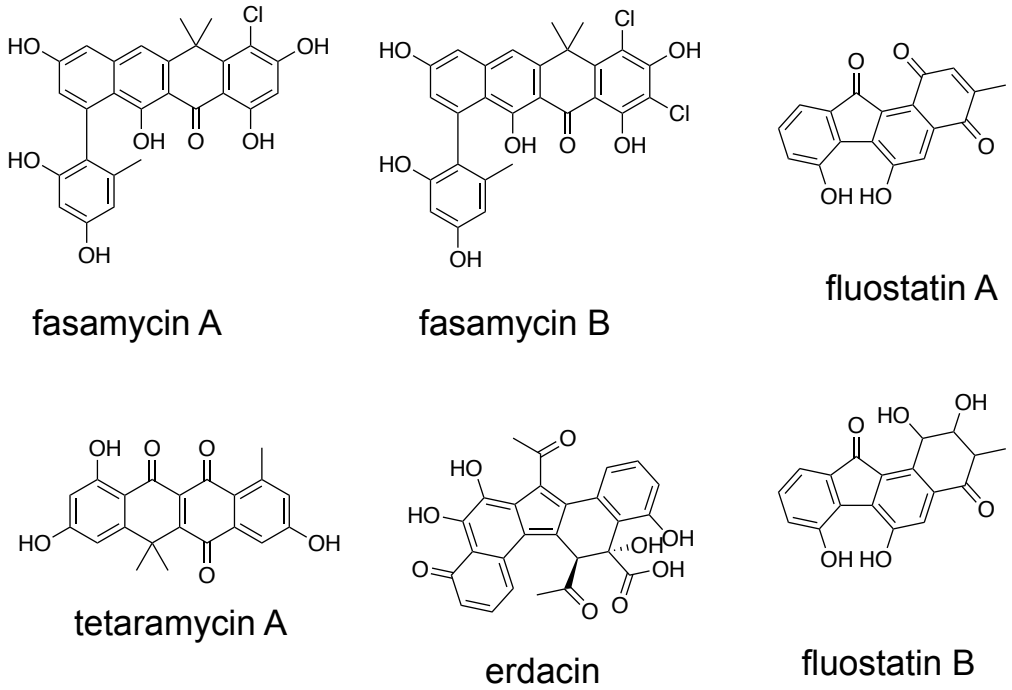


Figure 11: Metabolites isolated in homology-based metagenomic screening

Thus, homology-based metagenomic approaches have been used in a high number of studies to identify new derivatives of known compounds (Donia *et al.*, 2006; Ziemert *et al.*, 2010). Banik et al. applied a PCR screening targeting OxyC, a glycopeptide oxidative coupling enzyme, to detect new glycopeptide biosynthetic gene clusters from soil libraries (**Figure 10**) (Banik e Brady, 2008; Banik *et al.*, 2010). From this metagenomics library, novel mono, di- and trisulfated glycopeptide

congeners of teicoplanin aglycone were isolated (Figure 8) (Banik e Brady, 2008). In another homology-based metagenomic study, Kang et al. identified an eDNA-derived cluster encoding for arimetamycin A, an anthracycline obtained to be more efficient than clinically used natural anthracyclines against multidrug-resistant cancer cells (**Figure 10**) (Kang e Brady, 2013).

AIM OF THIS STUDY

2 Aim of this study

2.1 General goal

The main goal of this thesis was to use the metagenomics as a tool to acess and identify the chemo-geographic diversity of multiple samples collected in distinct Brazilian soil samples and develop bioinformatics tools for the molecular studies of natural products.

2.2 Objectives

- To collect, isolate and purify eDNA from different Brazilian samples analyzing the physicochemical properties;
- To sequence the eDNA isolated from distinct biomes using next general sequencing platform and analyze the data by bioinformatics approaches;
- To create a pipeline and a package capable to design degenerated primers targeting important natural products molecules;
- To validate the primers using a way to mimic the diversity naturally found in environmental samples;
- To analyze the Brazilian biogeography for nonribosomal peptide syntethases and polyketides and find important hotspots to search clinically relevant molecules.

MATERIAL AND METHODS

3 Material and Methods

3.1 General instrumentation and materials

This work was developed in a biosafety level 2 (BSL2) equipped lab containing all of the general equipment required for the molecular biology, microbiology and chemical biology research described in this proposal. This includes equipment for electrophoresis, centrifugation and fermentation of microorganisms as well as all of the equipment required for the isolation and purification of new natural products. Major equipment includes: an Eppendorf liquid handling station for library screening, 4 three-chamber temperature controlled ATR shakers for large scale culturing, a Personal Genome Machine next-generation sequencer, Illumina MiSeq, centrifuges, incubators, freezers, balances, water baths, electrophoresis equipment, thermocyclers, a laminar flow hood, biosafety cabinet and multiple fume hoods.

Computer/Bioinformatics resources: High-end bioinformatics analysis was carried out on the Rockefeller University Da Vinci server. Data was analyzed and stored on a Dell server (Tower-based PowerEdge T710, Red Hat Enterprise Linux OS, 80GB memory, 3.46 Ghz Intel primary and co-processors with RAID 5/6 Hard Drive, 10 TB Hard Disc Space). We used a MacBook Pro 15 Inch 2.2 Ghz Intel Core I7, 16 GB 1600 MHz DDR3 for the analysis.

Solvents and buffers were ordered from multiple companies. We have used pWEB cloning kit (Epicentre), Mini and MidiPREP (Qlagen), PowerClean MoBIO®, V3 Illumina. Degenerated primers were obtained from IDT technologies.

3.2 SecMetPrimer Pipeline

Our internal data was composed from 1310 biosynthetic gene clusters from MiBIG database (Medema *et al.*, 2015). We distributed these BGCs by natural products families. The pipeline consisted of: 1) to align the sequences from an specific target using MUSCLE (Edgar, 2004). Thus, the multiple sequence alignment (MSA) of all targets were created; 2) to trim the MSA to remove the gaps and provide more accurate file; 3) to pick primers in all positions and provide the degeneracy increment using DegePrime (Hugerth *et al.*, 2014); 4) to find the conserved sections with the association of coverage/degeneracy; 5) to Increase de primers degeneracy applying synonymous data; 6) to increase the degeneracy of the primers employing the CodonReplacement tool (**Figure 12**).

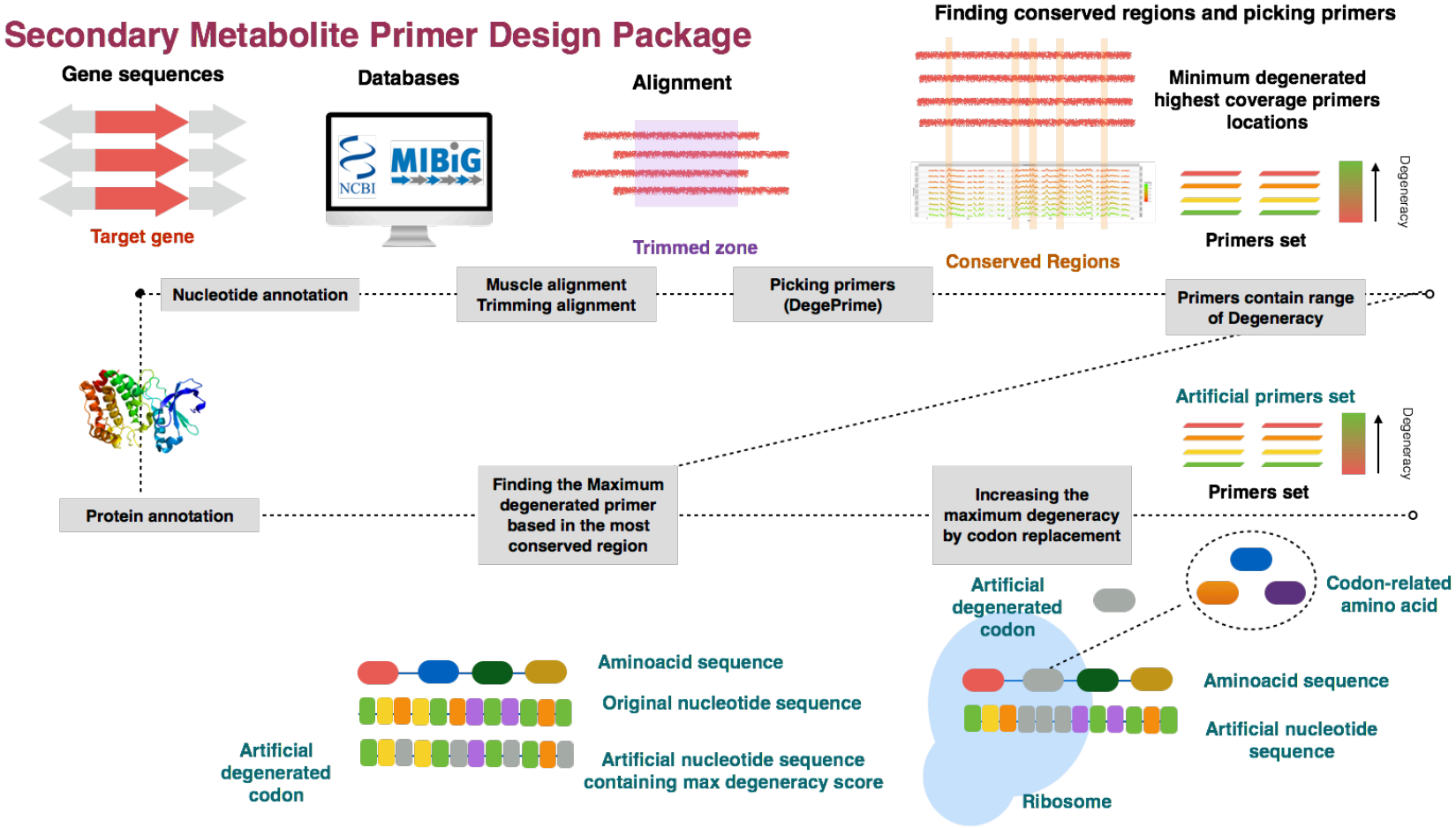


Figure 12: SecMetPrimer Pipeline
SecMetPrimer consist to design primers using information from databases. All steps in SecMetPrimer pipeline to design, align, increase the degeneracy, and annotation were summarized in this this figure.

The degeneracy in the SecMetPrimer was increased following the degenerate nucleotides available. The **Table 1** above showed all possibilities of the nucleotides and degenerated nucleotides used in the SecMetPrimer tool. For example, the degenerated nucleotide S has strong interaction and link to cytosine and guanine while the degenerate nucleotide W has weak interactions, linking to adenine and thymine.

Table 1: Degenerate nucleotides and their interactions

Symbol	***Meaning***	***Origin of designation***
G	G	Guanine
A	A	Adenine
T	T	Thymine
C	C	Cytosine
R	G or A	puRine
Y	T or C	pYrimidine
M	A or C	aMino
K	G or T	Keto
S	G or C	Strong interaction (3 H bonds)
W	A or T	Weak interaction (2 H bonds)
H	A or C or T	not-G, H follows G in the alphabet
B	G or T or C	not-A, B follows A
V	G or C or A	not-T (not-U), V follows U
D	G or A or T	not-C, D follows C
N	G or A or T or C	aNy

The codon replacement strategy were developed to replace the characteristics of the codon belongs to primers to artificial codons capable to amplify all the specific amino acid classes. For example, the hydrophobic aliphatic amino acids isoleucine, leucine and valine may be combined together to the artificial codon VTN (**Table 2**). As the same, the hydrophobic aromatic amino acids phenylalanine, tryptophan and tyrosine could be combined together to the artificial codon TWB (**Table 2**). By this approach was expected that increase the range of amplification of one target by

keeping the physicochemical properties of the enzyme. Naturally, the artificial codon has high degeneracy score. Thus, for this reason and to create heterogeneity, the ArtificialCodon replacement tools replace the original codon in two levels of degeneracy (low and high).

Table 2: Artificial amino acids used in the CodonReplacement tool

Class	***Low degeneracy***	***High degeneracy***	***Amino acids***
Hydrophobic aliphatic	VTS	VTN	Isoleucine Leucine Valine
Hydrophobic aromatic	TWS	TWB	Phenylalanine Tryptophan Tyrosine
Neutral	HCR	HCN	Serine Threonine
	MAS	MAN	Glutamine Asparagine
Eletricaly charged - acid	GAS	GAN	Aspartic acid Glutamic acid
Eletricaly charged - basic	MAS	MAN	Histidine Lysine

The CodonReplacement tool use degenerate nucleotides S

3.3 Targeting Natural Products Gene Clusters

Clinically relevant natural products families were selected to design new sets of primers. The selected families were type I polyketide (PK), nonribosomal peptide (NRP), aminoglycoside, type II polyketide, isonitriles, phosphono containing-NP and tryptophan dimmers (**Table 3**).

For type I PK was selected the ketosynthase (KS), acyltransferase (AT) domains and the gene 3-amino-5-hydroxybenzoic acid (AHBA) synthase that encode the AHBA group in the ansamycins group. For NRP was selected adenylation (AD), condensation (C) domains and the genes acetate decarboxylase, lysine cyclodeaminase and L-capreomycidine synthase from enduracididine-, pipecolic acid- and capreomycidine-containing NRP. For aminoglycoside was selected the gene 2-deoxy-streptamine inosose synthase (2-DOI), while for tryptophan dimmers (TD) the chromopyrrolic acid synthase. The β-ketoacyl synthase from T2 polyketide and phosphoenolpyruvate synthase (PEP) from phosphono containing-NP were also selected for design the primers.

Table 3: Selected natural products targets

Molecule family	***Target***	***Total of primers designed***	***Min degeneracy***	***Max degeneracy***
PK	KS domain	20	4	36
PK	AT domain	10	3	36
PK	AHBA	10	4	32
NRP	AD domain	9	2	37
NRP	C domain	13	3	39
NRP	capreomycidine	14	4	37
NRP	pipecolic acid	20	4	34
NRP	enduracidine	20	9	29
Aminoglycoside	2-DOI	10	4	38
Type II PKS	ketoacylsynthase	11	4	39
Isonitrille	PvcA	16	3	33
Phosphono	PEP mutase	11	2	38
Tryptophan dimmer	CPA synthase	13	4	37

In **Table 3** we observed the total number of primer designed for each class and the min and max degeneracy counted. All primers were obtained containing double-labeled barcode indexes.

3.4 Validating the primers sets

To validate the primers sets we needed a high volume of purified DNA. Unfortunately, we could not validate directly using eDNA samples due to the tremendous costs to isolate and purify the eDNA from soil samples. To simulate the soil diversity, we decided to build a pool of multiple libraries, called metalibrary. We began picking 10 distinct libraries, containing a total of 229.2 million of clones. The libraries were created following the protocol described by Brady (2007) and corresponding to soil samples from Arkansas, Alaska, Arizona, Hawaii, Oregon and New Mexico (**Table 4**).

Table 4: Metalibrary construction

Library	*Antibiotic*	*Megapools*	*Total cosmids*	*1x cover*
AR/AZ	chlor/amp	36	11.5 million	1
AZ-25	chlor/amp	8	10.7 million	1
HI-3	chlor/amp	16	20 million	2
OR-13	chlor/amp	16	20 million	2
AK-3	chlor/amp	16	18 million	2
AZ-52	chlor/amp	16	22 million	2
NM-5	chlor/amp	16	39 million	1
NM-18	chlor/amp	2	5 million	1
NM-8	chlor/amp	16	59 million	2
HI-4	chlor/amp	16	24 million	2
SUM		**158**	**229.2 million**	**24**

*These libraries were constructed in Sean Brady lab using different eDNA. AR/AZ–Arizona/Arkansas soil; Az–Arizona soil; HI–Hawaii soil; Or-Oregon soil; AK–Arkansas, NM–New Mexico.

For each library, all mega pools were added to the single tube and electrotransformed. For the transformation, 50 µL of EC100 cells were added to 600 ng of combined mega pools and a pulse was applied to shock the cells. Immediately 1000 µL of SOC medium were added into the curvet. The SOC-cells mixture was

incubated at 37 °C, 200 RPM for 1 hour to permit expression of antibiotic resistance gene. After that, 1000 µL of SOC-transformed cells expressed were added to 200 mL of LB/Chlor (12.5 ng/µL) for replication. The incubation was carried out overnight in the 200 RPM. The glycerol stocks were storaged for each sample. We MidPreped the samples following the commercial MIDIPREP (QIAGEN) protocol, and a high concentration of eDNA was obtained (**Figure 13**).

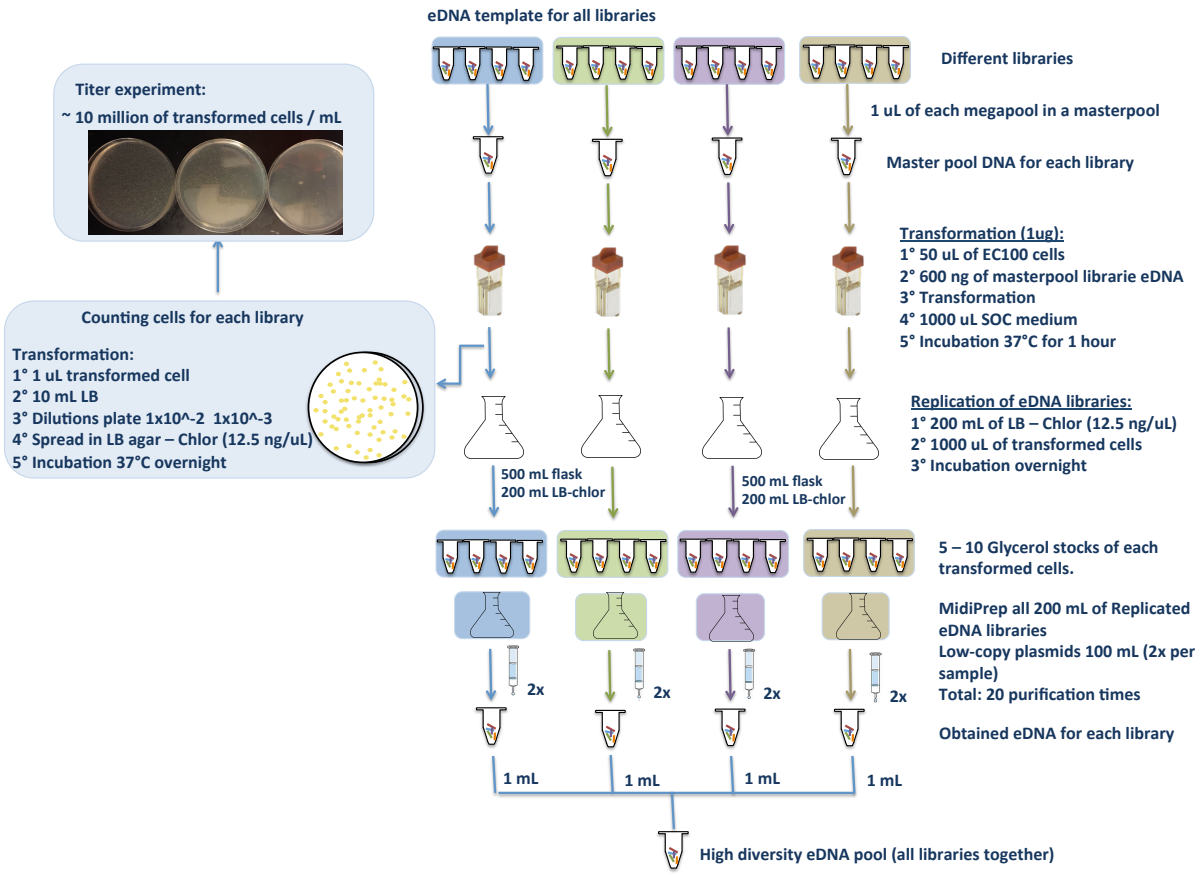


Figure 13: Electrotransformation protocol
The protocol was followed by electrotransformation of *E. coli* EC100 with eDNA from megapols libraries. After replication, clones were counted and the culture was scale up. The eDNA was isolated using MidiPrep kit. Different colos in figure represent different libraries. The dilution for counting is showed in pictures.

3.5 Pair-end double labeled barcoding strategy

We used a barcoding strategy based on the previous work from Jaques Ravel's group (Fadrosh *et al.*, 2014) where they used a dual barcode to create a unique per-sample barcode with a relatively small number of barcodes (**Figure 14**). The basic idea is the MiSeq invariant sequences (called 5' and 3' linker) are incorporated into the primer outside of the MiSeq sequencing primer sequences (called HP10 and HP11), as well as outside a barcode and a spacer to generate heterogenity. The full primer from MiSeq / primer is over than 90bp, which is prohibitively expensive when we need to order several hundreds, so we decided to break it up into a two-step PCR where the indexes and part of the MiSeq invariant sequences are incorporated in the first PCR round and the rest of the adaptor is incorporated in the second round of amplification.

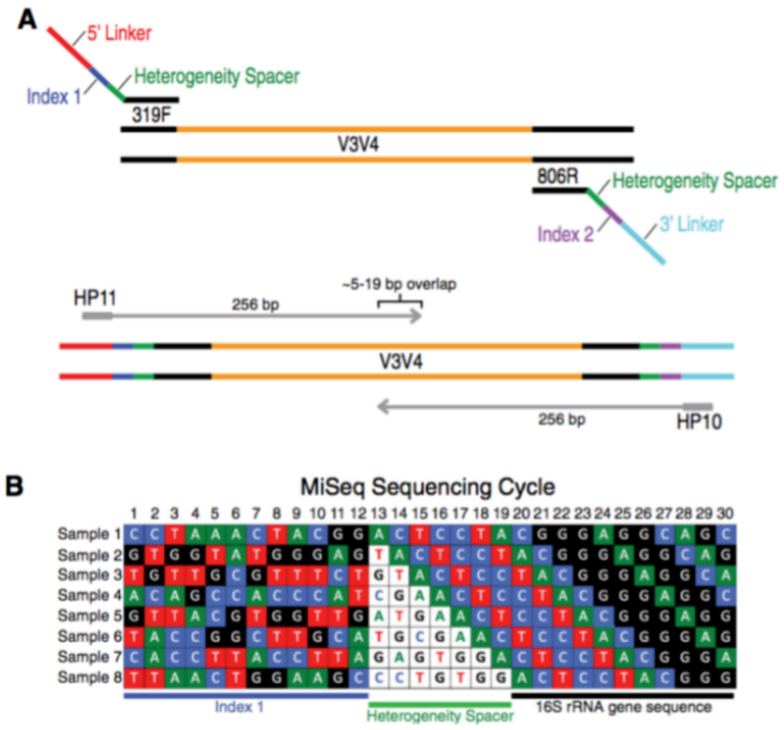


Figure 14: Barcoding labeling strategy
In A, the structure of amplicons was designed by the amplicon (yellow) and primers (black). In B, an example of how the heterogeneity works to avoid overlaps.

The **Figure 14**(A) showed in brown the amplicons regions and in the flanking regions the barcodes in black, heterogeneity spacer in green and part of MiSeq adaptors Index1 and Index2 in blue and purple respectively. In (B) we showed how the heterogeneity spacer helps to avoid the overlapping in the sequencing step. To observe the structure of the amplicons, please check the annex 1.

3.6 PCR amplification and sequencing

To the amplification reaction, 25 microliters PCR reactions contained 12.5 µL of FailSafe PCR Buffer G (Epicentre), 1 µL of Taq Polymerase (New England Lab), 0.5 µL of each forward and reverse primer (100 µM), 10.2 µL of water, and 200 ng of purified eDNA. Amplification conditions of the multiplex were 95 °C for 5 min, followed by 40 cycles of 95°C for 30s, 65.1°C for 30s, 72°C for 1min, and, finally, 72°C for 7 min. No PCR amplicon was seen in the negative control (water) well on the experiment plate. Amplicons at this stage contained unfinished Illumina adaptors and consequently demanded a second round of PCR to append the complete adaptor sequence. The addition of the Illumina adaptors also served to allow amplicon identification beyond 96 samples.

For the second PCR, the amplicons were pooled into groups and cleaned utilizing Agencourt Ampure XP magnetic beads (Beckman Coulter). The cleaned amplicon pools were later applied to the template with the Illumina tag/index primers. Second PCR conditions were: 10 µL of FailSafe Buffer G (Epicenter), 3.8 µL of water, 0.4 µL of each primer (100 µM), 0.4 µL of Taq Polymerase (New England Lab), and 5 µL of purified amplicon (50–100 ng). PCR proceeded as follows: 95 °C for 5 min, six cycles of 95 °C for 30 s, 70 °C for 30 s, and 72 °C for 45 s, and, lastly, 72 °C for 5

min. Second PCR amplicon pools were merged in an equimolar proportion for all targets. Each pool was cleaned doubly more with Agencourt Ampure XP magnetic beads (0.7:1 bead volume to DNA solution). The AD and KS cleaned amplicons were next sequenced employing Illumina MiSEq (2 × 300 technology (602 cycles: 301 × 301)) using V3 sequencing kit.

3.7 Data analysis

The raw MiSeq fastq files were demultiplexed using a publicly available python package developed for debarcoding Illumina paired-end reads (https://github.com/ esnapd/paired-end- debarcoder).

The fastq files were then processed using seqtk trimfq to remove low-quality bases with the Phred algorithm and the default error value cutoff = 0.05 (Li, 2016). Forward reads shorter than 240 bp and reverse reads shorter than 175 bp were dropped. Following this, paired reads were concatenated using a single “N” between each forward and reverse pair, resulting in a single fasta file with demultiplexed, quality-filtered, and uniform length sequences for each sample site.

For the clustering, OTUs were generated using the Usearch software with the clust_fast function on the concatenated reads from each site (Edgar, 2013). A dereplication step was applied, followed by a 97% identity clustering and removal of singleton clusters. Following these denoising steps, the 97% identical centroid sequences were pooled and subsequently clustered at 95% identity. Rarefaction curves were measured based in abundance of OTUs from each unique primer set. All plots were generated by GGPLOT2.

Annotation of NP diversity in the metalibrary for each target was queried using the MiBig internal database. For this analysis, we blast the sequences against the

database, and one e-value cutoff of 1e-40 was used to filter the results. Empirical studies using operational taxonomy units (OTUs) have shown that e-values of 1e-40 and lower yield reliable gene cluster annotation results (Reddy *et al.*, 2014; Charlop-Powers *et al.*, 2015).

3.8 Soil Collection and DNA Isolation.

Around 200 g of topsoil (maximum depth of 15 cm) were collected from 37 distinct Brazilian sites and stocked at the LQMo Lab at the School of Pharmaceutical Sciences of Ribeirão Preto - University of Sao Paulo. To reduce the possibility of cross-contamination, samples tools (trowel, paint scraper, and others) were physically washed with ethanol after the collection of samples. Samples were assigned with an individual barcode and instantly set in an identified bag. Soils were not touched during the collection process and were transported to the laboratory instantly. The samples were collected in the biomes Cerrado, Atlantic Forest and Marine sediment. The sampling locations were descried in the **Table 5**. All samples were collected according to CNPq (010214/2012-7), COTEC-IF (260108-012.192/2012) and SISBIO (22817-1) permits.

Table 5: Latitude and Longitude of the samples collected across different biomes

DFD_UID	***Latitude***	***Longitude***	***pH***	***Biome***
DFD000109	-216.107	-476.307	5	Cerrado
DFD000110	-21.643.833	-47.644.667	4	Cerrado
DFD000111	-216.487	-476.396	4.2	Cerrado
DFD000112	-21.648.767	-47.639.933	4.7	Cerrado
DFD000113	-21.692.417	-4.760.855	6.5	Atlantic Forest
DFD000114	-21.694.083	-47.615.533	6.6	Atlantic Forest
DFD000115	-2.173.895	-476.594	6.3	Atlantic Forest
DFD000116	-2.172.755	-47.649.933	7.2	Atlantic Forest
DFD000117	-21.730.167	-47.612.883	5.3	Atlantic Forest
DFD000118	-21.726.917	-47.611	6.6	Atlantic Forest
DFD000119	-217.225	-47.596.717	7	Atlantic Forest
DFD000120	-21.728.733	-47.583.833	5.8	Atlantic Forest
DFD000121	-22.965.167	-42.001	7.9	Marine environment
DFD000122	-229.705	-42.032.833	8	Marine environment
DFD000123	-22.984.833	-41.990.833	8	Marine environment
DFD000124	-22.960.833	-42.004	8	Marine environment
DFD000125	-229.681	-419.942	8	Marine environment
DFD000126	-22.980.617	-41.983.967	8	Marine environment
DFD000127	-18.290.983	-492.583	6.6	Cerrado
DFD000128	-18.290.983	-492.583	6.9	Cerrado
DFD000129	-2.292.075	-42.038.167	7	Cerrado
DFD000130	-22.820.567	-42.153.833	6.5	Sandbank
DFD000131	-229.418	-42.041.833	6.7	Atlantic Forest
DFD000132	-22.726.083	-41.966.833	7.1	Sandbank
DFD000133			8.1	Mangrove
DFD000134				Atlantic Forest
DFD000135				Atlantic Forest
DFD000136	-2.173.895	-476.594		Mangrove
DFD000137	-21.630.317	-47.632.217	4.6	Cerrado
DFD000138	-21.630.583	-47.631.967	4.4	Cerrado
DFD000139	-21.627.667	-47.634.833	4.7	Cerrado
DFD000140	-2.164.845	-47.636.283	4.3	Cerrado
DFD000141	-216.482	-476.365	4	Cerrado
DFD000142	-21.648.017	-47.636.633	4.4	Cerrado
DFD000143	-21.649.717	-47.639.767	5.2	Cerrado
DFD000144	-21.675.917	-475.948	5.8	Atlantic Forest

3.9 DNA extraction and purification

DNA was extracted from soil using a simplified version of DNA isolation protocol (Brady, 2007). The protocol was as follows: 250 grams of each soil sample was incubated at 70 °C in 150 ml of lysis buffer (2% sodium dodecyl sulfate [wt/vol], 100 mM Tris–HCl, 100 mM EDTA, 1.5 M NaCl, 1% cetyl trimethyl-ammonium bromide [wt/vol]) for 2 hr. Large particulates were then removed by centrifugation (4000×*g*, 30 min), and crude eDNA was precipitated from the resulting supernatant with the addition of 0.7 vol of isopropyl alcohol. Precipitated DNA was collected by centrifugation (4000×*g*, 30 min), washed with 70% ethanol and resuspended in a minimum volume of TE (10 mM Tris, 1 mM EDTA [pH 8]). Crude environmental DNA was passed through two rounds of column purification using the PowerClean system (MO BIO, Carlsbad, California).

3.10 Degenerate Primer Design

The following degenerate primer sets were employed to PCR-amplify AD and KS domains: AD domains A3F (5′-GCSTACSYSATSTACACSTCSGG) and A7R (5′-SASGTCVCCSGTSCGGTA) and KS domains KS2F (5′-GCNATGGAYCCNCARCARM GNVT) and KS2R (5′-GTNCNNGTNCCRTGNSCYTCNAC) (Ayuso-Sacido e Genilloud, 2005; Schirmer *et al.*, 2005). We designed the primer sets containing 8 forward and 8 reverse unique barcodes. The barcodes were applied to create single pairs allowing for association of amplicons from different places in a pooled sequencing run. Primer sequences added the Illumina p5 or p7 sequence, an 8-bp barcode sequence, a spacer sequence, and the degenerate primer itself.

3.11 PCR Amplification and Sequencing

Forty-microliters PCR reactions were conducted containing 20 µL of FailSafe PCR Buffer G (AD) or Buffer E (KS) (Epicentre), 1 µL of Taq Polymerase (England Lab), 1.25 µL of each primer (100 µM), 14.5 µL of water, and 2 µL of purified eDNA. Amplification conditions for AD domain primers were 95 °C for 5 min, followed by 40 cycles of 94°C for 30s, 67.5°C for 30s, 72°C for 1 min, and, finally, 72°C for 7 min.

Amplification conditions for KS domain primers were 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 56.3 °C for 30s, 72°C for 1 min, and, ultimately, 72°C for 7 min. No PCR amplicon was noted in the negative control (no soil). At this stage, the amplicons contained incomplete Illumina adaptors and consequently demanded a second round of PCR to supplement the whole adaptor sequence. The extension of the Illumina adaptors further helped to recognize amplicon identification. For the second PCR, the amplicons were pooled into groups and cleaned utilizing Agencourt Ampure XP magnetic beads (Beckman Coulter). The cleaned amplicon pools were later applied to the template with the Illumina tag/index primers.

Second PCR conditions were: 10 µL of FailSafe Buffer G (Epicenter), 3.8 µL of water, 0.4 µL of each primer (100 µM) , 0.4 µL of Taq Polymerase (New England Lab), and 5 µL of purified amplicon (50–100 ng). PCR proceeded as follows: 95 °C for 5 min, six cycles of 95 °C for 30 s, 70 °C for 30 s, and 72 °C for 45 s, and, lastly, 72 °C for 5 min. Second PCR amplicon pools were merged in an equimolar proportion within two separate final pools— AD and a KS pool. Each pool was cleaned doubly more with Agencourt Ampure XP magnetic beads (0.7:1 bead volume to DNA solution). The AD and KS cleaned amplicons were next sequenced employing Illumina MiSEq (2 × 300 technology (602 cycles: 301 × 301)).

3.12 Bioinformatics processing

The raw MiSeq fastq files were demultiplexed using a publicly available python package developed for debarcoding Illumina paired-end reads (https://github.com/ esnapd/paired-end- debarcoder).

For the filtering, we used fastqFilter() command. This tool provides the filtering of fastq files using the usearch fastq_filter command (http://www.drive5.com/ usearch/manual/cmd_fastq_filter.html) (Edgar, 2010). In summary, this function trims sequences to a particular length, pushes shorter sequences than that range, and filters based on the representation of ambiguous bases, a minimum quality score, and the expected errors in a read (Edgar, 2010).

For dereplication, we used derepFastq() on DADA2 (Callahan *et al.*, 2016). This command imports a fastq file and outputs a dereplicated table of individual sequences and their abundances. This tool further outputs consensus positional quality scores for any single sequence by considering the average (mean) of the positional qualities of the element of the reads. These consensus scores were used by the error model of the dada() function (Callahan *et al.*, 2016).

For the denoising process, dada() implements the core denoising algorithm. The function isBimeraDenovo() was used for chimeras identification. This identifies sequences are exact bimeras of more plentiful output sequences. Bimeras are recognized by a Needleman-Wunsch global alignment of every sequence to all more abundant sequences, and then seeking for combinations of a left-parent and a right-parent that cover the sequence without mismatches or internal indels. The isBimeraDenovo() function was applied following denoising, and suggested sample sequences, preferably than on noisy input reads or fuzzy operational taxonomy units

- OTUs. We used mergePairs() function to performs a global alignment connecting paired forward and reverse reads, and merges them synchronically (Callahan *et al.*, 2016).

In this way, R1 reads shorter than 240 bp and R2 reads shorter than 175 bp were then discarded. Following this, paired reads were concatenated using a single “N” between each R1 and R2 pair, resulting in a single fasta file with demultiplexed, quality-filtered, and uniform length sequences for each sample site.

Rarefaction curves were generated from the phyloseq R package by subsampling the DADA2 tables at several depths (1, 100, 500, and 1,000–10,000 by increments of 500) (Mcmurdie, P. J. e Holmes, S., 2013). For all depth, the Chao1 diversity metric was calculated over 10 iterations, and the mean of the richness calculated in each iteration was plotted. At a several depths, scatterplots were generated to display richness trends versus reported sites variables). Non-metric multidimentional scaling (NMDS) analysis (was performed with the phyloseq R package through the ordination methods using the Bray–Curtis distance. The DADA2 objects with abundance lower than three were not included in the analysis and data from each site were normalized according to the total number of total reads observed at the site.

RESULTS AND DISCUSSION

4 Results and discussion

The complexity of terrestrial microbiomes has made the analysis of total biosynthetic enzyme diversity challenging. Initial studies focused only on analyzing groups of sequences from single environmental samples (Metsa-Ketela *et al.*, 2002; Ginolhac *et al.*, 2004; Daniel, 2005; Gentry *et al.*, 2006). Several studies revealed that the accession distinct homologous BGCs could be developed based PCR approach (Seow *et al.*, 1997; Ginolhac *et al.*, 2004).

Due to exciting advances in high-throughput sequencing technology, the large-scale, conforming analysis of millions of sequences from distinct environmental samples can now be achieved (Margulies *et al.*, 2005). To identify the chemical heterogeneity present in our samples, we designed a panel of degenerated PCR primers using SecMetPrimer targeting different biosynthetic enzymes for high-throughput Illumina sequencing (Charlop-Powers, Milshteyn*, et al.*, 2014; Charlop-Powers, Owen*, et al.*, 2014; Charlop-Powers *et al.*, 2015).

4.1 Primers designed using SecMetPrimer

SecMetPrimer designed degenerated primers including domains of homology for the selected biosynthetic targets and an Illumina adapter oligomer, which allows for the direct sequencing of the PCR amplicons. Additionally, individual hamming barcodes were included in each primer for the multiplexing. Thus, the analysis and the identification were developed by the source of individual sequencing reads in the data set (Charlop-Powers et al., 2015; Lemetre et al., 2017). The targets chosen to investigate are described in **Table 3**. In summary, were clinically relevant classes such as PK, NRP, aminoglycosides, phosphono, tryptophan dimmers and other NP families. For each target, we obtained a FASTA file containing information about BGCs related to that family. The MSA supported us to find the conserved regions through the aligned sequences. For this end, SecMetPrimer designed 176 distinct sets of primers. All primers designed by SecMetPrimer were summarized in **Table 6**.

The conserved regions of all alignments were recognized by the correlation between maximum coverage and minimum degeracy. DegePrime provided degenerate oligomers with a defined length of 21 nucleotides and multiple degeneracies. A degenerate PCR primer is an oligomer sequence that supports some possible nucleotides in one or more positions. As the example, in the primer AT{C,G}T{A,T,C}G, the third location is C or G and the fifth is A, T or C. The degeneration of this primer is the total number of combination it contains. In this instance, the degeneracy is 6.

We observed in **Figure 15** the degeneracy needed to improve the increase of the coverage. In this specific target, four different conserved regions were discovered in the MSA. In this case, the regions were very conserved, which turns easy to

observe and detect in the alignment file. The same strategy was used in the TD MSA. In this case, only one conserved region was easily detected. Thus, for design of primers, SecMetPrimer apply DegePrime function to increase de degeneracy (**Figure 16**). The same strategy was used for all targets, but here, were described for capreomycidine-containing NRP (**Figure 17**) and phosphono-containing NP (**Figure 18**). In this last case, the PEP mutase gene showed initially very difficult to discover the conserved regions. However, the increment of the degeneracy provides means to detect them.

Thus, in this protocol, we were able to designed primers containing a series of desirable degeneracies: 4; 8; 12; 16; 48; 96; 128; 864 and 1296 (**Figure 19**) for all targets.

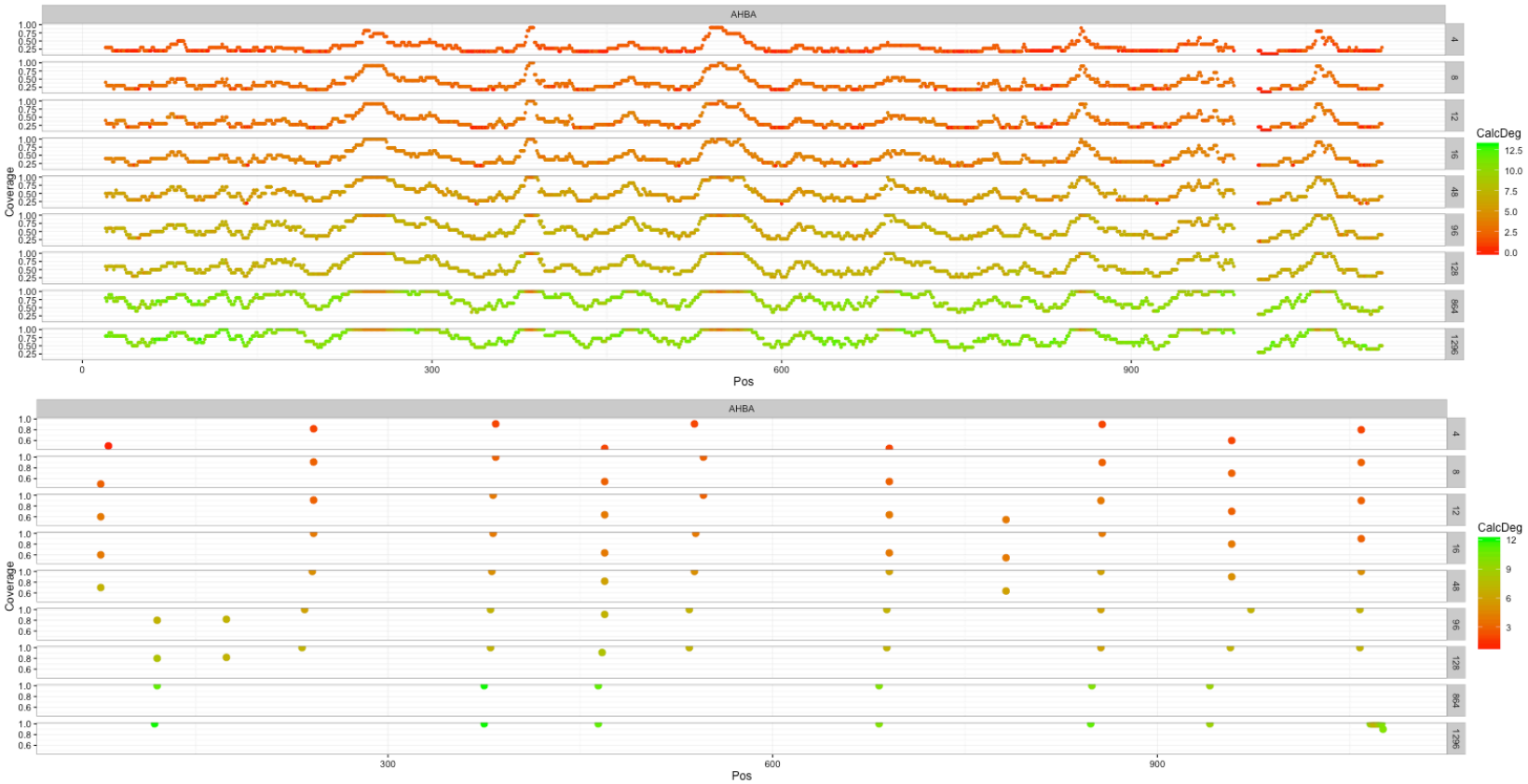


Figure 15: AHBA alignment across different degeneracy gradient
The x-axis correspond to the position of the primer sequence in the alignment and in y-axis the coverage of the primer in MSA. In top is observed the complete primer alignment and in bottom the positions containing the best coverage. Degeneracy colors range from red (minimum) to green (maximum).

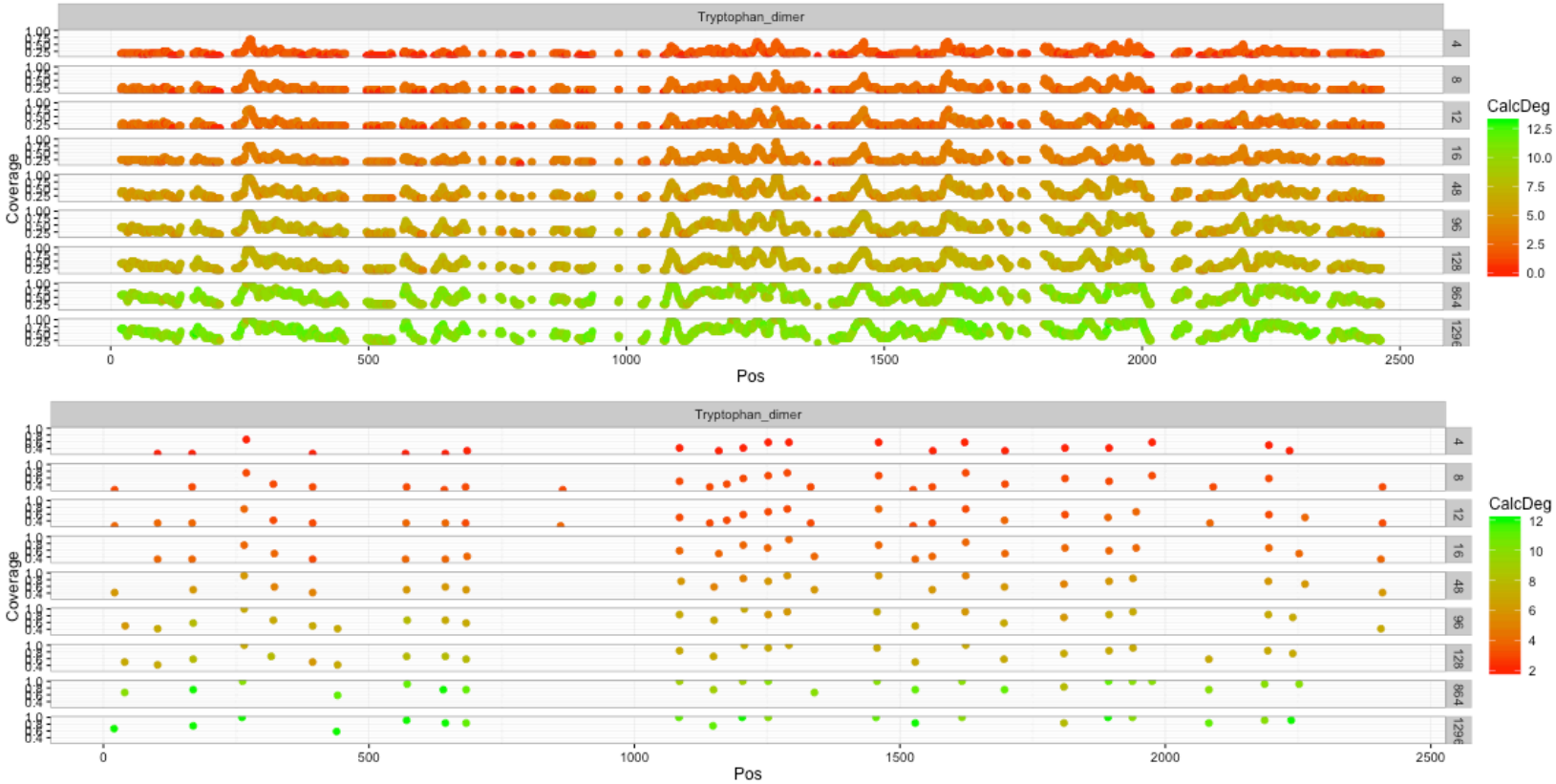


Figure 16: TD alignment across different degeneracy gradient
The x-axis correspond to the position of the primer sequence in the alignment and in y-axis the coverage of the primer in MSA. In top is observed the complete primer alignment and in bottom the positions containing the best coverage. Degeneracy colors range from red (minimum) to green (maximum).

Figure 17: Capreomycidine alignment across different degeneracy gradient
The x-axis correspond to the position of the primer sequence in the alignment and in y-axis the coverage of the primer in MSA. In top is observed the complete primer alignment and in bottom the positions containing the best coverage. Degeneracy colors range from red (minimum) to green (maximum).

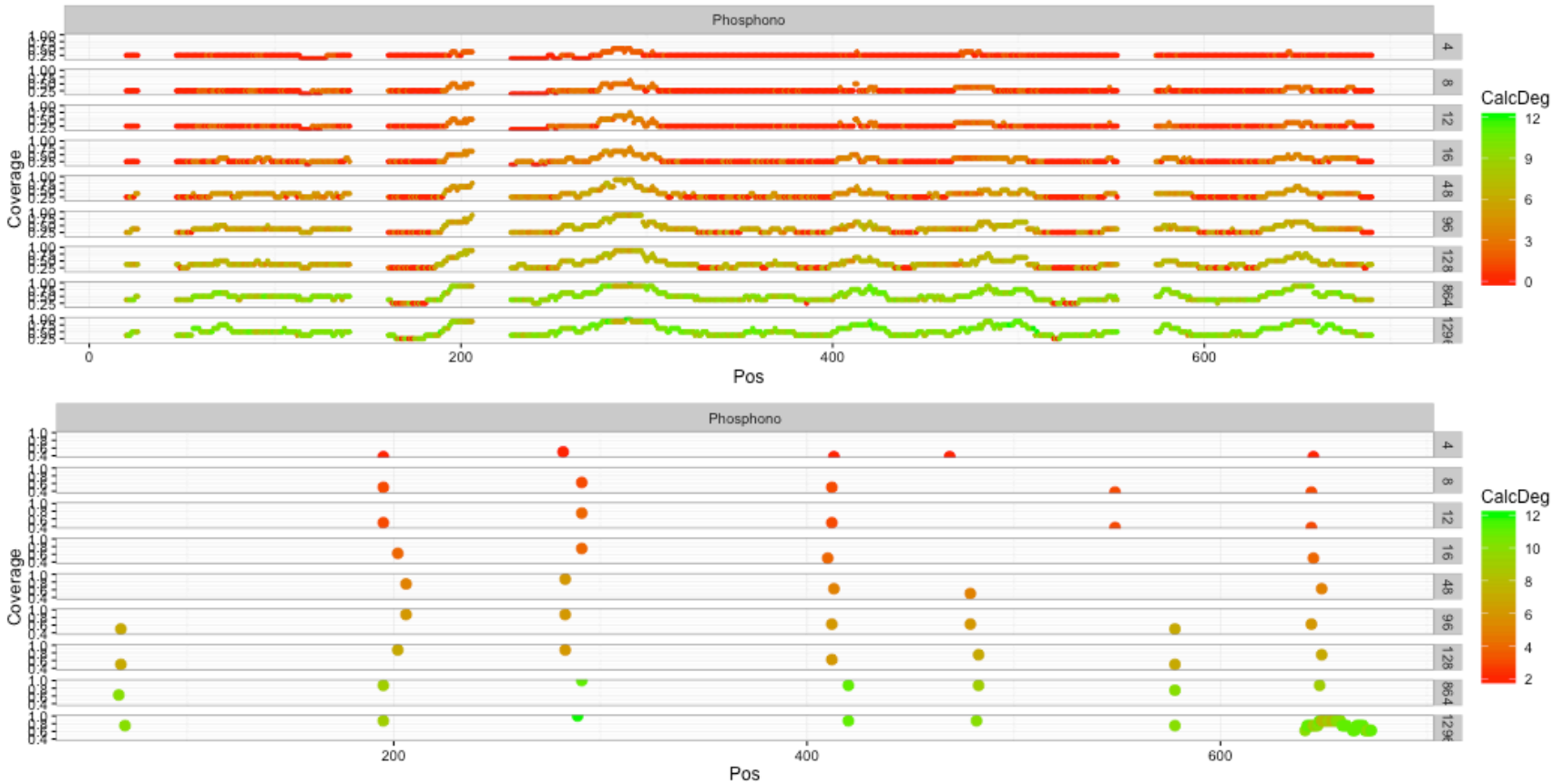


Figure 18: PEP alignment across multiple degeneracy gradient
The x-axis correspond to the position of the primer sequence in the alignment and in y-axis the coverage of the primer in MSA. In top is observed the complete primer alignment and in bottom the positions containing the best coverage. Degeneracy colors range from red (minimum) to green (maximum).

As before mentioned the interesting to increase the degeneracy of all designed primers were determined by (**Figure 19**) the relation between primer designed x degeneracy sum. We observed that SecMetPrimer was able to design smartly and quickly a range of primers for distinct families of NP and at the same time increase drastically the degeneracy.

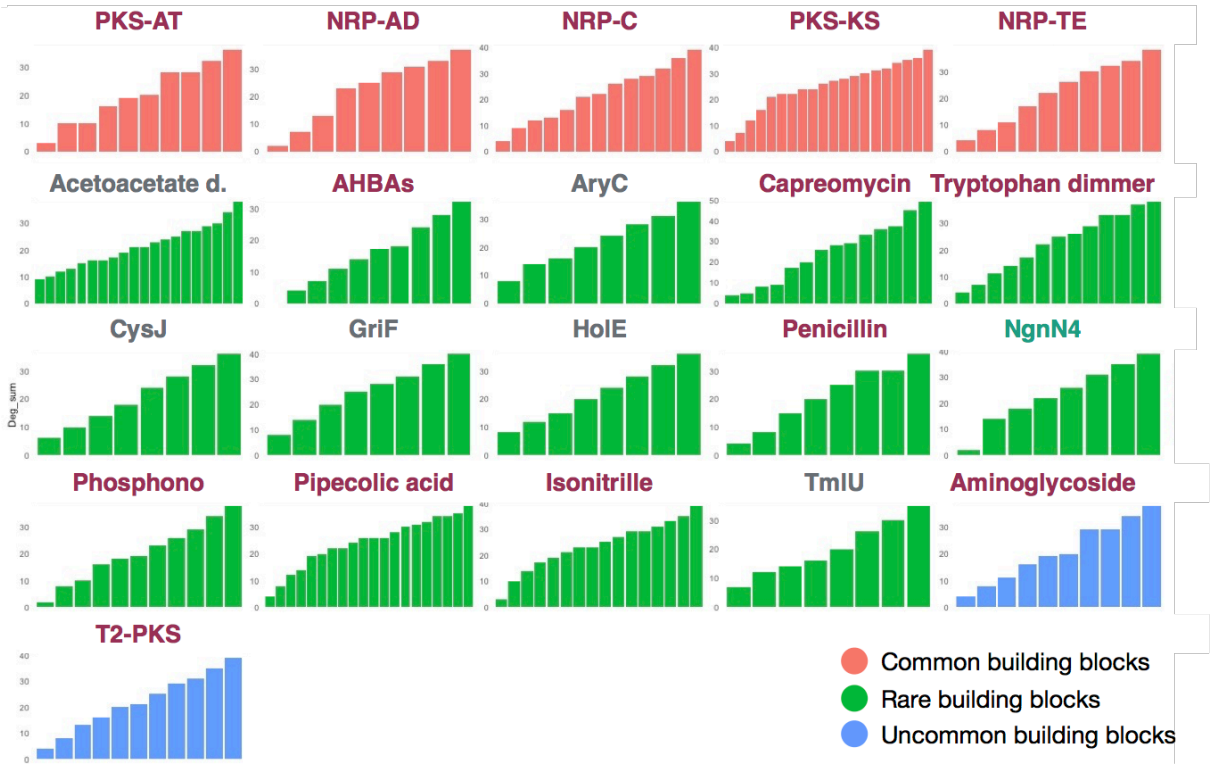


Figure 19: Primers designed for different targets
In x-axis an individual primer set is plotted and in y-axis the degeneracy sum of the primer set. The name color showed designed primers already screened and in grey primers sets designed to amplify another targets to validate SecMetPrimer. In this these we described AHBA and Tryptophan dimmer primers sets.

4.2 Amplicons obtained by degenerated primers

The amplicons were obtained by PCR, and their quality was measured by electrophoresis gel and High Sensitivity D1000 ScreenTap®. For each class, the amplicons were combined together, and this provides a way to multiplexing the samples. We obtained amplification for most of the targets (**Figure 20**). Sadly, all

primers designed for enduracididine containing-NRP did not show amplification. This is probably related to the conditions for amplification. DNA template with high diversity such as eDNA and meta libraries become difficult the amplification of rare genes in the background. Although, all amplicons length were obtained with similar size then expected through bioinformatics calculations. This suggested the presence of the expected amplicons. However, we observed some smears and very faint amplification and some targets (such as NRP-AD) (**Figure 20**). These results implied that some unrelated amplification might have occurred.

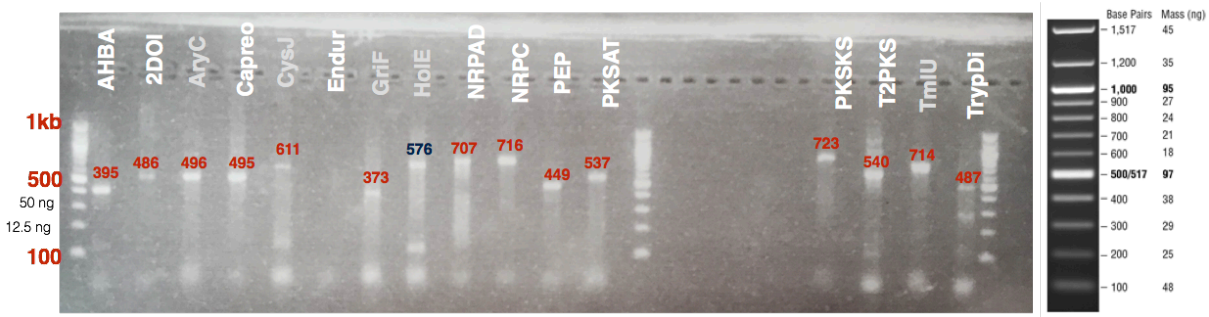


Figure 20: 2% agarose gel with pool of amplicons obtained using degenerated primers sets In red the expected amplicons size calculated bioinformatically. In left and right the 100 bp ladder with corresponded size. Amplicons were pooled together by target.

4.3 Bioinformatics results and best primer selection

The analysis of the sequence data was followed as mentioned pipeline in the material and methods. This pipeline was adapted from Lemetre (2017). The clusterization and annotation of the sequences with their unique ID occurred as expected and all data were mapped. For this thesis, we highlight the results in some classes. The OTUs obtained by clusterization of the data were analyzed, and the alpha diversity was measured by rarefaction analysis.

Initially, we observed in the **Figure 21** the relation between the depth x alpha diversity. We observed strange values of alpha diversity for the NP families analyzed.

For example, we observed a diversity of 60 different types of tryptophan dimmers and 100 different clusters of AHBA. Both are rare families of natural products in nature. Thus we expected a lower value. The same was observed for all other families.

For this reason, novel scripts for filtering were developed and included to SecMetPrimer tool. The script provides a way to remove unspecific sequences in data. To do this, all sequences were blasted against the Pfam family for all specific targets. After this, the sequences that passed through this filter were re-analyzed and the rarefaction curves obtained again.

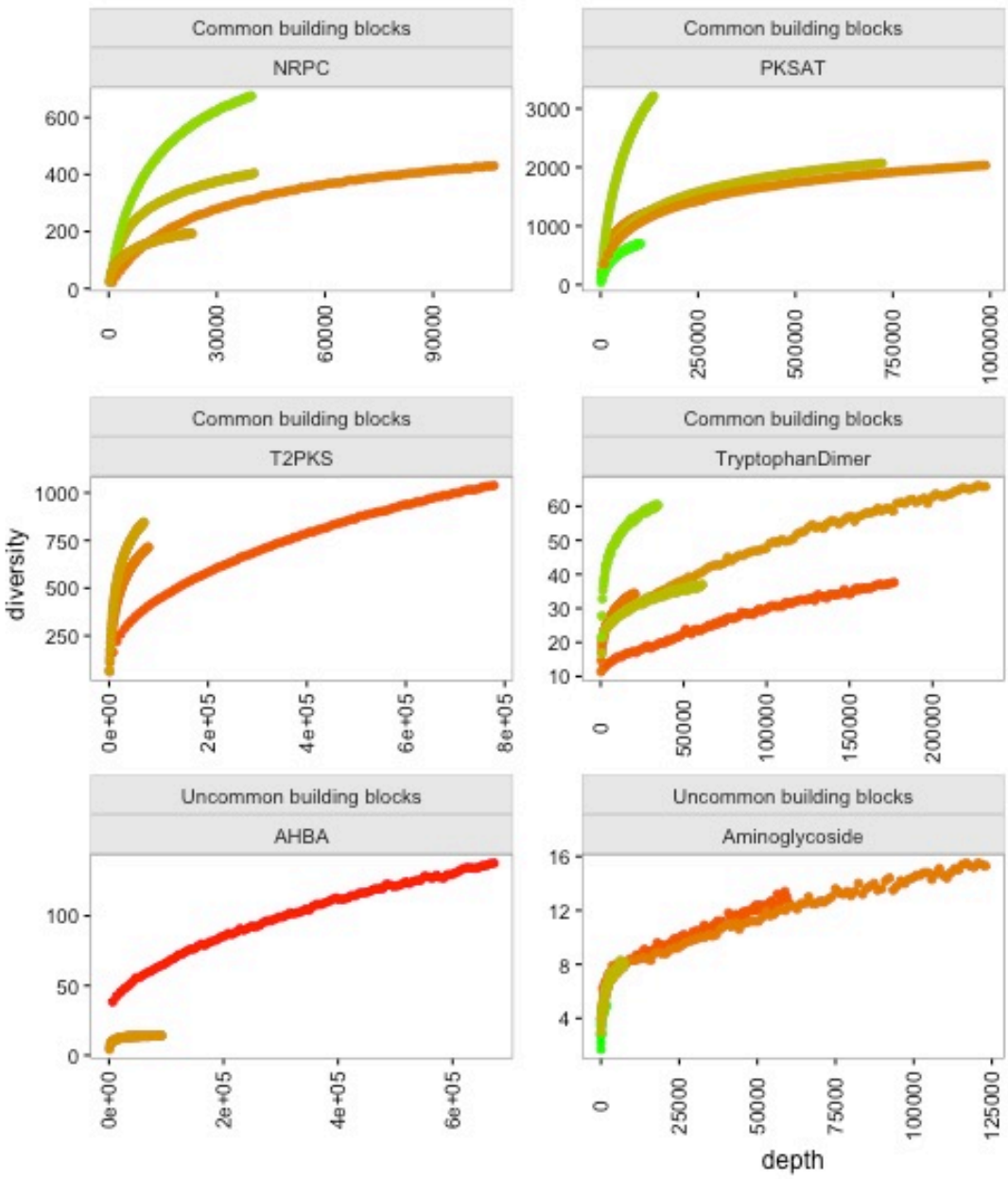


Figure 21: Rarefaction curves containing all alpha diversity calculated
In x-axis the depth were showed and in y-axis the calculated diversity. The degeneracy colors range from red (minimum) to green (maximum)

Table 6: Primers designed by SecMetPrimer with respectively barcodes and length

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_2DOIF01	TSAAGAACGCSCTGGCCATCC	ACGTATCA	GTAGCCSCGCTTGTTGTCGWA	ACGTATCA	Aminoglycoside	486
UF_miSeq_2DOIF02	TSAAGAACKCSCTSGCCRTCC	AGATCGCA	RTAGCCSCGCTTGTTGTCGWA	AGATCGCA	Aminoglycoside	486
UF_miSeq_2DOIF03	GTSAARAAYGCSCTSGCSATY	ATTGGCTC	GTASCCGCGCTTRTTRTCRTA	ATTGGCTC	Aminoglycoside	486
UF_miSeq_2DOIF04	GTVAARAAYGCVCTVGCVATH	CCAGTTCA	RTABCCGCKCTTRTTRTCRTA	CCAGTTCA	Aminoglycoside	486
UF_miSeq_2DOIF05	GTNAARAAYGCNYTNGCNATH	CTAAGGTC	RTANCCNCKYTTRTTRTCRTA	CTAAGGTC	Aminoglycoside	486
UF_miSeq_2DOIF06	GTSAARAAYGCSCTSGCVATY	ACGCTCGA	RTASCCSCKYTTRTTRTCRTA	CATCAAGT	Aminoglycoside	486
UF_miSeq_2DOIF07	GTVAARAAYGCSCTSGCVATH	GAACAGGC	RTABCCBCKYTTRTTRTCRTA	GATAGACA	Aminoglycoside	486
UF_miSeq_2DOIF08	GTNAARAAYGCVYTVGCVATH	AACGCTTA	VWASCCNCGSTKSTKSTCVWA	ACCACTGT	Aminoglycoside	486
UF_miSeq_2DOIF09	VTNMASMASGCSVTSGCSVTS	CACCTTAC	VWASCCNCGSTKSTKNTCVWA	ATTGAGGA	Aminoglycoside	486
UF_miSeq_2DOIF10	VTNMANMASGCSVTSGCSVTS	CGACACAC	VWASCCNCGSTKNTKNTCVWA	ATTGGCTC	Aminoglycoside	486
UF_miSeq_AHBAF01	TTCCAGAACGGMAARCTGATG	CGAACTTA	GGAACATSGCCATGTAGTGSG	CGAACTTA	AHBA	395
UF_miSeq_AHBAF02	TTCCAGAACGGMAARYTGATG	ACAGCAGA	GGAACATSGCCATGTAGYKSG	ACAGCAGA	AHBA	395
UF_miSeq_AHBAF03	TTYCARAAYGGSAARCTSATG	CGGATTGC	KRAACATSGCCATRTARTGVS	CGGATTGC	AHBA	395
UF_miSeq_AHBAF04	TTCCAGAACGGCAAACTGATG	GATGAATC	GAACATCGCCATGTAGTGCGA	GATGAATC	AHBA	395
UF_miSeq_AHBAF05	TTYCARAAYGGSAAAYTSATG	GACAGTGC	RAACATSGCCATRTARTGRGA	GACAGTGC	AHBA	395
UF_miSeq_AHBAF06	TTYCARAAYGGNAARYTNATG	ATTGAGGA	RAACATNGCCATRTARTGRCT	ATTGAGGA	AHBA	395
UF_miSeq_AHBAF07	TTYCARAAYGGVAARYTVATG	ATTGGCTC	SCGSWACATSGCCATSWASTK	CAAGGAGC	AHBA	395
UF_miSeq_AHBAF08	TWSMASMASGGSMASVTSATG	ATCATTCC	SCGVWACATSGCCATVWANTK	CAGCGTTA	AHBA	395
UF_miSeq_AHBAF09	TWNMASMASGGSMASVTSATG	TGGAACAA	SCGVWACATNGCCATVWANTK	CGCTGATC	AHBA	395
UF_miSeq_AHBAF10	TWNMANMASGGSMASVTSATG	AGAGTCAA	NCGVWACATNGCCATVWANTK	AGATGTAC	AHBA	395
UF_miSeq_CPF01	SGACATCAGCAGCAGCGGYG	CGCTGATC	CGSTSGGGTTGTGCGGGAAG	CGCTGATC	Capreomycidine	495
UF_miSeq_CPF02	SGACATCWSCAGCAGCGGCG	AACAACCA	CGSTSGGGTTGTGCGGGAAG	AACAACCA	Capreomycidine	495
UF_miSeq_CPF03	SGACATCWSCAGCAGCGGYG	AACGCTTA	CRSTSGGGTTGTGSGGGAAG	AACGCTTA	Capreomycidine	495
UF_miSeq_CPF04	SGACRTCWSCAGCAGCGGYG	AAGGTACA	CRSTSGGGTTGTGSGGGAAG	AAGGTACA	Capreomycidine	495
UF_miSeq_CPF05	SGACRTSWSCWSSWSCGGYGT	CTGAGCCA	CRSTNGGRTTRTGSGGRAAGT	CTGAGCCA	Capreomycidine	495
UF_miSeq_CPF06	GAYGTSWSSWSSWSSGGBGTN	CTGGCATA	NCCBGTNGGRTTRTGSGGRAA	CTGGCATA	Capreomycidine	495
UF_miSeq_CPF07	GAYGTBWSSWSBWSNGGNGTN	GAATCTGA	NCCNGTNGGRTTRTGNGGRAA	GAATCTGA	Capreomycidine	495

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_CPF08	GAYGTNWSNWSNWSNGGNGTN	CAAGACTA	NCCNGTNGGRTYNTBNGGRAA	CAAGACTA	Capreomycidine	495
UF_miSeq_CPF09	SGYGTNWSNWSNWSNGGNGTN	GAGCTGAA	NCCNGTNGGNTBNTBNGGRAA	GAGCTGAA	Capreomycidine	495
UF_miSeq_CPF10	GAYATHAGYAGYAGYGGVGTS	CCTCCTGA	RCTSGGRTTRTGBGGRAARTT	CCTCCTGA	Capreomycidine	495
UF_miSeq_CPF11	GAYGTSWSSWSSWSSGGSGTS	TAGGATGA	YGDSGGSTKSTKSGGSWASTK	TGGCTTCA	Capreomycidine	495
UF_miSeq_CPF12	GAYGTSWSSWSSWSSGGBGTV	GCTCGGTA	SGDSGGSTKSTKSGGVWASTK	GATAGACA	Capreomycidine	495
UF_miSeq_CPF13	GASVTSHCRHCRHCRGGSVTS	ATCATTCC	SGDSGGSTKSTKSGGVWANTK	CAAGACTA	Capreomycidine	495
UF_miSeq_CPF14	GANVTSHCSHCSHCSGGSVTS	AGCCATGC	SGDSGGSTKSTKNGGVWANTK	GTCGTAGA	Capreomycidine	495
UF_miSeq_EndurF01	AGGGSATGCCCAARCAGTTCG	ACAGATTC	ARSGTCTCSSCRTASGMGAAC	ACAGATTC	Enduracidine	464
UF_miSeq_EndurF02	CAGGGSATGCCSAARCARTTY	CGACTGGA	SAASGTYTCSCCRTASGARAA	CGACTGGA	Enduracidine	464
UF_miSeq_EndurF03	CARGGSATGCCVAARCARTTY	CAAGGAGC	BAABGTYTCSCCRTASGARAA	CAAGGAGC	Enduracidine	464
UF_miSeq_EndurF04	CARGGNATGCCNAARCARTTY	TGGCTTCA	NARNGTYTCNCCRTARCTRAA	TGGCTTCA	Enduracidine	464
UF_miSeq_EndurF05	CARGGSATGCCSAARCARTTY	GAGCTGAA	SAASGTTTCSCCGTASGAGAA	AAGAGATC	Enduracidine	464
UF_miSeq_EndurF06	CARGGSATGCCVAARCARTTY	CTGTAGCC	SAASGTYTCSCCRTASGAGAA	ACAGATTC	Enduracidine	464
UF_miSeq_EndurF07	CARGGVATGCCVAARCARTTY	ACAGCAGA	SAASGTYTCBCCRTASGARAA	TATCAGCA	Enduracidine	464
UF_miSeq_EndurF08	MASGGSATGCCSMASMASTWS	ACACAGAA	BARBGTYTCBCCRTARCTRAA	CGCATACA	Enduracidine	464
UF_miSeq_EndurF09	MANGGSATGCCSMASMASTWB	CTCAATGA	SABYGDSTCSCCSWARGDSAA	AAGGACAC	Enduracidine	464
UF_miSeq_EndurF10	MANGGNATGCCSMASMASTWB	CAAGGAGC	SABYGDSTCSCCSWAYGDSWA	CCATCCTC	Enduracidine	464
UF_miSeq_EndurF11	MANGGNATGCCSMANMASTWB	ATGCCTAA	SABSGDSTCSCCVWASGDVWA	CCAGTTCA	Enduracidine	464
UF_miSeq_EndurF12	MANGGNATGCCSMANMANTWB	CACCTTAC	SABSGDSTCSCCVWANGDVWA	ACCACTGT	Enduracidine	464
UF_miSeq_IsoNitriF01	CCAAGCATTACTKCGYTGRAT	AGGCTAAC	AGCCAATTCATTTTATTCTCC	AGGCTAAC	Isonitrile	
UF_miSeq_IsoNitriF02	GAYATYAGYAGCAGCGGSGTS	CCATCCTC	ACTSGGGTTRTGSGGRAARTT	CCATCCTC	Isonitrile	
UF_miSeq_IsoNitriF03	GAYATHAGYAGYAGYGGVGTS	GTACGCAA	RCTSGGRTTRTGBGGRAARTT	GTACGCAA	Isonitrile	
UF_miSeq_IsoNitriF04	GAYATHAGYAGYAGYGGNGTN	GTGTTCTA	RCTNGGRTTRTGNGGRAARTT	GTGTTCTA	Isonitrile	
UF_miSeq_IsoNitriF05	GAYATYAGYAGYAGYGGSGTS	GATAGACA	RCTSGGRTTRTGSGGRAARTT	TAGGATGA	Isonitrile	
UF_miSeq_IsoNitriF06	GASCCSVTSMASTWSVTSVTS	CAACCACA	SWASGCSTKYGDSCGSTKSAB	CAAGGAGC	Isonitrile	
UF_miSeq_IsoNitriF07	GANCCSVTSMASTWBVTSVTS	CAGATCTG	VWASGCSTKSGDSCGSTKNAB	AACGCTTA	Isonitrile	
UF_miSeq_IsoNitriF08	GANCCNVTSMASTWBVTSVTS	CGCATACA	VWASGCSTKSGDSCGNTKNAB	CAACCACA	Isonitrile	
UF_miSeq_IsoPenF01	GTCCTGGTGYTAYCTGAACCC	TGGAACAA	CTCGACCTGSAGGTTCTGYAC	TGGAACAA	Isopenicillin	
UF_miSeq_IsoPenF02	GTCCTKSTGYTAYCTGAACCC	CAGCGTTA	CTCGAMCTGSAGGKTCTGYAC	CAGCGTTA	Isopenicillin	

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_IsoPenF03	RTCCTKSYSYTAYCYGAACCC	TAGGATGA	CTCSAMCTGSAGGKTCKSYAC	TAGGATGA	Isopenicillin	
UF_miSeq_IsoPenF04	AGYTGGTGYTAYYTVAAYCCV	AAGAGATC	YTCBACYTGBARRTTYTGBAC	AAGAGATC	Isopenicillin	
UF_miSeq_IsoPenF05	AGYTGGTGYTAYYTNAAYCCN	AGTCACTA	YTCNACYTGNARRTTYTGNAC	AGTCACTA	Isopenicillin	
UF_miSeq_IsoPenF06	HCRTWSTGYTWSVTSMASCCS	ACAGCAGA	STCSABSTKSABSTKSTKSAB	AGCAGGAA	Isopenicillin	
UF_miSeq_IsoPenF07	HCSTWSTGYTWSVTSMASCCS	CCAGTTCA	STCSABSTKSABSTKSTKSAB	CAAGGAGC	Isopenicillin	
UF_miSeq_IsoPenF08	HCNTWBTGYTWBVTSMASCCS	AGTCACTA	STCSABSTKSABSTKSTKNAB	GCTCGGTA	Isopenicillin	
UF_miSeq_NRPADF01	CGTGATGTACACGTCCGGCTC	CGACACAC	CACGAACCGYTCCGCSGTCAG	CGACACAC	NRPAD	707
UF_miSeq_NRPADF02	CGTSATSTACACSTCCGGCTC	CCGAAGTA	CASGAACCGYTCSGCSGTCAG	CCGAAGTA	NRPAD	707
UF_miSeq_NRPADF03	YSTSAYSTACACSTCCGGCWC	AGCACCTC	SASGAAMCGYTCSGCSGTCAG	AGCACCTC	NRPAD	707
UF_miSeq_NRPADF04	GTVATGTAYACVAGYGGVAGY	GATGAATC	BACRAABCKYTCBGCBGTBAR	CTGAGCCA	NRPAD	707
UF_miSeq_NRPADF05	GTNATGTAYACNAGYGGNAGY	GACAGTGC	NACRAANCKYTCNGCNGTNAR	CTGGCATA	NRPAD	707
UF_miSeq_NRPADF06	VTSATGTWSHCRHCRGGSACY	AGTACAAG	SACRAASCKYTCBGCBGTBAR	AACCGAGA	NRPAD	707
UF_miSeq_NRPADF07	VTSATGTWSHCRHCRGGSHCR	GCTCGGTA	SABSWASCGSTCSGCYGDSAB	GTCGTAGA	NRPAD	707
UF_miSeq_NRPADF08	VTSATGTWBHCSHCSGGSHCS	CGCATACA	SABVWASCGSTCSGCSGDNAB	GAGTTAGC	NRPAD	707
UF_miSeq_NRPADF09	VTNATGTWBHCSHCSGGSHCS	AAACATCG	SABVWASCGSTCSGCNGDNAB	AGCCATGC	NRPAD	707
UF_miSeq_NRPCF01	YCGYCACGATATCCTGCGCAC	GCTAACGA	CAGCACSGTRCCGAACACCAC	GCTAACGA	NRPC	716
UF_miSeq_NRPCF02	NCGBCAYGAYATCCTGCGCAC	AATCCGTC	CARCACVGTRCCGAAYACCAC	AATCCGTC	NRPC	716
UF_miSeq_NRPCF03	NCGBCAYGAYATYCTGCGCAC	GTCTGTCA	CARCACVGWVCCGAABACCAC	GTCTGTCA	NRPC	716
UF_miSeq_NRPCF04	NCGBCAYGAYATYYTRCGYAC	GAGTTAGC	YARCACVGWVCCRAABACSAC	GAGTTAGC	NRPC	716
UF_miSeq_NRPCF05	MGVCAYGAYATHYTVMGVACV	ATTGAGGA	BARBACBGTBCCRAABACBAC	GAATCTGA	NRPC	716
UF_miSeq_NRPCF06	MGNCAYGAYATHYTNMGNACN	ATTGGCTC	NARNACNGTNCCRAANACNAC	CAAGACTA	NRPC	716
UF_miSeq_NRPCF07	CGSCAYGACATCCTSCGSACS	AAACATCG	SAGSACGGTSCCGAASACCAC	GAGTTAGC	NRPC	716
UF_miSeq_NRPCF08	MGSCAYGACATCCTSMGSACS	CAGATCTG	SAGSACSGTSCCGAASACSAC	TGAAGAGA	NRPC	716
UF_miSeq_NRPCF09	MGSCAYGAYATYYTSMGVACS	GCTCGGTA	SARSACSGTBCCRAASACBAC	GACTAGTA	NRPC	716
UF_miSeq_NRPCF10	MGSCAYGAYATHYTVMGVACS	AGATGTAC	BARBACBGTBCCRAASACBAC	AGATCGCA	NRPC	716
UF_miSeq_NRPCF11	CGSMASGASVTSVTSCGSHCR	AAGGACAC	BARBACBGTBCCRAASACNAC	CGCTGATC	NRPC	716
UF_miSeq_NRPCF12	CGNMASGASVTSVTSCGSHCS	ATAGCGAC	BARBACBGTBCCRAANACNAC	GGAGAACA	NRPC	716
UF_miSeq_NRPCF13	CGNMANGASVTSVTSCGSHCS	TCTTCACA	SABSABYGDSCCSWASABSAB	ACGCTCGA	NRPC	716
UF_miSeq_PEPF01	GAGGACAAGRTCTTCCCSAAG	TGGTGGTA	GCATGCTGTGGTTGGCCCAGA	TGGTGGTA	Phosphono	449

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_PEPF02	GARGACAAGVTSTTCCCSAAG	TATCAGCA	GCAKSSCSTGGTTSGCGTAGA	TATCAGCA	Phosphono	449
UF_miSeq_PEPF03	GARGAYAAGVWVTTYCCBAAR	ATCCTGTA	GCAKSSYSTGGTTSGCSYAGA	ATCCTGTA	Phosphono	449
UF_miSeq_PEPF04	GARGAYAARGTSTTYCCSAAR	CCAGTTCA	BARSGCRTGRTTSGCRTARAT	GAGCTGAA	Phosphono	449
UF_miSeq_PEPF05	GARGAYAARGTNTTYCCNAAR	CTAAGGTC	NARNGCRTGRTTNGCRTADAT	CTGAGCCA	Phosphono	449
UF_miSeq_PEPF06	GAGGACAAGSWSTTCCCSAAG	CACTTCGA	SAGSGCGTGGTTSGCGTARAT	AAGACGGA	Phosphono	449
UF_miSeq_PEPF07	GARGAYAAGSWVTTYCCSAAR	GAGTTAGC	SAGSGCRTKRTKSGCRTASAT	ACGCTCGA	Phosphono	449
UF_miSeq_PEPF08	GASGASMASVTSTWSCCSMAS	GCTAACGA	SABSGCSTKSTKSGCSTASAT	AGCAGGAA	Phosphono	449
UF_miSeq_PEPF09	GANGASMASVTSTWBCCSMAS	GATAGACA	SABSGCSTKSTKSGCSWASAB	TGAAGAGA	Phosphono	449
UF_miSeq_PEPF10	GANGANMASVTSTWBCCSMAS	ACAGCAGA	SABSGCSTKSTKSGCVWANAB	CCATCCTC	Phosphono	449
UF_miSeq_PEPF11	GANGANMANVTSTWBCCSMAS	ACAAGCTA	SABSGCSTKSTKNGCVWANAB	CCGACAAC	Phosphono	449
UF_miSeq_PIPF01	CTGCCGACSATCATCGGCACS	GCTCGGTA	TSGCSGTGCAGATGATGTCGG	GCTCGGTA	PipecolicAcid	
UF_miSeq_PIPF02	CTSCCSACSATCRTCGGVACS	CAATGGAA	TSRCSGTGCASAKSAYGTCGG	CAATGGAA	PipecolicAcid	
UF_miSeq_PIPF03	CTVCCVACVATVATHGGCACV	CCTCCTGA	SGTSGCBGTRCADATDATRTC	CTGGCATA	PipecolicAcid	
UF_miSeq_PIPF04	CTVCCVACNATVATHGGCACN	CGACTGGA	BGTBGCBGTRCADATDATRTC	GAATCTGA	PipecolicAcid	
UF_miSeq_PIPF05	YTNCCNACNATHATHGGNACN	CAAGGAGC	NGTNGCNGTRCADATDATRTC	CAAGACTA	PipecolicAcid	
UF_miSeq_PIPF06	CTGCCSACSATSATCGGCACS	ACGCTCGA	SGTSGCSGTRCAGATGATGTC	CCATCCTC	PipecolicAcid	
UF_miSeq_PIPF07	CTSCCSACSATVATCGGCACS	CGAACTTA	SGTSGCSGTRCARATRATGTC	ATCATTCC	PipecolicAcid	
UF_miSeq_PIPF08	CTVCCSACSATVATHGGCACS	AGATGTAC	BGTSGCSGTRCADATDATRTC	GTGTTCTA	PipecolicAcid	
UF_miSeq_PIPF09	YTSCCVACVATHATHGGVACV	TCTTCACA	BGTNGCSGTRCADATDATRTC	GACTAGTA	PipecolicAcid	
UF_miSeq_PIPF10	YTNCCVACVATHATHGGVACV	ACTATGCA	YGDSGCYGDRCASABSABSTC	CGAACTTA	PipecolicAcid	
UF_miSeq_PIPF11	VTSCCSHCRVTSVTSGGSHCR	GACAGTGC	SGDSGCSGDRCASABSABNTC	AGATGTAC	PipecolicAcid	
UF_miSeq_PIPF12	VTNCCSHCSVTSVTSGGSHCS	CTCAATGA	SGDSGCSGDRCASABNABNTC	AATGTTGC	PipecolicAcid	
UF_miSeq_PKSATF01	GCCGGGCACTCSATCGGTGAG	CCTCTATC	CGCAGGTTSCGGTACCAGTAK	CCTCTATC	PKSAT	537
UF_miSeq_PKSATF02	GTSGGSCAYTCSCAGGGBGAG	CATACCAA	CGCAGGTTSCKGWACCAGTAK	CATACCAA	PKSAT	537
UF_miSeq_PKSATF03	GYSGGSCAYTCSVWSGGBGAG	ACACGACC	CGCASRTKSCKGWMCCAGTAB	ACACGACC	PKSAT	537
UF_miSeq_PKSATF04	GTSGGSCAYAGYCARGGSGAR	TGGCTTCA	SCKSARRTTSCKRTACCARTA	GAGCTGAA	PKSAT	537
UF_miSeq_PKSATF05	GTNGGNCAYAGYCARGGNGAR	CCATCCTC	NCKNARRTTNCKRTACCARTA	GATAGACA	PKSAT	537
UF_miSeq_PKSATF06	GTSGGSCACAGCCARGGVGAG	ACCACTGT	SCGSARGTTSCGGTACCARTA	AGCAGGAA	PKSAT	537
UF_miSeq_PKSATF07	GTVGGVCAYAGYCARGGVGAG	ACAAGCTA	SCGSASSTTSCGSTASWASWA	ACGTATCA	PKSAT	537

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_PKSATF08	STSGGSCASHCRMASVTSCGH	GTCTGTCA	SCGSABSTKSCGSWASWAVWA	TGGCTTCA	PKSAT	537
UF_miSeq_PKSATF09	VTSGGSMASHCRMASVTSCGH	CCGACAAC	SCGSABSTKSCGVWAVWAVWA	CGACTGGA	PKSAT	537
UF_miSeq_PKSATF10	VTNGGSMASHCSMASVTSCGH	CAACCACA	SCGSABSTKNCGVWAVWAVWA	AAGAGATC	PKSAT	537
UF_miSeq_PKSKSF01	TTCGACGCCGRRTTCTTCGGC	TTCACGCA	TKCCGGTGCCGTGCAGYTCGA	TTCACGCA	PKSKS	723
UF_miSeq_PKSKSF02	TTCGACGCSGSMTTCTTCGRC	CGCATACA	TSCCGGTGCCGTGCSCCTCSA	CGCATACA	PKSKS	723
UF_miSeq_PKSKSF03	TTCGACGCBGVSTTCTTCGGS	ATCATTCC	TVCCGGTGCCGTGSSYCTCSA	ATCATTCC	PKSKS	723
UF_miSeq_PKSKSF04	TTCGACGCBGVSTTCTTCGVS	CTCAATGA	TNCCGGTGCCGTGCVSYTCSA	CTCAATGA	PKSKS	723
UF_miSeq_PKSKSF05	TTCGACGCBKVVTTCTTCGVS	AATGTTGC	TNCCSGTRCCGTGSVSYTCSA	AATGTTGC	PKSKS	723
UF_miSeq_PKSKSF06	TTCGASSCBKVVTTCTTCGVS	ATAGCGAC	TNCCSGTRCCGTGSVBYTCSA	ATAGCGAC	PKSKS	723
UF_miSeq_PKSKSF07	TTYGAYGCNGARTTYTTYGGN	GTACGCAA	NCCNGTNCCRTGNARYTCNAC	GCCACATA	PKSKS	723
UF_miSeq_PKSKSF08	TWSGASGCSGASTWSTWSGGS	AATGTTGC	SCCYGDSCCSTKSASSTCSAS	CCGAAGTA	PKSKS	723
UF_miSeq_PKSKSF09	TWBGASGCSGASTWBTWBGGS	AACCGAGA	SCCYGDSCCSTKSABSTCSAB	ACAGCAGA	PKSKS	723
UF_miSeq_PKSKSF10	TWBGANGCSGASTWBTWBGGS	CTCAATGA	SCCSGDSCCSTKSABSTCNAB	GTCTGTCA	PKSKS	723
UF_miSeq_PKSKSF11	TWBGANGCNGASTWBTWBGGS	GACAGTGC	SCCSGDSCCSTKSABNTCNAB	ACAAGCTA	PKSKS	723
UF_miSeq_PKSKSF12	TWBGANGCNGANTWBTWBGGS	AGTGGTCA	SCCSGDSCCSTKNABNTCNAB	GTACGCAA	PKSKS	723
UF_miSeq_T2PKSF01	TSGCCTGCTTCGACGCCATCM	CCGTGAGA	TAGTCSAGRTCGCACTCGGGG	CCGTGAGA	T2PKS	540
UF_miSeq_T2PKSF02	TSGCCTGYTTCGACGCSATCM	CCTAATCC	TAGTCSAGRTCGCASTCSGGG	CCTAATCC	T2PKS	540
UF_miSeq_T2PKSF03	TSGCSTGYTTCGAYGCSATCM	AACTCACC	TAGTCSAGRTCGCABKCSGGR	AACTCACC	T2PKS	540
UF_miSeq_T2PKSF04	TSKCSTGYTTCGAYRCSMTSM	AAGGACAC	TAGTCSAGRTCGMRBKSSGGR	AAGGACAC	T2PKS	540
UF_miSeq_T2PKSF05	GTNGCNTGYTTYGAYGCNATH	GTGTTCTA	RTARTCNARRTCNARYTGNGG	GCGAGTAA	T2PKS	540
UF_miSeq_T2PKSF06	GTVGCSTGYTTYGAYGCSATY	CAAGACTA	GTARTCSARRTCSARYTGSGG	CGCATACA	T2PKS	540
UF_miSeq_T2PKSF07	GTSGCVTGYTTYGAYGCVATH	AACCGAGA	RTARTCBARRTCBARYTGSGG	ACCTCCAA	T2PKS	540
UF_miSeq_T2PKSF08	VTSGCSTGYTWSGASGCSVTS	TGGTGGTA	SWASTCSABSTCSABSTKSGG	TAGGATGA	T2PKS	540
UF_miSeq_T2PKSF09	VTNGCSTGYTWBGASGCSVTS	CAGCGTTA	VWASTCSABSTCSABSTKNGG	CCGACAAC	T2PKS	540
UF_miSeq_T2PKSF10	VTNGCNTGYTWBGASGCSVTS	ACGCTCGA	VWASTCSABSTCSABNTKNGG	ATTGGCTC	T2PKS	540
UF_miSeq_T2PKSF11	VTNGCNTGYTWBGANGCSVTS	AACTCACC	VWASTCSABSTCNABNTKNGG	CAATGGAA	T2PKS	540
UF_miSeq_TDF01	STGCGGGTGCTGCCSGACGAC	AACGTGAT	SGCGTASAGGTACTGCAGC	AACGTGAT	TryptophanDimer	487
UF_miSeq_TDF02	STGCGGGTGCTGCCSGACSAC	AAACATCG	SGCGTASAGGTACTGVAGC	AAACATCG	TryptophanDimer	487
UF_miSeq_TDF03	STSCGGGTSCTGCCSGACSAC	ATGCCTAA	SGCGTASAGRTACTGVAKC	ATGCCTAA	TryptophanDimer	487

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_TDF04	GTSMGVGTSCTGCCSGACSAC	ACATTGGC	SGCGTASRKRTACTGVAGC	ACATTGGC	TryptophanDimer	487
UF_miSeq_TDF05	STSMGVGTSCTSCCSGACSAC	CAGATCTG	SGCGTASRKRTACTGVAKC	CAGATCTG	TryptophanDimer	487
UF_miSeq_TDF06	STNMGVGTSCTNCCSGAYSAY	GATAGACA	SGCGTAVRKRTAYTGVAKCAT	GATAGACA	TryptophanDimer	487
UF_miSeq_TDF07	GTNMGNGTNYTNCCNGAYCAY	GCCACATA	NGCRTANARRTAYTGNARCAT	GCCACATA	TryptophanDimer	487
UF_miSeq_TDF08	STNMGNGTNYTNCCNGAYSAY	GCGAGTAA	NGCRTANRDRTAYTGNADCAT	GCGAGTAA	TryptophanDimer	487
UF_miSeq_TDF09	STVMGSGTSCTSCCSGAYSAY	ATTGAGGA	BGCRTASRRRTAYTGBARCAT	ACACAGAA	TryptophanDimer	487
UF_miSeq_TDF10	STVMGVGTSCTVCCSGAYSAY	GACTAGTA	SGCSWASABSWASTKSABCAT	TAGGATGA	TryptophanDimer	487
UF_miSeq_TDF11	VTSCGSVTSVTSCCSGASMAS	ACCTCCAA	SGCVWASABVWASTKSABCAT	TCCGTCTA	TryptophanDimer	487
UF_miSeq_TDF12	VTNCGSVTSVTSCCSGASMAS	CGACACAC	SGCVWASABVWASTKNABCAT	AAGGACAC	TryptophanDimer	487
UF_miSeq_TDF13	VTNCGNVTSVTSCCSGASMAS	GTCGTAGA	SGCVWASABVWANTKNABCAT	ACATTGGC	TryptophanDimer	487
UF_miSeq_TeF01	TCTTCGGSCACAGCATGGGSG	GGAGAACA	AGTCGSCSCGCAGCGCGGGCA	GGAGAACA	Thioesterase	
UF_miSeq_TeF02	TSTTCGGGCACAGCMTSGGSG	CACCTTAC	AGTCGSCSCGCAGCGSSGGCA	CACCTTAC	Thioesterase	
UF_miSeq_TeF03	TSTTCGGSCACAGCMTSGGBG	GGTGCGAA	AGTCGSYSCGCAGCGSSGGCA	GGTGCGAA	Thioesterase	
UF_miSeq_TeF04	TBTWYGGSCACAGCMTSGGBG	TGAAGAGA	AGTCGSYSCGSASSGSSGGCA	TGAAGAGA	Thioesterase	
UF_miSeq_TeF05	TTYTTYGGNCAYAGYATGGGN	AAGAGATC	RTCNCCNCKNARNGCNGGNAR	GATAGACA	Thioesterase	
UF_miSeq_TeF06	TWSTWSGGSMASHCRATGGGS	TCCGTCTA	STCSCCSCGSABSGCSGGSAB	CACCTTAC	Thioesterase	
UF_miSeq_TeF07	TWBTWBGGSMASHCSATGGGS	AAGAGATC	STCSCCSCGSABSGCSGGNAB	CCTAATCC	Thioesterase	
UF_miSeq_TeF08	TWBTWBGGNMASHCSATGGGS	ACGTATCA	STCSCCSCGSABSGCNGGNAB	AGCCATGC	Thioesterase	
UF_miSeq_TeF09	TWBTWBGGNMANHCSATGGGS	AATGTTGC	STCSCCSCGSABNGCNGGNAB	GTCTGTCA	Thioesterase	
UF_miSeq_TeF10	TWBTWBGGNMANHCNATGGGS	CTCAATGA	STCSCCSCGNABNGCNGGNAB	AGGCTAAC	Thioesterase	
UF_miSeq_AryCF01	GASGAYCTGATCAGCMAGCTS	AGCACCTC	GCCSAGRCAGAARTGGATGCY	CCATCCTC	AryC	496
UF_miSeq_AryCF02	GASGATCTSAYCAGCMRSCTS	CCGTGAGA	RCCSAGRCARWARTGGATKCC	AAACATCG	AryC	496
UF_miSeq_AryCF03	GASGAYCTSAYCAGCMRSCTS	AAACATCG	RCCSAGRCARWARTGGATKCY	AATGTTGC	AryC	496
UF_miSeq_AryCF04	GASGASCTSATHAGYMRSCTS	GTGTTCTA	SCCSAGRCARAARTGDATNGA	ACGTATCA	AryC	496
UF_miSeq_AryCF05	GANGASCTSATHAGYMRSCTS	CGCATACA	SCCNAGRCARAARTGDATNGA	GACAGTGC	AryC	496
UF_miSeq_AryCF06	GANGANCTSATHAGYMRSCTS	CATCAAGT	NCCNAGRCARAARTGDATNGA	AGAGTCAA	AryC	496
UF_miSeq_AryCF07	GANGANCTNATHAGYMRSCTS	AGTACAAG	SCCSABRCASWASTKSABYGD	ACCACTGT	AryC	496
UF_miSeq_AryCF08	GANGASVTSVTSHCSMASVTS	CGCTGATC	SCCSABRCAVWASTKSABNGD	CGGATTGC	AryC	496
UF_miSeq_CysJF01	TGGCCSCGSAAGCARTGGTTC	ACTATGCA	CATGTTGTCSAGCATSACSAC	AGCCATGC	CysJ	611

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_CysJF02	TGGCCSCGSAARCARTGGTTY	CCTCTATC	CATRTTRTCSAGCATSACSAC	ATCCTGTA	CysJ	611
UF_miSeq_CysJF03	TGGCCNCGSAARCARTGGTTY	CTGGCATA	CATRTTRTCSAGCATSACNAC	CAAGGAGC	CysJ	611
UF_miSeq_CysJF04	TGGCCNCGNAARCARTGGTTY	TCCGTCTA	CATRTTRTCSAGCATNACNAC	CAACCACA	CysJ	611
UF_miSeq_CysJF05	TWSCCSCGSSTSMASTWSTWS	CAGCGTTA	CATSTKSTCSABCATSABSAB	ACAGCAGA	CysJ	611
UF_miSeq_CysJF06	TWBCCSCGSVTSMASTWSTWS	TGGCTTCA	CATSTKSTCSABCATSABNAB	TAGGATGA	CysJ	611
UF_miSeq_CysJF07	TWBCCNCGSVTSMASTWSTWS	AAGGACAC	CATSTKSTCSABCATNABNAB	TGGTGGTA	CysJ	611
UF_miSeq_CysJF08	TWBCCNCGNVTSMASTWSTWS	GATGAATC	CATSTKSTCNABCATNABNAB	AGCCATGC	CysJ	611
UF_miSeq_GriFF01	GTSCAGGGCAGCGGVCCGGTY	CGAACTTA	GTACTGSCCGAYGAYSAGGTA	TGGTGGTA	GriF	373
UF_miSeq_GriFF02	GTBCAGGGCAGYGGVCCSGTY	AGTGGTCA	GTACTGSCCSAYRAYSAGRTA	AGAGTCAA	GriF	373
UF_miSeq_GriFF03	GTNCARGGYWGYGGSCCSGTS	GACTAGTA	GTAMTGSCCRAYRAYVASRTA	CAAGGAGC	GriF	373
UF_miSeq_GriFF04	GTNCARGGYWGYGGSCCBGTB	CTGGCATA	RTAYTGNCCNACNACSAGRTA	GAACAGGC	GriF	373
UF_miSeq_GriFF05	GTNCARGGNWGYGGNCCSGTS	ACCACTGT	SWASTKSCCSASSASSABSWA	TTCACGCA	GriF	373
UF_miSeq_GriFF06	GTNCARGGNWGYGGNCCNGTS	GTGTTCTA	SWASTKSCCSASSABSABSWA	ATTGAGGA	GriF	373
UF_miSeq_GriFF07	GTNCARGGNWGYGGNCCNGTN	ATTGAGGA	VWASTKSCCSABSABSABVWA	CGCATACA	GriF	373
UF_miSeq_GriFF08	VTNMANGGNHCSGGSCCSVTS	GATAGACA	VWASTKSCCSABSABNABVWA	AAGGACAC	GriF	373
UF_miSeq_HolEF01	GMACGCTTYTTTAATRTYGAT	AGTCACTA	WGGCCCRAGYCGATAWCCATC	GAGCTGAA	HolE	576
UF_miSeq_HolEF02	GSSCATTTTTTYRMTGTMGAT	CTGGCATA	WGGCCCRASYCGATAHCCATC	ATAGCGAC	HolE	576
UF_miSeq_HolEF03	GMRYRYTTYTTTAATRTYGAT	CCGTGAGA	WGGSCCSAGSCGRTASCCRTC	CCTAATCC	HolE	576
UF_miSeq_HolEF04	GARCGNTTYTTYAAYATHGAY	ATGCCTAA	SGGSCCSAGSSGRTANCCRTC	TAGGATGA	HolE	576
UF_miSeq_HolEF05	GASCGSTWSTWSMASVTSGAS	ACCACTGT	SGGSCCSAGNSGRTANCCRTC	CGCATACA	HolE	576
UF_miSeq_HolEF06	GANCGSTWSTWSMASVTSGAS	ATCCTGTA	SGGSCCNAGNSGRTANCCRTC	GAGTTAGC	HolE	576
UF_miSeq_HolEF07	GANCGNTWSTWSMASVTSGAS	GGAGAACA	SGGNCCNAGNSGRTANCCRTC	TATCAGCA	HolE	576
UF_miSeq_HolEF08	GANCGNTWNTWSMASVTSGAS	GTGTTCTA	NGGNCCNAGNSGRTANCCRTC	AACTCACC	HolE	576
UF_miSeq_NgnN4F01	GCGATGCWGGCGGYGCTGCGC	AACTCACC	TTCGGTCGGGCCATACAGGTT	ATCATTCC	NgnN4	576
UF_miSeq_NgnN4F02	GCSATGCTSGCSGYSCTSCGS	AATGTTGC	YTCSGTSGGSCCRTASAGRTT	CACTTCGA	NgnN4	576
UF_miSeq_NgnN4F03	GCNATGCTSGCSGYSCTSCGS	AGTACAAG	YTCSGTSGGSCCRTANAGRTT	GATAGACA	NgnN4	576
UF_miSeq_NgnN4F04	GCNATGCTNGCSGYSCTSCGS	ACAGATTC	YTCSGTNGGSCCRTANAGRTT	AATGTTGC	NgnN4	576
UF_miSeq_NgnN4F05	GCNATGVTSGCSGYSVTSCGS	AAGGACAC	STCYGDSGGSCCSWASABSTK	ATTGAGGA	NgnN4	576
UF_miSeq_NgnN4F06	GCNATGCTNGCNGYNCTSCGS	CGCATACA	STCSGDSGGSCCVWASABNTK	AGCACCTC	NgnN4	576

PrimerNameF	PrimerF	BarcodeF	PrimerR	BarcodeR	Target	Length
UF_miSeq_NgnN4F07	GCNATGVTNGCNGYSVTSCGS	GCTCGGTA	STCSGDSGGSCCVWANABNTK	CAATGGAA	NgnN4	576
UF_miSeq_NgnN4F08	GCNATGVTNGCNGYNVTSCGS	CGCTGATC	STCSGDSGGNCCVWANABNTK	CAGCGTTA	NgnN4	576
UF_miSeq_TmlUF01	TTRCCGYTKTTCCACGTCAAY	GAGTTAGC	GTTGTRGCCSCCGCGRATGAT	CAAGACTA	TmlU	714
UF_miSeq_TmlUF02	TTGCCGYTVTTYCAYGTCAAY	GAACAGGC	GTTGTRKCCSCCRCGRATRAT	CGAACTTA	TmlU	714
UF_miSeq_TmlUF03	TTRCCGYTDTTYCAYGTCAAY	ACATTGGC	GTTGTRKCCRCCDCGRATRAT	CCTCTATC	TmlU	714
UF_miSeq_TmlUF04	TTRCCGYTNTTYCAYGTCAAY	AAGGTACA	RTTRTGSCCSCCSCGRATDAT	AGGCTAAC	TmlU	714
UF_miSeq_TmlUF05	TTRCCRYTNTTYCAYGTYAAY	CAGCGTTA	RTTGTRBCCVCCDCGRATRAT	GCGAGTAA	TmlU	714
UF_miSeq_TmlUF06	CTNCCNCTNTTYCAYGTSAAY	AAACATCG	RTTRTGSCCNCCNCGDATDAT	AATCCGTC	TmlU	714
UF_miSeq_TmlUF07	CTNCCNCTNTTYCAYGTNAAY	TCTTCACA	RTTRTGNCCNCCNCGDATDAT	GTCGTAGA	TmlU	714
UF_miSeq_TmlUF08	VTNCCSVTSTWSMASVTSMAS	ATCCTGTA	STKSTKSCCSCCSCGNABNAB	AACGTGAT	TmlU	714

Analyzing the rarefaction curves we observed the set of primer with the highest score of alpha diversity in the rarefaction curve. For AHBA the best primer set was AHBA04 while PEP04, TD09 and CP10 were the best primer sets for phophono-containing natural products, tryptophan dimmers and capreomycidine (**Table 7**).

Table 7: Best primer set for each target

Target	*Forward Primer*	*Reverse Primer*
AHBA	UF_miSeq_AHBAF04	UF_miSeq_AHBAR04
Aminoglycoside	UF_miSeq_2DOIF05	UF_miSeq_2DOIR05
AryC	UF_miSeq_AryCF01	UF_miSeq_AryCR01
Capreomycidine	UF_miSeq_CPF10	UF_miSeq_CPR10
CysJ	UF_miSeq_CysJF03	UF_miSeq_CysJR03
GriF	UF_miSeq_GriFF04	UF_miSeq_GriFR04
NRPAD	UF_miSeq_NRPADF03	UF_miSeq_NRPADR03
NRPC	UF_miSeq_NRPCF04	UF_miSeq_NRPCR04
Phosphono	UF_miSeq_PEPF04	UF_miSeq_PEPR04
PKSAT	UF_miSeq_PKSATF04	UF_miSeq_PKSATR04
PKSKS	UF_miSeq_PKSKSF04	UF_miSeq_PKSKSR04
T2PKS	UF_miSeq_T2PKSF03	UF_miSeq_T2PKSR03
Tryptophan Dimer	UF_miSeq_TDF09	UF_miSeq_TDR09
HolE	UF_miSeq_HolEF04	UF_miSeq_HolER04
TmlU	UF_miSeq_TmlUF06	UF_miSeq_TmlUR06

4.4 Mining natural products diversity in metalibrary

The nonribosomal peptides and polyketides are far away the most studied natural products families. Our focus was to design, identify and validate new sets of primers capable of amplifying rare natural products classes, including the usual PKS and NRPS. The multiple sequence alignment of each natural product enzyme provides us an initial view of conservation of each one of those related genes. For the validation of this tool, we selected to work with natural products containing AHBA and tryptophan dimmers.

The AHBA group is associated with the presence of Ansamycins. Ansamycins are a family of macrolactams that are biosynthesized by a type I PKS using a molecule of 3-amino-5-hydroxybenzoic acid as starter unit (Rinehart e Shield, 1976). Several AHBA-containing natural products molecules have shown powerful activities, such as antimicrobial, antifungal, anticancer and antiviral. Rifamycin is an ansamycin molecule originally biosynthesized by *Amycolatopsis rifamycinica* that showed potent antibacterial activity inhibiting bacterial DNA-dependent RNA polymerase, suppressing RNA synthesis (Sensi et al. 1959). Wang et al. (2013), designed primers set capable of amplifying AHBA synthase and revealed the biosynthetic potential of ansamycins and the complexity of AHBA-containing natural products in actinomycetes. They identified 25 ansamycins using PCR screening, including eight new compounds (Wang *et al.*, 2013). One endpoint was the degeneracy of the primers sets used in their work. The calculated degeneracy is low and provides a fractional access to an AHBA-synthase diversity. However, the ansamycins continue to be a mysterious group.

Table 8: AHBA families with acceptable e-value

AHBA group	***e-value***
Ansamitocin	1,00E-46
Ansatrienin	1,00E-46
Rifamycin	4,00E-86
Chaxamycin	2,00E-84
Divergolide	3,00E-73
Hygrocin	2,00E-49
Macbecin	5,00E-56
Napthomycin	4,00E-71
Neoansalactam	1,00E-62
Rubradirin	2,00E-55

Following the blast against MiBIG database (Medema *et al.*, 2015), we detected some AHBA-containing PK in the metalibrary DNA. All results were obtained after a filtering step to remove the unrelated sequences. We observed the presence of several groups of ansamycins including ansamitocin, ansatrienin, chaxamycin, divergolide, hygrocin, macbecin, napthmycin, rifamycin and rubradirin (**Figure 22**). It is important to highlight that e-value obtained from those AHBA-containing NP were low but representative to characterize families of molecules (**Table 8**). Rifamycin molecules obtained the lowest e-values, followed by chaxamycin and divergolide. Although, all AHBA-containing PK mined in the metalibray had an expected e-value.

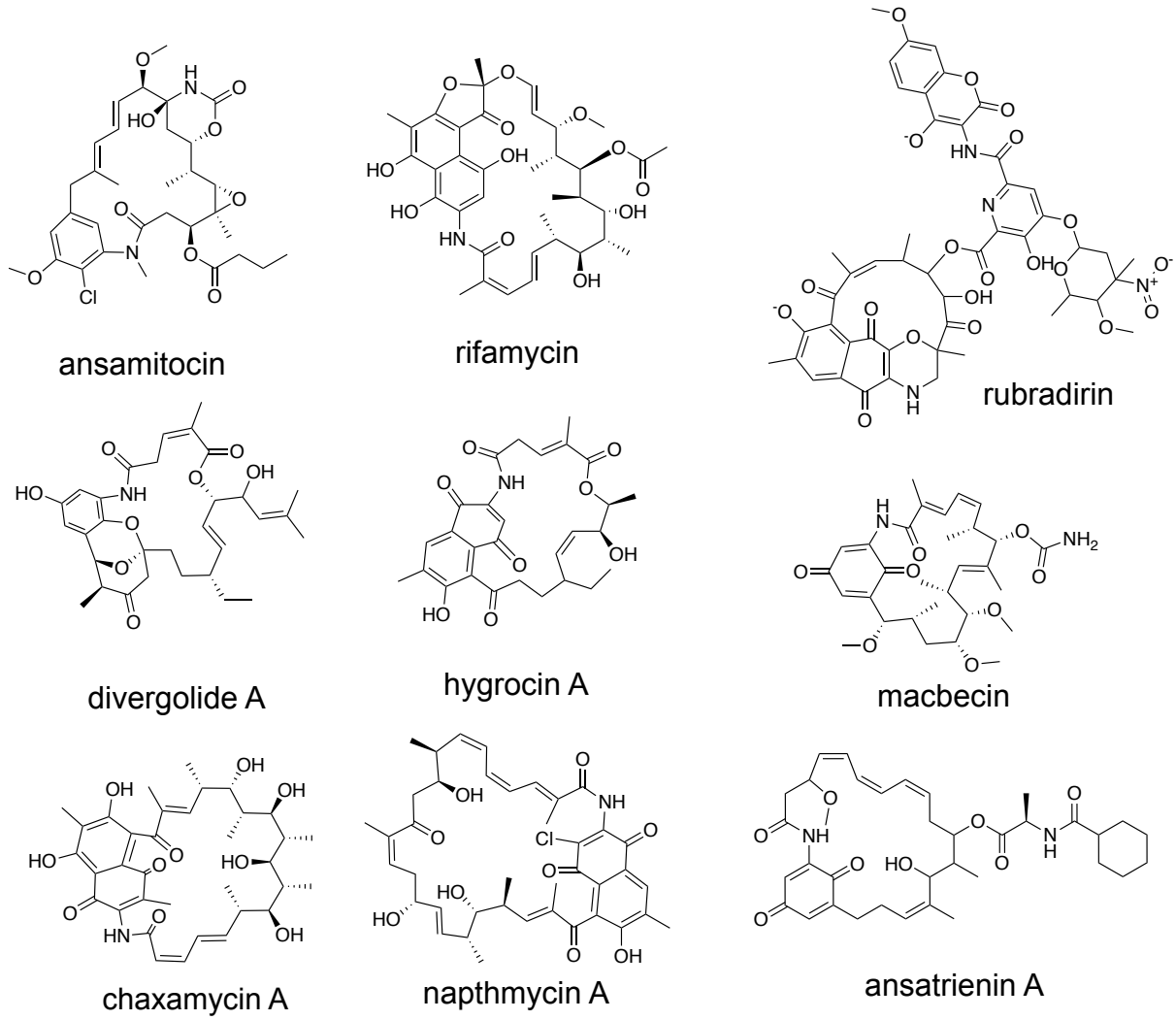


Figure 22: AHBA-containing PK inferred from metalibrary

A number of bacterial natural products result from the coupling of two tryptophan's (Ryan e Drennan, 2009). Attributed to as tryptophan dimers (TDs), these molecules take on a diversity of distinct chemical structures, depending on whereby the two tryptophans dimerize and functionalize (**Figure 23**).

TDs are also found to be frequently connected with clinically important biological activities, implying that the TD motif may be a privileged natural substructure (Sanchez *et al.*, 2006). The multiple bioactivities of TDs have been considered to come from their ability to target protein kinases or other target containing ATP-like binding sites, such as DNA topoisomerases, ABC transporter, and intercalative sites of DNA (Nakano e Omura, 2009). In special, staurosporine is an effective protein kinase inhibitor (Prade *et al.*, 1997), and its discovery has initiated the research of kinase inhibitors as potential anticancer drugs (Nakano e Omura, 2009). Rebeccamycin is a TD that is structurally similar to staurosporine, but it inhibits DNA topoisomerase I alternatively (Staker *et al.*, 2005). Natural and synthetic derivatives of staurosporine and rebeccamycin have registered clinical trials for cancer, neurodegenerative disorders, and diabetes-associated pathologies (**Figure 23**) (Sausville *et al.*, 2001; Carvajali *et al.*, 2007; Butler, 2008; Schwandt *et al.*, 2012). From a practical viewpoint, TD class is especially manageable to genome-based natural product discovery because the known TD biosynthetic gene clusters are comparatively short in length (10-30 kbp), giving them genetically moldable for small molecule expression (Sanchez *et al.*, 2006). Furthermore, the tryptophan family provides active UV compounds easily detect and isolate (Sanchez *et al.*, 2006).

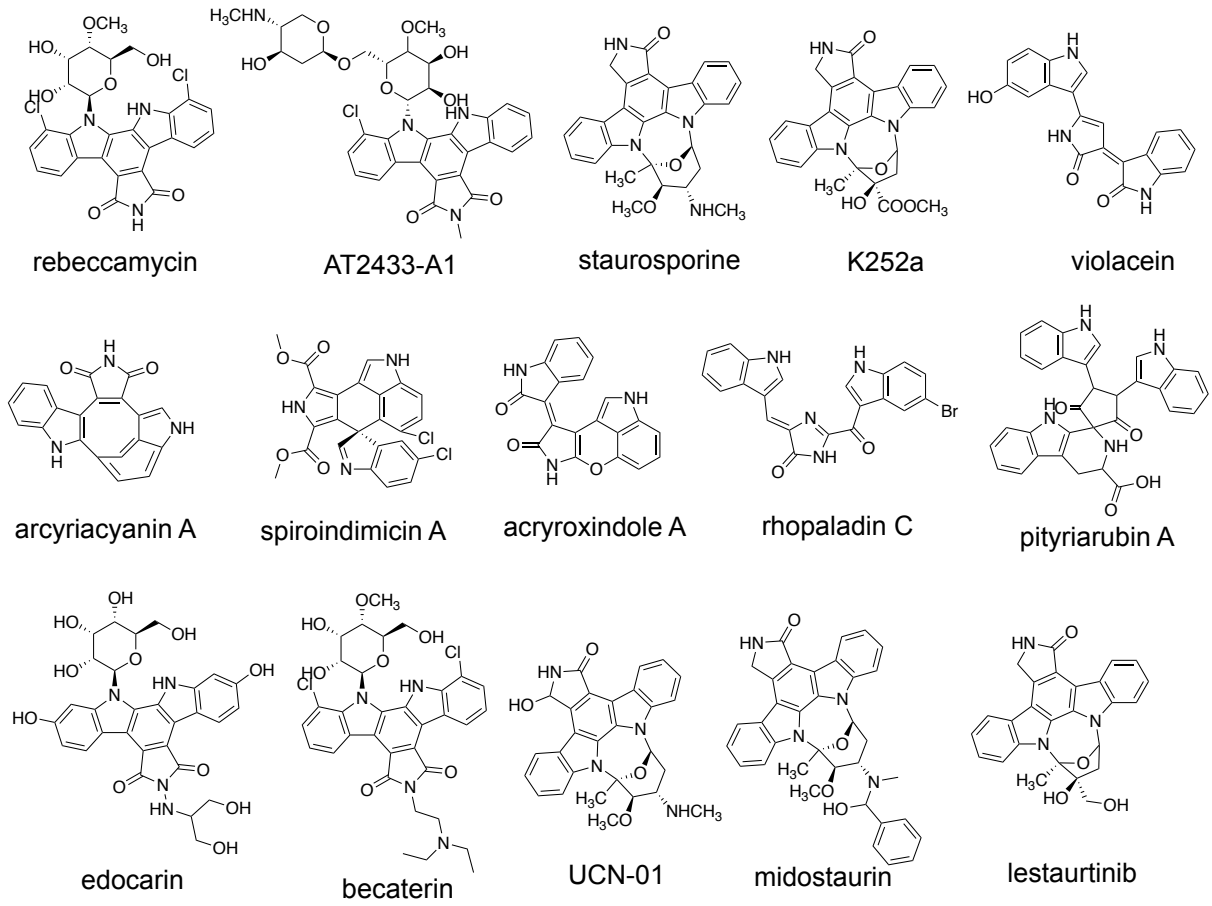


Figure 23: Tryptophan dimmers molecules inferred from metagenomic libraries

Our rarefaction curves allowed to identify the primer set which is responsible to obtain the highest diversity for each target. The degenerated primers designed to amplify AHBA synthase and CPA synthase were competent to amplify several molecules. The phylogeny of this molecules was defined by the alpha diversity in the rarefaction curve. Thus, we suggest that some of the designed primers were capable of increasing the range of distinct sequences to amplify. Thus, partially, for the ansamycins and tryptophan dimmers, the objectivity of the package was concluded.

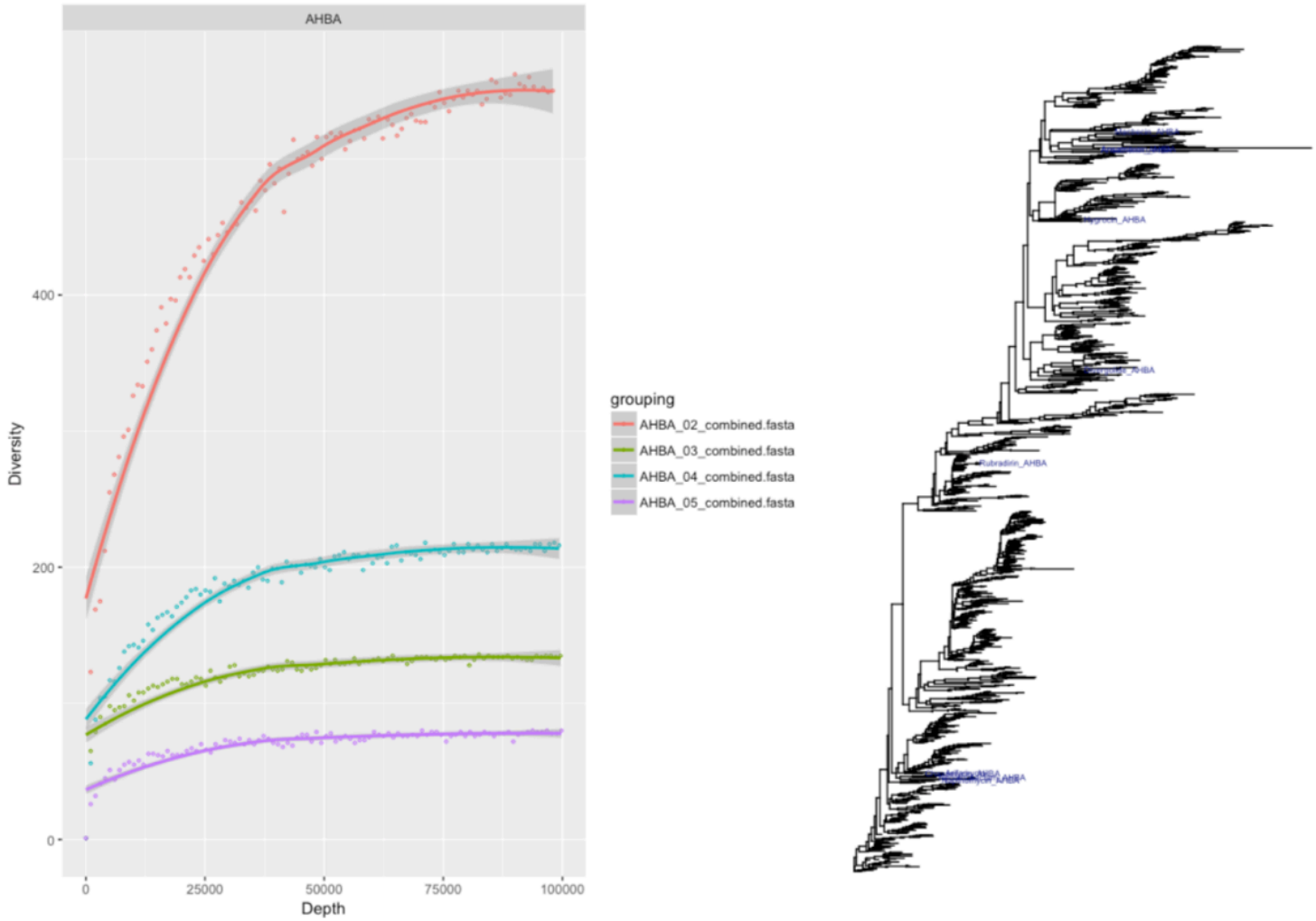


Figure 24: Phylogeny of AHBA obtained from degenerated primers
The rarefaction curves (left) highlight the diversity (y-axis) vs depth (x-axis). In right the phylogenetic analysis show the presence of new families in the clades. This may suggest the presence of new families of AHBA.

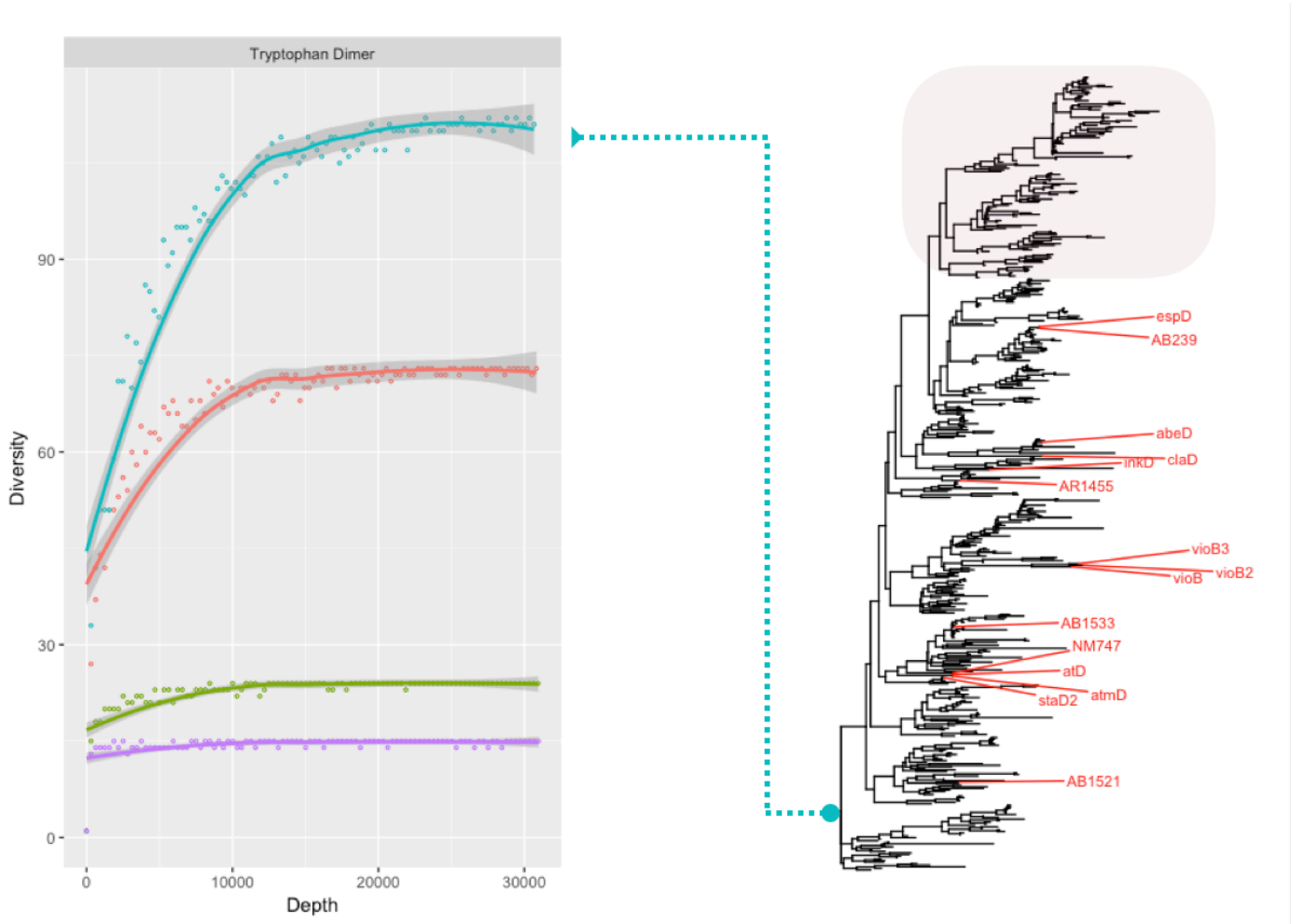


Figure 25: Phylogeny of TD obtained using degenerated primers
The rarefaction curves (left) highlight the diversity (y-axis) vs depth (x-axis). In right the phylogenetic analysis show the presence of new families in the clades. This may suggest the presence of new families of TD. The red box show an example of new clade annotated.

The analysis of phylogenetic distribution of the OTUs across the data permitted to infer the amplification of unknown sequences in the data. For AHBA-containing PK (**Figure 24**) we observed in the bottom of the phylogenetic tree the presence of clades unrelated to known sequences. The same results was observed for TD (**Figure 25**). For TD we observed the presence of a clade unrelated to the known sequences in the top of phylogenetic tree. These results suggested that SecMetPrimer was able to design primers capable to increase the range to recovery NP information from a target.

4.5 Sample collection and eDNA isolation

Twenty-five distinct soil/sediment samples were collected in the Cerrado, Atlantic Forest and Marine Sediments (**Figure 26**). The samples were randomly collected in environmental reserves. The competent offices declared all certificates for collection and access to genetic diversity.

Power Clean MoBio® kit was essential due to the concentration of humic acid in some samples, mainly collected in soils from Cerrado. Through the trials, the influence of humic acid may cause the inactivation of enzymes such as ligases and polymerases, thus endangering any subsequent experiment (Roh *et al.*, 2006).

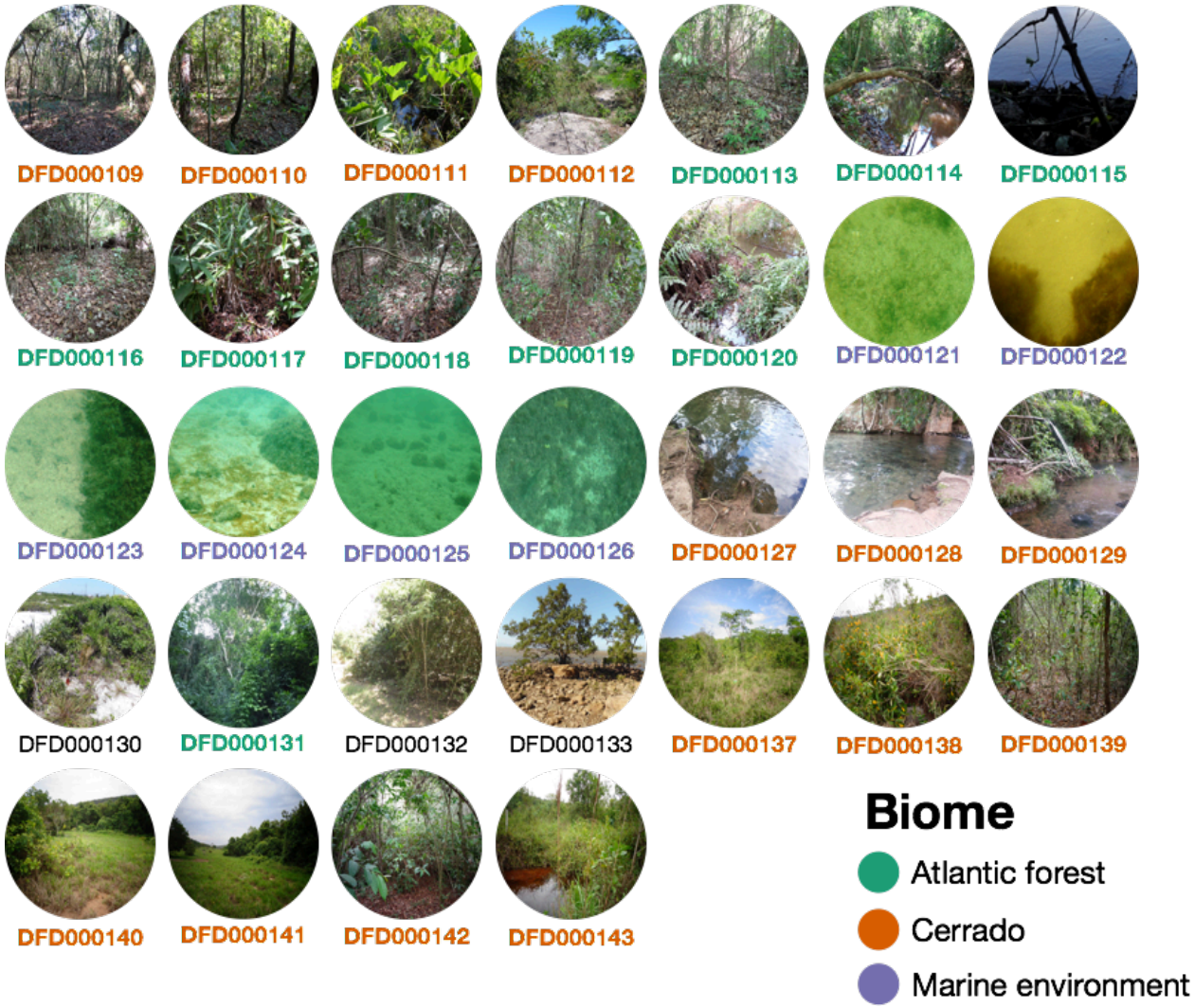


Figure 26: Samples collected into different places in Brazil
Samples with names in orange, green and purple were collect in Cerrado, Atlantic forest and Marine environment, respectively.

4.6 Amplification of eDNA using KS and AD degenerated primers

As previously mentioned, primers A3F and A7R were used for adenylation domains from NRPS and KS2F and KS2R for ketosynthase domains from PKS (Ayuso-Sacido e Genilloud, 2005; Schirmer *et al.*, 2005). NRPS-AD and PKS-KS domains were amplified using degenerate primers designed specifically for each sample, following previous biogeographic comparisons (Owen *et al.*, 2013; Charlop-Powers, Milshteyn*, et al.*, 2014; Charlop-Powers, Owen*, et al.*, 2014; Charlop-Powers *et al.*, 2015). In the beginning, an extensive study was carried out to define the best conditions for the amplification of the biosynthetic domains. For this purpose, three factors were selected: type of protocol; annealing temperature and buffer. Initially, the interference of the sample purification protocol was delimited. To this end, eDNA purification protocols were compared employing the PowerClean Mobio® and MiniPrep Qiagen and purification (Brady, 2007).

The best buffer for amplification process was determined by the comparison of commercial Epicenter buffers: Buffer E and Buffer G, and buffers built in the laboratory. To validate the best annealing temperature, gradient and touchdown PCR was executed. In these experiments the absence of primer dimers was analyzed; nonspecific amplification and formation of smears. Our results conclusively confirmed the use of the MoBio® Power Clean kit was required. The Buffer G developed a surpassing accessibility for the polymerase enzyme to improve the amplification of the eDNA samples. Thus, producing a higher concentration of amplicons with quality (**Figure 27**).

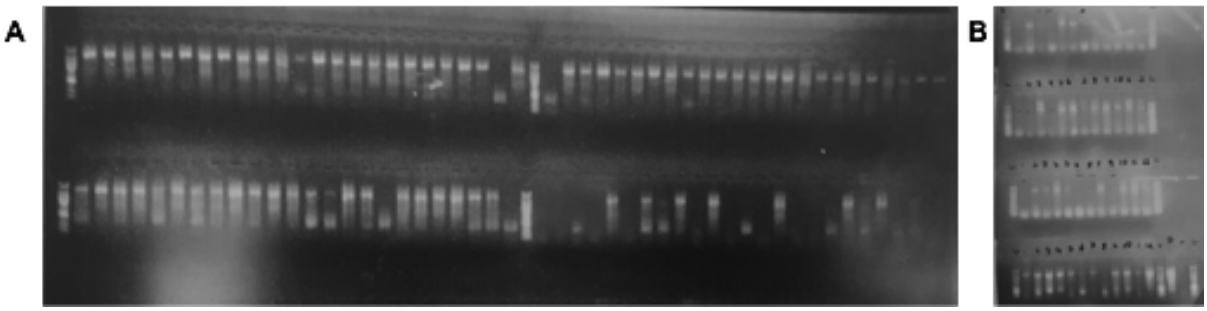


Figure 27: 2% Agarose gel containing AD amplicons (A) and ketosynthase (B)
100 bp Ladder were used as marker. All amplicons were observed with expected size.

For the subsequent sequencing, the NRPS-AD and PKS-KS amplicons were initially normalized. The normalization occurred by balancing the individual concentrations of each amplicon in a pool containing all amplifications (AD/ KS). The AD/KS pool was used as the template for the second round of PCR to complete the construction of the adapter for further sequencing. Following the amplification, magnetic beads purified the amplicons, and the sequencing was carried out through Illumina MiSEQ platform. The V3 sequencing kit was utilized, and the manufacturer protocols were followed.

4.7 Sequencing and bioinformatics analysis

We obtained NRP and KS data to 25 distinct metagenomes. Raw reads were annotated using a python package (https://github.com/esnapd/paired-end-debarcoder). Thus, the reads were obtained without the individual barcodes. The quality control process was conducted through the DADA2 package (Callahan *et al.*, 2016). The low-quality nucleotides were removed by standard error values less than 0.05. Reads forward smaller than 240 bp and reverse reads smaller than 160 bp were dropped (**Figure 28**). The reads were concatenated using degenerate nucleotide "N" between forward and reverse reads, resulting in a fasta file with

filtered sequences and uniform size. Usually, the forward reads displayed good quality while the reverse reads presented nucleotides with a low-quality score at the end of the reads. This is natural in Illumina sequencing methodologies (Kwon *et al.*, 2013).

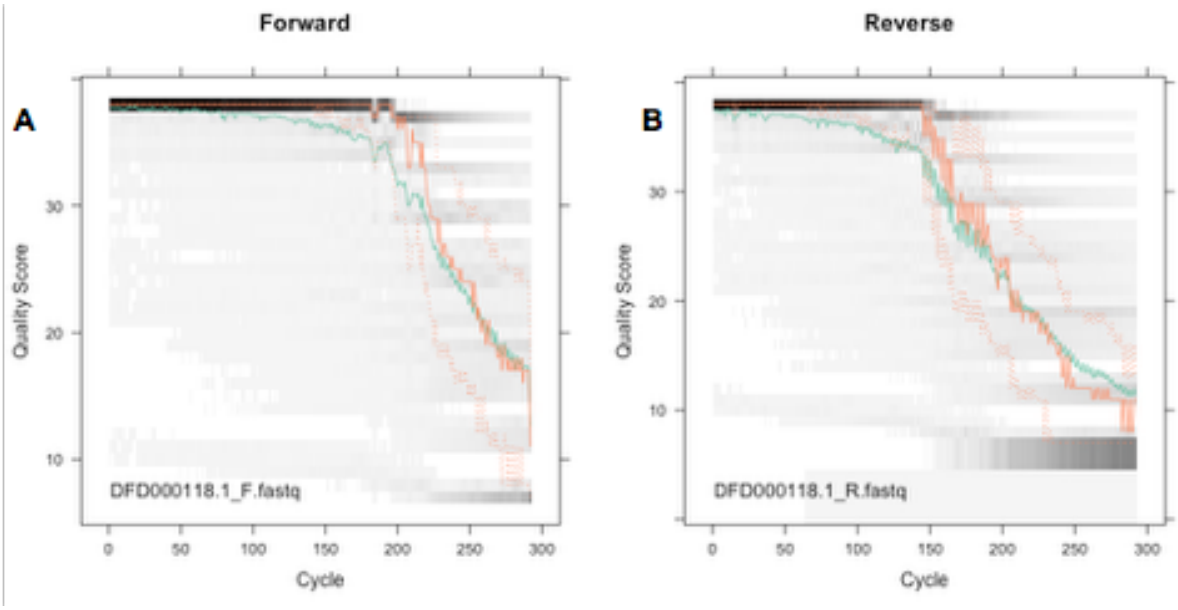


Figure 28: Quality scores of forward and reverse reads
The scores were plot in char containing quality score (y-axis) vs cycles (x-axis). The left char show the Forward reads while right char reverse reads

After the concatenation, the reads were annotated to the respective IDs using a data table, and the phyloseq object was assembled. For this end, the reads were aligned through MUSCLE (Edgar, 2004), and the taxonomic distances were delimited by Neibor-Joining (NJ). With the phyloseq object created, the following analyzes were then led applying the phyloseq package (Mcmurdie, Paul J. e Holmes, Susan, 2013).

4.8 NRP and KS biosynthetic domain richness and environmental factor analysis

Initially, the values of diversity across the samples were measured by the calculation of the alpha diversity. To this end, the observed values were compared

with those computationally calculated. By default, the plot_richness function of PhyloSeq initially allows the prediction of alpha diversity using multiple algorithms: Chao1, Ace, Shannon, Simpson, InvSimpson, and Fisher (**Figure 29**). For our purpose, we concentrate on the Chao1 and Shannon algorithms. Applying these measures, it was possible to predict the total number of taxa (biosynthetic molecules) from all distinct biomes. In this particular case, each taxon corresponded to a NRP target. These models have the advantage of the availability of equations for the calculation of confidence limits. The Chao1 method estimates the total richness using some species (taxon), represented by only one individual in the samples (singletons), and the number of species with only two individuals (doubletons). The Shannon method, on the other hand, rates the uniformity of the abundance of the taxa. Thus, it allows a representation of the rare individuals taxa.

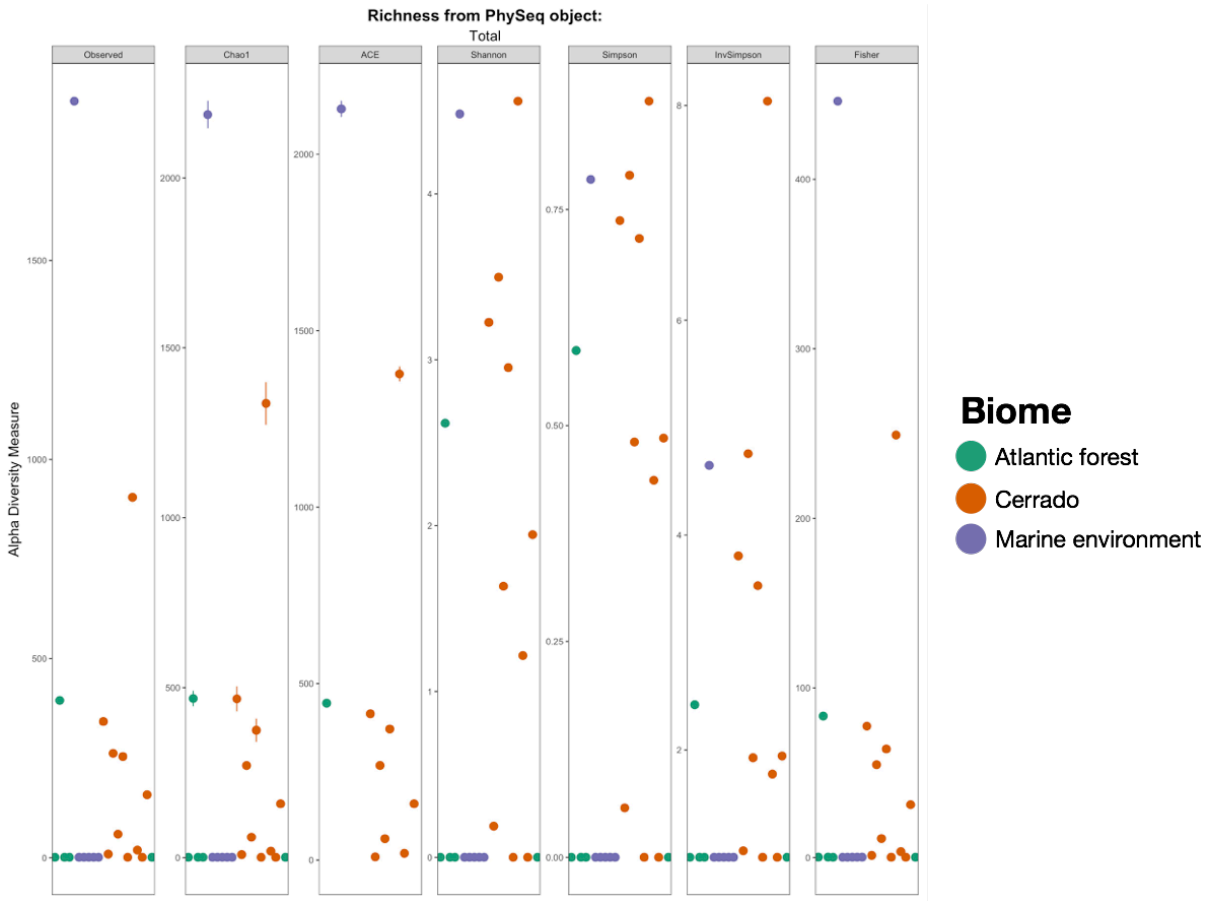


Figure 29: Alpha diversity measurement using Phyloseq. Obs, Chao1, Ace, Shannon, Simpson, InvSimpson, Fisher
In x-axis is described individual samples and y-axis the alpha diversity measure

The alpha-diversity scatter-plot showed no correlation between the taxa in the sequenced samples. The data showed a higher index of alpha diversity for the marine sediment DFD000121 collected in Macaé-RJ. Unexpectedly, this measurement may be induced by the high volume of unknown taxa in this sample. For the following analyses, the taxa unrelated to the biosynthetic clusters were excluded. The Chao1 distance exhibited a more similar profile than observed calculations. The DFD000121 sample showed a higher alpha diversity, accompanied sequentially by the samples DFD000140 and DFD000118, which were obtained from Cerrado and Atlantic forest in the State Park of Vassununga - SP. By Shannon's calculations, it was seen that the samples from the Cerrado tended to have a unique group. We noted some samples with a score of alpha diversity near to zero. Errors in

annotation, filtering process and concatenating the samples may cause this type of analytical disorder (Kwon *et al.*, 2013).

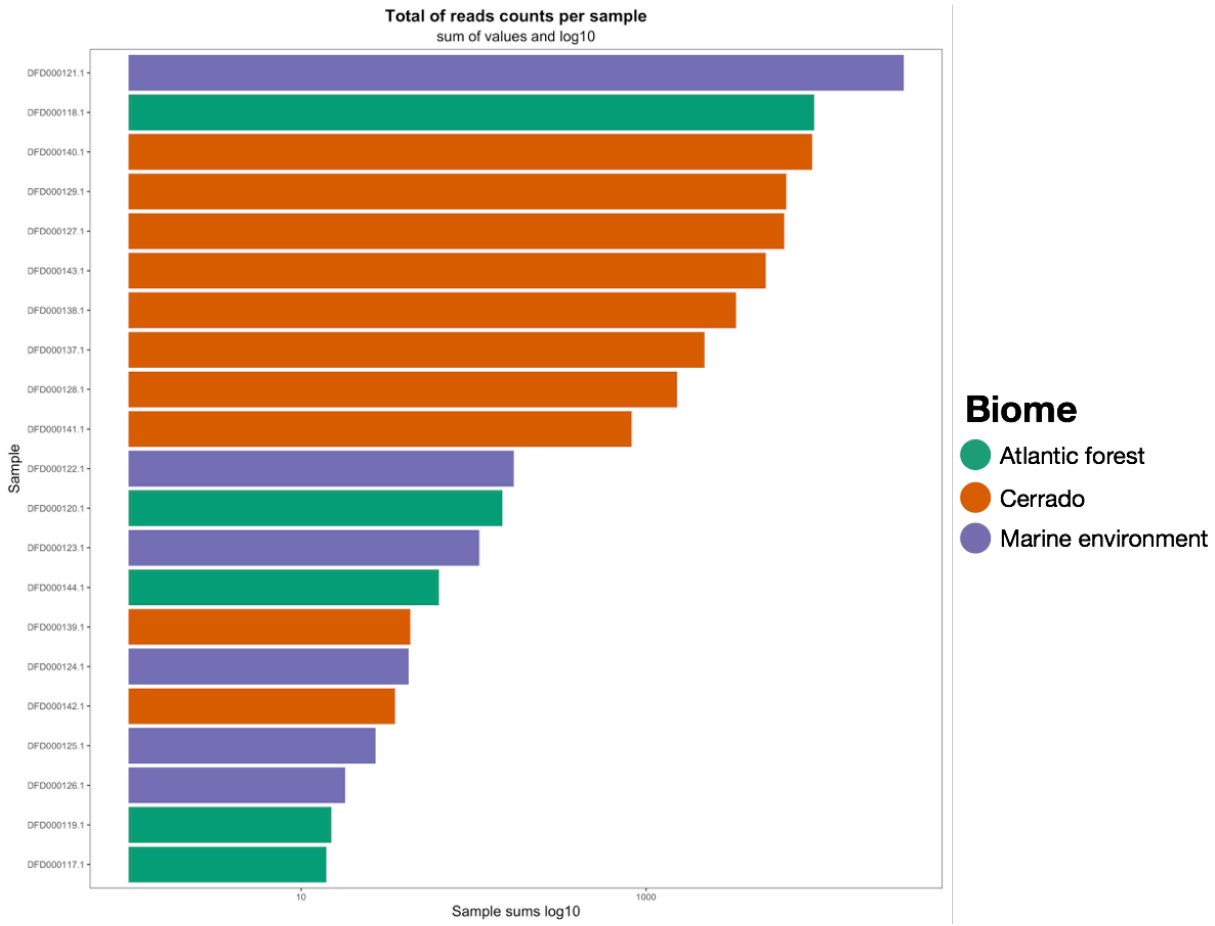


Figure 30: Total of counts per sample
In x-axis the log10 of the sequence sums of individual samples and y-axis sample

As previously mentioned, the sample DFD000121 from the Cananéia marine sediment showed the highest score in the number of sequences. This value does not necessarily expect to be considered in the diversity. The sample DFD000118 from Atlantic Forest, exhibited the second highest score. The samples obtained from cerrado soils DFD000140, DFD000129, DFD000127, DFD000143, DFD000138, DFD000137 and DFD000128 showed sequence scores counts greater than 1000 log10 respectively (**Figure 30**). As mentioned, there is a plenty of sequences

associated to Cerrado. High score of the sequence total associated with DFD000121 sample is noted. Certain values presumably are considered a reflection of the presence of amplicons unrelated to any target.

The similarities in the AD domains families across the samples were investigated by the creation of a non-chimerical annotated table in DADA2. First, rare or poorly atypical sequences were removed. The sample ordering analysis was performed using nonmetric multidimensional scaling (NMDS). The plots were originally plotted according to the biomes and pH values. No apparent clusterization was observed in the NMDS plots (**Figure 31**). However, a trivial clustering of the Cerrado biomes could be observed using the pH values as metrics (**Figure 31**).

Diverse factors could increase the uniqueness of the distinct microbiomes. Examples of factors that change the microbiome were geographic location, elevation, precipitation, mean temperature, and pH. For this end, only pH values were measured. It is recognized that the soil of Cerrado naturally exhibits an acid pH than the other soils, which, we believe, offer a fundamental role for the diversity of macro- and microorganisms between the analyzed biomes.

In a recent paper, Lemetre and Brady showed the occurrence of AD and KS domains in a continental scale (Lemetre *et al.*, 2017). They revealed the difficulty to create an association between the samples. However, when they analyzed longitudinally or latitudinally, spatial correlations may begin to emerge. This can be accomplished according to the dispersion process of specific taxonomic classes in the microbiome.

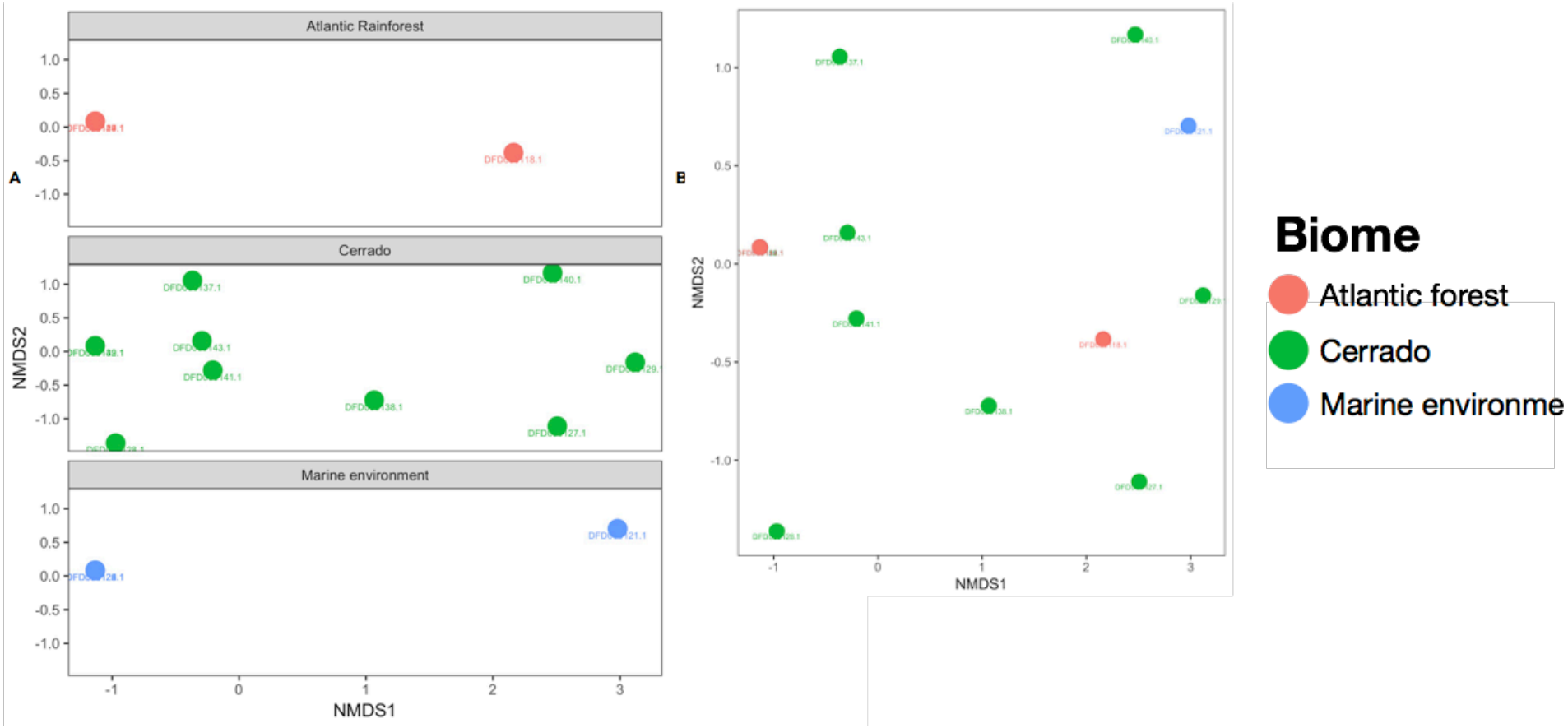


Figure 31: Ordination plot by Biome and pH
NMDS calculations showing the samples correlations individually (left) and together (right). The points were colored by biome type

By the estimation of the Jaccard score between the samples, a type of clusterization was observed based on eco-type instead the biome type. This result means that the biomes did not directly rule the NRP taxon diversity, however, the plant volume of the ecosystem, apparently guide. Samples of the Cerrado with similar ecological types showed correlations, while those forest environments were classified separately. We observed that samples collected at distant geographic areas, were clusterized based on ecological aspects (**Figure 32**).

Thus, the correlation-plot containing the association between the pairing of the samples in Km x calculated Jaccard scores, showed even geographically adjacent samples represent distinguishing unique samples (**Figure 33**). We observed no logical correlation between the points, suggesting that each sample was different. This highlights the taxonomic/microbiological potential of environmental samples. That is a significant result and shows the unusual ability of the microbiome to colonize the environment punctually. In this analysis, we did not include the diversity or abundance. Following, the Jaccard distances were conducted in R through phyloseq, and the results plotted using GGPLOT2.

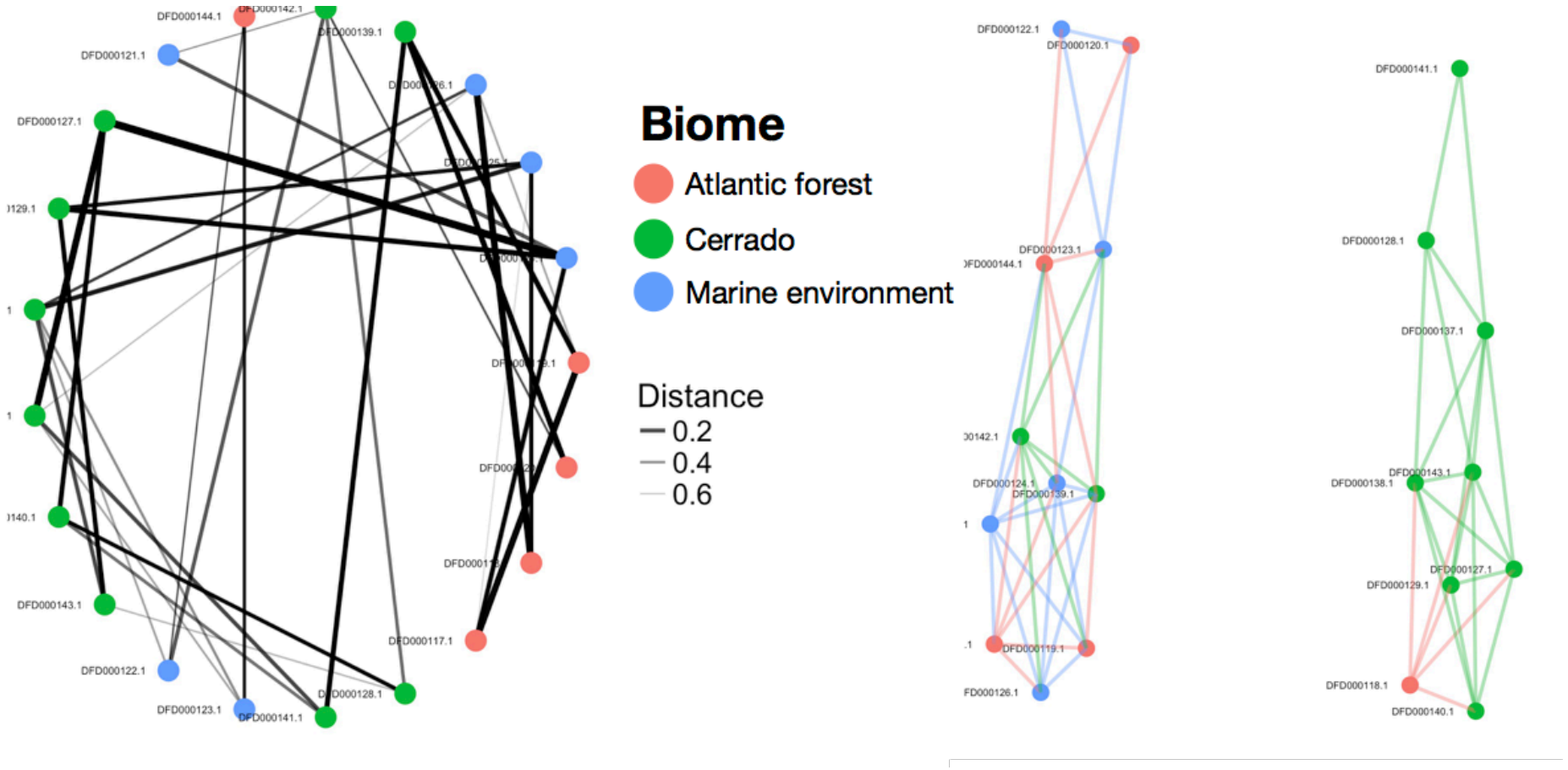


Figure 32: Clusterization of samples using eco-type information
Sample clusters showing statistic distances by biomes (left) and general network (right)

The rarefaction curves associated with these samples were generated by phyloseq. For this, a sub-sampling of the taxa was performed at different depths (1,10,100,500, and 1000-10000 in increments of 500 depths). For each point in the depth, the Chao1 diversity metric was calculated from 10 interactions, and the average of the calculated means in each interaction was plotted.

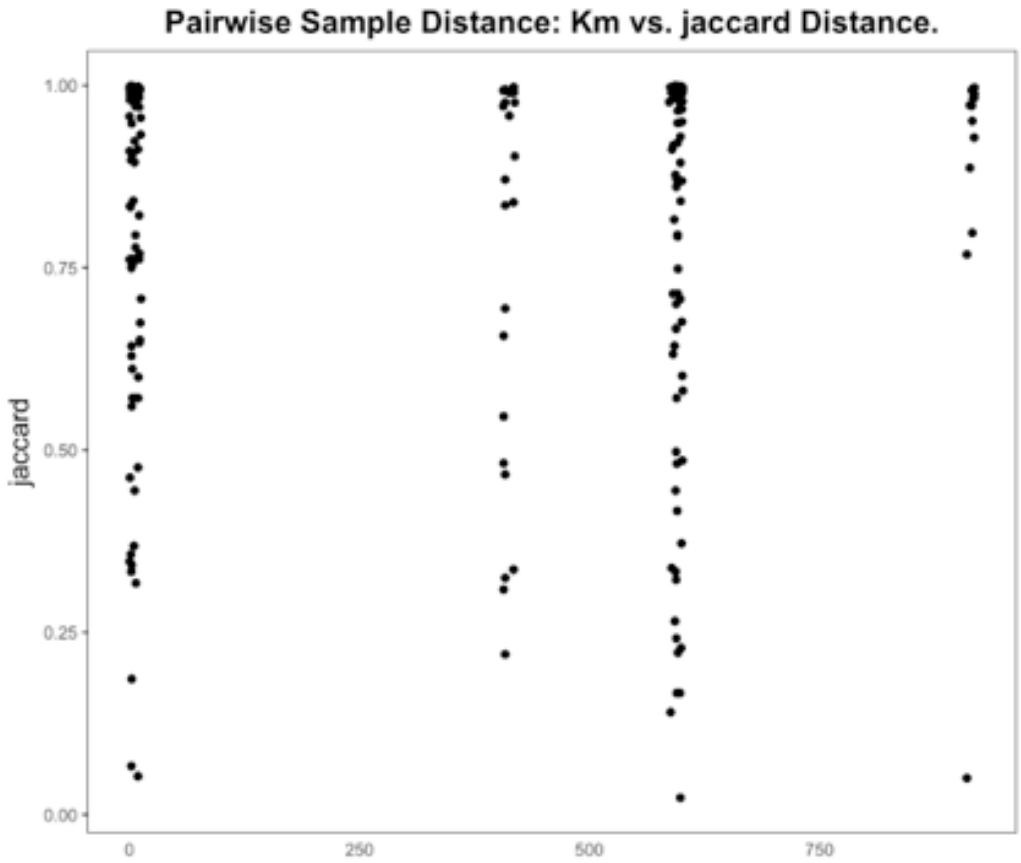


Figure 33: Jaccard distance between brazilian samples
In y-axis is described the jaccard distance between 0 and 1 and x-axis the distance of samples in km

The Chao1 calculations within the rarefaction curves allowed the groups to characterize the diversity of NRP and PK molecules in the samples. In this thesis, we are characterizing as an example the samples DFD000118.1 collected in the Cerrado of Goias and the sample DFD000140.1 collected in the marine sediment of Macaé. Although the DFD000121.1 was the one that presented the largest number of sequences, as previously described, it does not necessarily reflect the chemical

diversity that will be discovered. Samples DFD000118.1 and DFD000121.1 showed the highest diversity among the samples (**Figure 34**).

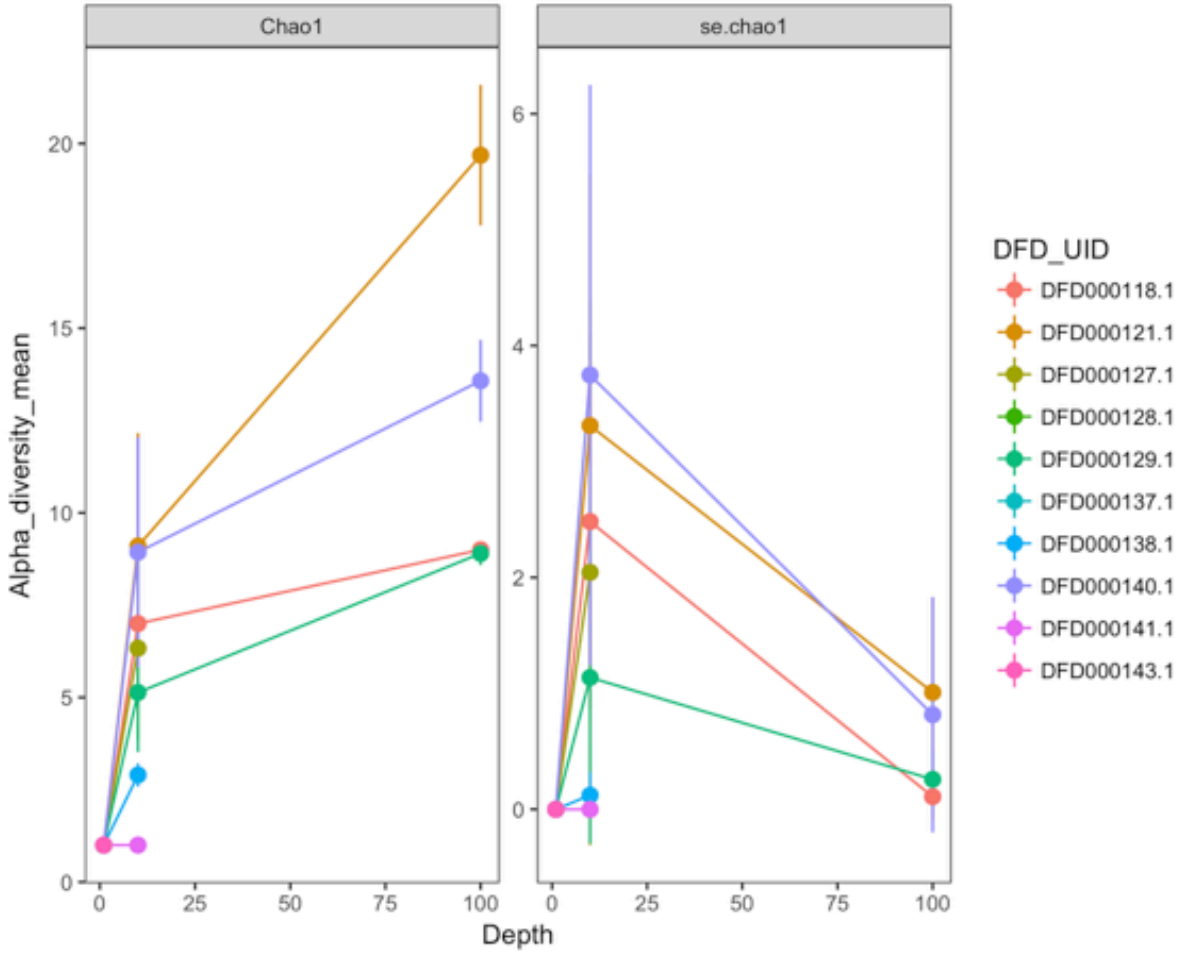


Figure 34: Rarefaction curves calculated by Chao1
Sub-sampling of the taxa was performed at different depths (1,10,100,500, and 1000-10000 in increments of 500 depths). For each point in the depth, the Chao1 diversity metric was calculated from 10 interactions

4.9 Mining clinically relevant gene cluster families in the metabiomes

In the last section we showed the abundance of NRP in the different samples. Now we search in the biomes for the classification of the natural products molecules. Following blast and filtering steps, we recovered a file containing all natural products molecules associated to e-values minor than e-40 and their respectively abundance. Highlighting the top molecule detected in have the calcium-dependent antibiotic in

the sample DFD000118.1 from Atlantic forest. The top hits for all samples could be visualized in the **Table 9**.

Table 9: Natural products molecules annotated in the distinct biomes

Sample	Biome	target	Abundance
DFD000118.1	Atlantic Forest	calcium-dependent	33
DFD000118.1	Atlantic Forest	oxazolomycin	32
DFD000118.1	Atlantic Forest	scabichelin	19
DFD000118.1	Atlantic Forest	didemnin	14
DFD000118.1	Atlantic Forest	auricin	13
DFD000118.1	Atlantic Forest	pyoverdine	12
DFD000118.1	Atlantic Forest	erythrochelin	6
DFD000118.1	Atlantic Forest	gobichelin	6
DFD000118.1	Atlantic Forest	myxoprincomide	5
DFD000127.1	Cerrado	myxoprincomide	32
DFD000127.1	Cerrado	pyoverdine	18
DFD000127.1	Cerrado	calcium-dependent	16
DFD000127.1	Cerrado	didemnin	6
DFD000127.1	Cerrado	mycobactin	6
DFD000127.1	Cerrado	oxazolomycin	4
DFD000127.1	Cerrado	A47934	2
DFD000127.1	Cerrado	actinomycin	2
DFD000127.1	Cerrado	griseobactin	1
DFD000128.1	Cerrado	serobactins	3
DFD000129.1	Cerrado	antimycin	56
DFD000129.1	Cerrado	griseobactin	51
DFD000129.1	Cerrado	oxazolomycin	46
DFD000129.1	Cerrado	didemnin	20
DFD000129.1	Cerrado	gobichelin	12
DFD000129.1	Cerrado	actinomycin	10
DFD000129.1	Cerrado	kirromycin	9
DFD000129.1	Cerrado	laspartomycin	9
DFD000129.1	Cerrado	syringomycin	4
DFD000137.1	Cerrado	polyoxypeptin	7
DFD000138.1	Cerrado	calcium-dependent	26
DFD000138.1	Cerrado	oxazolomycin	15
DFD000138.1	Cerrado	erythrochelin	10
DFD000140.1	Cerrado	polyoxypeptin	40
DFD000140.1	Cerrado	actinomycin	29
DFD000140.1	Cerrado	entolysin	13
DFD000140.1	Cerrado	oxazolomycin	13
DFD000140.1	Cerrado	antimycin	12
DFD000140.1	Cerrado	myxothiazol	12
DFD000140.1	Cerrado	scabichelin	11
DFD000140.1	Cerrado	laspartomycin	7

Sample	Biome	target	Abundance
DFD000140.1	Cerrado	orfamide	6
DFD000140.1	Cerrado	pyridomycin	5
DFD000140.1	Cerrado	WAP-8294A2	5
DFD000140.1	Cerrado	calcium-dependent	3
DFD000140.1	Cerrado	erythrochelin	3
DFD000140.1	Cerrado	gobichelin	3
DFD000141.1	Cerrado	oxazolomycin	12
DFD000143.1	Cerrado	oxazolomycin	4
DFD000121.1	Marine environment	oxazolomycin	196
DFD000121.1	Marine environment	actinomycin	57
DFD000121.1	Marine environment	entolysin	51
DFD000121.1	Marine environment	polyoxypeptin	41
DFD000121.1	Marine environment	myxoprincomide	38
DFD000121.1	Marine environment	taromycin	36
DFD000121.1	Marine environment	erythrochelin	35
DFD000121.1	Marine environment	orfamide	34
DFD000121.1	Marine environment	arthrofactin	33
DFD000121.1	Marine environment	gobichelin	25
DFD000121.1	Marine environment	griseobactin	23
DFD000121.1	Marine environment	microsclerodermins	22
DFD000121.1	Marine environment	didemnin	19
DFD000121.1	Marine environment	skyllamycin	17
DFD000121.1	Marine environment	pyoverdine	15
DFD000121.1	Marine environment	WLIP	12
DFD000121.1	Marine environment	laspartomycin	11
DFD000121.1	Marine environment	quinomycin	7
DFD000121.1	Marine environment	sessilin	7
DFD000121.1	Marine environment	massetolide	6
DFD000121.1	Marine environment	WAP-8294A2	5
DFD000121.1	Marine environment	chloroeremomycin	4

4.10 Tracking biomedically glycopeptide gene cluster and hotspot analysis

To recognize the significance of this diversity, we focused on the detection of clinically relevant secondary metabolites. Initially, the group of vancomycin-like glycopeptides was chosen (Van Wageningen *et al.*, 1998). Vancomycin was initially isolated from cultures of *Amycolatopsis* (actinobacteria). Vancomycin is a non-ribosomal peptide that displays in the biosynthetic cluster 3 genes annotated as non-

ribosomal peptide synthetases. They are vcm2, vcm3, and vcm4. In the total, they are responsible for incorporation of 7 amino acids. Vancomycin biosynthetic gene cluster comprised also 3 enzymes responsible for oxidative couplings. These enzymes were annotated as cytochrome P450 (**Figure 35**). The genes ox1, ox2, and ox3 encode the syntheses of them. The oxidative couplings provide by ox1, ox2 and ox3 increase the stabilization of the molecule, supporting in the antibiotic activity of vancomycin (Zerbe *et al.*, 2002; Pylypenko *et al.*, 2003; Zerbe *et al.*, 2004; Woithe *et al.*, 2007; Geib *et al.*, 2008; Woithe *et al.*, 2008).

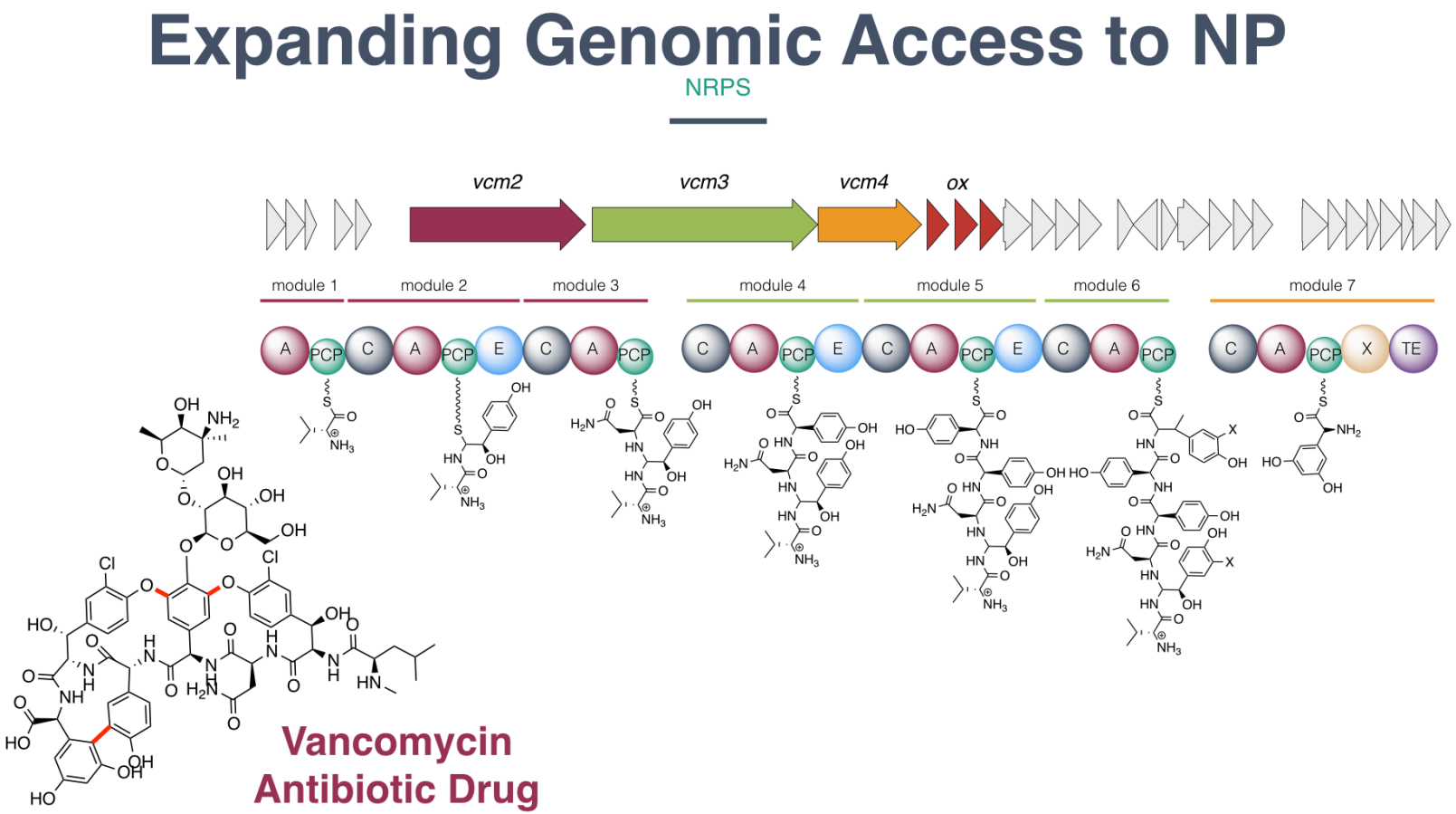


Figure 35: Vancomycin biosynthetic gene cluster
Vancomycin nucleotide sequence cluster was obtained from MiBIG database, analyzed in MacVector and design using ChemDraw

One internal database was created using annotated gene clusters from MiBIG to serve as a reference. The database was created through the makedb function and consisted of more than 1300 biosynthetic clusters related to different groups of secondary metabolites such as PK, NRP, terpenoids, and others.

After blast, a table containing the e-value of each taxon was created. As a confidence margin, sequences with e-values greater than 10e-40 were excluded. The presence of glycopeptides in samples of Cerrado and marine environment was observed. They are annotated as A47934 molecule in the Cerrado from Goias and chloroprenemomycin in the marine sediment from Macaé (**Figure 36**).

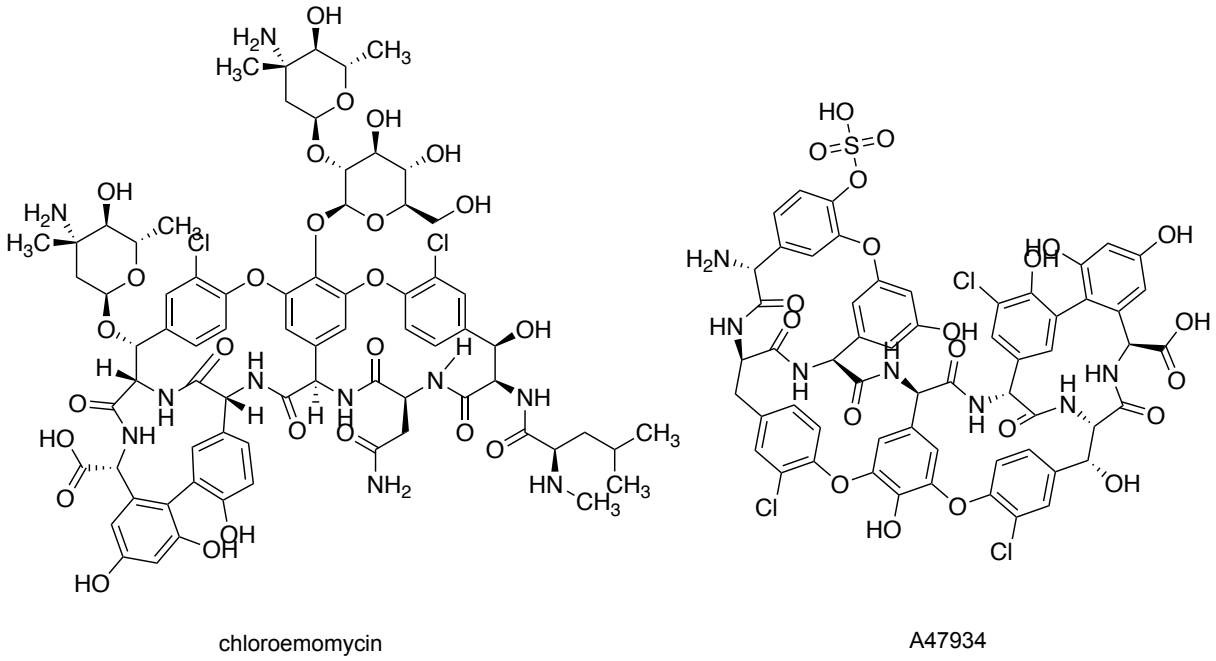


Figure 36: Clinically relevant glycopeptides families detected in some Brazilian samples Glycopeptides were detected using e-value e-40 or minor

The glycopeptide vancomycin and teicoplanin are important antibiotics showing good activity against Gram-positive pathogens including *Streptococcus*, *Staphylococcus*, and *Enterococcus* genera. The glycopeptide antibiotic A47934 is a teicoplanin-like molecule that was isolated from *Streptomyces toyocaensis* NRRL15009.

Glycopeptide antibiotics (GPAs) have been one of the pillars of antimicrobial therapy since their description in the mid-1950s. These molecules act exclusively on Gram-positive bacteria by creating a strong and distinct noncovalent complex with the D-Ala-D-Ala terminus of the peptidoglycan, inhibiting cell wall extension and crosslinking (**Figure 37**) (Barna e Williams, 1984). The first clinical resistance strains to GPAs was reported in the enterococci in 1988, and resistance has shown itself in the more virulent streptococci and staphylococci (Leclercq *et al.*, 1988; Pootoolal *et al.*, 2002). GPA resistance is presently a meaningful phenomenon that has critically impacted on the health care sector increasing the mortality and morbidity (Murray, 2000). The surpassing mechanism of protection is the synthesis of cell wall peptidoglycan terminating in D-Ala-D-lactate, which dramatically reduces the association of these molecules for their target (Murray, 2000).

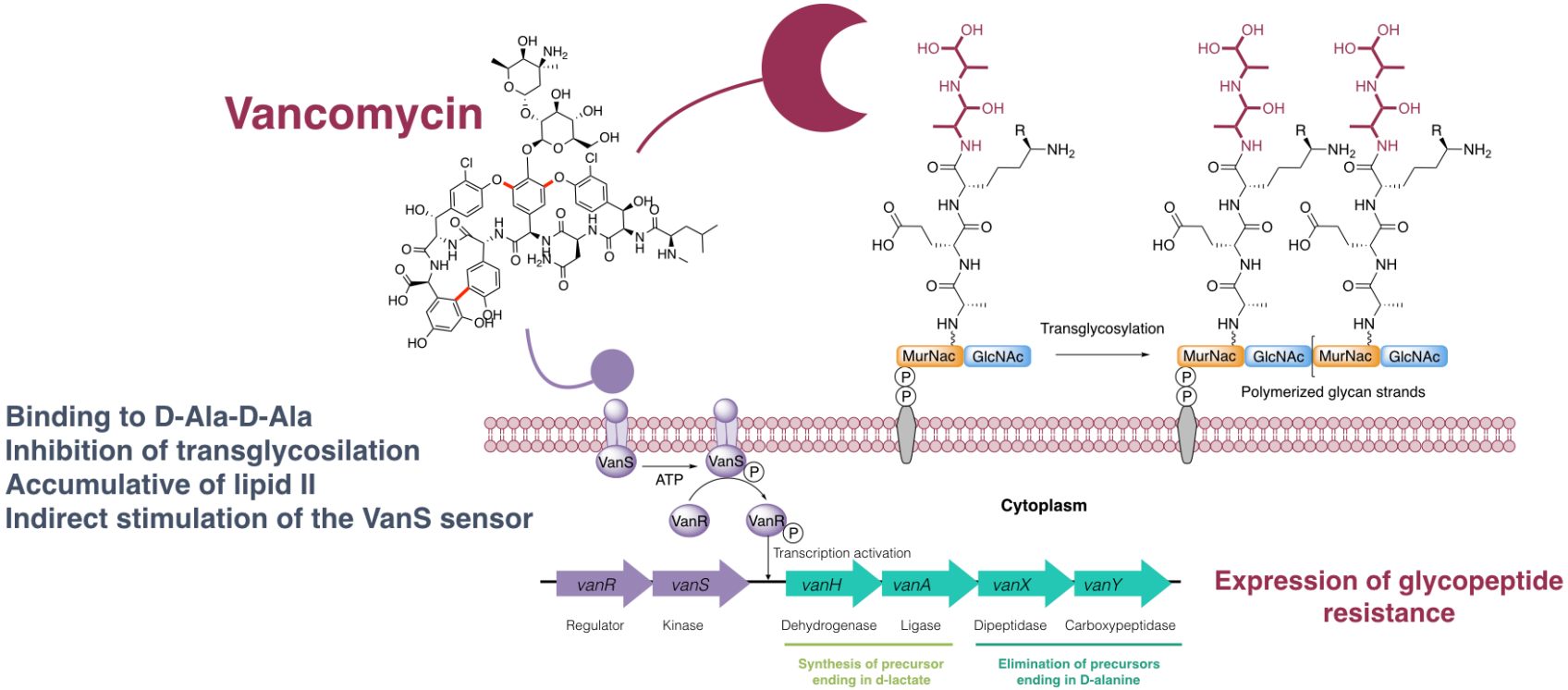


Figure 37: Mechanism of resistance to vancomycin
Vancomycin acts biding D-Ala-D-Ala (top) and resistance mechanism is described binding VanS (bellow)

The GPA is comprised of a heptapeptide core consisting of common and rare amino acids (Barna e Williams, 1984; Boeck e Mertz, 1986; Zmijewski *et al.*, 1987; Hadatsch *et al.*, 2007). Crosslinking of the amino acids in aryl ether and carbon-carbon bonds gives rigidity to the molecule. There are two major structural classes of GPAs based on the identification of the core, and the two clinically employed GPAs, vancomycin, and teicoplanin (Zmijewski *et al.*, 1987).

It is important to highlight these taxa were observed only in a few samples at non-geographically near locations. With these results, it is possible to observe how the samples collected at points may have the potential for the screening and search of new different glycopeptides of the vancomycin class.

CONCLUSIONS

5 Conclusions

The detection of innovative natural products molecules have demanded new approaches, and the metagenomic methods provided a powerful way to access a brand-new environment in the natural products research. Here we have applied two methods to search different molecules in nature. In the first part, we created the SecMetPrimer package to design efficiently degenerated primers for natural products targets. SecMetPrimer designed 165 set of primers for polyketide, nonribosomal peptide, aminoglycosides, phosphono, and tryptophan dimmers families. Besides, the validation of the primers was developed using eDNA as template a metalibrary with 223 millions of clones. For this work, the annotations of AHBA and tryptophan dimmers were developed. The data revealed a new clade of AHBA and tryptophan dimmers in the phylogenetic analysis, showing a new group of those molecules. We also described the presence the ansamycins ansamitocin, ansatrienin, chaxamycin, divergolide, hygrocin, macbecin, napthmycin, rifamycin and rubradirin. In the second section, we have used the biogeography approach to mining the information about biosynthetic gene clusters in Atlantic Forest, Cerrado, and Marine environment samples. The data revealed that no strong correlation of nonribosomal peptides was observed across the samples. This suggested that the microbial biodiversity is much complex than expected. Additionally, the search for clinically relevant nonribosomal peptides revealed the presence of the vancomycin-like glycopeptides in samples from Cerrado and Marine sediment. This is the first study to apply a bioinformatic approach to automatically design degenerated primers capable of amplifying unknown groups of natural products. In the same way, the use of biogeography approach to prioritizing samples for further metagenomic screening is novel, and this

is the first time this type of description was provided. In conclusion, this study demonstrated the creation and use of a bioinformatic pipeline to reach unknown groups of natural products and the biogeography of soil samples collected in different Brazilian biomes.

REFERENCES

6 References

AYUSO-SACIDO, A.; GENILLOUD, O. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. **Microbial ecology,** v. 49, n. 1, p. 10–24, 2005. Disponível em: < https://link.springer.com/content/pdf/10.1007%2Fs00248-004-0249-6.pdf >.

BANIK, J. J.; BRADY, S. F. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. **Proceedings of the National Academy of Sciences of the United States of America,** v. 105, n. 45, p. 17273-17277, Nov 2008. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS: 000260981800021 http://www.pnas.org/content/105/45/17273.full.pdf >.

BANIK, J. J. et al. Tailoring Enzyme-Rich Environmental DNA Clones: A Source of Enzymes for Generating Libraries of Unnatural Natural Products. **Journal of the American Chemical Society,** v. 132, n. 44, p. 15661-15670, Nov 2010. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000283955600048 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3111151/pdf/nihms-245676.pdf >.

BANIN, E.; HUGHES, D.; KUIPERS, O. P. Editorial: Bacterial pathogens, antibiotics and antibiotic resistance. **Fems Microbiology Reviews,** v. 41, n. 3, p. 450-452, May 2017. ISSN 0168-6445. Disponível em: < <Go to ISI>://WOS:000402064900012 https://academic.oup.com/femsre/article-abstract/41/3/450/3806588?redirectedFrom=fulltext >.

BARNA, J. C. J.; WILLIAMS, D. H. The Structure and Mode of Action of Glycopeptide Antibiotics of the Vancomycin Group. **Annual Review of Microbiology,** v. 38, p. 339-357, 1984. ISSN 0066-4227. Disponível em: < <Go to ISI>://WOS:A1984TL37000016 http://www.annualreviews.org/doi/pdf/10.1146/annurev.mi.38.100184.002011 >.

BEJA, O. et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. **Environmental Microbiology,** v. 2, n. 5, p. 516-529, Oct 2000. ISSN 1462-2912. Disponível em: < <Go to ISI>://WOS:000165322900005http://onlinelibrary.wiley.com/store/10.1046/j.1462-2920.2000.00133.x/asset/j.1462-2920.2000.00133.x.pdf?v=1&t=ja88gyrb&s=e44dc96b8cac67c035d88cd9938108c746bffc75 >.

BENTLEY, S. D. et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). **Nature,** v. 417, n. 6885, p. 141-147, May 9 2002. ISSN 0028-0836. Disponível em: < <Go to ISI>://WOS:000175460200034 http://www.nature.com/articles/417141a.pdf >.

BOECK, L. D.; MERTZ, F. P. A47934, a Novel Glycopeptide-Aglycone Antibiotic Produced by a Strain of Streptomyces-Toyocaensis Taxonomy and Fermentation Studies. **Journal of Antibiotics,** v. 39, n. 11, p. 1533-1540, Nov 1986. ISSN 0021-8820. Disponível em: < <Go to ISI>://WOS:A1986E973600004 https://www.jstage.jst.go.jp/article/antibiotics1968/39/11/39_11_1533/_pdf >.

BRADY, S. F. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. **Nature Protocols,** v. 2, n. 5, p. 1297-1305, 2007 2007. ISSN 1754-2189. Disponível em: < <Go to ISI>://WOS:000253138700029 http://www.nature.com/nprot/journal/v2/n5/pdf/nprot.2007.195.pdf >.

BRADY, S. F.; CHAO, C. J.; CLARDY, J. New natural product families from an environmental DNA (eDNA) gene cluster. **Journal of the American Chemical Society,** v. 124, n. 34, p. 9968-9969, Aug 28 2002. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000177576000005 http://pubs.acs.org/doi/pdfplus/10.1021/ja0268985 >.

_____. Long-chain N-acyltyrosine synthases from environmental DNA. **Applied and Environmental Microbiology,** v. 70, n. 11, p. 6865-6870, Nov 2004. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000225076100063 http://aem.asm.org/content/70/11/6865.full.pdf >.

BRADY, S. F. et al. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. **Organic Letters,** v. 3, n. 13, p. 1981-1984, Jun 2001. ISSN 1523-7060. Disponível em: < <Go to ISI>://WOS:000169487700005 http://pubs.acs.org/doi/abs/10.1021/ol015949k http://pubs.acs.org/doi/pdfplus/10.1021/ol015949k >.

BRADY, S. F.; CLARDY, J. CR377, a new pentaketide antifungal agent isolated from an endophytic fungus. **Journal of Natural Products,** v. 63, n. 10, p. 1447-1448, Oct 2000. ISSN 0163-3864. Disponível em: < <Go to ISI>://WOS:000165093900032 http://pubs.acs.org/doi/abs/10.1021/np990568p http://pubs.acs.org/doi/pdfplus/10.1021/np990568p >.

_____. N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in Escherichia coli. **Org Lett,** v. 7, n. 17, p. 3613-6, Aug 18 2005. ISSN 1523-7060 (Print) 1523-7052 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/16092832 http://pubs.acs.org/doi/pdfplus/10.1021/ol0509585 >.

BRADY, S. F. et al. Metagenomic approaches to natural products from free-living and symbiotic organisms. **Nat Prod Rep,** v. 26, n. 11, p. 1488-503, Nov 2009. ISSN 1460-4752 (Electronic) 0265-0568 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/19844642 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2919151/pdf/nihms171048.pdf >.

BREITBART, M. et al. Genomic analysis of uncultured marine viral communities. **Proceedings of the National Academy of Sciences of the United States of America,** v. 99, n. 22, p. 14250-14255, Oct 29 2002. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000178967400053 http://www.pnas.org/content/99/22/14250.full.pdf >.

BRULC, J. M. et al. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. **Proceedings of the National Academy of Sciences of the United States of America,** v. 106, n. 6, p. 1948-1953, Feb 10 2009. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000263252500051 http://www.pnas.org/content/106/6/1948.full.pdf >.

BUTLER, M. S. Natural products to drugs: natural product-derived compounds in clinical trials. **Natural Product Reports,** v. 25, n. 3, p. 475-516, 2008. ISSN 0265-0568. Disponível em: < <Go to ISI>://WOS:000256073800003 http://pubs.rsc.org/en/Content/ArticleLanding/2008/NP/b514294f >.

CALLAHAN, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. **Nature Methods,** v. 13, n. 7, p. 581-+, Jul 2016. ISSN 1548-7091. Disponível em: < <Go to ISI>://WOS:000383794500017 http://www.nature.com/articles/nmeth.3869.pdf >.

CARVAJALI, R. D.; ILSON, D. H.; NOY, A. Possible role of edotecarin, a novel topoisomerase I inhibitor, in therapy-related myelodysplastic syndrome. **Leukemia & Lymphoma,** v. 48, n. 1, p. 192-194, Jan 2007. ISSN 1042-8194. Disponível em: < <Go to ISI>://WOS:000244249900030 >.

CHANG, F.-Y.; BRADY, S. F. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. **Proceedings of the National Academy of Sciences of the United States of America,** v. 110, n. 7, p. 2478-2483, Feb 12 2013. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000315812800027 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574908/pdf/pnas.201218073.pdf >.

CHANG, F. Y.; BRADY, S. F. Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. **J Am Chem Soc,** v. 133, n. 26, p. 9996-9, Jul 6 2011. ISSN 1520-5126 (Electronic) 0002-7863 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21542592 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3126909/pdf/nihms-294136.pdf >.

_____. Characterization of an Environmental DNA-Derived Gene Cluster that Encodes the Bisindolylmaleimide Methylarcyriarubin. **Chembiochem,** v. 15, n. 6, p. 815-821, Apr 2014. ISSN 1439-4227. Disponível em: < <Go to ISI>://WOS:000333966100007 http://onlinelibrary.wiley.com/store/10.1002/cbic.201300756/asset/815_ftp.pdf?v=1&t=j6zfkgtq&s=c2a270e074ad406511d492e460917a9a0d5d2940 >.

CHANG, F. Y. et al. Discovery and Synthetic Refactoring of Tryptophan Dimer Gene Clusters from the Environment. **Journal of the American Chemical Society,** v. 135, n. 47, p. 17906-17912, Nov 2013. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000328099400042 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3878720/pdf/nihms540779.pdf >.

_____. Targeted Metagenomics: Finding Rare Tryptophan Dimer Natural Products in the Environment. **Journal of the American Chemical Society,** v. 137, n. 18, p. 6044-6052, May 2015. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000354910500031 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4839266/pdf/nihms727630.pdf >.

CHARLOP-POWERS, Z. et al. Selective Enrichment of Environmental DNA Libraries for Genes Encoding Nonribosomal Peptides and Polyketides by Phosphopantetheine Transferase-Dependent Complementation of Siderophore Biosynthesis. **Acs Chemical Biology,** v. 8, n. 1, p. 138-143, Jan 2013. ISSN 1554-8929. Disponível em: < <Go to ISI>://WOS:000314008000016 http://pubs.acs.org/doi/pdfplus/10.1021/cb3004918 >.

CHARLOP-POWERS, Z.; MILSHTEYN, A.; BRADY, S. F. Metagenomic small molecule discovery methods. **Current Opinion in Microbiology,** v. 19, p. 70-75, Jun 2014. ISSN 1369-5274. Disponível em: < <Go to ISI>://WOS:000340984200012 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4135586/pdf/nihms-606078.pdf >.

CHARLOP-POWERS, Z. et al. Global biogeographic sampling of bacterial secondary metabolism. **Elife,** v. 4, p. e05048, Jan 19 2015. ISSN 2050-084X (Electronic)2050-084X (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/25599565 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383359/pdf/elife05048.pdf >.

CHARLOP-POWERS, Z. et al. Chemical-biogeographic survey of secondary metabolism in soil. **Proceedings of the National Academy of Sciences of the United States of America,** v. 111, n. 10, p. 3757-3762, Mar 11 2014. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000332564800036 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3956145/pdf/pnas.201318021.pdf >.

CORDELL, G. A.; QUINN-BEATTIE, M. L.; FARNSWORTH, N. R. The potential of alkaloids in drug discovery. **Phytotherapy Research,** v. 15, n. 3, p. 183-205, May 2001. ISSN 0951-418x. Disponível em: < <Go to ISI>://WOS:000168713900001 http://onlinelibrary.wiley.com/store/10.1002/ptr.890/asset/890_ftp.pdf?v=1&t=ja88mx4j&s=a9d69b8cfb2d105c3bbbdd6560891410e0d9dac2 >.

CORSON, T. W.; CREWS, C. M. Molecular understanding and modern application of traditional medicines: Triumphs and trials. **Cell,** v. 130, n. 5, p. 769-774, Sep 7 2007. ISSN 0092-8674. Disponível em: < <Go to ISI>://WOS:000249581500007 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2507744/pdf/nihms56735.pdf >.

COURTOIS, S. et al. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. **Applied and Environmental Microbiology,** v. 69, n. 1, p. 49-55, Jan 2003. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000180328000007 http://aem.asm.org/content/69/1/49.full.pdf >.

COURTOIS, S. et al. Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. **Environmental Microbiology,** v. 3, n. 7, p. 431-439, Jul 2001. ISSN 1462-2912. Disponível em: < <Go to ISI>://WOS:000171022900002 http://onlinelibrary.wiley.com/store/10.1046/j.1462-2920.2001.00208.x/asset/j.1462-2920.2001.00208.x.pdf?v=1&t=ja88nywm&s=f8f09aab17a6a109ea2109844f6161b54458b3d6 >.

CRAGG, G.; NEWMAN, D. Natural products and drug discovery and development: A history of success and continuing promise for the future. **Planta Medica,** v. 80, n. 10, p. 750-750, Jul 2014. ISSN 0032-0943. Disponível em: < <Go to ISI>://WOS:000339781200003 >.

CRAGG, G. M.; GROTHAUS, P. G.; NEWMAN, D. J. Impact of Natural Products on Developing New Anti-Cancer Agents. **Chemical Reviews,** v. 109, n. 7, p. 3012-3043, Jul 2009. ISSN 0009-2665. Disponível em: < <Go to ISI>://WOS:000268090000010 >.

CRAGG, G. M. et al. The impact of the United Nations Convention on Biological Diversity on natural products research. **Natural Product Reports,** v. 29, n. 12, p. 1407-1423, 2012. ISSN 0265-0568. Disponível em: < <Go to ISI>://WOS:000310740700002 >.

CRAGG, G. M.; NEWMAN, D. J. Natural product drug discovery in the next millennium. **Pharmaceutical Biology,** v. 39, p. 8-17, 2001. ISSN 1388-0209. Disponível em: < <Go to ISI>://WOS:000175141800003 >.

CRAIG, J. W.; CHANG, F. Y.; BRADY, S. F. Natural Products from Environmental DNA Hosted in Ralstonia metallidurans. **Acs Chemical Biology,** v. 4, n. 1, p. 23-28, Jan 2009. ISSN 1554-8929. Disponível em: < <Go to ISI>://WOS:000262566300004 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2741005/pdf/nihms-91324.pdf >.

CRAIG, J. W. et al. Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria. **Applied and Environmental Microbiology,** v. 76, n. 5, p. 1633-1641, Mar 2010. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000274855800038 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2832356/pdf/2169-09.pdf >.

DANIEL, R. The metagenomics of soil. **Nature Reviews Microbiology,** v. 3, n. 6, p. 470-478, Jun 2005. ISSN 1740-1526. Disponível em: < <Go to ISI>://WOS:000229435100012 http://www.nature.com/articles/nrmicro1160.pdf >.

DAVIDSON, S. K. et al. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "Candidatus Endobugula sertula" of the bryozoan Bugula neritina. **Applied and Environmental Microbiology,** v. 67, n. 10, p. 4531-4537, Oct 2001. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000171237700018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC93199/pdf/am1001004531.pdf >.

DAVIDSON, S. K.; HAYGOOD, M. G. Identification of sibling species of the bryozoan Bugula neritina that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont "Candidatus endobugula sertula". **Biological Bulletin,** v. 196, n. 3, p. 273-280, Jun 1999. ISSN 0006-3185. Disponível em: < <Go to ISI>://WOS:000081104800006 >.

DESANTIS, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. **Applied and Environmental Microbiology,** v. 72, n. 7, p. 5069-5072, Jul 2006. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000238961000071 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489311/pdf/3006-05.pdf >.

DONIA, M. S. et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. **Nature Chemical Biology,** v. 2, n. 12, p. 729-735, Dec 2006. ISSN 1552-4450. Disponível em: < <Go to ISI>://WOS:000242168800018 http://www.nature.com/articles/nchembio829.pdf >.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research,** v. 32, n. 5, p. 1792-1797, Mar 2004. ISSN 0305-1048. Disponível em: < <Go to ISI>://WOS:000220487200025 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/pdf/gkh340.pdf >.

_____. Search and clustering orders of magnitude faster than BLAST. **Bioinformatics,** v. 26, n. 19, p. 2460-2461, Oct 2010. ISSN 1367-4803. Disponível em: < <Go to ISI>://WOS:000282170000016

https://watermark.silverchair.com/btq461.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAcwwggHIBgkqhkiG9w0BBwagggG5MIIBtQIBADCCAa4GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMF569icXK588sXza-AgEQgIIBf5nRkK9vQlReGIPL-DJez-kqRdXkD5Y146ZuM5gekTfdImFY666tCFgo-wezzz91dONDaWmMO5iQLt70BdCwTbHsrqyh70Pwxzp0D5skuFWq_eoHRGy_cLjxbWdc0sHqOB7EJ2_s0PKGgUKqrgsjLVlruLiO28rGlELC4qf6p-JJtMFbaSOHmzkRXCnMzvG2_DP8x44QFt-Wu4fG2kpiUVPMomh7q8Kheob3_haifhSNrCyXGu5AyZj3nOFLoxt_ZFeSxhxyT7lJRC4aBtR0mW-bg0xihAIUioIEIL0A9u4kJ5D6R >.

_____. UPARSE: highly accurate OTU sequences from microbial amplicon reads. **Nature Methods,** v. 10, n. 10, p. 996-+, Oct 2013. ISSN 1548-7091. Disponível em: < <Go to ISI>://WOS:000325073800023 http://www.nature.com/articles/nmeth.2604.pdf >.

FADROSH, D. W. et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. **Microbiome,** v. 2, Feb 24 2014. ISSN 2049-2618. Disponível em: < <Go to ISI>://WOS:000363189500001 https://microbiomejournal.biomedcentral.com/track/pdf/10.1186/2049-2618-2-6?site=microbiomejournal.biomedcentral.com >.

FENG, Z. Y. et al. Environmental DNA-Encoded Antibiotics Fasamycins A and B Inhibit FabF in Type II Fatty Acid Biosynthesis. **Journal of the American Chemical Society,** v. 134, n. 6, p. 2981-2987, Feb 2012. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000301161500035 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3335777/pdf/nihms354784.pdf >.

FENG, Z. Y.; KALLIFIDAS, D.; BRADY, S. F. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. **Proceedings of the National Academy of Sciences of the United States of America,** v. 108, n. 31, p. 12629-12634, Aug 2011. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000293385700025 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3150919/pdf/pnas.1103921108.pdf >.

FENG, Z. Y.; KIM, J. H.; BRADY, S. F. Fluostatins Produced by the Heterologous Expression of a TAR Reassembled Environmental DNA Derived Type II PKS Gene Cluster. **Journal of the American Chemical Society,** v. 132, n. 34, p. 11902-11903, Sep 2010. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000281296700030 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2930618/pdf/nihms227557.pdf >.

FOX, J. L. The business of developing antibacterials. **Nature Biotechnology,** v. 24, n. 12, p. 1521-1528, Dec 2006. ISSN 1087-0156. Disponível em: < <Go to ISI>://WOS:000242795800029 http://www.nature.com/articles/nbt1206-1521.pdf >.

GABOR, E. M.; ALKEMA, W. B. L.; JANSSEN, D. B. Quantifying the accessibility of the metagenome by random expression cloning techniques. **Environmental Microbiology,** v. 6, n. 9, p. 879-886, Sep 2004. ISSN 1462-2912. Disponível em: < <Go to ISI>://WOS:000223254700002 http://onlinelibrary.wiley.com/store/10.1111/j.1462-2920.2004.00640.x/asset/j.1462-2920.2004.00640.x.pdf?v=1&t=ja88pfsw&s=9108ffd5cc7c9cf4faa9bb1dde656190b867452f >.

GANS, J.; WOLINSKY, M.; DUNBAR, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. **Science,** v. 309, n. 5739, p. 1387-1390, Aug 26 2005. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000231543300048 http://science.sciencemag.org/content/309/5739/1387.long >.

GARCIA, J. A. L.; FERNANDEZ-GUERRA, A.; CASAMAYOR, E. O. A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes. **Molecular Phylogenetics and Evolution,** v. 61, n. 3, p. 650-658, Dec 2011. ISSN 1055-7903. Disponível em: < <Go to ISI>://WOS:000297387600005 >.

GEIB, N. et al. New insights into the first oxidative phenol coupling reaction during vancomycin biosynthesis. **Bioorganic & Medicinal Chemistry Letters,** v. 18, n. 10, p. 3081-3084, May 15 2008. ISSN 0960-894x. Disponível em: < <Go to ISI>://WOS:000255954000015 >.

GENTRY, T. J. et al. Microarray applications in microbial ecology research. **Microbial Ecology,** v. 52, n. 2, p. 159-175, Aug 2006. ISSN 0095-3628. Disponível em: < <Go to ISI>://WOS:000240481000001 https://link.springer.com/content/pdf/10.1007%2Fs00248-006-9072-6.pdf >.

GILLESPIE, D. E. et al. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. **Applied and Environmental Microbiology,** v. 68, n. 9, p. 4301-4306, Sep 2002. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000177718000018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC124076/pdf/0234.pdf >.

GINOLHAC, A. et al. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. **Applied and Environmental Microbiology,** v. 70, n. 9, p. 5522-5527, Sep 2004. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000223901100059 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC520897/pdf/0965-04.pdf >.

GUAN, C. H. et al. Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. **Applied and Environmental Microbiology,** v. 73, n. 11, p. 3669-3676, Jun 2007. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000247016600026 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1932686/pdf/2617-06.pdf >.

HADATSCH, B. et al. The biosynthesis of teicoplanin-type glycopeptide antibiotics: Assignment of p450 mono-oxygenases to side chain Cyclizations of glycopeptide a47934. **Chemistry & Biology,** v. 14, n. 9, p. 1078-1089, Sep 2007. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000249913600014 >.

HAMADA, T. et al. Solution Structure of Polytheonamide B, a Highly Cytotoxic Nonribosomal Polypeptide from Marine Sponge. **Journal of the American Chemical Society,** v. 132, n. 37, p. 12941-12945, Sep 22 2010. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000282013700045 http://pubs.acs.org/doi/pdfplus/10.1021/ja104616z >.

HANDELSMAN, J. et al. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. **Chemistry & Biology,** v. 5, n. 10, p. R245-R249, Oct 1998. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000076474700001 >.

HERTWECK, C. The Biosynthetic Logic of Polyketide Diversity. **Angewandte Chemie-International Edition,** v. 48, n. 26, p. 4688-4716, 2009 2009. ISSN 1433-7851. Disponível em: < <Go to ISI>://WOS:000267494500004 http://onlinelibrary.wiley.com/store/10.1002/anie.200806121/asset/4688_ftp.pdf?v=1&t=j6zbpimn&s=eed637e2c165384fce1d3adb4225c0534866c47f >.

HILDEBRAND, M. et al. bryA: An unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan Bugula neritina. **Chemistry & Biology,** v. 11, n. 11, p. 1543-1552, Nov 2004. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000225613800013 >.

HORI, M. et al. MYCALOLIDE-B, A NOVEL AND SPECIFIC INHIBITOR OF ACTOMYOSIN ATPASE ISOLATED FROM MARINE SPONGE. **Febs Letters,** v. 322, n. 2, p. 151-154, May 10 1993. ISSN 0014-5793. Disponível em: < <Go to ISI>://WOS:A1993LB10300014 http://onlinelibrary.wiley.com/store/10.1016/0014-5793(93)81557-G/asset/feb2001457939381557g.pdf?v=1&t=ja88qfse&s=5aa37c4cdbbad9f079b5f62f70678de57b0c5d61 >.

HOSZTAFI, S. The discovery of alkaloids. **Pharmazie,** v. 52, n. 7, p. 546-550, Jul 1997. ISSN 0031-7144. Disponível em: < <Go to ISI>://WOS:A1997XN60800016 >.

HUGERTH, L. W. et al. DegePrime, a Program for Degenerate Primer Design for Broad-Taxonomic-Range PCR in Microbial Ecology Studies. **Applied and Environmental Microbiology,** v. 80, n. 16, p. 5116-5123, Aug 2014. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000340038400036 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4135748/pdf/zam5116.pdf >.

JI, H.-F.; LI, X.-J.; ZHANG, H.-Y. Natural products and drug discovery Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? **Embo Reports,** v. 10, n. 3, p. 194-200, Mar 2009. ISSN 1469-221X. Disponível em: < <Go to ISI>://WOS:000263893900002 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658564/pdf/embor200912.pdf >.

KAEBERLEIN, T.; LEWIS, K.; EPSTEIN, S. S. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. **Science,** v. 296, n. 5570, p. 1127-1129, May 10 2002. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000175565000056 http://science.sciencemag.org/content/296/5570/1127.long >.

KAISER, H. From the plant to chemistry - the early history of "rheumatic medication". **Zeitschrift Fur Rheumatologie,** v. 67, n. 3, p. 252-262, May 2008. ISSN 0340-1855. Disponível em: < <Go to ISI>://WOS:000255629700012 https://link.springer.com/content/pdf/10.1007%2Fs00393-008-0257-x.pdf >.

KALLIFIDAS, D.; KANG, H. S.; BRADY, S. F. Tetarimycin A, an MRSA-Active Antibiotic Identified through Induced Expression of Environmental DNA Gene Clusters. **Journal of the American Chemical Society,** v. 134, n. 48, p. 19552-19555, Dec 2012. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000311869600014 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540986/pdf/nihms424875.pdf >.

KANG, H. S.; BRADY, S. F. Arimetamycin A: Improving Clinically Relevant Families of Natural Products through Sequence-Guided Screening of Soil Metagenomes. **Angewandte Chemie-International Edition,** v. 52, n. 42, p. 11063-11067, Oct 2013. ISSN 1433-7851. Disponível em: < <Go to ISI>://WOS:000328812600023 http://onlinelibrary.wiley.com/store/10.1002/anie.201305109/asset/11063_ftp.pdf?v=1&t=j6zfl2ei&s=4e349b4069adbf6e10b8c6df849adf6ae61fb1e3 >.

_____. Arixanthomycins A-C: Phylogeny-Guided Discovery of Biologically Active eDNA-Derived Pentangular Polyphenols. **Acs Chemical Biology,** v. 9, n. 6, p. 1267-1272, Jun 2014. ISSN 1554-8929. Disponível em: < <Go to ISI>://WOS:000337870500008 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4076013/pdf/cb500141b.pdf >.

KATO, Y. et al. THE BIOACTIVE MARINE METABOLITES .16. CALYCULIN-A, A NOVEL ANTITUMOR METABOLITE FROM THE MARINE SPONGE DISCODERMIA-CALYX. **Journal of the American Chemical Society,** v. 108, n. 10, p. 2780-2781, May 14 1986. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:A1986C377800061 http://pubs.acs.org/doi/pdfplus/10.1021/ja00270a061 >.

KATZ, M.; HOVER, B. M.; BRADY, S. F. Culture-independent discovery of natural products from soil metagenomes. **Journal of Industrial Microbiology & Biotechnology,** v. 43, n. 2-3, p. 129-141, Mar 2016. ISSN 1367-5435. Disponível em: < <Go to ISI>://WOS:000372538000004 https://link.springer.com/content/pdf/10.1007%2Fs10295-015-1706-6.pdf >.

KELLER, M.; ZENGLER, K. Tapping into microbial diversity. **Nature Reviews Microbiology,** v. 2, n. 2, p. 141-150, Feb 2004. ISSN 1740-1526. Disponível em: < <Go to ISI>://WOS:000220431700016 http://www.nature.com/articles/nrmicro819.pdf >.

KING, R. W.; BAUER, J. D.; BRADY, S. F. An Environmental DNA-Derived Type II Polyketide Biosynthetic Pathway Encodes the Biosynthesis of the Pentacyclic Polyketide Erdacin. **Angewandte Chemie-International Edition,** v. 48, n. 34, p. 6257-6261, 2009 2009. ISSN 1433-7851. Disponível em: < <Go to ISI>://WOS:000269089900011 http://onlinelibrary.wiley.com/store/10.1002/anie.200901209/asset/6257_ftp.pdf?v=1&t=j6za2owl&s=b83e89bf653f83a01a6f09a19061c78cd50a39d3 >.

KWON, S. et al. In-Depth Analysis of Interrelation between Quality Scores and Real Errors in Illumina Reads. **2013 35th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc)**, p. 635-638, 2013. ISSN 1557-170x. Disponível em: < <Go to ISI>://WOS:000341702101031 >.

LAURETI, L. et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. **Proceedings of the National Academy of Sciences of the United States of America,** v. 108, n. 15, p. 6258-6263, Apr 12 2011. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000289413600065 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3076887/pdf/pnas.201019077.pdf >.

LECLERCQ, R. et al. Plasmid-Mediated Resistance to Vancomycin and Teicoplanin in Enterococcus-Faecium. **New England Journal of Medicine,** v. 319, n. 3, p. 157-161, Jul 21 1988. ISSN 0028-4793. Disponível em: < <Go to ISI>://WOS:A1988P282300007 >.

LEMETRE, C. et al. 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 of the United States of America,** v. 114, n. 44, p. 11615-11620, Oct 1 2017. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000414127400046 http://www.pnas.org/content/114/44/11615.full.pdf >.

LI, H. **seqtk: A fast and lightweight tool for processing sequences**. Broad Institute., Cambridge, MA 2016.

LI, J. W. H.; VEDERAS, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? **Science,** v. 325, n. 5937, p. 161-165, Jul 10 2009. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000267802000035 >.

LI, Y. G. et al. Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of Rhizobium leguminosarum and of Escherichia coli reveals different classes of cloned trp genes. **Environmental Microbiology,** v. 7, n. 12, p. 1927-1936, Dec 2005. ISSN 1462-2912. Disponível em: < <Go to ISI>://WOS:000233313400008 http://onlinelibrary.wiley.com/store/10.1111/j.1462-2920.2005.00853.x/asset/j.1462-2920.2005.00853.x.pdf?v=1&t=ja88rnqc&s=dcef032424d63f72266634b70554b74cef542155 >.

LIM, H. K. et al. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on Escherichia coli. **Applied and Environmental Microbiology,** v. 71, n. 12, p. 7768-7777, Dec 2005. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000234417600015 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1317350/pdf/0827-05.pdf >.

LUO, Y. et al. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. **Nature Communications,** v. 4, Dec 2013. ISSN 2041-1723. Disponível em: < <Go to ISI>://WOS:000329396200012 http://www.nature.com/articles/ncomms3894.pdf >.

MACNEIL, I. A. et al. Expression and isolation of antimicrobial small molecules from soil DNA libraries. **Journal of Molecular Microbiology and Biotechnology,** v. 3, n. 2, p. 301-308, Apr 2001. ISSN 1464-1801. Disponível em: < <Go to ISI>://WOS:000167792000023 >.

MAPLESTONE, R. A.; STONE, M. J.; WILLIAMS, D. H. The Evolutionary Role of Secondary Metabolites - a Review. **Gene,** v. 115, n. 1-2, p. 151-157, Jun 15 1992. ISSN 0378-1119. Disponível em: < <Go to ISI>://WOS:A1992JB68100022 >.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature,** v. 437, n. 7057, p. 376-380, Sep 15 2005. ISSN 0028-0836. Disponível em: < <Go to ISI>://WOS:000231849100045 http://www.nature.com/articles/nature03959.pdf >.

MARTINEZ, A. et al. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. **Applied and Environmental Microbiology,** v. 70, n. 4, p. 2452-2463, Apr 2004. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000220792200068 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC383137/pdf/1697.pdf >.

MATZANKE, B. F.; MULLER, G. I.; RAYMOND, K. N. COORDINATION CHEMISTRY OF MICROBIAL IRON TRANSPORT COMPOUNDS .28. HYDROXAMATE SIDEROPHORE MEDIATED IRON UPTAKE IN ESCHERICHIA-COLI - STEREOSPECIFIC RECOGNITION OF FERRIC RHODOTORULIC ACID. **Biochemical and Biophysical Research Communications,** v. 121, n. 3, p. 922-930, 1984 1984. ISSN 0006-291X. Disponível em: < <Go to ISI>://WOS:A1984SY02000025 >.

MCMURDIE, P. J.; HOLMES, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **Plos One,** v. 8, n. 4, Apr 22 2013. ISSN 1932-6203. Disponível em: < <Go to ISI>://WOS:000317911500023 http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0061217&type=printable >.

MCMURDIE, P. J.; HOLMES, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **PLoS ONE,** v. 8, n. 4 %U http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3632530/, 2013.

MEDEMA, M. H. et al. Minimum Information about a Biosynthetic Gene cluster. **Nat Chem Biol,** v. 11, n. 9, p. 625-31, Sep 2015. ISSN 1552-4469 (Electronic) 1552-4450 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/26284661 http://www.nature.com/nchembio/journal/v11/n9/pdf/nchembio.1890.pdf >.

METSA-KETELA, M. et al. 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,** v. 68, n. 9, p. 4472-4479, Sep 2002. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000177718000041 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC124067/pdf/0143.pdf >.

MILSHTEYN, A.; SCHNEIDER, J. S.; BRADY, S. F. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. **Chemistry & Biology,** v. 21, n. 9, p. 1211-1223, Sep 2014. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000342626700016 >.

MONTIEL, D. et al. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. **Proceedings of the National Academy of Sciences of the United States of America,** v. 112, n. 29, p. 8953-8958, Jul 21 2015. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000358225100057 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4517240/pdf/pnas.201507606.pdf >.

MURRAY, B. E. Drug therapy: Vancomycin-resistant enterococcal infections. **New England Journal of Medicine,** v. 342, n. 10, p. 710-721, Mar 9 2000. ISSN 0028-4793. Disponível em: < <Go to ISI>://WOS:000085708800007 http://www.nejm.org/doi/pdf/10.1056/NEJM200003093421007 >.

NAKANO, H.; OMURA, S. Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. **Journal of Antibiotics,** v. 62, n. 1, p. 17-26, Jan 2009. ISSN 0021-8820. Disponível em: < <Go to ISI>://WOS:000263386900004 http://www.nature.com/articles/ja20084.pdf >.

NEWMAN, D. J.; CRAGG, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. **Journal of Natural Products,** v. 75, n. 3, p. 311-335, Mar 2012. ISSN 0163-3864. Disponível em: < <Go to ISI>://WOS:000301810700002 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3721181/pdf/nihms356104.pdf >.

NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. **J Nat Prod,** v. 75, n. 3, p. 311-35, Mar 23 2012. ISSN 1520-6025 (Electronic) 0163-3864 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22316239 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3721181/pdf/nihms356104.pdf >.

NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. **Journal of Natural Products,** v. 79, n. 3, p. 629-661, Mar 2016. ISSN 0163-3864. Disponível em: < <Go to ISI>://WOS:000373031200024 http://pubs.acs.org/doi/pdfplus/10.1021/acs.jnatprod.5b01055 >.

NEWMAN, D. J.; CRAGG, G. M.; SNADER, K. M. The influence of natural products upon drug discovery. **Natural Product Reports,** v. 17, n. 3, p. 215-234, 2000. ISSN 0265-0568. Disponível em: < <Go to ISI>://WOS:000087151900002 >.

NOUGAYREDE, J. P. et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. **Science,** v. 313, n. 5788, p. 848-851, Aug 2006. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000239671300064 http://science.sciencemag.org/content/313/5788/848.long >.

OMURA, S. et al. Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites. **Proceedings of the National Academy of Sciences of the United States of America,** v. 98, n. 21, p. 12215-12220, Oct 9 2001. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000171558900071 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC59794/pdf/pq012215.pdf >.

OWEN, J. G. et al. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. **Proceedings of the National Academy of Sciences of the United States of America,** v. 112, n. 14, p. 4221-4226, Apr 7 2015. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000352287800030 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394318/pdf/pnas.201501124.pdf >.

OWEN, J. G. et al. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. **Proceedings of the National Academy of Sciences of the United States of America,** v. 110, n. 29, p. 11797-11802, Jul 16 2013. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000322086100041 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3718090/pdf/pnas.201222159.pdf >.

POOTOOLAL, J.; NEU, J.; WRIGHT, G. D. Glycopeptide antibiotic resistance. **Annual Review of Pharmacology and Toxicology,** v. 42, p. 381-408, 2002. ISSN 0362-1642. Disponível em: < <Go to ISI>://WOS:000174038800017 http://www.annualreviews.org/doi/pdf/10.1146/annurev.pharmtox.42.091601.142813 >.

PRADE, L. et al. Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. **Structure,** v. 5, n. 12, p. 1627-1637, Dec 15 1997. ISSN 0969-2126. Disponível em: < <Go to ISI>://WOS:000071387800008 >.

PYLYPENKO, O. et al. Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. **Journal of Biological Chemistry,** v. 278, n. 47, p. 46727-46733, Nov 21 2003. ISSN 0021-9258. Disponível em: < <Go to ISI>://WOS:000186569400063 >.

REDDY, B. V. B. et al. eSNaPD: A Versatile, Web-Based Bioinformatics Platform for Surveying and Mining Natural Product Biosynthetic Diversity from Metagenomes. **Chemistry & Biology,** v. 21, n. 8, p. 1023-1033, Aug 14 2014. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000340947300013 >.

RINEHART, K. L.; SHIELD, L. S. Chemistry of the ansamycin antibiotics. **Fortsch Chem Org Naturst,** v. 33, p. 231 - 307, 1976.

ROBBINS, T. et al. Structure and mechanism of assembly line polyketide synthases. **Curr Opin Struct Biol,** v. 41, p. 10-18, Dec 2016. ISSN 0959-440x.

ROH, C. et al. Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. **Applied Biochemistry and Biotechnology,** v. 134, n. 2, p. 97-112, Aug 2006. ISSN 0273-2289. Disponível em: < <Go to ISI>://WOS:000239575800001 https://link.springer.com/content/pdf/10.1385%2FABAB%3A134%3A2%3A97.pdf >.

RONDON, M. R. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. **Applied and Environmental Microbiology,** v. 66, n. 6, p. 2541-2547, Jun 2000. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000087358700037 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC110579/pdf/am002541.pdf >.

RONDON, M. R. et al. Toward functional genomics in bacteria: Analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus. **Proceedings of the National Academy of Sciences of the United States of America,** v. 96, n. 11, p. 6451-6455, May 25 1999. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000080527100097 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC26902/pdf/pq006451.pdf >.

RYAN, K. S.; DRENNAN, C. L. Divergent Pathways in the Biosynthesis of Bisindole Natural Products. **Chemistry & Biology,** v. 16, n. 4, p. 351-364, Apr 24 2009. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000265816900002 >.

SANCHEZ, C.; MENDEZ, C.; SALAS, J. A. Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. **Natural Product Reports,** v. 23, n. 6, p. 1007-1045, 2006. ISSN 0265-0568. Disponível em: < <Go to ISI>://WOS:000242220600010 http://pubs.rsc.org/en/Content/ArticleLanding/2006/NP/B601930G >.

SAUSVILLE, E. A. et al. Phase I trial of 72-hour continuous infusion UCN-01 in patients with refractory neoplasms. **Journal of Clinical Oncology,** v. 19, n. 8, p. 2319-2333, Apr 15 2001. ISSN 0732-183x. Disponível em: < <Go to ISI>://WOS:000168178300025 >.

SCHIPPER, C. et al. Metagenome-Derived Clones Encoding Two Novel Lactonase Family Proteins Involved in Biofilm Inhibition in Pseudomonas aeruginosa. **Applied and Environmental Microbiology,** v. 75, n. 1, p. 224-233, Jan 2009. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000262084800025 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2612230/pdf/1389-08.pdf >.

SCHIRMER, A. et al. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. **Applied and Environmental Microbiology,** v. 71, n. 8, p. 4840-4849, Aug 2005. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000231165500085 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1183291/pdf/1865-04.pdf >.

SCHLOSS, P. D.; HANDELSMAN, J. Status of the microbial census. **Microbiology and Molecular Biology Reviews,** v. 68, n. 4, p. 686-+, Dec 2004. ISSN 1092-2172. Disponível em: < <Go to ISI>://WOS:000225854100006 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC539005/pdf/0024-04.pdf >.

SCHWANDT, A. et al. Phase-II Trial of Rebeccamycin Analog, a Dual Topoisomerase-I and -II Inhibitor, in Relapsed "Sensitive" Small Cell Lung Cancer. **Journal of Thoracic Oncology,** v. 7, n. 4, p. 751-754, Apr 2012. ISSN 1556-0864. Disponível em: < <Go to ISI>://WOS:000301866500017 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3310884/pdf/nihms-356604.pdf >.

SEOW, K. T. et al. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. **Journal of Bacteriology,** v. 179, n. 23, p. 7360-7368, Dec 1997. ISSN 0021-9193. Disponível em: < <Go to ISI>://WOS:A1997YJ68200020 http://jb.asm.org/content/179/23/7360.full.pdf >.

SERTUERNER, F. Ueber das Morphium, eine neue salzfähige Grundlage, und die Mekonsäure, als Hauptbestandtheile des Opiums. **Ann Physik,** v. 55, p. 56 - 89, 1817.

SHAO, Z. et al. Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. **Acs Synthetic Biology,** v. 2, n. 11, p. 662-669, Nov 2013. ISSN 2161-5063. Disponível em: < <Go to ISI>://WOS:000327175400005 http://pubs.acs.org/doi/pdfplus/10.1021/sb400058n >.

SNEADER, W. The discovery of aspirinL a reappraisal. **BMJ,** v. 321, n. 7276, p. 1591 - 1594, 2000. Disponível em: < https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 1119266/pdf/1591.pdf >.

SOGIN, M. L. et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". **Proceedings of the National Academy of Sciences of the United States of America,** v. 103, n. 32, p. 12115-12120, Aug 8 2006. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000239701900053 http://www.pnas.org/content/103/32/12115.full.pdf >.

STAKER, B. L. et al. Structures of three classes of anticancer agents bound to the human topoisomerase I-DNA covalent complex. **Journal of Medicinal Chemistry,** v. 48, n. 7, p. 2336-2345, Apr 7 2005. ISSN 0022-2623. Disponível em: < <Go to ISI>://WOS:000228111500013 http://pubs.acs.org/doi/pdfplus/10.1021/jm049146p >.

TORSVIK, V.; OVREAS, L. Microbial diversity and function in soil: from genes to ecosystems. **Curr Opin Microbiol,** v. 5, n. 3, p. 240-5, Jun 2002. ISSN 1369-5274 (Print) 1369-5274 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/ 12057676 >.

TORSVIK, V. et al. COMPARISON OF PHENOTYPIC DIVERSITY AND DNA HETEROGENEITY IN A POPULATION OF SOIL BACTERIA. **Applied and Environmental Microbiology,** v. 56, n. 3, p. 776-781, Mar 1990. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:A1990CQ73000031 http://aem.asm.org/content/56/3/776.full.pdf >.

TRINDADE, M. et al. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. **Frontiers in Microbiology,** v. 6, Aug 28 2015. ISSN 1664-302X. Disponível em: < <Go to ISI>://WOS:000360341800001 >.

TRINGE, S. G. et al. Comparative metagenomics of microbial communities. **Science,** v. 308, n. 5721, p. 554-557, Apr 22 2005. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000228810500053 http://science.sciencemag.org/content/308/5721/554.long >.

TYSON, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. **Nature,** v. 428, n. 6978, p. 37-43, Mar 4 2004. ISSN 0028-0836. Disponível em: < <Go to ISI>://WOS:000189363800028 http://www.nature.com/articles/nature02340.pdf >.

VAN WAGENINGEN, A. M. A. et al. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. **Chemistry & Biology,** v. 5, n. 3, p. 155-162, Mar 1998. ISSN 1074-5521. Disponível em: < <Go to ISI>://WOS:000072935700005 >.

VENTER, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. **Science,** v. 304, n. 5667, p. 66-74, Apr 2 2004. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000220567900037 http://science.sciencemag.org/content/304/5667/66.long >.

VOGET, S. et al. Prospecting for novel biocatalysts in a soil metagenome. **Applied and Environmental Microbiology,** v. 69, n. 10, p. 6235-6242, Oct 2003. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000185881300063 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC201203/pdf/0460.pdf >.

WANG, G. Y. S. et al. Novel natural products from soil DNA libraries in a streptomycete host. **Organic Letters,** v. 2, n. 16, p. 2401-2404, Aug 10 2000. ISSN 1523-7060. Disponível em: < <Go to ISI>://WOS:000088605200003 http://pubs.acs.org/doi/pdfplus/10.1021/ol005860z >.

WANG, H. X. et al. PCR screening reveals considerable unexploited biosynthetic potential of ansamycins and a mysterious family of AHBA-containing natural products in actinomycetes. **J Appl Microbiol,** v. 115, n. 1, p. 77-85, Jul 2013. ISSN 1365-2672 (Electronic)1364-5072 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/23594089 http://onlinelibrary.wiley.com/store/10.1111/jam.12217/asset/jam12217.pdf?v=1&t=j6zl6lfa&s=f344b4cb0370c8b45f33838f36abb3effafdd971 >.

WANG, K. et al. A novel metagenome-derived beta-galactosidase: gene cloning, overexpression, purification and characterization. **Applied Microbiology and Biotechnology,** v. 88, n. 1, p. 155-165, Sep 2010. ISSN 0175-7598. Disponível em: < <Go to ISI>://WOS:000280914900017 https://link.springer.com/content/pdf/10.1007%2Fs00253-010-2744-7.pdf >.

WILLIAMS, D. H. et al. Why Are Secondary Metabolites (Natural-Products) Biosynthesized. **Journal of Natural Products,** v. 52, n. 6, p. 1189-1208, Nov-Dec 1989. ISSN 0163-3864. Disponível em: < <Go to ISI>://WOS:A1989CK57600001 http://pubs.acs.org/doi/pdfplus/10.1021/np50066a001 >.

WILLIAMSON, L. L. et al. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. **Applied and Environmental Microbiology,** v. 71, n. 10, p. 6335-6344, Oct 2005. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000232504000082 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1265936/pdf/0287-05.pdf >.

WILSON, M. C. et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. **Nature,** v. 506, n. 7486, p. 58-+, Feb 6 2014. ISSN 0028-0836. Disponível em: < <Go to ISI>://WOS:000330648100030 http://www.nature.com/articles/nature12959.pdf >.

WOITHE, K. et al. Exploring the substrate specificity of OxyB, a phenol coupling P450 enzyme involved in vancomycin biosynthesis. **Organic & Biomolecular Chemistry,** v. 6, n. 16, p. 2861-2867, Aug 21 2008. ISSN 1477-0520. Disponível em: < <Go to ISI>://WOS:000258770000004 >.

WOITHE, K. et al. Oxidative phenol coupling reactions catalyzed by OxyB: A cytochrome p450 from the vancomycin producing organism. Implications for vancomycin biosynthesis. **Journal of the American Chemical Society,** v. 129, n. 21, p. 6887-6895, May 30 2007. ISSN 0002-7863. Disponível em: < <Go to ISI>://WOS:000246686700050 >.

WOLFGANG, M. C. et al. Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. **Proceedings of the National Academy of Sciences of the United States of America,** v. 100, n. 14, p. 8484-8489, Jul 2003. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000184222500080 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC166255/pdf/1008484.pdf >.

WYATT, M. A. et al. Staphylococcus aureus Nonribosomal Peptide Secondary Metabolites Regulate Virulence. **Science,** v. 329, n. 5989, p. 294-296, Jul 2010. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000279925900032 http://science.sciencemag.org/content/329/5989/294.long >.

ZENGLER, K. et al. Cultivating the uncultured. **Proceedings of the National Academy of Sciences of the United States of America,** v. 99, n. 24, p. 15681-15686, Nov 26 2002. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000179530000077 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC137776/pdf/pq2402015681.pdf >.

ZENK, M. H.; JUENGER, M. Evolution and current status of the phytochemistry of nitrogenous compounds. **Phytochemistry,** v. 68, n. 22-24, p. 2757-2772, Nov-Dec 2007. ISSN 0031-9422. Disponível em: < <Go to ISI>://WOS:000252586300011 >.

ZERBE, K. et al. Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. **Journal of Biological Chemistry,** v. 277, n. 49, p. 47476-47485, Dec 6 2002. ISSN 0021-9258. Disponível em: < <Go to ISI>://WOS:000179663700084 >.

ZERBE, K. et al. An oxidative phenol coupling reaction catalyzed by OxyB, a cytochrome p450 from the vancomycin-producing microorganism. **Angewandte Chemie-International Edition,** v. 43, n. 48, p. 6709-6713, 2004. ISSN 1433-7851. Disponível em: < <Go to ISI>://WOS:000225869900026 >.

ZHANG, Z.; PAN, H. X.; TANG, G. L. New insights into bacterial type II polyketide biosynthesis. **F1000Res,** v. 6, p. 172, 2017. ISSN 2046-1402 (Print) 2046-1402. Disponível em: < https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5321127/pdf/f1000research-6-11276.pdf >.

ZIEMERT, N. et al. Exploiting the Natural Diversity of Microviridin Gene Clusters for Discovery of Novel Tricyclic Depsipeptides. **Applied and Environmental Microbiology,** v. 76, n. 11, p. 3568-3574, Jun 2010. ISSN 0099-2240. Disponível em: < <Go to ISI>://WOS:000277871400020 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2876452/pdf/2858-09.pdf >.

ZMIJEWSKI, M. J. et al. Biosynthetic-Studies on Antibiotic A47934. **Antimicrobial Agents and Chemotherapy,** v. 31, n. 10, p. 1497-1501, Oct 1987. ISSN 0066-4804. Disponível em: < <Go to ISI>://WOS:A1987K355200010 http://aac.asm.org/content/31/10/1497.full.pdf >.

ANNEX

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_2DOIF01	TSAAGAACGCSCTGGCCATCC	CTACACGACGCTCTTCCGATCT	ACGTATCA	CTA	CTACACGACGCTCTTCCGATCTACGTATCACTATSAAGAACGCSCTGGCCATCC	UF_miSeq_2DOIR01	GTAGCCSCGCTTGTTGTCGWA	CAGACGTGTGCTCTTCCGATCT	ACGTATCA	CTA	CAGACGTGTGCTCTTCCGATCTACGTATCACTAGTAGCCSCGCTTGTTGTCGWA	Aminoglycoside	486
UF_miSeq_2DOIF02	TSAAGAACKCSCTSGCCRTCC	CTACACGACGCTCTTCCGATCT	AGATCGCA	GATA	CTACACGACGCTCTTCCGATCTAGATCGCAGATATSAAGAACKCSCTSGCCRTCC	UF_miSeq_2DOIR02	RTAGCCSCGCTTGTTGTCGWA	CAGACGTGTGCTCTTCCGATCT	AGATCGCA	GATA	CAGACGTGTGCTCTTCCGATCTAGATCGCAGATARTAGCCSCGCTTGTTGTCGWA	Aminoglycoside	486
UF_miSeq_2DOIF03	GTSAARAAYGCSCTSGCSATY	CTACACGACGCTCTTCCGATCT	ATTGGCTC	GATA	CTACACGACGCTCTTCCGATCTATTGGCTCGATAGTSAARAAYGCSCTSGCSATY	UF_miSeq_2DOIR03	GTASCCGCGCTTRTTRTCRTA	CAGACGTGTGCTCTTCCGATCT	ATTGGCTC	GATA	CAGACGTGTGCTCTTCCGATCTATTGGCTCGATAGTASCCGCGCTTRTTRTCRTA	Aminoglycoside	486
UF_miSeq_2DOIF04	GTVAARAAYGCVCTVGCVATH	CTACACGACGCTCTTCCGATCT	CCAGTTCA	A	CTACACGACGCTCTTCCGATCTCCAGTTCAAGTVAARAAYGCVCTVGCVATH	UF_miSeq_2DOIR04	RTABCCGCKCTTRTTRTCRTA	CAGACGTGTGCTCTTCCGATCT	CCAGTTCA	A	CAGACGTGTGCTCTTCCGATCTCCAGTTCAARTABCCGCKCTTRTTRTCRTA	Aminoglycoside	486
UF_miSeq_2DOIF05	GTNAARAAYGCNYTNGCNATH	CTACACGACGCTCTTCCGATCT	CTAAGGTC	TC	CTACACGACGCTCTTCCGATCTCTAAGGTCTCGTNAARAAYGCNYTNGCNATH	UF_miSeq_2DOIR05	RTANCCNCKYTTRTTRTCRTA	CAGACGTGTGCTCTTCCGATCT	CTAAGGTC	TC	CAGACGTGTGCTCTTCCGATCTCTAAGGTCTCRTANCCNCKYTTRTTRTCRTA	Aminoglycoside	486
UF_miSeq_2DOIF06	GTSAARAAYGCSCTSGCVATY	CTACACGACGCTCTTCCGATCT	ACGCTCGA	CTA	CTACACGACGCTCTTCCGATCTACGCTCGACTAGTSAARAAYGCSCTSGCVATY	UF_miSeq_2DOIR06	RTASCCSCKYTTRTTRTCRTA	CAGACGTGTGCTCTTCCGATCT	CATCAAGT	CTA	CAGACGTGTGCTCTTCCGATCTCATCAAGTCTARTASCCSCKYTTRTTRTCRTA	Aminoglycoside	486
UF_miSeq_2DOIF07	GTVAARAAYGCSCTSGCVATH	CTACACGACGCTCTTCCGATCT	GAACAGGC	GATA	CTACACGACGCTCTTCCGATCTGAACAGGCGATAGTVAARAAYGCSCTSGCVATH	UF_miSeq_2DOIR07	RTABCCBCKYTTRTTRTCRTA	CAGACGTGTGCTCTTCCGATCT	GATAGACA	GATA	CAGACGTGTGCTCTTCCGATCTGATAGACAGATARTABCCBCKYTTRTTRTCRTA	Aminoglycoside	486
UF_miSeq_2DOIF08	GTNAARAAYGCVYTVGCVATH	CTACACGACGCTCTTCCGATCT	AACGCTTA	A	CTACACGACGCTCTTCCGATCTAACGCTTAAGTNAARAAYGCVYTVGCVATH	UF_miSeq_2DOIR08	VWASCCNCGSTKSTKSTCVWA	CAGACGTGTGCTCTTCCGATCT	ACCACTGT	A	CAGACGTGTGCTCTTCCGATCTACCACTGTAVWASCCNCGSTKSTKSTCVWA	Aminoglycoside	486
UF_miSeq_2DOIF09	VTNMASMASGCSVTSGCSVTS	CTACACGACGCTCTTCCGATCT	CACCTTAC	TC	CTACACGACGCTCTTCCGATCTCACCTTACTCVTNMASMASGCSVTSGCSVTS	UF_miSeq_2DOIR09	VWASCCNCGSTKSTKNTCVWA	CAGACGTGTGCTCTTCCGATCT	ATTGAGGA	TC	CAGACGTGTGCTCTTCCGATCTATTGAGGATCVWASCCNCGSTKSTKNTCVWA	Aminoglycoside	486
UF_miSeq_2DOIF10	VTNMANMASGCSVTSGCSVTS	CTACACGACGCTCTTCCGATCT	CGACACAC	CTA	CTACACGACGCTCTTCCGATCTCGACACACCTAVTNMANMASGCSVTSGCSVTS	UF_miSeq_2DOIR10	VWASCCNCGSTKNTKNTCVWA	CAGACGTGTGCTCTTCCGATCT	ATTGGCTC	CTA	CAGACGTGTGCTCTTCCGATCTATTGGCTCCTAVWASCCNCGSTKNTKNTCVWA	Aminoglycoside	486
UF_miSeq_AHBAF01	TTCCAGAACGGMAARCTGATG	CTACACGACGCTCTTCCGATCT	CGAACTTA	TC	CTACACGACGCTCTTCCGATCTCGAACTTATCTTCCAGAACGGMAARCTGATG	UF_miSeq_AHBAR01	GGAACATSGCCATGTAGTGSG	CAGACGTGTGCTCTTCCGATCT	CGAACTTA	TC	CAGACGTGTGCTCTTCCGATCTCGAACTTATCGGAACATSGCCATGTAGTGSG	AHBA	395
UF_miSeq_AHBAF02	TTCCAGAACGGMAARYTGATG	CTACACGACGCTCTTCCGATCT	ACAGCAGA	CTA	CTACACGACGCTCTTCCGATCTACAGCAGACTATTCCAGAACGGMAARYTGATG	UF_miSeq_AHBAR02	GGAACATSGCCATGTAGYKSG	CAGACGTGTGCTCTTCCGATCT	ACAGCAGA	CTA	CAGACGTGTGCTCTTCCGATCTACAGCAGACTAGGAACATSGCCATGTAGYKSG	AHBA	395
UF_miSeq_AHBAF03	TTYCARAAYGGSAARCTSATG	CTACACGACGCTCTTCCGATCT	CGGATTGC	GATA	CTACACGACGCTCTTCCGATCTCGGATTGCGATATTYCARAAYGGSAARCTSATG	UF_miSeq_AHBAR03	KRAACATSGCCATRTARTGVS	CAGACGTGTGCTCTTCCGATCT	CGGATTGC	GATA	CAGACGTGTGCTCTTCCGATCTCGGATTGCGATAKRAACATSGCCATRTARTGVS	AHBA	395
UF_miSeq_AHBAF04	TTCCAGAACGGCAAACTGATG	CTACACGACGCTCTTCCGATCT	GATGAATC	A	CTACACGACGCTCTTCCGATCTGATGAATCATTCCAGAACGGCAAACTGATG	UF_miSeq_AHBAR04	GAACATCGCCATGTAGTGCGA	CAGACGTGTGCTCTTCCGATCT	GATGAATC	A	CAGACGTGTGCTCTTCCGATCTGATGAATCAGAACATCGCCATGTAGTGCGA	AHBA	395
UF_miSeq_AHBAF05	TTYCARAAYGGSAAAYTSATG	CTACACGACGCTCTTCCGATCT	GACAGTGC	TC	CTACACGACGCTCTTCCGATCTGACAGTGCTCTTYCARAAYGGSAAAYTSATG	UF_miSeq_AHBAR05	RAACATSGCCATRTARTGRGA	CAGACGTGTGCTCTTCCGATCT	GACAGTGC	TC	CAGACGTGTGCTCTTCCGATCTGACAGTGCTCRAACATSGCCATRTARTGRGA	AHBA	395
UF_miSeq_AHBAF06	TTYCARAAYGGNAARYTNATG	CTACACGACGCTCTTCCGATCT	ATTGAGGA	CTA	CTACACGACGCTCTTCCGATCTATTGAGGACTATTYCARAAYGGNAARYTNATG	UF_miSeq_AHBAR06	RAACATNGCCATRTARTGRCT	CAGACGTGTGCTCTTCCGATCT	ATTGAGGA	CTA	CAGACGTGTGCTCTTCCGATCTATTGAGGACTARAACATNGCCATRTARTGRCT	AHBA	395
UF_miSeq_AHBAF07	TTYCARAAYGGVAARYTVATG	CTACACGACGCTCTTCCGATCT	ATTGGCTC	TC	CTACACGACGCTCTTCCGATCTATTGGCTCTCTTYCARAAYGGVAARYTVATG	UF_miSeq_AHBAR07	SCGSWACATSGCCATSWASTK	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	TC	CAGACGTGTGCTCTTCCGATCTCAAGGAGCTCSCGSWACATSGCCATSWASTK	AHBA	395
UF_miSeq_AHBAF08	TWSMASMASGGSMASVTSATG	CTACACGACGCTCTTCCGATCT	ATCATTCC	CTA	CTACACGACGCTCTTCCGATCTATCATTCCCTATWSMASMASGGSMASVTSATG	UF_miSeq_AHBAR08	SCGVWACATSGCCATVWANTK	CAGACGTGTGCTCTTCCGATCT	CAGCGTTA	CTA	CAGACGTGTGCTCTTCCGATCTCAGCGTTACTASCGVWACATSGCCATVWANTK	AHBA	395
UF_miSeq_AHBAF09	TWNMASMASGGSMASVTSATG	CTACACGACGCTCTTCCGATCT	TGGAACAA	GATA	CTACACGACGCTCTTCCGATCTTGGAACAAGATATWNMASMASGGSMASVTSATG	UF_miSeq_AHBAR09	SCGVWACATNGCCATVWANTK	CAGACGTGTGCTCTTCCGATCT	CGCTGATC	GATA	CAGACGTGTGCTCTTCCGATCTCGCTGATCGATASCGVWACATNGCCATVWANTK	AHBA	395
UF_miSeq_AHBAF10	TWNMANMASGGSMASVTSATG	CTACACGACGCTCTTCCGATCT	AGAGTCAA	A	CTACACGACGCTCTTCCGATCTAGAGTCAAATWNMANMASGGSMASVTSATG	UF_miSeq_AHBAR10	NCGVWACATNGCCATVWANTK	CAGACGTGTGCTCTTCCGATCT	AGATGTAC	A	CAGACGTGTGCTCTTCCGATCTAGATGTACANCGVWACATNGCCATVWANTK	AHBA	395
UF_miSeq_CPF01	SGACATCAGCAGCAGCGGYG	CTACACGACGCTCTTCCGATCT	CGCTGATC	TC	CTACACGACGCTCTTCCGATCTCGCTGATCTCSGACATCAGCAGCAGCGGYG	UF_miSeq_CPR01	CGSTSGGGTTGTGCGGGAAG	CAGACGTGTGCTCTTCCGATCT	CGCTGATC	TC	CAGACGTGTGCTCTTCCGATCTCGCTGATCTCCGSTSGGGTTGTGCGGGAAG	Capreomycidine	495
UF_miSeq_CPF02	SGACATCWSCAGCAGCGGCG	CTACACGACGCTCTTCCGATCT	AACAACCA	CTA	CTACACGACGCTCTTCCGATCTAACAACCACTASGACATCWSCAGCAGCGGCG	UF_miSeq_CPR02	CGSTSGGGTTGTGCGGGAAG	CAGACGTGTGCTCTTCCGATCT	AACAACCA	CTA	CAGACGTGTGCTCTTCCGATCTAACAACCACTACGSTSGGGTTGTGCGGGAAG	Capreomycidine	495
UF_miSeq_CPF03	SGACATCWSCAGCAGCGGYG	CTACACGACGCTCTTCCGATCT	AACGCTTA	GATA	CTACACGACGCTCTTCCGATCTAACGCTTAGATASGACATCWSCAGCAGCGGYG	UF_miSeq_CPR03	CRSTSGGGTTGTGSGGGAAG	CAGACGTGTGCTCTTCCGATCT	AACGCTTA	GATA	CAGACGTGTGCTCTTCCGATCTAACGCTTAGATACRSTSGGGTTGTGSGGGAAG	Capreomycidine	495
UF_miSeq_CPF04	SGACRTCWSCAGCAGCGGYG	CTACACGACGCTCTTCCGATCT	AAGGTACA	A	CTACACGACGCTCTTCCGATCTAAGGTACAASGACRTCWSCAGCAGCGGYG	UF_miSeq_CPR04	CRSTSGGGTTGTGSGGGAAG	CAGACGTGTGCTCTTCCGATCT	AAGGTACA	A	CAGACGTGTGCTCTTCCGATCTAAGGTACAACRSTSGGGTTGTGSGGGAAG	Capreomycidine	495
UF_miSeq_CPF05	SGACRTSWSCWSSWSCGGYGT	CTACACGACGCTCTTCCGATCT	CTGAGCCA	TC	CTACACGACGCTCTTCCGATCTCTGAGCCATCSGACRTSWSCWSSWSCGGYGT	UF_miSeq_CPR05	CRSTNGGRTTRTGSGGRAAGT	CAGACGTGTGCTCTTCCGATCT	CTGAGCCA	TC	CAGACGTGTGCTCTTCCGATCTCTGAGCCATCCRSTNGGRTTRTGSGGRAAGT	Capreomycidine	495
UF_miSeq_CPF06	GAYGTSWSSWSSWSSGGBGTN	CTACACGACGCTCTTCCGATCT	CTGGCATA	CTA	CTACACGACGCTCTTCCGATCTCTGGCATACTAGAYGTSWSSWSSWSSGGBGTN	UF_miSeq_CPR06	NCCBGTNGGRTTRTGSGGRAA	CAGACGTGTGCTCTTCCGATCT	CTGGCATA	CTA	CAGACGTGTGCTCTTCCGATCTCTGGCATACTANCCBGTNGGRTTRTGSGGRAA	Capreomycidine	495
UF_miSeq_CPF07	GAYGTBWSSWSBWSNGGNGTN	CTACACGACGCTCTTCCGATCT	GAATCTGA	GATA	CTACACGACGCTCTTCCGATCTGAATCTGAGATAGAYGTBWSSWSBWSNGGNGTN	UF_miSeq_CPR07	NCCNGTNGGRTTRTGNGGRAA	CAGACGTGTGCTCTTCCGATCT	GAATCTGA	GATA	CAGACGTGTGCTCTTCCGATCTGAATCTGAGATANCCNGTNGGRTTRTGNGGRAA	Capreomycidine	495
UF_miSeq_CPF08	GAYGTNWSNWSNWSNGGNGTN	CTACACGACGCTCTTCCGATCT	CAAGACTA	A	CTACACGACGCTCTTCCGATCTCAAGACTAAGAYGTNWSNWSNWSNGGNGTN	UF_miSeq_CPR08	NCCNGTNGGRTYNTBNGGRAA	CAGACGTGTGCTCTTCCGATCT	CAAGACTA	A	CAGACGTGTGCTCTTCCGATCTCAAGACTAANCCNGTNGGRTYNTBNGGRAA	Capreomycidine	495
UF_miSeq_CPF09	SGYGTNWSNWSNWSNGGNGTN	CTACACGACGCTCTTCCGATCT	GAGCTGAA	TC	CTACACGACGCTCTTCCGATCTGAGCTGAATCSGYGTNWSNWSNWSNGGNGTN	UF_miSeq_CPR09	NCCNGTNGGNTBNTBNGGRAA	CAGACGTGTGCTCTTCCGATCT	GAGCTGAA	TC	CAGACGTGTGCTCTTCCGATCTGAGCTGAATCNCCNGTNGGNTBNTBNGGRAA	Capreomycidine	495

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_CPF10	GAYATHAGYAGYAGYGGVGTS	CTACACGACGCTCTTCCGATCT	CCTCCTGA	CTA	CTACACGACGCTCTTCCGATCTCCTCCTGACTAGAYATHAGYAGYAGYGGVGTS	UF_miSeq_CPR10	RCTSGGRTTRTGBGGRAARTT	CAGACGTGTGCTCTTCCGATCT	CCTCCTGA	CTA	CAGACGTGTGCTCTTCCGATCTCCTCCTGACTARCTSGGRTTRTGBGGRAARTT	Capreomycidine	495
UF_miSeq_CPF11	GAYGTSWSSWSSWSSGGSGTS	CTACACGACGCTCTTCCGATCT	TAGGATGA	GATA	CTACACGACGCTCTTCCGATCTTAGGATGAGATAGAYGTSWSSWSSWSSGGSGTS	UF_miSeq_CPR11	YGDSGGSTKSTKSGGSWASTK	CAGACGTGTGCTCTTCCGATCT	TGGCTTCA	GATA	CAGACGTGTGCTCTTCCGATCTTGGCTTCAGATAYGDSGGSTKSTKSGGSWASTK	Capreomycidine	495
UF_miSeq_CPF12	GAYGTSWSSWSSWSSGGBGTV	CTACACGACGCTCTTCCGATCT	GCTCGGTA	A	CTACACGACGCTCTTCCGATCTGCTCGGTAAGAYGTSWSSWSSWSSGGBGTV	UF_miSeq_CPR12	SGDSGGSTKSTKSGGVWASTK	CAGACGTGTGCTCTTCCGATCT	GATAGACA	A	CAGACGTGTGCTCTTCCGATCTGATAGACAASGDSGGSTKSTKSGGVWASTK	Capreomycidine	495
UF_miSeq_CPF13	GASVTSHCRHCRHCRGGSVTS	CTACACGACGCTCTTCCGATCT	ATCATTCC	CTA	CTACACGACGCTCTTCCGATCTATCATTCCCTAGASVTSHCRHCRHCRGGSVTS	UF_miSeq_CPR13	SGDSGGSTKSTKSGGVWANTK	CAGACGTGTGCTCTTCCGATCT	CAAGACTA	CTA	CAGACGTGTGCTCTTCCGATCTCAAGACTACTASGDSGGSTKSTKSGGVWANTK	Capreomycidine	495
UF_miSeq_CPF14	GANVTSHCSHCSHCSGGSVTS	CTACACGACGCTCTTCCGATCT	AGCCATGC	GATA	CTACACGACGCTCTTCCGATCTAGCCATGCGATAGANVTSHCSHCSHCSGGSVTS	UF_miSeq_CPR14	SGDSGGSTKSTKNGGVWANTK	CAGACGTGTGCTCTTCCGATCT	GTCGTAGA	GATA	CAGACGTGTGCTCTTCCGATCTGTCGTAGAGATASGDSGGSTKSTKNGGVWANTK	Capreomycidine	495
UF_miSeq_EndurF01	AGGGSATGCCCAARCAGTTCG	CTACACGACGCTCTTCCGATCT	ACAGATTC	CTA	CTACACGACGCTCTTCCGATCTACAGATTCCTAAGGGSATGCCCAARCAGTTCG	UF_miSeq_EndurR01	ARSGTCTCSSCRTASGMGAAC	CAGACGTGTGCTCTTCCGATCT	ACAGATTC	CTA	CAGACGTGTGCTCTTCCGATCTACAGATTCCTAARSGTCTCSSCRTASGMGAAC	Enduracidine	464
UF_miSeq_EndurF02	CAGGGSATGCCSAARCARTTY	CTACACGACGCTCTTCCGATCT	CGACTGGA	GATA	CTACACGACGCTCTTCCGATCTCGACTGGAGATACAGGGSATGCCSAARCARTTY	UF_miSeq_EndurR02	SAASGTYTCSCCRTASGARAA	CAGACGTGTGCTCTTCCGATCT	CGACTGGA	GATA	CAGACGTGTGCTCTTCCGATCTCGACTGGAGATASAASGTYTCSCCRTASGARAA	Enduracidine	464
UF_miSeq_EndurF03	CARGGSATGCCVAARCARTTY	CTACACGACGCTCTTCCGATCT	CAAGGAGC	A	CTACACGACGCTCTTCCGATCTCAAGGAGCACARGGSATGCCVAARCARTTY	UF_miSeq_EndurR03	BAABGTYTCSCCRTASGARAA	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	A	CAGACGTGTGCTCTTCCGATCTCAAGGAGCABAABGTYTCSCCRTASGARAA	Enduracidine	464
UF_miSeq_EndurF04	CARGGNATGCCNAARCARTTY	CTACACGACGCTCTTCCGATCT	TGGCTTCA	TC	CTACACGACGCTCTTCCGATCTTGGCTTCATCCARGGNATGCCNAARCARTTY	UF_miSeq_EndurR04	NARNGTYTCNCCRTARCTRAA	CAGACGTGTGCTCTTCCGATCT	TGGCTTCA	TC	CAGACGTGTGCTCTTCCGATCTTGGCTTCATCNARNGTYTCNCCRTARCTRAA	Enduracidine	464
UF_miSeq_EndurF05	CARGGSATGCCSAARCARTTY	CTACACGACGCTCTTCCGATCT	GAGCTGAA	A	CTACACGACGCTCTTCCGATCTGAGCTGAAACARGGSATGCCSAARCARTTY	UF_miSeq_EndurR05	SAASGTTTCSCCGTASGAGAA	CAGACGTGTGCTCTTCCGATCT	AAGAGATC	A	CAGACGTGTGCTCTTCCGATCTAAGAGATCASAASGTTTCSCCGTASGAGAA	Enduracidine	464
UF_miSeq_EndurF06	CARGGSATGCCVAARCARTTY	CTACACGACGCTCTTCCGATCT	CTGTAGCC	TC	CTACACGACGCTCTTCCGATCTCTGTAGCCTCCARGGSATGCCVAARCARTTY	UF_miSeq_EndurR06	SAASGTYTCSCCRTASGAGAA	CAGACGTGTGCTCTTCCGATCT	ACAGATTC	TC	CAGACGTGTGCTCTTCCGATCTACAGATTCTCSAASGTYTCSCCRTASGAGAA	Enduracidine	464
UF_miSeq_EndurF07	CARGGVATGCCVAARCARTTY	CTACACGACGCTCTTCCGATCT	ACAGCAGA	CTA	CTACACGACGCTCTTCCGATCTACAGCAGACTACARGGVATGCCVAARCARTTY	UF_miSeq_EndurR07	SAASGTYTCBCCRTASGARAA	CAGACGTGTGCTCTTCCGATCT	TATCAGCA	CTA	CAGACGTGTGCTCTTCCGATCTTATCAGCACTASAASGTYTCBCCRTASGARAA	Enduracidine	464
UF_miSeq_EndurF08	MASGGSATGCCSMASMASTWS	CTACACGACGCTCTTCCGATCT	ACACAGAA	GATA	CTACACGACGCTCTTCCGATCTACACAGAAGATAMASGGSATGCCSMASMASTWS	UF_miSeq_EndurR08	BARBGTYTCBCCRTARCTRAA	CAGACGTGTGCTCTTCCGATCT	CGCATACA	GATA	CAGACGTGTGCTCTTCCGATCTCGCATACAGATABARBGTYTCBCCRTARCTRAA	Enduracidine	464
UF_miSeq_EndurF09	MANGGSATGCCSMASMASTWB	CTACACGACGCTCTTCCGATCT	CTCAATGA	A	CTACACGACGCTCTTCCGATCTCTCAATGAAMANGGSATGCCSMASMASTWB	UF_miSeq_EndurR09	SABYGDSTCSCCSWARGDSAA	CAGACGTGTGCTCTTCCGATCT	AAGGACAC	A	CAGACGTGTGCTCTTCCGATCTAAGGACACASABYGDSTCSCCSWARGDSAA	Enduracidine	464
UF_miSeq_EndurF10	MANGGNATGCCSMASMASTWB	CTACACGACGCTCTTCCGATCT	CAAGGAGC	CTA	CTACACGACGCTCTTCCGATCTCAAGGAGCCTAMANGGNATGCCSMASMASTWB	UF_miSeq_EndurR10	SABYGDSTCSCCSWAYGDSWA	CAGACGTGTGCTCTTCCGATCT	CCATCCTC	CTA	CAGACGTGTGCTCTTCCGATCTCCATCCTCCTASABYGDSTCSCCSWAYGDSWA	Enduracidine	464
UF_miSeq_EndurF11	MANGGNATGCCSMANMASTWB	CTACACGACGCTCTTCCGATCT	ATGCCTAA	GATA	CTACACGACGCTCTTCCGATCTATGCCTAAGATAMANGGNATGCCSMANMASTWB	UF_miSeq_EndurR11	SABSGDSTCSCCVWASGDVWA	CAGACGTGTGCTCTTCCGATCT	CCAGTTCA	GATA	CAGACGTGTGCTCTTCCGATCTCCAGTTCAGATASABSGDSTCSCCVWASGDVWA	Enduracidine	464
UF_miSeq_EndurF12	MANGGNATGCCSMANMANTWB	CTACACGACGCTCTTCCGATCT	CACCTTAC	A	CTACACGACGCTCTTCCGATCTCACCTTACAMANGGNATGCCSMANMANTWB	UF_miSeq_EndurR12	SABSGDSTCSCCVWANGDVWA	CAGACGTGTGCTCTTCCGATCT	ACCACTGT	A	CAGACGTGTGCTCTTCCGATCTACCACTGTASABSGDSTCSCCVWANGDVWA	Enduracidine	464
UF_miSeq_IsoNitriF01	CCAAGCATTACTKCGYTGRAT	CTACACGACGCTCTTCCGATCT	AGGCTAAC	GATA	CTACACGACGCTCTTCCGATCTAGGCTAACGATACCAAGCATTACTKCGYTGRAT	UF_miSeq_IsoNitriR01	AGCCAATTCATTTTATTCTCC	CAGACGTGTGCTCTTCCGATCT	AGGCTAAC	GATA	CAGACGTGTGCTCTTCCGATCTAGGCTAACGATAAGCCAATTCATTTTATTCTCC	Isonitrile	
UF_miSeq_IsoNitriF02	GAYATYAGYAGCAGCGGSGTS	CTACACGACGCTCTTCCGATCT	CCATCCTC	CTA	CTACACGACGCTCTTCCGATCTCCATCCTCCTAGAYATYAGYAGCAGCGGSGTS	UF_miSeq_IsoNitriR02	ACTSGGGTTRTGSGGRAARTT	CAGACGTGTGCTCTTCCGATCT	CCATCCTC	CTA	CAGACGTGTGCTCTTCCGATCTCCATCCTCCTAACTSGGGTTRTGSGGRAARTT	Isonitrile	
UF_miSeq_IsoNitriF03	GAYATHAGYAGYAGYGGVGTS	CTACACGACGCTCTTCCGATCT	GTACGCAA	GATA	CTACACGACGCTCTTCCGATCTGTACGCAAGATAGAYATHAGYAGYAGYGGVGTS	UF_miSeq_IsoNitriR03	RCTSGGRTTRTGBGGRAARTT	CAGACGTGTGCTCTTCCGATCT	GTACGCAA	GATA	CAGACGTGTGCTCTTCCGATCTGTACGCAAGATARCTSGGRTTRTGBGGRAARTT	Isonitrile	
UF_miSeq_IsoNitriF04	GAYATHAGYAGYAGYGGNGTN	CTACACGACGCTCTTCCGATCT	GTGTTCTA	A	CTACACGACGCTCTTCCGATCTGTGTTCTAAGAYATHAGYAGYAGYGGNGTN	UF_miSeq_IsoNitriR04	RCTNGGRTTRTGNGGRAARTT	CAGACGTGTGCTCTTCCGATCT	GTGTTCTA	A	CAGACGTGTGCTCTTCCGATCTGTGTTCTAARCTNGGRTTRTGNGGRAARTT	Isonitrile	
UF_miSeq_IsoNitriF05	GAYATYAGYAGYAGYGGSGTS	CTACACGACGCTCTTCCGATCT	GATAGACA	TC	CTACACGACGCTCTTCCGATCTGATAGACATCGAYATYAGYAGYAGYGGSGTS	UF_miSeq_IsoNitriR05	RCTSGGRTTRTGSGGRAARTT	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	TC	CAGACGTGTGCTCTTCCGATCTTAGGATGATCRCTSGGRTTRTGSGGRAARTT	Isonitrile	
UF_miSeq_IsoNitriF06	GASCCSVTSMASTWSVTSVTS	CTACACGACGCTCTTCCGATCT	CAACCACA	CTA	CTACACGACGCTCTTCCGATCTCAACCACACTAGASCCSVTSMASTWSVTSVTS	UF_miSeq_IsoNitriR06	SWASGCSTKYGDSCGSTKSAB	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	CTA	CAGACGTGTGCTCTTCCGATCTCAAGGAGCCTASWASGCSTKYGDSCGSTKSAB	Isonitrile	
UF_miSeq_IsoNitriF07	GANCCSVTSMASTWBVTSVTS	CTACACGACGCTCTTCCGATCT	CAGATCTG	GATA	CTACACGACGCTCTTCCGATCTCAGATCTGGATAGANCCSVTSMASTWBVTSVTS	UF_miSeq_IsoNitriR07	VWASGCSTKSGDSCGSTKNAB	CAGACGTGTGCTCTTCCGATCT	AACGCTTA	GATA	CAGACGTGTGCTCTTCCGATCTAACGCTTAGATAVWASGCSTKSGDSCGSTKNAB	Isonitrile	
UF_miSeq_IsoNitriF08	GANCCNVTSMASTWBVTSVTS	CTACACGACGCTCTTCCGATCT	CGCATACA	A	CTACACGACGCTCTTCCGATCTCGCATACAAGANCCNVTSMASTWBVTSVTS	UF_miSeq_IsoNitriR08	VWASGCSTKSGDSCGNTKNAB	CAGACGTGTGCTCTTCCGATCT	CAACCACA	A	CAGACGTGTGCTCTTCCGATCTCAACCACAAVWASGCSTKSGDSCGNTKNAB	Isonitrile	
UF_miSeq_IsoPenF01	GTCCTGGTGYTAYCTGAACCC	CTACACGACGCTCTTCCGATCT	TGGAACAA	A	CTACACGACGCTCTTCCGATCTTGGAACAAAGTCCTGGTGYTAYCTGAACCC	UF_miSeq_IsoPenR01	CTCGACCTGSAGGTTCTGYAC	CAGACGTGTGCTCTTCCGATCT	TGGAACAA	A	CAGACGTGTGCTCTTCCGATCTTGGAACAAACTCGACCTGSAGGTTCTGYAC	Isopenicillin	
UF_miSeq_IsoPenF02	GTCCTKSTGYTAYCTGAACCC	CTACACGACGCTCTTCCGATCT	CAGCGTTA	TC	CTACACGACGCTCTTCCGATCTCAGCGTTATCGTCCTKSTGYTAYCTGAACCC	UF_miSeq_IsoPenR02	CTCGAMCTGSAGGKTCTGYAC	CAGACGTGTGCTCTTCCGATCT	CAGCGTTA	TC	CAGACGTGTGCTCTTCCGATCTCAGCGTTATCCTCGAMCTGSAGGKTCTGYAC	Isopenicillin	
UF_miSeq_IsoPenF03	RTCCTKSYSYTAYCYGAACCC	CTACACGACGCTCTTCCGATCT	TAGGATGA	CTA	CTACACGACGCTCTTCCGATCTTAGGATGACTARTCCTKSYSYTAYCYGAACCC	UF_miSeq_IsoPenR03	CTCSAMCTGSAGGKTCKSYAC	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	CTA	CAGACGTGTGCTCTTCCGATCTTAGGATGACTACTCSAMCTGSAGGKTCKSYAC	Isopenicillin	
UF_miSeq_IsoPenF04	AGYTGGTGYTAYYTVAAYCCV	CTACACGACGCTCTTCCGATCT	AAGAGATC	TC	CTACACGACGCTCTTCCGATCTAAGAGATCTCAGYTGGTGYTAYYTVAAYCCV	UF_miSeq_IsoPenR04	YTCBACYTGBARRTTYTGBAC	CAGACGTGTGCTCTTCCGATCT	AAGAGATC	TC	CAGACGTGTGCTCTTCCGATCTAAGAGATCTCYTCBACYTGBARRTTYTGBAC	Isopenicillin	

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_IsoPenF05	AGYTGGTGYTAYYTNAAYCCN	CTACACGACGCTCTTCCGATC	AGTCACTA	CTA	CTACACGACGCTCTTCCGATCTAGTCACTACTAAGYTGGTGYTAYYTNAAYCCN	UF_miSeq_IsoPenR05	YTCNACYTGNARRTTYTGNAC	CAGACGTGTGCTCTTCCGATCT	AGTCACTA	CTA	CAGACGTGTGCTCTTCCGATCTAGTCACTACTAYTCNACYTGNARRTTYTGNAC	Isopenicillin	
UF_miSeq_IsoPenF06	HCRTWSTGYTWSVTSMASCCS	CTACACGACGCTCTTCCGATC	ACAGCAGA	CTA	CTACACGACGCTCTTCCGATCTACAGCAGACTAHCRTWSTGYTWSVTSMASCCS	UF_miSeq_IsoPenR06	STCSABSTKSABSTKSTKSAB	CAGACGTGTGCTCTTCCGATCT	AGCAGGAA	CTA	CAGACGTGTGCTCTTCCGATCTAGCAGGAACTASTCSABSTKSABSTKSTKSAB	Isopenicillin	
UF_miSeq_IsoPenF07	HCSTWSTGYTWSVTSMASCCS	CTACACGACGCTCTTCCGATC	CCAGTTCA	GATA	CTACACGACGCTCTTCCGATCTCCAGTTCAGATAHCSTWSTGYTWSVTSMASCCS	UF_miSeq_IsoPenR07	STCSABSTKSABSTKSTKSAB	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	GATA	CAGACGTGTGCTCTTCCGATCTCAAGGAGCGATASTCSABSTKSABSTKSTKSAB	Isopenicillin	
UF_miSeq_IsoPenF08	HCNTWBTGYTWBVTSMASCCS	CTACACGACGCTCTTCCGATC	AGTCACTA	A	CTACACGACGCTCTTCCGATCTAGTCACTAAHCNTWBTGYTWBVTSMASCCS	UF_miSeq_IsoPenR08	STCSABSTKSABSTKSTKNAB	CAGACGTGTGCTCTTCCGATCT	GCTCGGTA	A	CAGACGTGTGCTCTTCCGATCTGCTCGGTAASTCSABSTKSABSTKSTKNAB	Isopenicillin	
UF_miSeq_NRPADF01	CGTGATGTACACGTCCGGCTC	CTACACGACGCTCTTCCGATC	CGACACAC	TC	CTACACGACGCTCTTCCGATCTCGACACACTCCGTGATGTACACGTCCGGCTC	UF_miSeq_NRPADR1	CACGAACCGYTCCGCSGTCAG	CAGACGTGTGCTCTTCCGATCT	CGACACAC	TC	CAGACGTGTGCTCTTCCGATCTCGACACACTCCACGAACCGYTCCGCSGTCAG	NRPAD	707
UF_miSeq_NRPADF02	CGTSATSTACACSTCCGGCTC	CTACACGACGCTCTTCCGATC	CCGAAGTA	CTA	CTACACGACGCTCTTCCGATCTCCGAAGTACTACGTSATSTACACSTCCGGCTC	UF_miSeq_NRPADR02	CASGAACCGYTCSGCSGTCAG	CAGACGTGTGCTCTTCCGATCT	CCGAAGTA	CTA	CAGACGTGTGCTCTTCCGATCTCCGAAGTACTACASGAACCGYTCSGCSGTCAG	NRPAD	707
UF_miSeq_NRPADF03	YSTSAYSTACACSTCCGGCWC	CTACACGACGCTCTTCCGATC	AGCACCTC	GATA	CTACACGACGCTCTTCCGATCTAGCACCTCGATAYSTSAYSTACACSTCCGGCWC	UF_miSeq_NRPADR03	SASGAAMCGYTCSGCSGTCAG	CAGACGTGTGCTCTTCCGATCT	AGCACCTC	GATA	CAGACGTGTGCTCTTCCGATCTAGCACCTCGATASASGAAMCGYTCSGCSGTCAG	NRPAD	707
UF_miSeq_NRPADF04	GTVATGTAYACVAGYGGVAGY	CTACACGACGCTCTTCCGATC	GATGAATC	GATA	CTACACGACGCTCTTCCGATCTGATGAATCGATAGTVATGTAYACVAGYGGVAGY	UF_miSeq_NRPADR04	BACRAABCKYTCBGCBGTBAR	CAGACGTGTGCTCTTCCGATCT	CTGAGCCA	GATA	CAGACGTGTGCTCTTCCGATCTCTGAGCCAGATABACRAABCKYTCBGCBGTBAR	NRPAD	707
UF_miSeq_NRPADF05	GTNATGTAYACNAGYGGNAGY	CTACACGACGCTCTTCCGATC	GACAGTGC	A	CTACACGACGCTCTTCCGATCTGACAGTGCAGTNATGTAYACNAGYGGNAGY	UF_miSeq_NRPADR05	NACRAANCKYTCNGCNGTNAR	CAGACGTGTGCTCTTCCGATCT	CTGGCATA	A	CAGACGTGTGCTCTTCCGATCTCTGGCATAANACRAANCKYTCNGCNGTNAR	NRPAD	707
UF_miSeq_NRPADF06	VTSATGTWSHCRHCRGGSACY	CTACACGACGCTCTTCCGATC	AGTACAAG	TC	CTACACGACGCTCTTCCGATCTAGTACAAGTCVTSATGTWSHCRHCRGGSACY	UF_miSeq_NRPADR06	SACRAASCKYTCBGCBGTBAR	CAGACGTGTGCTCTTCCGATCT	AACCGAGA	TC	CAGACGTGTGCTCTTCCGATCTAACCGAGATCSACRAASCKYTCBGCBGTBAR	NRPAD	707
UF_miSeq_NRPADF07	VTSATGTWSHCRHCRGGSHCR	CTACACGACGCTCTTCCGATC	GCTCGGTA	CTA	CTACACGACGCTCTTCCGATCTGCTCGGTACTAVTSATGTWSHCRHCRGGSHCR	UF_miSeq_NRPADR07	SABSWASCGSTCSGCYGDSAB	CAGACGTGTGCTCTTCCGATCT	GTCGTAGA	CTA	CAGACGTGTGCTCTTCCGATCTGTCGTAGACTASABSWASCGSTCSGCYGDSAB	NRPAD	707
UF_miSeq_NRPADF08	VTSATGTWBHCSHCSGGSHCS	CTACACGACGCTCTTCCGATC	CGCATACA	GATA	CTACACGACGCTCTTCCGATCTCGCATACAGATAVTSATGTWBHCSHCSGGSHCS	UF_miSeq_NRPADR08	SABVWASCGSTCSGCSGDNAB	CAGACGTGTGCTCTTCCGATCT	GAGTTAGC	GATA	CAGACGTGTGCTCTTCCGATCTGAGTTAGCGATASABVWASCGSTCSGCSGDNAB	NRPAD	707
UF_miSeq_NRPADF09	VTNATGTWBHCSHCSGGSHCS	CTACACGACGCTCTTCCGATC	AAACATCG	A	CTACACGACGCTCTTCCGATCTAAACATCGAVTNATGTWBHCSHCSGGSHCS	UF_miSeq_NRPADR09	SABVWASCGSTCSGCNGDNAB	CAGACGTGTGCTCTTCCGATCT	AGCCATGC	A	CAGACGTGTGCTCTTCCGATCTAGCCATGCASABVWASCGSTCSGCNGDNAB	NRPAD	707
UF_miSeq_NRPCF01	YCGYCACGATATCCTGCGCAC	CTACACGACGCTCTTCCGATC	GCTAACGA	A	CTACACGACGCTCTTCCGATCTGCTAACGAAYCGYCACGATATCCTGCGCAC	UF_miSeq_NRPCR01	CAGCACSGTRCCGAACACCAC	CAGACGTGTGCTCTTCCGATCT	GCTAACGA	A	CAGACGTGTGCTCTTCCGATCTGCTAACGAACAGCACSGTRCCGAACACCAC	NRPC	716
UF_miSeq_NRPCF02	NCGBCAYGAYATCCTGCGCAC	CTACACGACGCTCTTCCGATC	AATCCGTC	TC	CTACACGACGCTCTTCCGATCTAATCCGTCTCNCGBCAYGAYATCCTGCGCAC	UF_miSeq_NRPCR02	CARCACVGTRCCGAAYACCAC	CAGACGTGTGCTCTTCCGATCT	AATCCGTC	TC	CAGACGTGTGCTCTTCCGATCTAATCCGTCTCCARCACVGTRCCGAAYACCAC	NRPC	716
UF_miSeq_NRPCF03	NCGBCAYGAYATYCTGCGCAC	CTACACGACGCTCTTCCGATC	GTCTGTCA	CTA	CTACACGACGCTCTTCCGATCTGTCTGTCACTANCGBCAYGAYATYCTGCGCAC	UF_miSeq_NRPCR03	CARCACVGWVCCGAABACCAC	CAGACGTGTGCTCTTCCGATCT	GTCTGTCA	CTA	CAGACGTGTGCTCTTCCGATCTGTCTGTCACTACARCACVGWVCCGAABACCAC	NRPC	716
UF_miSeq_NRPCF04	NCGBCAYGAYATYYTRCGYAC	CTACACGACGCTCTTCCGATC	GAGTTAGC	GATA	CTACACGACGCTCTTCCGATCTGAGTTAGCGATANCGBCAYGAYATYYTRCGYAC	UF_miSeq_NRPCR04	YARCACVGWVCCRAABACSAC	CAGACGTGTGCTCTTCCGATCT	GAGTTAGC	GATA	CAGACGTGTGCTCTTCCGATCTGAGTTAGCGATAYARCACVGWVCCRAABACSAC	NRPC	716
UF_miSeq_NRPCF05	MGVCAYGAYATHYTVMGVACV	CTACACGACGCTCTTCCGATC	ATTGAGGA	TC	CTACACGACGCTCTTCCGATCTATTGAGGATCMGVCAYGAYATHYTVMGVACV	UF_miSeq_NRPCR05	BARBACBGTBCCRAABACBAC	CAGACGTGTGCTCTTCCGATCT	GAATCTGA	TC	CAGACGTGTGCTCTTCCGATCTGAATCTGATCBARBACBGTBCCRAABACBAC	NRPC	716
UF_miSeq_NRPCF06	MGNCAYGAYATHYTNMGNACN	CTACACGACGCTCTTCCGATC	ATTGGCTC	CTA	CTACACGACGCTCTTCCGATCTATTGGCTCCTAMGNCAYGAYATHYTNMGNACN	UF_miSeq_NRPCR06	NARNACNGTNCCRAANACNAC	CAGACGTGTGCTCTTCCGATCT	CAAGACTA	CTA	CAGACGTGTGCTCTTCCGATCTCAAGACTACTANARNACNGTNCCRAANACNAC	NRPC	716
UF_miSeq_NRPCF07	CGSCAYGACATCCTSCGSACS	CTACACGACGCTCTTCCGATC	AAACATCG	CTA	CTACACGACGCTCTTCCGATCTAAACATCGCTACGSCAYGACATCCTSCGSACS	UF_miSeq_NRPCR07	SAGSACGGTSCCGAASACCAC	CAGACGTGTGCTCTTCCGATCT	GAGTTAGC	CTA	CAGACGTGTGCTCTTCCGATCTGAGTTAGCCTASAGSACGGTSCCGAASACCAC	NRPC	716
UF_miSeq_NRPCF08	MGSCAYGACATCCTSMGSACS	CTACACGACGCTCTTCCGATC	CAGATCTG	GATA	CTACACGACGCTCTTCCGATCTCAGATCTGGATAMGSCAYGACATCCTSMGSACS	UF_miSeq_NRPCR08	SAGSACSGTSCCGAASACSAC	CAGACGTGTGCTCTTCCGATCT	TGAAGAGA	GATA	CAGACGTGTGCTCTTCCGATCTTGAAGAGAGATASAGSACSGTSCCGAASACSAC	NRPC	716
UF_miSeq_NRPCF09	MGSCAYGAYATYYTSMGVACS	CTACACGACGCTCTTCCGATC	GCTCGGTA	A	CTACACGACGCTCTTCCGATCTGCTCGGTAAMGSCAYGAYATYYTSMGVACS	UF_miSeq_NRPCR09	SARSACSGTBCCRAASACBAC	CAGACGTGTGCTCTTCCGATCT	GACTAGTA	A	CAGACGTGTGCTCTTCCGATCTGACTAGTAASARSACSGTBCCRAASACBAC	NRPC	716
UF_miSeq_NRPCF10	MGSCAYGAYATHYTVMGVACS	CTACACGACGCTCTTCCGATC	AGATGTAC	TC	CTACACGACGCTCTTCCGATCTAGATGTACTCMGSCAYGAYATHYTVMGVACS	UF_miSeq_NRPCR10	BARBACBGTBCCRAASACBAC	CAGACGTGTGCTCTTCCGATCT	AGATCGCA	TC	CAGACGTGTGCTCTTCCGATCTAGATCGCATCBARBACBGTBCCRAASACBAC	NRPC	716
UF_miSeq_NRPCF11	CGSMASGASVTSVTSCGSHCR	CTACACGACGCTCTTCCGATC	AAGGACAC	CTA	CTACACGACGCTCTTCCGATCTAAGGACACCTACGSMASGASVTSVTSCGSHCR	UF_miSeq_NRPCR11	BARBACBGTBCCRAASACNAC	CAGACGTGTGCTCTTCCGATCT	CGCTGATC	CTA	CAGACGTGTGCTCTTCCGATCTCGCTGATCCTABARBACBGTBCCRAASACNAC	NRPC	716
UF_miSeq_NRPCF12	CGNMASGASVTSVTSCGSHCS	CTACACGACGCTCTTCCGATC	ATAGCGAC	GATA	CTACACGACGCTCTTCCGATCTATAGCGACGATACGNMASGASVTSVTSCGSHCS	UF_miSeq_NRPCR12	BARBACBGTBCCRAANACNAC	CAGACGTGTGCTCTTCCGATCT	GGAGAACA	GATA	CAGACGTGTGCTCTTCCGATCTGGAGAACAGATABARBACBGTBCCRAANACNAC	NRPC	716
UF_miSeq_NRPCF13	CGNMANGASVTSVTSCGSHCS	CTACACGACGCTCTTCCGATC	TCTTCACA	A	CTACACGACGCTCTTCCGATCTTCTTCACAACGNMANGASVTSVTSCGSHCS	UF_miSeq_NRPCR13	SABSABYGDSCCSWASABSAB	CAGACGTGTGCTCTTCCGATCT	ACGCTCGA	A	CAGACGTGTGCTCTTCCGATCTACGCTCGAASABSABYGDSCCSWASABSAB	NRPC	716
UF_miSeq_PEPF01	GAGGACAAGRTCTTCCCSAAG	CTACACGACGCTCTTCCGATC	TGGTGGTA	GATA	CTACACGACGCTCTTCCGATCTTGGTGGTAGATAGAGGACAAGRTCTTCCCSAAG	UF_miSeq_PEPR01	GCATGCTGTGGTTGGCCCAGA	CAGACGTGTGCTCTTCCGATCT	TGGTGGTA	GATA	CAGACGTGTGCTCTTCCGATCTTGGTGGTAGATAGCATGCTGTGGTTGGCCCAGA	Phosphono	449
UF_miSeq_PEPF02	GARGACAAGVTSTTCCCSAAG	CTACACGACGCTCTTCCGATC	TATCAGCA	A	CTACACGACGCTCTTCCGATCTTATCAGCAAGARGACAAGVTSTTCCCSAAG	UF_miSeq_PEPR02	GCAKSSCSTGGTTSGCGTAGA	CAGACGTGTGCTCTTCCGATCT	TATCAGCA	A	CAGACGTGTGCTCTTCCGATCTTATCAGCAAGCAKSSCSTGGTTSGCGTAGA	Phosphono	449
UF_miSeq_PEPF03	GARGAYAAGVWVTTYCCBAAR	CTACACGACGCTCTTCCGATC	ATCCTGTA	TC	CTACACGACGCTCTTCCGATCTATCCTGTATCGARGAYAAGVWVTTYCCBAAR	UF_miSeq_PEPR03	GCAKSSYSTGGTTSGCSYAGA	CAGACGTGTGCTCTTCCGATCT	ATCCTGTA	TC	CAGACGTGTGCTCTTCCGATCTATCCTGTATCGCAKSSYSTGGTTSGCSYAGA	Phosphono	449

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_PEPF04	GARGAYAARGTSTTYCCSAAR	CTACACGACGCTCTTCCGATCT	CCAGTTCA	GATA	CTACACGACGCTCTTCCGATCTCCAGTTCAGATAGARGAYAARGTSTTYCCSAAR	UF_miSeq_PEPR04	BARSGCRTGRTTSGCRTARAT	CAGACGTGTGCTCTTCCGATCT	GAGCTGAA	GATA	CAGACGTGTGCTCTTCCGATCTGAGCTGAAGATABARSGCRTGRTTSGCRTARAT	Phosphono	449
UF_miSeq_PEPF05	GARGAYAARGTNTTYCCNAAR	CTACACGACGCTCTTCCGATCT	CTAAGGTC	A	CTACACGACGCTCTTCCGATCTCTAAGGTCAGARGAYAARGTNTTYCCNAAR	UF_miSeq_PEPR05	NARNGCRTGRTTNGCRTADAT	CAGACGTGTGCTCTTCCGATCT	CTGAGCCA	A	CAGACGTGTGCTCTTCCGATCTCTGAGCCAANARNGCRTGRTTNGCRTADAT	Phosphono	449
UF_miSeq_PEPF06	GAGGACAAGSWSTTCCCSAAG	CTACACGACGCTCTTCCGATCT	CACTTCGA	CTA	CTACACGACGCTCTTCCGATCTCACTTCGACTAGAGGACAAGSWSTTCCCSAAG	UF_miSeq_PEPR06	SAGSGCGTGGTTSGCGTARAT	CAGACGTGTGCTCTTCCGATCT	AAGACGGA	CTA	CAGACGTGTGCTCTTCCGATCTAAGACGGACTASAGSGCGTGGTTSGCGTARAT	Phosphono	449
UF_miSeq_PEPF07	GARGAYAAGSWVTTYCCSAAR	CTACACGACGCTCTTCCGATCT	GAGTTAGC	GATA	CTACACGACGCTCTTCCGATCTGAGTTAGCGATAGARGAYAAGSWVTTYCCSAAR	UF_miSeq_PEPR07	SAGSGCRTKRTKSGCRTASAT	CAGACGTGTGCTCTTCCGATCT	ACGCTCGA	GATA	CAGACGTGTGCTCTTCCGATCTACGCTCGAGATASAGSGCRTKRTKSGCRTASAT	Phosphono	449
UF_miSeq_PEPF08	GASGASMASVTSTWSCCSMAS	CTACACGACGCTCTTCCGATCT	GCTAACGA	A	CTACACGACGCTCTTCCGATCTGCTAACGAAGASGASMASVTSTWSCCSMAS	UF_miSeq_PEPR08	SABSGCSTKSTKSGCSTASAT	CAGACGTGTGCTCTTCCGATCT	AGCAGGAA	A	CAGACGTGTGCTCTTCCGATCTAGCAGGAAASABSGCSTKSTKSGCSTASAT	Phosphono	449
UF_miSeq_PEPF09	GANGASMASVTSTWBCCSMAS	CTACACGACGCTCTTCCGATCT	GATAGACA	TC	CTACACGACGCTCTTCCGATCTGATAGACATCGANGASMASVTSTWBCCSMAS	UF_miSeq_PEPR09	SABSGCSTKSTKSGCSWASAB	CAGACGTGTGCTCTTCCGATCT	TGAAGAGA	TC	CAGACGTGTGCTCTTCCGATCTTGAAGAGATCSABSGCSTKSTKSGCSWASAB	Phosphono	449
UF_miSeq_PEPF10	GANGANMASVTSTWBCCSMAS	CTACACGACGCTCTTCCGATCT	ACAGCAGA	CTA	CTACACGACGCTCTTCCGATCTACAGCAGACTAGANGANMASVTSTWBCCSMAS	UF_miSeq_PEPR10	SABSGCSTKSTKSGCVWANAB	CAGACGTGTGCTCTTCCGATCT	CCATCCTC	CTA	CAGACGTGTGCTCTTCCGATCTCCATCCTCCTASABSGCSTKSTKSGCVWANAB	Phosphono	449
UF_miSeq_PEPF11	GANGANMANVTSTWBCCSMAS	CTACACGACGCTCTTCCGATCT	ACAAGCTA	GATA	CTACACGACGCTCTTCCGATCTACAAGCTAGATAGANGANMANVTSTWBCCSMAS	UF_miSeq_PEPR11	SABSGCSTKSTKNGCVWANAB	CAGACGTGTGCTCTTCCGATCT	CCGACAAC	GATA	CAGACGTGTGCTCTTCCGATCTCCGACAACGATASABSGCSTKSTKNGCVWANAB	Phosphono	449
UF_miSeq_PIPF01	CTGCCGACSATCATCGGCACS	CTACACGACGCTCTTCCGATCT	GCTCGGTA	A	CTACACGACGCTCTTCCGATCTGCTCGGTAACTGCCGACSATCATCGGCACS	UF_miSeq_PIPR01	TSGCSGTGCAGATGATGTCGG	CAGACGTGTGCTCTTCCGATCT	GCTCGGTA	A	CAGACGTGTGCTCTTCCGATCTGCTCGGTAATSGCSGTGCAGATGATGTCGG	PipecolicAcid	
UF_miSeq_PIPF02	CTSCCSACSATCRTCGGVACS	CTACACGACGCTCTTCCGATCT	CAATGGAA	TC	CTACACGACGCTCTTCCGATCTCAATGGAATCCTSCCSACSATCRTCGGVACS	UF_miSeq_PIPR02	TSRCSGTGCASAKSAYGTCGG	CAGACGTGTGCTCTTCCGATCT	CAATGGAA	TC	CAGACGTGTGCTCTTCCGATCTCAATGGAATCTSRCSGTGCASAKSAYGTCGG	PipecolicAcid	
UF_miSeq_PIPF03	CTVCCVACVATVATHGGCACV	CTACACGACGCTCTTCCGATCT	CCTCCTGA	TC	CTACACGACGCTCTTCCGATCTCCTCCTGATCCTVCCVACVATVATHGGCACV	UF_miSeq_PIPR03	SGTSGCBGTRCADATDATRTC	CAGACGTGTGCTCTTCCGATCT	CTGGCATA	TC	CAGACGTGTGCTCTTCCGATCTCTGGCATATCSGTSGCBGTRCADATDATRTC	PipecolicAcid	
UF_miSeq_PIPF04	CTVCCVACNATVATHGGCACN	CTACACGACGCTCTTCCGATCT	CGACTGGA	CTA	CTACACGACGCTCTTCCGATCTCGACTGGACTACTVCCVACNATVATHGGCACN	UF_miSeq_PIPR04	BGTBGCBGTRCADATDATRTC	CAGACGTGTGCTCTTCCGATCT	GAATCTGA	CTA	CAGACGTGTGCTCTTCCGATCTGAATCTGACTABGTBGCBGTRCADATDATRTC	PipecolicAcid	
UF_miSeq_PIPF05	YTNCCNACNATHATHGGNACN	CTACACGACGCTCTTCCGATCT	CAAGGAGC	GATA	CTACACGACGCTCTTCCGATCTCAAGGAGCGATAYTNCCNACNATHATHGGNACN	UF_miSeq_PIPR05	NGTNGCNGTRCADATDATRTC	CAGACGTGTGCTCTTCCGATCT	CAAGACTA	GATA	CAGACGTGTGCTCTTCCGATCTCAAGACTAGATANGTNGCNGTRCADATDATRTC	PipecolicAcid	
UF_miSeq_PIPF06	CTGCCSACSATSATCGGCACS	CTACACGACGCTCTTCCGATCT	ACGCTCGA	A	CTACACGACGCTCTTCCGATCTACGCTCGAACTGCCSACSATSATCGGCACS	UF_miSeq_PIPR06	SGTSGCSGTRCAGATGATGTC	CAGACGTGTGCTCTTCCGATCT	CCATCCTC	A	CAGACGTGTGCTCTTCCGATCTCCATCCTCASGTSGCSGTRCAGATGATGTC	PipecolicAcid	
UF_miSeq_PIPF07	CTSCCSACSATVATCGGCACS	CTACACGACGCTCTTCCGATCT	CGAACTTA	CTA	CTACACGACGCTCTTCCGATCTCGAACTTACTACTSCCSACSATVATCGGCACS	UF_miSeq_PIPR07	SGTSGCSGTRCARATRATGTC	CAGACGTGTGCTCTTCCGATCT	ATCATTCC	CTA	CAGACGTGTGCTCTTCCGATCTATCATTCCCTASGTSGCSGTRCARATRATGTC	PipecolicAcid	
UF_miSeq_PIPF08	CTVCCSACSATVATHGGCACS	CTACACGACGCTCTTCCGATCT	AGATGTAC	GATA	CTACACGACGCTCTTCCGATCTAGATGTACGATACTVCCSACSATVATHGGCACS	UF_miSeq_PIPR08	BGTSGCSGTRCADATDATRTC	CAGACGTGTGCTCTTCCGATCT	GTGTTCTA	GATA	CAGACGTGTGCTCTTCCGATCTGTGTTCTAGATABGTSGCSGTRCADATDATRTC	PipecolicAcid	
UF_miSeq_PIPF09	YTSCCVACVATHATHGGVACV	CTACACGACGCTCTTCCGATCT	TCTTCACA	A	CTACACGACGCTCTTCCGATCTTCTTCACAAYTSCCVACVATHATHGGVACV	UF_miSeq_PIPR09	BGTNGCSGTRCADATDATRTC	CAGACGTGTGCTCTTCCGATCT	GACTAGTA	A	CAGACGTGTGCTCTTCCGATCTGACTAGTAABGTNGCSGTRCADATDATRTC	PipecolicAcid	
UF_miSeq_PIPF10	YTNCCVACVATHATHGGVACV	CTACACGACGCTCTTCCGATCT	ACTATGCA	TC	CTACACGACGCTCTTCCGATCTACTATGCATCYTNCCVACVATHATHGGVACV	UF_miSeq_PIPR10	YGDSGCYGDRCASABSABSTC	CAGACGTGTGCTCTTCCGATCT	CGAACTTA	TC	CAGACGTGTGCTCTTCCGATCTCGAACTTATCYGDSGCYGDRCASABSABSTC	PipecolicAcid	
UF_miSeq_PIPF11	VTSCCSHCRVTSVTSGGSHCR	CTACACGACGCTCTTCCGATCT	GACAGTGC	CTA	CTACACGACGCTCTTCCGATCTGACAGTGCCTAVTSCCSHCRVTSVTSGGSHCR	UF_miSeq_PIPR11	SGDSGCSGDRCASABSABNTC	CAGACGTGTGCTCTTCCGATCT	AGATGTAC	CTA	CAGACGTGTGCTCTTCCGATCTAGATGTACCTASGDSGCSGDRCASABSABNTC	PipecolicAcid	
UF_miSeq_PIPF12	VTNCCSHCSVTSVTSGGSHCS	CTACACGACGCTCTTCCGATCT	CTCAATGA	GATA	CTACACGACGCTCTTCCGATCTCTCAATGAGATAVTNCCSHCSVTSVTSGGSHCS	UF_miSeq_PIPR12	SGDSGCSGDRCASABNABNTC	CAGACGTGTGCTCTTCCGATCT	AATGTTGC	GATA	CAGACGTGTGCTCTTCCGATCTAATGTTGCGATASGDSGCSGDRCASABNABNTC	PipecolicAcid	
UF_miSeq_PKSATF01	GCCGGGCACTCSATCGGTGAG	CTACACGACGCTCTTCCGATCT	CCTCTATC	CTA	CTACACGACGCTCTTCCGATCTCCTCTATCCTAGCCGGGCACTCSATCGGTGAG	UF_miSeq_PKSATR01	CGCAGGTTSCGGTACCAGTAK	CAGACGTGTGCTCTTCCGATCT	CCTCTATC	CTA	CAGACGTGTGCTCTTCCGATCTCCTCTATCCTACGCAGGTTSCGGTACCAGTAK	PKSAT	537
UF_miSeq_PKSATF02	GTSGGSCAYTCSCAGGGBGAG	CTACACGACGCTCTTCCGATCT	CATACCAA	GATA	CTACACGACGCTCTTCCGATCTCATACCAAGATAGTSGGSCAYTCSCAGGGBGAG	UF_miSeq_PKSATR02	CGCAGGTTSCKGWACCAGTAK	CAGACGTGTGCTCTTCCGATCT	CATACCAA	GATA	CAGACGTGTGCTCTTCCGATCTCATACCAAGATACGCAGGTTSCKGWACCAGTAK	PKSAT	537
UF_miSeq_PKSATF03	GYSGGSCAYTCSVWSGGBGAG	CTACACGACGCTCTTCCGATCT	ACACGACC	A	CTACACGACGCTCTTCCGATCTACACGACCAGYSGGSCAYTCSVWSGGBGAG	UF_miSeq_PKSATR03	CGCASRTKSCKGWMCCAGTAB	CAGACGTGTGCTCTTCCGATCT	ACACGACC	A	CAGACGTGTGCTCTTCCGATCTACACGACCACGCASRTKSCKGWMCCAGTAB	PKSAT	537
UF_miSeq_PKSATF04	GTSGGSCAYAGYCARGGSGAR	CTACACGACGCTCTTCCGATCT	TGGCTTCA	A	CTACACGACGCTCTTCCGATCTTGGCTTCAAGTSGGSCAYAGYCARGGSGAR	UF_miSeq_PKSATR04	SCKSARRTTSCKRTACCARTA	CAGACGTGTGCTCTTCCGATCT	GAGCTGAA	A	CAGACGTGTGCTCTTCCGATCTGAGCTGAAASCKSARRTTSCKRTACCARTA	PKSAT	537
UF_miSeq_PKSATF05	GTNGGNCAYAGYCARGGNGAR	CTACACGACGCTCTTCCGATCT	CCATCCTC	TC	CTACACGACGCTCTTCCGATCTCCATCCTCTCGTNGGNCAYAGYCARGGNGAR	UF_miSeq_PKSATR05	NCKNARRTTNCKRTACCARTA	CAGACGTGTGCTCTTCCGATCT	GATAGACA	TC	CAGACGTGTGCTCTTCCGATCTGATAGACATCNCKNARRTTNCKRTACCARTA	PKSAT	537
UF_miSeq_PKSATF06	GTSGGSCACAGCCARGGVGAG	CTACACGACGCTCTTCCGATCT	ACCACTGT	A	CTACACGACGCTCTTCCGATCTACCACTGTAGTSGGSCACAGCCARGGVGAG	UF_miSeq_PKSATR06	SCGSARGTTSCGGTACCARTA	CAGACGTGTGCTCTTCCGATCT	AGCAGGAA	A	CAGACGTGTGCTCTTCCGATCTAGCAGGAAASCGSARGTTSCGGTACCARTA	PKSAT	537
UF_miSeq_PKSATF07	GTVGGVCAYAGYCARGGVGAG	CTACACGACGCTCTTCCGATCT	ACAAGCTA	CTA	CTACACGACGCTCTTCCGATCTACAAGCTACTAGTVGGVCAYAGYCARGGVGAG	UF_miSeq_PKSATR07	SCGSASSTTSCGSTASWASWA	CAGACGTGTGCTCTTCCGATCT	ACGTATCA	CTA	CAGACGTGTGCTCTTCCGATCTACGTATCACTASCGSASSTTSCGSTASWASWA	PKSAT	537
UF_miSeq_PKSATF08	STSGGSCASHCRMASVTSCGH	CTACACGACGCTCTTCCGATCT	GTCTGTCA	GATA	CTACACGACGCTCTTCCGATCTGTCTGTCAGATASTSGGSCASHCRMASVTSCGH	UF_miSeq_PKSATR08	SCGSABSTKSCGSWASWAVWA	CAGACGTGTGCTCTTCCGATCT	TGGCTTCA	GATA	CAGACGTGTGCTCTTCCGATCTTGGCTTCAGATASCGSABSTKSCGSWASWAVWA	PKSAT	537
UF_miSeq_PKSATF09	VTSGGSMASHCRMASVTSCGH	CTACACGACGCTCTTCCGATCT	CCGACAAC	A	CTACACGACGCTCTTCCGATCTCCGACAACAVTSGGSMASHCRMASVTSCGH	UF_miSeq_PKSATR09	SCGSABSTKSCGVWAVWAVWA	CAGACGTGTGCTCTTCCGATCT	CGACTGGA	A	CAGACGTGTGCTCTTCCGATCTCGACTGGAASCGSABSTKSCGVWAVWAVWA	PKSAT	537

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	Spacer F	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	Barcode R	Spacer R	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_PKSATF10	VTNGGSMASHCSMASVTSCGH	CTACACGACGCTCTTCCGATCT	CAACCACA	TC	CTACACGACGCTCTTCCGATCTCAACCACATCVTNGGSMASHCSMASVTSCGH	UF_miSeq_PKSATR10	SCGSABSTKNCGVWAVWAVWA	CAGACGTGTGCTCTTCCGATCT	AAGAGATC	TC	CAGACGTGTGCTCTTCCGATCTAAGAGATCTCSCGSABSTKNCGVWAVWAVWA	PKSAT	537
UF_miSeq_PKSKSF01	TTCGACGCCGRRTTCTTCGGC	CTACACGACGCTCTTCCGATCT	TTCACGCA	A	CTACACGACGCTCTTCCGATCTTTTCACGCAATTCGACGCCGRRTTCTTCGGC	UF_miSeq_PKSKSR01	TKCCGGTGCCGTGCAGYTCGA	CAGACGTGTGCTCTTCCGATCT	TTCACGCA	A	CAGACGTGTGCTCTTCCGATCTTTTCACGCAATKCCGGTGCCGTGCAGYTCGA	PKSKS	723
UF_miSeq_PKSKSF02	TTCGACGCSGSMTTCTTCGRC	CTACACGACGCTCTTCCGATCT	CGCATACA	TC	CTACACGACGCTCTTCCGATCTCGCATACATCTTCGACGCSGSMTTCTTCGRC	UF_miSeq_PKSKSR02	TSCCGGTGCCGTGCSCCTCSA	CAGACGTGTGCTCTTCCGATCT	CGCATACA	TC	CAGACGTGTGCTCTTCCGATCTCGCATACATCTSCCGGTGCCGTGCSCCTCSA	PKSKS	723
UF_miSeq_PKSKSF03	TTCGACGCBGVSTTCTTCGGS	CTACACGACGCTCTTCCGATCT	ATCATTCC	CTA	CTACACGACGCTCTTCCGATCTATCATTCCCTATTCGACGCBGVSTTCTTCGGS	UF_miSeq_PKSKSR03	TVCCGGTGCCGTGSSYCTCSA	CAGACGTGTGCTCTTCCGATCT	ATCATTCC	CTA	CAGACGTGTGCTCTTCCGATCTATCATTCCCTATVCCGGTGCCGTGSSYCTCSA	PKSKS	723
UF_miSeq_PKSKSF04	TTCGACGCBGVSTTCTTCGVS	CTACACGACGCTCTTCCGATCT	CTCAATGA	GATA	CTACACGACGCTCTTCCGATCTCTCAATGAGATATTCGACGCBGVSTTCTTCGVS	UF_miSeq_PKSKSR04	TNCCGGTGCCGTGCVSYTCSA	CAGACGTGTGCTCTTCCGATCT	CTCAATGA	GATA	CAGACGTGTGCTCTTCCGATCTCTCAATGAGATATNCCGGTGCCGTGCVSYTCSA	PKSKS	723
UF_miSeq_PKSKSF05	TTCGACGCBKVVTTCTTCGVS	CTACACGACGCTCTTCCGATCT	AATGTTGC	A	CTACACGACGCTCTTCCGATCTAATGTTGCATTCGACGCBKVVTTCTTCGVS	UF_miSeq_PKSKSR05	TNCCSGTRCCGTGSVSYTCSA	CAGACGTGTGCTCTTCCGATCT	AATGTTGC	A	CAGACGTGTGCTCTTCCGATCTAATGTTGCATNCCSGTRCCGTGSVSYTCSA	PKSKS	723
UF_miSeq_PKSKSF06	TTCGASSCBKVVTTCTTCGVS	CTACACGACGCTCTTCCGATCT	ATAGCGAC	TC	CTACACGACGCTCTTCCGATCTATAGCGACTCTTCGASSCBKVVTTCTTCGVS	UF_miSeq_PKSKSR06	TNCCSGTRCCGTGSVBYTCSA	CAGACGTGTGCTCTTCCGATCT	ATAGCGAC	TC	CAGACGTGTGCTCTTCCGATCTATAGCGACTCTNCCSGTRCCGTGSVBYTCSA	PKSKS	723
UF_miSeq_PKSKSF07	TTYGAYGCNGARTTYTTYGGN	CTACACGACGCTCTTCCGATCT	GTACGCAA	CTA	CTACACGACGCTCTTCCGATCTGTACGCAACTATTYGAYGCNGARTTYTTYGGN	UF_miSeq_PKSKSR07	NCCNGTNCCRTGNARYTCNAC	CAGACGTGTGCTCTTCCGATCT	GCCACATA	CTA	CAGACGTGTGCTCTTCCGATCTGCCACATACTANCCNGTNCCRTGNARYTCNAC	PKSKS	723
UF_miSeq_PKSKSF08	TWSGASGCSGASTWSTWSGGS	CTACACGACGCTCTTCCGATCT	AATGTTGC	CTA	CTACACGACGCTCTTCCGATCTAATGTTGCCTATWSGASGCSGASTWSTWSGGS	UF_miSeq_PKSKSR08	SCCYGDSCCSTKSASSTCSAS	CAGACGTGTGCTCTTCCGATCT	CCGAAGTA	CTA	CAGACGTGTGCTCTTCCGATCTCCGAAGTACTASCCYGDSCCSTKSASSTCSAS	PKSKS	723
UF_miSeq_PKSKSF09	TWBGASGCSGASTWBTWBGGS	CTACACGACGCTCTTCCGATCT	AACCGAGA	GATA	CTACACGACGCTCTTCCGATCTAACCGAGAGATATWBGASGCSGASTWBTWBGGS	UF_miSeq_PKSKSR09	SCCYGDSCCSTKSABSTCSAB	CAGACGTGTGCTCTTCCGATCT	ACAGCAGA	GATA	CAGACGTGTGCTCTTCCGATCTACAGCAGAGATASCCYGDSCCSTKSABSTCSAB	PKSKS	723
UF_miSeq_PKSKSF10	TWBGANGCSGASTWBTWBGGS	CTACACGACGCTCTTCCGATCT	CTCAATGA	A	CTACACGACGCTCTTCCGATCTCTCAATGAATWBGANGCSGASTWBTWBGGS	UF_miSeq_PKSKSR10	SCCSGDSCCSTKSABSTCNAB	CAGACGTGTGCTCTTCCGATCT	GTCTGTCA	A	CAGACGTGTGCTCTTCCGATCTGTCTGTCAASCCSGDSCCSTKSABSTCNAB	PKSKS	723
UF_miSeq_PKSKSF11	TWBGANGCNGASTWBTWBGGS	CTACACGACGCTCTTCCGATCT	GACAGTGC	CTA	CTACACGACGCTCTTCCGATCTGACAGTGCCTATWBGANGCNGASTWBTWBGGS	UF_miSeq_PKSKSR11	SCCSGDSCCSTKSABNTCNAB	CAGACGTGTGCTCTTCCGATCT	ACAAGCTA	CTA	CAGACGTGTGCTCTTCCGATCTACAAGCTACTASCCSGDSCCSTKSABNTCNAB	PKSKS	723
UF_miSeq_PKSKSF12	TWBGANGCNGANTWBTWBGGS	CTACACGACGCTCTTCCGATCT	AGTGGTCA	GATA	CTACACGACGCTCTTCCGATCTAGTGGTCAGATATWBGANGCNGANTWBTWBGGS	UF_miSeq_PKSKSR12	SCCSGDSCCSTKNABNTCNAB	CAGACGTGTGCTCTTCCGATCT	GTACGCAA	GATA	CAGACGTGTGCTCTTCCGATCTGTACGCAAGATASCCSGDSCCSTKNABNTCNAB	PKSKS	723
UF_miSeq_T2PKSF01	TSGCCTGCTTCGACGCCATCM	CTACACGACGCTCTTCCGATCT	CCGTGAGA	TC	CTACACGACGCTCTTCCGATCTCCGTGAGATCTSGCCTGCTTCGACGCCATCM	UF_miSeq_T2PKSR01	TAGTCSAGRTCGCACTCGGGG	CAGACGTGTGCTCTTCCGATCT	CCGTGAGA	TC	CAGACGTGTGCTCTTCCGATCTCCGTGAGATCTAGTCSAGRTCGCACTCGGGG	T2PKS	540
UF_miSeq_T2PKSF02	TSGCCTGYTTCGACGCSATCM	CTACACGACGCTCTTCCGATCT	CCTAATCC	CTA	CTACACGACGCTCTTCCGATCTCCTAATCCCTATSGCCTGYTTCGACGCSATCM	UF_miSeq_T2PKSR02	TAGTCSAGRTCGCASTCSGGG	CAGACGTGTGCTCTTCCGATCT	CCTAATCC	CTA	CAGACGTGTGCTCTTCCGATCTCCTAATCCCTATAGTCSAGRTCGCASTCSGGG	T2PKS	540
UF_miSeq_T2PKSF03	TSGCSTGYTTCGAYGCSATCM	CTACACGACGCTCTTCCGATCT	AACTCACC	GATA	CTACACGACGCTCTTCCGATCTAACTCACCGATATSGCSTGYTTCGAYGCSATCM	UF_miSeq_T2PKSR03	TAGTCSAGRTCGCABKCSGGR	CAGACGTGTGCTCTTCCGATCT	AACTCACC	GATA	CAGACGTGTGCTCTTCCGATCTAACTCACCGATATAGTCSAGRTCGCABKCSGGR	T2PKS	540
UF_miSeq_T2PKSF04	TSKCSTGYTTCGAYRCSMTSM	CTACACGACGCTCTTCCGATCT	AAGGACAC	A	CTACACGACGCTCTTCCGATCTAAGGACACATSKCSTGYTTCGAYRCSMTSM	UF_miSeq_T2PKSR04	TAGTCSAGRTCGMRBKSSGGR	CAGACGTGTGCTCTTCCGATCT	AAGGACAC	A	CAGACGTGTGCTCTTCCGATCTAAGGACACATAGTCSAGRTCGMRBKSSGGR	T2PKS	540
UF_miSeq_T2PKSF05	GTNGCNTGYTTYGAYGCNATH	CTACACGACGCTCTTCCGATCT	GTGTTCTA	GATA	CTACACGACGCTCTTCCGATCTGTGTTCTAGATAGTNGCNTGYTTYGAYGCNATH	UF_miSeq_T2PKSR05	RTARTCNARRTCNARYTGNGG	CAGACGTGTGCTCTTCCGATCT	GCGAGTAA	GATA	CAGACGTGTGCTCTTCCGATCTGCGAGTAAGATARTARTCNARRTCNARYTGNGG	T2PKS	540
UF_miSeq_T2PKSF06	GTVGCSTGYTTYGAYGCSATY	CTACACGACGCTCTTCCGATCT	CAAGACTA	A	CTACACGACGCTCTTCCGATCTCAAGACTAAGTVGCSTGYTTYGAYGCSATY	UF_miSeq_T2PKSR06	GTARTCSARRTCSARYTGSGG	CAGACGTGTGCTCTTCCGATCT	CGCATACA	A	CAGACGTGTGCTCTTCCGATCTCGCATACAAGTARTCSARRTCSARYTGSGG	T2PKS	540
UF_miSeq_T2PKSF07	GTSGCVTGYTTYGAYGCVATH	CTACACGACGCTCTTCCGATCT	AACCGAGA	TC	CTACACGACGCTCTTCCGATCTAACCGAGATCGTSGCVTGYTTYGAYGCVATH	UF_miSeq_T2PKSR07	RTARTCBARRTCBARYTGSGG	CAGACGTGTGCTCTTCCGATCT	ACCTCCAA	TC	CAGACGTGTGCTCTTCCGATCTACCTCCAATCRTARTCBARRTCBARYTGSGG	T2PKS	540
UF_miSeq_T2PKSF08	VTSGCSTGYTWSGASGCSVTS	CTACACGACGCTCTTCCGATCT	TGGTGGTA	CTA	CTACACGACGCTCTTCCGATCTTGGTGGTACTAVTSGCSTGYTWSGASGCSVTS	UF_miSeq_T2PKSR08	SWASTCSABSTCSABSTKSGG	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	CTA	CAGACGTGTGCTCTTCCGATCTTAGGATGACTASWASTCSABSTCSABSTKSGG	T2PKS	540
UF_miSeq_T2PKSF09	VTNGCSTGYTWBGASGCSVTS	CTACACGACGCTCTTCCGATCT	CAGCGTTA	GATA	CTACACGACGCTCTTCCGATCTCAGCGTTAGATAVTNGCSTGYTWBGASGCSVTS	UF_miSeq_T2PKSR09	VWASTCSABSTCSABSTKNGG	CAGACGTGTGCTCTTCCGATCT	CCGACAAC	GATA	CAGACGTGTGCTCTTCCGATCTCCGACAACGATAVWASTCSABSTCSABSTKNGG	T2PKS	540
UF_miSeq_T2PKSF10	VTNGCNTGYTWBGASGCSVTS	CTACACGACGCTCTTCCGATCT	ACGCTCGA	A	CTACACGACGCTCTTCCGATCTACGCTCGAAVTNGCNTGYTWBGASGCSVTS	UF_miSeq_T2PKSR10	VWASTCSABSTCSABNTKNGG	CAGACGTGTGCTCTTCCGATCT	ATTGGCTC	A	CAGACGTGTGCTCTTCCGATCTATTGGCTCAVWASTCSABSTCSABNTKNGG	T2PKS	540
UF_miSeq_T2PKSF11	VTNGCNTGYTWBGANGCSVTS	CTACACGACGCTCTTCCGATCT	AACTCACC	CTA	CTACACGACGCTCTTCCGATCTAACTCACCCTAVTNGCNTGYTWBGANGCSVTS	UF_miSeq_T2PKSR11	VWASTCSABSTCNABNTKNGG	CAGACGTGTGCTCTTCCGATCT	CAATGGAA	CTA	CAGACGTGTGCTCTTCCGATCTCAATGGAACTAVWASTCSABSTCNABNTKNGG	T2PKS	540
UF_miSeq_TDF01	STGCGGGTGCTGCCSGACGAC	CTACACGACGCTCTTCCGATCT	AACGTGAT	A	CTACACGACGCTCTTCCGATCTAACGTGATASTGCGGGTGCTGCCSGACGAC	UF_miSeq_TDR01	SGCGTASAGGTACTGCAGC	CAGACGTGTGCTCTTCCGATCT	AACGTGAT	A	CAGACGTGTGCTCTTCCGATCTAACGTGATASGCGTASAGGTACTGCAGC	TryptophanDimer	487
UF_miSeq_TDF02	STGCGGGTGCTGCCSGACSAC	CTACACGACGCTCTTCCGATCT	AAACATCG	TC	CTACACGACGCTCTTCCGATCTAAACATCGTCSTGCGGGTGCTGCCSGACSAC	UF_miSeq_TDR02	SGCGTASAGGTACTGVAGC	CAGACGTGTGCTCTTCCGATCT	AAACATCG	TC	CAGACGTGTGCTCTTCCGATCTAAACATCGTCSGCGTASAGGTACTGVAGC	TryptophanDimer	487
UF_miSeq_TDF03	STSCGGGTSCTGCCSGACSAC	CTACACGACGCTCTTCCGATCT	ATGCCTAA	CTA	CTACACGACGCTCTTCCGATCTATGCCTAACTASTSCGGGTSCTGCCSGACSAC	UF_miSeq_TDR03	SGCGTASAGRTACTGVAGC	CAGACGTGTGCTCTTCCGATCT	ATGCCTAA	CTA	CAGACGTGTGCTCTTCCGATCTATGCCTAACTASGCGTASAGRTACTGVAKC	TryptophanDimer	487
UF_miSeq_TDF04	GTSMGVGTSCTGCCSGACSAC	CTACACGACGCTCTTCCGATCT	ACATTGGC	GATA	CTACACGACGCTCTTCCGATCTACATTGGCGATAGTSMGVGTSCTGCCSGACSAC	UF_miSeq_TDR04	SGCGTASRKRTACTGVAGC	CAGACGTGTGCTCTTCCGATCT	ACATTGGC	GATA	CAGACGTGTGCTCTTCCGATCTACATTGGCGATASGCGTASRKRTACTGVAGC	TryptophanDimer	487
UF_miSeq_TDF05	STSMGVGTSCTSCCSGACSAC	CTACACGACGCTCTTCCGATCT	CAGATCTG	A	CTACACGACGCTCTTCCGATCTCAGATCTGASTSMGVGTSCTSCCSGACSAC	UF_miSeq_TDR05	SGCGTASRKRTACTGVAGC	CAGACGTGTGCTCTTCCGATCT	CAGATCTG	A	CAGACGTGTGCTCTTCCGATCTCAGATCTGASGCGTASRKRTACTGVAKC	TryptophanDimer	487

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_TDF06	STNMGVGTSCTNCCSGAYSAY	CTACACGACGCTCTTCCGATCT	GATAGACA	CTA	CTACACGACGCTCTTCCGATCTGATAGACACTASTNMGVGTSCTNCCSGAYSAY	UF_miSeq_TDR06	SGCGTAVRKRTAYTGVAKCAT	CAGACGTGTGCTCTTCCGATCT	GATAGACA	CTA	CAGACGTGTGCTCTTCCGATCTGATAGACACTASGCGTAVRKRTAYTGVAKCAT	TryptophanDimer	487
UF_miSeq_TDF07	GTNMGNGTNYTNCCNGAYCAY	CTACACGACGCTCTTCCGATCT	GCCACATA	GATA	CTACACGACGCTCTTCCGATCTGCCACATAGATAGTNMGNGTNYTNCCNGAYCAY	UF_miSeq_TDR07	NGCRTANARRTAYTGNARCAT	CAGACGTGTGCTCTTCCGATCT	GCCACATA	GATA	CAGACGTGTGCTCTTCCGATCTGCCACATAGATANGCRTANARRTAYTGNARCAT	TryptophanDimer	487
UF_miSeq_TDF08	STNMGNGTNYTNCCNGAYSAY	CTACACGACGCTCTTCCGATCT	GCGAGTAA	A	CTACACGACGCTCTTCCGATCTGCGAGTAAASTNMGNGTNYTNCCNGAYSAY	UF_miSeq_TDR08	NGCRTANRDRTAYTGNADCAT	CAGACGTGTGCTCTTCCGATCT	GCGAGTAA	A	CAGACGTGTGCTCTTCCGATCTGCGAGTAAANGCRTANRDRTAYTGNADCAT	TryptophanDimer	487
UF_miSeq_TDF09	STVMGSGTSCTSCCSGAYSAY	CTACACGACGCTCTTCCGATCT	ATTGAGGA	A	CTACACGACGCTCTTCCGATCTATTGAGGAASTVMGSGTSCTSCCSGAYSAY	UF_miSeq_TDR09	BGCRTASRRRTAYTGBARCAT	CAGACGTGTGCTCTTCCGATCT	ACACAGAA	A	CAGACGTGTGCTCTTCCGATCTACACAGAAABGCRTASRRRTAYTGBARCAT	TryptophanDimer	487
UF_miSeq_TDF10	STVMGVGTSCTVCCSGAYSAY	CTACACGACGCTCTTCCGATCT	GACTAGTA	CTA	CTACACGACGCTCTTCCGATCTGACTAGTACTASTVMGVGTSCTVCCSGAYSAY	UF_miSeq_TDR10	SGCSWASABSWASTKSABCAT	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	CTA	CAGACGTGTGCTCTTCCGATCTTAGGATGACTASGCSWASABSWASTKSABCAT	TryptophanDimer	487
UF_miSeq_TDF11	VTSCGSVTSVTSCCSGASMAS	CTACACGACGCTCTTCCGATCT	ACCTCCAA	GATA	CTACACGACGCTCTTCCGATCTACCTCCAAGATAVTSCGSVTSVTSCCSGASMAS	UF_miSeq_TDR11	SGCVWASABVWASTKSABCAT	CAGACGTGTGCTCTTCCGATCT	TCCGTCTA	GATA	CAGACGTGTGCTCTTCCGATCTTCCGTCTAGATASGCVWASABVWASTKSABCAT	TryptophanDimer	487
UF_miSeq_TDF12	VTNCGSVTSVTSCCSGASMAS	CTACACGACGCTCTTCCGATCT	CGACACAC	A	CTACACGACGCTCTTCCGATCTCGACACACAVTNCGSVTSVTSCCSGASMAS	UF_miSeq_TDR12	SGCVWASABVWASTKNABCAT	CAGACGTGTGCTCTTCCGATCT	AAGGACAC	A	CAGACGTGTGCTCTTCCGATCTAAGGACACASGCVWASABVWASTKNABCAT	TryptophanDimer	487
UF_miSeq_TDF13	VTNCGNVTSVTSCCSGASMAS	CTACACGACGCTCTTCCGATCT	GTCGTAGA	TC	CTACACGACGCTCTTCCGATCTGTCGTAGATCVTNCGNVTSVTSCCSGASMAS	UF_miSeq_TDR13	SGCVWASABVWANTKNABCAT	CAGACGTGTGCTCTTCCGATCT	ACATTGGC	TC	CAGACGTGTGCTCTTCCGATCTACATTGGCTCSGCVWASABVWANTKNABCAT	TryptophanDimer	487
UF_miSeq_TeF01	TCTTCGGSCACAGCATGGGSG	CTACACGACGCTCTTCCGATCT	GGAGAACA	A	CTACACGACGCTCTTCCGATCTGGAGAACAATCTTCGGSCACAGCATGGGSG	UF_miSeq_TeR01	AGTCGSCSCGCAGCGCGGGCA	CAGACGTGTGCTCTTCCGATCT	GGAGAACA	A	CAGACGTGTGCTCTTCCGATCTGGAGAACAAAGTCGSCSCGCAGCGCGGGCA	Thioesterase	
UF_miSeq_TeF02	TSTTCGGGCACAGCMTSGGSG	CTACACGACGCTCTTCCGATCT	CACCTTAC	TC	CTACACGACGCTCTTCCGATCTCACCTTACTCTSTTCGGGCACAGCMTSGGSG	UF_miSeq_TeR02	AGTCGSCSCGCAGCGSSGGCA	CAGACGTGTGCTCTTCCGATCT	CACCTTAC	TC	CAGACGTGTGCTCTTCCGATCTCACCTTACTCAGTCGSCSCGCAGCGSSGGCA	Thioesterase	
UF_miSeq_TeF03	TSTTCGGSCACAGCMTSGGBG	CTACACGACGCTCTTCCGATCT	GGTGCGAA	CTA	CTACACGACGCTCTTCCGATCTGGTGCGAACTATSTTCGGSCACAGCMTSGGBG	UF_miSeq_TeR03	AGTCGSYSCGCAGCGSSGGCA	CAGACGTGTGCTCTTCCGATCT	GGTGCGAA	CTA	CAGACGTGTGCTCTTCCGATCTGGTGCGAACTAAGTCGSYSCGCAGCGSSGGCA	Thioesterase	
UF_miSeq_TeF04	TBTWYGGSCACAGCMTSGGBG	CTACACGACGCTCTTCCGATCT	TGAAGAGA	GATA	CTACACGACGCTCTTCCGATCTTGAAGAGAGATATBTWYGGSCACAGCMTSGGBG	UF_miSeq_TeR04	AGTCGSYSCGSASSGSSGGCA	CAGACGTGTGCTCTTCCGATCT	TGAAGAGA	GATA	CAGACGTGTGCTCTTCCGATCTTGAAGAGAGATAAGTCGSYSCGSASSGSSGGCA	Thioesterase	
UF_miSeq_TeF05	TTYTTYGGNCAYAGYATGGGN	CTACACGACGCTCTTCCGATCT	AAGAGATC	A	CTACACGACGCTCTTCCGATCTAAGAGATCATTYTTYGGNCAYAGYATGGGN	UF_miSeq_TeR05	RTCNCCNCKNARNGCNGGNAR	CAGACGTGTGCTCTTCCGATCT	GATAGACA	A	CAGACGTGTGCTCTTCCGATCTGATAGACAARTCNCCNCKNARNGCNGGNAR	Thioesterase	
UF_miSeq_TeF06	TWSTWSGGSMASHCRATGGGS	CTACACGACGCTCTTCCGATCT	TCCGTCTA	GATA	CTACACGACGCTCTTCCGATCTTCCGTCTAGATATWSTWSGGSMASHCRATGGGS	UF_miSeq_TeR06	STCSCCSCGSABSGCSGGSAB	CAGACGTGTGCTCTTCCGATCT	CACCTTAC	GATA	CAGACGTGTGCTCTTCCGATCTCACCTTACGATASTCSCCSCGSABSGCSGGSAB	Thioesterase	
UF_miSeq_TeF07	TWBTWBGGSMASHCSATGGGS	CTACACGACGCTCTTCCGATCT	AAGAGATC	A	CTACACGACGCTCTTCCGATCTAAGAGATCATWBTWBGGSMASHCSATGGGS	UF_miSeq_TeR07	STCSCCSCGSABSGCSGGNAB	CAGACGTGTGCTCTTCCGATCT	CCTAATCC	A	CAGACGTGTGCTCTTCCGATCTCCTAATCCASTCSCCSCGSABSGCSGGNAB	Thioesterase	
UF_miSeq_TeF08	TWBTWBGGNMASHCSATGGGS	CTACACGACGCTCTTCCGATCT	ACGTATCA	TC	CTACACGACGCTCTTCCGATCTACGTATCATCTWBTWBGGNMASHCSATGGGS	UF_miSeq_TeR08	STCSCCSCGSABSGCNGGNAB	CAGACGTGTGCTCTTCCGATCT	AGCCATGC	TC	CAGACGTGTGCTCTTCCGATCTAGCCATGCTCSTCSCCSCGSABSGCNGGNAB	Thioesterase	
UF_miSeq_TeF09	TWBTWBGGNMANHCSATGGGS	CTACACGACGCTCTTCCGATCT	AATGTTGC	CTA	CTACACGACGCTCTTCCGATCTAATGTTGCCTATWBTWBGGNMANHCSATGGGS	UF_miSeq_TeR09	STCSCCSCGSABNGCNGGNAB	CAGACGTGTGCTCTTCCGATCT	GTCTGTCA	CTA	CAGACGTGTGCTCTTCCGATCTGTCTGTCACTASTCSCCSCGSABNGCNGGNAB	Thioesterase	
UF_miSeq_TeF10	TWBTWBGGNMANHCNATGGGS	CTACACGACGCTCTTCCGATCT	CTCAATGA	GATA	CTACACGACGCTCTTCCGATCTCTCAATGAGATATWBTWBGGNMANHCNATGGGS	UF_miSeq_TeR10	STCSCCSCGNABNGCNGGNAB	CAGACGTGTGCTCTTCCGATCT	AGGCTAAC	GATA	CAGACGTGTGCTCTTCCGATCTAGGCTAACGATASTCSCCSCGNABNGCNGGNAB	Thioesterase	
UF_miSeq_AryCF01	GASGAYCTGATCAGCMAGCTS	CTACACGACGCTCTTCCGATCT	AGCACCTC	TC	CTACACGACGCTCTTCCGATCT AGCACCTC TC GASGAYCTGATCAGCMAGCTS	UF_miSeq_AryCR01	GCCSAGRCAGAARTGGATGCY	CAGACGTGTGCTCTTCCGATCT	CCATCCTC	TC	CAGACGTGTGCTCTTCCGATCTCCATCCTCTCGCCSAGRCAGAARTGGATGCY	AryC	496
UF_miSeq_AryCF02	GASGATCTSAYCAGCMRSCTS	CTACACGACGCTCTTCCGATCT	CCGTGAGA	CTA	CTACACGACGCTCTTCCGATCTCCGTGAGACTAGASGATCTSAYCAGCMRSCTS	UF_miSeq_AryCR02	RCCSAGRCARWARTGGATKCC	CAGACGTGTGCTCTTCCGATCT	AAACATCG	CTA	CAGACGTGTGCTCTTCCGATCTAAACATCGCTARCCSAGRCARWARTGGATKCC	AryC	496
UF_miSeq_AryCF03	GASGAYCTSAYCAGCMRSCTS	CTACACGACGCTCTTCCGATCT	AAACATCG	GATA	CTACACGACGCTCTTCCGATCTAAACATCGGATAGASGAYCTSAYCAGCMRSCTS	UF_miSeq_AryCR03	RCCSAGRCARWARTGGATKCY	CAGACGTGTGCTCTTCCGATCT	AATGTTGC	GATA	CAGACGTGTGCTCTTCCGATCTAATGTTGCGATARCCSAGRCARWARTGGATKCY	AryC	496
UF_miSeq_AryCF04	GASGASCTSATHAGYMRSCTS	CTACACGACGCTCTTCCGATCT	GTGTTCTA	A	CTACACGACGCTCTTCCGATCTGTGTTCTAAGASGASCTSATHAGYMRSCTS	UF_miSeq_AryCR04	SCCSAGRCARAARTGDATNGA	CAGACGTGTGCTCTTCCGATCT	ACGTATCA	A	CAGACGTGTGCTCTTCCGATCTACGTATCAASCCSAGRCARAARTGDATNGA	AryC	496
UF_miSeq_AryCF05	GANGASCTSATHAGYMRSCTS	CTACACGACGCTCTTCCGATCT	CGCATACA	CTA	CTACACGACGCTCTTCCGATCTCGCATACACTAGANGASCTSATHAGYMRSCTS	UF_miSeq_AryCR05	SCCNAGRCARAARTGDATNGA	CAGACGTGTGCTCTTCCGATCT	GACAGTGC	CTA	CAGACGTGTGCTCTTCCGATCTGACAGTGCCTASCCNAGRCARAARTGDATNGA	AryC	496
UF_miSeq_AryCF06	GANGANCTSATHAGYMRSCTS	CTACACGACGCTCTTCCGATCT	CATCAAGT	GATA	CTACACGACGCTCTTCCGATCTCATCAAGTGATAGANGANCTSATHAGYMRSCTS	UF_miSeq_AryCR06	NCCNAGRCARAARTGDATNGA	CAGACGTGTGCTCTTCCGATCT	AGAGTCAA	GATA	CAGACGTGTGCTCTTCCGATCTAGAGTCAAGATANCCNAGRCARAARTGDATNGA	AryC	496
UF_miSeq_AryCF07	GANGANCTNATHAGYMRSCTS	CTACACGACGCTCTTCCGATCT	AGTACAAG	A	CTACACGACGCTCTTCCGATCTAGTACAAGAGANGANCTNATHAGYMRSCTS	UF_miSeq_AryCR07	SCCSABRCASWASTKSABYGD	CAGACGTGTGCTCTTCCGATCT	ACCACTGT	A	CAGACGTGTGCTCTTCCGATCTACCACTGTASCCSABRCASWASTKSABYGD	AryC	496
UF_miSeq_AryCF08	GANGASVTSVTSHCSMASVTS	CTACACGACGCTCTTCCGATCT	CGCTGATC	TC	CTACACGACGCTCTTCCGATCTCGCTGATCTCGANGASVTSVTSHCSMASVTS	UF_miSeq_AryCR08	SCCSABRCAVWASTKSABNGD	CAGACGTGTGCTCTTCCGATCT	CGGATTGC	TC	CAGACGTGTGCTCTTCCGATCTCGGATTGCTCSCCSABRCAVWASTKSABNGD	AryC	496
UF_miSeq_CysJF01	TGGCCSCGSAAGCARTGGTTC	CTACACGACGCTCTTCCGATCT	ACTATGCA	CTA	CTACACGACGCTCTTCCGATCTACTATGCACTATGGCCSCGSAAGCARTGGTTC	UF_miSeq_CysJR01	CATGTTGTCSAGCATSACSAC	CAGACGTGTGCTCTTCCGATCT	AGCCATGC	CTA	CAGACGTGTGCTCTTCCGATCTAGCCATGCCTACATGTTGTCSAGCATSACSAC	CysJ	611
UF_miSeq_CysJF02	TGGCCSCGSAARCARTGGTTY	CTACACGACGCTCTTCCGATCT	CCTCTATC	GATA	CTACACGACGCTCTTCCGATCTCCTCTATCGATATGGCCSCGSAARCARTGGTTY	UF_miSeq_CysJR02	CATRTTRTCSAGCATSACSAC	CAGACGTGTGCTCTTCCGATCT	ATCCTGTA	GATA	CAGACGTGTGCTCTTCCGATCTATCCTGTAGATACATRTTRTCSAGCATSACSAC	CysJ	611
UF_miSeq_CysJF03	TGGCCNCGSAARCARTGGTTY	CTACACGACGCTCTTCCGATCT	CTGGCATA	A	CTACACGACGCTCTTCCGATCTCTGGCATAATGGCCNCGSAARCARTGGTTY	UF_miSeq_CysJR03	CATRTTRTCSAGCATSACNAC	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	A	CAGACGTGTGCTCTTCCGATCTCAAGGAGCACATRTTRTCSAGCATSACNAC	CysJ	611

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_CysJF04	TGGCCNCGNAARCARTGGTTY	CTACACGACGCTCTTCCGATCT	TCCGTCTA	CTA	CTACACGACGCTCTTCCGATCTTCCGTCTACTATGGCCNCGNAARCARTGGTTY	UF_miSeq_CysJR04	CATRTTRTCSAGCATNACNAC	CAGACGTGTGCTCTTCCGATCT	CAACCACA	CTA	CAGACGTGTGCTCTTCCGATCTCAACCACACTACATRTTRTCSAGCATNACNAC	CysJ	611
UF_miSeq_CysJF05	TWSCCSCGSSTSMASTWSTWS	CTACACGACGCTCTTCCGATCT	CAGCGTTA	GATA	CTACACGACGCTCTTCCGATCTCAGCGTTAGATATWSCCSCGSSTSMASTWSTWS	UF_miSeq_CysJR05	CATSTKSTCSABCATSABSAB	CAGACGTGTGCTCTTCCGATCT	ACAGCAGA	GATA	CAGACGTGTGCTCTTCCGATCTACAGCAGAGATACATSTKSTCSABCATSABSAB	CysJ	611
UF_miSeq_CysJF06	TWBCCSCGSVTSMASTWSTWS	CTACACGACGCTCTTCCGATCT	TGGCTTCA	A	CTACACGACGCTCTTCCGATCTTGGCTTCAATWBCCSCGSVTSMASTWSTWS	UF_miSeq_CysJR06	CATSTKSTCSABCATSABNAB	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	A	CAGACGTGTGCTCTTCCGATCTTAGGATGAACATSTKSTCSABCATSABNAB	CysJ	611
UF_miSeq_CysJF07	TWBCCNCGSVTSMASTWSTWS	CTACACGACGCTCTTCCGATCT	AAGGACAC	TC	CTACACGACGCTCTTCCGATCTAAGGACACTCTWBCCNCGSVTSMASTWSTWS	UF_miSeq_CysJR07	CATSTKSTCSABCATNABNAB	CAGACGTGTGCTCTTCCGATCT	TGGTGGTA	TC	CAGACGTGTGCTCTTCCGATCTTGGTGGTATCCATSTKSTCSABCATNABNAB	CysJ	611
UF_miSeq_CysJF08	TWBCCNCGNVTSMASTWSTWS	CTACACGACGCTCTTCCGATCT	GATGAATC	CTA	CTACACGACGCTCTTCCGATCTGATGAATCCTATWBCCNCGNVTSMASTWSTWS	UF_miSeq_CysJR08	CATSTKSTCNABCATNABNAB	CAGACGTGTGCTCTTCCGATCT	AGCCATGC	CTA	CAGACGTGTGCTCTTCCGATCTAGCCATGCCTACATSTKSTCNABCATNABNAB	CysJ	611
UF_miSeq_GriFF01	GTSCAGGGCAGCGGVCCGGTY	CTACACGACGCTCTTCCGATCT	CGAACTTA	GATA	CTACACGACGCTCTTCCGATCTCGAACTTAGATAGTSCAGGGCAGCGGVCCGGTY	UF_miSeq_GriFR01	GTACTGSCCGAYGAYSAGGTA	CAGACGTGTGCTCTTCCGATCT	TGGTGGTA	GATA	CAGACGTGTGCTCTTCCGATCTTGGTGGTAGATAGTACTGSCCGAYGAYSAGGTA	GriF	373
UF_miSeq_GriFF02	GTBCAGGGCAGYGGVCCSGTY	CTACACGACGCTCTTCCGATCT	AGTGGTCA	A	CTACACGACGCTCTTCCGATCTAGTGGTCAAGTBCAGGGCAGYGGVCCSGTY	UF_miSeq_GriFR02	GTACTGSCCSAYRAYSAGRTA	CAGACGTGTGCTCTTCCGATCT	AGAGTCAA	A	CAGACGTGTGCTCTTCCGATCTAGAGTCAAAGTACTGSCCSAYRAYSAGRTA	GriF	373
UF_miSeq_GriFF03	GTNCARGGYWGYGGSCCSGTS	CTACACGACGCTCTTCCGATCT	GACTAGTA	CTA	CTACACGACGCTCTTCCGATCTGACTAGTACTAGTNCARGGYWGYGGSCCSGTS	UF_miSeq_GriFR03	GTAMTGSCCRAYRAYVASRTA	CAGACGTGTGCTCTTCCGATCT	CAAGGAGC	CTA	CAGACGTGTGCTCTTCCGATCTCAAGGAGCCTAGTAMTGSCCRAYRAYVASRTA	GriF	373
UF_miSeq_GriFF04	GTNCARGGYWGYGGSCCBGTB	CTACACGACGCTCTTCCGATCT	CTGGCATA	GATA	CTACACGACGCTCTTCCGATCTCTGGCATAGATAGTNCARGGYWGYGGSCCBGTB	UF_miSeq_GriFR04	RTAYTGNCCNACNACSAGRTA	CAGACGTGTGCTCTTCCGATCT	GAACAGGC	GATA	CAGACGTGTGCTCTTCCGATCTGAACAGGCGATARTAYTGNCCNACNACSAGRTA	GriF	373
UF_miSeq_GriFF05	GTNCARGGNWGYGGNCCSGTS	CTACACGACGCTCTTCCGATCT	ACCACTGT	A	CTACACGACGCTCTTCCGATCTACCACTGTAGTNCARGGNWGYGGNCCSGTS	UF_miSeq_GriFR05	SWASTKSCCSASSASSABSWA	CAGACGTGTGCTCTTCCGATCT	TTCACGCA	A	CAGACGTGTGCTCTTCCGATCTTTCACGCAASWASTKSCCSASSASSABSWA	GriF	373
UF_miSeq_GriFF06	GTNCARGGNWGYGGNCCNGTS	CTACACGACGCTCTTCCGATCT	GTGTTCTA	TC	CTACACGACGCTCTTCCGATCTGTGTTCTATCGTNCARGGNWGYGGNCCNGTS	UF_miSeq_GriFR06	SWASTKSCCSASSASBSWA	CAGACGTGTGCTCTTCCGATCT	ATTGAGGA	TC	CAGACGTGTGCTCTTCCGATCTATTGAGGATCSWASTKSCCSASSABSABSWA	GriF	373
UF_miSeq_GriFF07	GTNCARGGNWGYGGNCCNGTN	CTACACGACGCTCTTCCGATCT	ATTGAGGA	CTA	CTACACGACGCTCTTCCGATCTATTGAGGACTAGTNCARGGNWGYGGNCCNGTN	UF_miSeq_GriFR07	VWASTKSCCSABSABSABVWA	CAGACGTGTGCTCTTCCGATCT	CGCATACA	CTA	CAGACGTGTGCTCTTCCGATCTCGCATACACTAVWASTKSCCSABSABSABVWA	GriF	373
UF_miSeq_GriFF08	VTNMANGGNHCSGGSCCSVTS	CTACACGACGCTCTTCCGATCT	GATAGACA	GATA	CTACACGACGCTCTTCCGATCTGATAGACAGATAVTNMANGGNHCSGGSCCSVTS	UF_miSeq_GriFR08	VWASTKSCCSABSABNABVWA	CAGACGTGTGCTCTTCCGATCT	AAGGACAC	GATA	CAGACGTGTGCTCTTCCGATCTAAGGACACGATAVWASTKSCCSABSABNABVWA	GriF	373
UF_miSeq_HolEF01	GMACGCTTYTTTAATRTYGAT	CTACACGACGCTCTTCCGATCT	AGTCACTA	A	CTACACGACGCTCTTCCGATCTAGTCACTAAGMACGCTTYTTTAATRTYGAT	UF_miSeq_HolER01	WGGCCCRAGYCGATAWCCATC	CAGACGTGTGCTCTTCCGATCT	GAGCTGAA	A	CAGACGTGTGCTCTTCCGATCTGAGCTGAAAWGGCCCRAGYCGATAWCCATC	HolE	576
UF_miSeq_HolEF02	GSSCATTTTTTYRMTGTMGAT	CTACACGACGCTCTTCCGATCT	CTGGCATA	CTA	CTACACGACGCTCTTCCGATCTCTGGCATACTAGSSCATTTTTTYRMTGTMGAT	UF_miSeq_HolER02	WGGCCCRASYCGATAHCCATC	CAGACGTGTGCTCTTCCGATCT	ATAGCGAC	CTA	CAGACGTGTGCTCTTCCGATCTATAGCGACCTAWGGCCCRASYCGATAHCCATC	HolE	576
UF_miSeq_HolEF03	GMRYRYTTYTTTAATRTYGAT	CTACACGACGCTCTTCCGATCT	CCGTGAGA	GATA	CTACACGACGCTCTTCCGATCTCCGTGAGAGATAGMRYRYTTYTTTAATRTYGAT	UF_miSeq_HolER03	WGGSCCSAGSCGRTASCCRTC	CAGACGTGTGCTCTTCCGATCT	CCTAATCC	GATA	CAGACGTGTGCTCTTCCGATCTCCTAATCCGATAWGGSCCSAGSCGRTASCCRTC	HolE	576
UF_miSeq_HolEF04	GARCGNTTYTTYAAYATHGAY	CTACACGACGCTCTTCCGATCT	ATGCCTAA	A	CTACACGACGCTCTTCCGATCTATGCCTAAAGARCGNTTYTTYAAYATHGAY	UF_miSeq_HolER04	SGGSCCSAGSSGRTANCCRTC	CAGACGTGTGCTCTTCCGATCT	TAGGATGA	A	CAGACGTGTGCTCTTCCGATCTTAGGATGAASGGSCCSAGSSGRTANCCRTC	HolE	576
UF_miSeq_HolEF05	GASCGSTWSTWSMASVTSGAS	CTACACGACGCTCTTCCGATCT	ACCACTGT	TC	CTACACGACGCTCTTCCGATCTACCACTGTTCGASCGSTWSTWSMASVTSGAS	UF_miSeq_HolER05	SGGSCCSAGNSGRTANCCRTC	CAGACGTGTGCTCTTCCGATCT	CGCATACA	TC	CAGACGTGTGCTCTTCCGATCTCGCATACATCSGGSCCSAGNSGRTANCCRTC	HolE	576
UF_miSeq_HolEF06	GANCGSTWSTWSMASVTSGAS	CTACACGACGCTCTTCCGATCT	ATCCTGTA	CTA	CTACACGACGCTCTTCCGATCTATCCTGTACTAGANCGSTWSTWSMASVTSGAS	UF_miSeq_HolER06	SGGSCCNAGNSGRTANCCRTC	CAGACGTGTGCTCTTCCGATCT	GAGTTAGC	CTA	CAGACGTGTGCTCTTCCGATCTGAGTTAGCCTASGGSCCNAGNSGRTANCCRTC	HolE	576
UF_miSeq_HolEF07	GANCGNTWSTWSMASVTSGAS	CTACACGACGCTCTTCCGATCT	GGAGAACA	GATA	CTACACGACGCTCTTCCGATCTGGAGAACAGATAGANCGNTWSTWSMASVTSGAS	UF_miSeq_HolER07	SGGNCCNAGNSGRTANCCRTC	CAGACGTGTGCTCTTCCGATCT	TATCAGCA	GATA	CAGACGTGTGCTCTTCCGATCTTATCAGCAGATASGGNCCNAGNSGRTANCCRTC	HolE	576
UF_miSeq_HolEF08	GANCGNTWNTWSMASVTSGAS	CTACACGACGCTCTTCCGATCT	GTGTTCTA	A	CTACACGACGCTCTTCCGATCTGTGTTCTAAGANCGNTWNTWSMASVTSGAS	UF_miSeq_HolER08	NGGNCCNAGNSGRTANCCRTC	CAGACGTGTGCTCTTCCGATCT	AACTCACC	A	CAGACGTGTGCTCTTCCGATCTAACTCACCANGGNCCNAGNSGRTANCCRTC	HolE	576
UF_miSeq_NgnN4F01	GCGATGCWGGCGGYGCTGCGC	CTACACGACGCTCTTCCGATCT	AACTCACC	CTA	CTACACGACGCTCTTCCGATCTAACTCACCCTAGCGATGCWGGCGGYGCTGCGC	UF_miSeq_NgnN4R01	TTCGGTCGGGCCATACAGGTT	CAGACGTGTGCTCTTCCGATCT	ATCATTCC	CTA	CAGACGTGTGCTCTTCCGATCTATCATTCCCTATTCGGTCGGGCCATACAGGTT	NgnN4	576
UF_miSeq_NgnN4F02	GCSATGCTSGCSGYSCTSCGS	CTACACGACGCTCTTCCGATCT	AATGTTGC	GATA	CTACACGACGCTCTTCCGATCTAATGTTGCGATAGCSATGCTSGCSGYSCTSCGS	UF_miSeq_NgnN4R02	YTCSGTSGGSCCRTASAGRTT	CAGACGTGTGCTCTTCCGATCT	CACTTCGA	GATA	CAGACGTGTGCTCTTCCGATCTCACTTCGAGATAYTCSGTSGGSCCRTASAGRTT	NgnN4	576
UF_miSeq_NgnN4F03	GCNATGCTSGCSGYSCTSCGS	CTACACGACGCTCTTCCGATCT	AGTACAAG	A	CTACACGACGCTCTTCCGATCTAGTACAAGAGCNATGCTSGCSGYSCTSCGS	UF_miSeq_NgnN4R03	YTCSGTSGGSCCRTANAGRTT	CAGACGTGTGCTCTTCCGATCT	GATAGACA	A	CAGACGTGTGCTCTTCCGATCTGATAGACAAYTCSGTSGGSCCRTANAGRTT	NgnN4	576
UF_miSeq_NgnN4F04	GCNATGCTNGCSGYSCTSCGS	CTACACGACGCTCTTCCGATCT	ACAGATTC	TC	CTACACGACGCTCTTCCGATCTACAGATTCTCGCNATGCTNGCSGYSCTSCGS	UF_miSeq_NgnN4R04	YTCSGTNGGSCCRTANAGRTT	CAGACGTGTGCTCTTCCGATCT	AATGTTGC	TC	CAGACGTGTGCTCTTCCGATCTAATGTTGCTCYTCSGTNGGSCCRTANAGRTT	NgnN4	576
UF_miSeq_NgnN4F05	GCNATGVTSGCSGYSVTSCGS	CTACACGACGCTCTTCCGATCT	AAGGACAC	CTA	CTACACGACGCTCTTCCGATCTAAGGACACCTAGCNATGVTSGCSGYSVTSCGS	UF_miSeq_NgnN4R05	STCYGDSGGSCCSWASABSTK	CAGACGTGTGCTCTTCCGATCT	ATTGAGGA	CTA	CAGACGTGTGCTCTTCCGATCTATTGAGGACTASTCYGDSGGSCCSWASABSTK	NgnN4	576
UF_miSeq_NgnN4F06	GCNATGCTNGCNGYNCTSCGS	CTACACGACGCTCTTCCGATCT	CGCATACA	GATA	CTACACGACGCTCTTCCGATCTCGCATACAGATAGCNATGCTNGCNGYNCTSCGS	UF_miSeq_NgnN4R06	STCSGDSGGSCCVWASABNTK	CAGACGTGTGCTCTTCCGATCT	AGCACCTC	GATA	CAGACGTGTGCTCTTCCGATCTAGCACCTCGATASTCSGDSGGSCCVWASABNTK	NgnN4	576
UF_miSeq_NgnN4F07	GCNATGVTNGCNGYSVTSCGS	CTACACGACGCTCTTCCGATCT	GCTCGGTA	A	CTACACGACGCTCTTCCGATCTGCTCGGTAAGCNATGVTNGCNGYSVTSCGS	UF_miSeq_NgnN4R07	STCSGDSGGSCCVWANABNTK	CAGACGTGTGCTCTTCCGATCT	CAATGGAA	A	CAGACGTGTGCTCTTCCGATCTCAATGGAAASTCSGDSGGSCCVWANABNTK	NgnN4	576
UF_miSeq_NgnN4F08	GCNATGVTNGCNGYNVTSCGS	CTACACGACGCTCTTCCGATCT	CGCTGATC	CTA	CTACACGACGCTCTTCCGATCTCGCTGATCCTAGCNATGVTNGCNGYNVTSCGS	UF_miSeq_NgnN4R08	STCSGDSGGNCCVWANABNTK	CAGACGTGTGCTCTTCCGATCT	CAGCGTTA	CTA	CAGACGTGTGCTCTTCCGATCTCAGCGTTACTASTCSGDSGGNCCVWANABNTK	NgnN4	576

PrimerNameF	PrimerF	MiSeq_Fadaptor	BarcodeF	SpacerF	Full_SEQ_F	PrimerNameR	PrimerR	MiSeq_R_adaptor	BarcodeR	SpacerR	Full_SEQ_R	Target	AmpliconLenght
UF_miSeq_TmlUF01	TTRCCGYTKTTCCACGTCAAY	CTACACGACGCTCTTCCGATCT	GAGTTAGC	GATA	CTACACGACGCTCTTCCGATCTGAGTTAGCGATATTRCCGYTKTTCCACGTCAAY	UF_miSeq_TmlUR01	GTTGTRGCCSCCGCGRATGAT	CAGACGTGTGCTCTTCCGATCT	CAAGACTA	GATA	CAGACGTGTGCTCTTCCGATCTCAAGACTAGATAGTTGTRGCCSCCGCGRATGAT	TmlU	714
UF_miSeq_TmlUF02	TTGCCGYTVTTYCAYGTCAAY	CTACACGACGCTCTTCCGATCT	GAACAGGC	A	CTACACGACGCTCTTCCGATCTGAACAGGCATTGCCGYTVTTYCAYGTCAAY	UF_miSeq_TmlUR02	GTTGTRKCCSCCRCGRATRAT	CAGACGTGTGCTCTTCCGATCT	CGAACTTA	A	CAGACGTGTGCTCTTCCGATCTCGAACTTAAGTTGTRKCCSCCRCGRATRAT	TmlU	714
UF_miSeq_TmlUF03	TTRCCGYTDTTYCAYGTCAAY	CTACACGACGCTCTTCCGATCT	ACATTGGC	TC	CTACACGACGCTCTTCCGATCTACATTGGCTCTTRCCGYTDTTYCAYGTCAAY	UF_miSeq_TmlUR03	GTTGTRKCCRCCDCGRATRAT	CAGACGTGTGCTCTTCCGATCT	CCTCTATC	TC	CAGACGTGTGCTCTTCCGATCTCCTCTATCTCGTTGTRKCCRCCDCGRATRAT	TmlU	714
UF_miSeq_TmlUF04	TTRCCGYTNTTYCAYGTCAAY	CTACACGACGCTCTTCCGATCT	AAGGTACA	CTA	CTACACGACGCTCTTCCGATCTAAGGTACACTATTRCCGYTNTTYCAYGTCAAY	UF_miSeq_TmlUR04	RTTRTGSCCSCCSCGRATDAT	CAGACGTGTGCTCTTCCGATCT	AGGCTAAC	CTA	CAGACGTGTGCTCTTCCGATCTAGGCTAACCTARTTRTGSCCSCCSCGRATDAT	TmlU	714
UF_miSeq_TmlUF05	TTRCCRYTNTTYCAYGTYAAY	CTACACGACGCTCTTCCGATCT	CAGCGTTA	GATA	CTACACGACGCTCTTCCGATCTCAGCGTTAGATATTRCCRYTNTTYCAYGTYAAY	UF_miSeq_TmlUR05	RTTGTRBCCVCCDCGRATRAT	CAGACGTGTGCTCTTCCGATCT	GCGAGTAA	GATA	CAGACGTGTGCTCTTCCGATCTGCGAGTAAGATARTTGTRBCCVCCDCGRATRAT	TmlU	714
UF_miSeq_TmlUF06	CTNCCNCTNTTYCAYGTSAAY	CTACACGACGCTCTTCCGATCT	AAACATCG	A	CTACACGACGCTCTTCCGATCTAAACATCGACTNCCNCTNTTYCAYGTSAAY	UF_miSeq_TmlUR06	RTTRTGSCCNCCNCGDATDAT	CAGACGTGTGCTCTTCCGATCT	AATCCGTC	A	CAGACGTGTGCTCTTCCGATCTAATCCGTCARTTRTGSCCNCCNCGDATDAT	TmlU	714
UF_miSeq_TmlUF07	CTNCCNCTNTTYCAYGTNAAY	CTACACGACGCTCTTCCGATCT	TCTTCACA	CTA	CTACACGACGCTCTTCCGATCTTCTTCACACTACTNCCNCTNTTYCAYGTNAAY	UF_miSeq_TmlUR07	RTTRTGNCCNCCNCGDATDAT	CAGACGTGTGCTCTTCCGATCT	GTCGTAGA	CTA	CAGACGTGTGCTCTTCCGATCTGTCGTAGACTARTTRTGNCCNCCNCGDATDAT	TmlU	714
UF_miSeq_TmlUF08	VTNCCSVTSTWSMASVTSMAS	CTACACGACGCTCTTCCGATCT	ATCCTGTA	GATA	CTACACGACGCTCTTCCGATCTATCCTGTAGATAVTNCCSVTSTWSMASVTSMAS	UF_miSeq_TmlUR08	STKSTKSCCSCCSCGNABNAB	CAGACGTGTGCTCTTCCGATCT	AACGTGAT	GATA	CAGACGTGTGCTCTTCCGATCTAACGTGATGATASTKSTKSCCSCCSCGNABNAB	TmlU	714