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

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

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).

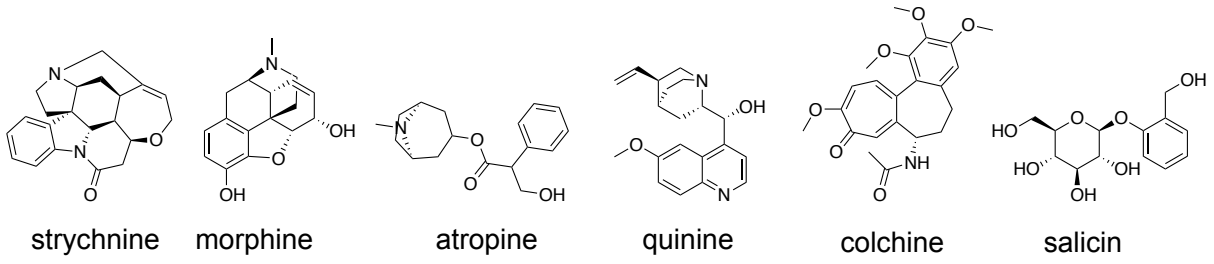


Figure 1: Natural Products breakthrough isolated from plants

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).

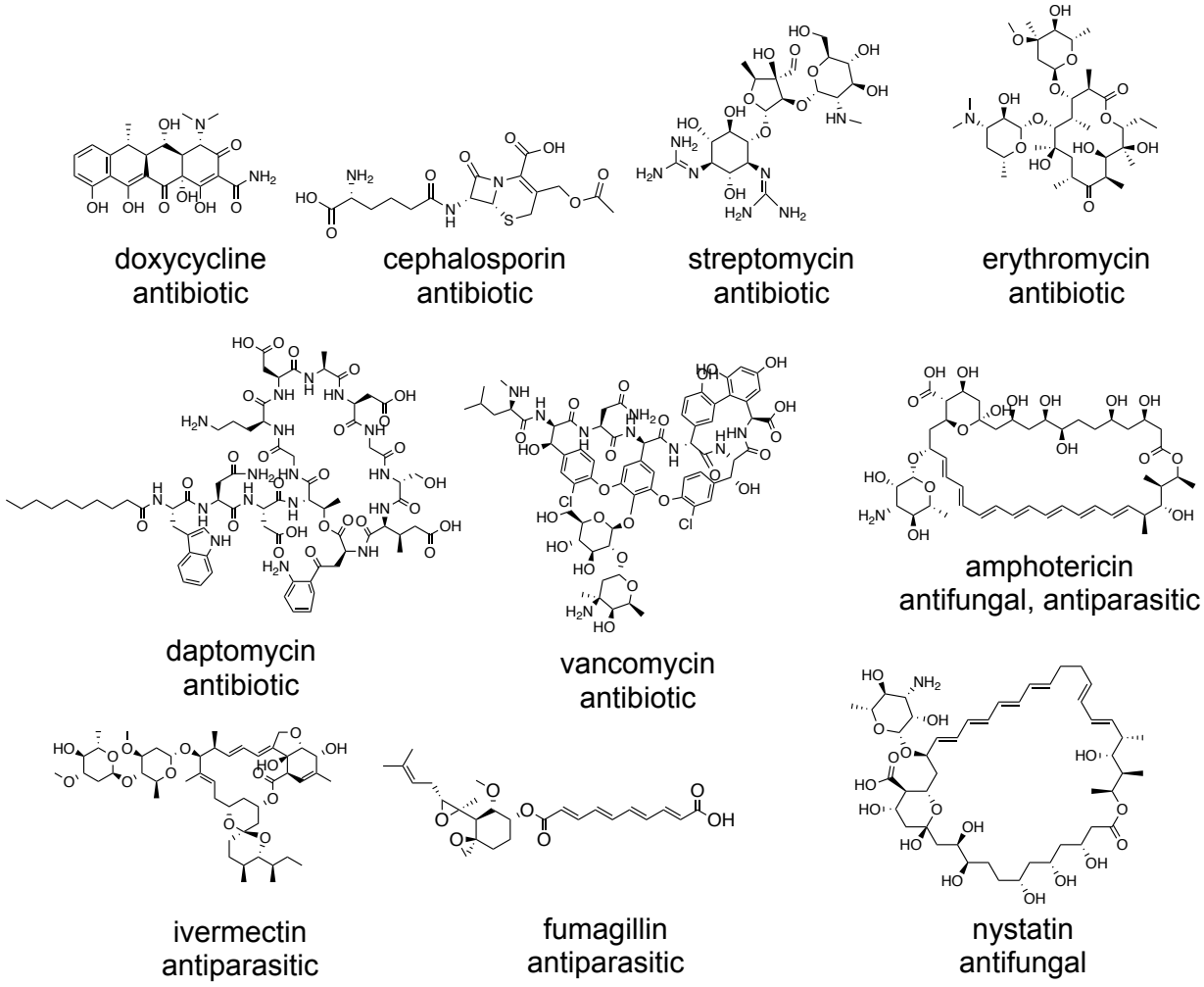
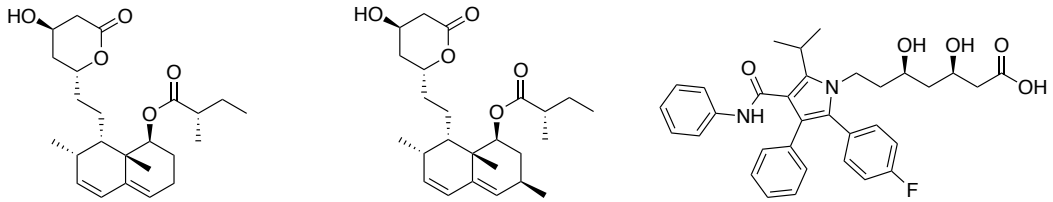


Figure 2: Natural Products as therapeutic agents

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).

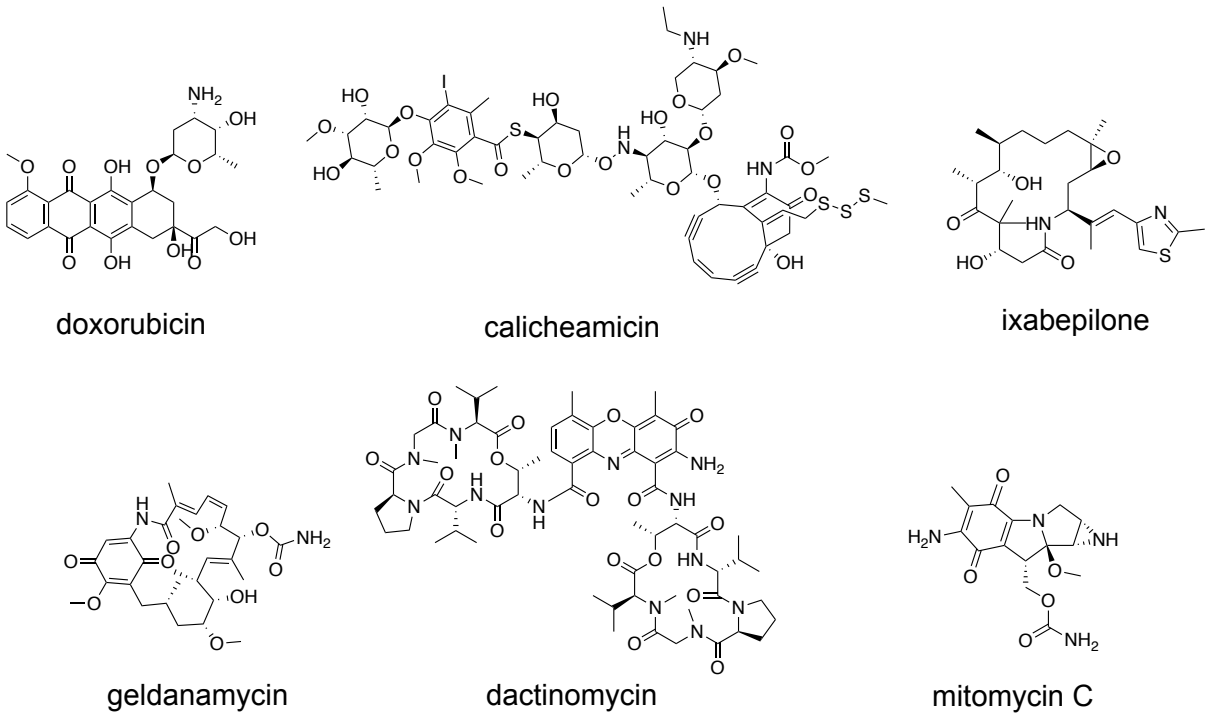


Figure 4: Therapeutic natural products isolated from bacteria

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.

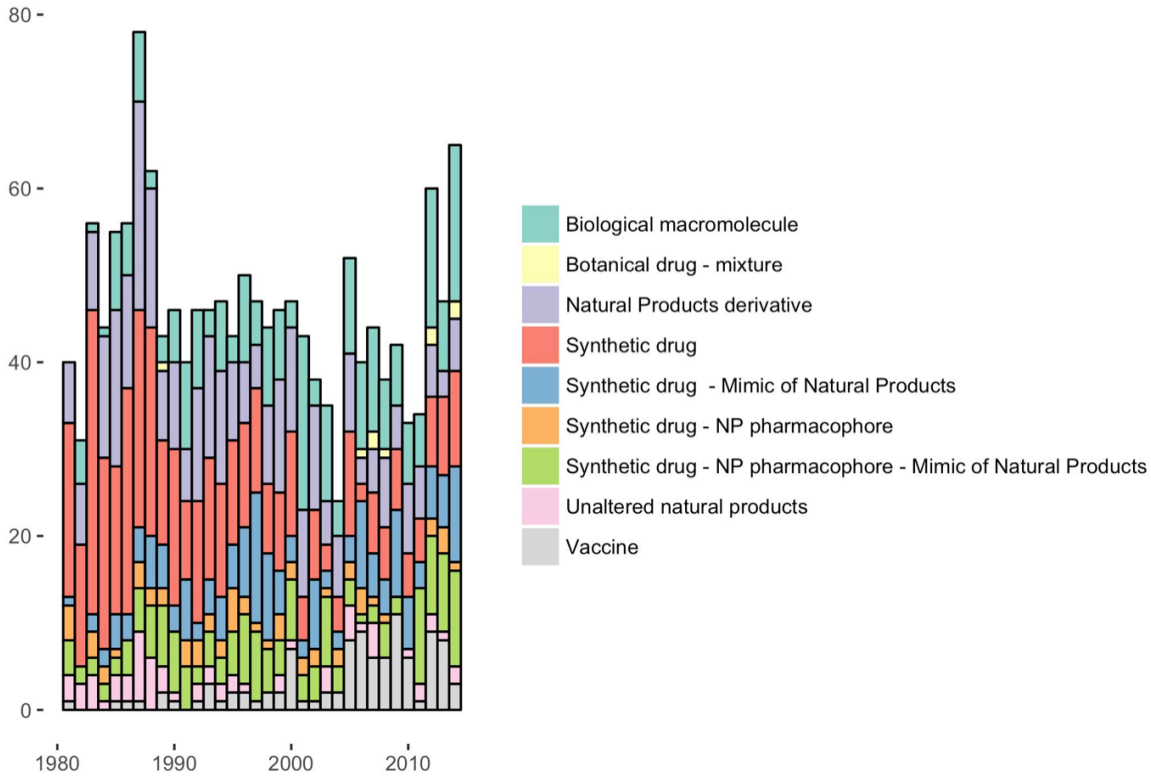


Figure 5: Distribution of 1562 drugs approved by FDA the during 1980 - 2014
This graphic represents the correlation of the amount of drugs approved by FDA (x-axis) by year (y-axis). The colors in legend represent drugs class. This data is available in reference - Newman and Cragg (2016).

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.

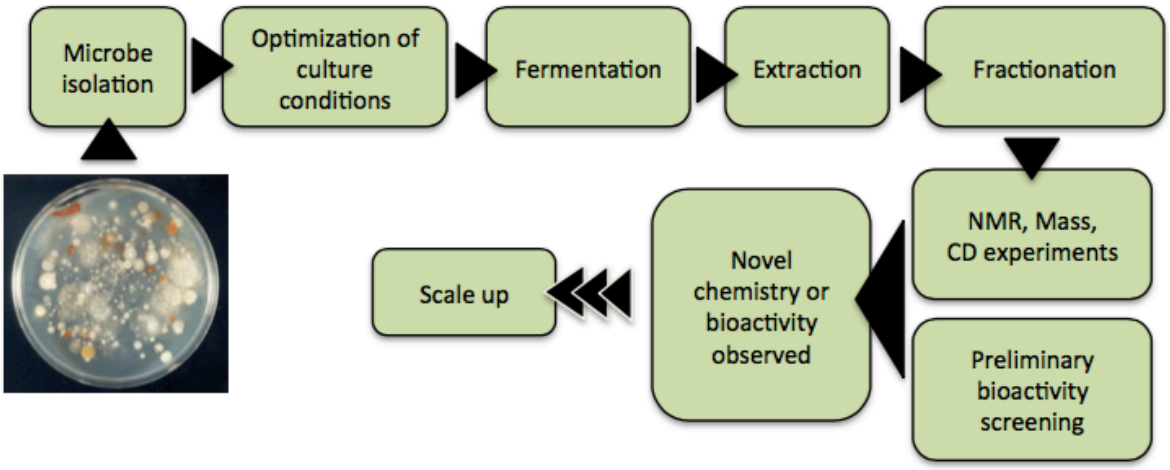


Figure 6: Traditional culture-based methods in microbial natural products isolation
Arrows guide the pipeline for isolation and characterization of natural products.
Methods could change across laboratories, however the general idea is maintained.

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).

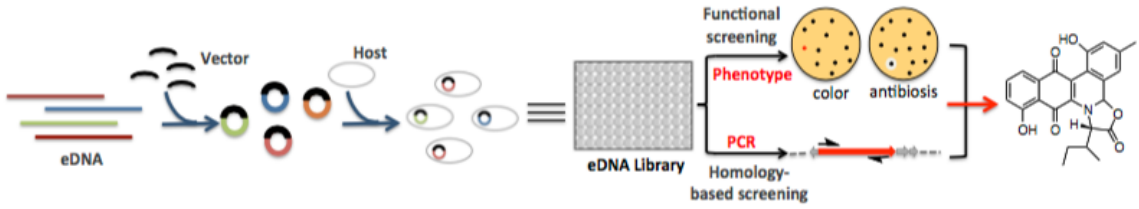


Figure 7: Metagenomic library construction and screening
The general idea is connect the eDNA with vectors using ligase and put inside a host (in general *E. coli*). The library is construct and screening by phenotypic screenings.

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**).

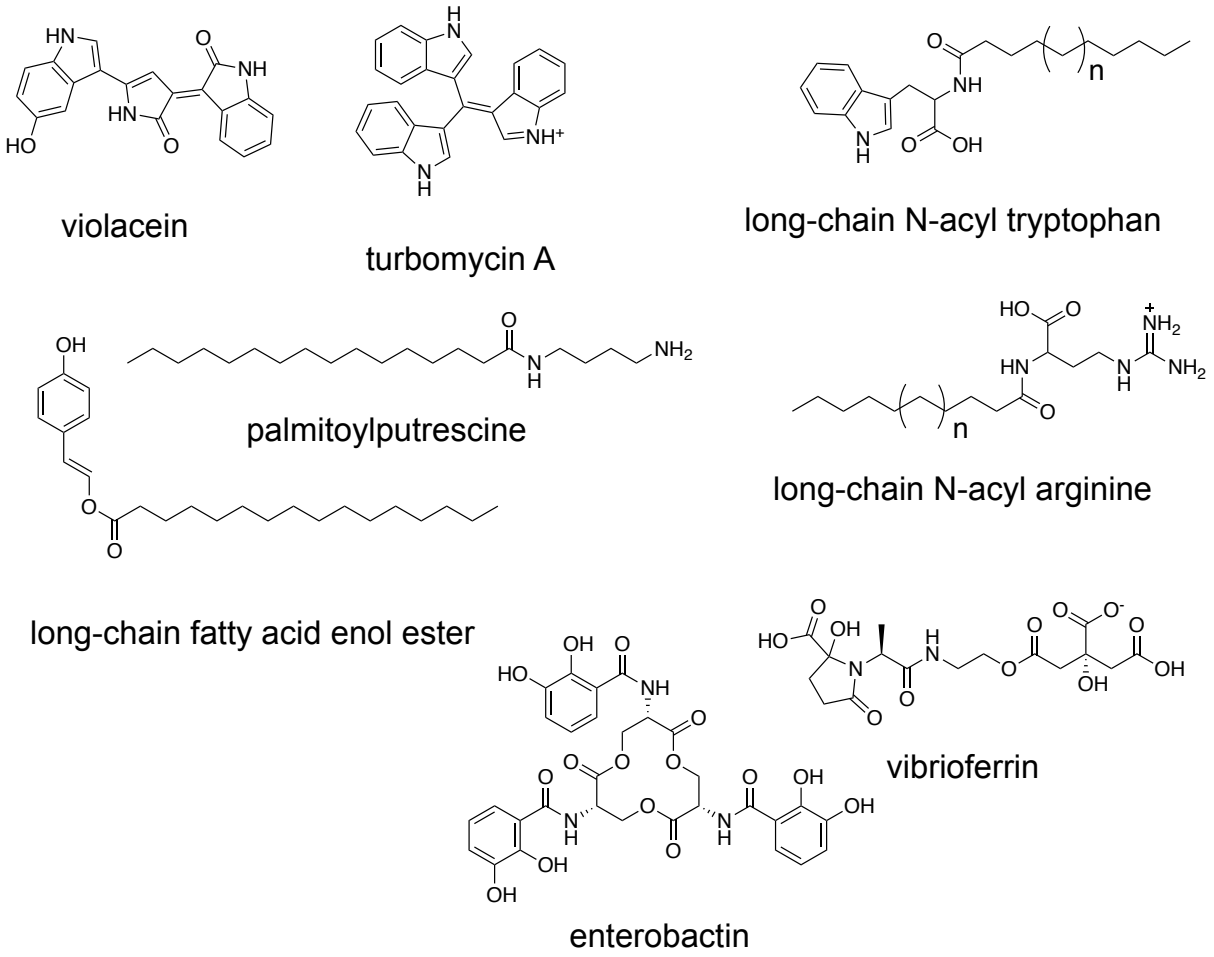


Figure 8: Natural products identified by functional metagenomic screening
This molecules were obtained from different libraries and screening using different host

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**).

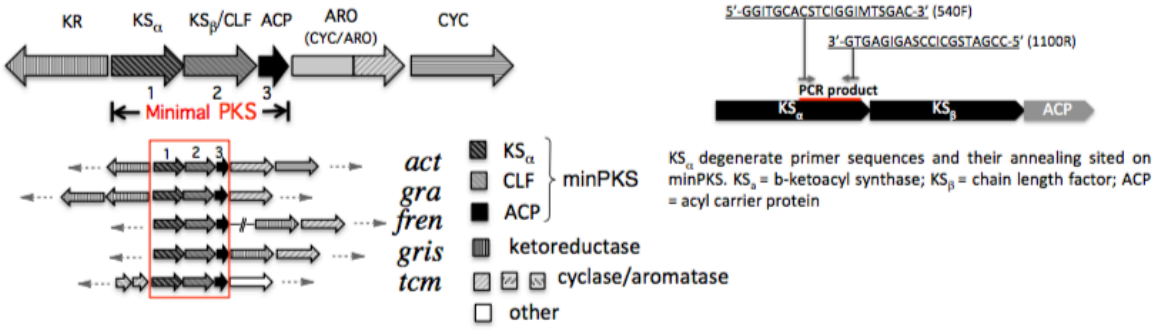


Figure 9: PCR-based screening for novel natural products
The type II PKS here described show an example of natural products gene cluster and how the PCR approach could be used to design primers targeting conserved genes.

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).

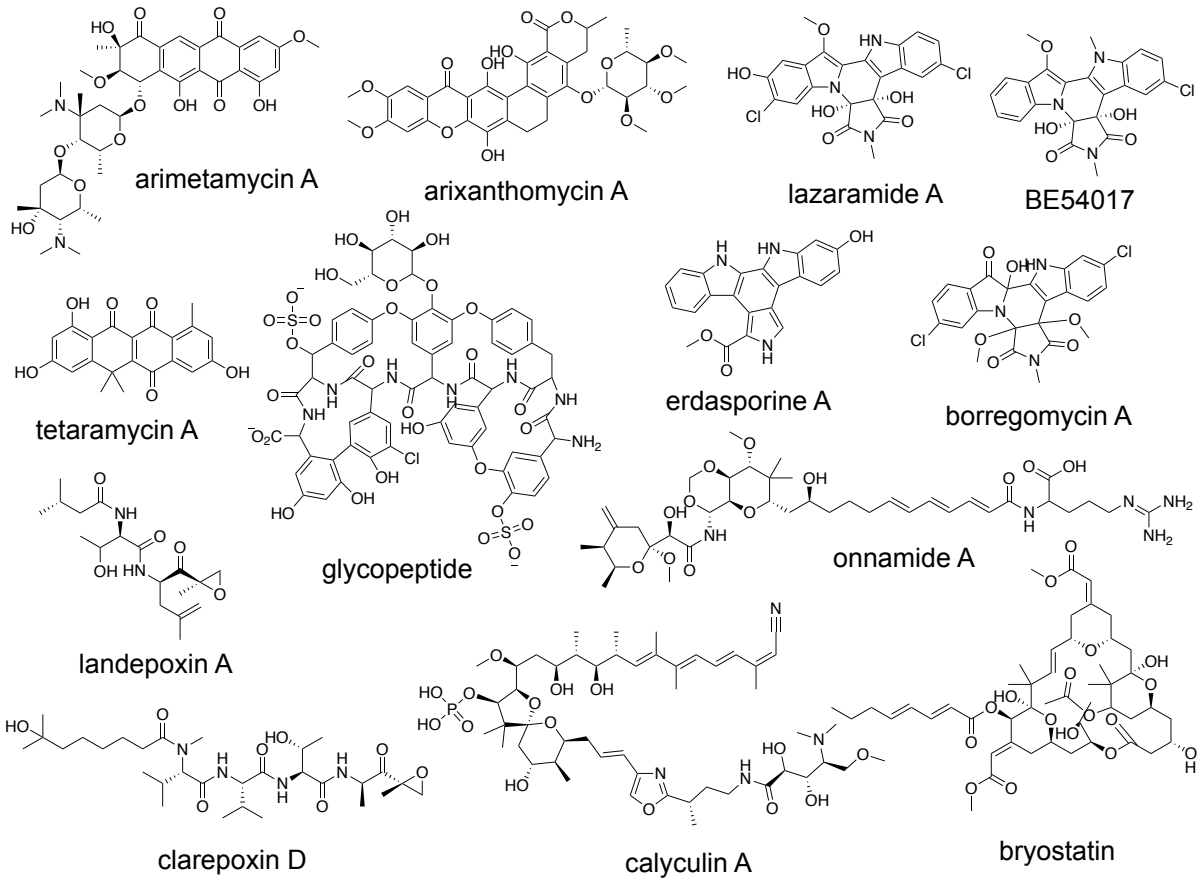


Figure 10: Some natural products molecules from homology screenings
The examples described were obtained from different metagenomic libraries. The homology were also detected using PCR-based approach

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).

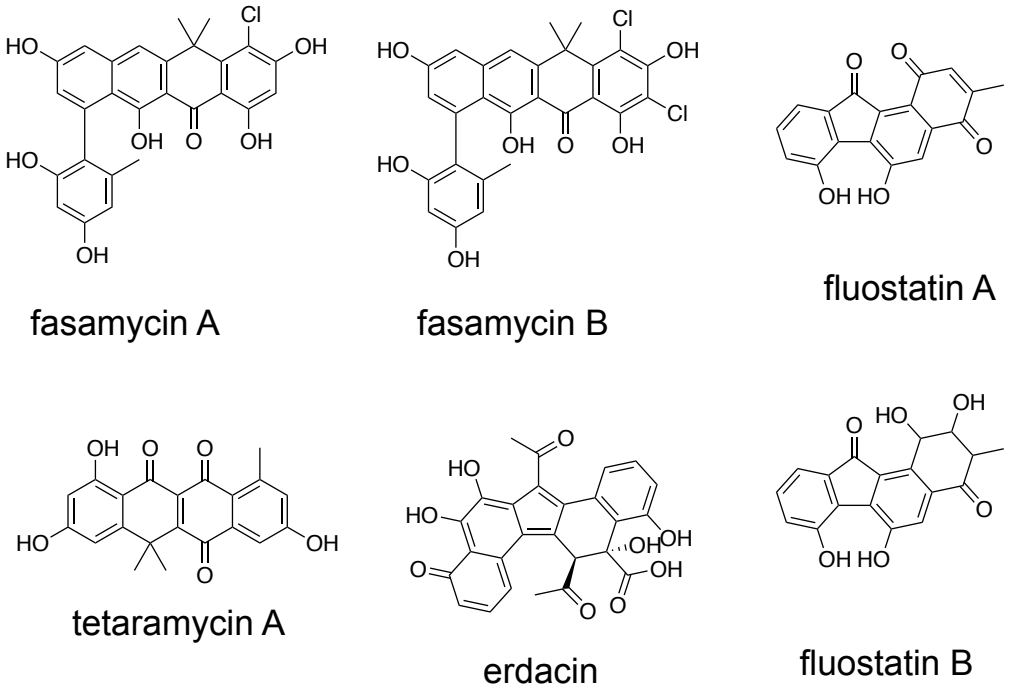


Figure 11: Metabolites isolated in homology-based metagenomic screening

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).

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/f1000 research-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 >.