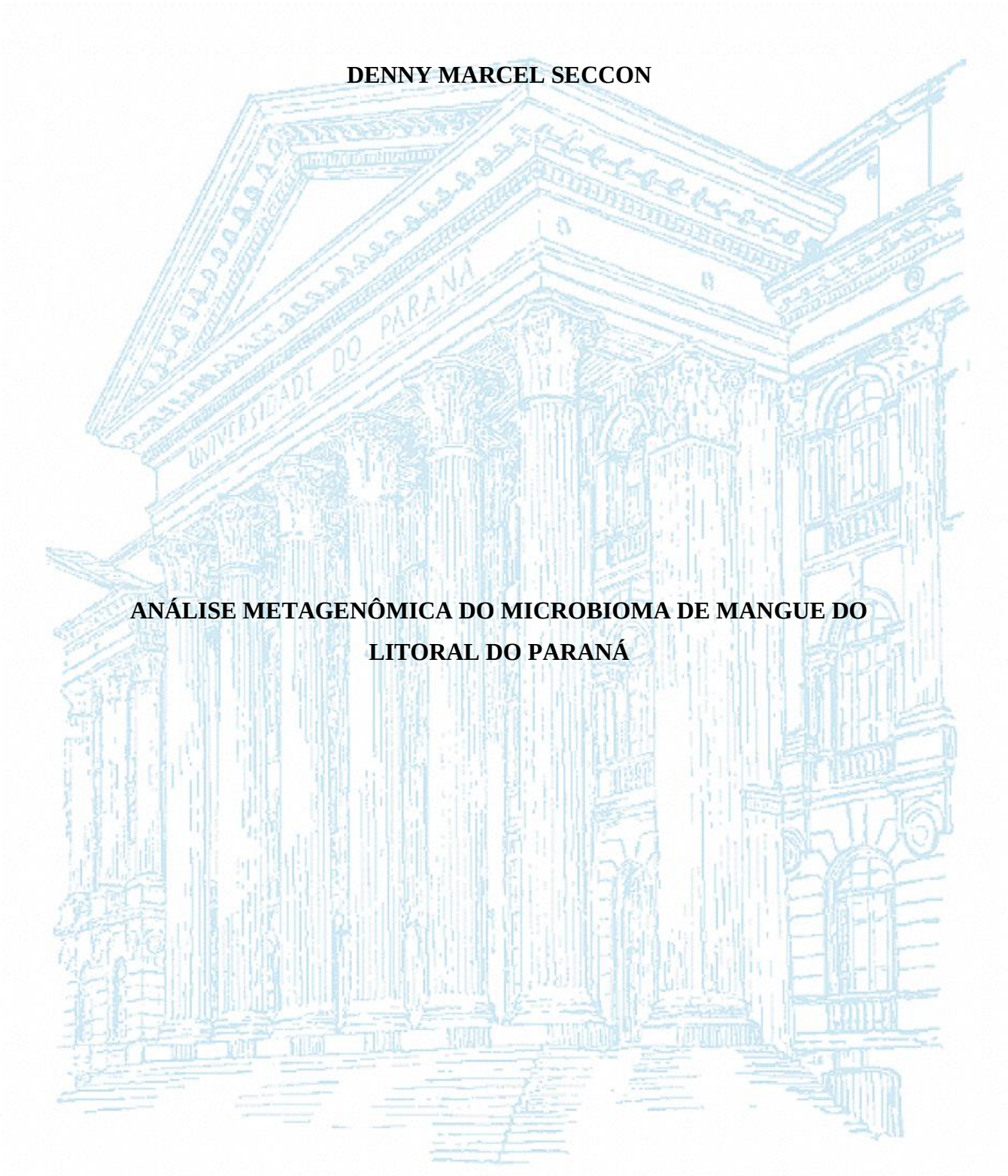
UNIVERSIDADE FEDERAL DO PARANÁ

DENNY MARCEL SECCON



ANÁLISE METAGENÔMICA DO MICROBIOMA DE MANGUE DO LITORAL DO PARANÁ

CURITIBA

2016

DENNY MARCEL SECCON

ANÁLISE METAGENÔMICA DO MICROBIOMA DE MANGUE DO LITORAL DO PARANÁ

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Ciências – Bioquímica, Curso de Pós-Graduação em Ciências – Bioquímica, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientador: Prof. Fabio Oliveira Pedrosa

Co-orientadores: Helisson Faoro
Profa. Roseli Wassem

CURITIBA

2016

TERMO DE APROVAÇÃO

DENNY MARCEL SECCON

ANÁLISE METAGENÔMICA DO MICROBIOMA DE MANGUE DO LITORAL DO PARANÁ

Tese aprovada como requisito parcial para obtenção do grau de Doutor em Ciências – Bioquímica no Curso de Pós-Graduação em Ciências – Bioquímica, Setor de Ciências Biológicas da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador: Prof. Fábio Oliveira Pedrosa
Departamento de Bioquímica e Biologia Molecular, UFPR

Profa. Emanuel Maltempi de Souza
Departamento de Bioquímica e Biologia Molecular, UFPR

Profa. Luciane Maria Pereira Passaglia
Departamento de Genética, UFRGS

Prof. Marcelo Müller dos Santos
Departamento de Bioquímica e Biologia Molecular, UFPR

Prof. Vasco Ariston de Carvalho Azevedo
Departamento de Biologia Geral, UFMG

Curitiba, 27 de setembro de 2016.

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

RESUMO

A descrição da penicilina revelou que microrganismos eram a fonte primária de muitos produtos naturais com aplicações farmacológicas. Sua bem-sucedida adoção medicamentosa foi seguida por um período prolífico de isolamento de novas drogas obtidas a partir de microrganismos encontrados no ambiente. Entretanto, a velocidade de novas descobertas diminuiu quando se observou que a grande maioria dos microrganismos não era prontamente cultivável em condições laboratoriais, ao mesmo tempo em que os organismos patogênicos alvo das drogas que vinham sendo usadas desenvolviam resistência. Logo ficou evidente que a empreitada deveria buscar novas estratégias. Pesquisadores passaram a estudar novas classes de metabólitos secundários, entre os quais chamou a atenção os produtos das sintetases de peptídeos não-ribossomais (NRPS) e sintases de policetídeos (PKS). Os módulos que compõem esses sistemas enzimáticos são formados por alguns domínios básicos e outros opcionais, que a partir de precursores simples – aminoácidos e acetato, respectivamente – são capazes de sintetizar moléculas com grande variedade estrutural, muitas com atividade biológica igualmente diversa. Outra frente de investigação se aproveitou dos avanços nas técnicas de isolamento e pesquisa independente de cultivo para explorar ambientes antes negligenciados, como o mar, que logo revelou hospedar uma nova gama de microrganismos e produtos derivados. Além de informações funcionais específicas, a exploração eficiente e sustentável desses ambientes exige uma melhor caracterização de suas comunidades microbianas; microrganismos são componentes essenciais de seu funcionamento, mas essa literatura ainda é escassa. Em virtude disso, esse trabalho propôs a investigação do microbioma de sedimentos de mangue da região da Baía de Paranaguá, Paraná, Brasil, através de metagenômica. No primeiro capítulo, usamos técnicas de sequenciamento de amplicons gerados a partir do gene de rRNA 16S para caracterizar a comunidade presente e determinar os fatores ambientais que moldam essa estrutura. Encontramos um predomínio dos filos *Proteobacteria, Bacteroidetes, Cyanobacteria, Chloroflexi* e *Acidobacteria,* quase todos significativamente afetando as estruturas intralocais. As classes mais abundantes foram *Deltaproteobacteria, Gammaproteobacteria* e *Alphaproteobacteria.* Revelamos forte influência da composição química derivada da proximidade do mar, sendo os fatores pH (SMP), H^++Al^{3+}, Ca^{2+}, Mg^{2+}, Na^+, T, P, Na% e Ca/Mg significativos, mas não encontramos efeito de presença de planta, possibilidade descrita na literatura. Comparamos essa comunidade com a de outros mangues e encontramos similaridades que apontam para um perfil característico dependente de bioma, dominado pelos mesmo filos abundantes encontrados no mangue paranaense. No segundo capítulo, realizamos o sequenciamento shotgun de duas amostras do mangue pertencentes a localizações quimicamente diversas para determinar seu perfil funcional. As funções prevalentes eram relacionadas a subsistemas baseados em formação de cluster, carboidratos, aminoácidos e derivados, e funções variadas. Em níveis hierárquicos mais baixos, a abundância relativa de funções é geralmente balanceada. As diferenças funcionais também se mostraram dependentes de fatores ambientais, mas essa influência é menor do que no perfil taxonômico, o que indica resiliência funcional frente a alterações externas. As similaridades do mangue frente a outros biomas se mostraram mais dependentes de tipo de matriz (onde solo > água) do que de composição química (onde salina > não-salina); as dissimilaridades funcionais também foram menores que as taxonômicas, mas sua intensidade manteve relação com o grau de diferença de ambiente. No terceiro capítulo, o sequenciamento e processamento de amplicons provindos dos domínios A e C de enzimas NRPS, e KS de enzimas PKS, confirmou o mangue estudado como fonte potencial de metabólitos secundários. As

medidas de diversidade não mostraram um padrão entre domínios enzimáticos, mas sugerem os melhores grupos para prospecção: verão | localização B para domínios A, inverno | localização B para domínios C, e inverno | localização A para domínios KS. Apesar de muitas sequências encontradas terem sido taxonomicamente afiliadas a gêneros reportadamente envolvidos na síntese desses metabólitos secundários, a descoberta de novos produtores é esperada dado o número de OTUs reportados e a presença de muitas sequências relacionadas a bactérias não caracterizadas. Os resultados apresentados devem servir de base para futuros estudos de prospecção direcionada.

ABSTRACT

The description of penicillin revealed that microorganisms were the primary source of many natural products with pharmacological application. Its successful adoption as a medicament was followed by a prolific period in the isolation of new drugs obtained from microorganisms found in the environment. However, the pace of new discoveries decreased when it was noticed that the vast majority of microorganisms was not readily cultivable in laboratorial conditions, while the pathogenic organisms targeted by drugs in use developed resistance. It soon became evident that the enterprise should try new strategies. Researchers started studying new classes of secondary metabolites, among which much attention was given to the products of non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). The modules building these enzymatic systems are formed by some basic and optional domains, which, from simple precursors – amino acids and acetate, respectively – are able to synthetize molecules with an incredible structural variety, many with equality diverse biological activity. Another front of investigation took advantage of the advances in isolation and cultivation-independent research techniques to explore environments previously neglected, like the ocean, which soon revealed to host a whole new range of microorganisms and derived products. Besides specific functional information, the efficient and sustainable exploration of such environments demands a better characterization of their microbial communities; microorganisms are essential components of their functioning, yet this literature is scarce. Following from these, this work proposed investigating the microbiome of mangrove sediments from the region of Paranaguá Bay, Paraná, Brazil, through metagenomics. In the first chapter, we used 16S rRNA gene-amplicon sequencing to characterize the community present and to determine the environmental factors shaping this structure. We found a predominance of phyla *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Chloroflexi* and *Acidobacteria*, almost all significantly affecting the intralocal structures. The most abundant classes were *Deltaproteobacteria*, *Gammaproteobacteria* and *Alphaproteobacteria*. We revealed a strong influence of the chemical composition derived from the proximity to sea, being significant the factors pH (SMP), H^{+}+Al^{3+}, Ca^{2+}, Mg^{2+}, Na^{+}, T, P, Na% and Ca/Mg, but we found no plant effect, a possibility described in the literature. We compared this community with those of other mangroves and found similarities that point to a biome-dependent characteristic profile, dominated by the same abundant phyla found in the mangrove from Paraná. In the second chapter, we performed the shotgun sequencing of two mangrove samples from different environments to determine their functional profile. Prevailing functions were related to clustering-based subsystems, carbohydrates, amino acids and derivatives, and miscellaneous functions. In lowest hierarchy levels, the relative abundance of functions is generally balanced. The functional differences were also dependent on environmental factors, but this influence was smaller than in the taxonomic profile, which indicates functional resilience in front of external changes. The similarities of mangroves to other biomes were more dependent on matrix type (where soil > water) than on chemical composition (where saline > non-saline); the functional dissimilarities were also smaller than the taxonomical, but the intensity kept its relation to the degree of environmental difference. In the third chapter, sequencing and processing of amplicons derived from A and C domains from NRPS, and KS from PKS, confirmed the studied mangrove as a potential source of secondary metabolites. The diversity measures showed no pattern among the enzymatic domains, but suggest the most suitable groups for prospection: summer | location B for A domains, winter | location B for C domains, and winter | location A for KS domains. Even though many sequences were taxonomically affiliated to genera reportedly involved in the

production of such secondary metabolites, the discovery of new producers is expected given the number of OTUs reported and the presence of many sequences affiliated to uncharacterized bacteria. The results presented will be the basis for future studies of directed prospection.

INTRODUÇÃO

A busca por antibióticos

A descoberta e introdução de antibióticos para o tratamento de doenças na década de 1940 foi um marco. Apesar de as drogas pioneiras dessa revolução, penicilina e estreptomicina, serem ainda importantes agentes farmacológicos, seu uso disseminado logo revelou que elas não eram a solução definitiva como seu sucesso inicial fez parecer: os microrganismos nos quais elas agem desenvolvem resistência. Assim, ficou evidente que a busca por novos antibióticos precisa ser constante. Por volta da década de 1960, muitos pesquisadores foram bem-sucedidos em descobrir novas moléculas derivadas de fontes naturais: 20 novas classes de antibióticos foram comercializadas no período (Coates, Halls e Hu, 2011), dentre elas as tetraciclinas, as cefalosporinas, os aminoglicosídeos e os macrolídeos. Entretanto, limitações nas técnicas para o isolamento e cultivo dos microrganismos produtores levou a iniciativa a um novo entrave. Desde a descoberta de lipopeptídeos em 1987, o lançamento de novas drogas dependeu quase que exclusivamente de desenvolvimento sintético ou da modificação de moléculas anteriores (Payne *et al.*, 2007), mas sua eficiência nunca se comparou à dos produtos naturais; por volta do ano 2000, essa estratégia se mostrou insuficiente para competir com a emergência de bactérias resistentes (Coates, Halls e Hu, 2011). Uma revisão recente aponta que todo antibiótico em uso clínico tem resistência conhecida (Banskota *et al.*, 2006), e o surgimento de bactérias multirresistentes apresenta sérias ameaças à saúde pública.

Novas tecnologias recentemente desenvolvidas trouxeram fôlego renovado a essa empreitada vital. Técnicas inovadoras de cultivo levaram ao isolamento de espécies até então não caracterizadas, o que possibilitou a descoberta da teixobactina (Ling *et al.*, 2015), a primeira nova classe de antibióticos em três décadas, sem resistência conhecida. A investigação de novos ambientes levou à descoberta da lugdunina (Zipperer *et al.*, 2016), outra molécula de alta eficiência obtida de uma bactéria comensal humana, o primeiro reporte do tipo. Esses resultados animam as principais vias atuais de prospecção *in silico*, a mineração genômica e a exploração metagenômica, que por sua vez ampliaram seu campo de investigação para incluir novos ambientes e novas classes de moléculas.

Metabólitos secundários

Produtos naturais são substâncias que tradicionalmente representam um dos mais importantes ramos da química orgânica. Eles podem ser derivados do metabolismo primário ou, mais comumente, secundário de organismos vivos. Os metabólitos primários (polissacarídeos, proteínas, ácidos nucleicos e ácidos graxos) são comuns em todos os sistemas biológicos. Os metabólitos secundários, entretanto, são química e taxonomicamente diversos e muitos têm função obscura (Bérdy, 2005), não possuindo um papel explícito na economia interna do organismo que os produz (Williams *et al.*, 1989). Essas moléculas são geralmente produtos distintivos de grupos particulares de organismos, algumas vezes mesmo de uma única cepa. Muitas, se não estruturalmente monoméricas, são oligômeros de massa molecular não maior que 3000 Da. A maioria é secretada pelos organismos produtores (Vining, 1990).

Quando primeiro descritos, os metabólitos secundários foram classificados como não funcionais. Entretanto, apesar de não conferirem funções metabólicas essenciais, organismos evoluíram a habilidade de produzir tais compostos devido às vantagens seletivas que eles conferem. Em alguns casos, por exemplo, essas moléculas podem atuar como sinais para desencadear processos de diferenciação celular, agindo como análogos de hormônios; em outros, podem servir à comunicação em processos de simbiose (Beppu, 1992). Em ambientes onde há competição, podem repelir outros organismos com diferentes intensidades, causando até morte celular (Stone e Williams, 1992). Nesse caso, organismos sem um sistema imunológico são prolíficos produtores de metabólitos secundários, os quais atuariam como um mecanismo de defesa alternativo. A ampla e sofisticada variedade estrutural desses produtos e modos de ação requer vias biossintéticas igualmente complexas, altamente coordenadas, as quais apresentam custo energético alto e, por isso, não teriam evoluído a menos que seus produtos tivessem funções vantajosas. Muitos metabólitos secundários são produzidos naturalmente, e não como artefatos de condições de cultivo, e apresentam atividade biológica derivada de interações específicas com receptores aos quais apresentam alta complementaridade. A organização em clusters de genes codificando a biossíntese, regulação e auto resistência a essas moléculas é evidência molecular em favor da evolução de metabólitos secundários funcionais.

Antes da descrição da penicilina, poucos produtos naturais microbianos foram descobertos em decorrência de sua atividade biológica. A percepção dessa função mudou, com sucesso, o foco de investigação: a pesquisa intensiva de microrganismos por substâncias de valor medicinal revelou uma ampla coleção de atividades biológicas apresentadas por metabólitos secundários. Atualmente, há vasta documentação sobre sua atividade tóxica, inseticida, nematicida, farmacológica e imunológica, além da antibiótica primeiramente reportada (Vining, 1990), sendo comumente empregados em terapia humana, veterinária e na agricultura. Essa gama funcional se deve à forma oportunística com que foram incorporados à maquinaria celular, cada um conferindo sua própria vantagem particular, e à variedade de composição estrutural. Estima-se que a eles é devida 40% da diversidade de produtos naturais, e que pelo menos metade nunca foi sintetizada artificialmente (Feher e Schmidt, 2003).

NRPS e PKS

Entre os metabólitos secundários com ampla variedade estrutural e atividade biológica, duas classes que vêm recebendo atenção são os produtos das sintetases de peptídeos não-ribossomais (NRPS) e das sintases de policetídeos (PKS), gerados a partir da condensação, elongação e modificação de precursores simples – aminoácidos e acetato, respectivamente. Apesar de usarem mecanismos químicos distintos para a manipulação de seus substratos, esses sistemas enzimáticos apresentam similaridades marcantes na arquitetura de seus domínios catalíticos e em seus mecanismos de linha de montagem. Ambos os sistemas são formados por módulos, cada um responsável por um passo discreto na cadeia de elongação, e a ordem específica dos módulos define, em princípio, a sequência de incorporação de blocos unitários. Os módulos, por sua vez, são compostos por domínios proteicos que lhes conferem especificidade.

Um módulo típico de NRPS tem em sua configuração mínima um domínio de adenilação (A) responsável pela ativação do aminoácido, um domínio de tiolação (T), também conhecido como proteína carreadora de peptidil (PCP), à qual o aminoácido ativado é covalentemente ligado, e um domínio de condensação (C) que catalisa a formação da ligação peptídica entre os substratos do módulo atual e do módulo anterior. O módulo de PKS contém três domínios análogos: um domínio aciltransferase (AT) para seleção e transferência da unidade de extensão à proteína carreadora de acil (ACP), e um domínio sintase de cetoacil (KS) para a condensação descarboxilativa da unidade

extensora com um tio éster acil. Nos dois casos, um domínio de terminação (TE) é responsável por liberar os produtos finais de síntese (Figura 1).

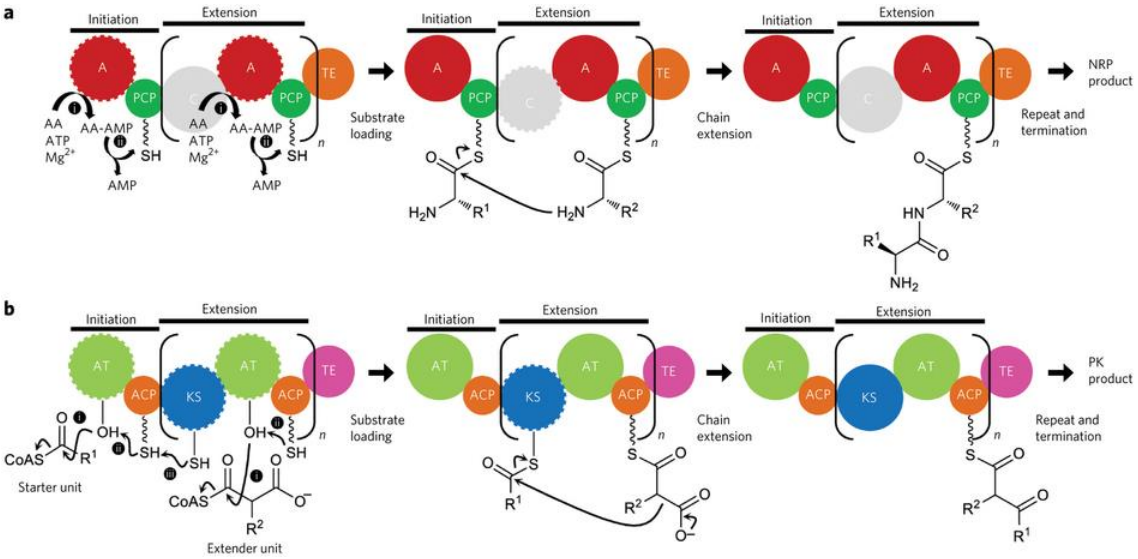


Figura 1. Estrutura básica dos módulos de iniciação e extensão de a) NRPS e b) PKS. Fonte: Weissman, 2014.

Os fatores que fazem com que os produtos de NRPS e PKS tenham uma grande diversidade estrutural são a associação combinatória de seus módulos, a possibilidade de operação modular iterativa resultando em produtos elongados ou de serem ignorados para gerar produtos encurtados, e a capacidade de interagir de forma não-linear (Wenzel e Müller, 2005). Adicionalmente, uma variedade de domínios opcionais é responsável por modificações diversas. Um novo nível de diversidade é possibilitado pela existência de sistemas NRPS-PKS híbridos, cujos mecanismos só recentemente passaram a ser investigados (Fisch, 2013).

A relevância desses complexos enzimáticos para a pesquisa de metabólitos secundários pode ser ilustrada pelo lançamento recente de bases de dados curadas como antiSMASH (Medema *et al.*, 2011), DoBISCUIT (Ichikawa *et al.*, 2012) e ClusterMine360 (Conway e Boddy, 2013), que buscam organizar o grande número de informações disponíveis em catálogos de domínios, módulos, clusters gênicos e produtos de síntese.

Prospecção de metabolismo secundário no ambiente

Apesar dos avanços nas técnicas de cultivo, a maioria dos microrganismos continua distante do isolamento. As técnicas moleculares são reconhecidas por terem tornado a exploração desse universo oculto possível, mas além disso, seu emprego aumentou a eficiência da empreitada por novos produtos naturais: antes mesmo de escrutinizar os produtos de um organismo – um processo trabalhoso e demorado –, a busca por seus genes produtores permite uma investigação mais direcionada (Payne *et al.*, 2007). Os avanços nas técnicas de bioinformática consolidaram a mineração genômica como ferramenta inovadora de exploração do potencial metabólito de organismos individuais (Winter, Behnken e Hertweck, 2011, Corre e Challis, 2009). Mas foi o estabelecimento da metagenômica que permitiu ampliar as possibilidades de pesquisa para incluir, por exemplo, o estudo de genomas de todos os organismos de um ambiente.

Estudos metagenômicos envolvem, basicamente, a extração e análise do DNA total de uma amostra ambiental. Estudos *in silico* dirigidos, onde os produtos de PCR de sequências alvo do metagenoma são sequenciados, possibilitam a análise do alcance de diferentes ortólogos daquela sequência dentro de uma comunidade, o que resulta num perfil da estrutura dessa comunidade. Estudos *in silico* não-dirigidos, chamados de sequenciamento aleatório "shotgun", se baseiam no sequenciamento integral do DNA ambiental, gerando um perfil dos genomas presentes (Gilbert e Dupont, 2011), que pode ser base tanto de inferências taxonômicas quanto funcionais. Se, por um lado, as sequências apresentam menor profundidade de sequenciamento e são mais fragmentadas que nos projetos de sequenciamento genômico, por outro se obtêm informações de um conjunto muito maior de organismos habitantes, o que pode direcionar estudos posteriores mais focalizados sem a necessidade de sequenciar todos os genomas individuais presentes.

Prospecção de metabolismo secundário no mangue

A importância de microrganismos terrestres como fonte de metabólitos bioativos é bem estabelecida por mais de meio século. Devido à facilidade de obtenção, eles foram os primeiros microrganismos a serem levados ao laboratório. Mas tão logo seu isolamento se mostrou um trabalho árduo e pouco eficiente, com diversidade limitada, novos

ambientes começaram a ser investigados. Além disso, a exploração de novos ambientes prometia novidade: ecossistemas únicos possuem organismos únicos com vias metabólicas únicas.

O início do interesse de cientistas pela busca de drogas no ambiente marinho não é recente, podendo ser oficialmente atribuído ao simpósio "Drugs from the Sea", acontecido nos Estados Unidos em 1967 (Fusetani, 2000). O isolamento alguns anos antes da cefalosporina, produzida por um fungo encontrado em água marinha, foi o primeiro sinal de que muitos metabólitos obtidos de algas e invertebrados seriam na verdade produzidos não por esses organismos mais complexos, mas por microrganismos associados. Como consequência, na década de 1990, o número de moléculas de origem marinha descobertas dobrou em relação ao período anterior, mas o aumento na fração correspondente àquelas produzidas por microrganismos foi de dez vezes (Kelecom, 2002). Até o ano 2000, aproximadamente 10000 metabólitos foram isolados (Fusetani, 2000), muitos com propriedades farmacológicas de amplo espectro de atividade: antibiótica, antifúngica, tóxica, citotóxica, neurotóxica, antimitótica, antiviral e antineoplásica (Kelecom, 2002). A adição recente de técnicas moleculares modernas envolvendo bioinformática e tecnologia recombinante foi instrumental no aprofundamento do entendimento e apreciação das vias microbianas marinhas complexas. Várias vias biossintéticas em bactérias marinhas e cianobactérias foram elucidadas pela primeira vez em nível molecular, incluindo aquelas de policetídeos, peptídeos e moléculas híbridas (Moore, 2005), abrindo a porta para novas oportunidades de pesquisa envolvendo engenharia metabólica e biocatálise.

Mangues são florestas altamente produtivas confinadas às regiões tropicais e subtropicais (Figura 2), onde a temperatura elevada favorece seu desenvolvimento (Alongi, 2002). O Brasil ocupa a segunda posição no ranking de maiores florestas (Spalding, Kainuma e Collins, 2010). Mangues possuem uma associação complexa e rica de espécies associadas, dando suporte a uma rede alimentar aquática complexa. São importantes refúgios para animais marinhos e terrestres, estabilizam costas e trocam nutrientes com ecossistemas adjacentes. Comunidades humanas em seu entorno fazem uso de mangues como provedores de material para construção, lenha, carvão, alimentos e medicamentos (Yan, Hong and Yu, 2006; Alongi, 2002).

Estando presentes na interface terrestre e oceânica, os mangues apresentam uma comunidade microbiana caracterizada por adaptações necessárias para a sobrevivência nos dois ambientes, o que lhes confere vias metabólicas únicas (Thatoi *et al.*, 2013). Não

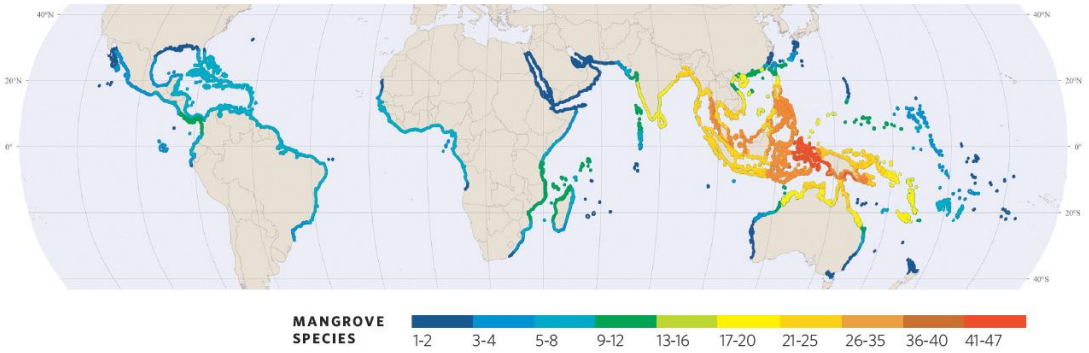


Figura 2. Distribuição global de mangues, e respectiva diversidade vegetal. Fonte: US Department of Commerce (2014).

por acaso, 23 novos metabólitos obtidos de bactérias associadas com mangues foram reportados em 2014, mais que o dobro do período anterior, entre eles a mangromicina A, exibindo potente atividade antitripanossômica e sequestrante de radicais (Nakashima *et al.*, 2014) e moléculas obtidas dos gêneros *Streptomyces, Jishengella, Micromonospora* e *Verrucosispora* (revisados em Blunt *et al.*, 2016). Produtos nitrogenados, incluindo peptídeos cíclicos, e derivados de acetato constituem as cinco principais classes de metabólitos encontrados em microrganismos marinhos, perfazendo mais de 60% do total. Quando a atividade biológica de produtos bacterianos é investigada, 45% tem ação antibiótica, sendo que desse total, 38% foram obtidas de sedimentos (Kelecom, 2002). Isso faz com que sedimentos marinhos sejam extremamente propícios para prospecção de metabólitos com atividade biológica. Entretanto, a exploração racional do mangue depende de um melhor esclarecimento sobre o frágil equilíbrio que mantém sua eficiência produtiva. Esse habitat é deficiente em alguns nutrientes como nitrogênio e fósforo, mas isso é compensado por um eficiente mecanismo de reciclagem, que depende de uma relação estreita entre microrganismos, plantas e o solo (Feller *et al.*, 2010); ainda não é claro, entretanto, como esses fatores definem a estrutura dessas comunidades (Nogueira *et al.*, 2015; Andreote *et al.*, 2012).

Apesar dos vários estudos sobre a biogeografia, a botânica, a zoologia, a ictiologia, a poluição ambiental e o impacto econômico, pouco se sabe sobre a vida microbiana em águas e sedimentos de mangue (Andreote *et al.*, 2012; Sahoo e Dhal, 2009; Gomes *et al.*, 2008). Esse levantamento e a subsequente comparação entre diferentes mangues pode permitir a descrição de um microbioma central composto por membros compartilhados, os quais podem ser críticos no funcionamento dessas comunidades. A definição de um microbioma central é o primeiro passo na definição de um ambiente

sadio, da onde é possível prever respostas a distúrbios (Shade e Handelsman, 2012) e desenvolver estratégias com interesse biotecnológico.

Este trabalho

A justificativa para este trabalho se baseou no evidente potencial biotecnológico presente no metabolismo secundário do mangue, cuja estrutura microbiana segue em grande parte desconhecida, sua elucidação podendo ainda agregar às estratégias voltadas a sua conservação. O objetivo principal foi utilizar técnicas metagenômicas para definir o perfil taxonômico e funcional desse ambiente, e os fatores ambientais que dirigem essa constituição, construindo a base conceitual necessária para estudos posteriores. Secundariamente, sugerimos a comparação dos perfis encontrados com outros mangues e outros biomas, de forma a apontar similaridades e diferenças que também podem ser atribuídas a características ambientais compartilhadas. Por fim, investigamos a diversidade de domínios proteicos relacionados à síntese de metabólitos secundários por enzimas NRPS e PKS, de forma a quantificar o potencial presente para síntese dessas moléculas, o que será útil na definição de futuras estratégias de prospecção direcionada.

Amostras de sedimento de mangue foram obtidas da região da Baía de Paranaguá, Paraná, Brasil, de acordo com três fatores ambientais: estação do ano (verão e inverno), região do estuário (mais próxima e mais distante do mar aberto) e proximidade das plantas prevalentes na região (*Avicennia schaueriana, Laguncularia racemora, Rhizophora mangle*, *Spartina alterniflora* e uma região livre de vegetação). As vinte condições, representadas individualmente por um homogeneizado de três pontos equidistantes até 30 cm ao redor do local de coleta, foram obtidas em triplicata, somando 60 amostras.

No Capítulo 1, objetivando a caracterização do microbioma, realizamos o sequenciamento de amplicons do gene de rRNA 16S das amostras obtidas no verão, revelando um perfil estrutural que pôde ser comparado com outros mangues para a definição de uma estrutural central; também pudemos avaliar a relevância dos fatores ambientais na definição dessa estrutura. No Capítulo 2, escolhemos duas amostras representativas de dois fatores ambientais que definem a estrutura da comunidade para realizar o sequenciamento integral de seus respectivos metagenomas, construindo um perfil funcional que também adiciona à caracterização pendente desse habitat, e a partir dele comparamos o mangue com outros biomas com diferentes graus de similaridade, reportando as relações encontradas. No Capítulo 3, buscando dimensionar o potencial biotecnológico presente no metabolismo secundário do mangue, sequenciamos amplicons correspondentes a sequências dos domínios proteicos A e C de enzimas NRPS, e KS de enzimas PKS, revelando a diversidade desses complexos catalíticos no ambiente, o que

representa o primeiro passo na definição de estratégias de prospecção focalizada de metabólitos secundários no mangue.

CAPÍTULO 1 – Fatores ambientais formatam a estrutura da comunidade do microbioma em sedimentos de mangue na costa do sul do Brasil

Environmental factors shape microbiome community structure in mangrove sediments in the Southern Brazilian Coast

Abstract

Mangroves are highly efficient ecosystems in which a diverse array of microorganisms interact with the few plant species present to maintain an economic energy flux, in which recycling is crucial. Being present in the interface between land and sea, its microbial community has evolved to cope with its very characteristic features: varying degrees of salinity and constant anaerobicity. Despite its ecological and economical importance, knowledge about the microbial world inhabiting mangroves and factors affecting its structure is still scarce. We collected samples from the sediments of mangroves in Paranaguá Bay, Brazil, in two locations representing the extremes of a salinity gradient; within each spot, samples were taken in the proximity to four prevailing plants and in an open area. Location was a strong factor shaping community; the chemical characterization of the sediments showed that many components usually associated with soil nutritional status drive community patterns. We did not observe a plant effect. The mangrove showed a prevalence of phyla *Proteobacteria*, *Actinobacteria*, *Chloroflexi*, *Bacteroidetes*, *Acidobacteria* and *Cyanobacteria*, which are commonly associated with soils; furthermore, the comparison with other mangrove areas showed similar relative abundances that suggest a specific mangrove profile. The consolidation of such a pattern signature will be helpful as a fingerprint of the mangrove among other biomes and as a fitness indicator.

Introduction

Mangrove ecosystems are wetlands present in tropical and sub-tropical coastal regions, covering 60-70% of its coastlines worldwide (Thatoi *et al.*, 2013). Their greatest areal extent is found in the estuarine areas of the tropics (Saseeswari, Kanimozhi and Panneerselvam, 2016). Brazil harbors 8.5% of all mangrove forests, ranking second after Indonesia (Spalding, Kainuma and Collins, 2010); most areas are present in the north region, while in the south, they are sparsely found in river deltas, coastal lagoons and the inner parts of bays (Spalding, Blasco and Field, 1997).

Being a transition zone between earth and sea, they are characterized by gradients in salinity caused by high regular rainfall, groundwater flows and rivers diluting seawater to different extents; and by anaerobic soils except for the sediment surface, since they are constantly submerged (Bouchez *et al.*, 2013; Duke, Ball and Ellison, 1998). These contrasts make them heterogeneous habitats. In estuaries, the spatial distribution of species might be classified into downstream, closer to open sea, where salinity and tide variation are higher, thus sediment deposition and nutrient exchange are more dynamic; and upstream, the upper part, where rivers flow into and organic matter content is higher (Duke, Ball and Ellison, 1998). The nutrient availability is variable within and among mangrove ecosystems, ranging from extremely low in oceanic settings to very high in accreting muddy systems; it can vary spatially along tidal gradients and temporally with season and year (Feller *et al.*, 2010).

The most prevalent mangrove trees found in Brazil account for only seven species (Spalding, Blasco and Field, 1997). In Paranaguá Bay, located in the southern Brazilian coast, four species are dominant: *Avicennia schaueriana, Laguncularia racemosa, Rhizophora mangle* and *Spartina alterniflora.* Even though tree diversity is low, the microbial diversity can be high (Andreote *et al.*, 2012). Bacteria seem to have greater importance: they tend to be prevalent (Nogueira *et al.*, 2015) for, due to their small size, being able to colonize a greater variety of niches (Ghizelini, Mendonça-Hagler and Macrae, 2012), and their trophic functions therein are more relevant than in non-tropic habitats (Alongi, 1994). In sediments, high bacterial activity coincide with the presence of plants (Holguin *et al.*, 2001). Plant-microbe-soil relations are tightly linked (Feller *et al.*, 2010) and help conserve scarce nutrients through cycling (Alongi, 2009, Lovelock, 2008, Holguin *et al.*, 2001). A diverse microbial community continuously transforms dead vegetation in nutrients that are assimilated by plants (Feller *et al.*, 2010; Reef, Feller and Lovelock, 2010; Holguin *et al.*, 2001). In exchange, plant factors such as root exudates influence rhizosphere communities by serving as a nutrient source for microorganisms (Holguin *et al.*, 2001). The physicochemical properties of the soil also influence microbes, and since plants can alter these, all factors may interact. It is not yet quite clear how these factors define the structure of such communities (Nogueira *et al.*, 2015; Andreote *et al.*, 2012; Marschner, Crowley and Yang, 2004).

Natural benthic communities are stable as long as they are not disturbed (Holguin *et al.*, 2001), thus eutrophication has negative consequences for mangrove growth (Reef, Feller and Lovelock, 2010, Spalding, Kainuma and Collins, 2010). The major problem in

predicting mangrove responses to human impact is the lack of long-term data, and the ability to distinguish natural from anthropogenic change. Many environmental projects have been launched to promote research activities and the conservation, rational management and sustainable utilization of mangroves (Spalding, Blasco and Field, 1997). Conservation strategies for mangroves should consider the ecosystem as a biological entity, which includes all the processes that maintain productivity, functionality and resilience. Despite of various studies on the biogeography, botany, zoology, ichthyology, environmental pollution and economic impact, little is known about the microbial life in mangrove waters and sediments (Andreote *et al.*, 2012, Sahoo and Dhal, 2009; Gomes *et al.*, 2008). The comparison of mangroves from different geographic regions may enable the description of shared members (core microbiome), which might be critical to the community function. Recognizing this signature structure is the first step to define a healthy environment and to predict responses to disturbance (Shade and Handelsman, 2012).

We herein describe the microbial communities inhabiting mangroves in the southeastern coast of Brazil. We consider environmental factors such as chemical composition of the sediments and presence of plants to determine their influence in community structure. By comparing our results with other studies, we contribute to the generation of a microbiome fingerprint of the habitat.

Materials and methods

Sample extraction and treatment

Samples were extracted from the sediments of mangroves close to Paranaguá Bay, Paraná, Brazil, during the summer season (March 14th, 2014), from two locations within the estuary (level I), upstream and downstream. Within each location, three sampling spots (level III) were chosen, so that we had triplicates for each environmental factor tested (Figure 1). Within each spot, the individual samples were retrieved from the

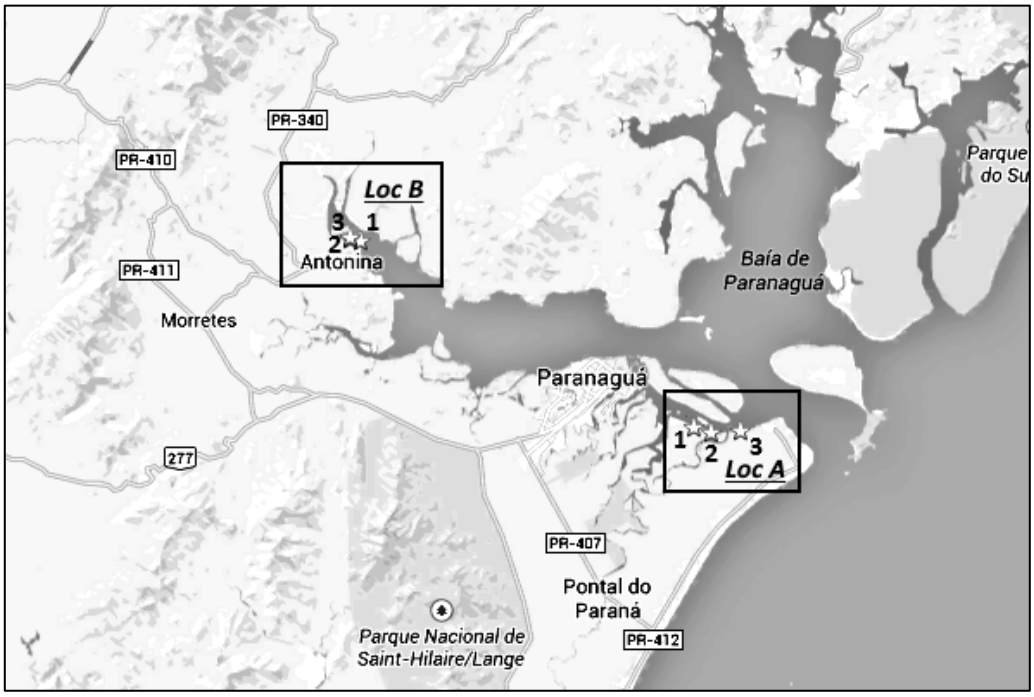


Figure 1. Sampling scheme. Source: Google Maps.

surroundings (up to 30 cm away from the roots) of the plants evaluated, as well as from an area free of vegetation (level II). Samples were coded with a four-digit name given by the letter V (standing for Summer)+(I)+(II)+(III). For example, the sample taken from the second spot in the upstream region around *S. alterniflora* got the code VBS2. Table 1 shows the information regarding the levels of sampling, the codes used (in bold) and their locations. The upper layer of sediment (around 10 cm) was discarded and around 200 grams of the surfaced material was aseptically collected and stored in ice until processing.

Table 1. Sampling levels (I, II, III), codes (in bold) and coordinates.

Estuary region (I)	Vegetation nearby (II)	Spot (III)	Coordinates
Upstream **(B)**	*Avicennia schaueriana* **(A)** *Laguncularia racemosa* **(L)** *Rhizophora mangle* **(R)**	**1**	25°25'02.5"S 48°42'29.9"W
		2	25°24'59.8"S 48°42'57.2"W
		3	25°30'31.4"S 48°28'07.4"W
Downstream **(A)**	*Spartina alterniflora* **(S)** Non-vegetated **(N)**	**1**	25°33'04.5"S 48°26'35.2"W
		2	25°33'19.1"S 48°25'46.1"W
		3	25°33'15.6"S 48°24'20.2"W

Environmental characterization

For the chemical characterization of the sediments, 100 grams of sample was quantified for the following routinely analyzed parameters to assess nutritional status of soils: pH (in $CaCl_2$ 0.01M), pH (by the SMP method), $H^+ + Al^{3+}$, Ca^{2+}, Mg^{2+}, K^+, Na^+, SB (sum of bases, given by $Ca^{2+} + Mg^{2+} + K^+ + Na^+$), T (cationic exchange capacity, given by $SB + H^+ + Al^{3+}$), P, C, V% (percent base saturation, given by SB/T), Na% (percent sodium saturation, given by Na^+/T) and Ca/Mg. The analyzes were performed by the Laboratory of Soil Chemistry and Fertility, from the Department of Soils and Agricultural Engineering, from Universidade Federal do Paraná, Brazil, according to Marques and Motta (2003).

DNA extraction, 16S rRNA amplification and sequencing

DNA was extracted from the samples using the PowerSoil DNA Isolation Kit (MO Bio Laboratories). For the 16S rRNA amplification, a PCR protocol was conducted. The system of 10 µL, containing 10 ng DNA, 1 µM primer 515F, 1 µM primer 806R which included one specific barcode for each sample (as described in Caporaso *et al.*, 2012), and 5 µL KlenTaq DV ReadyMix (Sigma-Aldrich), was submitted to the cycle 94ºC 3 min, 18x (94ºC 45 s, 50ºC 30 s, 68ºC 1 min), 72ºC 10 min. The resulting amplicons were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies) and their quality was assessed in the 2100 Bioanalyzer (Agilent). Sequencing was performed in the MiSeq platform using the MiSeq v2 Reagent Kit 500 Cycles PE (Illumina).

Reads processing

The sequences obtained were used for diversity assessment. For such, the R1 reads were processed according to Pylro *et al.* (2014). Briefly, the reads were quality filtered, chimeras were discarded, the remaining reads were clustered into OTUs, which had then taxonomy assigned to. The pipeline was run using softwares Qiime v1.8.0 (Caporaso *et al.*, 2010) and Usearch v8.0.1623 (Edgar, 2010) and the toolkit Fastx v0.0.14 (not published, available at http://hannonlab.cshl.edu/fastx_toolkit/). Taxonomy was derived from the 16S Greengenes database v13.8 (DeSantis *et al.*, 2006). At the end of this process, we obtained an OTU table, listing the sample counts of all OTUs found with at

least 97% identity, and a phylogenetic tree relating them, which were used in the next step.

Statistical analysis of diversity

The statistical analyses were performed in R v3.3.0 (R Core Team, 2016), using the packages phangorn (Schliep, 2011) and GUniFrac (Chen, 2012). The first is used to root the phylogenetic tree obtained previously. The second uses the rooted tree and the OTU table to calculate the distances between each pair of samples in GUniFrac units; we worked with the d(0.5) matrix obtained, since it is more robust and fitter to test the association of microbiome composition with environmental covariates, once it avoids over- or underestimation of abundant or rare OTUs thus preventing missing important findings (Chen *et al.*, 2012). Previously to this, each sample was subsampled by the script to normalize the number of reads at the minimum sample read count. The statistical significance of all environmental factors was assessed by PERMANOVA (n=999). A PCoA plot was designed based on the distance matrix, to show a graphical representation of the relationship among the samples. The environmental factors and taxa that showed statistical significance in shaping community were fitted into the ordination by function envfit from package vegan (Oksanen *et al.*, 2016).

Results

Phylogenetic affiliation of the samples

The analysis was performed on 2,066,259 reads, 240-bp long each, with counts ranging from 37,192 to 347,721 per sample, rarefied at the minimum sample count. Figure 2 presents a boxplot with the 15 most abundant taxa in the mangrove studied, from phylum to genus level. Among all samples, the proportion of *Bacteria* is (95.3±2.3)% (given as mean ± sd), *Archaea* sums (1.7±1.2)%, and (2.9±1.3)% cannot be assigned to any organism. At the phylum level (sample relative abundances represented at Supplementary Figure 1), the samples contain (53.5±4.2)% *Proteobacteria*, followed by (9.6±2.4)% *Bacteroidetes*, (5.4±2.8)% *Cyanobacteria*, (3.8±1.3)% *Chloroflexi*, (3.8±1.2)% *Acidobacteria* and (3.6±1.2)% *Planctomycetes*, which sum 79.6% of the sequences. The next phylum in relative abundance, (2.9±1.3)%, could not be assigned to

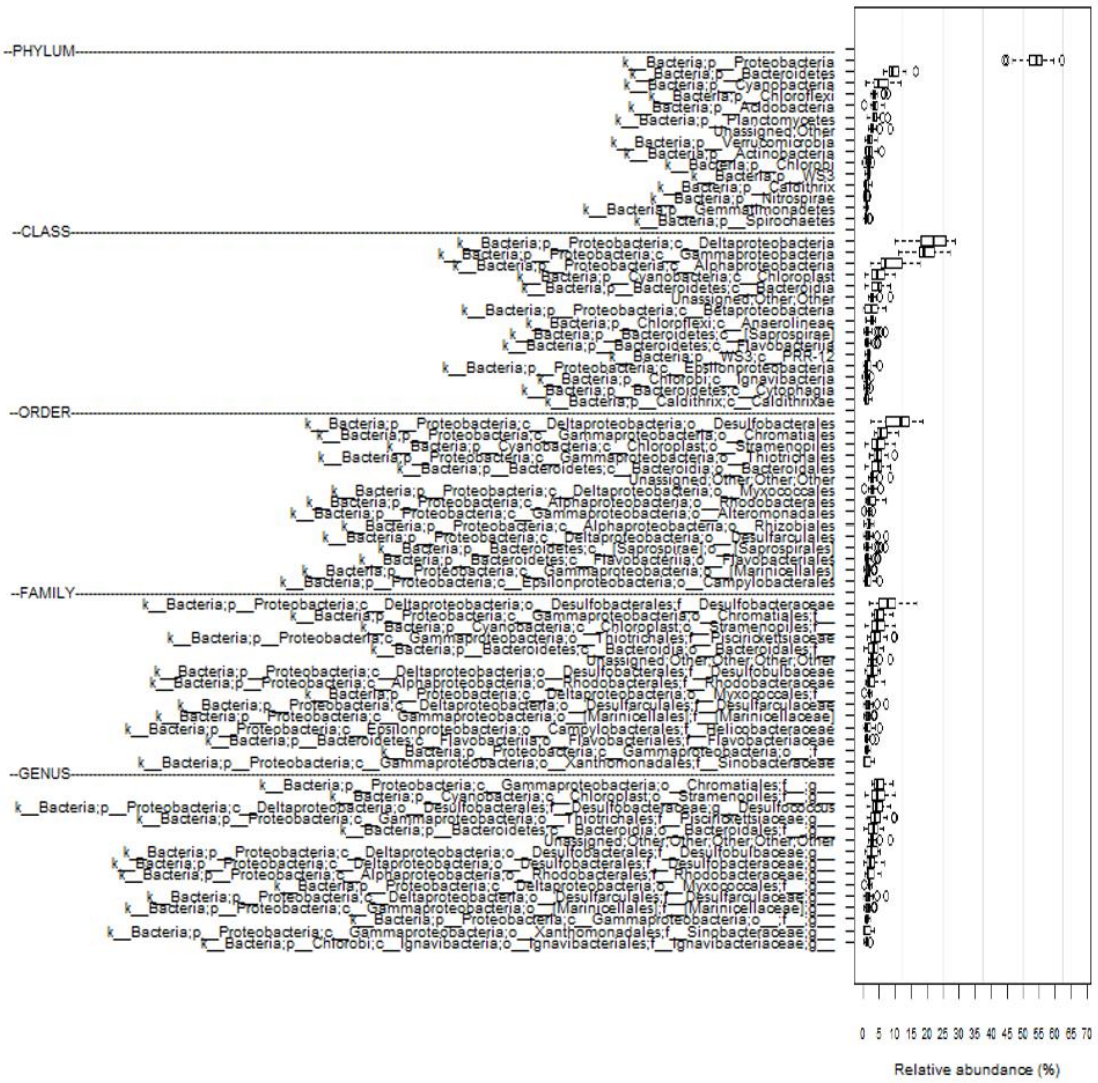


Figure 2. Boxplot summarizing the relative abundances of the fifteen most abundant taxa, by taxon level, from phylum to genus, present in the samples from mangrove sediment.

known organisms, possibly having members not yet described. For the classes, there is a huge prevalence of (21.2±5.3)% *Deltaproteobacteria*, (19.3±3.6)% *Gammaproteobacteria*, (8.4±3.9)% and *Alphaproteobacteria*, which account for almost 50% of the classes found. The top 25% orders are (11.1±4.2)% *Desulfobacterales*, (6.1±2.0)% *Chromatiales*, (4.9±2.5)% *Stramenopiles* and (4.8±2.1)% *Thiotrichales*. The top 20% families are (8.2±4.1)% *Desulfobacteraceae*, (5.0±1.8)% families not identified from the orders *Chromatiales*, (4.9±2.5)% *Stramenopiles*, and (4.3±2.2)% *Piscirickettsiaceae*. The top 10% genera are (5.0±1.8)% genera not identified from order

Chromatiales, (4.9±2.5)% genera not identified from order *Stramenopiles*, and (4.6±2.0)% *Desulfococcus.*

Environmental factors affecting community structure

Supplementary Table 1 summarizes the chemical characteristics of the sediments. Once the distance among samples was calculated, PERMANOVA allowed us to determine which factors influence the community composition (Table 2). Among the levels of sampling adopted for this study, sample, which is a combination of location and plant type, proved to be significant in determining the separation of groups. Location alone is the factor most strongly driving the differences observed, as given by the pseudo-F index. Plant has no effect on the community structure. Among the chemical parameters measured, community structure is influenced by pH (SMP), H^{+}+Al^{3+}, Ca^{2+}, Mg^{2+}, Na^{+}, T, P, Na% and Ca/Mg. Though the strength of the influence does not vary much, parameters Na% and Na^{+} were the strongest.

The distance matrix was plotted in a PCoA graph where the environmental factors significantly affecting community were overlaid (Figure 3). The separation between location sites is evident, and is mostly represented along the horizontal axis. The communities in location A were most influenced by pH (SMP) and Na%, while those of location B were influenced by all other parameters.

Table 2. PERMANOVA evaluation of the environmental factors affecting community structure.

Factors	Pseudo-F	p-value
Discrete factors		
Sample	1.47	0.013*
Location	7.22	0.001***
Plant	0.67	0.987

Factors	Pseudo-F	p-value
Continuous factors		
pH ($CaCl_2$)	1.26	0.056
pH (SMP)	1.66	0.003**
H^++Al^{3+}	2.10	0.001***
Ca^{2+}	1.59	0.003**
Mg^{2+}	1.52	0.005**
K^+	1.33	0.104
Na^+	2.17	0.016*
SB	0	1
T	1.78	0.022*
P	1.69	0.03*
C	1.24	0.13
V	0.81	0.95
Na%	2.32	0.042*
Ca/Mg	1.38	0.04*

208 Note: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

209

210 **Taxa affecting community structure**

211

212 A similar approach revealed which taxa drives the differences among samples.
213 Considering phyla, 56 categories present significant influence (Supplementary Table 2);
214 from the overlay of the corresponding vectors (data not shown), we observed that samples
215 from location A are most enriched for *Firmicutes,* while samples from location B are for
216 *BRC1*, *WS2* and *WS3.* Among the genera (data not shown), the number is 530; from those
217 that could be assigned to genus level, location A is most influenced by *Cyanobacterium*,
218 *Marinicella*, *B-42* and *Microbulbifer,* while location B is best characterized by an
219 enrichment in *Anaerovorax.*

220

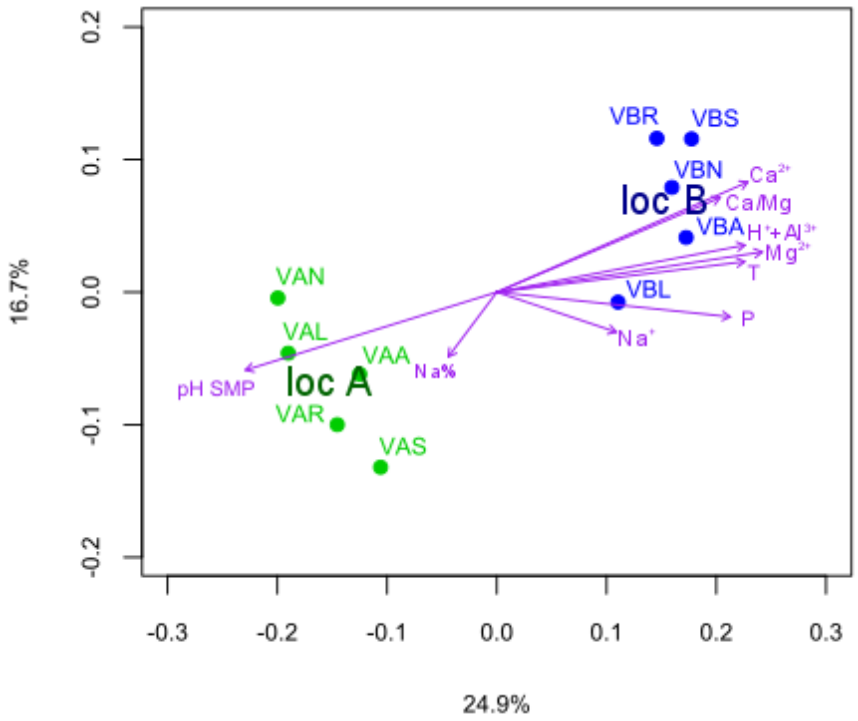


Figure 3. PCoA representing the relative distance among samples, only the centroids are shown. Samples are colored according to the estuarine location from which they were collected: downstream/location A (green), upstream/location B (blue). The purple arrows represent the chemical parameters that were found to significantly affect the diversity; their length is a measure of the correlation. The axis represent the first and the second components of the analysis and their relative weight explaining the variation observed.

221

222 **Community comparison with other mangroves**

223

224 The relative phyla abundance was compared with those revealed by other studies
225 in Brazilian pristine mangrove sediments (Figure 4). The eight most abundant phyla in
226 our study were also the most prominent in other mangroves, with a prevalence of
227 *Proteobacteria* (47-63%). The relative proportion of all phyla is similar, with the
228 following exceptions: two studies reported no sequences from *Cyanobacteria*, and while
229 *Firmicutes* is among the ten most abundant in most mangroves, it was practically absent
230 in our study and in Peixoto *et al.* (2011). *Delta-*, *Gamma-* and *Alphaproteobacteria* are
231 also the most abundances classes, in this order, in Andreote *et al.* (2012), dos Santos *et*

al. (2011) and Nogueira *et al.* (2015); Peixoto *et al.* (2011) shows a prevalence of *Gamma* over *Deltaproteobacteria.*

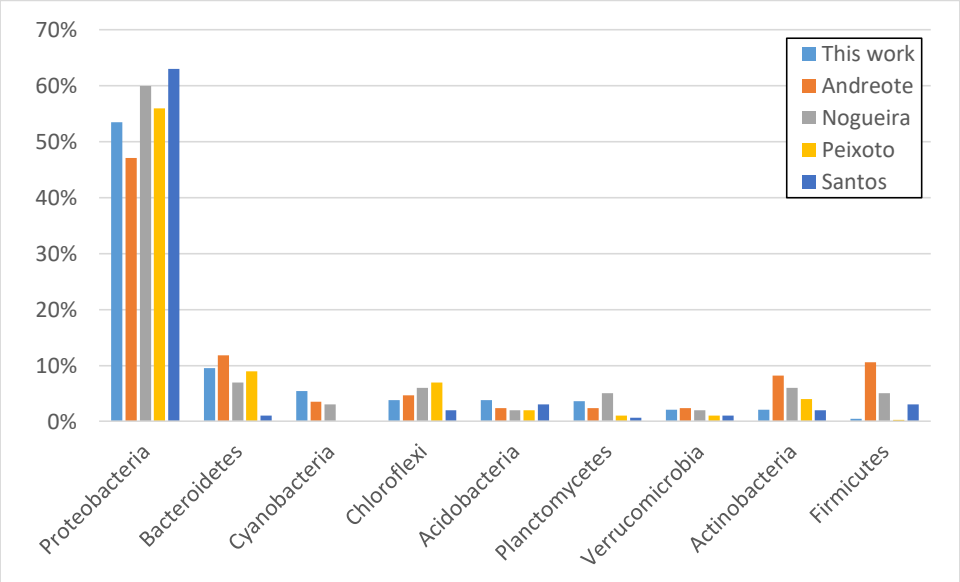


Figure 4. Relative abundance of the most abundant phyla in Brazilian pristine mangrove sediments.

Discussion

Some of the phyla most abundant in the mangrove studied were also prevalent in other soils. Xu *et al.* (2014) showed that environments as distinct as desert, forest, grassland, artic and mangrove share a prevalence of *Proteobacteria, Actinobacteria, Chloroflexi, Bacteroidetes, Acidobacteria* and *Cyanobacteria.* The dominance of *Proteobacteria* is common in all these soils, probably because it is a group with great metabolic diversity (Nogueira *et al.*, 2015), including many nitrogen-fixing organisms. Albeit so widely present, the relative phyla abundances vary depending on the local conditions, so that a profile in such a high taxonomic level seems to be sufficient to characterize these different biomes. The comparison of the Paranaguá Bay mangrove to others (Figure 4) points to average abundances that could be considered as a fingerprint of this habitat. Considering that environmental changes, such as those caused by eutrophication, disturb these structures (Peixoto *et al.*, 2011, Andreote *et al.*, 2012, Nogueira *et al.*, 2015), a single phylum profile could be enough to evaluate the fitness of any given mangrove.

The prevalence of *Deltaproteobacteria* in Paranaguá Bay is also not surprising. This class contains most organisms responsible for sulphate reduction in mangroves, the process mainly responsible for organic matter degradation and sulphur cycling in anoxic environments (Andreote *et al.*, 2012), being also abundant in the other mangroves cited. Their importance is further emphasized by the presence of many orders (*Desulfobacterales, Myxococcales, Desulfarculales*), families (*Desulfobacteraceae, Desulfobulbaceae, Desulfarculaceae*) and genera (as *Desulfococcus*) belonging to this class among the most abundant categories (Figure 2). In fact, 4.6% of all sequences belong to the nutritionally versatile genus *Desulfococcus,* one of the most relevant lineages for carbon degradation in marine sediments (Kleindienst *et al.*, 2014).

When we considered the environmental parameters, location revealed to be the strongest factor affecting community structure. Given the difference in distance to open sea thus in salinity of water, a shift in community to accommodate the diverse conditions was expected. Location B, farther from the ocean, was most influenced by the increased concentration of ions in the sediments; waterlogged soils with lower salinity have higher organic matter content, which retain these ions more efficiently, reason why they show a higher nutritional state. Location A was influenced by the slight shift in pH and by the higher relative concentration of sodium; since the region is more subject to sea water, sodium apparently binds relatively more to the sediment matrix. We did not observe plant effect on community structure, even though specific chemistry microenvironments are expected around roots. While some works report such an effect in mangroves (Jiang *et al.*, 2013, Chen *et al.*, 2016, Rocha *et al.*, 2016), the influence of soil may be greater than that of the plant, masquerading the last (Marschner, Crowley and Yang, 2004).

The community inhabiting location A was most characterized by an enrichment in *Firmicutes.* Many organisms belonging to this phylum produce endospores, showing resistance in extreme environments such as that presented therein. Among the genera enriched, *Cyanobacterium* comprises photosynthetic, nitrogen-fixing cyanobacteria, which are common in moist soils and water, and *Microbulbifer* contains cellulose- and chitin-degrading organisms found in high-salt environments. The phyla enriched in location B, *BRC1*, *WS2* and *WS3* were only recently designated, having no cultivated representatives. From the enriched genus *Anaerovoras*, the only species described, *A. odorimutans*, is a strictly anaerobic chemoorganotrophic bacteria considered to have a voracious putrescine-fermentative metabolism (Matthies *et al.*, 2000).

The analysis of 16S rRNA sequences from environmental DNA samples by high-throughput sequencing technologies is an efficient way to assess their microbial diversity. We could categorize several samples from mangrove sediments according to their community profile, relating it with the environmental factors measured. We could also determine that mangroves seem to have a similar phylum distribution that might serve as a reliable measure for identification and evaluation of the health status of this habitat. Microbial communities are important factors shaping ecosystems, thus their description aids the development of strategies for better handling, conservation and exploitation of such rich, though delicate, sources of natural products.

References

Alongi, D. M. (1994) The role of bacteria in nutrient recycling in tropical mangrove and other coastal benthic ecosystems. **Hydrobiol.**, v. 285, p. 19-32.

Alongi, D. M. (2009) Paradigm shifts in mangrove biology. In **Coastal wetlands: an integrated ecosystem approach.** Perillo, G. M. E.; Wolanski, E.; Cahoon, D. R.; Brinson, M. M. (eds.), Elsevier.

Andreote, F. D.; Jimenez, D. J.; Chaves, D.; Dias, A. C. F. D.; Luvizotto, D. M.; Dini-Andreote, F.; Fasanella, C. C.; Lopez., M. V.; Baena, S.; Taketani, R. G.; Melo, I. S. (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. **PLoS ONE**, v. 7(6), p. 1-14.

Bouchez, A.; Pascault, N.; Chardon, C.; Bouvy, M.; Cecchi, P.; Lambs, L. (2013) Mangrove microbial diversity and the impact of trophic contamination. **Marine Poll. Bull.**, v. 66, p. 39-46.

Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, v. 7., p. 335-336.

Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M.; Gormley, N.; Gilbert, J. A.; Smith, Geoff.;

315 Knight, R. (2012) Ultra-high-throughput microbial community analysis on the Illumina
316 HiSeq and MiSeq platforms. **ISME J.**, v. 6, p. 1621-1624.

317 Chen, J. (2012). GUniFrac: Generalized UniFrac distances. R package version 1.0.
318 https://CRAN.R-project.org/package=GUniFrac

319 Chen, J.; Bittinger, K.; Charlson, E. S.; Hoffmann, C.; Lewis, J.; Wu, G. D.; Collman, R.
320 G.; Bushman, F. D.; Li, H. (2012) Associating microbiome composition with
321 environmental covariates using generalized UniFrac distances. **Bioinformatics**, v.
322 28(16), p. 2106-2113.

323 Chen, Q., Zhao, Q., Li, J., Jian, S., Ren, H. (2016) Mangrove succession enriches the
324 sediment microbial community in South China. **Sci. Rep.**, v. 6, 27468. doi:
325 10.1038/srep27468

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

330 Dos Santos, H.; Cury, J. C.; Do Carmo, F. L.; Dos Santos, A. L. ; Tiedje, J; Van Elsas, J.
331 D.; Rosado, A. S.; Peixoto, R. S. (2011) Mangrove bacterial diversity and the impact of
332 oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. **PLOS**
333 **ONE**, v. 6(3), 416943.

334 Duke, N. C.; Ball, M. C.; Ellison, J.C. (1998) Factors influencing biodiversity and
335 distributional gradients in mangroves. **Global Ecol. Biogeo. Lett.**, v. 7(1), p. 27-47.

336 Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST,
337 **Bioinformatics**, v. 26(19), p.2460-2461. doi: 10.1093/bioinformatics/btq461

338 Feller, I. C.; Lovelock, C. E.; Berger, U.; McKee, K. L. (2010) Biocomplexity in
339 mangrove ecosystems. **Annu. Rev. Mar. Sci.**, v. 2, p. 395-417.

340 Ghizelini, A. M.; Mendonça-Hagler, L. C. S.; Macrae, A. (2012) Microbial diversity in
341 Brazilian mangrove sediments – a mini review. **Brazil. J. Microbiol.**, p. 1242-1254.

342 Gomes, N. C. M.; Borges, L. R.; Paranhos, R.; Pinto, F. N.; Mendonça-Hagler, L. C. S.;
343 Smalla, K. (2008) Exploring the diversity of bacterial communities in sediments of urban
344 mangrove forests. **FEMS Microbiol. Ecol.**, v. 66, p. 96-109.

345 Holguin, G.; Vazquez, P.; Bashan, Y. (2001) The role of sediment microorganisms in the
346 productivity, conservation and rehabilitation of mangrove ecosystems: an overview. **Biol.**
347 **Fertil. Soils.**, v. 33, p. 265-278.

348 Jiang, X.T., Peng, X., Deng, G.H., Sheng, H.F., Wang, Y., Zhou, H.W. Tam, N.F.Y.
349 (2013) Illumina sequencing of 16S rRNA tag revealed spatial variations of bacterial
350 communities in mangrove wetland. **Microbial Ecol.**, v. 66(1), p.96-104.

351 Kleindienst, S., Herbst, F.A., Stagars, M., von Netzer, F., von Bergen, M., Seifert, J. *et*
352 *al.* (2014) Diverse sulfate-reducing bacteria of the *Desulfosarcina/Desulfococcus* clade
353 are the key alkane degraders at marine seeps. **ISME J.**, v. 10, p.2029-2044.

354 Lovelock, C. E. (2008) Soil respiration and belowground carbon allocation in mangrove
355 forests. **Ecosystems**, v. 11, p.342–354.

356 Marques, R.; Motta, A. C. V. (2003) Análise química do solo para fins de fertilidade. In:
357 **Manual de diagnóstico da fertilidade e manejo dos solos agrícolas.** Lima, R. M. (Org.)
358 Universidade Federal do Paraná, Departamento de Solos e Engenharia Agrícola, p. 81-
359 102.

360 Marschner, P.; Crowley, D.; Yang, C. H. (2004) Development of specific rhizosphere
361 bacterial communities in relation to plant species, nutrition and soil type. **Plant and Soil**,
362 v. 261, p. 199-208.

363 Matthies, C., Evers, S., Ludwig, W., Schink, B. (2000) *Anaerovorax odorimutans* gen.
364 nov., sp. Nov., a putrescine-fermenting, strictly anaerobic bacterium. **Int. J. Syst. Evol.**
365 **Microbiol**,, v. 50(4), p. 1591-1594.

366 Nogueira, V. L. R.; Rocha, L. L.; Colares, G. B.; Angelima, A. L.; Normando, L. R. O.;
367 Cantão, M. E.; Agnez-Lima, L. F.; Andreote, F. D.; Melo, V. M. M. (2015) Microbiomes
368 and potential metabolic pathways of pristine and anthropized Brazilian mangroves. **Reg.**
369 **Stud. Marine Sci.**, v. 2, p. 56-64.

370 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B.,
371 Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Wagner, H. (2016) vegan:
372 Community Ecology Package. R package version 2.3-5. https://CRAN.R-
373 project.org/package=vegan

374 Peixoto, R.; Chaer, G. M.; Carmo, F. L.; Araújo, F. V.; Paes, J. E.; Volpon, A.; Santiago,
375 G. A.; Rosado, A. S. (2011) Bacterial communities reflect the spatial variation in pollutant
376 levels in Brazilian mangrove sediments. **Antonie van Leeuwenhoek**, v. 99, p. 341-354.

377 Pylro, V.S., Roesch, L.F.W., Morais, D.K., Clark, I.A., Hirsch, P.R., Tótola, M.R. (2014)
378 Data analysis for 16S microbial profiling from different benchtop sequencing platforms.
379 **J. Microbiol. Methods**, v. 107, p.30-37. doi:10.1016/j.mimet.2014.08.018

380 R Core Team (2016). R: A language and environment for statistical computing. R
381 Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

382 Reef, R.; Feller, I. C.; Lovelock, C. E. (2010) Nutrition of mangroves. **Tree Physiol.**, v.
383 30, p. 1148-1160.

384 Rocha, L.L., Colares, G.B., Nogueira, V.L.R., Paes, F.A., Melo, V.M.M. (2016) Distinct
385 habitats select particular bacterial communities in mangrove sediments. **Int. J.**
386 **Microbiol.**, v. 2016, article ID 3435809, 6 pages. doi:10.1155/2016/3435809

387 Sahoo, K.; Dhal, N. K. (2009) Potential microbial diversity in mangrove ecosystems: a
388 review. **Ind. J. Marine Sci.**, v. 38(2), p. 249-256.

389 Saseeswari, A., Kanimozhi, G., Panneerselvam, A. (2016) Bacterial diversity of
390 mangrove soil in Karankadu from East Coast of Tamil Nadu, India. **Int. J. Curr.**
391 **Microbiol. App Sci.**, v. 5(4), p.750-756.

392 Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. **Bioinformatics**, v. 27(4),
393 p.592-593

394 Shade, A.; Handelsman, J. (2012) Beyond the Venn diagram: the hunt for a core
395 microbiome. **Environ. Microbiol.**, v. 14(1), p. 4-12.

396 Spalding, M. D.; Blasco, E., Field, C.D. (Eds). (1997) **World mangrove atlas.** The
397 International Society for Mangrove Ecosystems, Okinawa, Japan. 178 pp.

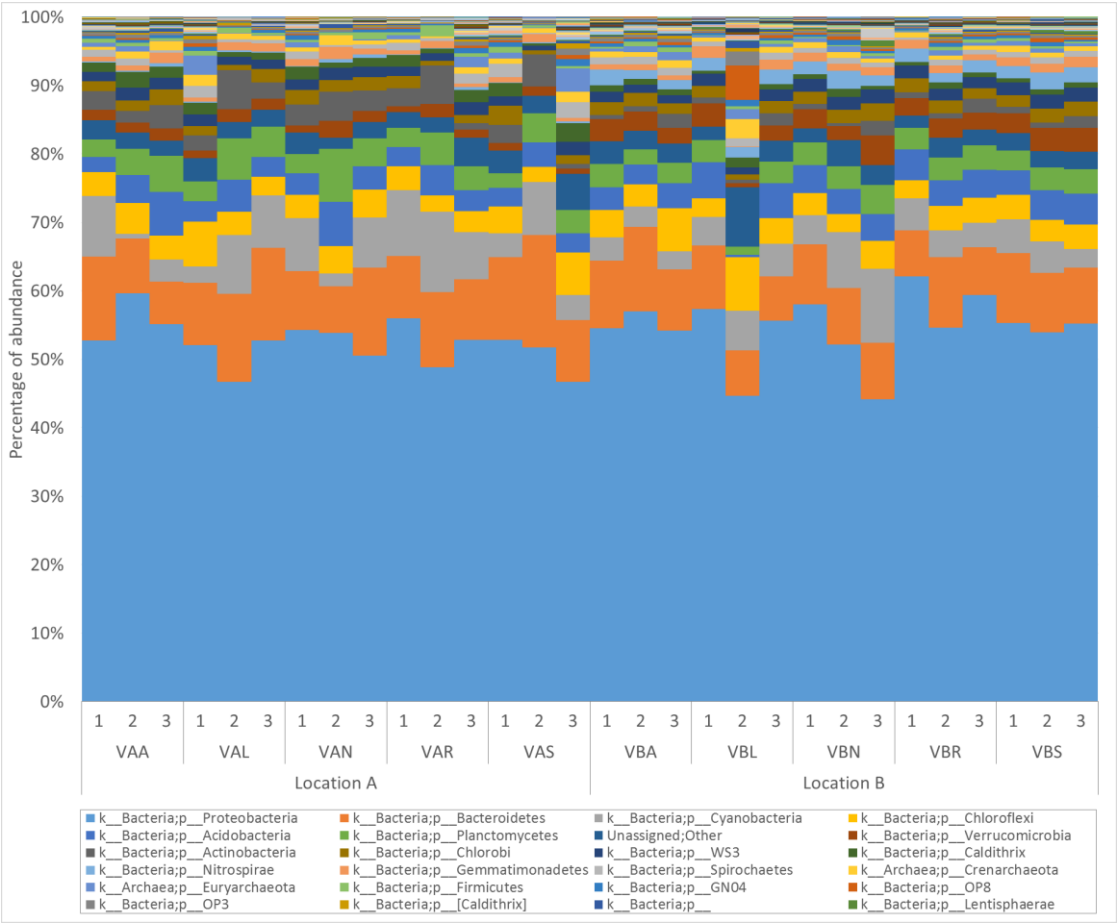
398 Spalding, M. D.; Kainuma, M.; Collins, L. (2010) **World atlas of mangroves.** Earthscan,
399 London, UK. 338 pp.

400 Thatoi, H., Behera, B.C., Mishra, R.R., Dutta, S.K. (2013) Biodiversity and
401 biotechnological potential of microorganisms from mangrove ecosystems: a review. **Ann.**
402 **Microbiol.**, v. 63, p.1-19.

403 Xu, Z., Hansen, M.A., Hansen, L.H., Jacquiod, S., Sørensen, S.J. (2014) Bioinformatic
404 approaches reveal metagenomic characterization of soil microbial community. **PLoS**
405 **ONE**, v. 9(4), e93445. doi:10.1371/journal.pone.0093445

406

Supplementary figures



Supplementary Figure 1. Relative abundance of the phyla for each sample analyzed. The first row in the horizontal axis represents the replicate number; the second represents the sample according to location in the estuary and the plant species from which rhizosphere it was taken. Only the 24 most abundant phyla are given in the legend.

Supplementary tables

Supplementary Table 1. Chemical data of the soil samples.

Sample	pH ($CaCl_2$)	pH (SMP)	$H^+ + Al^{3+}$ ($cmol_c/dm^3$)	Ca^{2+} ($cmol_c/dm^3$)	Mg^{2+} ($cmol_c/dm^3$)	K^+ ($cmol_c/dm^3$)	Na^+ ($cmol_c/dm^3$)	SB ($cmol_c/dm^3$)	T ($cmol_c/dm^3$)	P (mg/dm^3)	C (g/dm^3)	V (%)	Na% (%)	Ca/Mg
Location A														
VAA1	6.1	7.2	2.0	1.6	1.4	0.45	2.73	6.18	8.18	6.1	20.2	76	33.4	1.14
---2	6.1	7.2	2.0	2.1	3.9	0.36	2.35	8.71	10.71	4.4	19.2	81	21.9	0.54
---3	6.2	7.4	1.8	1.1	2.2	0.35	2.26	5.91	7.71	4.2	6.0	77	29.3	0.50
VAL1	5.3	6.7	3.0	2.5	4.2	0.57	2.96	10.23	13.23	8.0	29.6	77	22.4	0.60
---2	6.4	7.6	1.5	1.1	2.7	0.40	2.92	7.12	8.62	4.3	5.1	83	33.9	0.41
---3	6.5	7.5	1.6	1.0	2.0	0.33	2.25	5.58	7.18	4.3	5.1	78	31.3	0.50
VAR1	6.0	7.2	2.0	1.4	3.0	0.46	2.60	7.46	9.46	6.5	13.3	79	27.5	0.47
---2	6.8	7.6	1.5	1.0	2.7	0.35	2.52	6.57	8.07	4.1	3.3	81	31.2	0.37
---3	6.0	7.0	2.4	3.2	4.7	1.23	11.70	20.83	23.23	21.2	32.9	90	50.4	0.68
VAS1	5.4	7.1	2.2	1.4	2.9	0.38	2.58	7.26	9.46	5.9	12.4	77	27.3	0.48
---2	6.8	7.6	1.5	0.8	1.9	0.28	2.30	5.28	6.78	4.3	4.2	78	33.9	0.42
---3	5.6	7.6	3.0	3.2	4.8	1.13	13.50	22.63	25.63	22.2	39.7	88	52.7	0.67
VAN1	5.9	7.1	2.2	2.1	4.0	0.54	3.19	9.83	12.03	7.1	22.2	82	26.5	0.53
---2	6.4	7.4	1.8	1.4	2.9	0.29	2.26	6.85	8.65	3.7	6.9	79	26.1	0.48
---3	6.8	7.5	1.6	1.4	3.0	0.34	2.77	7.51	9.11	3.1	7.8	82	30.4	0.47
Location B														
VBA1	5.9	6.6	3.2	5.7	5.0	1.23	3.26	15.19	18.39	19.0	40.9	83	17.7	1.14
---2	5.8	6.3	4.0	5.6	5.4	1.19	3.34	15.53	19.53	12.2	45.9	80	17.1	1.04
---3	5.5	6.0	5.0	6.7	5.2	1.02	3.13	16.05	21.05	13.2	53.2	76	14.9	1.29
VBL1	6.1	6.6	3.2	4.8	5.0	1.27	9.50	20.57	23.77	15.2	29.6	87	40.0	0.96
---2	6.2	6.6	3.2	5.2	5.1	1.24	3.59	15.13	18.33	14.1	32.9	83	19.6	1.02
---3	6.4	6.8	2.7	5.3	5.0	1.34	3.54	15.18	17.88	8.9	34.0	85	19.8	1.06
VBR1	6.0	6.5	3.4	7.4	5.1	1.05	3.30	16.85	20.25	8.8	43.3	83	16.3	1.45
---2	6.3	6.7	3.0	5.6	5.3	1.23	12.9	25.03	28.03	17.9	45.7	89	46.0	1.06
---3	6.0	6.5	3.4	5.8	5.1	1.12	3.10	15.12	18.52	18.2	49.4	82	16.7	1.14
VBS1	6.3	6.8	2.7	5.5	5.3	1.26	12.10	24.16	26.86	11.9	49.4	90	45.0	1.04
---2	5.6	6.2	4.3	4.6	5.1	1.18	3.42	14.30	18.60	16.5	37.4	77	18.4	0.90
---3	5.8	6.2	4.3	4.2	5.0	1.21	3.43	13.84	18.14	17.0	36.3	76	18.9	0.84
VBN1	6.4	6.9	2.5	4.5	5.2	1.22	10.10	21.02	23.52	17.2	29.6	89	42.9	0.87
---2	5.9	6.5	3.4	4.6	5.2	1.27	3.41	14.48	17.88	13.7	25.3	81	19.1	0.88
---3	5.5	6.3	4.0	4.1	5.1	1.25	3.43	13.88	17.88	19.0	27.4	78	19.2	0.80

Supplementary Table 2. Phyla that are significantly affecting community structure.

Domain | Phylum	p-value	Domain | Phylum	p-value
Archaea | [Parvarchaeota]	0.001	Bacteria | Lentisphaerae	0.001
Archaea | Crenarchaeota	0.036	Bacteria | MVS-104	0.002
Archaea | Euryarchaeota	0.001	Bacteria | NC10	0.001
Bacteria | Other	0.003	Bacteria | Nitrospirae	0.001
Bacteria | [Caldithrix]	0.001	Bacteria | OC31	0.03
Bacteria | [Thermi]	0.001	Bacteria | OP11	0.031
Bacteria | AC1	0.002	Bacteria | OP3	0.001
Bacteria | Acidobacteria	0.001	Bacteria | OP8	0.005
Bacteria | Actinobacteria	0.001	Bacteria | OP9	0.001
Bacteria | Armatimonadetes	0.001	Bacteria | Planctomycetes	0.001
Bacteria | Bacteroidetes	0.035	Bacteria | Poribacteria	0.001
Bacteria | BHI80-139	0.015	Bacteria | Proteobacteria	0.043
Bacteria | BRC1	0.001	Bacteria | SAR406	0.001
Bacteria | Caldithrix	0.001	Bacteria | SBR1093	0.007
Bacteria | Chlamydiae	0.027	Bacteria | SC4	0.001
Bacteria | Chloroflexi	0.002	Bacteria | Spirochaetes	0.001
Bacteria | Fibrobacteres	0.001	Bacteria | SR1	0.032
Bacteria | Firmicutes	0.017	Bacteria | Thermotogae	0.038
Bacteria | Fusobacteria	0.014	Bacteria | TM7	0.001
Bacteria | Gemmatimonadetes	0.001	Bacteria | TPD-58	0.028
Bacteria | GN02	0.001	Bacteria | Verrucomicrobia	0.001
Bacteria | GN04	0.001	Bacteria | VHS-B3-43	0.012
Bacteria | H-178	0.003	Bacteria | WPS-2	0.007
Bacteria | Hyd24-12	0.001	Bacteria | WS1	0.031
Bacteria | Kazan-3B-28	0.003	Bacteria | WS2	0.001
Bacteria | KSB3	0.001	Bacteria | WS3	0.011
Bacteria | LCP-89	0.001	Bacteria | WWE1	0.001
Bacteria | LD1	0.001	Unassigned | Other	0.001

CAPÍTULO 2 – Análise da comunidade e análise funcional do microbioma do mangue da costa sul do Brasil

Community and Functional Analysis of the Mangrove Microbiome from the Southern Brazilian Coast

Abstract

Mangroves are habitats that present harsh conditions for microorganisms due to fluctuations in salinity gradients and in anaerobicity of the soils. Nevertheless, they are very active environments, in which a small diversity of plants interacts with a high diversity of microbes by means of a complex biochemical web designed to conserve scarce nutrients through efficient recycling. This equilibrium is fragile, thus the rational management of mangroves requires a better understanding of how the players act and interact. In this work, we investigated the microbiome of mangrove sediments from Paranaguá Bay, Brazil, sampled from two chemically distinct locations, by means of metagenomic whole genome shotgun (mWGS) sequencing. The results revealed a predominance of *Bacteria*, with most representatives from phyla *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes.* The major classes are *Beta-*, *Gamma-*, *Alpha-* and *Deltaproteobacteria*, and around one-fourth of the sequences are related to the genera *Achromobacter* and *Stenotrophomonas.* Prevalent functions are related to clustering-based subsystems, carbohydrate metabolism, amino acids and derivatives, and miscellaneous functions; at the lowest functional levels, abundances are lower and shifts are subtler. Although the high similarity in all comparisons between the two locations, significant differences in microbiological structures were identified, which might be explained by differences in the environmental conditions. The microbial taxonomic profiles are more dissimilar than the functional profiles, which indicates the latter to be redundant; this feature might aid in its resiliency when the habitat is faced with small disturbances. By adding publicly available data to our analysis, we showed that samples from different biomes tend to cluster according to biome and to biome type, both taxonomically and functionally; the similarity of mangroves to other biomes seems to be driven more by matrix type (where soil > water) than to chemical composition (where saline > non-saline). Finally, we observed that metabolic functions are more conserved than organisms across all biomes, which indicates that core functions must be preserved regardless of the specific organisms harboring them.

Introduction

Mangroves are highly productive forests largely confined to tropical and subtropical regions, where the warm temperature favors its development (Alongi, 2002). Its estimated 152,000-km^2 area covers around 60-75% of the world's coastlines (Holguin *et al.,* 2001), usually estuarine (Taketani *et al.,* 2010). Some of the largest forests are found in Indonesia, India, Bangladesh, and Brazil (Spalding *et al.,* 2010). In Brazil, which occupies the second position harboring 8.5% of this total, the most extensive areas are present in the north region; in the southeastern coast, their distribution is sparser, often restricted to river deltas, coastal lagoons and the inner parts of bays (Spalding *et al.,* 1997).

Some characteristics make mangroves structurally and functionally unique. The species present have evolved to cope with the physiological challenges of variable conditions: gradients in salinity caused by high regular rainfall, groundwater flows and rivers diluting seawater to different extents; and anaerobic soils except for the sediment surface, since they are constantly waterlogged (Duke *et al.,* 1998, Bouchez *et al.,* 2013). These contrasts are readily observable in estuaries, where the spatial distribution of species might be classified into downstream, closer to open sea, where salinity and tide variation are higher, thus sediment deposition and nutrient exchange are more dynamic; and upstream, where rivers flow into and organic matter content is higher (Duke *et al.,* 1998). The soil is composed of thick organic matter mixed with sediment (Holguin *et al.,* 2001). The sediments are the foundation for mangrove forests and all that live in them (Ghizelini *et al.,* 2012). Its nutrient availability is variable within and among mangrove ecosystems, ranging from extremely low in oceanic settings to very high in accreting muddy systems (Feller *et al.,* 2010).

The species hosted form a rich array of association, supporting a complex aquatic food web. The microbial diversity can be high (Andreote *et al.,* 2012), especially in shallow sediments, decreasing with increasing depth (Taketani *et al.,* 2010). Many of the microorganisms play a critical role in the final stages of decomposition, helping to build highly efficient energetic processes (Yan *et al.,* 2006, Spalding *et al.,* 2010). The detritus generated is the base of this habitat's food web (Holguin *et al.,* 2001). Bacteria seem to have greater importance: they tend to be prevalent (Nogueira *et al.,* 2015) for, due to their small size, being able to colonize a greater proportion of niches (Ghizelini *et al.,* 2012); their trophic functions therein are also more relevant than in non-trophic habitats (Alongi,

1994). Despite various studies on the biogeography, botany, zoology, ichthyology, environmental pollution and economic impact, little is known about the microbial life in mangrove waters and sediments (Marcial Gomes *et al.*, 2008, Dhal and Sahoo, 2009, Andreote *et al.*, 2012). Nonetheless, understanding the biological and functional diversity of such systems is crucial to maintain ecosystem dynamics and to sustainably manage land use (Pylro *et al.*, 2014).

We performed metagenomic whole genome shotgun (mWGS) sequencing to describe the microbial communities present in two environmentally distinct locations in a mangrove in the southeastern coast of Brazil, according to their taxonomic and functional hierarchies. We compared the mangrove communities with those inhabiting other biomes with diverse characteristics. We established the level of similarity among them and speculated on what parameters might be driving these indices. Finally, we discussed the relations between the taxonomic and functional similarities observed.

Materials and methods

Sample extraction and environment characterization

The work described in Chapter 1 showed location exerted a strong influence in the community composition. Therefore, in order to represent the mangrove in this shotgun study, we chose one sample from an open area from each location, VAN2 and VBN2, henceforth mentioned as VA and VB for simplicity. The chemical characterization of the sediments was also imported from that study. The comparison of these parameters between regions was done by two-sided Welch's t-test.

DNA extraction, metagenome sequencing and annotation

Environmental DNA was extracted using the PowerSoil DNA Isolation Kit (MO Bio Laboratories). The library was prepared with the Nextera XT DNA Library Preparation Kit (Illumina) and the sequencing was performed in the MiSeq platform (Illumina) using the MiSeq Reagent Kit v3 600 (Illumina). After initial trimming for barcode removal and standard quality control, the R1 reads were uploaded to the MG-RAST online platform under accession numbers 4604820.3 (VA) and 4604821.3 (VB), and then processed according to the published pipeline (Meyer *et al.*, 2008). Taxonomic

and functional hierarchies were determined based on all reads annotated and their closest relatives (max. e-value 1E-5, min. identity 60%, min. alignment length 15 cut-offs applied) in databases M5NR for taxonomy, and Subsystems for function. The taxonomic (from domain to genus level) and the functional (from level 1 to level 4/function) category hits distribution tables were selected as profiling matrices of the samples.

Taxonomic and functional analysis

The profiling matrices for samples VA and VB were compared by fitting a linear model with forced intercept at 0 (VA ~ 0 + VB), and evaluating the calculated correlation value, as implemented in R v3.2.2 (R Core Team, 2016). The most abundant categories, ordered decreasingly by average, for both hierarchies and all levels, were plotted. Effect sizes by category were determined as the ratio between the absolute difference in abundance of VA and VB, and the abundance in VA, being the highest the value, the highest the shift in that category between the two locations. Publicly available metagenomes representing biomes estuarine, freshwater, Mediterranean Sea, river, temperate forest, temperate grassland, terrestrial and tropical forest were retrieved from the MG-RAST webserver for comparison with our mangrove metagenomes (Supplementary Table 1). Bias was minimized by selecting only mWGS projects, sequenced in the Illumina platform, containing at least 1,000,000 reads in the range of 200 and 300 bp long each. Their taxonomic (at domain and genus level) and functional (at level 1 and level 4/function) abundance tables were downloaded into R v3.2.2 (R Core Team, 2016). They were normalized by rarefaction at the minimum metagenomic total abundance using package GUniFrac (Chen, 2012), and Bray-Curtis distances among metagenomes were calculated with package vegan (Oksanen *et al.*, 2016). Heatmaps were generated using package heatmap.plus (Day, 2012). Statistics for the comparison of final abundance tables, taking biome as a factor, were calculated with PERMANOVA as implemented by package vegan (Oksanen *et al.*, 2016). All other statistical analyses and graphs not mentioned above were generated using the basic R package (R Core Team, 2016).

Results

Mangrove environmental characterization

First, we aimed at establishing if the environments from which the two sample were taken were different according to the chemical composition of the sediments. The statistical comparison of the measured parameters revealed significant difference of VA and VB, specifically in the contents of Ca^{2+}, Mg^{2+}, K^{+} and P, and in SB, T and Ca/Mg (Table 1). pH was not significantly different between VA and VB.

Table 1. Chemical parameters for the sediment samples triplicates.

Parameter	Sample		p-value
	VA	VB	
pH ($CaCl_2$)	6.4 ± 0.5	5.9 ± 0.5	0.493
pH (SMP)	7.3 ± 0.2	6.6 ± 0.3	0.122
Al^{+3} ($cmol_c/dm^3$)	0	0	-
$H^{+} + Al^{+3}$ ($cmol_c/dm^3$)	1.9 ± 0.3	3.3 ± 0.7	0.144
Ca^{2+} ($cmol_c/dm^3$)	1.6 ± 0.4	4.4 ± 0.3	0.007*
Mg^{2+} ($cmol_c/dm^3$)	3.3 ± 0.6	5.2 ± 0.1	0.031*
K^{+} ($cmol_c/dm^3$)	0.39 ± 0.13	1.25 ± 0.03	0.011*
Na^{+} ($cmol_c/dm^3$)	2.74 ± 0.47	5.65 ± 3.86	0.284
SB ($cmol_c/dm^3$)	8.06 ± 1.57	16.46 ± 3.96	0.028*
T ($cmol_c/dm^3$)	9.93 ± 1.83	19.76 ± 3.26	0.007*
P (mg/dm^3)	4.6 ± 2.2	16.6 ± 2.7	0.025*
C (g/dm^3)	12.3 ± 8.6	27.4 ± 2.1	0.060
V (%)	81 ±2	83 ± 6	0.709
Na (%)	27.7 ± 2 .38	27.1 ± 13.7	0.949
Ca/Mg	0.49 ± 0.03	0.85 ± 0.04	0.004*

Note: the values represent mean ± standard deviation for three independent measures at each environment. (*) = $p < 0.05$.

Comparison of mangrove microbial communities through metagenomics

The total DNA of both mangrove locations was sequenced and annotated (summary in Table 2).

Table 2. Sequencing and annotation report for the metagenomes sequenced.

Parameter	Estuarine location	
	Downstream (VA)	Upstream (VB)
Sequencing		
Number of reads	5,284,657	6,418,039
Mean read length (bp)	283 ± 46	280 ± 48
Total bases count	1,499,227,644	1,803,223,762
Annotation		
Post QC sequences count	3,695,582 (69.9%)	3,843,928 (59.9%)
Protein features		
Predicted	2,128,959 (57.6%)	2,611,090 (67.9%)
Identified	1,124,894 (52.8%)	978,732 (37.5%)
Identified functional categories	878,849 (78.1%)	768,433 (78.5%)
rRNA features		
Predicted	39,598 (1.1%)	44,076 (0.7%)
Identified	2,582 (6.5%)	2,739 (6.2%)

Note: the values in brackets indicate the percentage of features retained from the previous step.

The microbial communities from samples VA and VB were compared at six taxonomic levels, and category averages were plotted to assess relative abundance in the biome (Figure 1). At the domain level, with a correlation coefficient of R^2=0.999, the data suggest that both environments are virtually identical. There is a huge predominance of over 90% of *Bacteria*, though the larger effect sizes are observed in *Archaea* (2.2), unclassified sequences (1.8) and *Eukaryota* (0.5). The correlation is also high at the phylum level (R^2=0.999); *Proteobacteria* represents more than 79% of the sequences, followed by unassigned phyla (6%), *Bacteroidetes* (5%) and a smaller percentage of

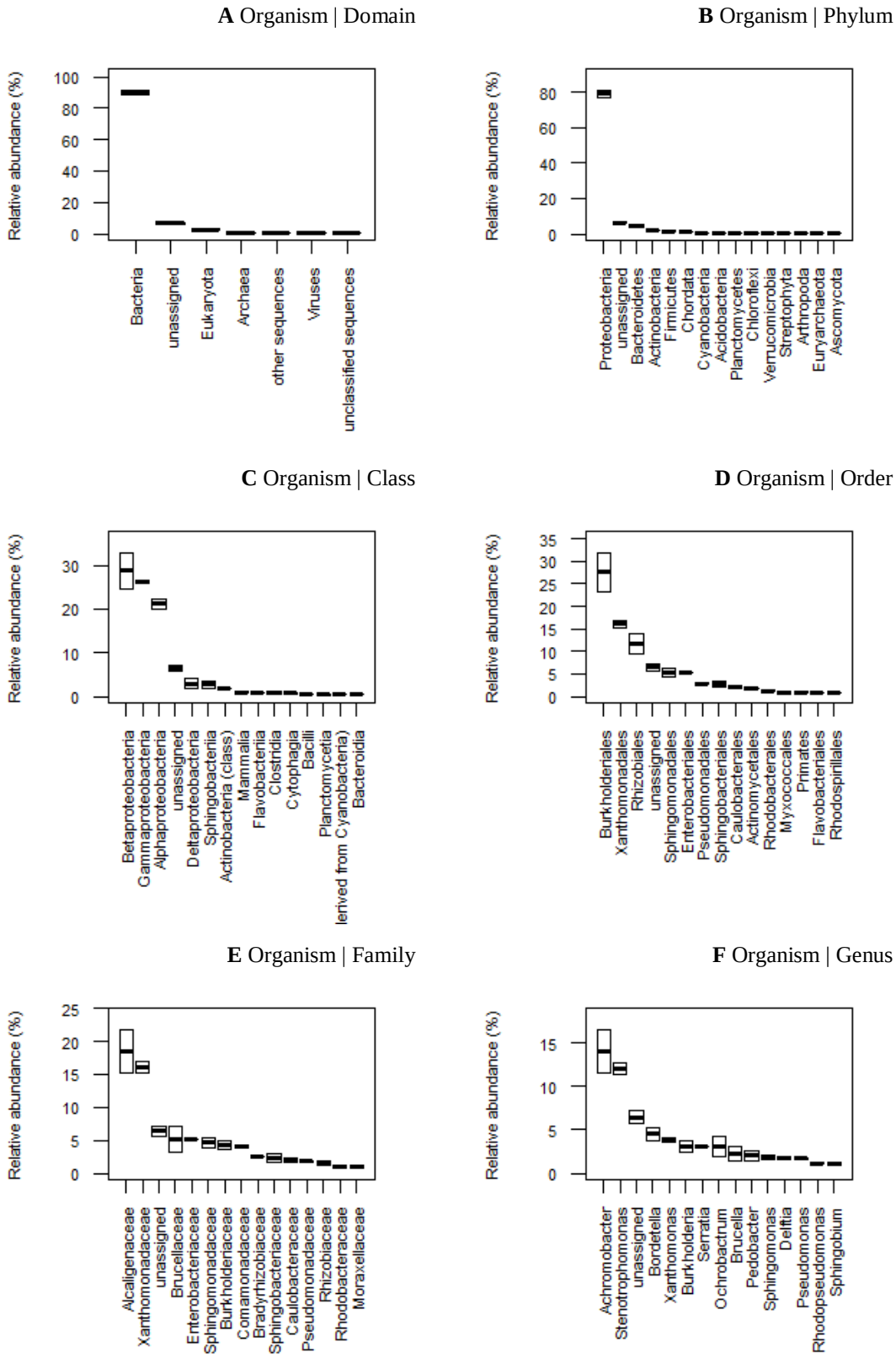


Figure 1. Plots showing the relative abundances (%) of the 15 most abundant categories in each taxonomical level considered in the mangrove locations. The boxes represent the values for VA and VB as the shorter sides and the mean as the thicker line in the middle.

164 *Actinobacteria*, *Firmicutes* and *Chordata*. The correlation coefficient was lower at the

class level (R^2=0.966), with over 75% of the sequences belonging to *Beta-*, *Gamma-*, *Alpha-* and *Deltaproteobacteria*. Family (R^2=0.938) and genus (R^2=0.937) ranks showed the lowest correlation coefficients. Among the genera, the biome is enriched for the presence of *Achromobacter* (14%), *Stenotrophomonas* (12%) and unassigned genera (6%). Its highest effect sizes were observed decreasingly in *Ochrobactrum, Brucella, Pedobacter, Burkholderia* and *Achromobacter*, ranging from 1.2 to 0.3.

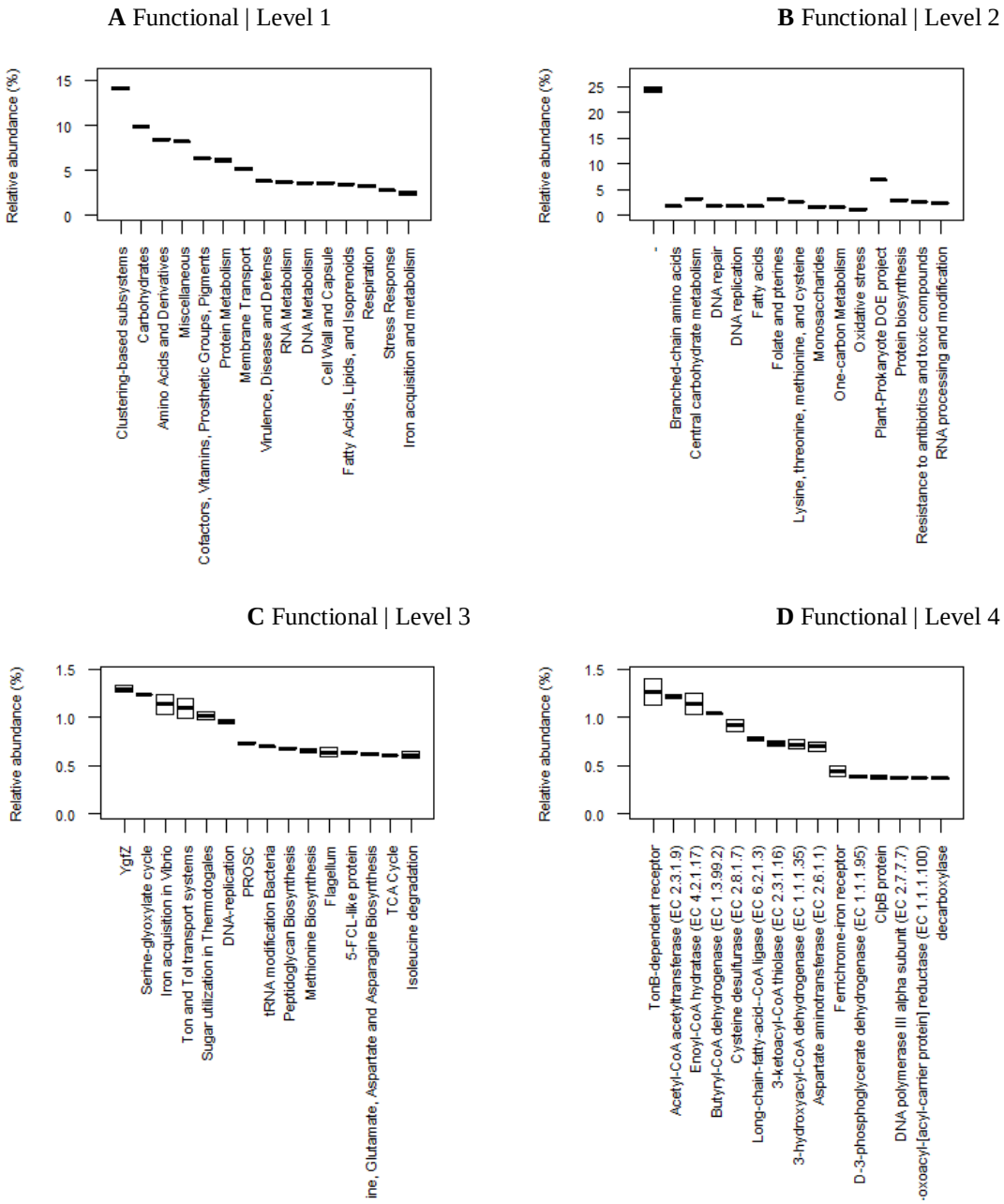


Figure 2. Plots showing the relative abundances (%) of the 15 most abundant categories in each functional level considered in the mangrove locations. The boxes represent the values for VA and VB as the shorter sides and the mean as the thicker line in the middle.

The same datasets were similarly analyzed and compared according to four levels of functional hierarchy (Figure 2). At level 1, R^2=0.999, which again suggests a close resemblance; the effect sizes are small, not higher than 0.15. The distribution of the functions at this level is quite balanced: 14% of the sequences are related to clustering-based subsystems, 10% to carbohydrate metabolism, 8% to amino acids and derivatives and 8% to miscellaneous functions. At level 2, the correlation factor is the same, followed by R^2=0.989 at level 3 and R^2=0.975 at level 4, the effect size never exceeding 0.23. At this last level, no function is represented by more than 1.3% of the sequences; the highest effect sizes, albeit overall small (ranging from 0.23 to 0.13), are related to ferrichrome-iron receptor, TonB-dependent receptor, enoyl-CoA hydratase, cysteine desulfurase and 3-hydroxyacyl-CoA dehydrogenase.

When both metagenomes are superimposed in a KEGG map representation, they overlap extensively, which means the large majority of functions are shared by the microbiomes of both sites, with only few functions specific to each site, which occur in small abundance (Supplementary Figure 1).

Comparison among biomes

When comparing the different biomes by their taxonomic similarity, first we represented them in a heatmap, for the highest and lowest hierarchies considered, domain and genus, after metagenome clustering (Figure 3). The individual metagenomes tend to group according to their biome, but at the domain level, we noticed some scattering within datasets representing freshwater, mangrove and Mediterranean Sea. Main groups are formed by 1) estuarine, 2) temperate forest, and 3) the remaining biomes. As observed for the mangrove assessment, *Bacteria* is the majoritarian taxon. At the genus level, all metagenomes cluster according to their respective biome, being the groups represented by: 1) river, 2) mangrove, 3) freshwater, 4) temperate forest, temperate grassland, terrestrial and tropical forest, and 5) estuarine and Mediterranean Sea. PERMANOVA analysis taking biome as a factor shows that the differences reported are significant in both levels (Table 3).

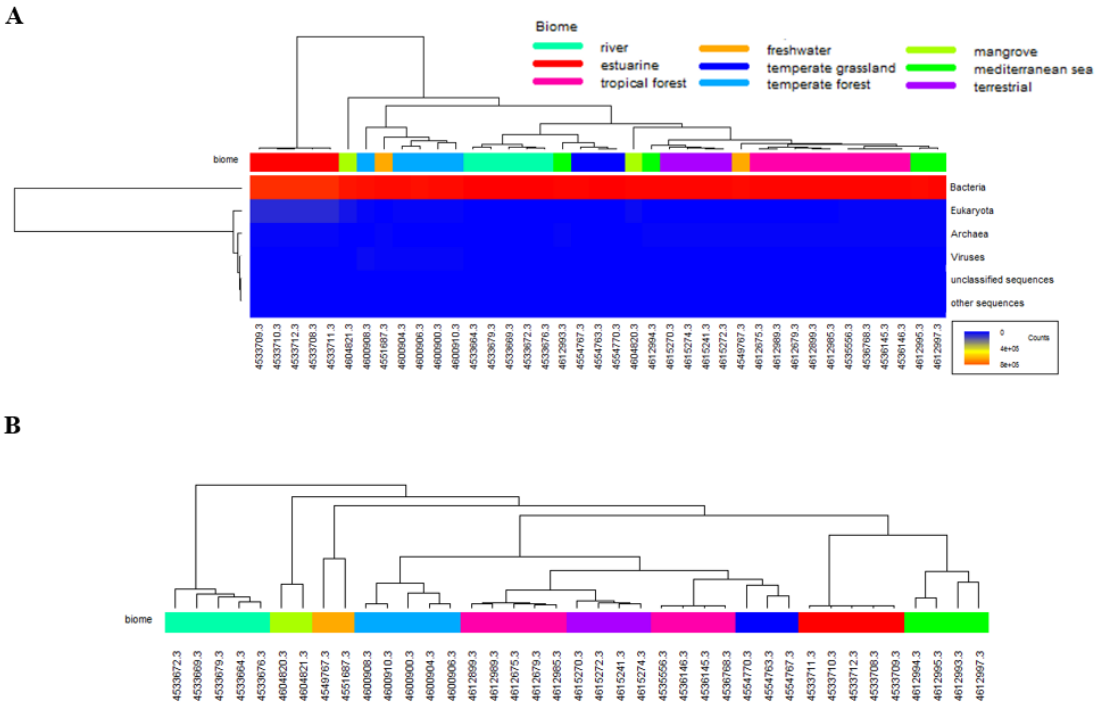


Figure 3. Taxonomic clustering of the evaluated metagenomes as revealed at the domain (A) and genus (B) level. The corresponding biomes are color tagged. The heatmap at panel (B) was omitted due to the large number of genera identified, which compromised the resolution to identify color changes.

Table 3. Difference in hierarchy structure in the evaluated biomes as determined by PERMANOVA.

Hierarchy	Level	Pseudo-F	P
Taxonomic	Domain	205.59	0.001*
	Genus	88.14	0.001*
Functional	Level 1	108.77	0.001*
	Function	39.21	0.001*

Note: (*) $p < 0.05$.

Similarity indices were drawn from the pairwise comparison within and among biomes after clustering (Table 4, Figure 4). The degree of similarity lowers considerably when changing from domain to genus level; the biomes are more diverse the farther we advance in taxonomic classification. At the genus level, the biomes with significant difference in similarity to mangrove are ordered as such: temperate forest, tropical forest,

estuarine and river. However, as revealed by the corrected p-value, none of these pairwise differences is driving the difference in the population.

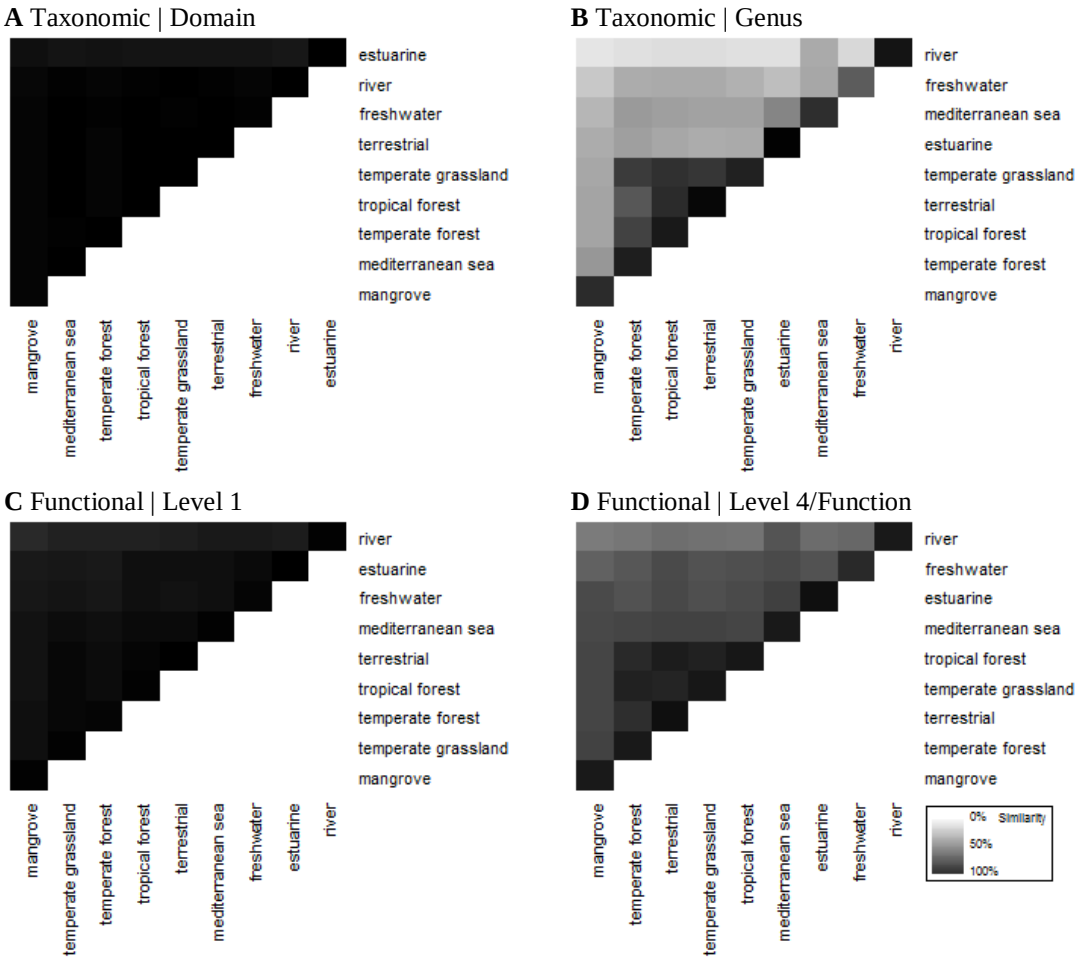


Figure 4. Degrees of similarity within and between biomes after clustering by abundance of the hierarchic classes at the given level. The biomes are ordered in decreasing degree of similarity compared to mangrove. Only the upper triangular matrix is shown.

214 **Table 4.** Pairwise PERMANOVA comparison among mangrove and the other biomes
215 after clustering.

Hierarchy|Level	Biome	Similarity (%)	p-value	p-value corrected
Taxonomy|Domain	Mediterranean Sea	97.4	0.075	1
	Temperate forest	97.4	0.06	1
	Tropical forest	97.4	0.019*	0.684
	Temperate grassland	97.4	0.114	1
	Terrestrial	97.4	0.073	1
	Freshwater	97.2	0.363	1
	River	96.9	0.001*	0.036*
	Estuarine	93.4	0.058	1
Taxonomy / Genus	Temperate forest	40.4	0.048*	1
	Tropical forest	35.6	0.021*	0.756
	Terrestrial	35.3	0.073	1
	Temperate grassland	34.7	0.091	1
	Estuarine	32.1	0.046*	1
	Mediterranean Sea	28.8	0.069	1
	Freshwater	21.2	0.338	1
	River	10.0	0.043*	1
Function / Level 1	Temperate grassland	93.5	0.109	1
	Temperate forest	93.4	0.054	1
	Tropical forest	92.7	0.014*	0.504
	Terrestrial	92.4	0.078	1
	Mediterranean Sea	92.3	0.065	1
	Freshwater	90.2	0.327	1
	Estuarine	89.2	0.052	1
	River	83.5	0.046*	1

216

Hierarchy|Level	Biome	Similarity (%)	p-value	p-value corrected
Function / Level 4	Temperate forest	73.5	0.053	1
	Terrestrial	72.8	0.059	1
	Temperate grassland	72.6	0.090	1
	Tropical forest	72.3	0.025*	0.9
	Mediterranean Sea	71.2	0.073	1
	Estuarine	70.1	0.040*	1
	Freshwater	61.7	0.367	1
	River	51.5	0.053	1

Note: (*) $p < 0.05$.

We performed an analogous analysis considering the functional similarities. Heatmaps were designed for the highest and the lowest hierarchies considered, level 1 and level 4/function (Figure 5). At level 1, metagenomes cluster within their respective biomes. We observe the formation of groups 1) river, 2) estuarine, and 3) the remaining biomes. We also notice that the functions cluster as well, forming three well-defined groups based in abundance. As seen for mangroves, functions related to clustering-based subsystems and carbohydrates are majoritarian. At level 4, all metagenomes but those from tropical forest, which formed two subsets, remain together; main groups are formed in the fashion 1) river, 2) mangrove, temperate forest, temperate grassland, terrestrial and tropical forest, and 3) freshwater, estuarine and Mediterranean Sea. These differences also proved to be significant, as determined by PERMANOVA with biome taken as a factor (Table 3).

The pattern of similarity indices for the pairwise comparison within and among biomes after clustering (Table 4, Figure 4) showed the same decreasing tendency when moving from level 1 to level 4, though much less accentuated. At level 4, only tropical forest and estuarine are significantly different from mangrove, in decreasing order. Again, these pairs did not translate as significant when the whole population was considered.

Taxonomy versus functional hierarchy

When comparing structures within mangrove, the ratio between its taxonomic and functional correlation coefficients at the lowest hierarchic level studied (R^2_{genus}/R^2_{level4}, which will henceforth be called measure of redundancy, R^2_{ratio}), is 0.961, which means at

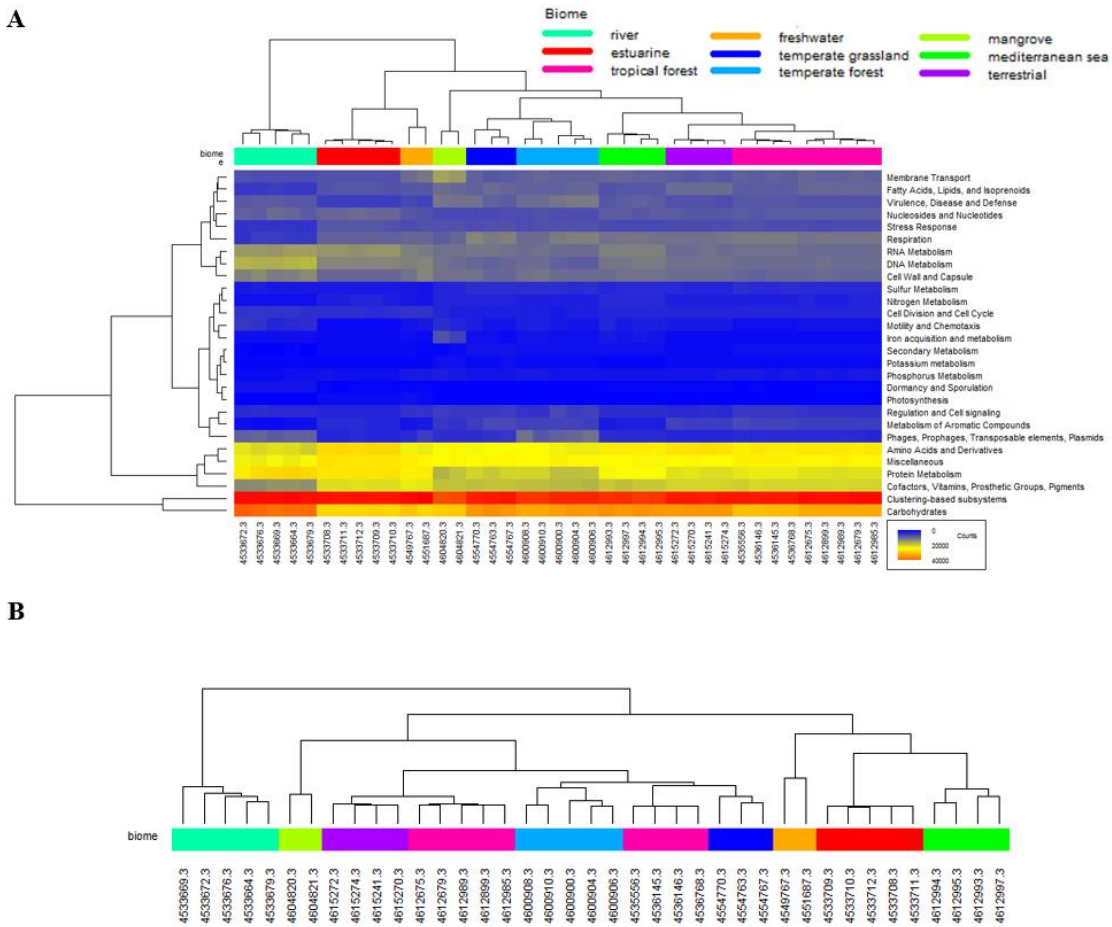


Figure 5. Functional clustering of the evaluated metagenomes as revealed at level 1 (A) and level 4/function (B) level. The corresponding biomes are color tagged. The heatmap at panel (B) was omitted due to the large number of functions identified which compromised the resolution to identify color changes.

this resolution, the difference in taxonomic profile is higher than its functional correlate. For the pairwise comparison between mangrove and the other biomes, this pattern remains, but the indices are also lower than that reported for mangrove alone (Table 5). Thus, different biomes tend to become taxonomically apart while keeping a slower pace of functional dissimilarity. By the three measures considered herein, non-saline soils are closer to mangrove, then saline environments, then non-saline water.

When all biomes are considered in respect to both hierarchies (Figure 3, panel B and Figure 5, panel B), the groups formed at the lowest hierarchic level are similar, with two reallocations: mangrove and freshwater, which formed unitary taxonomic groups, now cluster functionally with other groups, represented by non-saline soils and saline

environments, respectively. Temperate forest is the closest and river is the farthest biome to mangrove by both hierarchy systems (Figure 4, panels B and D).

Table 5. R^2 indices for the biomes considered, in decreasing order of measure of redundancy.

Biome type	Biome	R^2_{genus}	R^2_{level4}	R^2_{ratio}
Non-saline soils	Temperate forest	0.404	0.735	0.550
	Tropical forest	0.356	0.723	0.492
	Terrestrial	0.353	0.728	0.485
	Temperate grassland	0.347	0.726	0.478
Saline environments	Estuarine	0.321	0.701	0.458
	Mediterranean Sea	0.288	0.712	0.404
Non-saline water	Freshwater	0.212	0.617	0.344
	River	0.100	0.515	0.194

Discussion

The two mangrove sites within Paranaguá Bay differ slightly by the taxonomic and functional hierarchic systems. Although the correlation coefficients are high for both systems, they present a smaller taxonomic than functional resemblance, which indicates functional redundancy, which in turn may account for its resiliency. The effect sizes are also higher for taxonomy, meaning the shifts in function abundance are weaker. The same patterns were found in another Brazilian mangrove (Nogueira *et al.*, 2015).

Nogueira *et al.* (2015) also reported a similar community inhabiting mangrove sediments, with a predominance of around 90% of *Bacteria*, most sequences representing phyla *Proteobacteria*, *Bacteroidetes, Actinobacteria, Chloroflexi, Firmicutes* and *Plantomycetes*. The ten most abundant phyla in the current study are predominant in other mangroves (Jiang *et al.*, 2013, Nair *et al.*, 2013, Mendes and Tsai, 2014, Alzubaidy *et al.*, 2016), though the relative abundances appear to be specific to location. Xu *et al.* (2014), when comparing several biomes, found mangroves to be the most enriched in *Proteobacteria*, tendency that we also observed. *Betaproteobacteria* was also found to be the dominant class in a mangrove in southeastern Brazil (Mendes and Tsai, 2014), region close to the one studied herein, and is also cited, along with *Alpha-*, *Delta-* and

Gammaproteobacteria, as dominant in other mangroves in the country (Andreote *et al.*, 2012, Nair *et al.*, 2013, Nogueira *et al.*, 2015). *Deltaproteobacteria* seems to be impoverished in the Paranaguá Bay mangrove according to this system of annotation. The dominating genera, *Achromobacter* and *Stenotrophomonas*, representing together one-fourth of all sequences, are commonly found in soils. Although classified as an aerobic organism, *Achromobacter* may also thrive in anaerobic environments (Swenson and Sadikot, 2015), so their presence is not surprising. It also showed to be one of the genera which abundance is most affected between the two locations.

Functionally, Nogueira *et al.* (2015) report a similar profile as well: a balanced prevalence of sequences related to clustering-based subsystems carbohydrate metabolism, amino acids and derivatives and miscellaneous functions, followed also by cofactors and vitamins, and protein metabolism. The next category (3% of sequences, versus 4% in our analysis) is virulence and disease, which is relevant since human activity is common inside this habitat. For instance, the abundant genus *Achromobacter* is known as an opportunistic pathogen in humans (Swenson and Sadikot, 2015), as well as the species *Stenotrophomonas maltophilia* from the other abundant genus. Another study of mangroves (Andreote *et al.*, 2012) based their functional analysis in NCBI-NR categories, but some similarities are still observed: most of the functions pertain to carbohydrate metabolism, amino acid metabolism, energy metabolism and metabolism of cofactors and vitamins. Within Paranaguá Bay, as shown by the present study, the small effect sizes at levels 1 and 2 mirror their high correlation factors, being comparatively larger only from level 3 on (Figure 2). This observation supports some degree of resiliency, the changes having potentially only a fine-tuning role.

The community comparison among biomes showed that, at the distance considered at the genus level, not only samples from the same biome cluster, but a pattern is also seen for biome type: non-saline soils cluster together, and saline environments and non-saline water, while not forming unitary groups, show subsets of their own. The order of similarity of biomes to mangroves (Table 5) suggests matrix type (soil/water) is more deterministic than chemical composition (saline/non-saline), since the highest degrees are observed for forest soils, then saline environments, than non-saline water. When we assessed functional similarity, the same patterns were observed (Figure 5), with one exception: at level 4, freshwater is closer to saline environments than to river. The order of similarity to mangroves differs slightly by biome but stays the same by biome type (Table 5). The similarity in measure of redundancy (Table 5) follows the same biome

type dependency, which suggests redundancy, therefore resiliency, are also driven by the same factors. The absence of good quality metadata hampers investigating higher-level factors driving clustering. Since environmental conditions determine the degree of similarity among biomes, it is reasonable that the same is true within them, thus for the two mangrove locations from Paranaguá Bay. These factors are not expected to change drastically within the same biome, which explains why similarities within are higher than among. All chemical parameters significantly different are higher in VB than VA (Table 1); the higher mineral content in the sediments in VB should be responsible for at least some of the dissimilarity found.

When compared within itself and to other biomes, the analyzed mangrove is less dissimilar functionally than taxonomically, maintaining a considerable metabolic resemblance even to the most extreme environments despite the higher dissimilarity in community structure. In fact, even a biome as different taxonomically from mangrove as river (R^2_{genus} = 0.100) can be relatively similar at the functional level (R^2_{level4} = 0.515). This indicates that the microbial communities from all these biomes share most functions. Though internal redundancy is characteristic of natural, undisturbed, benthic environments (Holguin *et al.*, 2001), we observed the same at a larger scale. In practice, this can be achieved by the presence of a plethora of organisms possessing the same functions, adapted to the prevailing environmental conditions. It is reasonable to assume that some of these would represent core functions needed for sustainability in any of the given contexts, despite of the organisms harboring them. Altogether, our results suggest that the environment determines the structure of the community as a side effect; the core functions should be adapted to the environment by means of fortuitous organisms, not specific organisms relying on different functions.

All functional comparisons are based solely in the presence of genes, but presence alone does not result in expression. Approaches such as metatranscriptomics might result in a more complete description of the active functions present.

References

Alongi, D.M. (1994) The role of bacteria in nutrient recycling in tropical mangrove and other coastal benthic ecosystems. *Hydrobiologia* 285(1),19-32.

342 Alongi, D.M. (2002) Present state and future of the world's mangrove forests.
343 *Environmental Conservation* 29(03), 331-349.

344 Alzubaidy, H., Essack, M., Malas, T.B., Bokhari, A., Motwalli, O., Kamanu, F. K. *et al.*
345 (2016) Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*)
346 in the Red Sea. *Gene* 576(2), 626-636.

347 Andreote, F.D., Jiménez, D.J., Chaves, D., Dias, A.C., Luvizotto, D.M., Dini-Andreote,
348 F., Fasanella, C.C., Lopez, M.V., Baena, S., Taketani, R.G. (2012) The microbiome of
349 Brazilian mangrove sediments as revealed by metagenomics. *PLoS One* 7(6), e38600.

350 Bouchez, A., Pascault, N., Chardon, C., Bouvy, M., Cecchi, P., Lambs, L., Herteman, M.,
351 Fromard, F., Got, P., Leboulanger, C. (2013) Mangrove microbial diversity and the
352 impact of trophic contamination. *Mar Pollut Bull* 66(1-2), 39-46.

353 Chen, J. (2012) *GUniFrac: Generalized UniFrac distances.* R package version 1.0.
354 https://CRAN.R-project.org/package=GUniFrac

355 Day, A. (2012) *heatmap.plus: Heatmap with more sensible behavior.* R package version
356 1.3. https://CRAN.R-project.org/package=heatmap.plus

357 Dhal, N.K., Sahoo, K. (2009) Potential microbial diversity in mangrove ecosystems: a
358 review. *Indian Journal of Marine Sciences* 38(2), 249-256.

359 Duke, N.C., Ball, M.C., Ellison, J.C. (1998) Factors influencing biodiversity and
360 distributional gradients in mangroves. *Global Ecology and Biogeography Letters* 7(1),
361 27-47.

362 Feller, I.C., Lovelock, C.E., Berger, U., McKee, K.L., Joye, S.B., Ball, M.C. (2010)
363 Biocomplexity in mangrove ecosystems. *Ann Rev Mar Sci* 2, 395-417.

364 Ghizelini, A.M., Mendonça-Hagler, L.C., Macrae, A. (2012) Microbial diversity in
365 Brazilian mangrove sediments - a mini review. *Braz J Microbiol* 43(4), 1242-54.

366 Holguin, G., Vazquez, P., Bashan, Y. (2001) The role of sediment microorganisms in the
367 productivity, conservation, and rehabilitation of mangrove ecosystems: an overview.
368 *Biology and Fertility of Soils* 33(4), 265-278.

369 Jiang, X.T., Peng, X., Deng, G.H., Sheng, H.F., Wang, Y., Zhou, H.W. Tam, N.F.Y.
370 (2013) Illumina sequencing of 16S rRNA tag revealed spatial variations of bacterial
371 communities in mangrove wetland. *Microbial Ecol* 66(1), 96-104.

Marcial-Gomes, N.C., Borges, L.R., Paranhos, R., Pinto, F.N., Mendonça-Hagler, L.C.S., Smalla, K. (2008) Exploring the diversity of bacterial communities in sediments of urban mangrove forests. *FEMS Microbiology Ecology* 66(1), 96-109.

Marques, R., and Motta, A.C.V. (2003) "Análise química do solo para fins de fertilidade", in *Manual de diagnóstico da fertilidade e manejo dos solos agrícolas*, ed. R. M. Lima (Curitiba, Brazil: Universidade Federal do Paraná), 81-102.

Mendes, L. W., Tsai, S.M. (2014) Variations of bacterial community structure and composition in mangrove sediment at different depths in Southeastern Brazil. *Diversity* 6(4), 827-843.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A. et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386.

Nair, H. P., Vincent, H., Bhat, S.G. (2013) Culture independent analysis of the soil microbiome to assess microbial diversity of mangrove soil. *Bio-Genetics J* 1(1), 1-4.

Nogueira, V.L.R., Rocha, L.L., Colares, G.B., Angelim, A.L., Normando, L.R.O., Cantão, M.E., Agnez-Lima, L.F., Andreote, F.D., Melo, V.M.M. (2015) Microbiomes and potential metabolic pathways of pristine and anthropized Brazilian mangroves. *Regional Studies in Marine Science* 2, 56-64.

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2016) *vegan: Community Ecology Package.* R package version 2.3-5. https://CRAN.R-project.org/package=vegan

Pylro, V.S., Roesch, L.F.W., Ortega, J.M., Amaral, A.M., Tótola, M.R., Hirsch, P.R. *et al.* (2014) Brazilian Microbiome Project: revealing the unexplored microbial diversity – challenges and prospects. *Microbial Ecol* 67(2), 237-241.

R Core Team (2016) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Spalding, M., Blasco, F., Field, C. (1997) *World Mangrove Atlas*. Okinawa, Japan: International Society for Mangrove Ecosystems.

Spalding, M., Kainuma, M., Collins, L. (2010) *World Atlas of Mangroves*. UK: Earthscan from Routledge.

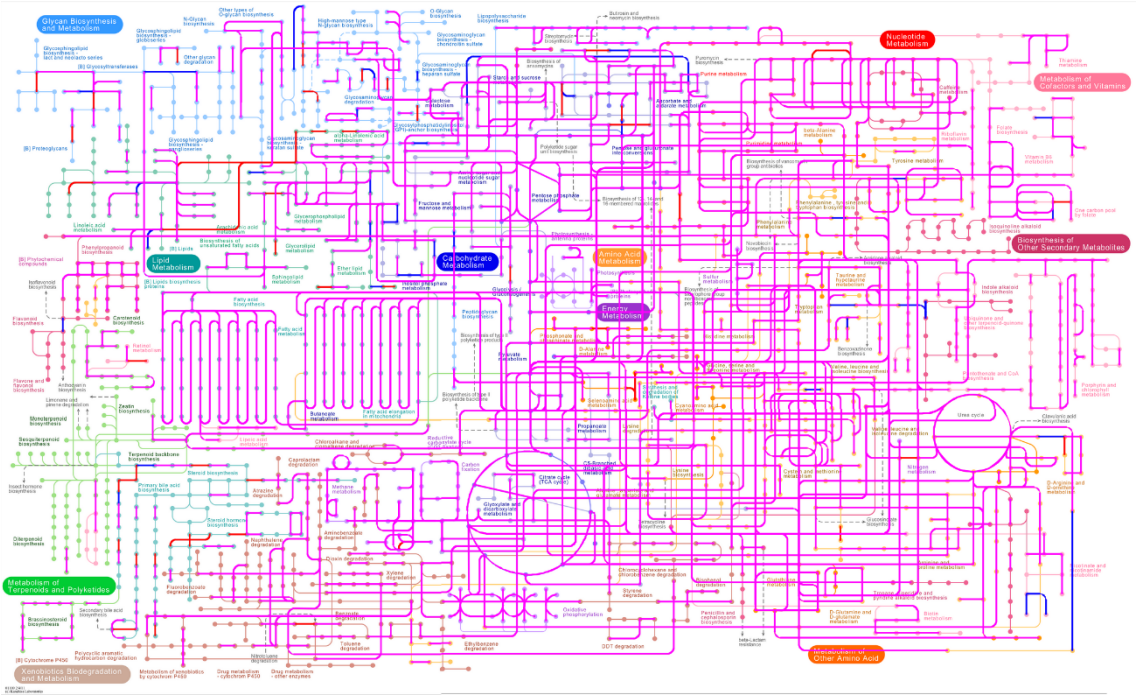
Swenson, C.E., Sadikot, R.T. (2015) Achromobacter respiratory infections. *Annals of the American Thoracic Society* 12(2), 252-258.

Taketani, R.G., Yoshiura, C.A., Dias, A.C., Andreote, F.D., Tsai, S.M. (2010) Diversity and identification of methanogenic archaea and sulphate-reducing bacteria in sediments from a pristine tropical mangrove. *Antonie Van Leeuwenhoek* 97(4), 401-411.

Xu, Z., Hansen, M.A., Hansen, L.H., Jacquiod, S., Sørensen, S.J. (2014) Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE* 9(4), e93445. doi:10.1371/journal.pone.0093445

Yan, B., Hong, K., Yu, Z.N. (2006) Archaeal communities in mangrove soil characterized by 16S rRNA gene clones. *J Microbiol* 44(5), 566-571.

Supplementary figures



Supplementary Figure 1. KEGG map showing functions shared (bold pink lines) and specific in both VA (bold blue) and VB (bold orange) metagenomes.

Supplementary tables

Supplementary Table 1. Metagenomes retrieved from MG-Rast.

Biome	Accession number	Sequences uploaded	Average sequence length (bp)	Identified rRNA features	Identified protein features	Identified functional categories
Estuarine	4533708.3	3,627,820	201.75	8,397	713,263	517,917
	4533709.3	3,657,703	201.75	8,365	722,740	564,546
	4533710.3	3,615,217	201.81	8,322	711,715	516,574
	4533711.3	3,635,746	201.84	8,494	715,239	519,671
	4533712.3	3,628,047	201.61	8,481	716,336	520,328
Freshwater	4549767.3	5,127,922	300	7,957	2,769,992	2,339,931
	4551687.3	4,366,557	300	5,327	2,036,129	1,733,889
Mediterranean Sea	4612993.3	13,224,240	238.26	5,688	2,850,775	2,203,236
	4612994.3	13,978,721	236.66	5,783	2,787,074	2,103,283
	4612995.3	24,768,866	223.48	8,253	4,807,752	3,618,624
	4612997.3	9,373,858	230.34	5,109	2,143,592	1,647,135
River	4533664.3	13,224,240	251	6,346	882,998	424,518
	4533669.3	13,978,721	251	2,236	368,370	181,124
	4533672.3	24,768,866	251	3,184	438,832	205,582
	4533676.3	9,373,858	251	2,661	381,987	181,641
	4533679.3	3,614,659	251	6,885	1,049,242	509,791
Temperate forest	4600900.3	3,172,850	279.72	1,580	1,163,711	886,294
	4600904.3	2,608,969	281.82	924	851,586	648,834
	4600906.3	4,354,330	257.38	1,410	1,273,270	963,225
	4600908.3	3,591,383	280.17	2,433	1,269,169	965,343
	4600910.3	4,680,725	246.85	3,030	1,656,396	1,262,438
Temperate grassland	4554763.3	1,545,693	229.13	1,634	582,042	451,834
	4554767.3	2,378,586	220.83	2,076	841,027	643,608
	4554770.3	1,035,089	235.83	1,067	402,662	314,415
Terrestrial	4615241.3	1,065,596	226.24	699	446,769	355,300
	4615270.3	1,083,940	230.76	755	453,075	359,550
	4615272.3	1,080,277	230.55	766	459,310	365,736
	4615274.3	1,062,667	233.01	836	452,277	359,354
Tropical forest	4535556.3	8,966,780	209.24	3,462	2,534,011	1,972,899
	4536145.3	70,234,163	208.25	32,260	21,740,745	17,023,785
	4536146.3	71,189,722	206.33	27,299	20,054,261	15,598,276
	4536768.3	71,157,224	206.38	32,009	21,420,956	16,757,572
	4612675.3	1,093,748	230.73	666	447,927	357,477
	4612679.3	1,035,679	209.61	552	386,840	308,686
	4612899.3	1,033,077	214.94	527	392,132	313,949
	4612985.3	1,124,037	201.07	431	405,825	324,897
	4612989.3	1,453,973	222.67	715	554,954	442,864

CAPÍTULO 3 – Sequenciamento de amplicons de domínios proteicos pela plataforma Illumina prediz a diversidade de metabolismo secundário em sedimentos do mangue do sul do Brasil

Illumina amplicon-sequencing of protein domains predicts the diversity of secondary metabolism in mangrove sediments from southern Brazil

Abstract

The shortage of new drugs from natural sources has been confronted by the development of new strategies and the search of new molecule types. In this respect, marine microorganisms harbor much potential yet to be unraveled. Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) are enzymatic machineries responsible for the production of secondary metabolites with great structural diversity and bioactivity range. These enzymes are composed of modules containing specific domains. By the amplicon sequencing of such domains, it is possible to predict the diversity of secondary metabolism present in a given sample. We report the diversity of the domains A and C from NRPS, and KS from PKS, as found in mangrove sediments collected from Paranaguá Bay, Brazil. Samples represented three environmental factors: season, location (distance to the ocean) and vegetation present in the sampling spot. Season and location proved to be statistically significant in defining domain-based community structure, while plant had no effect. Diversity measures show no pattern among domains, but hint the most suited groups for prospection: summer | location B for A domains, winter | location B for C domains, and winter | location A for KS domains. Given the possibility of engineering individual modules into functional clusters, even mixing NRPS and PKS components into hybrid systems, the separate investigation of domains adds to the practical value of such approach. Most sequences found are affiliated to genera reportedly involved in NRP and PK synthesis, but novelty is expected given the number of OTUs reported, and the presence of many sequences assigned to uncharacterized environmental bacteria where C and KS domains are concerned. Besides justifying further NRPS and PKS prospection in this mangrove, this study is useful in designing strategies with such goal.

Introduction

After the successful introduction of antibiotics for treating infections in the 1940s, scientists were soon faced with an obstacle, the development of resistance by the pathogenic organisms at which they were aimed. The first efforts for new antibiotics discovery were directed at the isolation of new molecules from cultivable organisms, but this source soon exhausted due to the high rediscovery rate. From late 1980s, new marketed products were developed synthetically or by the modification of previous molecules, but their efficacy was never comparable to the novelty introduced by natural products. Nowadays, the emergence of multi-drug resistant organisms is considered a serious threat, pressing research enterprises into developing new strategies.

The development of new sequencing technologies showed that the diversity of microorganisms far exceeds that revealed by classic cultivation, since most cannot be maintained in artificial conditions. This, however, motivated researchers into exploring the biotechnological potential hidden in this unknown universe. The improvement of isolation techniques (Vartoukian, Palmer and Wade, 2010) unraveled teixobactin (Ling *et al.*, 2015), the first new class of antibiotic discovered in three decades. One year later, lugdunin, a novel cyclic peptide antibiotic produced from a human commensal bacteria, was also reported (Zipperer *et al.*, 2016).

Marine microorganisms are promising sources of new bioactive molecules, producing a large repertoire of pharmaceutically and biotechnologically useful metabolites including antimicrobial, anticancer, anti-inflammatory and antioxidant compounds, antifouling and nutraceuticals agents (reviewed in Reen *et al.*, 2015, Nagarajan *et al.*, 2015, Montaser and Luesch, 2011, Liu *et al.*, 2010). Among these, molecules from secondary metabolism are especially targeted; although they are naturally produced in small amounts, they possess diverse chemical structures, thus broad potential activity spectrum. Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) are enzymatic machineries involved in the synthesis of secondary metabolites with increasing pharmaceutical interest. Though depending on different substrates, both possess structural and functional analogies, being composed of catalytic modules, each one responsible for extending the resulting metabolite by the addition of one new residue. The modules possess domains which confer specificity. In NRPS, the adenylation (A) domain is responsible for recognition and activation of an amino acid, which is transferred to the thiolation (PCP) domain, then attached to the elongating chain by the condensation

(C) domain. In PKS, the acyl-transferase (AT) domain handles the substrate, acetate, then transferred to an analogous thiolation (ACP) domain, and finally ligated to the chain by the ketosynthase (KS) domain. The combinatorial nature of these domains and modules when forming enzymatic assembly lines accounts for the diversity of NRP and PK structures (Rodrigues and McDaniel, 2001).

NRPS and PKS belonging to cultivable marine microorganisms have been described (Nunner, Mevers and Gerwick, 2010), including from mangrove origin (Li *et al.*, 2013). Nevertheless, the fraction of uncultivated microorganisms is still predominant. Current works, many focused on symbiotic microorganisms, are based mostly on genome mining (Baltz, 2008) or whole metagenome sequencing (Schirmer *et al.*, 2005), but targeted strategies such as amplicon-sequencing approaches can achieve higher efficiency (Trindade *et al.*, 2015). By amplifying and sequencing variable regions of C and KS domains in the 454 (Roche) platform, and further comparing with metagenomic whole genome sequencing approaches, Woodhouse *et al.* (2013) reported an unprecedented diverse collection present in marine sponges, many belonging to novel sequences.

We herein described the diversity of A, C and KS domains present in a mangrove from southern Brazil, as revealed by amplicon-sequencing in the MiSeq (Illumina) platform. From metadata analysis, we defined environmental factors driving domain-based community structures. When groups were formed based on these factors, the estimation of diversity measures enabled us to elect the best conditions for prospection of these domains. We also presented the taxonomy assignment of such domains, and how they are related to the groups formed, which aids predicting taxa present in such conditions. This represents our first effort into designing a focused prospection strategy for novel NRPS and PKS enzymes in this environment.

Materials and methods

Sample collection

The sampling scheme was the same adopted for the study presented in Chapter 1, with the addition of samples collected from the same locations and plant surroundings in the winter (September 22nd, 2014). Only one sample representing each season, location and tree was analyzed; in total, the study involved 20 samples. Table 1 shows the information regarding the levels of sampling, the codes used (in bold) and their

coordinates. The upper layer of sediment (around 10 cm) was discarded and around 200 grams of material was aseptically collected and stored in ice until processing.

Table 1. Sampling levels (I, II, III), codes (in bold) and coordinates.

Season (I)	Estuary region (II)	Plant nearby (III)	Coordinates
Summer **(V)** (March 14th, 2014)	Upstream **(B)**	*Avicennia schaueriana* **(A)** *Laguncularia racemosa* **(L)** *Rhizophora mangle* **(R)** *Spartina alterniflora* **(S)** Non-vegetated **(N)**	25°25'02.5"S 48°42'29.9"W
	Downstream **(A)**		25°33'04.5"S 48°26'35.2"W
Winter **(I)** (September 22nd, 2014)	Upstream **(B)**		25°24'27.1"S 48°43'14.6"W
	Downstream **(A)**		25°33'15.6"S 48°24'20.2"W

DNA extraction, amplification and sequencing

The environmental DNA was extracted using the PowerSoil DNA Isolation Kit (MO Bio Laboratories). The 20 samples were amplified for sequencing, targeting the A, C (NRPS) and KS (PKS) domains, in two PCR steps. In the first step, the primers designed contained sequences flanking these domains (see below), and also a T7 promoter- or M13R-sequence 5'-attached to the forward and reverse primers, respectively. In the second, these amplicons had sample-specific barcodes inserted at each end by using primers containing the complementary T7 or M13R sequence as well as a 5'-tail containing the barcode sequence. By using a different collection of indexes at each end, it was possible to pool all samples at the end of the amplification and perform a multiplex (dual) sequencing. The A-domain primer sequences corresponded to that of primers A3F and A7R (Ayuso-Sacido and Genilloud, 2005); the amplification occurred in 35 cycles at an annealing temperature of 57°C. The C-domain sequences were those of primers CnDm and DCCR (Woodhouse *et al.,* 2013), used in 35 cycles at the annealing temperature of 45°C. The KS-domain was amplified with two sets of primers, independently: primer set KS(1), which contained the sequences from degKS2F.i and degKS5R.i (Schirmer *et al.,* 2005), in 35 cycles at an annealing temperature of 56.3°C; and primer set KS(2), with sequences from DKF and DKR (Moffitt & Neilan, 2003), in

35 cycles at an annealing temperature of 55ºC. The PCR products from this first system were resolved in an agarose gel, and then extracted with the Zymoclean Gel DNA Recovery Kit (Zymo Research). 10 ng of each purified amplicon was then submitted to the second PCR protocol, consisting of 30 cycles at the annealing temperature of 65ºC. The resulting products were normalized using a SequalPrep Normalization Plate Kit (Thermo Fisher Scientific). The pool generated by combining all 80 normalized amplifications (20 samples x 4 sets of primers) was sequenced at the Ramaciotti Centre for Genomics (Sydney, Australia), using the MiSeq Platform (Illumina) set in a 2x 300 bp run format.

Reads processing

The reads related to each protein domain should be aligned to a reference aligment. For the construction of reference alignments, we retrieved amino acid sequences corresponding to the domains studied. Sequences for the A domain were obtained from the Conserved Domain Database from NCBI, searching for string "A_NRPS". The 297 aligned sequences were downloaded and trimmed for keeping only the region overlapping the first sequence given, 1AMU_A, which is a well characterized sequence for this domain. Curated sequences for the C and KS domains were retrieved from the Natural Product Domain Seeker (NaPDoS) online tool, totalizing 189 and 458, respectively. For each domain, these amino acid sequences were blasted against the NCBI nucleotide database, using the tblastn tool from blast+ v2.2.31 (Altschul *et al.*, 1990), and the aligned regions of the results with the highest hit score for each pair query/subject with maximum e-value 1E-3 were downloaded. Since the variability of these domains at amino acid level is preserved when clustering nucleotides with a 90% identity cutoff (Varaljay et al., 2010), the downloaded sequences were clustered at a 0.10 distance threshold using usearch v8.0.1623 (Edgar, 2010), and then finally aligned within themselves using muscle v3.8.31 (Edgar, 2004) set to perform three iterations.

Mangrove reads were processed as implemented in mothur v1.37.0 (Schloss *et al.*, 2009). They were first combined by domain. After filtering for quality (window average 30, window size 50, min. length 200, max. homopolymers 8, max. ambiguity 0), those corresponding to each domain were aligned to their specific reference aligment generated previously. Reads that aligned to the densest region, as given by the number of reads covering that region, were maintained. Empty base information from the extremities was

removed and the reads were pre-clustered to remove sequences resulting from unspecific amplification, then sub-sampled at the level of the sample with fewer reads. Uncorrected pairwise distance was calculated within each dataset to allow its posterior clustering, and finally OTU representatives were designated from the resulting clustered files. At the end, we obtained a OTU table representing the profile of each sample as given by the number of OTUs at 0.10 distance threshold that they contained.

Taxonomy assignment and profiling

In order to assign taxonomy, a reference database was also necessary. First, the same reference amino acid sequences for the A, C and KS domains described above were blasted against the NCBI NR database, using the blastp tool from blast+ v2.2.31 (Altschul *et al.,* 1990). The aligned regions of the results with the highest hit score for each pair query/subject with maximum e-value 1E-3 were downloaded, while the sequence ID was maintained at the sequence header, from which taxonomy can be later derived. Tool makedb from DIAMOND v0.8.4 (Buchfink, Xie and Huson, 2014) was used to generate a database file from the sequences retrieved.

The quality-filtered mangrove reads, separated by sample, were blasted against the corresponding database using tool blastx from DIAMOND v0.8.4 (Buchfink, Xie and Huson, 2014). The resulting files had taxonomy assigned using tool daa2rma from MEGAN v6.4.5 (Huson *et al.,* 2011), using the gi2tax mapping file provided by the software dated February 2016. All samples from the same protein domain were taxonomically compared after loading the taxonomy-assigned files in the graphical version of MEGAN v6.4.5 (Huson *et al.* 2011). The analysis was performed after normalization at the minimum read count; unassigned sequences were removed. A taxonomic profile, given as the relative abundance at the genera level, was generated for each protein domain.

Comparisons and statistics

The OTU tables and the taxonomic profiles generated were loaded into R v3.3.0 (R Core Team, 2016). OTU tables were subsampled at the minimum per-sample count using package GUniFrac (Chen, 2012). Package vegan (Oksanen *et al.,* 2016) was required to calculate Bray-Curtis distances among samples, to perform clustering using

the average neighbor method, and to derive species richness and α-diversity measures within sample groups. The impact of environmental factors was calculated by PERMANOVA (n=999). PCoA plots were designed based on the distance matrix, to show a graphical representation of the relationship among the samples. The environmental factors and taxa that showed statistical significance in shaping community structure were fitted into the ordination by function envfit from package vegan (Oksanen *et al.*, 2016).

Results

Reads processing

The OTU table for the A domain summed 2130 different OTUs, with counts ranging from 157 to 189 per sample. The C domain had 2758 OTUs, counts in the range of 128 to 277. Primer set KS(1) resulted in 2092 OTUs, counting from 326 to 667, while primer set KS(2) was not as efficient, showing only 458 OTUs in total. For this reason, the OTU table resulting from primer set KS(1) was chosen as representative of this protein domain.

Factors driving diversity

Average neighbor clustering derived from the Bray-Curtis distance calculated among sample, for each of the domains addressed (Figure 1), showed that clusters are formed separating downstream (A) and upstream (B) samples; within each location, samples are further separated by the season of collection. For the A domain, samples VA are most similar, while samples IB are the most disperse. The C domain shows samples VB most clustered, with samples IA presenting the largest distances. For the KS domain, the most clustered group is VA, and IA the most dispersed. When the plants from which samples were taken are concerned, no pattern of distribution is discernible.

The PERMANOVA evaluation of the three environmental factors shows that season and location had statistical significance in defining groups, while plant had no effect (Table 2) confirming the observation on the clustering patterns. For this reason, samples were considered as belonging to four groups (season x location) for the remaining analyses. For all three protein domains, location is a stronger predictor of group

separation, as given by the pseudo-F coefficients. The separation is more pronounced for the KS domain for both significant factors, as given by the pseudo-F factor.

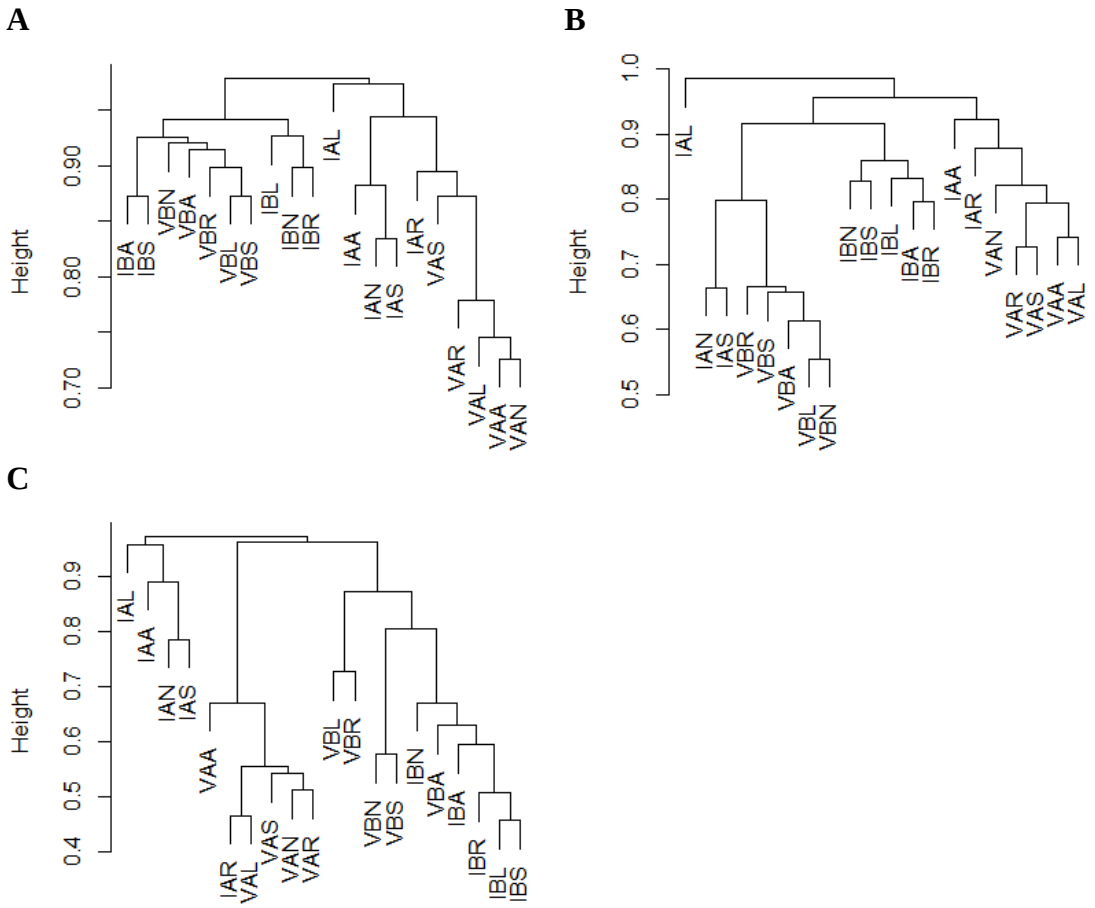


Figure 1. Cluster plots by the average neighbor method calculated on Bray-Curtis distance matrixes. A: A domain, B: C domain, C: KS domain.

Table 2. PERMANOVA test for assessing the statistical significance of the factors driving community structure.

	A domain		C domain		KS domain	
Factor	Pseudo-F	p-value	Pseudo-F	p-value	Pseudo-F	p-value
Season	1.34	0.029*	1.54	0.031*	2.03	0.028*
Location	2.39	0.001***	3.11	0.001***	5.23	0.001***
Plant	0.84	0.999	0.74	0.998	0.63	0.997

Note: (*) $p < 0.05$, (***) $p < 0.001$.

Diversity measures

The highest species richness and α-diversity measures were found for the A domain within location B, where diversity far exceeds that from all other groups. Diversity is second highest for the same A domain within location A, but richness is second highest for the KS domain within location A in winter. The lowest richness indices were reported for the C domain, while diversity is smallest for the KS domain within location A in the summer. When comparing group changes of the same domain, both indices are directly proportional for the whole dataset, but there are patterns domain-specific (Table 3). For the A domain, both indices are higher in location B within the same season, though the difference is larger in the summer. The indices for the C domain do not differ in the summer, but both are higher in location B in the winter. For the KS domain, the shift in location in the summer causes both indices to increase, while the same shift in the winter makes them to decrease.

Table 3. Richness and diversity measures for the groups assembled.

Factor	A domain		C domain		KS domain	
	Richness	α-diversity	Richness	α-diversity	Richness	α-diversity
Loc A						
Summer	465	479.6	276	184.2	390	162.4
Winter	515	651.0	326	266.0	585	326.9
Loc B						
Summer	606	1219.2	276	184.2	563	304.5
Winter	593	1104.3	378	387.8	494	241.1

Taxonomic affiliation

While primers KS(1) resulted in a more comprehensive OTU table, primers KS(2) performed better for taxonomy assignment; since the primers covered the same domain, the taxonomic profile generated from KS(2) was chosen to complement the data from the OTU assignment. The A domain has 99.6% of sequences assigned to *Bacteria*. Among the genera identified, there is a prevalence of (30.4±11.3)% *Myxococcus,* (14.2±5.3)% *Brevibacillus,* (11.5±6.0)% *Streptoalloteichus,* (6.8±3.8)% *Bacillus,* (5.7±3.2)%

Sinorhizobium, (4.6±4,6)% *Saccharothrix* and (4.3±2.1)% *Streptomyces,* which represent more than 75% of the sequences (Figure 2). The C domain shows 100% of sequences assigned to *Bacteria.* Among the abundant genera, (37.8±21.3)% are *Moorea,* (15.5±13.6)% are annotated as bacterial environmental samples, (7.6±7.8)% *Silvibacterium,* (5.6±6.1)% *Paenibacillus,* (5.1±13.2)% *Pseudoalteromonas* and (3.9±6.2)% *Streptomyces,* which also account for more than 75% of the sequences (Figure 3). The KS domain has also 100% of sequences belonging to *Bacteria*. Among the genera, there is a high prevalence of (84.4±11.4)% of sequences annotated as environmental bacterial samples, then (4.0±10.8)% belonging to *Lysobacter* and (3.6±3.6)% of unclassified *Peptococcaceae,* making up more than 90% of the sequences (Figure 4).

The taxonomic profile at the genus level for each domain was overlaid on PCoA plots representing the distance among samples calculated after the OTU tables (Figure 5). Only genera that showed statistical significance on the configuration of sample distribution were plotted. The centroids for the four environmental factors are also shown. For the A domain (panel A), we notice that *Myxococcus* is enriched in the summer, *Brevibacillus* is enriched in the group winter | location B, *Streptomyces* is more related to winter, while the other abundant genera are not driving community structure. Among the less abundant genera, we still observe a tight correlation between *Legionella* and location B, *Actinoplanes* and summer, and *Planktothris* and winter. For the C domain (panel B), *Moorea* is somewhat related to summer, while among minoritarian genera we see that location A is characterized by an enrichment in *Burkholderia*, and winter by an enrichment in *Candidatus Magnetomorum, Luteibacter* and bacterium UASB270. The KS domain (panel C) shows that the abundant environmental taxa are tightly correlated with summer, which is also enriched for minor genera *Labrenzia, Sorangum,* unclassified *Gammaproteobacteria* and *Rhodopirellula;* other minor group, *Streptomyces*¸ shows a tight correlation with location A.

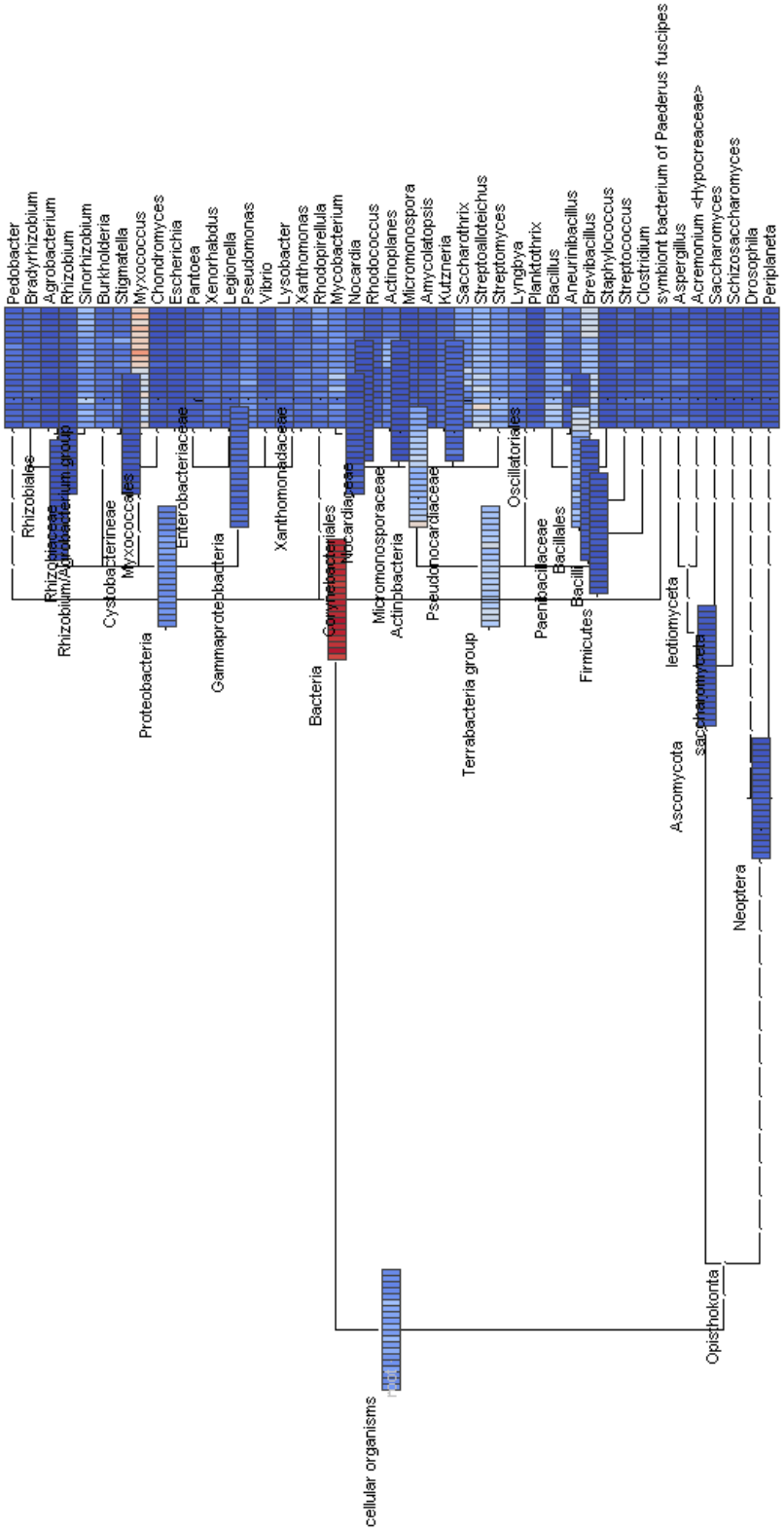


Figure 2. Taxonomic affiliation for the A domain at the genus level. Each square represents one sample in the order IAA, IAL, IAN, IAR, IAS, IBA, IBL, IBN, IBR, IBS, VAA, VAL, VAN, VAR, VAS, VBA, VBL, VBN, VBR, VBS. The colors represent the abundance assigned at that taxonomic level; blue = 0, red = 100%.

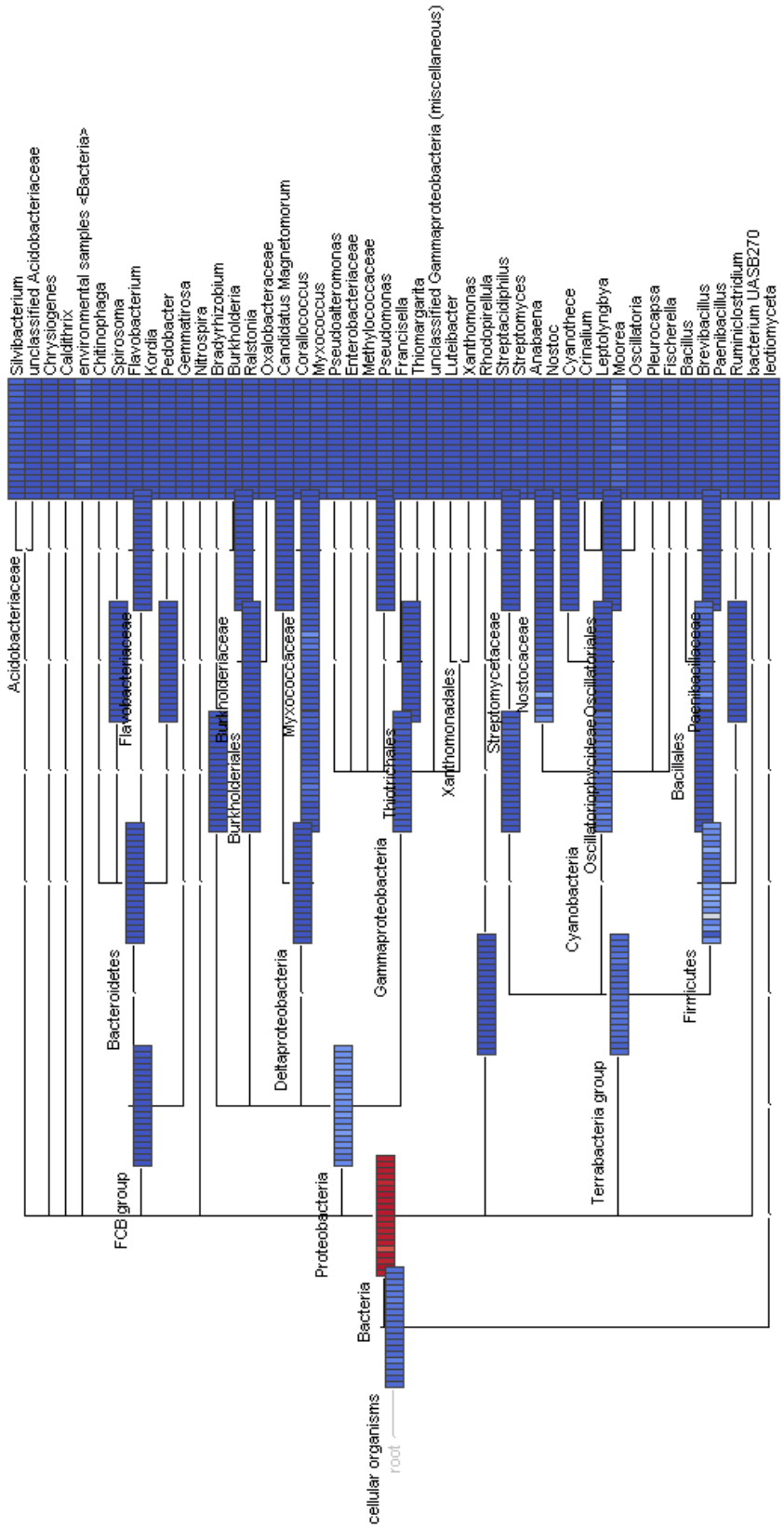


Figure 3. Taxonomic affiliation for the C domain at the genus level. Each square represents one sample in the order IAA, IAL, IAN, IAR, IAS, IBA, IBL, IBN, IBR, IBS, VAA, VAL, VAN, VAR, VAS, VBA, VBL, VBN, VBR, VBS. The colors represent the abundance assigned at that taxonomic level; blue = 0, red = 100%.

285

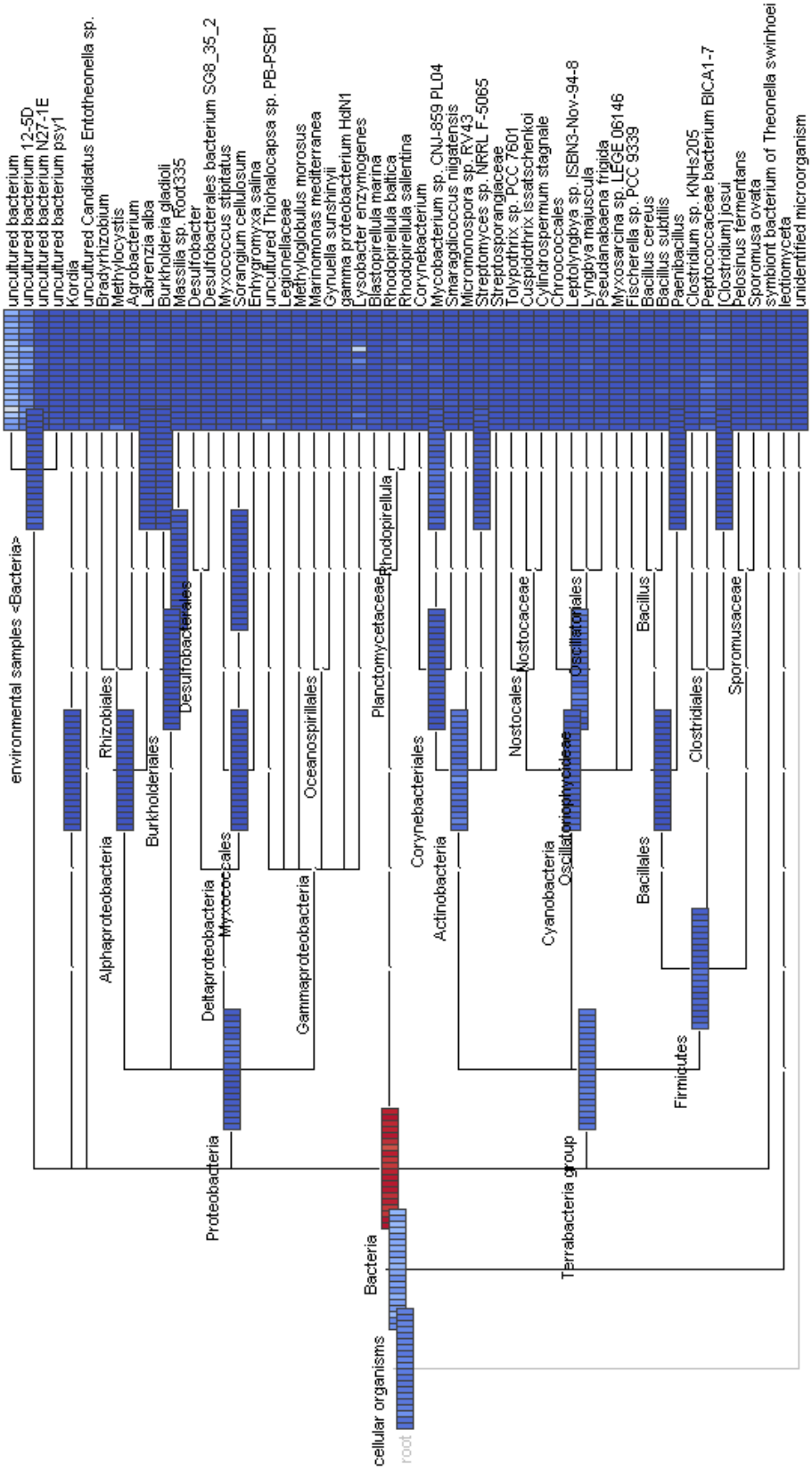
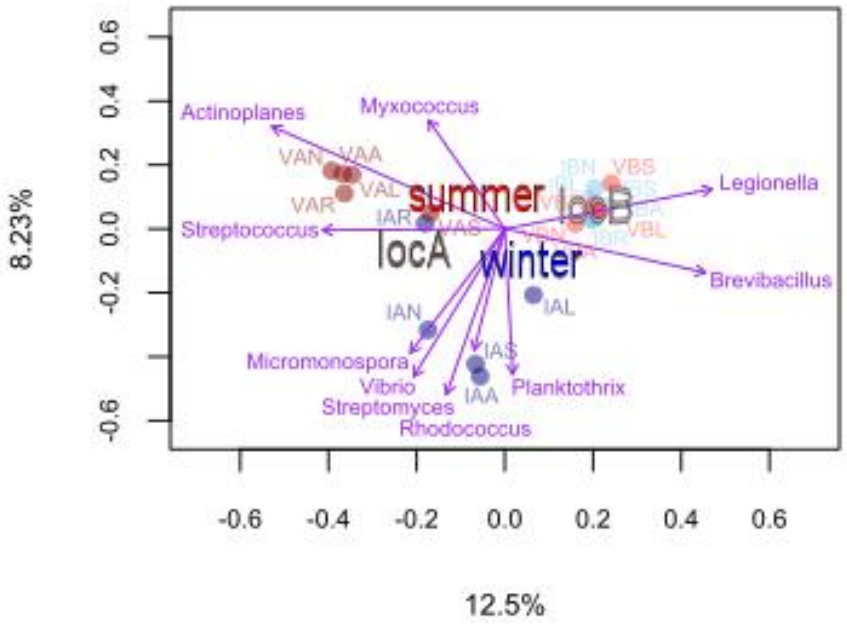


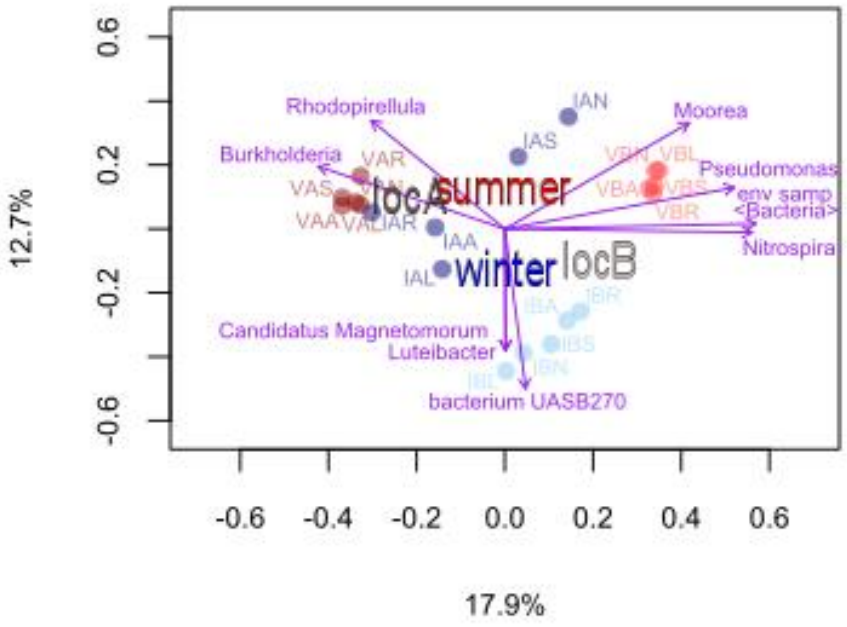
Figure 4. Taxonomic affiliation for the KS domain at the genus level. Each square represents one sample in the order IAA, IAL, IAN, IAR, IAS, IBA, IBL, IBN, IBR, IBS, VAA, VAL, VAN, VAR, VAS, VBA, VBL, VBN, VBR, VBS. The colors represent the abundance assigned at that taxonomic level; blue = 0, red = 100%.

286

A


8.23%
12.5%
Actinoplanes
Myxococcus
Legionella
Streptococcus
Brevibacillus
Micromonospora
Vibrio
Streptomyces
Rhodococcus
Planktothrix
summer
winter
locA
locB
VAN
VAA
VAL
VAR
IAR
IAS
IAA
IAL
IAN
VBS
VBL
VBR
IBR
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

B


12.7%
17.9%
Rhodopirellula
Moorea
Burkholderia
Pseudomonas
env samp
<Bacteria>
Nitrospira
Candidatus Magnetomorum
Lutelbacter
bacterium UASB270
summer
winter
locA
locB
IAN
IAS
IAA
IAL
VAR
VAS
VAA
VAL
IAR
VBN
VBL
VBA
VBS
VBR
IBA
IBR
IBS
IBN
IBE
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

287

C

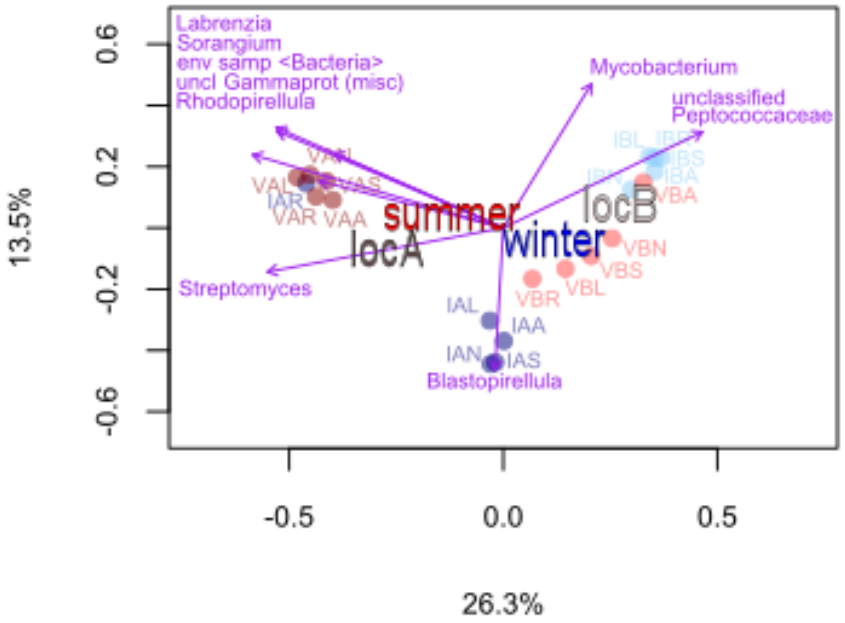


Figure 5. PCoA plots representing the relative distance among samples; centroids for the significant factors shaping community structure are also shown. A: A domain, B: C domain, C: KS domain. The purple arrows represent the genera found to significantly affect the community structure; their length is a measure of the correlation in that direction. The axis represent the first and the second components of the analysis and their relative weight explaining the variation observed.

Discussion

This work expanded one previous report (Woodhouse *et al.*, 2013) which investigated the diversity of C and KS domains in marine sponges by next generation sequencing. We applied the same approach to a larger environment. The number of OTUs obtained herein is considerably higher, possibly reflecting the more diverse nature of the organisms inhabiting mangroves. Even though reads and reference files were the same for both protocols, the generation of an OTU table using primers KS(2) (DKF/DKR) proved to be inefficient with the pipeline adopted, while taxonomy could not be assigned with the reads provided by primers KS(1) (degKS2F.i /degKS5R.i). Since the primers chosen for domains A and C worked well for both processes, we believe the protocols to be solid, but we could not answer why the issue above happened. Since primers KS(1)

and KS(2) cover the same domain, we decided to join the results to represent the KS domain in our study.

We found samples to cluster by factors season and location, which is evidenced in the cluster plots (Figure 1), in the PERMANOVA analysis (Table 2) and in the PCoA plots (Figure 5). PERMANOVA showed location is a better predictor of group separation, which is again evident in the PCoA plots by the larger distance between locations than between seasons. The plants from which rhizosphere samples were collected showed no effect on diversity of these domains, mirroring the results obtained for 16S diversity (Chapter 1 of this thesis); soil characteristics in this habitat seem to make any plant effect irrelevant (Marschner, Crowley and Yang, 2004). The delimitation of groups is valuable to reduce sampling and processing efforts if the prospection of novel domains is to follow up.

While Woodhouse *et al.* (2013) report greater richness for the C domain and greater diversity for the KS domain in sponges, we observed the KS domain presenting higher richness, while diversity is context dependent: the C domain is more diverse in groups winter | location A and summer | location B, while the KS domain in the other groups. In the mangrove, α-diversity is expected to increase with species richness for all domains. This is the best scenario for prospection, since new species resulting from shifts in environmental factors are to be of a diverse nature than those already present. If taxonomy is not considered, this makes samples retrieved in the summer from location B ideal for prospection of A domains; samples from winter | location B preferred for prospection of C domains; and samples from winter | location A best suited for KS domains. Moreover, the best groups for each domain are different, which possibly reflects the modular nature of the enzymatic machineries containing them.

The fact that nearly 100% of sequences were assigned to *Bacteria* is not surprising, since the primers are designed for these organisms; this at least confirms their efficiency in amplifying with the desired specificity. Nevertheless, the profiles at genus level are very diverse for the enzyme domains. Among the most abundant genera reported, only *Streptomyces* is present in more than one protein domain study, while posing a great promise for the prospection of secondary metabolites (Bérdy, 2005). Some genera from the A-domain study reportedly producing NRPs are *Myxococcus* (Cortina *et al.*, 2012, Cortina *et al.*, 2011), *Brevibacillus* (Djukic *et al.*, 2011), *Streptoalloteichus* (Lohman *et al.*, 2013, Tao *et al.*, 2007) and *Bacillus* (Chen *et al.*, 2006, Ehling-Schulz *et al.*, 2005, Emmert *et al.*, 2004). The abundant *Moorea* in the C-domain study is also

implicated in the production of such molecules (Thomburg *et al.*, 2013, Choi *et al.*, 2012), including new strains with prolific secondary metabolism (Engene *et al.*, 2012), as well as genus *Lysobacter* (Xie *et al.,* 2012), found in the KS-domain research. Sequences annotated as belonging to yet uncharacterized environmental bacteria are present in the C and in the KS-domain datasets, being particularly abundant in the last, possibly representing novelty. If one now considers taxonomy, choosing prospection sites based on the diversity measures does not help much predicting the genera that are expected to be enriched; among the three protein domains, only the minor taxon *Blastopirellula* lies close to the group suggested for KS-domain prospection, winter | location A; no information regarding PK synthesis by this genus was found. The sequences annotated as environmental bacteria will be enriched in the summer | location B group for the C domain, and in the summer groups for the KS domain.

The technique applied proved to be efficient in evaluating the diversity of NRPS and PKS domains in an environment for which no prior knowledge regarding secondary metabolite production was available. The mangrove studied revealed to be a promising source for prospection of NRPS and PKS enzymes. A first screening according to sample type was suggested. Sequence reads may now be retrieved and used for designing primers aimed at recovering the enzymatic clusters in which these domains are integrated, which will aid the prospection in a fosmid library and the extraction of full sequences for heterologous expression of their products.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. **J. Mol. Biol.**, v. 215, p. 403-410.

Ayuso-Sacido, A., Genilloud, O. (2005) 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. **Microb. Ecol.**, v. 49(1), p. 10-24.

Baltz, R.H. (2008) Renaissance in antibacterial discovery from actinomycetes. **Current Opinion in Pharmacology**, v. 8(5), p. 557-563.

Bérgy, J. (2005) Bioactive microbial metabolites: a personal view. **J. Antibiot.**, v. 58(1), p. 1-26.

Buchfink, B., Xie, C., Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, p. 59-60. doi:10.1038/nmeth.3176

Chen, J. (2012) GUniFrac: Generalized UniFrac distances. R package version 1.0. https://CRAN.R-project.org/package=GUniFrac

Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Gottschalk, G. (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. **Journal of Bacteriology**, v. 188(11), p. 4024-4036.

Choi, H., Mevers, E., Byrum, T., Valeriote, F. A., Gerwick, W. H. (2012) Lyngbyabellins K–N from two Palmyra Atoll collections of the marine cyanobacterium *Moorea bouillonii*. **European Journal of Organic Chemistry**, v. 2012(27), p. 5141-5150.

Coates, A.R.M., Halls, G., Hu, Yanmin. (2011) Novel classes of antibiotics or more of the same? **Br. J. Pharmacol.**, v.163(1), p. 184–194. doi: 10.1111/j.1476-5381.2011.01250.x

Cortina, N.S., Krug, D., Plaza, A., Revermann, O., Müller, R. (2012) Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. **Angewandte Chemie International Edition**, v. 51(3), p. 811-816.

Cortina, N.S., Revermann, O., Krug, D., Müller, R. (2011) Identification and characterization of the althiomycin biosynthetic gene cluster in *Myxococcus xanthus* DK897. **Chembiochem**, v. 12(9), p. 1411-1416.

Djukic, M., Poehlein, A., Thürmer, A., Daniel, R. (2011) Genome sequence of *Brevibacillus laterosporus* LMG 15441, a pathogen of invertebrates. **Journal of Bacteriology**, v. 193(19), p. 5535-5536.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32(5), p. 1792-1797.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. **Bioinformatics**, v. 26(19), p. 2460-2461.

Ehling-Schulz, M., Vukov, N., Schulz, A., Shaheen, R., Andersson, M., Märtlbauer, E., Scherer, S. (2005) Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*. **Applied and Environmental Microbiology**, v. 71(1), 105-113.

Emmert, E.A., Klimowicz, A.K., Thomas, M.G., Handelsman, J. (2004) Genetics of zwittermicin A production by *Bacillus cereus*. **Applied and Environmental Microbiology**, v. 70(1), p. 104-113.

Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., Gerwick, W. H. (2012) *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. **International Journal of Systematic and Evolutionary Microbiology**, v. 62(5), p. 1171-1178.

Huson, D.H.,, Mitra, S., Ruscheweyh, H.J., Weber, N., Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. **Genome Res.**, v. 21, p. 1552-1560.

Li, X., Tang, X., Xiao, J., Ma, G., Xu, L., Xie, S. (2013) Harnessing the potential of halogenated natural product biosynthesis by mangrove-derived actinomycetes. **Mar. Drugs**, v. 11(10), p. 3875-3890. doi:10.3390/md11103875

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L, Engels, I., Conlon, B.P. *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. **Nature**, v. 517(7535), p. 455-459. doi:10.1038/nature14098

Liu, X., Ashforth, E., Ren, B., Song, F., Dai, H., Liu, M., *et al.* (2010) Bioprospecting microbial natural product libraries from the marine environment for drug discovery. **The Journal of Antibiotics**, v. 63, p. 415-422. doi:10.1038/ja.2010.56

Lohman, J.R., Huang, S.X., Horsman, G.P., Dilfer, P.E., Huang, T., Chen, Y., Shen, B. (2013) Cloning and sequencing of the kedarcidin biosynthetic gene cluster from *Streptoalloteichus* sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics. **Molecular BioSystems**, v. 9(3), p. 478-491.

Marschner, P.; Crowley, D.; Yang, C. H. (2004) Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. **Plant and Soil**, v. 261, p. 199-208.

Moffitt, M.C., Neilan, B.A. (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. **J. Mol. Evol.**, v. 56, p. 446-457.

Montaser, R., Luesch, H. (2011) Marine natural products: a new wave of drugs? **Future Medicinal Chemistry**, v. 3(12), p. 1475-1489.

Nagarajan, M., Rajesh Kumar, R., Meenakshi Sundaram, K., Sundararaman. (2015) Marine biotechnology: potentials of marine microbes and algae with reference to pharmacological and commercial values. In “Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology”, Bahadur, B., Venkat, R. Manchikatla, Sahijram, L., Krishnamurthy, V. K. (eds)", Springer India, New Delhi, p. 685-723. doi:10.1007/978-81-322-2283-5_35

Nunnery, J. K., Mevers, E., Gerwick, W.H. (2010) Biologically active secondary metabolites from marine cyanobacteria. **Current Opinion in Biotechnology**, v. 21(6), p. 787-793.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al.* (2016) vegan: Community Ecology Package. R package version 2.3-5. https://CRAN.R-project.org/package=vegan

R Core Team. (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Reen, F. J., Gutiérrez-Barranquero, J. A., Dobson, A. D. W., Adams, C., O’Gara, F. (2015) Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. **Marine Drugs**, v. 13(5), p. 2924-2954. doi:10.3390/md13052924

Rodriguez, E., McDaniel, R. (2001) Combinatorial biosynthesis of antimicrobials and other natural products. **Current Opinion in Microbiology**, v. 4(5), p. 526-534.

Schirmer, A., Gadkari, R., Reeved, C.D., Ibrahim, F., DeLong, EF., Hutchinson, C.R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissolute*. **Appl. Environ. Microbiol.**, v. 71(8), p. 4840-4849. doi:10.1128/AEM.71.8.4840-4849.2005

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E., Hutchinson, C.R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. **Appl. Environ. Microbiol.**, v. 71(8), p. 4840-4849. doi:10.1128/AEM.71.8.4840-4849.2005

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B, *et al.* (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. **Appl. Environ. Microbiol.**, v. 75(23), p. 7537-7541.

Tao, M., Wang, L., Wendt-Pienkowski, E., George, N.P., Galm, U., Zhang, G., Shen, B. (2007) The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. **Molecular BioSystems**, v. 3(1), p. 60-74.

Thornburg, C. C., Cowley, E. S., Sikorska, J., Shaala, L. A., Ishmael, J. E., Youssef, D. T., McPhail, K. L. (2013) Apratoxin H and apratoxin A sulfoxide from the Red Sea cyanobacterium Moorea producens. **Journal of Natural Products**, v. 76(9), p. 1781-1788.

Trindade, M., van Zyl, L. J., Navarro-Fernández, J., Abd Elrazak, A. (2015) 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, p. 890. doi:10.3389/fmicb.2015.00890

Varaljay, V.A., Howard, E.C., Sun, S., Moran, M.A. (2010) Deep sequencing of a dimethylsulfoniopropionate-degrading gene (dmdA) by using PCR primer pairs designed on the basis of marine metagenomic data. **Appl. Environ. Microbiol.**, v. 76, p. 609-617.

Vartoukian, S.R., Palmer, R.M., Wade, W.G. (2010) Strategies for culture of 'unculturable' bacteria. **FEMS Microbiol. Lett.**, v. 309(1), p. 1-7. doi:10.1111/j.1574-6968.2010.02000.x

Woodhouse, J.N., Fan, L. Brown, M.V., Thomas, T., Neilan, B.A. (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. **ISME J.**, v. 7, p. 1842-1851.

Xie, Y., Wright, S., Shen, Y., Du, L. (2012) Bioactive natural products from *Lysobacter*. **Nat. Prod. Rep.**, v. 29, p. 1277-1287. doi:0.1039/c2np20064c

Zipperer, A., Konnerth, M.C., Laux, C., Berscheid, A., Janek, D., Weidenmaier, C., *et al.* (2016) Human commensals producing a novel antibiotic impair pathogen colonization. **Nature**, v. 535 (7613), p. 511-516. doi:10.1038/nature18634

CONSIDERAÇÕES FINAIS

Através do uso de ferramentas metagenômicas, pudemos caracterizar o microbioma do mangue selecionado de acordo com três categorias: a comunidade microbiana revelada por sequências do gene de rRNA 16S, o perfil funcional revelado pelo sequenciamento do metagenoma, e a diversidade de domínios relacionados às enzimas NRPS e PKS, produtoras de metabólitos secundários.

A estrutura da comunidade se mostrou influenciada por fatores ambientais relacionados à distância do mar, que se refletiram na quantidade de íons retidos pela matriz orgânica presente nos sedimentos. Não observamos efeito de rizosfera das plantas características do local. Uma vez que o funcionamento do mangue depende de uma relação fina entre microbiota e vegetação, acreditamos que a comunidade presente é capaz de interagir com a mesma eficiência com as diferentes plantas.

Como apenas a localização influenciou a identidade dos microrganismos habitantes, escolhemos uma amostra representando cada localização para o sequenciamento do metagenoma integral. Os dados gerados nos permitiram acessar a estrutura da comunidade por outro sistema de anotação e o perfil funcional do habitat. O perfil taxonômico é diferente daquele apresentado no estudo baseado em rRNA 16S, o que pode ser atribuído aos vieses típicos de cada abordagem. Atualmente, a anotação taxonômica por 16S em estudos metagenômicos têm sido preferida, já que a relação dessas sequências com árvores filogenéticas é bem consolidada. Comparando os perfis taxonômico e funcional do mangue com outros biomas publicamente disponíveis, pudemos observar uma tendência de agrupamento baseada em tipo de bioma, o que indica que níveis hierárquicos superiores de fatores ambientais também conduzem a estrutura da comunidade microbiana. Essa constatação dá suporte à geração de perfis baseados em informação genética para caracterizar um ambiente, e, consequentemente, seu estado frente a distúrbios como os causados por eutrofização. Assim como a medicina caminha para o diagnóstico genético, estratégias de conservação e recuperação de ambientes também podem se beneficiar dessa abordagem. Todos os biomas investigados são funcionalmente menos dissimilares que taxonomicamente, o que indica a existência de diferentes microrganismos adaptados para as condições ambientais presentes mas realizando as mesmas funções.

A diversidade dos domínios de NRPS e PKS reportada confirma as expectativas encontradas na literatura sobre o potencial biotecnológico dormente em ambientes

marinhos. A adição de amostras provenientes de outra estação do ano mostrou que, quando esses domínios proteicos são considerados, fatores sazonais também têm influência na definição das funções presentes. Medidas de riqueza e diversidade baseadas em estação do ano e localização nos permitem sugerir em quais coleções de amostras a prospecção de metabólitos secundários sintetizados por essas enzimas pode ser mais prolífica. Levando em consideração que, comparativamente com outras funções, essas tendem a estar presentes em baixa quantidade, estudos direcionados tendem a ser mais eficientes, exigindo menos recursos e tendo maior taxa de sucesso.

O trabalho apresentado agrega à caracterização do microbioma do mangue numa época onde pesquisadores se esforçam para criar bases de dados ambientais numa resolução inédita, e propõe as condições a serem adotadas em futuros estudos de prospecção.

REFERÊNCIAS

Alongi, D. M. (1994) The role of bacteria in nutrient recycling in tropical mangrove and other coastal benthic ecosystems. **Hydrobiol.**, v. 285, p. 19-32.

Alongi, D.M. (2002) Present state and future of the world's mangrove forests. **Environmental Conservation**, 29(03), 331-349.

Alongi, D. M. (2009) Paradigm shifts in mangrove biology. In **Coastal wetlands: an integrated ecosystem approach.** Perillo, G. M. E.; Wolanski, E.; Cahoon, D. R.; Brinson, M. M. (Eds.), Elsevier.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. **J. Mol. Biol.**, v. 215, p. 403-410.

Alzubaidy, H., Essack, M., Malas, T.B., Bokhari, A., Motwalli, O., Kamanu, F. K. *et al.* (2016) Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea. *Gene* 576(2), 626-636.

Andreote, F. D.; Jiménez, D. J.; Chaves, D.; Dias, A. C. F. D.; Luvizotto, D. M.; Dini-Andreote, F.; Fasanella, C. C.; Lopez., M. V.; Baena, S.; Taketani, R. G.; Melo, I. S. (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. **PLoS ONE**, v. 7(6), p. 1-14.

Ayuso-Sacido, A., Genilloud, O. (2005) 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. **Microb. Ecol.**, v. 49(1), p. 10-24.

Baltz, R.H. (2008) Renaissance in antibacterial discovery from actinomycetes. **Current Opinion in Pharmacology**, v. 8(5), p. 557-563.

Banskota, A.H., McAlpine, J.B., Sorensen, D., Ibrahim, A., Aouidate, M., Piraee, M. *et al.* (2006) Genomic analyses lead to novel secondary metabolites. **J. Antibiot.**, v. 59(9), p. 533-542.

Beppu, T. (1992) Secondary metabolites as chemical signals for cellular differentiation. **Gene**, v. 115, p. 159-165.

Bérdy, J. (2005) Bioactive microbial metabolites a personal view. **J. Antibiot.**, v. 58(1), p. 1-26.

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munroa, M.H.G., Prinsep, M.R. (2016) Marine natural products. **Nat. Prod. Rep.**, v. 33, p. 382. doi:10.1039/c5np00156k

Bouchez, A.; Pascault, N.; Chardon, C.; Bouvy, M.; Cecchi, P.; Lambs, L. (2013) Mangrove microbial diversity and the impact of trophic contamination. **Marine Poll. Bull.**, v. 66, p. 39-46.

Buchfink, B., Xie, C., Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, p. 59-60. doi:10.1038/nmeth.3176

Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, v. 7., p. 335-336.

Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M.; Gormley, N.; Gilbert, J. A.; Smith, Geoff.; Knight, R. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. **ISME J.**, v. 6, p. 1621-1624.

Chen, J. (2012) GUniFrac: Generalized UniFrac distances. R package version 1.0. https://CRAN.R-project.org/package=GUniFrac

Chen, J.; Bittinger, K.; Charlson, E. S.; Hoffmann, C.; Lewis, J.; Wu, G. D.; Collman, R. G.; Bushman, F. D.; Li, H. (2012) Associating microbiome composition with environmental covariates using generalized UniFrac distances. **Bioinformatics**, v. 28(16), p. 2106-2113.

Chen, Q., Zhao, Q., Li, J., Jian, S., Ren, H. (2016) Mangrove succession enriches the sediment microbial community in South China. **Sci. Rep.**, v. 6, 27468. doi: 10.1038/srep27468

Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Gottschalk, G. (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. **Journal of Bacteriology**, v. 188(11), p. 4024-4036.

Choi, H., Mevers, E., Byrum, T., Valeriote, F. A., Gerwick, W. H. (2012) Lyngbyabellins K–N from two Palmyra Atoll collections of the marine cyanobacterium *Moorea bouillonii*. **European Journal of Organic Chemistry**, v. 2012(27), p. 5141-5150.

Coates, A.R.M., Halls, G., Hu, Yanmin. (2011) Novel classes of antibiotics or more of the same? **Br. J. Pharmacol.**, v. 163(1), p. 184-194. doi:10.1111/j.1476-5381.2011.01250.x

Conway, K.R., Boddy, C.N. (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. **Nucleic Acids Research**, v. 41, Database issue. doi:10.1093/nar/gks993

Corre, C., Challis, G.L. (2009) New natural product biosynthetic chemistry discovered by genome mining. **Nat. Prod. Rep.**, v. 26, p. 977-986. doi:10.1039/b713024b

Cortina, N.S., Krug, D., Plaza, A., Revermann, O., Müller, R. (2012) Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. **Angewandte Chemie International Edition**, v. 51(3), p. 811-816.

Cortina, N.S., Revermann, O., Krug, D., Müller, R. (2011) Identification and characterization of the althiomycin biosynthetic gene cluster in *Myxococcus xanthus* DK897. **Chembiochem**, v. 12(9), p. 1411-1416.

Day, A. (2012) *heatmap.plus: Heatmap with more sensible behavior.* R package version 1.3. https://CRAN.R-project.org/package=heatmap.plus

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

Dhal, N.K., Sahoo, K. (2009) Potential microbial diversity in mangrove ecosystems: a review. *Indian Journal of Marine Sciences* 38(2), 249-256.

Djukic, M., Poehlein, A., Thürmer, A., Daniel, R. (2011) Genome sequence of *Brevibacillus laterosporus* LMG 15441, a pathogen of invertebrates. **Journal of Bacteriology**, v. 193(19), p. 5535-5536.

Dos Santos, H.; Cury, J. C.; Do Carmo, F. L.; Dos Santos, A. L. ; Tiedje, J; Van Elsas, J. D.; Rosado, A. S.; Peixoto, R. S. (2011) Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. **PLOS ONE**, v. 6(3), 416943.

Duke, N. C.; Ball, M. C.; Ellison, J.C. (1998) Factors influencing biodiversity and distributional gradients in mangroves. **Global Ecol. Biogeo. Lett.**, v. 7(1), p. 27-47.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32(5), p. 1792-1797.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST, **Bioinformatics**, v. 26(19), p.2460-2461. doi: 10.1093/bioinformatics/btq461

Ehling-Schulz, M., Vukov, N., Schulz, A., Shaheen, R., Andersson, M., Märtlbauer, E., Scherer, S. (2005) Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*. **Applied and Environmental Microbiology**, v. 71(1), 105-113.

Emmert, E.A., Klimowicz, A.K., Thomas, M.G., Handelsman, J. (2004) Genetics of zwittermicin A production by *Bacillus cereus*. **Applied and Environmental Microbiology**, v. 70(1), p. 104-113.

Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., Gerwick, W. H. (2012) *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. **International Journal of Systematic and Evolutionary Microbiology**, v. 62(5), p. 1171-1178.

Feher, M., Schmidt, J.M. (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. **J. Chem. Inf. Comput. Sci.**, v. 43(1), p. 218-227.

Feller, I. C.; Lovelock, C. E.; Berger, U.; McKee, K. L. (2010) Biocomplexity in mangrove ecosystems. **Annu. Rev. Mar. Sci.**, v. 2, p. 395-417.

Fish, K.M. (2013) Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. **RSC Advances**, v. 3, p. 18228. doi:10.1039/c3ra42661k

Fusetani, N. (2000) Introduction. In: Fusetani, N. (ed.), **Drugs from the Sea**, Basel, Karger, p. 1-5.

Ghizelini, A. M.; Mendonça-Hagler, L. C. S.; Macrae, A. (2012) Microbial diversity in Brazilian mangrove sediments – a mini review. **Brazil. J. Microbiol.**, p. 1242-1254.

Gilbert, J.A., Dupont, C.L. (2011) Microbial metagenomics: beyond the genome. **Annu. Rev. Mar. Sci.**, v. 3, p. 347-371. doi: 10.1146/annurev-marine-120709-142811

Gomes, N. C. M.; Borges, L. R.; Paranhos, R.; Pinto, F. N.; Mendonça-Hagler, L. C. S.; Smalla, K. (2008) Exploring the diversity of bacterial communities in sediments of urban mangrove forests. **FEMS Microbiol. Ecol.**, v. 66, p. 96-109.

Holguin, G.; Vazquez, P.; Bashan, Y. (2001) The role of sediment microorganisms in the productivity, conservation and rehabilitation of mangrove ecosystems: an overview. **Biol. Fertil. Soils.**, v. 33, p. 265-278.

Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. **Genome Res.**, v. 21, p. 1552-1560.

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S., Fujita, N. (2013) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. **Nucleic Acids Research**, 2013, v. 41, Database issue. doi:10.1093/nar/gks1177

Jiang, X.T., Peng, X., Deng, G.H., Sheng, H.F., Wang, Y., Zhou, H.W. Tam, N.F.Y. (2013) Illumina sequencing of 16S rRNA tag revealed spatial variations of bacterial communities in mangrove wetland. **Microbial Ecol.**, v. 66(1), p.96-104.

Kelecom, A. (2002) Secondary metabolites from marine microorganisms. Annals of the Brazilian Academy of Sciences, v. 74(1), p. 151-170.

Kleindienst, S., Herbst, F.A., Stagars, M., von Netzer, F., von Bergen, M., Seifert, J. *et al.* (2014) Diverse sulfate-reducing bacteria of the *Desulfosarcina/Desulfococcus* clade are the key alkane degraders at marine seeps. **ISME J.**, v. 10, p.2029-2044.

Li, X., Tang, X., Xiao, J., Ma, G., Xu, L., Xie, S. (2013) Harnessing the potential of halogenated natural product biosynthesis by mangrove-derived actinomycetes. **Mar. Drugs**, v. 11(10), p. 3875-3890. doi:10.3390/md11103875

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L, Engels, I., Conlon, B.P. *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. **Nature**, v. 517(7535), p. 455-459. doi:10.1038/nature14098

Liu, X., Ashforth, E., Ren, B., Song, F., Dai, H., Liu, M., *et al.* (2010) Bioprospecting microbial natural product libraries from the marine environment for drug discovery. **The Journal of Antibiotics**, v. 63, p. 415-422. doi:10.1038/ja.2010.56

Lohman, J.R., Huang, S.X., Horsman, G.P., Dilfer, P.E., Huang, T., Chen, Y., Shen, B. (2013) Cloning and sequencing of the kedarcidin biosynthetic gene cluster from *Streptoalloteichus* sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics. **Molecular BioSystems**, v. 9(3), p. 478-491.

Lovelock, C. E. (2008) Soil respiration and belowground carbon allocation in mangrove forests. **Ecosystems**, v. 11, p.342–354.

Marcial-Gomes, N.C., Borges, L.R., Paranhos, R., Pinto, F.N., Mendonça-Hagler, L.C.S., Smalla, K. (2008) Exploring the diversity of bacterial communities in sediments of urban mangrove forests. **FEMS Microbiology Ecology**, v. 66(1), p. 96-109.

Marques, R.; Motta, A. C. V. (2003) Análise química do solo para fins de fertilidade. In: **Manual de diagnóstico da fertilidade e manejo dos solos agrícolas.** Lima, R. M. (Org.) Universidade Federal do Paraná, Departamento de Solos e Engenharia Agrícola, p. 81-102.

Marschner, P.; Crowley, D.; Yang, C. H. (2004) Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. **Plant and Soil**, v. 261, p. 199-208.

Matthies, C., Evers, S., Ludwig, W., Schink, B. (2000) *Anaerovorax odorimutans* gen. nov., sp. Nov., a putrescine-fermenting, strictly anaerobic bacterium. **Int. J. Syst. Evol. Microbiol**,, v. 50(4), p. 1591-1594.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fishbach, M.A., *et al.* (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic Acids Research**, v. 39, Web Server issue W339-W346. doi:10.1093/nar/gkr466

Mendes, L. W., Tsai, S.M. (2014) Variations of bacterial community structure and composition in mangrove sediment at different depths in Southeastern Brazil. **Diversity**, v. 6(4), p. 827-843.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A. et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v. 9, p. 386.

Moffitt, M.C., Neilan, B.A. (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. **J. Mol. Evol.**, v. 56, p. 446-457.

Montaser, R., Luesch, H. (2011) Marine natural products: a new wave of drugs? **Future Medicinal Chemistry**, v. 3(12), p. 1475-1489.

Moore, B.S. (2005) Biosynthesis of marine natural products: microorganisms (part A) **Nat. Prod. Rep.**, v. 22, p. 580-593. doi:10.1039/b404737k

Nagarajan, M., Rajesh Kumar, R., Meenakshi Sundaram, K., Sundararaman. (2015) Marine biotechnology: potentials of marine microbes and algae with reference to

pharmacological and commercial values. In **Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology**, Bahadur, B., Venkat, R. Manchikatla, Sahijram, L., Krishnamurthy, V. K. (eds), Springer India, New Delhi, p. 685-723. doi:10.1007/978-81-322-2283-5_35

Nair, H. P., Vincent, H., Bhat, S.G. (2013) Culture independent analysis of the soil microbiome to assess microbial diversity of mangrove soil. **Bio-Genetics J.**, v. 1(1), p. 1-4.

Nakashima, T., Iwatsuki, M., Ochiai, J., Kamiya, Y., Nagai, K., Matsumoto, A., *et al*. (2014) Mangromicins A and B: structure and antitrypanosomal activity of two new cyclopentadecane compounds from Lechevalieria aerocolonigenes K10-0216. **J Antibiot (Tokyo)**, v. 67(3), p. 253-260. doi: 10.1038/ja.2013.129

Nogueira, V. L. R.; Rocha, L. L.; Colares, G. B.; Angelima, A. L.; Normando, L. R. O.; Cantão, M. E.; Agnez-Lima, L. F.; Andreote, F. D.; Melo, V. M. M. (2015) Microbiomes and potential metabolic pathways of pristine and anthropized Brazilian mangroves. **Reg. Stud. Marine Sci.**, v. 2, p. 56-64.

Nunnery, J. K., Mevers, E., Gerwick, W.H. (2010) Biologically active secondary metabolites from marine cyanobacteria. **Current Opinion in Biotechnology**, v. 21(6), p. 787-793.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*. (2016) vegan: Community Ecology Package. R package version 2.3-5. https://CRAN.R-project.org/package=vegan

Payne, D.J., Gwynn, M.N., Holmes, D.J., Pompliano, D.L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. **Nature Reviews**, v. 6, p. 29-40. 2007. doi:10.1038/nrd2201

Peixoto, R.; Chaer, G. M.; Carmo, F. L.; Araújo, F. V.; Paes, J. E.; Volpon, A.; Santiago, G. A.; Rosado, A. S. (2011) Bacterial communities reflect the spatial variation in pollutant levels in Brazilian mangrove sediments. **Antonie van Leeuwenhoek**, v. 99, p. 341-354.

Pylro, V.S., Roesch, L.F.W., Morais, D.K., Clark, I.A., Hirsch, P.R., Tótola, M.R. (2014) Data analysis for 16S microbial profiling from different benchtop sequencing platforms. **J. Microbiol. Methods**, v. 107, p.30-37. doi:10.1016/j.mimet.2014.08.018

Pylro, V.S., Roesch, L.F.W., Ortega, J.M., Amaral, A.M., Tótola, M.R., Hirsch, P.R. *et al.* (2014) Brazilian Microbiome Project: revealing the unexplored microbial diversity – challenges and prospects. **Microbial Ecol.**, v. 67(2), p. 237-241.

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reef, R.; Feller, I. C.; Lovelock, C. E. (2010) Nutrition of mangroves. **Tree Physiol.**, v. 30, p. 1148-1160.

Reen, F. J., Gutiérrez-Barranquero, J. A., Dobson, A. D. W., Adams, C., O'Gara, F. (2015) Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. **Marine Drugs**, v. 13(5), p. 2924-2954. doi:10.3390/md13052924

Rocha, L.L., Colares, G.B., Nogueira, V.L.R., Paes, F.A., Melo, V.M.M. (2016) Distinct habitats select particular bacterial communities in mangrove sediments. **Int. J. Microbiol.**, v. 2016, article ID 3435809, 6 pages. doi:10.1155/2016/3435809

Rodriguez, E., McDaniel, R. (2001) Combinatorial biosynthesis of antimicrobials and other natural products. **Current Opinion in Microbiology**, v. 4(5), p. 526-534.

Sahoo, K.; Dhal, N. K. (2009) Potential microbial diversity in mangrove ecosystems: a review. **Ind. J. Marine Sci.**, v. 38(2), p. 249-256.

Saseeswari, A., Kanimozhi, G., Panneerselvam, A. (2016) Bacterial diversity of mangrove soil in Karankadu from East Coast of Tamil Nadu, India. **Int. J. Curr. Microbiol. App Sci.**, v. 5(4), p.750-756.

Schimd, I.I., Sattler, I.I., Grabley, S., Thiericke, R. (1999) Natural products in High Throughput Screening: automated high-quality sample preparation. **J. Biomol. Screen.**, v. 4(1) p. 15-25.

Schirmer, A., Gadkari, R., Reeved, C.D., Ibrahim, F., DeLong, EF., Hutchinson, C.R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissolute*. **Appl. Environ. Microbiol.**, v. 71(8), p. 4840-4849. doi:10.1128/AEM.71.8.4840-4849.2005

Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. **Bioinformatics**, v. 27(4), p.592-593

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B, *et al.* (2009). Introducing mothur: Open-source, platform-independent, community-supported

software for describing and comparing microbial communities. **Appl. Environ. Microbiol.**, v. 75(23), p. 7537-7541.

Shade, A.; Handelsman, J. (2012) Beyond the Venn diagram: the hunt for a core microbiome. **Environ. Microbiol.**, v. 14(1), p. 4-12.

Spalding, M. D.; Blasco, E., Field, C.D. (Eds). (1997) **World mangrove atlas.** The International Society for Mangrove Ecosystems, Okinawa, Japan. 178 pp.

Spalding, M. D.; Kainuma, M.; Collins, L. (2010) **World atlas of mangroves.** Earthscan, London, UK. 338 pp.

Stone, M.J., Williams, D.H. (1992) On the evolution of functional secondary metabolites (natural products). **Molecular Microbiology**, v. 6(1), p. 29-34.

Swenson, C.E., Sadikot, R.T. (2015) Achromobacter respiratory infections. **Annals of the American Thoracic Society**, v. 12(2), p. 252-258.

Taketani, R.G., Yoshiura, C.A., Dias, A.C., Andreote, F.D., Tsai, S.M. (2010) Diversity and identification of methanogenic archaea and sulphate-reducing bacteria in sediments from a pristine tropical mangrove. **Antonie Van Leeuwenhoek**, v. 97(4), p. 401-411.

Tao, M., Wang, L., Wendt-Pienkowski, E., George, N.P., Galm, U., Zhang, G., Shen, B. (2007) The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. **Molecular BioSystems**, v. 3(1), p. 60-74.

Thatoi, H., Behera, B.C., Mishra, R.R., Dutta, S.K. (2013) Biodiversity and biotechnological potential of microorganisms from mangrove ecosystems: a review. **Ann. Microbiol.**, v. 63, p. 1-19. doi:10.1007/s13213-012-0442-7

Thornburg, C. C., Cowley, E. S., Sikorska, J., Shaala, L. A., Ishmael, J. E., Youssef, D. T., McPhail, K. L. (2013) Apratoxin H and apratoxin A sulfoxide from the Red Sea cyanobacterium Moorea producens. **Journal of Natural Products**, v. 76(9), p. 1781-1788.

Trindade, M., van Zyl, L. J., Navarro-Fernández, J., Abd Elrazak, A. (2015) 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, p. 890. doi:10.3389/fmicb.2015.00890

Varaljay, V.A., Howard, E.C., Sun, S., Moran, M.A. (2010) Deep sequencing of a dimethylsulfoniopropionate-degrading gene (dmdA) by using PCR primer pairs designed on the basis of marine metagenomic data. **Appl. Environ. Microbiol.**, v. 76, p. 609-617.

Vartoukian, S.R., Palmer, R.M., Wade, W.G. (2010) Strategies for culture of 'unculturable' bacteria. **FEMS Microbiol. Lett.**, v. 309(1), p. 1-7. doi:10.1111/j.1574-6968.2010.02000.x

Vining, L.C. (1990) Functions of secondary metabolites. **Annu. Rev. Microbial.**, v. 44, p. 395-427.

Weissman, K. J. (2014) The structural biology of biosynthetic megaenzymes. **Nature Chemical Biology**, v. 11, p. 660-670. doi:10.1038/nchembio.1883

Wenzel, S.C., Müller, R. (2005) Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. **Current Opinion in Chemical Biology**, v. 9, p. 447-458.

Williams, D.H., Stone, M.J., Hauck, P.R., Rahman, S.K. (1989) Why are secondary metabolites (natural products) biosynthesized? **Journal of Natural Products**, v. 52(6), p. 1189-1208.

Winter, J.M., Behnken, S., Hertweck, C. (2011) Genomics-inspired discovery of natural products. **Current Opinion in Chemical Biology**, v. 15(1), p. 22-31. doi:10.1016/j.cbpa.2010.10.020

Woodhouse, J.N., Fan, L. Brown, M.V., Thomas, T., Neilan, B.A. (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. **ISME J.**, v. 7, p. 1842-1851.

Xie, Y., Wright, S., Shen, Y., Du, L. (2012) Bioactive natural products from *Lysobacter.* **Nat. Prod. Rep.**, v. 29, p. 1277-1287. doi:0.1039/c2np20064c

Xu, Z., Hansen, M.A., Hansen, L.H., Jacquiod, S., Sørensen, S.J. (2014) Bioinformatic approaches reveal metagenomic characterization of soil microbial community. **PLoS ONE**, v. 9(4), e93445. doi:10.1371/journal.pone.0093445

Yan, B., Hong, K., Yu, Z.N. (2006) Archaeal communities in mangrove soil characterized by 16S rRNA gene clones. **J. Microbiol.**, v. 44(5), p. 566-571.

Zipperer, A., Konnerth, M.C., Laux, C., Berscheid, A., Janek, D., Weidenmaier, C., *et al.* (2016) Human commensals producing a novel antibiotic impair pathogen colonization. **Nature**, v. 535 (7613), p. 511-516. doi:10.1038/nature18634