UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia

SANDERSON TARCISO PEREIRA DE SOUSA

INVESTIGAÇÃO DE GENES ENVOLVIDOS NA BIODEGRADAÇÃO DE HIDROCARBONETOS AROMÁTICOS A PARTIR DO METAGENOMA DE MANGUEZAL IMPACTADO COM PETRÓLEO

INVESTIGATION OF GENES INVOLVED IN THE BIODEGRADATION OF AROMATIC HYDROCARBONS FROM THE OIL-IMPACTED MANGROVE METAGENOME

Campinas

2018

SANDERSON TARCISO PEREIRA DE SOUSA

INVESTIGAÇÃO DE GENES ENVOLVIDOS NA BIODEGRADAÇÃO DE HIDROCARBONETOS AROMÁTICOS A PARTIR DO METAGENOMA DE MANGUEZAL IMPACTADO COM PETRÓLEO

INVESTIGATION OF GENES INVOLVED IN THE BIODEGRADATION OF AROMATIC HYDROCARBONS FROM THE OIL-IMPACTED MANGROVE METAGENOME

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Genética e Biologia Molecular na Área de Genética de Micro-organismos.

Thesis presented to the Institute of Biology of the University of Campinas as partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology in the area of Genetics of Microorganisms.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO SANDERSON TARCISO PEREIRA DE SOUSA E ORIENTADO PELA DRA. VALÉRIA MAIA MERZEL.

Orientador (a): DRA. VALÉRIA MAIA MERZEL

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2013/22555-4
ORCID: https://orcid.org/0000-0002-3928-240X

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

So85i

Sousa, Sanderson Tarciso Pereira de, 1986-
Investigação de genes envolvidos na biodegradação de hidrocarbonetos aromáticos a partir do metagenoma de manguezal impactado com petróleo / Sanderson Tarciso Pereira de Sousa. – Campinas, SP : [s.n.], 2018.

Orientador: Valéria Maia Merzel.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Manguezais. 2. Hidrocarbonetos policiclicos aromáticos. 3. Metagenômica. 4. Biodegradação. 5. Biogeografia. I. Oliveira, Valeria Maia de, 1966-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Investigation of genes involved in the biodegradation of aromatic hydrocarbons from the oil-impacted mangrove metagenome
Palavras-chave em inglês:
Mangrove swamps
Polycyclic aromatic hydrocarbons
Metagenomics
Biodegradation
Biogeography
Área de concentração: Genética de Microorganismos
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Valéria Maia Merzel [Orientador]
Cynthia Canedo da Silva
Tiago Palladino Delforno
Fabiana Fantinatti Garboggini
Geizecler Tomazetto
Data de defesa: 20-04-2018
Programa de Pós-Graduação: Genética e Biologia Molecular

Campinas, 20/04/2018.

COMISSÃO EXAMINADORA

Prof.(a) Dr.(a). Valéria Maia Merzel (Presidente)

Prof.(a). Dr.(a) Cynthia Canedo da Silva

Prof.(a) Dr(a). Tiago Palladino Delforno

Prof.(a) Dr(a). Fabiana Fantinatti Garboggini

Prof.(a) Dr(a). Geizecler Tomazetto

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

DEDICATÓRIA...

Dedico essa tese a minha querida avó Antonina do Carmo Ribeiro, que hoje, infelizmente, já não está entre nós, mas tenho certeza que continua olhando e orando por mim de onde ela estiver.

Dedico também a minha noiva Jussimara Teles e a todos os meus familiares que me apoiaram de todas as formas para a realização desse trabalho, pois, se não fosse por eles, acho que não conseguiria estar aqui.

AGRADECIMENTOS...

À Universidade Estadual de Campinas (UNICAMP), em especial ao CPQBA, pela oportunidade e infraestrutura disponibilizada para execução deste trabalho;

À orientadora Profa. Dra. Valéria Maia Merzel, pelo incentivo, confiança, aprendizado e amizade em todos os momentos do Doutorado;

Aos membros da Banca, Profa. Dra. Cynthia Canêdo Da Silva, Dr. Tiago Palladino Delforno, Profa. Dra. Fabianna Fantinatti Garboggini e Profa. Dra. Marta Cristina Teixeira Duarte, pela disponibilidade e valiosas correções e sugestões realizadas nesta Tese;

Aos professores, técnicos e demais funcionários de todas as divisões do CPQBA e CBMEG/UNICAMP pelos ensinamentos e amizade;

A todos os amigos e colegas do Doutorado pela amizade, conversas e colaboração nos trabalhos;

À secretaria da pós-graduação em genética e biologia molecular pelas instruções, paciência e colaboração;

À minha noiva JUSSIMARA TELES PEREIRA, pelo amor e paciência;

À toda minha família, em especial a ANTONINA DO CARMO†, ABEL MARTINS, DALVA DE SOUSA, JUSSIARA SILVA, LUÍZ LÁZARO, MARCO AURÉLIO E GIOVANI RIBEIRO, por tudo;

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão da bolsa de estudo e apoio financeiro ao projeto “Investigação de genes envolvidos na biodegradação de hidrocarbonetos aromáticos a partir do metagenoma de mangue impactado com petróleo”. Processo nº 2013/22555-4.

Resumo

O manguezal é um dos biomas mais importantes para a manutenção da vida nos mares e o equilíbrio da biosfera. O sedimento do mangue é uma grande fonte de novos recursos genéticos para pesquisa. Estes ambientes têm sofrido impacto antropogênico por muito tempo, principalmente a poluição causada por derramamento de petróleo. Diversos estudos estão sendo conduzidos para explorar o potencial genético destas comunidades com a finalidade de desenvolver novos processos biotecnológicos para o tratamento de ambientes contaminados. Dentro deste contexto, este trabalho visou identificar e caracterizar genes que codificam dioxigenases, proteínas envolvidas na degradação de hidrocarbonetos aromáticos, a partir de clones fosmidiais de uma biblioteca metagenômica construída de sedimento de mangue impactado por petróleo do município de Bertioga-SP. A coleta e processamento das amostras ambientais, juntamente com a construção da biblioteca metagenômica fosmidial, foi realizada previamente no âmbito do projeto FAPESP "Prospecção de metagenoma microbiano de sedimentos de manguezais na busca por novos compostos bioativos" (Processo nº. 2011/50809-5). O DNA metagenômico das amostras de sedimento, coletadas em triplicata, foi extraído e usado em reações de PCR para a construção de bibliotecas dos genes de dioxigenases que hidroxilam anéis aromáticos (ARHDs) e análise de seus padrões de distribuição biogeográfico. O DNA total da biblioteca fosmidial foi isolado e sequenciado em plataforma Illumina para prospecção *in silico* de genes envolvidos no metabolismo de compostos aromáticos. Ainda, os clones fosmidiais foram submetidos a uma triagem preliminar por colorimetria, baseada no uso do composto MTT [3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2 H-tetrazolium bromide] (Merck) e, posteriormente, ensaios de degradação de hidrocarbonetos por cromatografia gasosa acoplada à espectrometria de massas (CG-EM) para confirmar e quantificar a atividade de degradação de hidrocarbonetos aromáticos (fenol, naftaleno, fenantreno, pireno e benzopireno) dos clones. O DNA fosmidial dos clones com melhor potencial de degradação foi sequenciado e processado para montagem de grandes fragmentos metagenômicos. Caracterização estrutural e funcional dos insertos metagenômicos foram feitos nas plataformas RAST e IMG, para elucidação de clusters gênicos envolvidos na degradação de hidrocarbonetos aromáticos. Alvos gênicos foram isolados, ligados em vetor de expressão pET28a(+) e transformados em cepas de E. coli BL21(DE3) e Rosetta. A análise da diversidade dos genes α-ARHD revelou que *Pseudomonas*, *Streptomyces*, *Variovorax*, *Bordetella* e *Rhodococcus* foram os cinco gêneros mais abundantes entre os 15 sítios analisados. A análise de distribuição biogeográfica revelou forte endemismo de algumas *Operational Protein Families* (OPF) de α-ARHD em cada região geográfica e uma considerável proximidade evolutiva entre OPFs encontradas em sítios da Antártica e da América do Sul. O sequenciamento da biblioteca fosmidial permitiu identificar um total de 71.729 sequências metagenômicas parciais na categoria SEED para metabolismo de compostos aromáticos (1% do total). Estas foram classificadas nas subcategorias *Anaerobic degradation of aromatic compounds* (20,34%), *Metabolism of central aromatic intermediates* (35,40%), e *Peripheral pathways for catabolism of aromatic compounds* (22,56%). Além disso, foi evidenciado o predomínio das dioxigenases de Catecol, Protocatecuato e Homogentisato, envolvidas no metabolismo central, e das dioxigenases de Fenilpropionato, Lignostilbeno, Bifenil e Benzoato envolvidas nas vias periféricas de degradação de aromáticos. As análises de bioinformática dos dados de sequenciamento dos clones fosmidiais potencialmente degradadores revelaram uma interessante diversidade de genes envolvidos direta e indiretamente no metabolismo de compostos aromáticos. Os ensaios de cromatografia mostraram a eficiência de 3 clones na degradação de fenol (98%), pireno (75%) e naftaleno (71%). Os clones contendo os genes alvos para dioxigenases apresentaram sinais positivos nos testes de expressão. Os resultados deste trabalho permitiram elucidar a riqueza, distribuição e relação genética entre os genes ARHD de diferentes sítios ao redor do mundo. Além disso, a avaliação da biblioteca metagenômica pelas diferentes técnicas empregadas, revelou seu extenso potencial genético para o metabolismo de compostos aromáticos.

Abstract

Mangrove is one of the most important biomes for the maintenance of life in the seas and the balance of the biosphere. Mangrove sediment is a major source of new genetic resources for research. These environments have long been suffering anthropogenic impacts, especially pollution caused by oil spills. Many microbial species are able to degrade petroleum compounds, and several studies are being conducted to explore the genetic potential of these communities with the purpose of developing new biotechnological processes for the treatment of contaminated environments. In this context, this work aimed to identify and characterize genes encoding dioxygenases, proteins involved in the degradation of aromatic hydrocarbons, from fosmid clones of a metagenomic library constructed from oil-impacted mangrove sediments located in the city of Bertioga-SP. Sampling and further processing of environmental samples, together with the metagenomic fosmid library construction, were performed previously under the project "Bioprospecting microbial metagenome from mangrove sediments in the search for novel bioactive compounds" (FAPESP Process nº. 2011/50809-5). Metagenomic DNA from sediment samples, collected in triplicate, was extracted and used in PCR reactions for the assembly of Aromatic Ring-Hydroxylating Dioxygenases (ARHDs) gene library and analysis of their biogeographic distribution patterns. Total fosmid DNA was isolated and sequenced on the Illumina platform for in silico prospecting of genes involved in the metabolism of aromatic compounds. Further, fosmid clones were subjected to a preliminary colorimetric screening, based on the use of the compound 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2 H-tetrazolium bromide (MTT, Merck) and, subsequently, to hydrocarbon degradation assays by gas chromatography coupled to mass spectrometry (GC-MS) to confirm and quantify the aromatic hydrocarbon degradation activity (phenol, naphthalene, phenanthrene, pyrene and benzopyrene) of the clones. Fosmid DNA from clones with the best degradation potential was sequenced and processed for the assembly of large metagenomic fragments. Further structural and functional characterization of the metagenomic inserts was performed in the RAST and IMG platforms, to elucidate gene clusters involved in the aromatic hydrocarbon degradation. Gene targets were isolated, ligated into pET28a (+) expression vector and transformed into E. coli strains BL21 (DE3) and Rosetta. Diversity analysis of the α-ARHD genes revealed that *Pseudomonas*, *Streptomyces*, *Variovorax*, *Bordetella* and *Rhodococcus* were the five most abundant genera harboring dioxygenases among the 15 sites analyzed. The biogeographic distribution analysis revealed strong endemism of some Operational Protein Families (OPF) of α-ARHD in each geographic region and a significant evolutionary relatedness between OPFs found in sites of Antarctica and South America. Sequencing of the fosmid library allowed the identification of a total of 71,729 partial metagenomic sequences in the SEED category for the metabolism of aromatic compounds (1% of the total). These were classified in the categories of *Anaerobic degradation of aromatic compounds* (20.34%), *Metabolism of central aromatic intermediates* (35.40%), and *Peripheral pathways for catabolism of aromatic compounds* (22.56%). In addition, we have demonstrated the predominance of Catechol, Protocatechuate and Homogentisate dioxygenases involved in central metabolism, and the dioxygenases of Phenylpropionate, Lignostilbene, Biphenyl and Benzoate involved in the peripheral pathways of aromatics degradation. Bioinformatics analyzes of sequencing data from potentially degrading fosmid clones revealed an interesting diversity of genes involved directly or indirectly in the aromatic compounds metabolism. Chromatography assays showed the efficiency of three clones in the degradation of phenol (98%), pyrene (75%) and naphthalene (71%). Clones containing the target genes for dioxygenase showed positive signals in expression tests. The results of this work allowed to elucidate the richness, distribution and genetic relationships among ARHD genes from different sites around the world. In addition, the evaluation of the metagenomic library by the different techniques employed revealed its extensive genetic potential for the aromatic compounds metabolism.

ÍNDICE

1. INTRODUÇÃO

Atualmente há um grande interesse na investigação de recursos genéticos e metabólicos contidos na microbiota presente em ambientes inexplorados ou pouco explorados ao redor do mundo. Além de gerar novos conhecimentos científicos e acadêmicos, os recursos microbianos permitem avanços tecnológicos na medicina, agricultura e no setor industrial, propiciando um consequente crescimento econômico e benefícios para a sociedade.

O aumento na emissão de diversos poluentes tóxicos e recalcitrantes (que persistem por longos períodos na natureza), decorrente das atividades de diversos setores da sociedade moderna, como o avanço da indústria, turismo, transporte, dentre outras, configura em uma das maiores preocupações atuais. O desenvolvimento da indústria petrolífera, mais especificamente, trouxe muitas consequências para a natureza, como o impacto causado pelos vazamentos de petróleo e seus derivados, ao longo de toda a cadeia de extração, refino, transporte, consumo e descarte de resíduos.

Os hidrocarbonetos aromáticos (HAs), incluindo hidrocarbonetos aromáticos policíclios (HAPs), derivados de processos naturais e antropogênicos (como manufatura de produtos derivados de petróleo) são perigosos contaminantes ambientais devido ao seu caráter recalcitrante, cumulativo, mutagênico e carcinogênico. Os HAs podem comprometer a integridade de ambientes aquáticos, do ar, do solo, sedimentos e dos seres vivos em geral. A microbiota ambiental é sensivelmente alterada na presença de HAs, afetando o funcionamento normal dos ciclos biogeoquímicos. Processos físico-químicos como volatização, foto-oxidação, oxidação química, bioacumulação e adsorção a partículas sólidas podem contribuir para a remoção de HAPs do meio ambiente. Entretanto, na tentativa de desenvolver novas tecnologias para viabilizar a remoção de HAPs do meio ambiente, a biologia molecular tem sido empregada para entender a fina relação entre o genoma e o metabolismo de microrganismos capazes de degradar HAPs através da biorremediação.

As ARHDs (*Aromatic Ring-Hydroxylating Dioxygenases* - Dioxigenases que Hidroxilam Anéis Aromáticos) e ARCDs (*Aromatic ring-cleaving dioxygenases* – Dioxigenases que clivam anéis aromáticos) são as principais enzimas envolvidas no metabolismo de diferentes compostos aromáticos, tais como benzeno, tolueno, xileno, bifenilo, fenol, naftaleno, fenantreno, pireno, benzopireno, dentre outros. Regiões conservadas dos genes que codificam para dioxigenases são usadas como marcadores genéticos de poluição ambiental por hidrocarbonetos aromáticos. Além disso, a diversidade genética para dioxigenases tem sido

muito explorada, considerando seu amplo potencial biotecnológico para a remediação de ambientes impactados por petróleo.

O bioma manguezal representa um dos ambientes naturais mais afetados pela atividade antropogênica, resultando no impacto por poluentes industriais e domésticos. Dados literários já haviam revelado que áreas de manguezais apresentam uma vasta riqueza animal, vegetal e microbiana. Além do mais, estudos empregando isolamento de microrganismos e técnicas independentes de cultivo (como a metagenômica) tem revelado que o microbioma detém uma grande diversidade funcional e taxonômica. Visto que os manguezais são ambientes de importância econômica, ecológica e cultural, muitos projetos estão sendo desenvolvidos para sua preservação. O impacto da ação antrópica sobre esses ambientes vem reduzindo suas áreas de abrangência e acabando com sua biodiversidade, antes mesmo de serem estudadas. Um dos principais problemas enfrentados por esse bioma vem da contaminação por derramamento de petróleo. Diante deste problema, há uma grande mobilização científica em descobrir novos genes e vias metabólicas para a definição de novas estratégias, a partir da engenharia genética e biologia sintética, para remediação de ambientes contaminados.

A biorremediação faz uso de organismos que têm a capacidade de acelerar o processo natural de degradação de poluentes, reduzindo o tempo de danos à diversidade biológica e os recursos financeiros para a descontaminação. Outra vantagem em se estudar a microbiota presente em sedimento de manguezal é a possibilidade de encontrar genes que codificam para enzimas adaptadas às condições flutuantes de oxigenação, pH, temperatura, nutrientes e luminosidade características desse ambiente.

Neste trabalho foi utilizada uma biblioteca metagenômica construída a partir da microbiota de sedimento de manguezal impactado com petróleo, localizado no município de Bertioga-SP, no âmbito do projeto “Prospecção de metagenoma microbiano de sedimentos de manguezais na busca por novos compostos bioativos” (Processo FAPESP 2011/50809-5). Nosso objetivo geral foi a identificação e caracterização de genes que codificam enzimas (principalmente dioxigenases) envolvidas no metabolismo de hidrocarbonetos aromáticos, um dos principais contaminantes do petróleo. Após a identificação e anotação funcional, os genes de interesse foram quantificados em subgrupos funcionais e submetidos à classificação taxonômica para elucidação dos prováveis clados microbianos envolvidos na degradação de hidrocarbonetos aromáticos. Para a realização do trabalho, foram empregadas as abordagens de sequenciamento em larga escala, prospecção *in silico* e *screening* funcional dos clones fosmidiais de uma biblioteca metagenômica previamente construída.

De acordo com a literatura, o impacto causado pelo derramamento de petróleo no ambiente leva à alteração da estrutura da comunidade microbiana, favorecendo a ocorrência de grupos taxonômicos mais tolerantes e especializados na degradação de hidrocarbonetos aromáticos (Head *et al*. 2006). Sendo assim, acreditamos no propósito vantajoso de investigar os recursos genéticos provenientes de ambientes já contaminados, pois, teoricamente, o enriquecimento com hidrocarbonetos aromáticos propicia a seleção de populações microbianas envolvidas no seu metabolismo. Além disso, foi realizada uma comparação entre as sequências nucleotídicas de uma região conservada dos genes para dioxigenases (α-ARHD) obtidas de vários sítios ao redor do mundo, com base na hipótese da existência de uma correlação genética entre eles, movida por pressões evolutivas. Este trabalho ampliou o conhecimento científico sobre a diversidade de genes (principalmente os genes para dioxigenases) envolvidos no metabolismo de hidrocarbonetos aromáticos, mostrou quais são os principais grupos microbianos envolvidos com os processos de degradação destes compostos, permitiu a identificação e isolamento de prováveis novos genes para dioxigenases (não depositados no *GenBank*), e abriu a perspectiva de usá-los como ferramentas para a biologia sintética aplicada a processos de biorremediação no futuro.

2. REVISÃO DE LITERATURA

2.1 Aspectos gerais sobre o manguezal

O manguezal é um ecossistema de fundamental importância para a manutenção da vida nos mares e o equilíbrio da biosfera. Este bioma é composto por plantas com característica lenhosa que apareceram, primeiramente, no entorno do mar de Tethys, divergindo de parentes terrestres durante o Cretáceo Superior e Terciário (Ricklefs, Schwarzbach e Renner 2006). Atualmente, estima-se que haja 70 espécies de plantas geneticamente distintas compondo a flora do mangue (Spalding, Kainuma e Collins 2010). Destas 70 espécies, sete ocorrem no Brasil e estão distribuídas em 4 gêneros: *Laguncularia*, *Rhizophora*, *Conocarpus* e *Avicennia* (Spalding, Kainuma e Collins 2010). Dentre as espécies de mangue encontradas na Baixada Santista, Bertioga-SP, as mais comuns são: *Avicennia schauriana*, *Laguncularia racemosa* e *Rhizophora mangle* (Lacerda, 2003; Schuler *et al,* 2000). Estudos recentes constataram que o manguezal possui uma biomassa média global de 247 toneladas por hectare, valor similar ao apresentado por florestas tropicais úmidas (Camacho *et al.* 2011; Donato *et al.* 2012). A geomorfologia, salinidade e regime das marés do bioma mangue orientou, no curso da evolução, a formação de características estruturais e funcionais específicas em sua vegetação: raízes aéreas, embriões vivíparos e mecanismos eficientes de retenção de nutrientes (Alongi 2014). A energia gerada pelo movimento das ondas e a dinâmica das marés permite o transporte e o estoque de carbono, sedimentos, alimento, nutrientes e resíduos, além de promover a troca de calor com a atmosfera e águas costeiras adjacentes (Vo-Luong and Massel 2008; Barr *et al.* 2013). Curiosamente, a maior parte do carbono presente no manguezal é estocado no sedimento, em grandes áreas subterrâneas de raízes mortas, o que auxilia na retenção e reciclagem de nutrientes para a vegetação (Alongi *et al.* 2003, 2004). Dentre outros motivos, os manguezais precisam ser preservados porque desempenham um grande papel na regulação climática, absorvendo e armazenando uma grande quantidade de carbono, amenizando o impacto causado pela emissão de CO_2 pelo homem (Mcleod *et al.* 2011; Siikamäki, Sanchirico e Jardine 2012).

A grande biomassa do manguezal pode ser encontrada ao longo da costa continental, em regiões estuarinas, deltas de rios, lagoas e ilhas (Lugo e Snedaker 1974). Os manguezais estão localizados, normalmente, entre o mar e a terra, nos trópicos e subtrópicos do planeta entre as latitudes 30°N e 30°S. Suas condições ambientais podem incluir alta salinidade, altas temperaturas, grandes variações no nível das marés, muita sedimentação e flutuações na taxa de oxigênio no solo. É um dos biomas mais produtivos e relevantes para a biologia, prestando um grande serviço tanto para a sociedade quanto para os ecossistemas costeiros e marinhos

(Giri *et al.* 2011). Abriga uma enorme diversidade de seres vivos como peixes, crustáceos, moluscos, aves, répteis e mamíferos, além dos micro-organismos. Esse bioma ajuda a proteger as zonas costeiras, reduzindo impactos causados por *tsunamis* e furacões, fornece proteção para espécies marinhas e pelágicas se reproduzirem e evitarem seus predadores, é uma fonte de alimentos, de matéria-prima para remédios artesanais, e de combustível e materiais de construção para comunidades humanas locais (Giri *et al.* 2011).

Os habitats costeiros sofrem muita pressão pela instalação de grandes populações, pelo impacto causado por grandes tempestades, pelo aumento relativo do nível do mar e pelo desenvolvimento econômico (Gilman *et al.* 2008; Giri *et al.* 2011). Outras atividades, como a agricultura, a aquacultura, o turismo e a expansão urbana também contribuem para a degradação dos manguezais (Alongi 2002). De forma agravante, os manguezais sofrem com frequentes impactos decorrentes de vazamentos de petróleo, emissão de dejetos, metais pesados, esgotos, depósito de lixo urbano-industrial e desmatamento para a construção de indústrias e domicílios (Oliveira & Netto, 1989). Devido à sua posição de contato direto com as águas oceânicas (fonte poluidora ativa), somado à alta taxa de matéria orgânica, insuficiência de oxigênio e regime das marés, o manguezal apresenta grande susceptibilidade para o acúmulo de compostos do petróleo em seu sedimento (Kathiresan e Bingham 2001).

Estudos mostraram que de 1980 a 2000, 35% das florestas de manguezais já haviam desaparecido do mundo, uma taxa de declínio mais rápida do que a das florestas tropicais e dos recifes de coral (Duke *et al.* 2007). Baseadas nestas taxas de declínio, projeções sugerem que em 100 anos esse bioma desaparecerá totalmente, seguido da diminuição na biodiversidade, no sequestro de carbono e na sua função como barreira natural ao longo da linha costeira (Duke *et al.* 2007).

O manguezal tem uma distribuição mundial de aproximadamente 137.760 km^2, e o Brasil detém a terceira maior porção dessa área (7%), atrás apenas da Indonésia (22,6%) e Austrália (7,1%) (Giri *et al.* 2011) (Figura 1). A área de 20.000 km^2 de manguezal no território brasileiro está distribuída entre Cabo Orange, no Amapá, até o município de Laguna em Santa Catarina. A região da Baixada Santista, no Estado de São Paulo, é composta pelos municípios de Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos e São Vicente, e detém 125 km^2 de toda a área de manguezal do Estado (Holguin et al. 2001; Cury, 2002). A região da Baixada Santista é conhecida pela sua grande importância econômica, abrigando um dos maiores portos do mundo, indústrias metalúrgicas e de mineração, centros produtores de fertilizantes e o maior polo petroquímico do Brasil (Bonetti 2000; Lima 2012). Em decorrência da grande atividade econômica, fatores como a aumento da densidade

populacional, urbanização desordenada e poluição ambiental tem descaracterizado áreas de manguezal da Baixada Santista (Bonetti 2000). Em 14 de outubro de 1983 ocorreu o rompimento de um oleoduto da PETROBRAS (Petróleo Brasileiro S.A.), ligando o Terminal Marítimo Almirante Barroso (TEBAR) à Refinaria Presidente Bernardes em Cubatão. Este acidente culminou no derramamento de 3 milhões de litros de petróleo que se alastrou pelo Rio Iriri até o canal de Bertioga, acarretando em um impacto ambiental sem precedentes na área de manguezal, com 60 km de extensão (Coimbra 2006).

Neste contexto, o presente estudo foi motivado pelo interesse em determinar o perfil funcional e taxonômico da comunidade microbiana do sedimento de manguezal localizado em Bertioga-SP, após, aproximadamente, 30 anos do impacto causado pelo vazamento de petróleo. Nosso plano de trabalho visou identificar e quantificar genes envolvidos nas vias de degradação de hidrocarbonetos aromáticos, elucidar o potencial genético natural para a atenuação da poluição ambiental e explorar o arsenal gênico da microbiota para o desenvolvimento de futuros processos biotecnológicos.

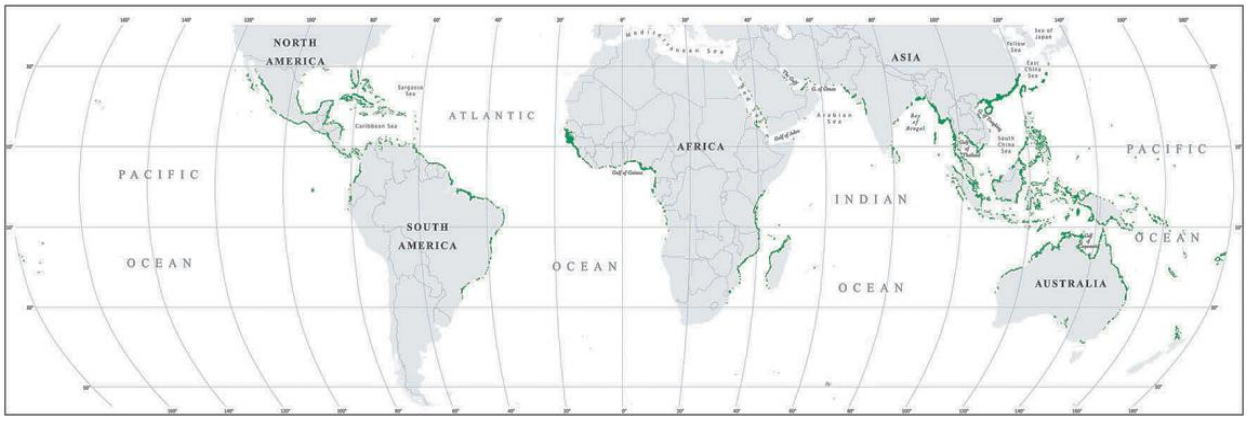


Figura 1. Distribuição dos manguezais ao redor do mundo (marcado em verde). Fonte: (Giri *et al.* 2011).

2.2 Microbiota do manguezal

Em função das altas temperaturas, grandes variações de salinidade, ventos fortes, baixas concentrações de oxigênio e influência das marés, os organismos que habitam o manguezal incluindo as comunidades microbianas, desenvolveram adaptações para a sua sobrevivência (Kathiresan e Bingham 2001). A diversidade microbiana desempenha um papel muito importante para a manutenção dos ciclos biogeoquímicos, conservação e recuperação do manguezal como um grande ecossistema (Holguin, Vazquez e Bashan 2001; Alongi 2002). Foi relatado que bactérias, arqueias e fungos formam 91% da biomassa de micro-organismos

existentes nesse bioma, seguido de 7% formado por microalgas e 2% por protozoários (Alongi 2002).

O sedimento de manguezal apresenta intensa atividade metabólica realizada pelos micro-organismos presentes, atuando tanto de maneira individual quanto através de relações sintróficas. Esses micro-organismos têm fundamental importância para a manutenção desse ecossistema, uma vez que participam de processos biogeoquímicos cruciais para o seu funcionamento (Santos *et al.* 2011). A microbiota do manguezal desempenha atividades como fixação de nitrogênio, redução de sulfato, ciclo do fósforo, transformação dos nutrientes, fotossíntese, metanogênese, além da produção de compostos de interesse biotecnológico, como antibióticos, enzimas e biossurfactantes (Holguin, Vazquez and Bashan 2001; Dias *et al.* 2009; Aniszewski *et al.* 2010). Apesar de ter muita matéria orgânica disponível em seu sedimento, o manguezal é acometido pela escassez de nitrogênio e fosfato, porém sua alta produtividade é atribuída aos micro-organismos que realizam o processo eficiente de ciclagem de nutrientes (Vazquez *et al.* 2000; Holguin, Vazquez and Bashan 2001).

A comunidade microbiana tem sofrido impacto direto dos contínuos derramamentos de petróleo nas áreas de manguezal, tornando-o, de certa forma, um importante alvo para estudos biológicos (Marcial Gomes *et al.* 2008; Brito *et al.* 2009). É provado que ambientes que apresentam condições extremas como baixas ou elevadas temperaturas, pH ácidos ou alcalinos, anaerobiose, altas concentrações de sais e elevada pressão, são ideais para atividades de bioprospecção de genes para extremoenzimas com potencial biotecnológico (Hough e Danson 1999; van den Burg 2003; Schiraldi e De Rosa 2002). Portanto, por apresentarem elevada salinidade, baixa concentração de O_2, déficit de alguns nutrientes (nitrogênio e fósforo), variações do pH, e muitas vezes altas concentrações de poluentes orgânicos e inorgânicos, os manguezais são uma promissora fonte para prospecção de micro-organismos capazes de tolerar valores acima ou abaixo da média considerada normal para os parâmetros mencionados, como por exemplo, algas, bactérias e arqueias halotolerantes ou halofílicas (que suportam ou necessitam de concentrações elevadas de sal) ou bactérias capazes de degradar altas concentrações de hidrocarbonetos poluentes (Holguin, Vazquez e Bashan 2001; Lewis, Pryor e Wilking 2011; Alongi 2014; Wang *et al.* 2017).

Nos últimos anos, estudos têm sido feitos na tentativa de caracterizar a diversidade taxonômica e o papel funcional dos micro-organismos em ambientes de manguezal no Brasil (Gomes *et al.* 2007; Marcial Gomes *et al.* 2008; Brito *et al.* 2009; Taketani *et al.* 2010b; dos Santos *et al.* 2011; Andreote *et al.* 2012). Brito *et al*. (2009) avaliaram a degradação *in situ* de hidrocarbonetos do petróleo em sedimentos de manguezal por um consórcio de bactérias

hidrocarbonoclásticas. Os autores evidenciaram que as bactérias não corresponderam à expectativa de degradação do óleo, provavelmente devido à presença de matéria orgânica no ambiente natural. Os mesmos autores publicaram um outro trabalho no qual revelaram uma rica diversidade microbiana (incluindo espécies dos gêneros *Pseudomonas*, *Marinobacter*, *Alcanivorax*, *Microbulbifer*, *Sphingomonas*, *Micrococcus*, *Cellulomonas*, *Dietzia* e *Gordonia*) em sedimento de manguezal da Baia da Guanabara (Rio de Janeiro) com potencial para degradação de hidrocarbonetos e possível aplicação em processo de remediação de poluentes (Brito *et al.* 2006). Outros trabalhos publicados por Gomes *et al*. (2007; 2010) revelaram novos genótipos relacionados a genes *ndo* (principalmente genes *nagAc-like*), para naftaleno dioxigenase, em sedimento de manguezal próximo a áreas urbanas e contaminado com hidrocarbonetos aromáticos policíclicos (HAPs). Em adição, os autores evidenciaram um aumento na abundância de genes envolvidos na degradação de HA (naftaleno, extradiol e intradiol dioxigenases) em amostras de manguezal enriquecidas com contaminantes. Taketani *et al*. (2010) usaram a abordagem baseada em PCR-DGGE para avaliar as diferenças entre as comunidades microbianas presentes em mesocosmos contendo sedimento de manguezal impactado com petróleo e sedimento pristino. De acordo com os resultados, houve a emergência de bandas para comunidades microbianas em resposta à exposição ao petróleo, após 75 dias de incubação. Outro estudo revelou uma forte relação entre a comunidade bacteriana e a variação espacial na concentração de poluentes do petróleo em sedimentos de manguezal (Peixoto *et al.* 2011). No Estado de São Paulo, região de Cananéia, Taketani *et al*. (2010b) conduziram um estudo sobre diversidade microbiana em sedimento de manguezal pristino, identificando bactérias redutoras de sulfato e arqueias metanogênicas. Outro estudo realizado na Ilha do Cardoso, extremo sul de São Paulo, revelou mudanças na comunidade microbiana do sedimento de manguezal pristino em função de fatores como a localização, a variação sazonal e a profundidade das amostras coletadas (Dias *et al.* 2010). Os mesmos autores publicaram um trabalho no qual analisaram a diversidade de arqueias em três manguezais diferentes: pristino, antropogênico e afetado pelo vazamento de petróleo. Os resultados mostraram que os filos *Thaumarchaeota* (53.1%, 24 OTUs) e *Euryarchaeota* (29,6%, 14 OTUs) foram os mais abundantes, seguidos de 16,1% das sequências sem correspondência com nenhum banco de dados (Dias *et al.* 2011). O estudo concluiu que os sistemas de manguezal analisados servem como reservatório da diversidade de arqueias, as quais estão envolvidas no ciclo biogeoquímico do nitrogênio, porém esta diversidade é afetada por pressões antrópicas (Dias *et al*. 2011).

O primeiro trabalho que investigou o metagenoma de sedimento de manguezal, usando plataforma de sequenciamento “454”, resultou em um total de 905.521 sequências de DNA

(Andreote *et al*. 2012). Estas sequências foram analisadas contra banco de dados, por similaridade, e agrupadas conforme seus grupos taxonômicos e funcionais correspondentes (Andreote *et al.* 2012). Esse estudo revelou a dominância das classes bacterianas *Deltaproteobacteria* e *Gammaproteobacteria* nos sedimentos de manguezal. Com relação à análise funcional foi demonstrada uma abundância maior de genes envolvidos com metabolismo de metano, formaldeído, dióxido do carbono, redução de nitrato, imobilização de nitrogênio, desnitrificação, produção de adenilsulfato, sulfito e H_2S, envolvidos nos ciclos biogeoquímicos (Andreote *et al.* 2012).

De acordo com uma análise comparativa entre um estudo de diversidade microbiana da rizosfera de mangue e outro estudo de diversidade microbiana em sedimento de manguezal natural constatou-se uma interessante similaridade no aumento de grupos bacterianos específicos: *Acidobacteria*, *Actinobacteria*, *Verrucromicrobia*, *Burkholderiales*, *Caulobacterales* e *Rhizobiales,* e uma redução dos grupos *Chloroflexi*, *Firmicutes* e *Desulfobacterales*. Essa similaridade sugere uma relação intrínseca desses grupos de bactérias com as características específicas do ambiente de manguezal, sugerindo a possibilidade de uso desses grupos como marcadores para caracterização de sedimento de mangue e qualidade ambiental (Andreote *et al.* 2012). Podemos supor que em áreas com uma grande densidade de plantas de mangue o nível de oxigênio no sedimento é maior devido à presença de raízes que ajudam na difusão do mesmo em uma profundidade maior. Essa penetração de oxigênio em camadas mais profundas do sedimento torna o ambiente parcialmente aeróbico, permitindo o crescimento de grupos microbianos que estão adaptados a essas condições como Actinomycetales e Planctomycetales. Por outro lado, a limitação de oxigênio em áreas com pouca vegetação de mangue e ausência de rizosfera, contribuiu para a ocorrência de grupos bacterianos anaeróbicos, tais como Deltaproteobacterias (sulfato-redutoras) (Gomes *et al.* 2010; Andreote *et al.* 2012).

Um estudo realizado por Santos *et al*. (2011) usou pirosequenciamento de genes RNAr 16S para comparar a diversidade microbiana entre microcosmos contendo sedimento de manguezal pristino e impactado com óleo. Os resultados mostraram a dominância de bactérias das classes Deltaproteobacteria e Gammaproteobacteria em microcosmo com óleo, descoberta confirmada pelo estudo metagenômico realizado por Andreote *et al.* (2012). Nos sedimentos afetados pelo óleo houve um aumento dos gêneros bacterianos *Marinobacterium*, *Marinobacter* e *Cycloclasticus* e uma redução de bactérias da ordem *Chromatiales* e do gênero *Haliea*. Resultados como esses levam a iniciativas para a elaboração de técnicas e protocolos com a

finalidade de realizar biomonitoramento de áreas de manguezal afetadas ou sob o risco de contaminação por óleo (Santos *et al.* 2011).

Apesar dos recentes trabalhos terem apontado o manguezal como detentor de uma grande diversidade funcional e taxonômica de micro-organismos, muitos trabalhos ainda precisam ser feitos para ampliar o conhecimento nesta área através do isolamento, identificação e exploração funcional de novas espécies. A presença de uma grande diversidade de micro-organismos, tanto em ambientes pristinos quanto em ambientes contaminados pelo derramamento de petróleo, sugere a possível presença de genes e rotas metabólicas envolvidas na biodegradação de compostos orgânicos poluentes.

2.3 Petróleo e hidrocarbonetos aromáticos

O petróleo é uma mistura complexa e variável de compostos de origem natural, em sua maioria hidrocarbonetos (97%) seguido de orgânicos sulfurados, nitrogenados, oxigenados e organometálicos, formado pela combinação de fatores geológicos, biológicos e físico-químicos (Sugiura *et al.* 1997). Em estudo realizado por Marshall e Rodgers (2004), objetivando a identificação por espectrometria de massas de alto desempenho dos componentes do petróleo, foi revelada a presença de mais de 17.000 substâncias diferentes em solução. O petróleo e seus subprodutos são altamente recalcitrantes e podem demorar anos para se decompor, podendo afetar o equilíbrio ambiental, alterando os ciclos biogeoquímicos e a cadeia trófica (Holguin, Vazquez e Bashan 2001; Tam, Wong e Wong 2005; Marcial Gomes *et al.* 2008). Um dos principais problemas encontrados após um derramamento de petróleo é o efeito tóxico dos poluentes orgânicos que ficam impregnados nos solos, ambientes aquáticos e sedimentos impactados pela extração, refino, transporte e uso do petróleo, comprometendo a viabilidade da fauna, flora e microbiota presentes (Margesin *et al.* 2003; Le Borgne, Paniagua e Vazquez-Duhalt 2008).

Estima-se que 7.847 derramamentos de óleo ocorreram no mundo ao longo de 39 anos, dos quais 79,5% tinham menos de 7 t, 15,9% tinham de 7 a 700 t e 4,6% tinham mais de 700 t (Wang *et al.* 2014). Milhões de dólares são empregados todo ano para conter os impactos causados pelo derramamento de petróleo ao redor do mundo (Smits *et al.* 2002; Kubota *et al.* 2008; Xu, Xiao e Wang 2008). A biorremediação é uma das ferramentas mais promissoras no tratamento de ambientes impactados por esses poluentes orgânicos, uma vez que aproveita o potencial adaptativo de micro-organismos, incluindo sua rica diversidade metabólica, para o desenvolvimento de produtos e processos para a descontaminação ambiental. A implantação de

estratégias microbiológicas para a biorremediação de ambientes contaminados por petróleo está focada na capacidade desses organismos em transformar ou mineralizar poluentes orgânicos em compostos de toxicidade mais branda, e que possam fluir através dos sistemas biogeoquímicos (Margesin e Schinner 2001; Lovley 2003). Portanto, a pesquisa e o entendimento dos genes e enzimas das vias metabólicas envolvidas na biodegradação de compostos orgânicos, como os hidrocarbonetos derivados do petróleo, são de suma importância para a criação de métodos de resposta rápida a acidentes que possam causar desastres ambientais (Seo, Keum e Li 2009).

Os hidrocarbonetos constituem um dos grupos de substâncias mais abundantes do planeta, presentes nos flavonóides, quinonas, aminoácidos aromáticos, combustíveis fósseis e como componente da lignina, que, por sua vez, é o segundo composto mais abundante do mundo, atrás apenas da celulose (Carmona *et al.* 2009). Estes compostos são classificados em cíclicos ou alifáticos/saturados (alcanos, alcenos e alcinos), aromáticos (benzênicos e policíclicos) e compostos polares (resinas e asfaltenos) (Yender *et al.* 2002). Os compostos aromáticos são moléculas que possuem anéis aromáticos, como o benzeno (Figura 2), e são divididos em três grandes categorias: hidrocarbonetos aromáticos policíclicos (HAPs), heterocíclicos e aromáticos substituídos (Seo, Keum e Li 2009). Os anéis benzênicos dos compostos aromáticos lhes dão estabilidade térmica, tornando-os vantajosos para aplicações industriais, porém prejudiciais quando liberados no meio ambiente, devido à sua alta taxa de aderência a material particulado suspenso ou depositado e baixa solubilidade, favorecendo sua longa permanência em solos e sedimentos (Ghosal *et al.* 2016).

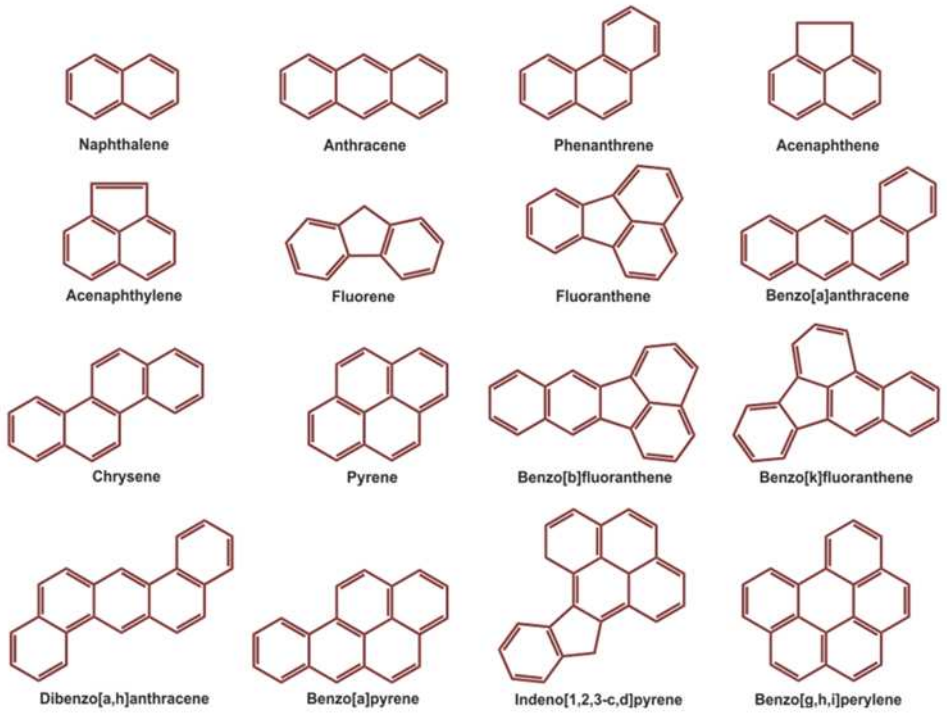


Figura 2. Estrutura molecular de 16 hidrocarbonetos aromáticos (HAs) considerados fortes poluentes ambientais de acordo com "United States Environmental Protection Agency - US EPA". Fonte: (Ghosal *et al.* 2016).

Os HAPs são compostos orgânicos hidrofóbicos com dois ou mais anéis benzênicos, ligações duplas e simples que se alternam em sua estrutura molecular com arquitetura linear, angular ou agrupada (Cerniglia 1992; Ghosal *et al.* 2016). Propriedades químicas, tais como solubilidade, volatilidade e reatividade mudam em função do seu peso molecular e número de anéis benzênicos (Soclo, Garrigues e Ewald 2000; Seo, Keum e Li 2009). Os HAPs podem ser classificados em dois grupos de acordo com seu peso molecular e propriedade química: i) HAPs de 2 a 3 anéis aromáticos e com baixo peso molecular (naftaleno, bifenilo, fluoreno, acenafteno, acenaftileno, dibenzotiofeno, fenantreno e antraceno); ii) HAPs com 4 ou mais anéis aromático e com alto peso molecular (fluoranteno, criseno, benzo(e)pireno, pireno, benzo(a)antraceno, benzo(b)fluoranteno, benzo(k)fluoranteno, benzo(a)pireno, indeno[1,2,3-cd] pireno, dibenzo(a,h)antraceno, benzo(g,h,i)pirileno) (Tam *et al.* 2001).

Os HAPs estão presentes no meio ambiente por conta da atividade biogênica (transformações naturais da biomassa vegetal e animal) e antropogênica (combustão veicular, queimadas não naturais, exploração, processamento industrial e transporte de petróleo e seus subprodutos, efluentes industriais e domésticos, etc) (Seo, Keum e Li 2009; Ghosal *et al.* 2016) (Figura 3). A manipulação incorreta de produtos contendo resíduos de HAPs pode causar tanto a contaminação direta do ser humano quanto a contaminação de ambientes aquáticos, do ar, do

solo e dos seres vivos em geral (Haritash e Kaushik 2009). O contato com HAPs pode causar risco à saúde do homem devido ao seu caráter lipolítico, cumulativo, mutagênico e carcinogênico (Seo, Keum e Li 2009; Rengarajan *et al.* 2015).

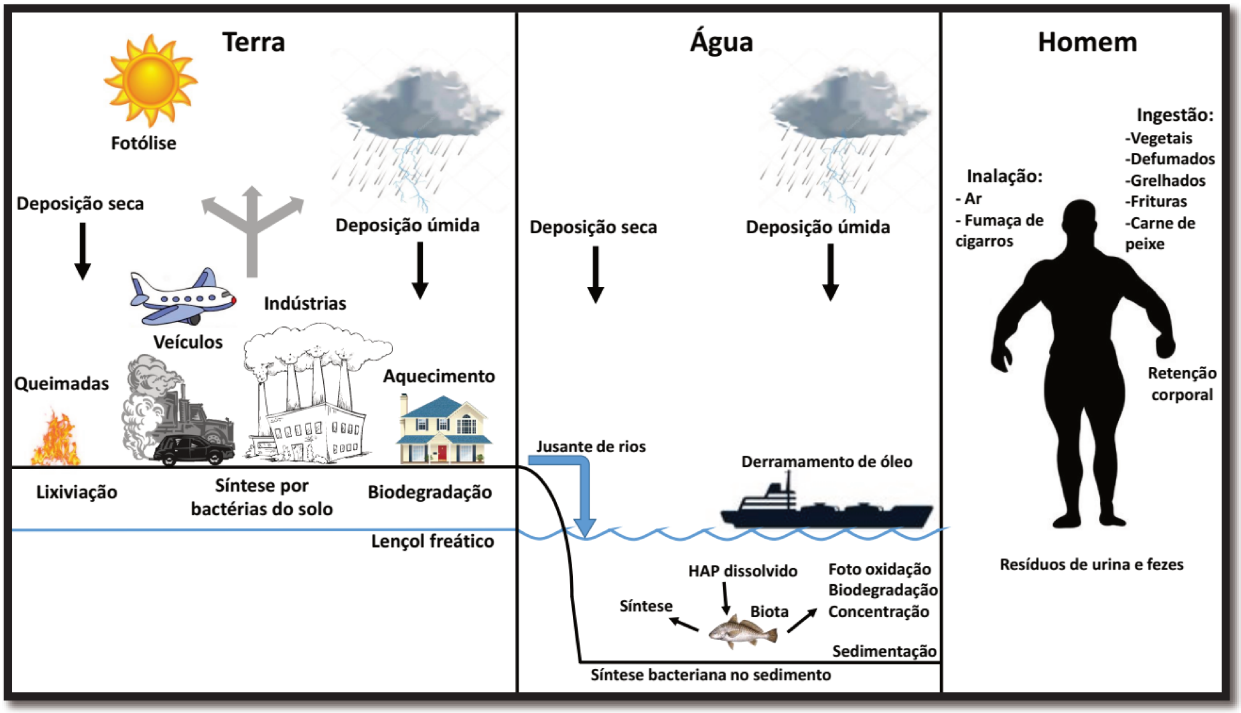


Figura 3. Dispersão dos hidrocarbonetos aromáticos através da terra, do ar e da água até chegar ao homem. Esquema extraído da seguinte fonte: Abdel-Shafy e Mansour (2015).

Os HAPs podem ser eliminados do ambiente através da volatização, foto-oxidação, oxidação química, bioacumulação, adsorção a partículas sólidas ou degradação enzimática por microrganismos (Semple, Reid e Fermor 2001; Singh *et al.* 2008; Megharaj *et al.* 2011). Os microrganismos são usados como alternativa mais barata em processos de biorremediação, tais como bioestimulação, bioaumento, humificação, *landfarming*, compostagem, dentre outros (Ghosal *et al.* 2016). Novos estudos com engenharia genética e biologia sintética têm sido feitos na tentativa de aumentar o potencial de microrganismos para degradação de hidrocarbonetos do petróleo em ambientes contaminados, vislumbrando seu uso em processos de biorremediação (Singh *et al.* 2008; Ghosal *et al.* 2016).

2.4 Biodegradação de hidrocarbonetos aromáticos

A degradação microbiana de compostos aromáticos depende de fatores bióticos (diversidade microbiana com competência metabólica para degradação de HAs) e abióticos (aceptores de elétrons, pH, temperatura, nutrientes e salinidade) (Semple, Reid e Fermor 2001; Ghosal *et al.* 2016). Forças evolutivas (mutação, seleção natural, deriva genética e fluxo gênico)

têm dirigido o estabelecimento de comunidades microbianas com habilidades para degradação de HAs, tais como crescimento rápido sob diferentes condições ambientais, plasticidade genética e versatilidade metabólica (Diaz 2004; Head, Jones e Roling 2006). Algumas revisões têm relatado vários estudos que usaram a abordagem clássica de cultivo bacteriano para evidenciar a habilidade de vários gêneros na degradação de compostos aromáticos, a saber: *Arthrobacter*, *Acidovorax*, *Bacillus*, *Burkholderia*, *Marinobacter*, *Xanthomonas*, *Mycobacterium, Ralstonia*, *Pseudomonas*, *Rhodococcus*, *Stenotrophomonas*, *Janibacter*, *Sphingomonas*, *Achromobacter*, *Alcaligenes*, *Brevibacterium*, *Chryseobacterium*, *Cycloclasticus*, *Nocardioides*, *Pasteurella*, *Polaromonas*, *Rhodanobacter*, *Staphylococcus* e *Terrabacter* (Carmona *et al.* 2009; Seo, Keum e Li 2009).

Na natureza, os compostos aromáticos podem ser degradados na ausência ou presença de oxigênio, por bactérias anaeróbicas e aeróbicas, respectivamente (Carmona *et al.* 2009). No processo de degradação aeróbico, o oxigênio atua como aceptor de elétrons e co-substrato durante a atividade enzimática de oxigenases que clivam ou hidroxilam anéis aromáticos (Parales e Resnick 2006; Vaillancourt, Bolin e Eltis 2006). No catabolismo anaeróbico ocorrem reações de redução molecular seguida pelo ataque ao anel aromático. Nessas reações os aceptores finais de elétrons podem ser íons férricos, nitratos e sulfatos, e são realizadas por bactérias redutoras de íon férrico, desnitrificantes ou redutoras de sulfato, respectivamente (Heider e Fuchs 1997; Van Hamme, Singh e Ward 2003). Dentre os hidrocarbonetos que podem ser degradados na ausência de oxigênio temos: alcilbenzenos monocíclicos, incluindo tolueno, etilbenzeno, propilbenzeno, p-cimeno, isômeros de xileno e etiltolueno, além de fenol, benzeno e naftaleno (Heider *et al.* 1998).

Os compostos aromáticos são moléculas com alta energia de ressonância estabilizante, fato que as torna difíceis para o consumo microbiano como substrato energético. Para contornar este problema, os micro-organismos desenvolveram vias metabólicas de degradação ao longo do curso evolutivo, iniciando o processo de ativação molecular de compostos aromáticos através de reações de oxidação e redução. As vias de degradação periféricas conduzem o processo de transformação estrutural de compostos aromáticos complexos até um número reduzido de intermediários simples que entram nas rotas do catabolismo central (Fuchs, Boll e Heider 2011) (Figura 4).

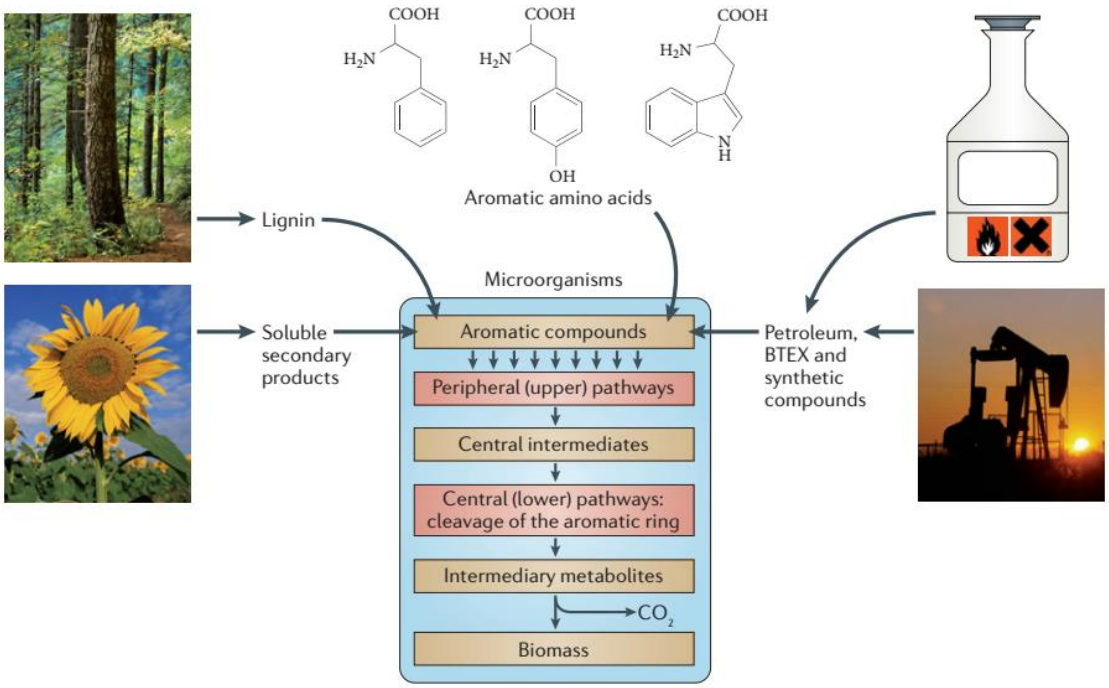


Figura 4. Fontes ambientais de compostos aromáticos e principais rotas de transformação metabólica. Fonte: Fuchs, Boll e Heider (2011).

Na via periférica aeróbica, as oxigenases ativam metabólitos intermediários (catecol, protocatecuato, gentisato ou homogentisato) por hidroxilação, que serão catabolizados nas vias centrais (Gibson e Parales 2000; Vaillancourt, Bolin e Eltis 2006; Lipscomb 2008; Fuchs, Boll e Heider 2011) (Figura 5). Na via periférica de degradação anaeróbica, grupos químicos substituintes são adicionados através de reações de redução, facilitando a transferência de elétrons até os anéis aromáticos (Fuchs, Boll e Heider 2011). O benzoil-CoA é o principal intermediário metabólico gerado neste processo, o qual seguirá o processo de degradação ao longo do catabolismo central (Carmona *et al.* 2009) (Figura 5). Além do benzoil-CoA, as vias periféricas também podem gerar outros intermediários centrais, tais como floroglucinol (1,3,5-trihidroxibenzeno), hidroxihidroquinona (HHQ) (1,2,4,-trihidroxibenzeno), resorcinol (1,3-dihidroxibenzeno), 6-hidroxinicotinato, hidroxibenzoil-CoA, metilbenzoil-CoA e aminobenzoil-CoA (Heider e Fuchs 1997; Fuchs 2008).

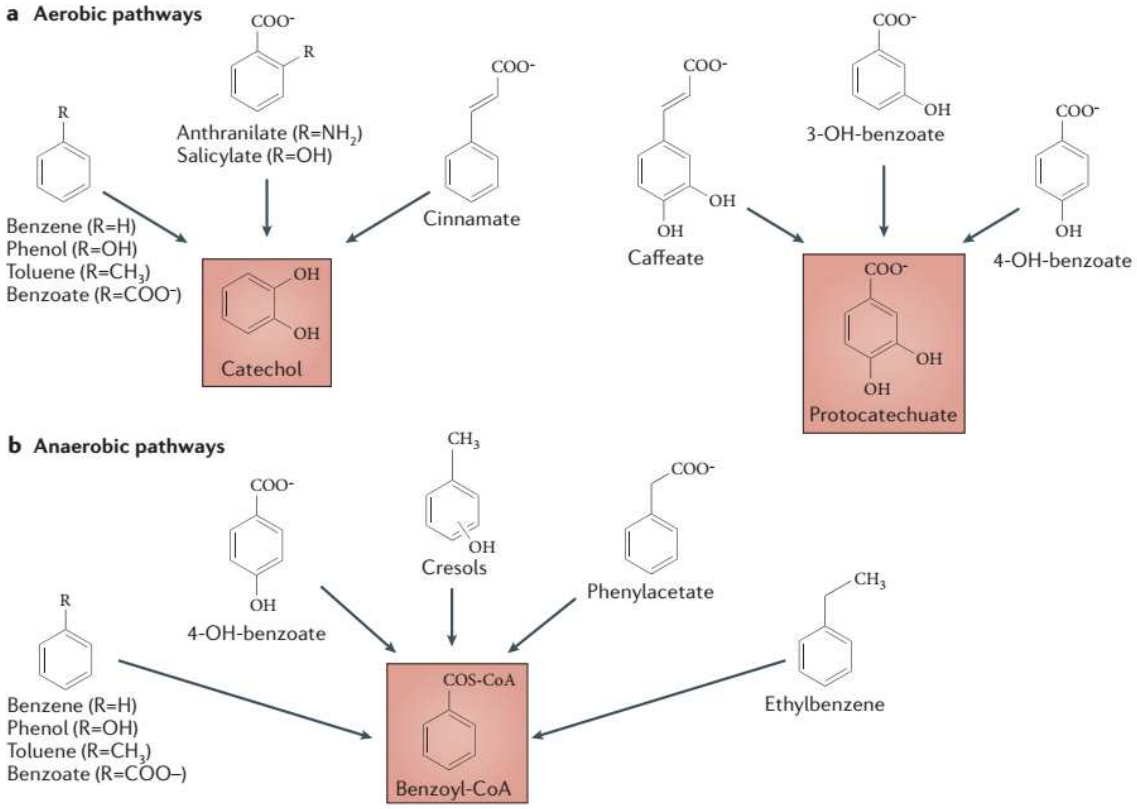


Figura 5. Principais intermediários metabólitos das rotas do catabolismo central, derivados das vias de degradação (a) aeróbica e (b) anaeróbica. Fonte: Fuchs, Boll e Heider (2011).

Durante a degradação aeróbica de compostos aromáticos, tais como HAPs, micro-organismos usam enzimas dioxigenases que inserem grupos hidroxilas nos anéis aromáticos, formando cis-dihidrodiol. Após sua formação, este composto é convertido em um intermediário diol por desidrogenases, voltando à sua forma aromática. Nesta forma, o intermediário diol pode ser clivado via orto- ou meta-clivagem pela dioxigenase que cliva anéis aromáticos intradiol ou extradiol (do inglês *intradiol* or *extradiol ring-cleaving dioxygenases*), respectivamente. Por fim, o metabólito intermediário clivado entra no ciclo dos Ácidos Tricarboxílicos (ATC) (Mallick, Chakraborty e Dutta 2011; Ghosal *et al.* 2016).

O benzoato é um composto aromático monocíclico e produto metabólico da via periférica do catabolismo de compostos aromáticos. No início do processo de degradação aeróbica deste composto, dioxigenases e monoxigenases adicionam grupos hidroxila em seu anel aromático, originando intermediários reativos que levam à formação de catecol e protocatecuato, respectivamente. Estes compostos podem ser gerados não apenas a partir do benzoato mas também de um arsenal de compostos fenólicos presentes na natureza. A estrutura molecular do catecol pode ser clivada por dioxigenases na posição entre os dois grupos hidroxila ou em uma posição adjacente aos grupos hidroxila, através das vias de orto- e meta-

clivagem, respectivamente. Processo similar ocorre com o protocatecuato. Na rota da orto-clivagem, tanto o catecol quanto o protocatecuato são clivados por dioxigenases específicas gerando metabólitos intermediários que seguem até a formação β-cetoadipato que, por sua vez, é convertido em succinato, entrando na via do ATC. Na rota da meta-clivagem, o catecol e o protocatecuato não seguem pela via do β-cetoadipato, mas são clivados por dioxigenases em metabólitos intermediários que seguem, respectivamente, até a formação *de* acetaldeído + piruvato ou apenas piruvato, os quais entram na via do ATC (Fuchs, Boll e Heider 2011) (Figura 6).

Existe um interesse muito grande na busca e entendimento de novos genes e vias metabólicas de degradação de compostos aromáticos, a fim de desenvolver processos biotecnológicos mais eficientes para a biorremediação de áreas contaminadas (Gkorezis *et al.* 2016). Por serem ambientes ainda pouco explorados, sedimentos de manguezais vêm sendo vistos como uma provável fonte de novos genes e enzimas com potencial para a biorremediação. E, generalizando, poderíamos dizer que sedimentos de manguezais contaminados por petróleo se constituem em uma fonte ainda mais promissora desses genes, uma vez que estão enriquecidos por compostos aromáticos, selecionando grupos específicos de micro-organismos capazes de degradar tais compostos e utilizá-los como fonte de energia.

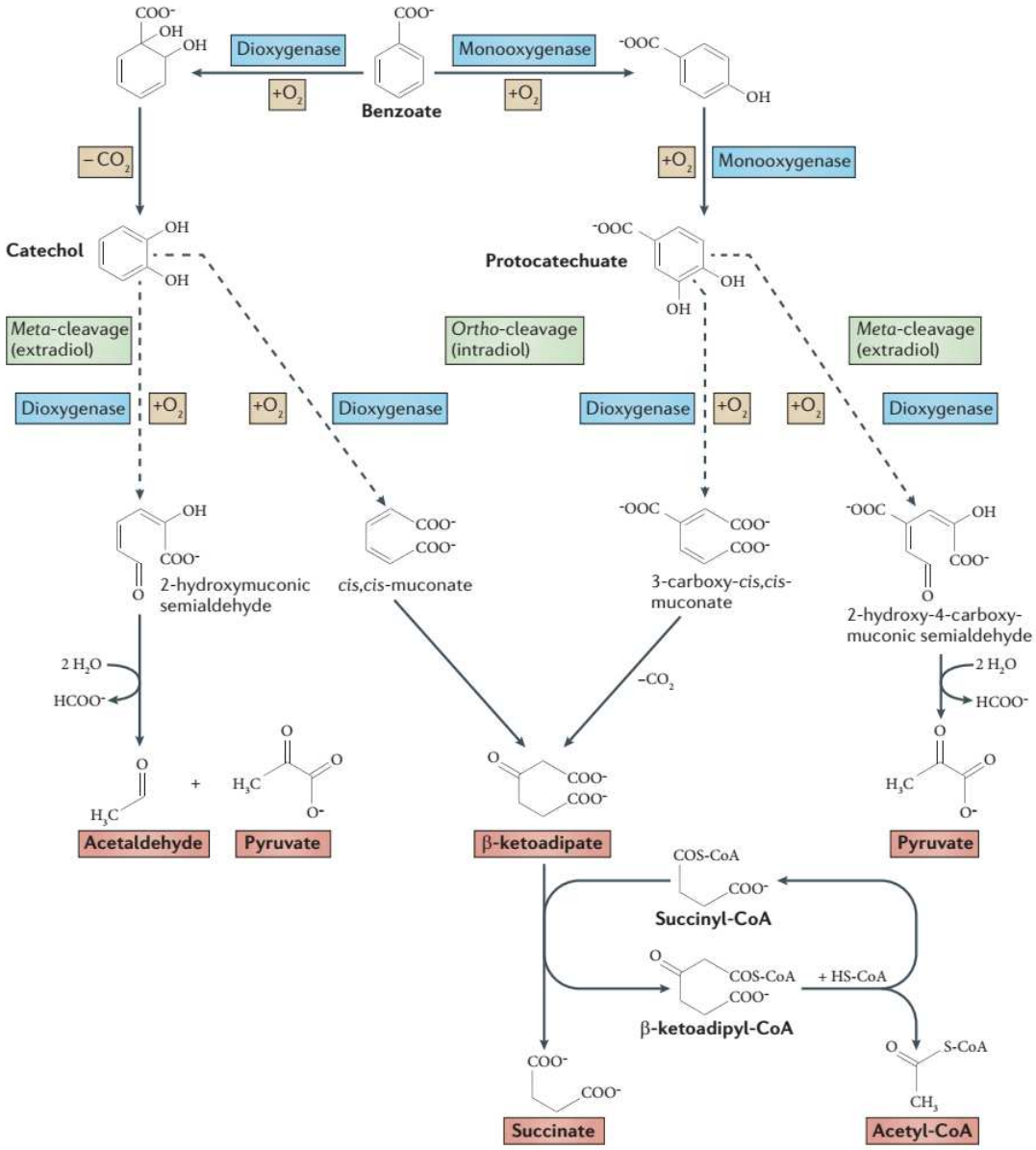


Figura 6. Principais etapas do metabolismo central da degradação de compostos aromáticos através da orto ou meta-clivagem do catecol e protocatecuato, derivados do benzoato, conduzidas por dioxigenases e monoxigenases. Fonte: Fuchs, Boll e Heider (2011).

2.5 Genes para ARHD (*Aromatic Ring Hydroxylating Dioxygenases*)

Estudos têm sido realizados para identificação e caracterização de novos genes de dioxigenases, tais como as dioxigenases que hidroxilam anéis aromáticos (*Aromatic Ring Hydroxylating Dioxygenases* - ARHDs) (Martin *et al.* 2013; Wu *et al.* 2014; Zeng *et al.* 2017). As ARHDs são enzimas de múltiplas subunidades, geralmente responsáveis pelo primeiro passo do processo de degradação de compostos aromáticos. A reação consiste basicamente na incorporação de dois átomos de oxigênio molecular ao anel benzênico do substrato, com a participação dos cofatores NADH e NADPH e liberação de CO_2 e água, durante o processo

(Lu, Zhang e Fang 2011). Estas enzimas consistem de um componente catalítico para oxigenases, contendo uma porção terminal com função de hidroxilação ligado a um domínio ferrodoxina e a uma flavoproteína redutase. A porção terminal contém as subunidades alfa e beta, das quais as α-subunidades são as mais conservadas pela evolução. A α-subunidade contém um sítio mononuclear de ferro e um centro catalítico Rieske (2Fe-2S), os quais determinam a especificidade da enzima pelo substrato (Jouanneau *et al.* 2011) (Figura 7). A nomenclatura dos genes é determinada conforme o nome do substrato usado pelas suas correspondentes enzimas: genes *bph*A (bifenilo), *bnz*A (benzeno), *ndo*B (naftaleno), *cba*A (clorobenzoato), *xyl*X (toluato), *tod*C1 (tolueno), *Cum*A1 (cumeno), *ipb*A1 (isopropilbenzeno), *edo*A (etilbenzeno), *ebd*A (alquilbenzeno), dentre outros (Yeates, Holmes e Gillings 2000).

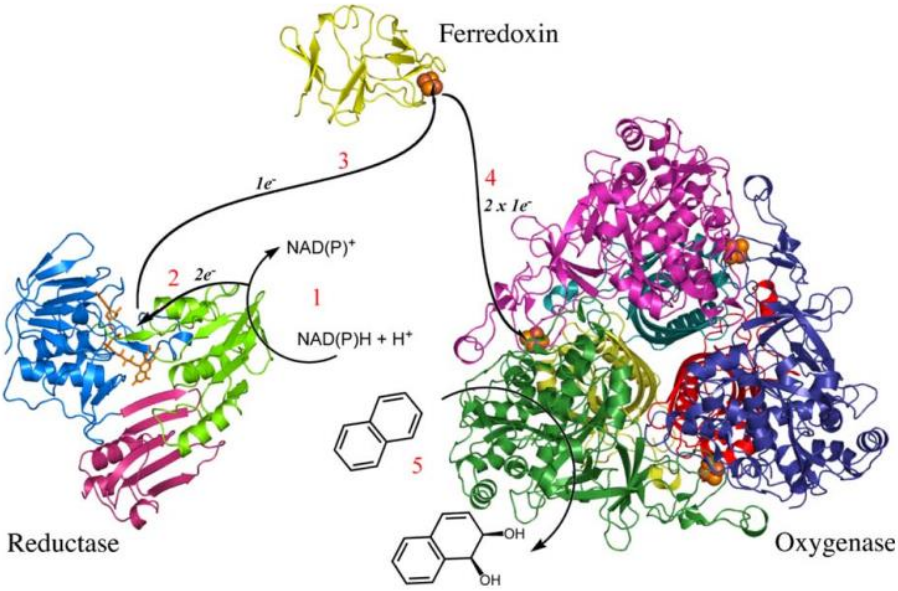


Figura 7. Modelo estrutural dos três componentes presentes em um sistema oxigenase Rieske. (1) Oxidação do NAD(P)H a NADP+ pela redutase, capturando 2 elétrons; (2) os elétrons são armazenados em uma flavina; (3) os elétrons são enviados ao componente ferrodoxina um de cada vez; (4) a ferrodoxina envia os elétrons recebidos ao centro catalítico Rieske da oxigenase; (5) esta etapa ocorre duas vezes para cada produto final formado no sítio de ferro mononuclear. Fonte: Ferraro, Gakhar e Ramaswamy 2005.

Geralmente, o sistema de classificação das dioxigenases em famílias tem como base sua especificidade ao substrato e a similaridade entre as sequências (Werlen, Kohler e van der Meer 1996; Gibson e Parales 2000; Nam *et al.* 2001) (Figura 8). Batie e colaboradores (1987) sugeriram um dos primeiros sistemas de classificação para os genes ARHD, baseado na combinação de seus componentes protéicos. Gibson e Parales (2000) classificaram as dioxigenases em quatro famílias: Tolueno/bifenil, Naftaleno, Benzoato e fitalate. As famílias

Tolueno/bifenil (T/B) e Naftaleno (PAH) são as mais bem estudadas até agora, mostrando uma grande diversidade genética (Iwai *et al.* 2011). Nam *et al.* (2001) também propuseram uma classificação para os genes ARHD, considerando o alinhamento de sequências de aminoácidos a partir da região da α-subunidade das dioxigenases. Este esquema de classificação é baseado em quatro grupos de oxigenases: I (oxigenases de baixa homologia), II (dioxigenases de benzoato/toluato), III (dioxigenases de hidrocarbonetos aromático policíclico/naftaleno) e IV (dioxigenases de benzeno/tolueno/bifenil). Atualmente, o trabalho de Iwai *et al.* (2011) tem sido visto como o melhor estudo sobre a diversidade e classificação filogenética de dioxigenases oriundas de diferentes amostras ambientais. De acordo com suas análises, os genes para dioxigenases foram agrupados em cinco subclados: PAH-GN (dioxigenases para PAHs de bactérias Gram-negativas), PAH-GP (dioxigenases *para* PAH de bactérias Gram-positivas), T/B (dioxigenases de tolueno/bifenil), OT-I (outras dioxigenases I) e OT-II (outras dioxigenases II) (Figura 8). Além disso, um grande grupo gênico foi classificado como benzoato dioxigenases ou dioxigenases tipo Rieske não caracterizadas funcionalmente (Figura 8). Meynet *et al.* (2015) publicaram uma reavaliação filogenética baseada em genes para dioxigenases a qual mostrou, por um lado, subclados com genes proximamente relacionados, derivados de micro-organismos que degradam o mesmo substrato e, por outro lado, clados englobando genes para dioxigenases derivados de micro-organismos capazes de degradar diferentes substratos.

Devido ao pequeno volume de estudos realizados sobre a microbiota de manguezais, existe um grande interesse em explorar seus recursos genéticos para a prospecção dessas ARHDs, oriundas de micro-organismos adaptados às condições particulares deste ecossistema.

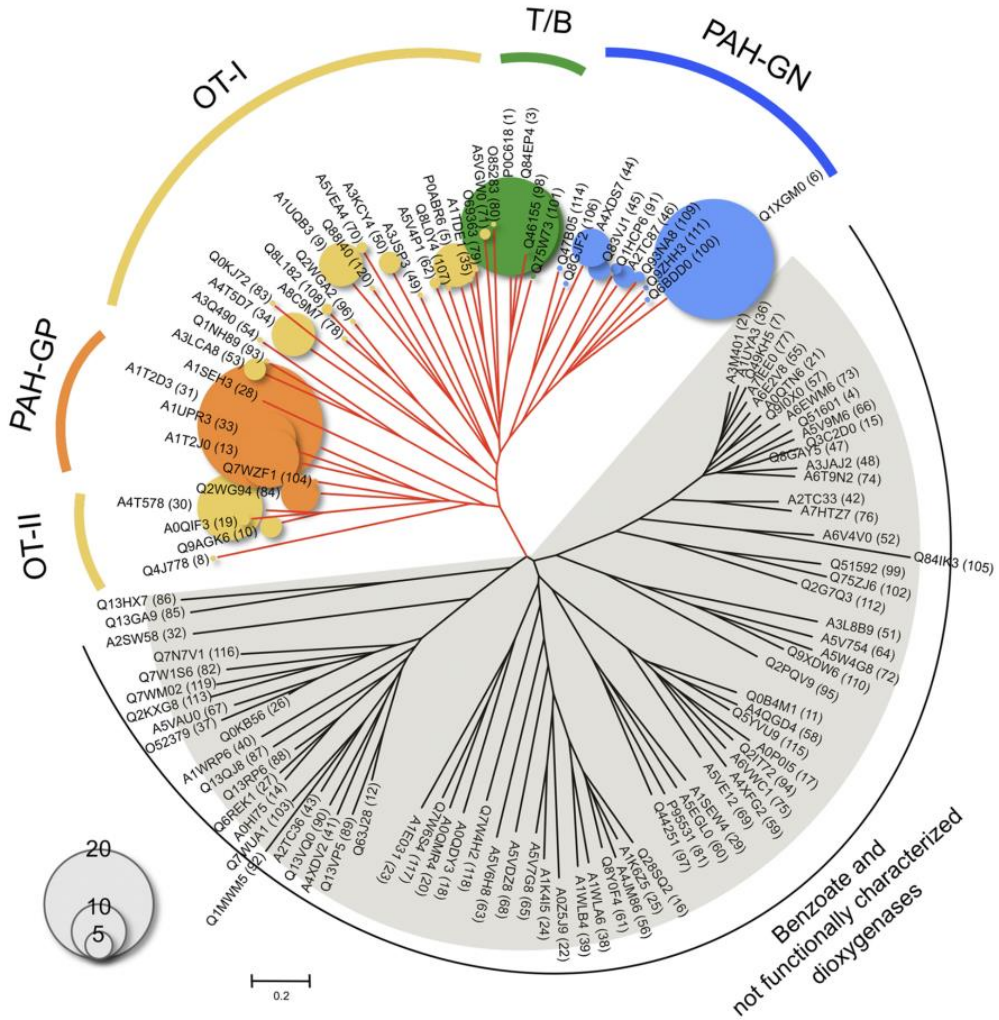


Figura 8. Análise filogenética a partir de 120 grupos de dioxigenases representando 464 genes. A árvore filogenética está dividida em cinco subclados representativos (em inglês: PAH-GP, *PAH dioxygenases from Gram-positive bacteria*; T/B, *toluene/biphenyl dioxygenases*; OT-I, *other dioxygenases I*; PAH-GN, *PAH dioxygenases from Gram-negative bacteria*; OT-II, *other dioxygenases II*) e um grande grupo de sequências representando dioxigenases funcionalmente não caracterizadas. Fonte: Iwai *et al.* (2011).

2.6 Abordagem metagenômica para o estudo da microbiota

Uma das maiores dificuldades para estudar ecologia microbiana e novos genes envolvidos em uma gama de vias metabólicas de interesse biotecnológico, como as vias de degradação de hidrocarbonetos aromáticos do petróleo, é a grande limitação dos métodos clássicos de microbiologia baseados no plaqueamento e cultivo de micro-organismos. Uma pequena parcela do total de espécies microbianas presentes no meio ambiente é acessada através dos métodos tradicionais de isolamento, em torno de 1 a 10%. Nas últimas décadas, muitos protocolos experimentais foram estabelecidos e/ou aprimorados na área de biologia molecular, tais como extração de ácidos nucleicos, PCR de genes alvos (RNAr 16S), clonagem, construção

de bibliotecas gênicas e sequenciamento, superando limitações impostas pela abordagem clássica de estudo de populações microbianas, eliminando as etapas de isolamento e cultivo dos micro-organismos (Rappé e Giovannoni 2003). O uso destas metodologias permitiu uma mudança drástica na perspectiva de diversidade microbiana ambiental, com a descoberta de novos grupos de organismos unicelulares nunca antes cultivados (Hugenholtz, Goebel e Pace 1998; Rappé e Giovannoni 2003). Entretanto, estas metodologias não nos permitem ter acesso ao potencial metabólico destes novos organismos, uma vez que as etapas de isolamento e cultivo são suprimidas nos estudos. O uso de protocolos de extração de DNA ambiental de boa qualidade, associado à estratégia de construção de bibliotecas metagenômicas, vem permitindo o acesso ao potencial metabólico de novos microrganismos, sendo uma alternativa viável para a descoberta de novas enzimas e vias metabólicas (Handelsman *et al.* 1998).

A abordagem metagenômica pode ser descritiva e/ou quantitativa, gerando dados a respeito da diversidade e abundância funcional/taxonômica de microrganismos presentes em ambientes como solo, água, sedimento, ar, superfícies e entranhas biológicas (Quince *et al.* 2017). Sucintamente, o fluxo de trabalho da metagenômica compreende o sequenciamento em larga escala do DNA ambiental e a análise *in silico* dos dados de sequências utilizando diversas plataformas ou pipelines desenvolvidos recentemente para esta finalidade (Figura 9; Tabela 1) (Simon e Daniel 2011; De Filippo *et al.* 2012). Neste contexto, o DNA metagenômico pode ser diretamente sequenciado ou fragmentado para inserção em vetores de clonagem como plasmídeos (insertos de até 8 kb), fosmídeo (de 30 – 45 kb) e *bacterial artificial chromosomes* - BACs (até 300 kb), para a construção de bibliotecas metagenômicas. Essas bibliotecas são formadas por clones contendo vetores com fragmentos de DNA exógeno inseridos em células competentes (modificadas geneticamente para armazenar e/ou expressar o inserto metagenômico) (Giovannoni *et al.* 1990; Amann, Ludwig e Schleifer 1995; Hugenholtz, Goebel e Pace 1998). Os clones de uma biblioteca podem ser triados com base em suas sequências de DNA exógeno (ex. estudos de caracterização genética de operons e diversidade funcional e taxonômica de comunidades microbianas em amostras ambientais) ou com base em sua atividade biológica, como a produção de compostos bioativos para aplicações biotecnológicas (ex. enzimas que degradam hidrocarbonetos aromáticos para biorremediação de ambientes contaminados por derramamento de pretróleo) (Uchiyama e Miyazaki 2009; Silva *et al.* 2013; Lam *et al.* 2015).

Estudos metagenômicos requerem a aplicação de tecnologias de sequenciamento em larga escala e bastante acuradas, pois o objetivo é acessar o máximo de informação possível e de alta qualidade, do conjunto total de sequências amostradas. Tecnologias de sequenciamento

têm sido aprimoradas ao longo dos anos, e vem auxiliando imensamente nos estudos metagenômicos. Neste cenário, as seguintes plataformas de sequenciamento têm se destacado: GS FLX e GS Junior (Roche 454); GA II, HiSeq, NextSeq, e MiSeq (Illumina); SOLiD (Applied Biosystems); Ion PGM e Ion Proton (Ion Torrent); PacBio RS (Pacific Biosciences, mas agora como parte da Roche); Polonator (Dover); HeliScope (Helicos BioSciences); e, mais recentemente, MinION (Oxford Nanopore) (Branton *et al.* 2008; Shendure e Ji 2008; Levy e Myers 2016). Entretanto, como consequência do avanço das plataformas de sequenciamento, um volume enorme de dados de sequências está sendo gerado a partir dos estudos metagenômicos. Assim, a bioinformática surge como ferramenta essencial de processamento e exposição dos resultados de forma legível (Teeling e Glöckner 2012). Diversos programas e pipelines têm sido desenvolvidos para ajudar a responder questões biológicas ancoradas em hipóteses científicas, através de processamento dos dados gerados por procedimentos experimentais padronizados e validados (Segata *et al.* 2013). Para o processamento de dados metagenômicos, um grande arsenal de *softwares* está disponível na internet, incluindo aqueles para montagem de sequências, anotação funcional, classificação taxonômica, comparação, análises estatísticas, dentre outros (Tabela 1) (De Filippo *et al.* 2012; Teeling e Glöckner 2012).

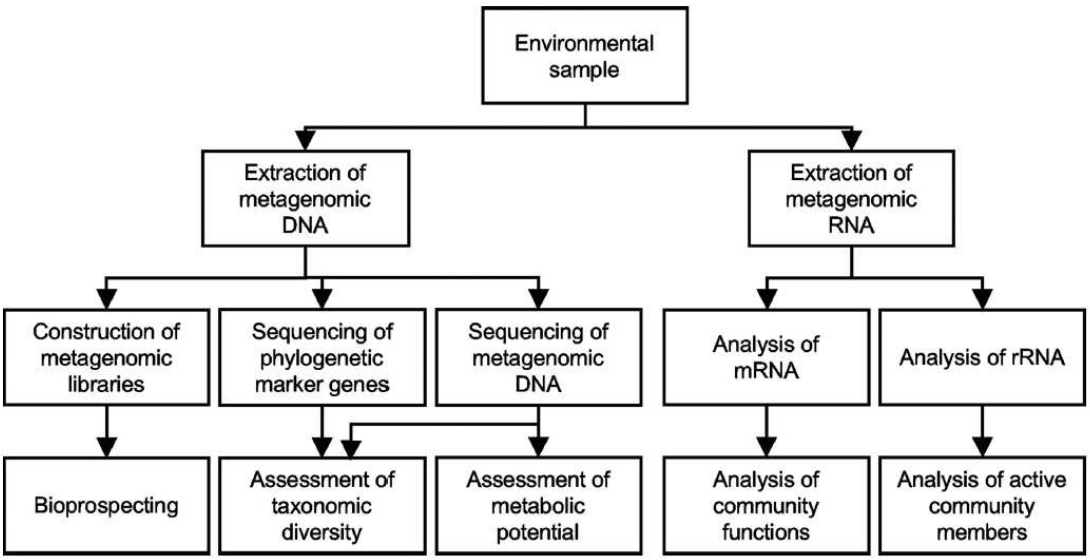


Figura 9. Fluxograma de uma análise metagenômica de comunidades microbianas ambientais a partir da extração de ácidos nucléicos totais. Fonte: Simon *et al*. (2011).

Tabela 1. Relação de alguns programas para análise de bioinformática de dados metagenômicos. Fonte: Filippo *et al.* (2012).

Scope	Name	Link to program
Recruitment	BWA	bio-bwa.sourceforge.net
	Bowtie	bowtie-bio.sourceforge.net
	FR-HIT	weizhong-lab.ucsd.edu/frhit
Assembly	Meta-Velvet	metavelvet.dna.bio.keio.ac.jp
	META-IDBA	i.cs.hku.hk/~alse/hkubrg/projects/metaidba/
	IDBA-UD	i.cs.hku.hk/~alse/hkubrg/projects/idba.ud/
	Genovo	cs.stanford.edu/group/genovo/
Genes	FragGeneScan	omics.informatics.indiana.edu/FragGeneScan
	MGA	whale.bio.titech.ac.jp/metagene
	Glimmer-MG	www.cbcb.umd.edu/software/glimmer-mg
	GeneMark	exon.gatech.edu/metagenome
Annotation	RPSBlast	www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
	HMMer3	hmmer.janelia.org
	BLAST	blast.ncbi.nlm.nih.gov
	RAPSearch2	omics.informatics.indiana.edu/mg/RAPSearch2
	RAST	rast.nmpdr.org/
Taxonomy	RDPclassifier	rdp.cme.msu.edu
	NBC	nbc.ece.drexel.edu
	CARMA3	webcarma.cebitec.uni-bielefeld.de
	MEGAN	ab.inf.uni-tuebingen.de/software/megan
	SOrt-ITEMS	metagenomics.atc.tcs.com/binning/SOrt-ITEMS
Servers	MG-RAST	metagenomics.anl.gov
	IMG/M	img.jgi.doe.gov/
	EBI metagenomics	https://www.ebi.ac.uk/metagenomics/
Models	PathwayTools	bioinformatics.ai.sri.com/ptools/
	Model SEED	seed-viewer.theseed.org/seedviewer.cgi?page=ModelView
Analysis	GSEA	www.broadinstitute.org/gsea/
	ShotgunFunctionalizeR	http://shotgun.math.chalmers.se/
	MetaPath	www.cbcb.umd.edu/~boliu/metapath/
	STAMP	http://kiwi.cs.dal.ca/Software/STAMP
	HUMAnN	uttenhower.sph.harvard.edu/humann

2.7 Aspectos sobre biogeografia microbiana

Biogeografia é a ciência que investiga a distribuição dos organismos no espaço e ao longo do tempo, e seus padrões de diversidade. Esses estudos ajudam a elucidar o processo evolutivo dos seres vivos através da compreensão dos mecanismos de extinção, especiação, dispersão e interação entre as espécies (Martiny *et al.* 2006). Apesar da maioria das respostas sobre biogeografia ter sido gerada a partir de estudos sobre plantas e animais, vários trabalhos com bactérias, arqueias, vírus, fungos e outros micro-organismos eucariontes tem sido realizados recentemente (Martiny *et al.* 2006; Hanson *et al.* 2012). Muitos desses trabalhos revelaram uma sensível correlação entre padrões biogeográficos de macro-organismos e micro-organismos, ajudando os microbiologistas a entenderem melhor os índices de riqueza e composição microbiana (Martiny *et al.* 2006; Hanson *et al.* 2012). Apesar dessas correlações,

é preciso ter cautela no estabelecimento de fundamentos teóricos sobre os padrões biogeográficos microbianos. A maioria dos estudos de padrões biogeográficos microbianos não compreende os padrões de distribuição espacial em grandes escalas de tempo, ao contrário dos estudos sobre plantas e animais. Portanto, quase sempre a diversidade microbiana é avaliada sob efeito de sazonalidade, mudanças naturais incomuns ou impactos antropogênicos em seu ecossistema. Normalmente, os estudos sobre biogeografia microbiana são suportados por evidências genéticas. Por esse motivo, *taxa* microbianos são definidos pela variação genética de *loci* genômicos (principalmente os mais conservados) agrupados como Unidades Taxonômicas Operacionais (do inglês "*operational taxonomic units*" - OTUs) (Hanson *et al.* 2012).

Os estudos filogenéticos e biogeográficos podem contribuir para um melhor entendimento sobre a história evolutiva, estratégias de sobrevivência e papel ecossistêmico dos microrganismos. Neste sentido, trabalhos focados em correlacionar os resultados de congruência filogenética (genômica estrutural) e coerência ecológica (atividade biológica) são fundamentais para entender o processo evolutivo dos microrganismos.

No geral, o presente estudo se baseou no propósito vantajoso de investigar os recursos genéticos provenientes de ambientes já contaminados, pois, teoricamente, o enriquecimento com hidrocarbonetos aromáticos propicia a seleção de populações microbianas envolvidas no seu metabolismo. Neste sentido, o plano de trabalho visou identificar e quantificar genes envolvidos nas vias de degradação de hidrocarbonetos aromáticos no sedimento de manguezal impactado, investigar o padrão de distribuição destes genes pelo mundo e seus principais mecanismos de dispersão, elucidar o potencial genético natural para a atenuação da poluição ambiental e explorar o arsenal gênico da microbiota para o desenvolvimento de futuros processos biotecnológicos

3. OBJETIVO GERAL

Análise da diversidade, biogeografia e caracterização estrutural de genes envolvidos no metabolismo de compostos aromáticos em amostras de sedimento de manguezal impactado por petróleo.

3.1. Objetivos específicos

- ✓ Determinação de padrões biogeográficos de genes para dioxigenases obtidos de biblioteca plasmidial do sedimento de manguezal impactado e de outros ambientes investigados;
- ✓ Análise *in silico* dos dados gerados pelo sequenciamento (em plataforma *454* e *Illumina*) da biblioteca metagenômica do mangue impactado por petróleo (Bertioga-SP) para a determinação da diversidade funcional e taxonômica de genes envolvidos no metabolismo de compostos aromáticos;
- ✓ Triagem dos clones da biblioteca metagenômica para atividade de degradação de hidrocarbonetos aromáticos através de ensaios funcionais de alto desempenho (HTS - *High Throughput Screening*);
- ✓ Avaliação do potencial de clones positivos para degradação de hidrocarbonetos aromáticos através de cromatografia gasosa acoplada a espectrometria de massas (CG-EM);
- ✓ Sequenciamento dos clones positivos para caracterização estrutural de *clusters* envolvidos na degradação de hidrocarbonetos aromáticos;
- ✓ Amplificação, clonagem, expressão dos genes alvos envolvidos no metabolismo de compostos aromáticos (ex. dioxigenases), para purificação e caracterização molecular de suas respectivas proteínas.

4. ESTRUTURA DA TESE

Este trabalho visou analisar e explorar a diversidade de genes para dioxigenases e vias metabólicas envolvidas no metabolismo de compostos aromáticos a partir de clones fosmidiais de uma biblioteca metagenômica de sedimento de manguezal impactado por petróleo. Além disso, foi investigada a filogenia e biogeografia de genes para dioxigenases que hidroxilam anéis aromáticos em uma perspectiva global, através da comparação entre diversos estudos realizados em diferentes regiões geográficas do planeta.

No capítulo 1 estão apresentados os resultados da determinação de diversidade, riqueza, distribuição, padrões biogeográficos e similaridade genética de genes ARHD (*aromatic ring-hydroxylating dioxygenases*) obtidos do metagenoma de manguezal impactado com petróleo do presente estudo e de outros ambientes investigados. No **Capítulo 2** estão apresentados os resultados da análise dos dados de sequências derivados do sequenciamento da biblioteca metagenômica construída a partir de DNA do sedimento de manguezal impactado por petróleo (Bertioga) para levantamento da diversidade de genes envolvidos na degradação de hidrocarbonetos aromáticos, através de análises *in silico*. Foram mostrados também a triagem dos clones da biblioteca metagenômica para atividade de degradação de hidrocarbonetos aromáticos através de ensaios funcionais com MTT; o sequenciamento dos clones positivos para caracterização estrutural de genes envolvidos direta e indiretamente no metabolismo de compostos aromáticos; e a avaliação do potencial de clones positivos para degradação de hidrocarbonetos aromáticos através de cromatografia gasosa acoplada à espectrometria de massas (CG-EM). Finalmente, no **Capítulo 3** estão apresentados os resultados parciais das etapas de identificação, isolamento e clonagem dos genes individuais para prováveis dioxigenases, em vetor de expressão (pET), para purificação e caracterização molecular de proteínas envolvidas na biodegradação de hidrocarbonetos aromáticos. Ao final, está apresentada uma **Discussão Geral** integrando os resultados obtidos nos três capítulos seguida das **Conclusões Gerais** e das **Perspectivas Futuras**.

CAPÍTULO I

Diversity of aromatic hydroxylating dioxygenase genes in mangrove microbiome and their biogeographic patterns across global sites

Sanderson Tarciso Pereira de Sousa[1], Lucélia Cabral[1], Gileno Vieira Lacerda Júnior[1], Valéria Maia de Oliveira[1]

[1]Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas (UNICAMP), Campinas, São Paulo, Brasil.

Artigo publicado na revista MicrobiologyOpen

Received: 19 December 2016 | Revised: 16 March 2017 | Accepted: 23 March 2017

DOI: 10.1002/mbo3.490

ORIGINAL RESEARCH

WILEY MicrobiologyOpen Open Access

Diversity of aromatic hydroxylating dioxygenase genes in mangrove microbiome and their biogeographic patterns across global sites

Sanderson T. P. de Sousa[1,2] | Lucélia Cabral[1] | Gileno Vieira Lacerda Júnior[1,2] | Valéria M. Oliveira[1]

[1]Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas (UNICAMP), Paulínia, São Paulo, Brazil

[2]Institute of Biology, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil

Correspondence
Sanderson T. P. de Sousa, Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas (UNICAMP), Paulínia, São Paulo, Brazil.
Email: sousa.tarciso@gmail.com

Funding information
Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: Processes no. 2011/50809-5; 2013/22555-4

Abstract

Aromatic hydrocarbons (AH), such as polycyclic aromatic hydrocarbons, are compounds largely found in nature. Aromatic-ring-hydroxylating dioxygenases (ARHD) are proteins involved in AH degradation pathways. We used ARHD functional genes from an oil-impacted mangrove area and compared their diversity with other sites around the world to understand the ARHD biogeographic distribution patterns. For this, a comprehensive database was established with 166 operational protein families (OPFs) from 1,758 gene sequences obtained from 15 different sites worldwide, of which twelve are already published studies and three are unpublished. Based on a deduced ARHD peptide sequences consensus phylogeny, we examined trends and divergences in the sequence phylogenetic clustering from the different sites. The taxonomic affiliation of the OPF revealed that *Pseudomonas*, *Streptomyces*, *Variovorax*, *Bordetella* and *Rhodococcus* were the five most abundant genera, considering all sites. The functional diversity analysis showed the enzymatic prevalence of benzene 1,2-dioxygenase, 3-phenylpropionate dioxygenase and naphthalene 1,2-dioxygenase, in addition to 10.98% of undefined category ARHDs. The ARHD gene correlation analysis among different sites was essentially important to gain insights on spatial distribution patterns, genetic congruence and ecological coherence of the bacterial groups found. This work revealed the genetic potential from the mangrove sediment for AH biodegradation and a considerable evolutionary proximity among the dioxygenase OPFs found in Antarctica and South America sites, in addition to high level of endemism in each continental region.

KEYWORDS
dioxygenases, Gene library, Microbial biogeography, Phylogeny

1 | INTRODUCTION

Mangrove is a particular ecosystem located in the intertidal zone of marine coastal environments and estuarine margins. The importance of this ecosystem lies in its great biological productivity, serving as shelter for several species of fish, crustaceans, mollusks, birds, reptiles, mammals, and microorganisms. However, mangroves around the world have suffered due to anthropogenic activities and consequent

contamination with heavy metals, pesticides, polychlorinated biphenyls (PCBs), and other industrial pollutants, including oil spill (Sandilyan & Kathiresan, 2014). PAHs are compounds largely found in the environment (e.g., fossil fuels and plants), which are significantly toxic to the organisms, due to its lipophilic, mutagenic and carcinogenic character (Rengarajan et al., 2015). The PHAs are molecules which contain two or more benzene rings, with low solubility in water, high boiling and melting points, in addition to the highly recalcitrant and bioaccumulative properties (Haritash & Kaushik, 2009).

Aromatic-ring-hydroxylating dioxygenases (ARHD) are multicomponent proteins responsible for the first step of Polycyclic Aromatic Hydrocarbons (PAHs) degradation (Lu, Zhang, & Fang, 2011). They consist of a catalytic component with a terminal ring-hydroxylating portion linked to a ferredoxin domain. The terminal portion consists of an α- and a β-subunits, of which the α-subunit is the most highly conserved (Jouanneau, Martin, Krivobok, & Willison, 2011). The ARHDs are present in environmental microbial communities and play a crucial role in the aromatic compound degradation (including phenol, naphthalene, phenanthrene, pyrene, benzopyrene, among others), especially in oil-contaminated sediments or soils (Lu et al., 2011; Mrozik & Piotrowska-Seget, 2010; Olajire & Essien, 2014; Seo, Keum, & Li, 2009). The contaminated environments exert a strong selective pressure on the microbial community, altering its taxonomic composition and defining its ARHDs gene diversity (Gutierrez et al., 2013; Liang et al., 2011; Tan et al., 2015). The description of the structural and functional diversity of ARHD genes has been important in the establishment of genetic markers for contaminated environments followed by new insights for bioremediation processes (Gutierrez et al., 2013; Liang et al., 2011; Tan et al., 2015).

Molecular methods have been employed in the studies of microbial biogeography to evaluate taxonomic diversity. Microbial taxa are defined by genetic variation of any genomic locus grouped as operational taxonomic units (OTUs) (Koeppel & Wu, 2013). In recent decades, the development of cultivation-independent techniques associated with next-generation sequencing (NGS) have stimulated the investigation of the microbial diversity involved in the PAHs degradation (Flocco, Gomes, Mac Cormack, & Smalla, 2009; Gomes et al., 2007; Jurelevicius et al., 2012; Ma et al., 2015; Wu et al., 2014). In addition, biogeography studies has helped to elucidate the evolutionary relationship (governed by selection, drift, dispersal, and mutation processes) of living creatures through genetic proximity between samples from different environments, and thus understand their distribution patterns in space and over time (Hanson, Fuhrman, Horner-Devine, & Martiny, 2012). Therefore, the microbial diversity can be evaluated under seasonality effect, unusual natural changes or anthropogenic impact in the ecosystems (Lynch & Neufeld, 2015).

The scarcity of studies on evolutionary relationships and distribution patterns of ARHD genes around the world motivated the development of this work. Since there is a reasonable number of published studies that generated amplicons of ARHD genes, we decided to compile all of this information to try to understand the genetic relatedness among these sequences and their geographic distribution across the globe. For this purpose, sediment samples from an oil-impacted mangrove site was analyzed through the construction and sequencing of one α-ARHD gene library. The diversity of ARHD genes involved in hydrocarbon biodegradation pathway in the mangrove sediment was determined and compared to ARHD gene sequences derived from other studies, in an attempt to elucidate patterns of biogeographic distribution of such enzymes at a global scale.

2 | MATERIAL AND METHODS

2.1 | Sampling

Sediment samples were collected in a mangrove area affected by an oil spill in the 1980s, located in the Bertioga city, São Paulo State (23°53′49′′ S, 46°12′28′′W) (Andreote et al., 2012). Samples were collected perpendicularly in the mangrove transect (approximately 300 m in total), from three sites separated from each other by at least 30 m as described previously (Andreote et al., 2012). Sampling was performed in triplicate using sterile 50 ml-polypropylene tubes, which were completely filled with sediment to prevent the entry of air, cooled, and transported to the laboratory for further processing.

2.2 | Community DNA extraction

The metagenomic DNA from 0.3 g of oil-impacted mangrove sediment was extracted in triplicate with Isolation PowerSoil® kit (Mobio Labs, Inc. Solana Beach, USA), according to the manufacturer's instructions. DNA quantity and quality were determined by agarose gel electrophoresis (1% w/v) and Nanodrop (Thermo Scientific). After the evaluation, DNA samples were stored at −20°C for subsequent analyzes.

2.3 | PCR amplification of the ARHD genes

The mangrove metagenomic DNA was used as template for ARHD gene PCR-based amplification using the degenerate primers: ARHDf (TTY RYI TGY AII TAY CAY GGI TGG G) and ARHDr (AAI TKY TCI GCI GSI RMY TTC CA) (Bellicanta, 2005). These primers were designed to flank a highly conserved region of the alpha subunit from ARHD gene, with an expected amplicon size ranging between 300 and 329 bp. The PCR reaction was prepared to a final volume of 50 µl containing 1X *Taq* buffer (Invitrogen), 1.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTP mix (GE Healthcare), 1.2 µmol/L of each primer, 1 U Platinum *Taq* DNA polymerase (Invitrogen) and 2 µl (12 ng/µl) of eDNA. The reaction was conducted in an Eppendorf Mastercycler Gradient (Eppendorf Scientific, New York, USA) and the program consisted of an initial denaturation at 97°C for 3 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 57°C for 1 min and extension at 72°C for 1 min, and a final extension step at 72°C for 5 min. PCR products were separated on 1% agarose gel (Fisher Scientific, MA) in 1X TAE buffer, using the molecular weight marker 1 Kb Plus DNA Ladder (Invitrogen) for size estimation and observed under UV light using a ImageQuant LAS 4000 system (GE Healthcare Life Sciences).

2.4 | Construction of the ARHD gene library

PCR products were excised from the gel and subsequently purified using Illustra GFX PCR DNA Purification Kit (GE Healthcare Life Sciences). The amplicons were inserted into the cloning vector pTZ57R/T and transformed into *Escherichia coli* JM 109 cells using InsTAclone PCR Cloning Kit (Thermo Scientific), according to the manufacturer's protocols. Transformed cells were plated onto LB agar containing 50 μg/ml of ampicillin, 40 μl of 5-bromo-4-chloro-3-indo lyl-β-D-galactopyranoside [X-Gal (20 mg/ml)] and 40 μl of isopropyl β-D-thiogalactopiranoside [IPTG (100 mmol/L)], and incubated at 37°C for 18 hr. Positive (white) colonies were transferred to a new LB agar plate containing the same concentrations of ampicillin, X-Gal and IPTG, and incubated for 20 hr at 37°C in order to confirm the success of the ligation/transformation procedure. White colonies were transferred to 96-deep well plates with 1 ml of LB broth medium, following incubation at 37°C for 18 hr. After that, aliquots of cell growth were transferred to 96 well-ELISA plate and added of 10% glycerol for storage at −80°C.

2.5 | Sequencing of the ARHD gene library

PCR amplification of the ARHD genes from the plasmid clones was performed in 96-well PCR plates in a final volume of 25 μl using the primers M13F (5′ CGC CAG GGT TTT CCC AGT CAC GAC 3′) and M13R (5′ TTT CAC ACA GGA AAC AGC TAT GAC 3′). PCR reaction contained 1X *Taq* buffer (Invitrogen), 1.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTP mix (GE Healthcare), 0.4 μmol/L of each primer, 2 U Platinum *Taq* DNA polymerase (Invitrogen) and 1 μl (10 ng/μl) of plasmid DNA. The reaction was conducted in an Eppendorf Mastercycler Gradient (Eppendorf Scientific, New York, USA) and the amplification program consisted of an initial denaturation at 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 20 s and extension at 72°C for 90 s, and a final extension step at 72°C for 5 min. The PCR products were purified using the Illustra GFX 96 PCR Purification Kit (GE Healthcare Life Sciences) and checked on 1% agarose gel electrophoresis (Fisher Scientific, MA). The purified PCR products were then used as template for sequencing reaction using Big Dye kit (Life Technologies) and the primers M13R, according to manufacturer's guidelines. The sequencing of ARHD gene library was performed using the ABI 3500 XL platform (Applied Biosystems® 3500 XL).

2.6 | Bioinformatic analysis of ARHD gene sequences

ARHD gene sequences obtained were up-loaded in Phred software (Ewing & Green, 1998), for the evaluation of the quality of the base calls and sequences with score <20 were removed. The filtered sequences were edited using the Bioedit software (Hall, 1999) to remove the remaining primer and vector sequences. After treatment, the sequences were classified using Blastx tool (Altschul et al., 1990) from National Center for Biotechnology Information (NCBI) against "Reference Protein" database (Refseq_protein). The best hits (returned sequence with best identity) were considered as reference for taxonomic classification.

2.7 | Phylogenetic analysis

Phylogenetic analyses were performed using the following sequence datasets: (1) ARHD gene sequences generated in this study, and (2) Homologous sequences of public domain obtained from other environments around the world. The following public sequence datasets were used: Pearl River in China (Wu et al., 2014); oil reservoir in Brazil (Silva, Verde, Santos Neto, & Oliveira, 2013); rhizosphere and bulk soil samples from Antarctica (Flocco et al., 2009) Antarctic ice (Kuhn & Pellizari, unpublished); microbial mats from France (Bordenave et al., 2008); coastal marine sediments of Patagonia (Lozada et al., 2008); biphenyl-contaminated river sediment from United States (Sul et al., 2009); anthropogenic dark earth from Brazilian Amazon (Germano et al., 2012); Pine root zone from Czech Republic (Leigh et al., 2007); soils from King George Bay, Antarctic (Jurelevicius et al., 2012); PAH-polluted soil from France (Cébron et al., 2011); and soils and sediments from Antarctica (Muangchinda et al., 2015). The main goal of this step was to conduct a biogeographical analysis using a conserved region of the ARHD gene as marker. Sequence alignment was performed using the ClustalW tool (Larkin et al., 2007) in Bioedit v.7.2.5. The aligned sequences were analyzed using Mothur v.1.33.3 (Schloss et al., 2009) to determine the OPFs (Operational Protein Families) at the evolutionary distances (D) 0.03, 0.05, 0.10 and 0.20. The EMBOSS Transeq software was used to convert the OPF DNA sequences to amino acid sequences in the correct reading frame. The best OPF representatives (D = 0.05) were submitted in MEGA v5.0 program (Tamura, Dudley, Nei, & Kumar, 2007) for the construction of the phylogenetic trees. The phylogenetic trees were constructed using Neighbor-joining method and Kimura 2-parameter substitution model (Kimura, 1980), with bootstrap values calculated from 1,000 replicate runs.

2.8 | Statistical analysis

ARHDs richness and diversity were analyzed in Mothur v.1.33.3 and were calculated based on the sequence datasets using Shannon and Simpson diversity indexes, and Chao1 and ACE richness estimators. In addition, rarefaction curves were constructed to assess the observed richness (D = 0.05). The distribution of ARHD OPFs (D = 0.05), considering the sequences from all sites and based on geographic origin, was compared and displayed in a Venn diagram (Fauth et al., 1996). Principal Coordinates Analysis (PCoA) of Bray–Curtis distances was conducted in Past3 software version 3.14, and assessed the genetic correlation among the sites.

3 | RESULTS

3.1 | A brief description of the analyzed sites

The biogeographic distribution patterns of the ARHD genes worldwide were obtained through the combined analysis of twelve datasets

resulting from already published studies and two unpublished, in addition to the dataset from this study. Of all studies, one was carried out in Asia (River estuary, China), one in North America (River sediment, USA), three in South America (petroleum reservoirs, Brazil; coastal marine sediments, Argentina; dark earth of Amazonia, Brazil), three in Europe (lagoon, south-east France; pine root zone, Czech Republic; PAH-contaminated soil, France) and six in Antarctica (Figure 1) (Table 1). All the sequences derived from these studies are available for public access in the Genbank database and were downloaded to proceed with the bioinformatics analyses performed in this study.

3.2 | Description of ARHD gene library

The gene library assembled in this study yielded 129 high quality inserts for ARHD genes (approximately 329 bp in length) from the 288 sequenced clones (44.8%). The functional classification was confirmed by BLASTx and revealed that all sequences were more closely related with the benzene 1,2-dioxygenase enzyme retrieved from *Pseudomonas* (Figure 2). The average identity between the sequences was 77%, with values ranging from 45% to 94%. Therefore, the majority of OPFs represent new types of ARHD gene, emphasizing their singularity and ecological importance to the mangrove. The search for functional domains revealed the presence of the dioxygenase alpha subunit containing the "Rieske" conserved domain in all sequences of interest.

3.3 | Phylogenetic analysis of ARHD genes

Eleven OPFs were identified among the 129 gene sequences determined in this study, considering 0.05 dissimilarity level as cutoff. The deduced ARHD peptide sequences were clustered into five distinct phylogenetic groups (Figure 3). This phylogenetic arrangement showed an unknown catabolic diversity in the mangrove sediment represented by new putative gene sequences. The OPFs 01 e 02 were the most abundant ones, accounting for 78% and 15% of all analyzed clones, respectively. These two OPFs were grouped in different clusters, the OPF01 presented 89% sequence similarity with a benzene 1,2-dioxygenase of *Pseudomonas*, while the OPF02 showed 78% sequence similarity with the same peptide sequence. The remaining OPFs (03, 04, 05, 06, 07, 09, 10, and 11) were represented by one single clone, accounting for 7% of the total sequences. OPF07 showed a greater relatedness to the OPF01, and they were both recovered into a monophyletic cluster containing two reference sequences obtained from Genbank, one from *Pseudomonas* sp. and other from *Bordetella petrii*. OPFs 02, 03, 04, 05, 06, 09, 10, and 11 formed distinct clusters more distantly related to the reference sequences, with special emphasis to the OPF10, which was recovered in a completely separate branch distantly related to all other groups.

3.4 | Biogeographic patterns of ARHD genes worldwide

Sequence analysis in the Mothur software allowed the identification of the OPF representatives for each site analyzed, totalizing 166 OPFs from 1,758 gene sequences (Table S1). The taxonomic assignment of all OPFs found, considering the five most abundant genera, revealed that 52% of all sequences matched to the genus *Pseudomonas*, 12% to *Streptomyces*, 5.7% to *Variovorax*, 5.4% to *Bordetella* and 4.8% to *Rhodococcus* (Figure 4a). The taxonomic assignment and relative abundance of all sequences analyzed are detailed in Supplementary Table 1. Functional annotation of OPF sequences against the NCBI RefSeq database revealed seven distinct functions. The biogeographic phylogenetic tree was divided into eleven groups considering the OPFs arrangement and their gene function (Figure 4b). Relative abundances were considerably different among the distinct functions, as follows:

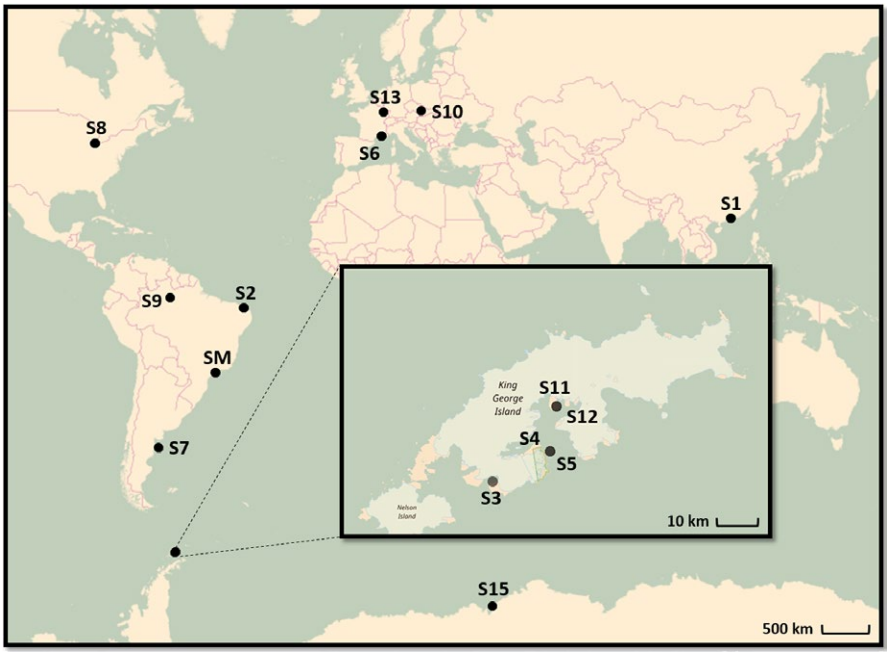


FIGURE 1 Sampling sites around the world

TABLE 1 Worldwide sites analyzed in this study

Code	Sites	Reference
SM	Mangrove Sediment, Bertioga city, Brazil	This study
S1	Sediments from the Pearl River estuary, China	Wu et al. (2014)
S2	Oil samples (naturally mixed with formation water) of Potiguar Basin (Northeast, Brazil)	Lima Verde et al. 2013
S3	Soil from Jubany station, King George island, Antarctic	Flocco et al. (2009)
S4	Botany Point, King George island, Antarctica	unpublished
S5	Brazilian station, King George island, Antarctic	unpublished
S6	Pristine and oil contaminated microbial mats, locations in south-east France	Bordenave et al. (2008)
S7	Coastal marine sediments of Patagonia	Lozada et al. (2008)
S8	Polychlorinated Biphenyl-Contaminated River Raisin Sediment, river in southeastern Michigan, United States	Sul et al. (2009)
S9	Anthropogenic dark earth of Amazonia	Germano et al. (2012)
S10	Pine root zone contaminated with polychlorinated biphenyls (PCBs), Czech Republic	Leigh et al. (2007)
S11	Different soils from King George Bay, Antarctic Peninsula (sA_sB_sC) - Diesel oil-contaminated soils	Jurelevicius et al. (2012)
S12	Different soils from King George Bay, Antarctic Peninsula (rookery, Ipanema, yellow soil) - Pristine soils	Jurelevicius et al. (2012)
S13	PAH-polluted soil, Neuves-Maisons, France	Cébron et al. 2011
S15	Antarctic soils and sediments around Syowa Station, East Ongul Island, Antarctica	Muangchinda et al. 2015

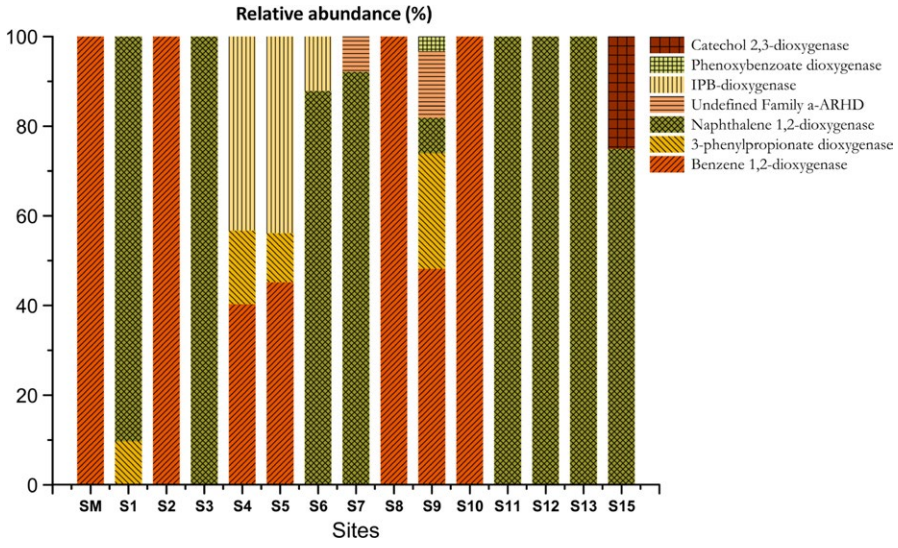


FIGURE 2 Relative abundance of ARHD sequences in each site based on the functional classification against the NCBI RefSeq database by BLAST

41.47% of all the sequences matched to benzene 1,2-dioxygenase, 29.95% to 3-phenylpropionate dioxygenase, 13.82% to naphthalene 1,2-dioxygenase, 10.98% to an undefined ARHD family, 3.87% to IPB-dioxygenase, 0.85% to phenoxybenzoate dioxygenase and 0.06% to catechol 2,3-dioxygenase (Figure 4c). The sequences assigned to the benzene 1,2-dioxygenase were distributed in the sites SM, S2, S4, S5, S8, S9, and S10, representing all continents except Asia (Figure 2 and 4b). Despite being the second most abundant function in general, 3-phenylpropionate dioxygenase was distributed only in sites S1, S4, S5, and S9, representing Asia, Antarctica and South America (Figure 2). The sequences assigned to the naphthalene 1,2-dioxygenase were distributed in more than half of the analyzed sites, representing the Asia (S1), South America (S7 and S9) Antarctica (S3, S11, S12, and S15) and Europe (S6 and S13) (Figure 2). The sequences related to the undefined ARHD Family were distributed in two sites in South America (S7 and S9), while the sequences assigned to the IPB-dioxygenase were distributed in sites located in Antarctica (S4 and S5) and Europe (S6) (Figure 2). Finally, the sequences corresponding to the phenoxybenzoate dioxygenase and catechol 2,3-dioxygenase were distributed only in one site in South America (S9) and Antarctica (S15), respectively (Figure 2). The sites S4, S5 (Antarctica) and S9 (South America) showed greater functional diversity (Figure 2).

The majority of OPFs in each site (S4, S5, S9, S3, S7, S11, S12, S1 e SM) were clustered closely together, suggesting an overall level

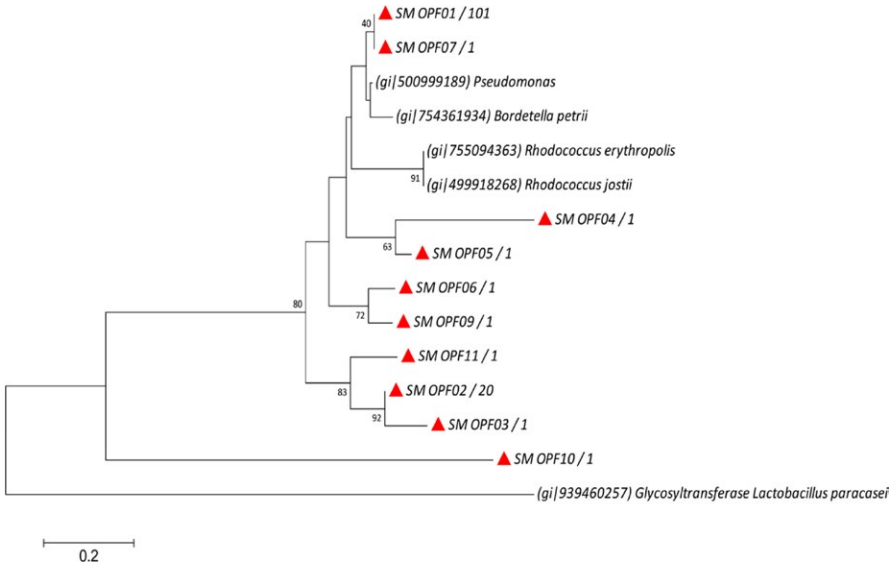


FIGURE 3 Phylogenetic tree representing the relationships among the OPF representatives for ARHD genes recovered from mangrove sediment and the most related hits from the GenBank database. Bootstrap values were obtained from 1,000 replicates and only values above 40% are shown. The red triangle represents the observed OTUs. The number of sequences in each OPF is shown after the OPF name. The gray scale bar represents 20% sequence divergence. The sequence of the glycosyltransferase from Lactobacillus paracasei was used as an outgroup

of endemism among the distinct OPFs (Figure 4b). The OPFs 05, 02, 03, 11, 01 e 07 determined in this study were grouped with OPFs of the sites S9 (Amazonia, OPFs 42, 10 and 65), S2 (Northeast, Brazil, OPF 01), S4 (Botany Point, Antarctica OPF05) and S5 (Brazilian station, Antarctica, OPF05). The phylogenetic tree showed closer evolutionary relatedness among the OPFs of the sites S4, S5, SM, S2, S9, and S8, representing Antarctica, South America and North America (Figure 4b). Curiously, the OPFs of the site S10 (Europe) also showed evolutionary relatedness to the OPFs of Antarctica and South America sites (Figure 4b).

Of the five OPFs identified among the eight sequences recovered from the S15 site (Antarctic soils), only two were annotated as putative dioxygenases. The OPF05 showed 100% of sequence similarity to a catechol 2,3-dioxygenase obtained from *Sphingobium yanoikuyae*. Some OPFs of the sites S4 (Antarctica, OPFs 13, 12, 14, 07, and 11), S5 (Antarctica, OPFs 13, 11, 06, and 18) and S9 (Amazonia, OPF 44) were recovered in a tight cluster, except for OPFs 44 and 13, and classified as "undefined ARHD family" because of the low average percentage of identity (<60%) with their closest bacterial hits (Figure 4b). Similarly, a layer of the tree with eighteen OPFs (52% average identity) from the site S9 (Amazonia) and other (52% average identity) containing 4, 2 and 1 OPFs from the sites S9 (Amazonia), S1 (China) and SM (Mangrove sediment), respectively, were classified as "undefined ARHD family" and clustered separately from the other ARHDs (Figure 4b). These OPFs encompassed a large number of sequences, apparently endemic to the Amazonia site.

3.5 | ARHD gene diversity worldwide

Genetic diversity of the ARHD genes was analyzed at 0.02 and 0.05 distance (D) cutoff (Table 2). However, based on the clusters observed from an initial phylogenetic tree performed with all nucleotide sequences and BLASTP analysis to identify similarities, the cutoff value of 0.05 distance for OPF definition was chosen for subsequent analyses. At 5% distance, 93–100% coverage values were observed for the ARHD gene diversity in all sites examined, except for sites S10 (Czech Republic) and S13 (France), which showed coverage lower than 60% (Table 2). Shannon index calculations showed that ARHD diversity in the mangrove soil under study was relatively higher than in the sites S1 (China), S2 (Northeast, Brazil) and S12 (Antarctica), although lower than the other sites (Table 2). The Simpson index measures the probability that two randomly sorted sequences within a sample will be of the same species (Simpson, 1949). The observed value for the Simpson index (62%) also demonstrated that the mangrove sediment (SM) has a relatively higher diversity (lower probability) than the sites S1, S2, and S12. The Chao 1 estimator, that measures the "species" (in this case, OPFs) richness based on the distribution of "singletons" and "doubletons" (Chao, 1984), revealed that the richness of ARHD genes in the mangrove sediment was higher than in all other sites, except for site S9 (Amazonia, Table 2). Indeed, almost 82% of OPFs determined in this study were represented by a single clone.

The rarefaction curve expressed the observed number (richness) of OPFs for each site as a function of the total sequences recovered from each site (Sanders, 1968). The complete coverage of dataset is

FIGURE 4 Taxonomic (a) and functional (c) abundance of the total ARHD gene sequences of all sites under study. Phylogenetic analysis (b) of ARHD sequences representing 166 OPF among 1758 ARHD sequences from 15 sites around the world. Bootstrap values were derived from 1,000 replicates and black scale bar represents 1% sequence divergence. Background colors represents the functional categories of ARHD gene sequences based on the classification against GenBank database. The values in double quotes represent the averages of similarity of the OPFs with their closest hits, from the GenBank database, in each functional group of the tree. Total numbers of OPFs are shown in parentheses. Total numbers of gene sequences are shown in brackets. The sequence of the glycosyltransferase from Lactobacillus paracasei was used as an outgroup

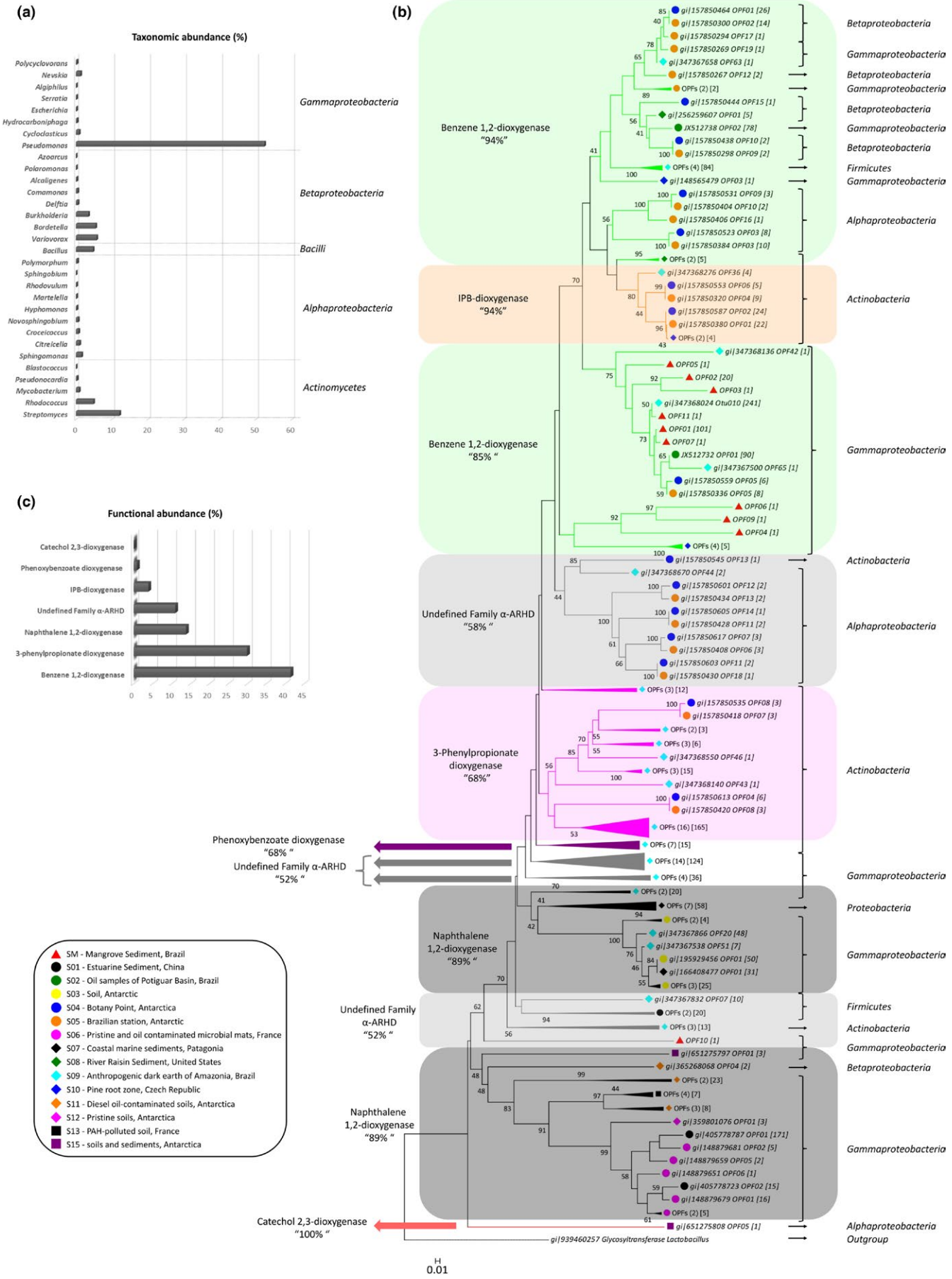

(a)
Taxonomic abundance (%)
Polycyclovorans
Nevskia
Algiphilus
Serratia
Escherichia
Hydrocarboniphaga
Cycloclasticus
Pseudomonas
Azoarcus
Polaromonas
Alcaligenes
Comamonas
Delftia
Burkholderia
Bordetella
Variovorax
Bacillus
Polymorphum
Sphingobium
Rhodovulum
Martelella
Hyphomonas
Novosphingobium
Croceicoccus
Citreicella
Sphingomonas
Blastococcus
Pseudonocardia
Mycobacterium
Rhodococcus
Streptomyces
0 10 20 30 40 50 60
Gammaproteobacteria
Betaproteobacteria
Bacilli
Alphaproteobacteria
Actinomycetes
(c)
Functional abundance (%)
Catechol 2,3-dioxygenase
Phenoxybenzoate dioxygenase
IPB-dioxygenase
Undefined Family α-ARHD
Naphthalene 1,2-dioxygenase
3-phenylpropionate dioxygenase
Benzene 1,2-dioxygenase
0 5 10 15 20 25 30 35 40 45
(b)
Benzene 1,2-dioxygenase "94%"
IPB-dioxygenase "94%"
Benzene 1,2-dioxygenase "85% "
Undefined Family α-ARHD "58% "
3-Phenylpropionate dioxygenase "68%"
Phenoxybenzoate dioxygenase "68% "
Undefined Family α-ARHD "52% "
Naphthalene 1,2-dioxygenase "89% "
Undefined Family α-ARHD "52% "
Naphthalene 1,2-dioxygenase "89% "
Catechol 2,3-dioxygenase "100% "
gi|157850464 OPF01 [26]
gi|157850300 OPF02 [14]
gi|157850294 OPF17 [1]
gi|157850269 OPF19 [1]
gi|347367658 OPF63 [1]
gi|157850267 OPF12 [2]
OPFs (2) [2]
gi|157850444 OPF15 [1]
gi|256259607 OPF01 [5]
JX512738 OPF02 [78]
gi|157850438 OPF10 [2]
gi|157850298 OPF09 [2]
OPFs (4) [84]
gi|148565479 OPF03 [1]
gi|157850531 OPF09 [3]
gi|157850404 OPF10 [2]
gi|157850406 OPF16 [1]
gi|157850523 OPF03 [8]
gi|157850384 OPF03 [10]
OPFs (2) [5]
gi|347368276 OPF36 [4]
gi|157850553 OPF06 [5]
gi|157850320 OPF04 [9]
gi|157850587 OPF02 [24]
gi|157850380 OPF01 [22]
OPFs (2) [4]
gi|347368136 OPF42 [1]
OPF05 [1]
OPF02 [20]
OPF03 [1]
gi|347368024 Otu010 [241]
OPF11 [1]
OPF01 [101]
OPF07 [1]
JX512732 OPF01 [90]
gi|347367500 OPF65 [1]
gi|157850559 OPF05 [6]
gi|157850336 OPF05 [8]
OPF06 [1]
OPF09 [1]
OPF04 [1]
OPFs (4) [5]
gi|157850545 OPF13 [1]
gi|347368670 OPF44 [2]
gi|157850601 OPF12 [2]
gi|157850434 OPF13 [2]
gi|157850605 OPF14 [1]
gi|157850428 OPF11 [2]
gi|157850617 OPF07 [3]
gi|157850408 OPF06 [3]
gi|157850603 OPF11 [2]
gi|157850430 OPF18 [1]
OPFs (3) [12]
gi|157850535 OPF08 [3]
gi|157850418 OPF07 [3]
OPFs (2) [3]
OPFs (3) [6]
gi|347368550 OPF46 [1]
OPFs (3) [15]
gi|347368140 OPF43 [1]
gi|157850613 OPF04 [6]
gi|157850420 OPF08 [3]
OPFs (16) [165]
OPFs (7) [15]
OPFs (14) [124]
OPFs (4) [36]
OPFs (2) [20]
OPFs (7) [58]
OPFs (2) [4]
gi|347367866 OPF20 [48]
gi|347367538 OPF51 [7]
gi|195929456 OPF01 [50]
gi|166408477 OPF01 [31]
OPFs (3) [25]
gi|347367832 OPF07 [10]
OPFs (2) [20]
OPFs (3) [13]
OPF10 [1]
gi|651275797 OPF01 [3]
gi|365268068 OPF04 [2]
OPFs (2) [23]
OPFs (4) [7]
OPFs (3) [8]
gi|359801076 OPF01 [3]
gi|405778787 OPF01 [171]
gi|148879681 OPF02 [5]
gi|148879659 OPF05 [2]
gi|148879651 OPF06 [1]
gi|405778723 OPF02 [15]
gi|148879679 OPF01 [16]
OPFs (2) [5]
gi|651275808 OPF05 [1]
gi|939460257 Glycosyltransferase Lactobacillus
Betaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Firmicutes
Gammaproteobacteria
Alphaproteobacteria
Actinobacteria
Gammaproteobacteria
Actinobacteria
Alphaproteobacteria
Actinobacteria
Gammaproteobacteria
Proteobacteria
Gammaproteobacteria
Firmicutes
Actinobacteria
Gammaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Alphaproteobacteria
Outgroup
SM - Mangrove Sediment, Brazil
S01 - Estuarine Sediment, China
S02 - Oil samples of Potiguar Basin, Brazil
S03 - Soil, Antarctic
S04 - Botany Point, Antarctica
S05 - Brazilian station, Antarctic
S06 - Pristine and oil contaminated microbial mats, France
S07 - Coastal marine sediments, Patagonia
S08 - River Raisin Sediment, United States
S09 - Anthropogenic dark earth of Amazonia, Brazil
S10 - Pine root zone, Czech Republic
S11 - Diesel oil-contaminated soils, Antarctica
S12 - Pristine soils, Antarctica
S13 - PAH-polluted soil, France
S15 - soils and sediments, Antarctica
0.01

TABLE 2 Diversity and richness indices of ARHD genes in the sites under study

Sites	Cutoff	OPFs	Clone numbers	ACE	Shannon	Simpson	Chao	Coverage (%)
SM	0.02	23	129	89.85	1.70	0.35	91.00	0.87
	0.05	11	129	0.00	0.82	0.62	47.00	0.93
S1	0.02	8	206	14.90	0.71	0.70	9.50	0.99
	0.05	4	206	4.00	0.64	0.70	4.00	1.00
S2	0.02	7	168	10.69	0.92	0.45	8.50	0.98
	0.05	1	168	0.00	0.00	1.00	1.00	1.00
S3	0.02	12	79	97.01	1.57	0.28	40.00	0.90
	0.05	6	79	13.60	0.94	0.48	9.00	0.96
S4	0.02	28	93	45.73	2.76	0.09	46.20	0.85
	0.05	15	93	16.74	2.15	0.16	15.75	0.97
S5	0.02	21	88	29.13	2.51	0.11	24.00	0.92
	0.05	19	88	29.87	2.42	0.12	21.50	0.93
S6	0.02	9	29	12.56	1.58	0.31	10.50	0.86
	0.05	6	29	6.46	1.35	0.33	6.00	0.97
S7	0.02	14	89	16.53	2.16	0.16	17.00	0.96
	0.05	8	89	8.45	1.74	0.21	8.00	0.99
S8	0.02	5	10	5.56	1.56	0.13	5.00	0.90
	0.05	3	10	3.00	1.03	0.31	3.00	1.00
S9	0.02	217	810	376.45	4.26	0.05	330.19	0.87
	0.05	78	810	103.21	3.07	0.11	99.43	0.97
S10	0.02	8	10	73.05	1.97	0.07	29.00	0.30
	0.05	7	10	16.44	1.83	0.09	12.00	0.50
S11	0.02	10	33	25.75	1.35	0.44	17.00	0.79
	0.05	6	33	6.50	1.14	0.45	6.00	0.97
S12	0.02	3	3	1.00	0.00	1.00	1.00	1.00
	0.05	3	3	1.00	0.00	1.00	1.00	1.00
S13	0.02	9	12	26.79	2.09	0.06	19.50	0.42
	0.05	7	12	17.45	1.75	0.14	17.00	0.58
S15	0.02	5	8	8.69	1.49	0.17	6.50	0.63
	0.05	5	8	8.69	1.49	0.14	6.50	0.63

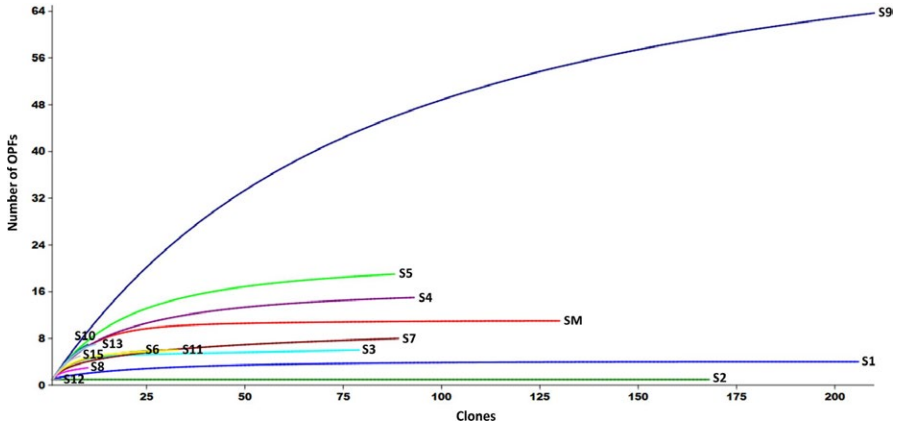


FIGURE 5 Rarefaction curves representing the richness of ARHDs (number of OPFs) as a function of the number of cloned gene sequences analyzed for each site. Information about the sites is detailed in Table 1

expected to result in a plateau-shaped curve. The results showed that the sampling effort performed in this study was sufficient to reveal all the OPFs present in the samples from sites SM (Mangrove sediment), S1 (China), S2 (Northeast Brazil), S3 (Maritime Antarctica) and S9 (Amazonia) (Figure 5). For the sites S4, S5, S6, S7 and S11, the curve clearly tends to an asymptote and additional sampling will probably

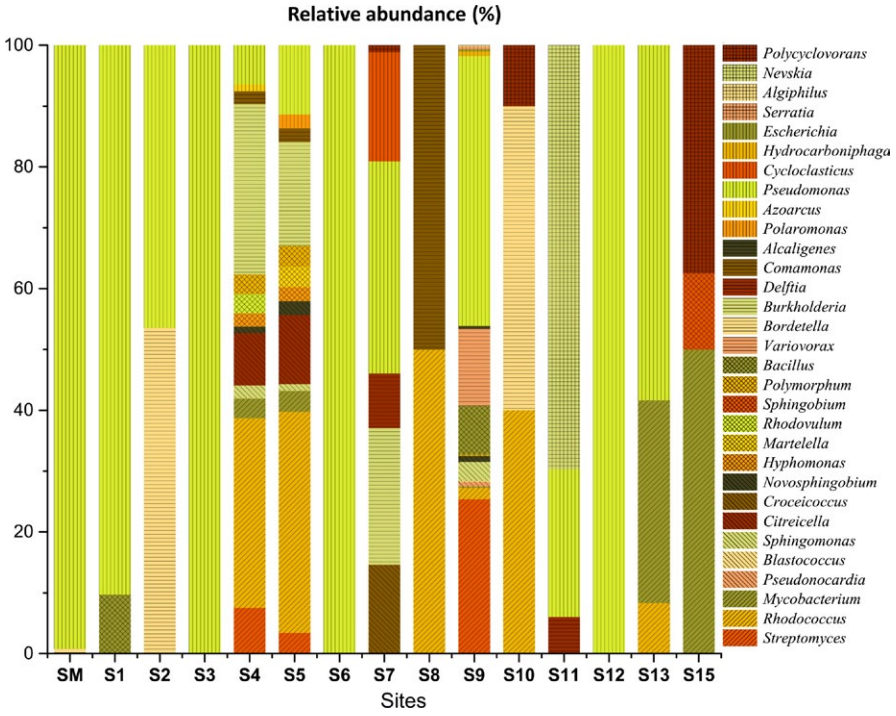


FIGURE 6 Distribution of genera in each site based on the relative abundance of ARHD sequences annotated against the NCBI RefSeq database by BLAST

reveal only a few number of additional OPFs. Finally, for the sites S8, S10, S12, S13, and S15, the weakly curvilinear plots showed that OPFs richness is far from being complete and further sampling may reveal potential untapped diversity (Figure 5).

The analysis of taxonomic diversity and abundance showed that *Pseudomonas* was the most common genus related to the ARHDs among the sites examined (Figure 6). The sites S4, S5, and S9 showed the highest richness of bacterial genera, suggesting that the higher diversity of dioxygenases observed in these sites (Table 2) is spread among distinct taxonomic groups. Curiously, there was not much taxonomic disparity between sites S4 and S5, which are geographically closely located. Nevertheless, detailed analysis of these sites revealed small differences in their taxonomic abundance. The *Rhodococcus*, *Citreicella*, and *Pseudomonas* genera were more abundant in the site S5 (Antarctica) than in the S4 (Antarctica), whereas *Burkholderia* and *Streptomyces* showed the opposite pattern. In addition, the genera *Martelella* and *Polaromonas* (S5), and *Rhodovulum* and *Azoarcus* (S4) appear only on these sites. The diversity of genera inferred from the dioxygenase sequences found in the mangrove sediment was low, although approximately half of the OPFs found presented average identity lower than 80% with their closest bacterial hits.

The Venn diagram showed the shared OPFs among sites of different continental regions. According to the data, 3, 160, 6, 85, and 23 OPFs showed endemism in Asia (site S1), South America (sites S2, S7, and S9), North America (site S8), Antarctica (sites S3, S4, S5, S11, S12, and S15) and Europe (sites S6, S10, and S13), respectively. In contrast, Europe shared one OPF with Antarctica and four OPFs with Asia, and South America shared six OPFs with Antarctica (Figure 7).

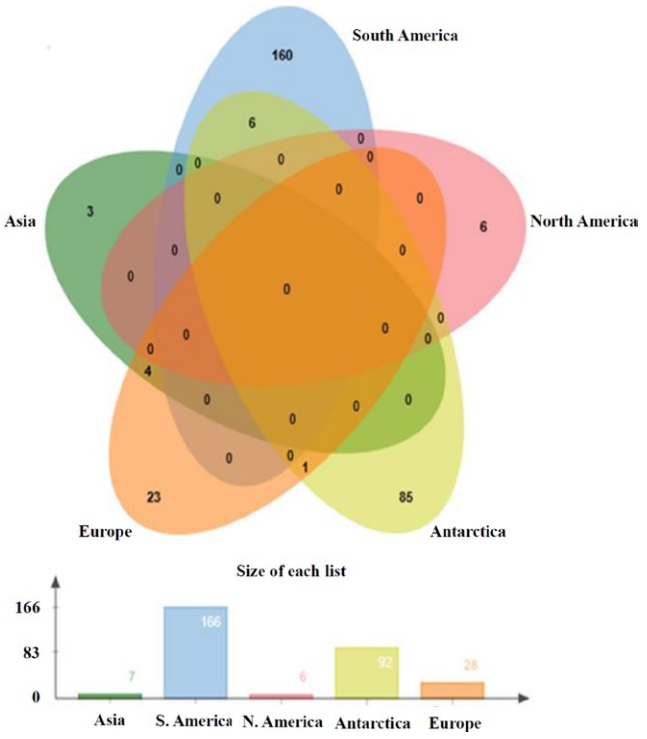


FIGURE 7 Venn diagram of shared OPFs among sites of different continental regions including Asia (site S1), North America (site S8), South America (sites S2, S7, and S9), Europe (sites S6, S10, and S13) and Antarctica (sites S3, S4, S5, S11, S12, and S15). The bar graph shows the number of OPFs found in each region. Detailed information on the sites is shown in Table 1. This analysis was run using the jvenn online platform (Bardou et al., 2014)

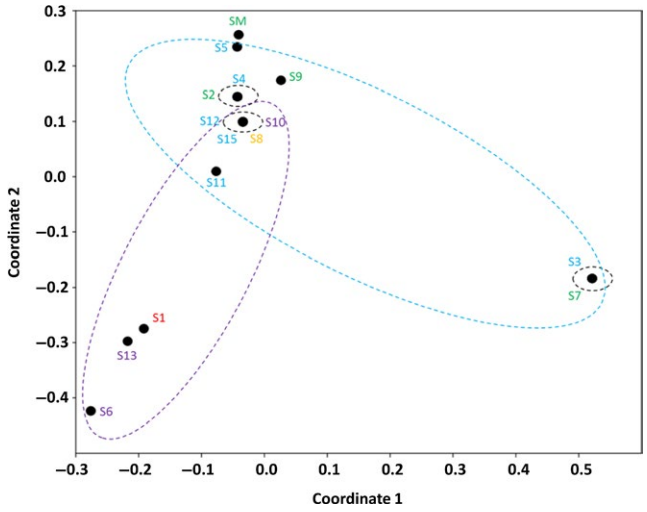


FIGURE 8 The Principal Coordinates Analysis (PCoA) based on Bray–Curtis distances showing the genetic correlation among the ARHD gene sequences worldwide. The clustering of the samples from Asia (red), North America (orange), South America (green), Europe (purple), and Antarctic (blue) are highlighted by the colored circles

3.6 | Principal coordinates analysis of the ARHD genes

Principal coordinates analysis (PCoA) of Bray–Curtis distances revealed significant correlations among sites as a function of shared OPFs. The majority of the sites were closely grouped respecting their continental origin. The sites of South America (SM, S9 e S2), Antarctica (S4, S5, S11, S12, and S15) and Europe (S6 and S13) grouped together. Although S3 and S7 sites are located in Antarctica and South America (Patagonia Coast), respectively, they showed larger discrepancy from the other sites with the same geographical location. The site S8, the only one located in North America, grouped together with the majority of the sites from South America and Antarctica. On the other hand, the site S1, the only one located in Asia, grouped together with the sites S6 and S13, both located in France (Figure 8).

4 | DISCUSSION

In this study, the biogeographic distribution patterns of ARHD functional genes obtained from 15 different sites worldwide was investigated, in addition to their diversity at the functional and taxonomic levels.

According to the functional assignment of the OPFs, benzene 1,2-dioxygenase, 3-phenylpropionate dioxygenase, naphthalene 1,2-dioxygenase, IPB-dioxygenase, phenoxybenzoate dioxygenase and catechol 2,3-dioxygenase were the most abundant enzyme sequences reported in this study, all involved in the aromatic hydrocarbon degradation. All these functions have already been grouped into dioxygenase families from the classification systems proposed previously (Batie, LaHaie, & Ballou, 1987; Gibson & Parales, 2000; Nam et al., 2001). Recently, the work of Iwai et al. (2011) has been suggested as the best study exploring the diversity and phylogenetic classification of dioxygenases derived from different environmental samples. According with their analysis, the dioxygenase genes were grouped into five subclades: PAH-GN (PAH dioxygenases from Gram-negative bacteria), PAH-GP (PAH dioxygenases from Gram-positive bacteria), T/B (toluene/biphenyl dioxygenases), OT-I (other dioxygenases I) and OT-II (other dioxygenases II). In addition, a large gene cluster was classified as benzoate dioxygenases or not functionally characterized Rieske-type dioxygenases. More recently, Meynet, Head, Werner, and Davenport (2015) published a re-evaluation of dioxygenase gene-based phylogeny that showed, on one side, subclades with closely related genes derived from microorganisms that degrade the same substrate, and, on the other side, clades encompassing dioxygenase genes derived from microorganisms able to degrade different substrates. In the present study, lineages 1, 2, and 3 represented PAH-GN, T/B and PAH-GP dioxygenases, respectively, similarly grouped in previous articles (Gibson & Parales, 2000; Iwai et al., 2011), whereas the lineage 4 showed similarity with phenylpropionate/phthalate/aromatic dioxygenases and benzoate/aromatic dioxygenases. All enzymes showed correlation with the dioxygenase groups indicated by the studies cited above. Benzene 1,2-dioxygenase and IPB-dioxygenase were showed as the closest clades in the phylogenetic tree, as suggested by Gibson and Parales (2000) and Nam et al., 2001;. Naphthalene 1,2-dioxygenase, phenoxybenzoate dioxygenase, 3-phenylpropionate dioxygenase and catechol 2,3-dioxygenase represented the groups III, I (Nam et al., 2001), clade C into lineage 2 and clade D into lineage 3 respectively, according to previous classifications (Meynet et al., 2015).

The present work revealed interesting patterns of ARHD biogeographic distribution. Integrated phylogenetic analysis of deduced peptide sequences from OPFs of each site examined revealed two aspects: (1) site-specific genetic endemism for some of the ARHD families (e.g., some benzene 1,2- dioxygenases, IPB dioxygenases, phenoxybenzoate dioxygenases and some undefined α-ARHD families), especially in Antarctic and Amazonian sites; and (2) genetic congruence among several ARHDs genes obtained from different environments, even separated by large distances. Müller et al. (2015) carried out an investigation about phylogenetic diversity and environmental distribution of *dsrAB* genes, coding for a reductase, and reported that members of the same *dsrAB* lineage were allocated in contrasting environments. On the other hand, the same work reported that there are unique lineages inhabiting environments with specific biogeochemical properties. In this study, some ARHD OPFs obtained from geographically different environments presented genetic relatedness. In some cases, such as the clustering between the sites SM (oil-spilled mangrove sediments from Brazilian southeast coast) and S5 (sediments from Admiralty Bay, King George Island), or between the sites S3 (soil from Potter Peninsula & Jubany Station, King George Island) and S7 (coastal marine sediments from Patagonia), the evolutionary relatedness among OPFs might be explained by transport of microorganisms between these sites through ocean currents. In other cases, such as the clustering between the sites S8 (Polychlorinated Biphenyl-Contaminated River

Raisin Sediment, southeastern Michigan, United States) and S10 (Pine root zone contaminated with PCBs, Czech Republic), the relatedness among ARHDs OPFs could be explained by evolutionary convergence as a function of the substrate nature (PCBs). On the other hand, clustering of the sites S1 (sediments from the Pearl River estuary, China) and S13 (PAH-polluted soil, Neuves-Maisons, France), or of the sites S2 (oil samples from Potiguar Basin, Northeast Brazil) and S4 (sediments from Botany Point, King George island, Antarctica) showed no apparent correlation, suggesting that, in some cases, environmental factors do not define the evolution of genetic lineages. Finally, there were some OPFs that clustered together according to their biogeochemical origins, demonstrating ecological coherence (OPFs from sites S5, S4, S12, S15 and S11; and OPFs from sites S13 and S6). The concept of ecological coherence was well employed in the study of Oton et al., 2015, when they discovered a correlation between physicochemical properties and genetic diversity in soil samples.

Pseudomonas, *Streptomyces*, *Variovorax*, *Bordetella*, and *Rhodococcus* were the most abundant genera inferred from ARHD sequences recovered from all sites examined. However, the different protocols of DNA extraction, primers used, Polymerase Chain Reaction (PCR), number of sequences and other experimental procedures used may have limited the access to the taxonomic diversity in sites SM, S1, S2, S3, S6, S8 e S12. On the other hand, although the inferred taxonomic diversity has been low at some sites (SM, S1, S2, and S3), the richness of OPFs determined was high, as shown in the rarefaction curves (Figure 5). Despite different primer sets have been used across all the analyzed studies, all of them targeted the same conserved gene region for ARHDs.

Several species of the genus *Pseudomonas* have been extensively reported with potential for degradation of aromatic hydrocarbons (De Lima-Morales et al., 2016). Metagenomic analysis carried out in the same site investigated in this study revealed the predominance of Gammaproteobacteria (Andreote et al., 2012), an important class of bacteria that encompasses *Pseudomonas*, known for their aromatic hydrocarbon degradation ability. Similarly, the ability of Actinobacteria members, such as *Streptomyces* and *Rhodococcus*, to metabolize hydrocarbons, including toluene, ethylbenzene, xylenes, biphenyl, polycyclic aromatic compounds and phenolic compounds, and their bioremediation potential has been well documented (Balachandran, Duraipandiyan, Balakrishna, & Ignacimuthu, 2012; Ishiyama, Vujaklija, & Davies, 2004; Martínková et al., 2009). The potential for aromatic compound degradation, based on functional and genetic evidences, is also well established for members of the Betaproteobacteria class, especially from the Burkholderiales order, such as the genera *Variovorax* and *Bordetella* (Satola, Wübbeler, & Steinbüchel, 2013). Thus, the data obtained in this study on the taxonomic affiliation of the ARHD sequences are strongly corroborated by previous findings.

The Venn diagram revealed that the South America and Antarctic sites shared microorganisms containing the same allele gene for dioxygenase in their genomes. Similar situation occurred between Asia and Europe, which shared four alleles. This finding suggests that some dispersal mechanisms (such as wind, marine currents or host organisms) may have allowed the transfer of gene content between these two regions. On the other hand, the endemism seen in the continental regions could be explained by their intrinsic environmental characteristics and/or by the long distance among them. The oceans can also play the role of geographic barrier, favoring the speciation process of microbial groups by mutations, isolation and reproduction. These long distances and geographic barriers could answer why there were no alleles shared between Asia and North America, Europe and North America, Europe and South America, Antarctica and Asia and Antarctica and North America.

The phylogenetic analysis of the ARHD OPFs was in accordance with the clustering of sites by PCoA. The grouping of the Antarctic samples could be explained by the influence of environmental factors, such as temperature and nutrient availability, which might have shaped the genetic diversity of the ARHDs in the microbial community. Previous studies have already showed that water depth, temperature, nutrient composition and salinity can shape the microbial community in nature (Cao, Hong, Li, & Gu, 2012; Oton et al., 2015). The close genetic relatedness among different environments (such as S4, S5, S11, S12, and S15) could be explained by microbial dispersal across ocean currents. According to Hanson et al. (2012), most studies show that microorganisms show limited dispersal due to their low mobility, favoring its establishment in restricted areas. On the other hand, the movement of microorganisms across space may occur by passive dispersal (e.g., via wind and air) or active dispersal (e.g., via ride in larger organisms), facilitating gene flow among different geographic regions (Martiny et al., 2006; Nemergut et al., 2013). Smith et al. (2013) revealed the dispersal of microorganisms across air masses and evidenced the presence of phyla with adaptations for atmospheric displacement. Microbial dispersion can also occur across ocean currents and thermohaline circulation (McGillicuddy et al., 2007). Hamdan et al. (2013) showed that hydrography can shape local and global microbial communities.

The phylogenetic tree showed some divergences in the distribution of the taxonomic classes inferred from the ARHD sequences. These divergences may be related to two factors: (1) low percentage of genetic similarity with their closest bacterial hits in the database; or (2) horizontal gene transfer. The lack of accuracy and representativeness of the available databases may be one of the causes for the first factor. The second factor is a natural biological phenomenon driven by selective pressures and may lead to taxonomic divergences. The study conducted by Jacob Parnell et al. (2010) showed that microbial biogeography may be affected by horizontal gene transfer, resulting in an average similarity of functional genes higher than the taxonomic average similarity evaluated based on the 16S ribosomal gene (Jacob Parnell et al., 2010). It is already known that horizontal gene transfer raises doubts about the veracity of the likely taxonomic classification inferred from the genetic elements that move among bacterial and archaeal species. In this study, a more robust analysis would be necessary to evaluate the percentage of phylogenetic congruence between 16S rRNA and ARHD genes from several bacterial genera. Oton et al. (2015) reported 85% of phylogenetic congruence between 16S rRNA and *amoA* genes of Thaumarchaeota, demonstrating low horizontal transfer rate of these genes.

In conclusion, this study contributed to the knowledge on the richness, distribution, and genetic relatedness of the ARHD genes recovered from fifteen different sites, as well as the taxonomic groups likely associated with such functional genes. A moderate genetic congruence was evidenced among ARHDs worldwide and, in some cases, biogeographic endemism patterns were revealed, presumably modeled by biotic and abiotic environmental factors. Results suggest that there was a connectivity between South America and Antarctica sites, probably caused by sea currents, air masses or host organisms. This could have facilitated the dispersion of microorganisms between these two regions, helping to explain their evolutionary history.

ACKNOWLEDGMENTS

The authors are grateful to the São Paulo Research Foundation - FAPESP for the financial support and scholarship (Processes no. 2011/50809-5; 2013/22555-4).

CONFLICTS OF INTEREST

None.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Andreote, F.D., Jiménez, D.J., Chaves, D., Dias, A. C. F., Luvizotto, D. M., Dini-Andreote, F. … de Melo, I. S. (2012). The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS ONE*, *7*, e38600.

Balachandran, C., Duraipandiyan, V., Balakrishna, K., & Ignacimuthu, S. (2012). Petroleum and polycyclic aromatic hydrocarbons (PAHs) degradation and naphthalene metabolism in Streptomyces sp. (ERI-CPDA-1) isolated from oil contaminated soil. *Bioresource Technology*, *112*, 83–90.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., & Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, *15*, 293. https://doi.org/10.1186/1471-2105-15-293

Batie, C. J., LaHaie, E., & Ballou, D. P. (1987). Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from Pseudomonas cepacia. *Journal of Biological Chemistry*, *262*, 1510–1518.

Bellicanta, G. S. 2005. Diversity of catabolic genes in sediment samples of Santos and São Vicente Estuarine Systems, SP (Doctoral dissertation) Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo.

Bordenave, S., Goñi-Urriza, M., Vilette, C., Blanchard, S., Caumette, P., & Duran, R. (2008). Diversity of ring-hydroxylating dioxygenases in pristine and oil contaminated microbial mats at genomic and transcriptomic levels. *Environmental Microbiology*, *10*, 3201–3211.

Cao, H., Hong, Y., Li, M., & Gu, J. D. (2012). Community shift of ammonia-oxidizing bacteria along an anthropogenic pollution gradient from the Pearl River Delta to the South China Sea. *Applied Microbiology and Biotechnology*, *94*, 247–259.

Cébron, A., Louvel, B., Faure, P., France-Lanord, C., Chen, Y., Murrell, J. C., & Leyval, C. (2011). Root exudates modify bacterial diversity of phenanthrene degraders in PAH-polluted soil but not phenanthrene degradation rates. *Environmental Microbiology*, *13*, 722–736.

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population Author. *Scandinavian Journal of Statistics*, *11*, 265–270.

De Lima-Morales, D., Chaves-Moreno, D., Wos-Oxley, M. L., Jáuregui, R., Vilchez-Vargas, R., & Pieper, D. H. (2016). Degradation of benzene by Pseudomonas veronii 1YdBTEX2 and 1YB2 is catalyzed by enzymes encoded in distinct catabolism gene clusters. *Applied and Environment Microbiology*, *82*, 167–173.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome Research*, *8*, 186–194.

Fauth, J. E., Bernardo, J., Camara, M., Resetarits, W. J. Jr, Buskirk, J. V., & McCollum, S. A. (1996). Simplifying the jargon of community ecology: A conceptual approach. *American Naturalist*, *147*, 282.

Flocco, C. G., Gomes, N. C. M., Mac Cormack, W., & Smalla, K. (2009). Occurrence and diversity of naphthalene dioxygenase genes in soil microbial communities from the Maritime Antarctic. *Environmental Microbiology*, *11*, 700–714.

Germano, M. G., de Souza Cannavan, F., Mendes, L. W., Lima, A. B., Teixeira, W. G., Pellizari, V. H., & Tsai, S. M. (2012). Functional diversity of bacterial genes associated with aromatic hydrocarbon degradation in anthropogenic dark earth of Amazonia. *Pesqui Agropecuária Bras*, *47*, 654–664.

Gibson, D. T., & Parales, R. E. (2000). Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Current Opinion in Biotechnology*, *11*, 236–243.

Gomes, N. C. M., Borges, L. R., Paranhos, R., Pinto, F. N., Krögerrecklenfort, E., Mendonça-Hagler, L. C. S., & Smalla, K. (2007). Diversity of ndo genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. *Applied and Environment Microbiology*, *73*, 7392–7399.

Gutierrez, T., Singleton, D. R., Berry, D., Yang, T., Aitken, M. D., & Teske, A. (2013). Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. *ISME Journal*, *7*, 2091–2104.

Hall, T. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hamdan, L. J., Coffin, R. B., Sikaroodi, M., Greinert, J., Treude, T., & Gillevet, P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *ISME Journal*, *7*, 685–696.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. H. (2012).Beyond biogeographic patterns: Processes shaping the microbial landscape.*Nature Reviews Microbiology*, *10*, 1–10.

Haritash, A. K., & Kaushik, C. P. (2009). Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. *Journal of Hazardous Materials*, *169*, 1–15.

Ishiyama, D., Vujaklija, D., & Davies, J. (2004). Novel Pathway of Salicylate Degradation by Streptomyces sp. Strain WA46. *Applied and Environment Microbiology*, *70*, 1297–1306.

Iwai, S., Johnson, T. A., Chai, B., Hashsham, S. A., & Tiedje, J. M. (2011). Comparison of the Specificity and Efficacy of Primers for Aromatic Dioxygenase Gene Analysis of Environmental Samples. *Applied and Environment Microbiology*,*77*, 3551–3557.

Jacob Parnell, J., Rompato, G., Latta IV, L. C., Pfrender, M. E., van Nostrand, J. D., He, Z. … Weimer, B. C. (2010). Functional biogeography as evidence of gene transfer in hypersaline microbial communities. *PLoS ONE*, *5*, e12919.

Jouanneau, Y., Martin, F., Krivobok, S., & Willison, J. C. (2011).Ring-hydroxylating Dioxygenases Involved in PAH Biodegradation: Structure, Function, and Biodiversity. In A.-I. Koukkou (Ed.), *Microbial Bioremediation of Non-metals: Current Research* pp 149–175. Norfolk, UK: Caister Academic Press.

Jurelevicius, D., Alvarez, V. M., Peixoto, R., Rosado, A. S., & Seldin, L. (2012). Bacterial polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases (PAH-RHD) encoding genes in different soils from King George Bay, Antarctic Peninsula. *Applied Soil Ecology*, *55*, 1–9.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111–120.

Koeppel, A. F., & Wu, M. (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Research*, *41*, 5175–5188.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H. … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*, 2947–2948.

Leigh, M. B., Pellizari, V. H., Uhlík, O., Sutka, R., Rodrigues, J., Ostrom, N. E. … Tiedje, J. M. (2007). Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME Journal*, *1*, 134–148.

Liang, Y., Van Nostrand, J. D., Deng, Y., He, Z., Wu, L., Zhang, X. … Zhou, J. (2011). Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *ISME Journal*, *5*, 403–413.

Lima Verde, L. C., e Silva, T. R., Dellagnezze, B. M., dos Santos Neto, E. V., & Oliveira, V. M. (2013). Diversity of Hydrocarbon-Related Catabolic Genes in Oil Samples from Potiguar Basin (Rn, Brazil). *Journal of Petroleum & Environmental Biotechnology*, *4*, 138.https://doi.org/10.4172/2157- 7463.1000138.

Lozada, M., Riva Mercadal, J. P., Guerrero, L. D., Di Marzio, W. D., Ferrero, M. A., & Dionisi, H. M. (2008). Novel aromatic ring-hydroxylating dioxygenase genes from coastal marine sediments of Patagonia. *BMC Microbiology*, *8*, 50.

Lu, X. Y., Zhang, T., & Fang, H. H. P. (2011). Bacteria-mediated PAH degradation in soil and sediment. *Applied Microbiology and Biotechnology*, *89*, 1357–1371.

Lynch, M. D. J., & Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, *13*, 217–229.

Ma, B., Lyu, X. F., Zha, T., Gong, J., He, Y., & Xu, J. M. (2015). Reconstructed metagenomes reveal changes of microbial functional profiling during PAHs degradation along a rice (Oryza sativa) rhizosphere gradient. *Journal of Applied Microbiology*, *118*, 890–900.

Martínková, L., Uhnáková, B., Pátek, M., Nešvera, J., & Křen, V. (2009). Biodegradation potential of the genus Rhodococcus. *Environment International*, *35*, 162–177.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L. … Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102–112.

McGillicuddy, D. J., Anderson, L. A., Bates, N. R., Bibby, T., Buesseler, K. O., Carlson, C. A. … Steinberg, D. K. (2007). Eddy/Wind Interactions Stimulate Extraordinary Mid-Ocean Plankton Blooms. *Science* (80-) *316*:1021–1026.

Meynet, P., Head, I. M., Werner, D., & Davenport, R. J. (2015). Re-evaluation of dioxygenase gene phylogeny for the development and validation of a quantitative assay for environmental aromatic hydrocarbon degraders. *FEMS Microbiology Ecology*, *91*, 1–11.

Mrozik, A., & Piotrowska-Seget, Z. (2010). Bioaugmentation as a strategy for cleaning up of soils contaminated with aromatic compounds. *Microbiological Research*, *165*, 363–375.

Muangchinda, C., Chavanich, S., Viyakarn, V., Watanabe, K., Imura, S., Vangnai, A. S., & Pinyakong, O. (2015). Abundance and diversity of functional genes involved in the degradation of aromatic hydrocarbons in Antarctic soils and sediments around Syowa Station. *Environmental Science and Pollution Research*, *22*, 4725–4735.

Müller, A. L., Kjeldsen, K. U., Rattei, T., Pester, M., & Loy, A. (2015). Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME Journal*, *9*, 1152–1165.

Nam, J. W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H., & Omori, T. (2001). New classification system for oxygenase components involved in ring-hydroxylating oxygenations. *Bioscience, Biotechnology, and Biochemistry*, *65*, 254–263.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F. … Ferrenberg, S. (2013). Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews*, *77*, 342–356.

Olajire, A. A., & Essien, J. P. (2014). Aerobic Degradation of Petroleum Components by Microbial Consortia. *Journal of Petroleum Environmental Biotechnology*, *5*, 1–22.

Oton, E. V., Quince, C., Nicol, G. W., Prosser, J. I., & Gubry-Rangin, C. (2015). Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *ISME Journal*, *10*, 85–96.

Rengarajan, T., Rajendran, P., Nandakumar, N., Lokeshkumar, B., Rajendran, P., & Nishigaki, I. (2015). Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. *Asian Pacific Journal of Tropical Biomedicine*, *5*, 182–189.

Sanders, H. L. (1968). Marine Benthic Diversity: A Comparative Study. *American Naturalist*, *102*, 243–282.

Sandilyan, S., & Kathiresan, K. (2014). Decline of mangroves - A threat of heavy metal poisoning in Asia. *Ocean and Coastal Management*, *102*, 161–168.

Satola, B., Wübbeler, J. H., & Steinbüchel, A. (2013). Metabolic characteristics of the species Variovorax paradoxus. *Applied Microbiology and Biotechnology*, *97*, 541–560.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B. … Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environment Microbiology*, *75*, 7537–7541.

Seo, J.-S., Keum, Y.-S., & Li, Q. X. (2009). Bacterial degradation of aromatic compounds. *International Journal of Environmental Research and Public Health*, *6*, 278–309.

Silva, T. R., Verde, L. C. L., Santos Neto, E. V., & Oliveira, V. M. (2013). Diversity analyses of microbial communities in petroleum samples from Brazilian oil fields. *International Biodeterioration \& Biodegradation*, *81*, 57–70.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*, 688.

Smith, D. J., Timonen, H. J., Jaffe, D. A., Griffin, D. W., Birmele, M. N., Perry, K. D. … Roberts, M. S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. *Applied and Environment Microbiology*, *79*, 1134–1139.

Sul, W. J., Park, J., Quensen, J. F., Rodrigues, J. L. M., Seliger, L., Tsoi, T. V. … Tiedje, J. M. (2009). DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Applied and Environment Microbiology*, *75*, 5501–5506.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*, 1596–1599.

Tan, B., Jane Fowler, S., Laban, N. A., Dong, X., Sensen, C. W., Foght, J., & Gieg, L. M. (2015). Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. *ISME Journal*, *9*, 2028–2045.

Wu, P., Wang, Y. S., Sun, F. L., Wu, M. L., & Peng, Y. L. (2014). Bacterial polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases in the sediments from the Pearl River estuary, China. *Applied Microbiology and Biotechnology*, *98*, 875–884.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: de Sousa STP, Cabral L, Lacerda Júnior GV, Oliveira VM. Diversity of aromatic hydroxylating dioxygenase genes in mangrove microbiome and their biogeographic patterns across global sites. *MicrobiologyOpen*. 2017;00:e490. https://doi.org/10.1002/mbo3.490

Table S1: Functional and taxonomic annotation of all OPFs.

OPF IDs	Abundance	Enzyme family	Closest bacterial hit	Ident	Accession
>SM_OPF001	101	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	97%	WP_012052601.1
>SM_OPF002	20	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	92%	WP_012052601.1
>SM_OPF003	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	74%	WP_012052601.1
>SM_OPF004	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	68%	WP_012052601.1
>SM_OPF005	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	82%	WP_012052601.1
>SM_OPF006	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	94%	WP_012052601.1
>SM_OPF007	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	89%	WP_012052601.1
>SM_OPF008	1	benzene 1,2-dioxygenase	Bordetella petrii	70%	WP_041863076.1
>SM_OPF009	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	68%	WP_012052601.1
>SM_OPF010	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	45%	WP_012052601.1
>SM_OPF011	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	64%	WP_012052601.1
>S1_OPF001	171	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	91%	WP_011154413.1
>S1_OPF002	15	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	98%	WP_011154413.1
>S1_OPF003	10	3-phenylpropionate dioxygenase	Bacillus massiliosenegalensis	58%	WP_019153061.1
>S1_OPF004	10	3-phenylpropionate dioxygenase	Bacillus aidingensis	52%	WP_026702167.1
>S2_OPF001	90	benzene 1,2-dioxygenase	Bordetella petrii	88%	WP_041863076.1
>S2_OPF002	78	benzene 1,2-dioxygenase	Pseudomonas frederiksbergensis	93%	WP_039594666.1
>S3_OPF01	50	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	100%	WP_011154413.1
>S3_OPF02	23	naphthalene dioxygenase iron sulfur protein large subunit	Pseudomonas	96%	WP_011475377.1
>S3_OPF03	3	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	90%	WP_011154413.1
>S3_OPF04	1	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	96%	WP_011117400.1
>S3_OPF05	1	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	86%	WP_011154413.1
>S3_OPF06	1	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	98%	WP_011154413.1
>S4_OPF01	26	biphenyl dioxygenase subunit alpha	Burkholderia xenovorans	100%	WP_011494299.1
>S4_OPF02	24	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	100%	WP_011133490.1
>S4_OPF03	8	benzene 1,2-dioxygenase	Citreicella sp. 357	99%	WP_009505788.1
>S4_OPF04	6	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	54%	WP_055520670.1
>S4_OPF05	6	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	100%	WP_012052601.1
>S4_OPF06	5	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	91%	WP_011133490.1
>S4_OPF07	3	benzene 1,2-dioxygenase	Rhodovulum sp. NI22	56%	WP_037209959.1
>S4_OPF08	3	3-phenylpropionate dioxygenase	Mycobacterium	81%	WP_011781569.1
>S4_OPF09	3	benzene 1,2-dioxygenase	Polymorphum gilvum	91%	WP_041375675.1
>S4_OPF10	2	benzene 1,2-dioxygenase	Comamonas testosteroni	99%	WP_043371571.1
>S4_OPF11	2	benzene 1,2-dioxygenase	Sphingomonas sp. Ant20	64%	WP_037529249.1
>S4_OPF12	2	benzene 1,2-dioxygenase	Hyphomonas chukchiensis	57%	WP_051615035.1
>S4_OPF13	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	55%	WP_055520670.1
>S4_OPF14	1	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	53%	WP_008070358.1
>S4_OPF15	1	benzene 1,2-dioxygenase	Azoarcus sp. CIB	91%	WP_050417871.1
>S5_OPF01	22	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	100%	WP_011133490.1
>S5_OPF02	14	biphenyl dioxygenase subunit alpha	Burkholderia xenovorans	100%	WP_011494299.1
>S5_OPF03	10	benzene 1,2-dioxygenase	Citreicella sp. 357	99%	WP_009505788.1
>S5_OPF04	9	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	91%	WP_011133490.1
>S5_OPF05	8	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	100%	WP_012052601.1

OPF IDs	Abundance	Enzyme family	Closest bacterial hit	Ident	Accession
>S5_OPF06	3	benzene 1,2-dioxygenase	Martelella sp. AD-3	56%	WP_036236102.1
>S5_OPF07	3	3-phenylpropionate dioxygenase	Mycobacterium	80%	WP_011781569.1
>S5_OPF08	3	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	54%	WP_055520670.1
>S5_OPF09	2	benzene 1,2-dioxygenase	Comamonas testosteroni	99%	WP_043371571.1
>S5_OPF10	2	benzene 1,2-dioxygenase	Polymorphum gilvum	91%	WP_041375675.1
>S5_OPF11	2	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	53%	WP_008070358.1
>S5_OPF12	2	benzene 1,2-dioxygenase	Polaromonas naphthalenivorans	94%	WP_011797818.1
>S5_OPF13	2	benzene 1,2-dioxygenase	Hyphomonas chukchiensis	57%	WP_051615035.1
>S5_OPF14	1	benzene 1,2-dioxygenase	Pseudomonas toyotomiensis	81%	WP_059391947.1
>S5_OPF15	1	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	85%	WP_011133490.1
>S5_OPF16	1	benzene 1,2-dioxygenase	Polymorphum gilvum	95%	WP_041375675.1
>S5_OPF17	1	biphenyl dioxygenase subunit alpha	Burkholderia xenovorans	98%	WP_011494299.1
>S5_OPF18	1	benzene 1,2-dioxygenase	Sphingomonas sp. Ant20	64%	WP_037529249.1
>S5_OPF19	1	benzene 1,2-dioxygenase	Pseudomonas aeruginosa	99%	WP_059400052.1
>S6_OPF01	16	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	100%	WP_011117400.1
>S6_OPF02	5	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	91%	WP_011117400.1
>S6_OPF03	3	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	95%	WP_011117400.1
>S6_OPF04	2	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	94%	WP_011117400.1
>S6_OPF05	2	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	91%	WP_011117400.1
>S6_OPF06	1	naphthalene dioxygenase iron sulfur protein large subunit	Pseudomonas	90%	WP_011475377.1
>S7_OPF01	31	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	100%	WP_011154413.1
>S7_OPF02	20	naphthalene 1,2-dioxygenase	Burkholderia sp. Ch1-1	99%	WP_007179244.1
>S7_OPF03	13	naphthalene 1,2-dioxygenase	Croceicoccus naphthovorans	60%	WP_047820944.1
>S7_OPF04	8	naphthalene 1,2-dioxygenase	Delftia sp. Cs1-4	69%	WP_013801305.1
>S7_OPF05	7	PAH dioxygenase iron sulfur protein large subunit	Cycloclasticus	100%	WP_016391028.1
>S7_OPF06	7	naphthalene 1,2-dioxygenase	Cycloclasticus	61%	WP_015007028.1
>S7_OPF07	2	naphthalene 1,2-dioxygenase	Cycloclasticus	61%	WP_015005786.1
>S7_OPF08	1	naphthalene 1,2-dioxygenase	Polycyclovorans algicola	96%	WP_029889175.1
>S8_OPF01	5	benzene 1,2-dioxygenase	Comamonas testosteroni	92%	WP_057092204.1
>S8_OPF02	3	benzene 1,2-dioxygenase	Rhodococcus erythropolis	82%	WP_042445407.1
>S8_OPF03	2	benzene 1,2-dioxygenase	Rhodococcus opacus	76%	WP_012687209.1
>S9_OPF001	48	benzene 1,2-dioxygenase	Bacillus	93%	WP_048680352.1
>S9_OPF002	73	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	56%	WP_021006872.1
>S9_OPF003	39	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	62%	WP_055520670.1
>S9_OPF004	34	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	64%	WP_055520670.1
>S9_OPF005	3	biphenyl 2,3-dioxygenase subunit alpha	Rhodococcus rhodnii	58%	WP_010839881.1
>S9_OPF006	13	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	94%	WP_055520670.1
>S9_OPF007	10	3-phenylpropionate dioxygenase	Bacillus massiliogorillae	55%	WP_042349570.1
>S9_OPF008	10	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	63%	WP_055520670.1
>S9_OPF009	8	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	59%	WP_055520670.1
>S9_OPF010	297	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	95%	WP_012052601.1
>S9_OPF011	47	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	62%	WP_055520670.1
>S9_OPF012	2	3-phenylpropionate dioxygenase	Streptomyces zinciresistens	72%	WP_007499173.1
>S9_OPF013	2	benzene 1,2-dioxygenase	Polymorphum gilvum	48%	WP_041375675.1

>S9_OPF014	2	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	61%	WP_055520670.1
>S9_OPF015	3	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	59%	WP_055520670.1
OPF IDs	**Abundance**	**Enzyme family**	**Closest bacterial hit**	**Ident**	**Accession**
>S9_OPF016	2	benzene 1,2-dioxygenase	Bacillus	90%	WP_048680352.1
>S9_OPF017	4	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	57%	WP_021006872.1
>S9_OPF018	5	phenoxybenzoate dioxygenase	Streptomyces torulosus	62%	WP_055713550.1
>S9_OPF019	11	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	58%	WP_021006872.1
>S9_OPF020	48	naphthalene dioxygenase iron sulfur protein large subunit	Pseudomonas	97%	WP_011475377.1
>S9_OPF021	3	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	61%	WP_008070358.1
>S9_OPF022	2	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	58%	WP_021006872.1
>S9_OPF023	2	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	58%	WP_021006872.1
>S9_OPF024	2	naphthalene 1,2-dioxygenase	Algiphilus aromaticivorans	56%	WP_052367460.1
>S9_OPF025	2	biphenyl 2,3-dioxygenase subunit alpha	Rhodococcus rhodnii	56%	WP_010839881.1
>S9_OPF026	2	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	60%	WP_055520670.1
>S9_OPF027	1	phenylpropionate dioxygenase	Streptomyces acidiscabies	63%	WP_010355984.1
>S9_OPF028	25	benzene 1,2-dioxygenase	Sphingomonas sp. Ant20	47%	WP_037529249.1
>S9_OPF029	1	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	59%	WP_008070358.1
>S9_OPF030	2	benzene 1,2-dioxygenase	Bacillus sp. UNC41MFS5	95%	WP_026564730.1
>S9_OPF031	3	3-phenylpropionate dioxygenase	Escherichia coli	51%	WP_000211164.1
>S9_OPF032	3	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	56%	WP_021006872.1
>S9_OPF033	10	phenylpropionate dioxygenase	Streptomyces acidiscabies	60%	WP_010355984.1
>S9_OPF034	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	62%	WP_055520670.1
>S9_OPF035	3	3-phenylpropionate dioxygenase	Streptomyces zinciresistens	75%	WP_007499173.1
>S9_OPF036	4	IPB-dioxygenase ISP large subunit (IpbA1)	Rhodococcus erythropolis	85%	WP_011133490.1
>S9_OPF037	2	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	56%	WP_021006872.1
>S9_OPF038	6	(2Fe-2S) ferredoxin	Pseudonocardia acaciae	62%	WP_028923111.1
>S9_OPF039	4	3-phenylpropionate dioxygenase	Alcaligenes faecalis	47%	WP_060186372.1
>S9_OPF040	1	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	60%	WP_008070358.1
>S9_OPF041	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	60%	WP_055520670.1
>S9_OPF042	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	98%	WP_012052601.1
>S9_OPF043	1	3-phenylpropionate dioxygenase	Streptomyces zinciresistens	57%	WP_007499173.1
>S9_OPF044	2	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	67%	WP_008070358.1
>S9_OPF045	1	3-phenylpropionate dioxygenase	Mycobacterium	67%	WP_011781569.1
>S9_OPF046	1	Large subunit naph/bph dioxygenase	Rhodococcus sp. EsD8	70%	WP_006933068.1
>S9_OPF047	1	phenylpropionate dioxygenase	Streptomyces acidiscabies	62%	WP_010355984.1
>S9_OPF048	1	Large subunit naph/bph dioxygenase	Rhodococcus sp. EsD8	77%	WP_006933068.1
>S9_OPF049	1	ring-hydroxylating dioxygenase	Enterobacteriaceae bacterium strain FGI 57	52%	WP_015963414.1
>S9_OPF050	1	(2Fe-2S)-binding protein	Blastococcus saxobsidens	45%	WP_014378242.1
>S9_OPF051	7	naphthalene dioxygenase Fe-S protein large subunit	Pseudomonas putida	97%	WP_011154413.1
>S9_OPF052	6	2Fe-2S ferredoxin	Streptomyces scabiei	51%	WP_059078277.1
>S9_OPF053	6	naphthalene 1,2-dioxygenase	Hydrocarboniphaga effusa	55%	WP_007184423.1
>S9_OPF054	2	3-phenylpropionate dioxygenase	Streptomyces zinciresistens	70%	WP_007499173.1
>S9_OPF055	3	3-phenylpropionate dioxygenase	Serratia fonticola	54%	WP_024530775.1
>S9_OPF056	3	phenoxybenzoate dioxygenase	Streptomyces torulosus	63%	WP_055713550.1
>S9_OPF057	2	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	57%	WP_021006872.1

>S9_OPF058	2	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	57%	WP_055520670.1
>S9_OPF059	2	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	59%	WP_055520670.1
>S9_OPF060	2	benzene 1,2-dioxygenase	Bacillus	89%	WP_048680352.1
OPF IDs	**Abundance**	**Enzyme family**	**Closest bacterial hit**	**Ident**	**Accession**
>S9_OPF061	2	phenoxybenzoate dioxygenase	Streptomyces torulosus	62%	WP_055713550.1
>S9_OPF062	2	phenoxybenzoate dioxygenase	Rhodococcus rhodnii	83%	WP_051110981.1
>S9_OPF063	1	benzene 1,2-dioxygenase	Pseudomonas toyotomiensis	95%	WP_059391947.1
>S9_OPF064	1	phenoxybenzoate dioxygenase	Rhodococcus rhodnii	67%	WP_051110981.1
>S9_OPF065	1	benzene 1,2-dioxygenase subunit alpha	Pseudomonas	92%	WP_012052601.1
>S9_OPF066	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	89%	WP_055520670.1
>S9_OPF067	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	60%	WP_055520670.1
>S9_OPF068	1	3-phenylpropionate dioxygenase	Novosphingobium nitrogenifigens	60%	WP_008070358.1
>S9_OPF069	1	phenoxybenzoate dioxygenase	Mycobacterium abscessus	83%	WP_052617997.1
>S9_OPF070	1	3-phenylpropionate dioxygenase	Streptomyces ossamyceticus	86%	WP_055520670.1
>S9_OPF071	1	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	57%	WP_021006872.1
>S9_OPF071	1	aromatic-ring-hydroxylating dioxygenase subunit alpha	Variovorax paradoxus	53%	WP_021006872.1
>S10_OPF01	3	benzene 1,2-dioxygenase	Rhodococcus opacus	99%	WP_012687209.1
>S10_OPF02	2	benzene 1,2-dioxygenase	Bordetella petrii	58%	WP_041863076.1
>S10_OPF03	1	benzene 1,2-dioxygenase	Polycyclovorans algicola	86%	WP_029889979.1
>S10_OPF04	1	benzene 1,2-dioxygenase	Bordetella petrii	61%	WP_041863076.1
>S10_OPF05	1	benzene 1,2-dioxygenase	Bordetella petrii	57%	WP_041863076.1
>S10_OPF06	1	benzene 1,2-dioxygenase	Rhodococcus erythropolis	96%	WP_042445407.1
>S10_OPF07	1	benzene 1,2-dioxygenase	Bordetella petrii	57%	WP_041863076.1
>S11_OPF01	22	naphthalene 1,2-dioxygenase	Nevskia ramosa	92%	WP_022978279.1
>S11_OPF02	4	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	99%	WP_011117400.1
>S11_OPF03	2	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	87%	WP_011117400.1
>S11_OPF04	2	naphthalene 1,2-dioxygenase	Delftia sp. Cs1-4	96%	WP_013801305.1
>S11_OPF04	2	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	93%	WP_011117400.1
>S11_OPF06	1	naphthalene 1,2-dioxygenase	Nevskia ramosa	88%	WP_022978279.1
>S12_OPF1	3	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	89%	WP_011117400.1
>S13_OPF01	4	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	97%	WP_011117400.1
>S13_OPF02	3	ribosomal subunit interface protein	Mycobacterium gilvum	100%	WP_013470369.1
>S13_OPF03	1	ribosomal subunit interface protein	Mycobacterium rhodesiae	99%	WP_014211473.1
>S13_OPF04	1	ribosomal subunit interface protein	Rhodococcus	85%	WP_017681826.1
>S13_OPF05	1	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	90%	WP_011117400.1
>S13_OPF06	1	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	98%	WP_011117400.1
>S13_OPF07	1	naphthalene 1,2-dioxygenase subunit alpha	Pseudomonas	87%	WP_011117400.1
>S15_OPF1	3	naphthalene 1,2-dioxygenase	Polycyclovorans algicola	45%	WP_029889175.1
>S15_OPF2	2	ribosomal subunit interface protein	Mycobacterium gilvum	96%	WP_013470369.1
>S15_OPF3	1	ribosomal subunit interface protein	Mycobacterium	100%	WP_011559048.1
>S15_OPF4	1	ribosomal subunit interface protein	Mycobacterium gilvum	100%	WP_011891532.1
>S15_OPF5	1	catechol 2,3-dioxygenase	Sphingobium yanoikuyae	100%	WP_037508695.1

CAPÍTULO II

Exploring the genetic potential of a fosmid metagenomic library from an oil-impacted mangrove sediment for metabolism of aromatic compounds.

Sanderson Tarciso Pereira de Sousa[1], Lucélia Cabral[1], Gileno Vieira Lacerda Júnior[1], Melline Noronha[1], Júlia Ronzella Ottoni[1], Adilson Sartoratto[1], Valéria Maia de Oliveira[1]

[1]Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas (UNICAMP), Campinas, São Paulo, Brasil.

Artigo submetido na revista Applied Microbiology and Biotechnology

Exploring the genetic potential of a fosmid metagenomic library from an oil-impacted mangrove sediment for metabolism of aromatic compounds

Sanderson Tarciso Pereira de Sousa[1], Lucélia Cabral[1], Gileno Vieira Lacerda Júnior[1], Melline Fontes Noronha[1], Júlia Ronzella Ottoni[1], Adilson Sartoratto[1], Valéria Maia de Oliveira [1]

[1]Research Center for Chemistry, Biology and Agriculture (CPQBA), Institute of Biology, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

Corresponding author. Sanderson Tarciso Pereira de Sousa (sousa.tarciso@gmail.com) Fax: +55 19 2139 2852.

E-mail addresses: sousa.tarciso@gmail.com (Sanderson Tarciso Pereira de Sousa), luc.g.cabral@gmail.com (Lucélia Cabral), gilenolacerdajr@gmail.com (Gileno Vieira Lacerda Júnior), melfontes@gmail.com (Melline Fontes Noronha), ottoni.julia@gmail.com (Júlia Ronzella Ottoni), adilson@cpqba.unicamp.br (Adilson Sartoratto), vmaia@cpqba.unicamp.br (Valéria Maia de Oliveira).

ABSTRACT

Aromatic hydrocarbons (AH) are widely distributed in nature, many of which have been reported as relevant environmental pollutants and valuable carbon sources for different microorganisms. Here we used high-throughput sequencing of a metagenomic fosmid library to evaluate the functional and taxonomic diversity of genes involved in aromatic compounds degradation in oil-impacted mangrove sediments. In addition, we used activity-based approach and gas chromatography to assess the degradation potential of fosmid clones. Our results indicated that AH degradation genes, such as monooxygenases and dioxygenases, were grouped into the categories anaerobic degradation of aromatic compounds (20.34%), metabolism of central aromatic intermediates (35.40%) and peripheral pathways for catabolism of aromatic compounds (22.56%). Taxonomic affiliation of genes related to aromatic compounds metabolism revealed the prevalence of the classes *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* and *Deltaproteobacteria*. Aromatic hydrocarbons (phenol, naphthalene, phenanthrene, pyrene and benzopyrene) were used as the only source of carbon for screening of clones with degradation potential. Of the 2,500 clones tested, 48 showed some respiratory activity in at least one of the five sources of carbon used. The hydrocarbon degradation ability of the ten best performing fosmid clones was confirmed by GC-MS. Further, annotation of assembled metagenomic fragments revealed ORFs corresponding to proteins and functional domains directly or indirectly involved in the aromatic compounds metabolism, such as catechol 2,3-dioxygenase and ferredoxin oxidoreductase. Finally, these data suggest that the indigenous mangrove sediment microbiota developed essential mechanisms towards ecosystem remediation of petroleum hydrocarbon impact.

Keywords: Mangrove, Oil Pollution, Metagenomic library, Functional and taxonomic diversity, Aromatic hydrocarbon degradation.

INTRODUCTION

In the last years, the negative effects of aromatic hydrocarbons (AHs) on environmental integrity and human health have gained prominence in the scientific community. The toxic, genotoxic, mutagenic, carcinogenic and bioaccumulative properties of AHs can impair the balance of biological communities, increasing the mortality rates of several species (Haritash and Kaushik 2009; Seo et al. 2009; Rengarajan et al. 2015). The AHs are non-polar organic compounds that have molecular structure formed by aromatic rings in linear, angular or cluster architecture (Cerniglia 1992; Ghosal et al. 2016). Commonly, AHs are found in the environment (air, soil, aquatic environments and sediments) by natural (forest fires, volcanic eruptions, plant and bacterial metabolism) and anthropic processes (incomplete burning of coal, tar, oil and gas, automobile exhaust, tobacco, smoked food, industrial activities and production of fossil fuels) (Holman et al. 1999; Ohkouchi et al. 1999; Lim et al. 1999; Ghosal et al. 2016).

Generally, the bacteria can take advantage of the AHs as a source of energy in the presence or absence of oxygen, based on their adaptive physiological mechanisms (Fuchs et al. 2011). During aerobic catabolism, the oxygen acts as final electron acceptor, entering in the reaction as substrate for the hydroxylation or oxygenolytic ring cleavage (Parales and Resnick 2006; Vaillancourt et al. 2006). On the other hand, the bacteria that perform anaerobic catabolism use nitrate, sulfate or ferric ions as final electron acceptors in reductive reactions (Gibson and S. Harwood 2002; Fuchs 2008). In oxygenated environments, AH metabolism is usually started by monooxygenase or dioxygenase enzymes (Fuchs et al. 2011; Jouanneau et al. 2011). Ring-hydroxylating dioxygenases (RHDs) and ring-cleaving dioxygenases (RCDs) are multicomponent enzymes normally formed by reductase, ferredoxin, and terminal oxygenase subunits, which play a crucial role in the AH biodegradation (Lipscomb 2008; Lu et al. 2011).

Since the beginning of life on earth, evolutionary forces (mutation, natural selection, genetic drift and gene flow) have driven the establishment of microbial populations capable of degrading aromatic hydrocarbons. In the last decades, several approaches have been employed to access microorganisms with potential to degrade these compounds and describe their metabolic pathways, in an attempt to develop technologies do mitigate AH environmental impact (Pace et al. 1986; Gallego et al. 2014; Vila et al. 2015). Many previous studies based on traditional culturing techniques have reported the ability of several bacterial genera (*Arthrobacter*, *Acidovorax*, *Bacillus*, *Burkholderia*, *Marinobacter*, *Mycobacterium*, *Pseudomonas*, *Rhodococcus*, *Stenotrophomonas*, *Sphingomonas*, *Xanthamonas*) for aromatic hydrocarbons degradation (Seo et al. 2009; Carmona et al. 2009). However, it is now widely known that the largest portion of the functional and taxonomic diversity in nature remain inaccessible through cultivation approaches (Steele and Streit 2005; Hugenholtz and Tyson 2008). To circumvent this problem and to better understand *in situ* hydrocarbon degradation microbial processes, cultivation-independent techniques, such as metagenomics, have been employed (Liang et al. 2011; Andreote et al. 2012; Guazzaroni et al. 2013; Mason et al. 2014; Tan et al. 2015).

Functional metagenomics, based on metagenomic library construction and subsequent screening, is a powerful strategy that allows one to isolate new genes of interest and to exploit the catalytic potential of their respective enzymes (Popovic et al. 2017). Clear examples of the successful use of activity-based approach for prospecting of novel genes of biotechnological interest have been published in literature, mainly for hydrolases (Kanokratana et al. 2015) (Bashir et al. 2014). Other studies have contributed to

elucidate novel metabolic pathways in the biodegradation of aromatic compounds using activity-based approach associated with functional classification of annotated genes in sequenced fosmid inserts (Suenaga et al. 2009; Silva et al. 2013; Sierra-García et al. 2014). Large pieces of metagenomic DNA from fosmid libraries can provide a unique source of information about new catabolic gene clusters and their regulatory mechanisms related to aromatic compounds degradation pathways.

Prospecting new genes from the metagenome of contaminated environments and microbial consortia, combined with synthetic biology techniques, have allowed the implementation of bioremediation technologies in polluted sites (Semple et al. 2001; Straube et al. 2003; Ang et al. 2005; Singh et al. 2008; Megharaj et al. 2011). In addition, the discovery of new genes may contribute to the establishment of genetic markers for monitoring environments contaminated with aromatic hydrocarbons (Rabus et al. 2016; von Netzer et al. 2016). The mangrove sediment of Bertioga-SP was chosen due to its history of contamination by oil spill and its conditions of high salinity. Mangroves enriched with AHs, under high salt concentrations, may favor the acquisition of genes for halophilic / halotolerant enzymes involved in the metabolism of aromatic compounds, which can be advantageous in biotechnological processes for decontamination of halophilic impacted environments. Based on the considerations above, a metagenomic fosmid library from oil-impacted mangrove sediment was constructed in a previous study (Ottoni et al., 2017) and used herein in the search for pathways of aromatic compounds metabolism using large-scale sequencing and activity-based approach.

MATERIAL AND METHODS

Metagenomic fosmid library

A metagenomic fosmid library consisting of 12,960 clones was previously constructed from community DNA extracted from oil-impacted mangrove sediment using the CopyControl™ HTP Fosmid Library Production Kit (Epicentre, USA). Details of mangrove sediment sampling, DNA extraction and library construction are given in Ottoni et al. (2017).

Total Fosmid DNA Sequencing and data processing

Total fosmid DNA from metagenomic library was extracted as described by Ottoni et al. (2017). In short, fosmid clones were grown in LB culture medium containing chloramphenicol (12,5 µg/mL) and L-arabinose (0,02%), centrifuged and subjected to DNA extraction using QIAGEN® Large-construct Kit (Qiagen Sample & Assay Technologies, Germany), according to the manufacturer's instructions. High throughput sequencing of total fosmid DNA was conducted using one lane of the Illumina HiSeq 2000 platform, by the facility for Functional Genomics for Agricultural Studies (ESALQ / University de São Paulo, Piracicaba, Brasil), and the 454 GS FLX Titanium platform (Macrogen, Korea). Detailed sequencing data are provided by Ottoni et al. (2017). All metagenomic sequencing reads were firstly trimmed for exclusion of genomic DNA from *E. coli* EPI-300 host and pCC2FOS vector sequences using the CLC's Workbench software version 6.5.1 (CLC Bio-Qiagen, Aarhus, Denmark). Remaining sequences were subjected to MG-RAST (Metagenomics Analysis Server, Argonne National Laboratory) for quality control, automated gene calling, annotation and taxonomic classification according to the following parameters: SEED and Refseq annotation, e-value $\leqslant 1 \times 10^{-5}$, alignment $\geqslant$ 15 bp, identity $\geqslant$ 60% (Meyer et al. 2008).

High quality sequences generated by the Illumina and 454 sequencing platforms were deposited in the MG-RAST database (http://metagenomics.anl.gov) under the access of MG-RAST ID's 4561201.3 and 4558576.3, respectively. In the current study, we focused on genes encoding proteins involved in the metabolism of aromatic compounds, mainly dioxygenases. Taxonomic profiling was performed for metagenomic reads previously annotated for the metabolism of aromatic compounds.

Preliminary screening of clones for AH degradation

Preliminary functional screening of fosmid clones for the ability to use AH was carried out using a colorimetric assay, according to the methodology developed by Johnsen et al. (2002), with modifications implemented by Silva et al. (2013). For this, fosmid clones were grown in 96 deep-well plates with 1 mL of LB broth culture medium containing chloramphenicol (12.5 µg/mL) and L-arabinose (0.02%), incubated for 18 h at 37°C and 180 rpm. Fifteen µL-aliquots of the pre-inoculum were transferred to 96-well microplate containing 150 µL BH (Bushnell Haas) (Bushnell and Haas 1941) mineral medium with chloramphenicol (12,5 µg/mL), supplemented with one of the following hydrocarbons: phenol (0.2 mg/mL) (Silva et al. 2013a); naphthalene (0.1 mg/mL), phenanthrene (0.1 mg/mL) and pyrene (0.1 mg/mL) (Zhou et al. 2008); benzopyrene (0.05 mg/mL) (Juhasz et al. 2000). The plates were incubated at 37°C and 180 rpm for 48 h, and *E. coli* EPI-300 host cells, without exogenous DNA, was used as a negative control. After incubation, 30 µL of 3-(4,5-dimethyl-2-thiazolyl)-2,5 -diphenyl-2 H-tetrazolium bromide (MTT) (Merck) (1 mg/mL) was added to each well. Clone cultures with purple color indicated respiratory activity and were scored as positive hits to be further confirmed by chromatographic analyses (Bicalho et al. 2003).

GC-MS analyses of AH degradation

Pre-screening with MTT was used to select the top 10 most promising fosmid clones for PAHs degradation analysis by GC-MS. Each clone culture was grown in 1 L Schott flasks containing 100 mL of LB medium supplemented with chloramphenicol (12.5 µg/mL) and L-arabinose (0.02%) for 16 h at 37°C. After cultivation, 500 µL-aliquots of cell suspension (10^8 CFU/mL) were transferred for 250-mL Erlenmeyer flasks containing 50 mL of BH minimal medium supplemented with chloramphenicol (12.5 µg/mL), L-arabinose (0.02%), yeast extract (0.1%), 50 mL of micronutrients solution (10.95 g $ZnSO_4.7H_2O$; 5 g $FeSO_4.7H_2O$; 1.54 g $MnSO_4.7H_2O$; 0.39 g $CuSO_4.5H_2O$; 0.25 g $Co(NO_3).6H_2O$; 0.17 g $Na_2B_4O_7.10H_2O$) and a PAHs cocktail as sole carbon source. Phenol, naphthalene, phenanthrene, pyrene and benzopyrene were dissolved in acetone and added to the culture to the final concentrations 0.2, 0.1, 0.1, 0.1 and 0.05 mg/mL, respectively. The assays were conducted in triplicate and cells were incubated for 16 days at 37°C and 140 rpm. Grown cultures (triplicate) were sampled every 4 days and then extracted with two volumes of ethyl acetate. The organic phase was separated, dewatered over anhydrous Na_2SO_4, filtered through a cotton layer and concentrated by rotary evaporator. The parameters used to run the samples in CG-MS were defined according to Vasconcellos et al. (2010).

Sequencing of Fosmid DNA from positive clones

Fosmid clones that showed the strongest colorimetric reaction in the functional assay with MTT were subjected to DNA sequencing. Before extraction, ten positive clones were cultivated in 2 mL of LB broth

culture medium containing chloramphenicol (12.5 µg/mL) and L-arabinose (0.02%) for 16 h at 37°C and 180 rpm. Aliquots of 500 µL of each culture were added to 100 mL of the same LB medium in Erlenmeyer flask (250 mL), followed by incubation at 37°C and 180 rpm for 16 h. After growth, cells were centrifuged and the pellet subjected to extraction. Fosmid DNA extraction was performed using QIAGEN® Large-construct Kit (Qiagen Sample & Assay Technologies, Germany), according to the manufacturer's instructions. All fosmid DNA samples were quantified using NanoDrop Nucleic Acid Quantification (Thermo Scientific, USA), according to the manufacturer's protocols. Finally, one µg of each fosmid DNA was sent for high throughput sequencing using Illumina MiSeq System platform, at MR DNA Laboratories (Molecular Research LP, USA).

Structural and functional characterization of fosmid inserts

Raw reads obtained from sequencing were processed using the following workflow: (i) paired-end sequences (2x300 pb) were combined using Flash v. 1.2.11 software (Magoč and Salzberg 2011); (ii) sequences were trimmed using Trimmomatic v. 0.33 software (Bolger et al. 2014) and grouped in FASTqc file for quality control; (iii) digital normalization of the filtered paired-end reads was made using Trinity v.2.0.6 software (Grabherr et al. 2011); (iv) identification and cleaning of unwanted sequences from *E. coli* and host cloning vectors was performed against a hybrid database (Bowtie2 + *E. coli* databank + UniVEC); and finally (v) remaining sequences were assembled using Spades v.3.11.1 software (Bankevich et al. 2012). Assembled sequences were submitted to RAST 2.0 (The Rapid Annotation using Subsystem Technology) (Aziz et al. 2008) and IMG (Integrated Microbial Genomes (IMG accession number: Ga0101857)) (Markowitz et al. 2014) computing platforms for mapping of open reading frames (ORFs), calculating the GC content (G+C%) and automatic annotation from predicted genes against public database. The ORFs were grouped into functional categories by COG (Clusters of Orthologous Groups) database (Tatusov et al. 2000) and assigned to putative metabolic pathways by KEGG (Kyoto Encyclopedia for Genes and Genomes) database (Ogata et al. 1999).

Phylogenetic analysis of dioxygenase sequences

Annotated data from the metagenomic library 454 sequencing were used for phylogenetic analyzes of the following gene sequences: catechol dioxygenase, protocatechuate dioxygenase and aromatic ring-hydroxylating dioxygenase. The reads generated by 454 sequencing were chosen for assembly into contigs due to their average length of 460 bp. Assembly was conducted in BIOEDIT software (Hall 1999) using CAP3 tool (Huang and Madan 1999). The contigs were aligned against "Reference Protein" database (Refseq_protein) using BLASTx tool (Altschul et al. 1990), from the National Center for Biotechnology Information (NCBI). The returned best hits were used to determine correct reading frame and as reference for taxonomic classification. The EMBOSS Transeq software (Rice et al. 2000) was used to convert the nucleotide sequences into amino acid sequences. A dataset containing contigs and their best hit sequences was aligned using ClustalW software (Larkin et al. 2007). Phylogenetic trees were constructed using MEGA v6.06 software (Tamura et al. 2013), according to the following parameters: Neighbor-joining method (Saitou and Nei 1987); Kimura 2-parameter substitution model (Kimura 1980); and bootstrap values calculated from 1,000 replicate runs.

RESULTS

Taxonomic and functional assessment

The main purpose of this work was to evaluate the genetic potential of a fosmid library constructed from the environmental DNA of oil-impacted mangrove sediments for the metabolism of aromatic compounds. Sequencing of the total metagenomic DNA contained in the fosmid library yielded 930,817 and 12,205,480 raw sequences by 454 and Illumina platforms, respectively (Ottoni et al., 2017). Detailed statistical data from sequencing can be viewed in Table 1. The genetic potential of the environmental microorganisms was analyzed through the functional annotation of the metagenomic sequences against the SEED database. The focus of this work was to analyze the genetic diversity and abundance of the SEED category for the metabolism of aromatic compounds.

Functional analysis

A total of 71,729 metagenomic sequences were annotated for metabolism of aromatic compounds, representing approximately 1% of all dataset (Fig. 1). These sequences were distributed into three main categories: Peripheral pathways for catabolism of aromatic compounds (22.56%); anaerobic degradation of aromatic compounds (20.34%); metabolism of central aromatic intermediates (35.40%) (Fig. 1). At the level of the COG category for anaerobic degradation of aromatic compounds, a predominance of sequences annotated for anaerobic benzoate metabolism (19.66%) was observed for the mangrove fosmid library. According to the functional classification, several sequences corresponded to the Benzoate-CoA ligase enzyme (4.47%), which is involved with the reduction of benzoate to benzoyl-CoA. The metagenomic sequence annotation also revealed many genes involved in the central metabolism pathways of aromatic compounds. Most of the genes in this category were related to central meta-cleavage pathway (12.21%), homogentisate pathway (7.73%), protocatechuate (6.04%) and catechol (5.27%) branches of beta-ketoadipate pathway, N-heterocyclic aromatic compound degradation (2.82%), 4-Hydroxyphenylacetic acid catabolic pathway (0.87%) and salicylate and gentisate catabolism (0.46%) (Fig. 1).

Based on the data processing against the SEED databases, only sequences related to dioxygenase genes involved in the metabolism of aromatic compounds were selected (Table 2). Several dioxygenases that play an important role in the peripheral and central pathways of aromatic compound catabolism were found in the mangrove fosmid library dataset. The enzymes, such as catechol, protocatechuate, homogentisate, hydroxyphenylpyruvate and dihydroxyphenylacetate dioxygenases, involved in central metabolism, were the most abundant (Table 2). Nonetheless, the dioxygenases for Phenylpropionate, Lignostilbene, Biphenyl, Benzoate and dihydroxybiphenyl, involved in the peripheral pathways of aromatics degradation, were also found (Table 2).

Taxonomic analysis

Taxonomic affiliation of the metagenomic fosmid library sequences related to the metabolism of aromatic compounds revealed that members of the Bacteria domain comprised 98.13% of the database. The main bacterial classes harboring the genes involved in the metabolism of aromatic compounds were *Alphaproteobacteria* (41.52%), *Actinobacteria* (16.54%), *Betaproteobacteria* (14.22%),

Gammaproteobacteria (10.16%) and *Deltaproteobacteria* (5.45%). Another 22 bacterial classes showed abundance below 3% (Fig. 2). A deeper analysis of the bacterial community involved in hydrocarbon degradation showed significant percentages of sequences classified as *Rhodopseudomonas* (11.11%), *Bradyrhizobium* (6.58%), *Ruegeria* (3.47%), *Mycobacterium* (3.21%), *Burkholderia* (3.18%), *Maritimibacter* (2.81%), *Frankia* (2.61%), *Pseudomonas* (2.53%), *Novosphingobium* (2.30%), *Magnetospirillum* (2.23%) and *Roseobacter* (2.08%), followed by other minor groups with relative abundance lower than 2% (Fig. 2). Ternary plot analysis showed the distribution of classes related to the following functional categories: Anaerobic degradation of aromatic compounds (A); Metabolism of central aromatic intermediates (B); Peripheral pathways for catabolism of aromatic compounds (C). The distribution analysis of sequence data revealed that *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria* and *Flavobacteria* were the most dominant classes in A, B and C categories, whereas unclassified (derived from *Cyanobacteria*), *Chlorobia*, *Dehalococcoidetes*, *Epsilonproteobacteria* and *Elusimicrobia* were detected only in C category, but less frequently. Several less abundant classes as *Chloroflexi*, *Bacteroidia*, *Gloeobacteria*, *Spirochaetes*, *Opitutae*, *Dictyoglomia* and *Planctomycetacia* were rarely or not detected in A and B categories, and only the last two classes were not found in category C (Fig. 2). Several specific bacterial genera, including those involved in aerobic and anaerobic degradation of aromatic compounds, were observed. *Rhodopseudomonas* was the most abundant genus, followed by *Bradyrhizobium* and *Ruegeria*.

Mapping and functional characterization of fosmid inserts

Fosmid inserts were sequenced by Illumina MISEQ platform and the sequence assembly process yielded ten contigs. Length of the contigs obtained ranged from 27,296 to 43,182 bp, with 12 X coverage (Table 3). The size of the contigs confirmed the efficiency of the cloning process in fosmid vector, which can harbor DNA inserts around 40 kb. The contigs processing by IMG (Integrated Microbial Genomes) platform helped to elucidate the number of ORFs contained in each fosmid insert as well as its reading direction and its functional classification against the COG database (Fig. 3; Supplementary Table S1). The number of annotated and non-annotated ORFs was determined from the functional classification against the COG, Pfam and KEGG databases (Table 3; Supplementary Table S1). The contigs were also submitted to the IMG platform for functional and taxonomic classification (Markowitz et al. 2014). The used algorithm for taxonomic classification revealed the closest microbial group for each metagenomic fragment (Table 3). Fosmids 01, 05 and 08 were classified only at the domain level (Bacteria); fosmid 06 was classified at the phylum level (Firmicutes); fosmids 02 and 10 were classified at the class level (*Actinobacteria* and *Deltaproteobacteria*, respectively); fosmids 04, 07 and 09 were classified at the order level (*Rhizobiales*, *Planctomycetales*); and only fosmid 03 was classified at the genus level (*Rhodopirellula*) (Table 3).

Functional annotation of the ORFs in each fosmid insert against the KEGG database allowed classification of gene sequences in different categories (Table S1). In general, most of the ORFs of fosmids (78.61%) did not match any correspondent in KEGG database. The remaining sequences were assigned to amino acid metabolism (12%), carbohydrate metabolism (10%), energy metabolism (7%), nucleotide metabolism (6%), xenobiotic metabolism (6%) and membrane transport (5%). Enzymes related to the

metabolism of aromatic hydrocarbons were found within the category of xenobiotics metabolism: Predicted acyltransferases - EC:3.8.1.3 (FOS1; FOS4) (Chlorocyclohexane and chlorobenzene degradation); Tautomerase enzyme - EC:5.3.2.6 (FOS10) (Benzoate degradation, Xylene degradation); Zn-dependent alcohol dehydrogenases - EC:1.1.1.1 (FOS8) (Naphthalene degradation); Acyl CoA:acetate / 3-ketoacid CoA transferase - EC:2.8.3.12 (FOS6) (Styrene degradation); and Acetyltransferases (FOS2; FOS7) (Aminobenzoate degradation, Ethylbenzene degradation) (Supplementary Table S1).

In an attempt to extend the coverage of functionally annotated ORFs, alignment was also performed with the COG database. In this case, 29.64% of the ORFs were not classified in any functional category while 70.36% had at least one hit. The best represented categories within each fosmid clone are shown in Fig. 3. We observed the predominance of sequences related to the transport of organic and inorganic molecules (e.g. ABC-type transport system) and to energy production and conversion. In addition, genes encoding proteins involved with signal transduction mechanisms, chemotaxis signaling, stress response, sporulation control, biofilm development, as well as kinases and phosphatases were also observed. According to the functional classification performed in the IMG platform, against the COG database, ORFs contained in the metagenomic DNA fragments were annotated for likely conserved regions of dioxygenases: Polyferredoxin (FOS9); Catechol 2,3-dioxygenase (FOS5); Fe-S oxidoreductase (FOS10); Ferredoxin oxidoreductase (FOS6); Predicted Fe-S-cluster oxidoreductase (FOS10); Predicted oxidoreductases (FOS9); and Aromatic ring hydroxylase (FOS4) (Supplementary Table S1). Interestingly, one gene for Epoxide hydrolase (EC 3.3.2.9) and another gene encoding for non-ribosomal peptide synthetase were also found. Some ORFs were annotated as proteins involved in the biosynthesis of Molybdopterin-guanine dinucleotide and Aldehyde dehydrogenase, which, to a certain extent, are related to the metabolism of aromatic hydrocarbons.

Although the main goal of this work was to find new dioxygenases involved in the aromatic hydrocarbons degradation, the results of the functional classification of the fosmid inserts revealed potentially new genes for interesting enzymes with several industrial applications. For instance, hydrolases as CAAX amino terminal protease family protein, Zn-dependent aminopeptidase, Putative cysteine protease, Lipase/esterase (EC 3.1.1.-), Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8), Carboxylesterase, Carboxyl-terminal protease (EC 3.4.21.102) and Membrane alanine aminopeptidase N (EC 3.4.11.2) (Busto et al. 2010). Detailed information on the annotated genes contained in all ten fosmid clones are shown in Supplementary Table S1.

Hydrocarbon degradation assays

The activity-based assays using MTT revealed 50 clones with respiratory activity, detected through colorimetric signal (Supplementary Table S2). The 10 best fosmid clones selected in the MTT assays based on their tolerance to different classes of hydrocarbons were subjected to quantitative analysis of aromatic hydrocarbon degradation by GC-MS. The analyses showed that clones 01, 08 and 10 showed high degradation activity for phenol (98%), pyrene (75%) and naphthalene (71%), respectively (Table 4). On the other hand, clones 03, 05, 06, 07 and 09 showed low extent of hydrocarbon degradation (<30%; Table 4). However, despite the low degradation percentage, clones 03 and 09 were able to consume more than one hydrocarbon dissolved in the same culture medium. Clone 03 degraded phenol (13.99%), phenanthrene

(27.32%), pyrene (16.1%) and benzopyrene (13.12%). Clone 09 consumed pyrene (10.55%) and benzopyrene (15.14%). In addition, two clones (02 and 04) showed no significant degradation activity for the hydrocarbons used. Although they had respiratory activity in MTT assays, clones 02 and 04 showed no detectable degradation activity by GC-MS, at least between 12 and 16 days of monitoring.

Interestingly, the same clone carrying the metagenomic insert containing two ORFs characterized with functional domain for Predicted Fe-S-cluster oxidoreductase (FOS10) showed degradation activity for naphthalene. In addition to ORFs with functional domains for oxidoreductase, FOS10 clone also revealed ORF for Tautomerase enzyme, which is related to the aromatic compounds degradation, such as Benzoate and Xylene. Clones FOS1 and FOS8 contained ORFs annotated for Predicted acyltransferases and alcohol dehydrogenases, respectively, also involved in the aromatic compounds metabolism. Although clones FOS3, 5, 6, 7 and 9 have shown low degradation rates for the hydrocarbons analyzed, they harbored inserts containing ORFs for enzymes and functional domains supposedly involved with aromatic compounds metabolism: Polyferredoxin and Predicted oxidoreductases (FOS9); Catechol 2,3-dioxygenase (FOS5); Ferredoxin oxidoreductase and Acyl CoA:acetate/3-ketoacid CoA transferase (FOS6); Acetyltransferases (FOS07).

Phylogenetic analysis of putative genes for dioxigenases

The assembly of reads yielded by 454 platform allowed to recover 10, 7 and 12 contigs for phylogenetic analysis of genes coding for catechol dioxygenase, protocatechuate dioxygenase and aromatic ring-hydroxylating dioxygenase, respectively. All contigs, except one catechol dioxygenase-related contig (contig4), were affiliated to *Alphaproteobacteria*, *Actinobacteria*, *Gammaproteobacteria* e *Betaproteobacteria* (Fig. 5a, b, c). As discussed above, these classes were already indicated as the most abundant among the metagenomic library sequences for aromatic compound metabolism. Curiously, contig4 showed 64% homology to a catechol dioxygenase present in the genome of *Thermus thermophilus*, an extremely thermophilic and halotolerant species of *Deinococci* phylum (Fig. 5a; Supplementary Table S3). The phylogenetic tree of each gene function revealed some contigs in separate branches: contigs 6, 7, 9, 15 and 16 for catechol dioxygenase; contigs 2, 11 and 12 for protocatechuate dioxygenase; and contigs 17 and 28 for aromatic ring-hydroxylating dioxygenase.

DISCUSSION

Mangrove microbiome involved in aromatic compounds metabolism

Benzoyl-CoA plays a crucial role in the anaerobic benzoate metabolism pathway and is part of a limited number of intermediate aromatic metabolites that precede the pathways of the central metabolism of aromatic compounds (Carmona et al. 2009). The presence of the Benzoate-CoA ligase enzyme is indicative that a significant fraction of the microorganisms present in the mangrove sediment are genetically prepared to use the anaerobic metabolic pathway in the degradation process of aromatic compounds by converting benzoate to benzoyl-CoA. This finding is consistent with the characteristics of the sampled environment, since the mangrove sediment presents fluctuating levels of dissolved oxygen.

Our data also revealed many genes involved in the central metabolism pathways of aromatic compounds. During the aerobic degradation of polycyclic aromatic hydrocarbons, bacteria use dioxygenase

enzymes to add hydroxyl groups to the aromatic rings, forming cis-dihydrodiol. After its formation, this compound is converted into a diol intermediate by a dehydrogenase enzyme, returning to its aromatic form. The intermediate diol can then be cleaved via ortho- or meta-cleavage by intradiol or extradiol ring-cleaving dioxygenases, respectively. Finally, the intermediate cleaved metabolites enter the pathway of the Tricarboxylic Acid (TCA) cycle (Mallick et al. 2011; Ghosal et al. 2016). Benzoate is a monocyclic aromatic compound and a metabolic product of the peripheral pathway of the aromatic compounds catabolism. At the beginning of the benzoate aerobic degradation process, dioxygenases and monooxygenases add hydroxyl groups into their aromatic ring, producing reactive intermediates that lead to the formation of catechol and protocatechuate, respectively. Catechol and protocatechuate can be generated not only from benzoate but also from an arsenal of phenolic compounds present in nature. The molecular structure of catechol can be cleaved by dioxygenases, either at the position between the two hydroxyl groups or at a position adjacent to the hydroxyl groups, through the ortho- and meta-cleavage pathways, respectively. Similar process occurs with the protocatechuate. In the ortho-cleavage route, both catechol and protocatechuate are cleaved by specific dioxygenases generating intermediate metabolites that follow up to β-ketoadipate formation, which in turn is converted to succinate, entering the TCA pathway. In the meta-cleavage route, catechol and protocatechuate do not follow the β-ketoadipate pathway, but are cleaved by dioxygenases into intermediate metabolites that follow, respectively, until the formation of acetaldehyde and pyruvate, which enter the TCA pathway (Fuchs et al. 2011).

Through deeper analysis of the data, we find only sequences related to dioxygenase genes involved in the metabolism of aromatic compounds. These enzymes play key role in the aromatic hydrocarbons degradation and their presence reveal the significant genetic potential of the metagenomic library for prospecting degradative genes and pathways for further biotechnological application.

According to the taxonomic analysis, bacteria domain play a crucial biochemical role in the cycling of the aromatic compounds present in the mangrove sediment under study. A previous study based on the taxonomic affiliation of the environmental DNA collected from the same sample used in this study revealed that *Alphaproteobacteria* is the third most abundant bacterial class found in the oil-impacted mangrove sediment (Andreote et al. 2012). Our study showed the dominance of sequences affiliated with *Alphaproteobacteria*, considering only those involved in the metabolism of aromatic compounds. These results suggest that species of the *Alphaproteobacteria* class play an important role in the cycling of aromatic hydrocarbons in the analyzed mangrove environment. The outstanding function of *Alphaproteobacterias* in AH-contaminated environments is corroborated by other studies that evaluated the composition of the microbial community in soils, sea surface and sediment and beach sands (Viñas et al. 2005; Kostka et al. 2011; Liu and Liu 2013). *Actinobacteria* is a phylum widely known for its metabolic versatility, especially the production of diverse metabolites with antimicrobial activity and the ability for recalcitrant compounds degradation (Lewin et al. 2016). According to a review published by Seo et al. (2009), some species of *Actinobacteria*, such as *Mycobacterium* spp., *Rhodococcus* spp. and *Streptomyces* spp., have been isolated from environments containing naphthalene as the sole source of carbon. A recent study revealed synergistic and antagonistic effects in the single and mixed PAH degradation by four *Mycobacterium* species with potential for bioremediation (Hennessee and Li 2016). *Betaproteobacteria* is a widespread class, which also contains members involved in processes of AH degradation. Through a

diversity analysis based on the study of 16S rRNA gene clone libraries from metagenomic DNA recovered from deep subsea-floor rock samples of a Brazilian oil reservoir, Weid et al. (2008) revealed high frequency of Operational Taxonomic Units (OTUs) belonging to *Betaproteobacteria* class. Additionally, bacterial isolation combined to cultivation-independent techniques have showed the prominent presence of *Betaproteobacteria* members, such as *Petrobacter*, *Thauera*, *Dechlorosoma*, *Hydrogenophilus*, *Azoarcus* and *Burkholderia*, in oil reservoir samples (Vasconcellos et al. 2010; Verde et al. 2013; Silva et al. 2013b).

Among the most abundant bacterial genera found, *Rhodopseudomonas* presents phototrophic properties and great metabolic versatility, able to anaerobically degrade different aromatic compounds (Larimer et al. 2004). The high occurrence of sequences affiliated with *Rhodopseudomonas* can be linked to the fact that the mangrove sediment is an environment with oxygen restriction. Our results are corroborated by a previous study that found isolates of *Rhodopseudomonas* and *Bradyrhizobium* harboring genes for the α-subunit of benzoyl-CoA reductase (BCR) (Song and Ward 2005). Benzoyl-CoA reductase is a key enzyme present in bacteria with capability for the anaerobic degradation of aromatic compounds. Song and Ward (2005) detected a high degree of similarity among sequences of Benzoyl-CoA reductase obtained from *Bradyrhizobium* and *Rhodopseudomonas* isolates, suggesting the occurrence of horizontal gene transfer between these taxa. The presence of genera that are found as free-living bacteria and in symbiotic associations with plants, such as *Frankia* and *Bradyrhizobium*, suggests a systemic relationship among plant-bacteria-environment in the metabolism of aromatic compounds in the mangrove sediment. The molecular synergy between plants and rhizospheric microorganisms has been extensively studied in an attempt to develop more efficient models for biological remediation of polluted soils (Segura et al. 2009). Previous reports have showed that the degradation of petroleum hydrocarbons by plant-associated bacteria can be 10-1,000 times higher than by free soil-dwelling bacteria (Gkorezis et al. 2016). *Pseudomonas* is also a well-known genus for their ability to degrade aromatic hydrocarbons, and strains of *Pseudomonas putida* have been used as a model organism in synthetic biology in order to increase degradation efficiency (Seo et al. 2009; Loeschcke and Thies 2015).

Determination of the genetic potential of fosmid clones

The taxonomic classification of the fosmid inserts are in accordance with literature, which have showed that most of the bacteria with genetic activity and/or predisposition for AHs degradation belong to Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Actinobacteria and Firmicutes (Ghosal et al. 2016). The ability of AH degradation may constitute a key metabolic process for bacteria of the order *Rhizobiales*, which live in association with plant roots (rhizobacteria) and may need to use plant aromatic compounds as carbon source. A study published by Keum et al. (2006) showed that a bacterium of the genus *Sinorhizobium*, isolated from soil contaminated with PAHs, can use phenanthrene as the only source of carbon for growth. The interaction between bacteria and plant roots is a very studied and promising phenomenon for the development of bioremediation technologies of contaminated areas (Haritash and Kaushik 2009). A previous study carried out by Yang et al. (2014) revealed an increased occurrence of bacteria belonging to the phyla Proteobacteria, Actinobacteria and Firmicutes in crude oil-amended soils. In another research, bacteria from the same phyla were enriched in oil sand tailing ponds and crude oil degrading microcosms under anaerobic conditions (Sherry et al. 2013; An et al. 2013). The taxonomic

profile obtained in the above studies was corroborated by the one observed in this work, which is consistent with the fact that the mangrove sediment presents limited oxygen concentration. To date, there are no reports that members of the genus *Rhodopirellula* possess AH degradation abilities. However, different genes coding for dioxygenases, mono-oxygenase and epoxide hydrolase have been found in *Rhodopirellula baltica* genome in distinct studies (Glöckner et al. 2003). These evidences may point to this genus as a potential new candidate for AHs degradation.

The high occurrence of hypothetical proteins against the KEGG database suggests the existence of potentially new genes. Similar results were shown in a previous work, which also analyzed metagenomic DNA from a fosmid library, but with environmental samples of different origin (Sierra-García et al. 2014). On the other hand, the functional annotation against the COG database increased the number of ORFs with matching. The presence of several genes related to the transport of organic and inorganic molecules, shows a great versatility of the metagenomic inserts in the processes of energy metabolism. Previous studies have revealed the presence of genes encoding transport proteins within gene clusters involved in the metabolism of aromatic compounds (Barragán et al. 2004; Rabus et al. 2005; Carmona et al. 2009).

This study revealed other interesting genes encoding proteins, as detailed below. Genes involved with signal transduction mechanisms, which play an important role in the recognition of environmental characteristics, conditioned by biotic and abiotic factors, and trigger a cascade of molecular processes that transfer information to the cell interior, changing its behavior (Kiel et al. 2010). Genes for proteins related to processes of chemotaxis signaling, stress response, sporulation control, biofilm development, as well as kinases and phosphatases enzymes can be related to resistance or bacterial tolerance to environments contaminated by aromatic compounds. Bacteria subjected to high concentrations of aromatic compounds may either use them as nutrients or trigger oxidative stress signals due to the level of toxicity (Carmona et al. 2009). However, some bacteria may first trigger a short-term molecular response as the activation of stress tolerance genes and then initiate the expression of genes for biodegradation (Velázquez et al. 2006). Recent studies have revealed significant physiological changes at genomic and proteomic levels in PAH-exposed bacteria, including up-regulated genes involved in primary carbon metabolism, stress response, PAH catabolism and transport of organic molecules (Denef et al. 2005; Seo et al. 2009). We detect ORFs for Epoxide hydrolase (EC 3.3.2.9), which is involved in detoxification processes of xenobiotic compounds, as well as in the degradation of aromatic compounds (Arand et al. 2005). Gene encoding for non-ribosomal peptide synthetase can be involved in the biosynthesis of biosurfactants that emulsify complex mixtures of water insoluble aromatic hydrocarbons (Roongsawang et al. 2011).

Our data also revealed ORFs annotated to proteins involved in the biosynthesis of Molybdopterin-guanine dinucleotide and Aldehyde dehydrogenase. Molybdopterin-guanine dinucleotide is a cofactor that binds to the active site of the Ethylbenzene dehydrogenase enzyme, which contains electron transfer chains with five Fe-S groups and one heme B group (Kloer et al. 2006). Some special cases of anaerobic aromatic compounds degradation involve the hydroxylation of aromatic rings by Ethylbenzene dehydrogenase enzyme, using water as co-substrate in place of oxygen (Rabus and Heider 1998; Fuchs et al. 2011). Aldehyde dehydrogenases are enzymes present in a large part of living organisms and are involved in the aldehydes detoxification, as well as in the degradation of many hydrocarbons (Jia et al. 2017). In a proteomic study, Kim et al. (2004) demonstrated the overexpression of aldehyde dehydrogenase and heat

shock chaperones, along with small subunits of dioxygenases, in *Mycobacterium vanbaalenii* PYR-1 grown in culture medium supplemented with pyrene, pyrene-4,5-quinoline, phenanthrene, anthracene and fluoranthene.

Dioxygenases are multi-component enzymes that act in the first step of the aromatic hydrocarbons degradation and consist of a catalytic component with a portion that promotes the hydroxylation or cleavage of rings attached to a ferredoxin domain (Jouanneau et al. 2011). Although only one dioxygenase sequence (Catechol 2,3-dioxygenase) has been found, the annotation results obtained do not exclude the possibility that putative new genes for dioxygenases are present in the metagenomic fragments and were eventually classified as hypothetical proteins due to low homology to known proteins.

Phenotypic plasticity for aromatic hydrocarbons degradation

Apparently, complete gene clusters known to be involved in aromatic hydrocarbon degradation were not found in the fosmids. However, the ORFs annotated as hypothetical proteins, found in the fosmid inserts, may be part of potentially new genetic organizations involved in the aerobic or anaerobic degradation of aromatic compounds. The occurrence of fosmid clones with degradation activity highlights their potential use in the creation of new biotechnological processes. GS-MS results further suggested that either clones without detectable degradation activity are only tolerant to the hydrocarbons dissolved in the culture medium, and thus may use efflux pump to get rid of the hydrocarbons that reach the cell interior, or they need more incubation time to initiate metabolic degradation activity. The presence of ORFs for transport proteins, such as ABC-type transport system, helps to strengthen the hypothesis that clones were actually tolerant to high concentrations of hydrocarbons in the culture medium. The tolerance phenotype has been previously reported in *Pseudomonas putida*, which was able to thrive in the presence of high concentrations of aromatic hydrocarbons (Kimura 1980; Ramos et al. 1995; Huertas et al. 1998; Roma-Rodrigues et al. 2010). During the evolutionary process, the bacteria developed a series of physiological strategies and genetic control of tolerance to toxic organic compounds as: activation of stress response, membrane impermeability mechanisms and molecular control of efflux pumps (Fillet et al. 2012).

So far, few studies have been published showing degradation activity of aromatic and aliphatic hydrocarbons by fosmid clones from metagenomic library (Vasconcellos et al. 2010; Silva et al. 2013a; Sierra-García et al. 2014). In this study, the low degradation activity observed for clones 03, 05, 06, 07 and 09 may be linked to the use of heterologous expression systems in *E. coli*, together with the fact that, generally, aromatic hydrocarbons are more difficult to degrade than aliphatic organic compounds (Uchiyama and Miyazaki 2009). Perhaps, the use of more efficient heterologous expression systems, hosted in *Pseudomonas putida*, added to a longer incubation time, could produce better results (Loeschcke and Thies 2015).

Phylogenetic considerations about genes for dioxigenases

Analysis of the complete genome of *Thermus thermophilus* revealed the presence of genes for enzymes involved in the degradation of phenylacetate, such as ring-cleaving catechol dioxygenases (Henne et al. 2004). The presence of genes encoding enzymes tolerant to saline environments, observed in halotolerant bacteria, is consistent with the saline character of the mangrove environment, which governs the selection

of specific microbial groups (Piedad Díaz et al. 2000; Andreote et al. 2012). The recovery of such AH-degradation genes in distinct clusters, distantly related to known sequences, strongly suggests that they correspond to new protein sequences.

The lack of pattern in the distribution of phyla along the topology of the phylogenetic trees indicates the occurrence of horizontal gene transfer among species of distinct taxonomic groups. The low similarity between most contigs and their best hits (Supplementary Table S3), retrieved from the database, suggests the presence of new genes in the mangrove sediment, making it an interesting target for understanding the metabolic processes responsible for environmental hydrocarbon degradation and carbon cycling. Similar results with respect to the divergences in the taxonomic distribution pattern of genes for dioxigenases in the phylogenetic tree were obtained by de Sousa et al. (2017).

In conclusion, functional and taxonomic investigation of the metagenome showed that the microbial community found in the oil-impacted mangrove sediment has the genetic capacity to metabolize *in-situ* aromatic hydrocarbons. It became clear that species of classes Alphaproteobacteria, Actinobacteria, Betaproteobacteria, Gammaproteobacteria e Deltaproteobacteria have genetic potential and possibly play a role in the following pathways involved in hydrocarbon degradation in the mangrove sediment: Anaerobic degradation of aromatic compounds; Metabolism of central aromatic intermediates; e Peripheral pathways for catabolism of aromatic compounds. Data analysis also showed the potential of the metagenomic library for prospecting new genes involved in the degradation of aromatic compounds. More details on the genetic structure and functional potential were obtained from sequencing and assembly of large metagenomic fragments extracted from fosmid clones with hydrocarbon degradation ability. The integration of the results obtained by the structural and functional characterization of the fosmid inserts with the results achieved in the GC-MS assays, suggests the occurrence of new molecular mechanisms involved in the process of tolerance and/or degradation of aromatic hydrocarbons. The occurrence of fosmid clones degrading pyrene and naphthalene can be explained by environmental enrichment of the mangrove sediment with AHs (Cabral et al. 2018). To date, this is the first report on the isolation of gene clusters harboring enzymes putatively involved in the degradation of AHs in a Brazilian mangrove impacted by oil spill. Finally, a better understanding of the pathways involved in the aromatic hydrocarbons degradation, along with the application of synthetic biology, may allow the development of important biotechnological tools in both production of industrial chemicals and treatment of contaminated environments.

Acknowledgments

The authors are grateful to the São Paulo Research Foundation - FAPESP for the financial support and scholarship (Processes no. 2011/50809-5; 2013/22555-4).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest

Human studies and informed consent This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

An D, Brown D, Chatterjee I, Dong X, Ramos-Padron E, Wilson S, Bordenave S, Caffrey SM, Gieg LM, Sensen CW, Voordouw G (2013) Microbial community and potential functional gene diversity involved in anaerobic hydrocarbon degradation and methanogenesis in an oil sands tailings pond. Genome 56:612–618 . doi: 10.1139/gen-2013-0083

Andreote FD, Jiménez DJ, Chaves D, Dias ACF, Luvizotto DM, Dini-Andreote F, Fasanella CC, Lopez MV, Baena S, Taketani RG, de Melo IS (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. PLoS One 7:e38600 . doi: 10.1371/journal.pone.0038600

Ang EL, Zhao H, Obbard JP (2005) Recent advances in the bioremediation of persistent organic pollutants via biomolecular engineering. Enzyme Microb Technol 37:487–496 . doi: https://doi.org/10.1016/j.enzmictec.2004.07.024

Arand M, Cronin A, Adamska M, Oesch FBT-M in E (2005) Epoxide Hydrolases: Structure, Function, Mechanism, and Assay. In: Phase II Conjugation Enzymes and Transport Systems. Academic Press, pp 569–588

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M (2008) The RAST Server: Rapid Annotations using Subsystems Technology BT. BMC Genomics 9:75 . doi: 10.1186/1471-2164-9-75

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5): 455–477 . doi: 10.1089/cmb.2012.0021

Barragán MJL, Carmona M, Zamarro MT, Thiele B, Boll M, Fuchs G, García JL, Díaz E (2004) The bzd Gene Cluster, Coding for Anaerobic Benzoate Catabolism, in *Azoarcus* sp. Strain CIB. J Bacteriol 186:5762–5774 . doi: 10.1128/JB.186.17.5762-5774.2004

Bashir Y, Singh SP, Konwar BK (2014) Metagenomics: An Application Based Perspective. Chinese J Biol . doi: 10.1155/2014/146030

Bicalho B, Goncalves RAC, Zibordi APM, Manfio GP, Marsaioli AJ (2003) Antimicrobial compounds of fungi vectored by *Clusia* spp. (*Clusiaceae*) pollinating bees. Z Naturforsch C 58:746–751

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120 . doi: 10.1093/bioinformatics/btu170

Bushnell LD, Haas HF (1941) The Utilization of Certain Hydrocarbons by Microorganisms . J Bacteriol 41:653–673

Busto E, Gotor-Fernandez V, Gotor V (2010) Hydrolases: catalytically promiscuous enzymes for non-conventional reactions in organic synthesis. Chem Soc Rev 39:4504–4523 . doi: 10.1039/C003811C

Carmona M, Zamarro MT, Blázquez B, Durante-Rodríguez G, Juárez JF, Valderrama JA, Barragán MJL, García JL, Díaz E (2009) Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. Microbiol Mol Biol Rev 73:71–133 . doi: 10.1128/MMBR.00021-08

Cerniglia CE (1992) Biodegradation of polycyclic aromatic hydrocarbons. Biodegradation 3:351–368 . doi: 10.1007/BF00129093

de Sousa STP, Cabral L, Lacerda Júnior GV, Oliveira VM (2017) Diversity of aromatic hydroxylating dioxygenase genes in mangrove microbiome and their biogeographic patterns across global sites. Microbiologyopen 6:e00490 . doi: 10.1002/mbo3.490

Denef VJ, Patrauchan MA, Florizone C, Park J, Tsoi T V, Verstraete W, Tiedje JM, Eltis LD (2005) Growth Substrate- and Phase-Specific Expression of Biphenyl, Benzoate, and C1 Metabolic Pathways in *Burkholderia xenovorans* LB400. J Bacteriol 187:7996–8005 . doi: 10.1128/JB.187.23.7996-8005.2005

Fillet S, Daniels C, Pini C, Krell T, Duque E, Bernal P, Segura A, Lu D, Zhang X, Ramos J-L (2012) Transcriptional control of the main aromatic hydrocarbon efflux pump in *Pseudomonas*. Environ Microbiol Rep 4:158–167 . doi: 10.1111/j.1758-2229.2011.00255.x

Fuchs G (2008) Anaerobic Metabolism of Aromatic Compounds. Ann N Y Acad Sci 1125:82–99 . doi: 10.1196/annals.1419.010

Fuchs G, Boll M, Heider J (2011) Microbial degradation of aromatic compounds — from one strategy to four. Nat Rev Microbiol 9:803–816 . doi: 10.1038/nrmicro2652

Gallego S, Vila J, Tauler M, Nieto JM, Breugelmans P, Springael D, Grifoll M (2014) Community structure and PAH ring-hydroxylating dioxygenase genes of a marine pyrene-degrading microbial consortium. Biodegradation 25:543–556 . doi: 10.1007/s10532-013-9680-z

Ghosal D, Ghosh S, Dutta TK, Ahn Y (2016) Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. Front Microbiol 7:1369 . doi: 10.3389/fmicb.2016.01369

Gibson J, S. Harwood C (2002) Metabolic Diversity in Aromatic Compound Utilization by Anaerobic Microbes. Annu Rev Microbiol 56:345–369 . doi: 10.1146/annurev.micro.56.012302.160749

Gkorezis P, Daghio M, Franzetti A, Van Hamme JD, Sillen W, Vangronsveld J (2016) The Interaction between Plants and Bacteria in the Remediation of Petroleum Hydrocarbons: An Environmental Perspective. Front Microbiol 7:1836 . doi: 10.3389/fmicb.2016.01836

Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc Natl Acad Sci U S A 100:8298–8303 . doi: 10.1073/pnas.1431443100

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644–652

Guazzaroni M-E, Herbst F-A, Lores I, Tamames J, Peláez AI, López-Cortés N, Alcaide M, Del Pozo M V, Vieites JM, von Bergen M, Gallego JLR, Bargiela R, López-López A, Pieper DH, Rosselló-Móra R, Sánchez J, Seifert J, Ferrer M (2013) Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. ISME J 7:122–136 . doi: 10.1038/ismej.2012.82

Haritash AK, Kaushik CP (2009) Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. J Hazard Mater 169:1–15

Henne A, Brüggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, Jacobi C, Starkuviene V, Schlenczeck S, Dencker S, Huber R, Klenk H-P, Kramer W, Merkl R, Gottschalk G, Fritz H-J (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. Nat Biotechnol 22(5):547-53

Hennessee CT, Li QX (2016) Effects of Polycyclic Aromatic Hydrocarbon Mixtures on Degradation, Gene Expression, and Metabolite Production in Four *Mycobacterium* Species. Appl Environ Microbiol 82:3357–3369 . doi: 10.1128/AEM.00100-16

Holman H-YN, Tsang YW, Holman WR (1999) Mineralization of Sparsely Water-Soluble Polycyclic Aromatic Hydrocarbons in a Water Table Fluctuation Zone. Environ Sci Technol 33:1819–1824 . doi: 10.1021/es980701k

Huertas M-J, Duque E, Marqués S, Ramos JL (1998) Survival in Soil of Different Toluene-Degrading *Pseudomonas* Strains after Solvent Shock. Appl Environ Microbiol 64:38–42

Hugenholtz P, Tyson GW (2008) Microbiology: Metagenomics. Nature 455:481–483

Jia B, Jia X, Hyun Kim K, Ji Pu Z, Kang M-S, Ok Jeon C (2017) Evolutionary, computational, and biochemical studies of the salicylaldehyde dehydrogenases in the naphthalene degradation pathway. Sci Rep 7:43489 . doi: 10.1038/srep43489

Johnsen AR, Bendixen K, Karlson U (2002) Detection of Microbial Growth on Polycyclic Aromatic Hydrocarbons in Microtiter Plates by Using the Respiration Indicator WST-1. Appl Environ Microbiol 68:2683–2689 . doi: 10.1128/AEM.68.6.2683-2689.2002

Jouanneau Y, Martin F, Krivobok S, Willison JC (2011) Ring-hydroxylating Dioxygenases Involved in PAH Biodegradation: Structure, Function, and Biodiversity. In: Microbial Bioremediation of Non-metals: Current Research. pp 149–175

Juhasz AL, Stanley GA, Britz ML (2000) Microbial degradation and detoxification of high molecular weight polycyclic aromatic hydrocarbons by *Stenotrophomonas maltophilia* strain VUN 10,003. Lett Appl Microbiol 30:396–401 . doi: 10.1046/j.1472-765x.2000.00733.x

Keum Y-S, Seo J-S, Hu Y, Li QX (2006) Degradation pathways of phenanthrene by *Sinorhizobium* sp. C4. Appl Microbiol Biotechnol 71:935–941 . doi: 10.1007/s00253-005-0219-z

Kiel C, Yus E, Serrano L (2010) Engineering Signal Transduction Pathways. Cell 140:33–47 . doi: https://doi.org/10.1016/j.cell.2009.12.028

Kim S-J, Jones RC, Cha C-J, Kweon O, Edmondson RD, Cerniglia CE (2004) Identification of proteins induced by polycyclic aromatic hydrocarbon in *Mycobacterium vanbaalenii* PYR-1 using two-dimensional polyacrylamide gel electrophoresis and de novo sequencing methods. Proteomics 4:3899–3908 . doi: 10.1002/pmic.200400872

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120 . doi: 10.1007/BF01731581

Kloer DP, Hagel C, Heider J, Schulz GE (2017) Crystal Structure of Ethylbenzene Dehydrogenase from *Aromatoleum aromaticum*. Structure 14:1377–1388 . doi: 10.1016/j.str.2006.07.001

Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, Delgardio J, Norton N, Hazen TC, Huettel M (2011) Hydrocarbon-Degrading Bacteria and the Bacterial Community Response in Gulf of Mexico Beach Sands Impacted by the Deepwater Horizon Oil Spill. Appl Environ Microbiol

77:7962–7974 . doi: 10.1128/AEM.05402-11

Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA, Beatty JT, Lang AS, Tabita FR, Gibson JL, Hanson TE, Bobst C, Torres JLT y, Peres C, Harrison FH, Gibson J, Harwood CS (2004) Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat Biotech 22:55–61

Lewin GR, Carlos C, Chevrette MG, Horn HA, McDonald BR, Stankey RJ, Fox BG, Currie CR (2016) Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. Annu Rev Microbiol 70:235–254 . doi: 10.1146/annurev-micro-102215-095748

Liang Y, Van Nostrand JD, Deng Y, He Z, Wu L, Zhang X, Li G, Zhou J (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. Isme J 5:403–413 . doi: 10.1038/ismej.2010.142

Lim LH, Harrison RM, Harrad S (1999) The Contribution of Traffic to Atmospheric Concentrations of Polycyclic Aromatic Hydrocarbons. Environ Sci Technol 33:3538–3542 . doi: 10.1021/es990392d

Lipscomb JD (2008) Mechanism of extradiol aromatic ring-cleaving dioxygenases. Curr Opin Struct Biol 18:644–649 . doi: https://doi.org/10.1016/j.sbi.2008.11.001

Liu Z, Liu J (2013) Evaluating bacterial community structures in oil collected from the sea surface and sediment in the northern Gulf of Mexico after the Deepwater Horizon oil spill. Microbiologyopen 2:492–504 . doi: 10.1002/mbo3.89

Loeschcke A, Thies S (2015) *Pseudomonas putida*—a versatile host for the production of natural products. Appl Microbiol Biotechnol 99:6197–6214 . doi: 10.1007/s00253-015-6745-4

Lu XY, Zhang T, Fang HHP (2011) Bacteria-mediated PAH degradation in soil and sediment. Appl Microbiol Biotechnol 89:1357–1371

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963 . doi: 10.1093/bioinformatics/btr507

Mallick S, Chakraborty J, Dutta TK (2011) Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular-weight polycyclic aromatic hydrocarbons: A review. Crit Rev Microbiol 37:64–90 . doi: 10.3109/1040841X.2010.512268

Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res 42:D568–D573

Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Balum J, Kimbrel J, Bouskill NJ, Prestat E, Borglin S, Joyner DC, Fortney JL, Jurelevicius D, Stringfellow WT, Alvarez-Cohen L, Hazen TC, Knight R, Gilbert JA, Jansson JK (2014) Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J 8:1464–1475

Megharaj M, Ramakrishnan B, Venkateswarlu K, Sethunathan N, Naidu R (2011) Bioremediation approaches for organic pollutants: A critical perspective. Environ Int 37:1362–1375 . doi: https://doi.org/10.1016/j.envint.2011.06.003

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27:29–34

Ohkouchi N, Kawamura K, Kawahata H (1999) Distributions of Three- to Seven-Ring Polynuclear Aromatic Hydrocarbons on the Deep Sea Floor in the Central Pacific. Environ Sci Technol 33:3086–3090 . doi: 10.1021/es981181w

Ottoni JR, Cabral L, de Sousa STP, Júnior GVL, Domingos DF, Soares Junior FL, da Silva MCP, Marcon J, Dias ACF, de Melo IS, de Souza AP, Andreote FD, de Oliveira VM (2017) Functional metagenomics of oil-impacted mangrove sediments reveals high abundance of hydrolases of biotechnological interest. World J Microbiol Biotechnol 33:141 . doi: 10.1007/s11274-017-2307-5

Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986) The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences BT - Advances in Microbial Ecology. In: Marshall KC (ed). Springer US, Boston, MA, pp 1–55

Parales RE, Resnick SM (2006) Aromatic Ring Hydroxylating Dioxygenases. In: Ramos J-L, Levesque RC (eds) *Pseudomonas*: Volume 4 Molecular Biology of Emerging Issues. Springer US, Boston, MA, pp 287–340

Pauling L, Wheland GW (1933) The Nature of the Chemical Bond. V. The Quantum-Mechanical Calculation of the Resonance Energy of Benzene and Naphthalene and the Hydrocarbon Free Radicals. J Chem Phys 1:362–374 . doi: 10.1063/1.1749304

Piedad Díaz M, Grigson SJW, Peppiatt CJ, Burgess JG (2000) Isolation and Characterization of Novel Hydrocarbon-Degrading Euryhaline Consortia from Crude Oil and Mangrove Sediments. Mar Biotechnol 2:522–532 . doi: 10.1007/s101260000037

Popovic A, Hai T, Tchigvintsev A, Hajighasemi M, Nocek B, Khusnutdinova AN, Brown G, Glinos J, Flick R, Skarina T, Chernikova TN, Yim V, Brüls T, Paslier D Le, Yakimov MM, Joachimiak A, Ferrer M, Golyshina O V, Savchenko A, Golyshin PN, Yakunin AF (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci Rep 7:44103 . doi: 10.1038/srep44103

Rabus R, Boll M, Heider J, Meckenstock RU, Buckel W, Einsle O, Ermler U, Golding BT, Gunsalus RP, Kroneck PMH, Krüger M, Lueders T, Martins BM, Musat F, Richnow HH, Schink B, Seifert J, Szaleniec M, Treude T, Ullmann GM, Vogt C, von Bergen M, Wilkes H (2016) Anaerobic Microbial Degradation of Hydrocarbons: From Enzymatic Reactions to the Environment. J Mol Microbiol Biotechnol 26:5–28

Rabus R, Heider J (1998) Initial reactions of anaerobic metabolism of alkylbenzenes in denitrifying and sulfate-reducing bacteria. Arch Microbiol 170:377–384 . doi: 10.1007/s002030050656

Rabus R, Kube M, Heider J, Beck A, Heitmann K, Widdel F, Reinhardt R (2005) The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. Arch Microbiol 183:27–36 . doi: 10.1007/s00203-004-0742-9

Ramos JL, Duque E, Huertas MJ, Haïdour A (1995) Isolation and expansion of the catabolic potential of a *Pseudomonas putida* strain able to grow in the presence of high concentrations of aromatic hydrocarbons. J Bacteriol 177:3911–3916

Rengarajan T, Rajendran P, Nandakumar N, Lokeshkumar B, Rajendran P, Nishigaki I (2015) Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. Asian Pac J Trop Biomed 5:182–189 . doi: 10.1016/S2221-1691(15)30003-4

Roma-Rodrigues C, Santos PM, Benndorf D, Rapp E, Sá-Correia I (2010) Response of *Pseudomonas putida* KT2440 to phenol at the level of membrane proteome. J Proteomics 73:1461–1478 . doi: https://doi.org/10.1016/j.jprot.2010.02.003

Roongsawang N, Washio K, Morikawa M (2011) Diversity of Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Lipopeptide Biosurfactants. Int J Mol Sci 12:141–172 . doi: 10.3390/ijms12010141

Segura A, Rodríguez-Conde S, Ramos C, Ramos JL (2009) Bacterial responses and interactions with plants during rhizoremediation. Microb Biotechnol 2:452–464 . doi: 10.1111/j.1751-7915.2009.00113.x

Semple KT, Reid BJ, Fermor TR (2001) Impact of composting strategies on the treatment of soils contaminated with organic pollutants. Environ Pollut 112:269–283 . doi: https://doi.org/10.1016/S0269-7491(00)00099-3

Seo J-S, Keum Y-S, Li QX (2009) Bacterial degradation of aromatic compounds. Int J Environ Res Public Health 6:278–309 . doi: 10.3390/ijerph6010278

Sherry A, Gray ND, Ditchfield AK, Aitken CM, Jones DM, Röling WFM, Hallmann C, Larter SR, Bowler BFJ, Head IM (2013) Anaerobic biodegradation of crude oil under sulphate-reducing conditions leads to only modest enrichment of recognized sulphate-reducing taxa. Int Biodeterior Biodegradation 81:105–113 . doi: https://doi.org/10.1016/j.ibiod.2012.04.009

Sierra-García IN, Alvarez JC, Vasconcellos SP de, Souza AP de, Neto EV dos S, de Oliveira VM (2014) New Hydrocarbon Degradation Pathways in the Microbial Metagenome from Brazilian Petroleum Reservoirs. PLoS One 9:e90087

Silva CC, Hayden H, Sawbridge T, Mele P, De Paula SO, Silva LCF, Vidigal PMP, Vicentini R, Sousa MP, Torres APR, Santiago VMJ, Oliveira VM (2013a) Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. PLoS One 8:e61811

Silva TR, Verde LCL, Santos Neto E V., Oliveira VM (2013b) Diversity analyses of microbial communities in petroleum samples from Brazilian oil fields. Int Biodeterior Biodegrad 81:57–70 . doi: 10.1016/j.ibiod.2012.05.005

Singh S, Kang SH, Mulchandani A, Chen W (2008) Bioremediation: environmental clean-up through pathway engineering. Curr Opin Biotechnol 19:437–444 . doi: https://doi.org/10.1016/j.copbio.2008.07.012

Song B, Ward BB (2005) Genetic Diversity of Benzoyl Coenzyme A Reductase Genes Detected in Denitrifying Isolates and Estuarine Sediment Communities. Appl Environ Microbiol 71:2036–2045 . doi: 10.1128/AEM.71.4.2036-2045.2005

Steele HL, Streit WR (2005) Metagenomics: Advances in ecology and biotechnology. FEMS Microbiol Lett 247:105–111

Straube WL, Nestler CC, Hansen LD, Ringleberg D, Pritchard PH, Jones-Meehan J (2003) Remediation of Polyaromatic Hydrocarbons (PAHs) through Landfarming with Biostimulation and Bioaugmentation. Acta Biotechnol 23:179–196 . doi: 10.1002/abio.200390025

Suenaga H, Koyama Y, Miyakoshi M, Miyazaki R, Yano H, Sota M, Ohtsubo Y, Tsuda M, Miyazaki K

(2009) Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. ISME J 3:1335–1348

Tan B, Jane Fowler S, Laban NA, Dong X, Sensen CW, Foght J, Gieg LM (2015) Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. ISME J 9:2028–2045 . doi: 10.1038/ismej.2015.22

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36

Tsai YL, Olson BH (1991) Rapid method for direct extraction of DNA from soil and sediments. Appl Environ Microbiol 57:1070–1074

Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol 20:616–622 . doi: https://doi.org/10.1016/j.copbio.2009.09.010

Vaillancourt FH, Bolin JT, Eltis LD (2006) The Ins and Outs of Ring-Cleaving Dioxygenases. Crit Rev Biochem Mol Biol 41:241–267 . doi: 10.1080/10409230600817422

Vasconcellos SP de, Angolini CFF, García INS, Martins Dellagnezze B, Silva CC da, Marsaioli AJ, Neto EV dos S, de Oliveira VM (2010) Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. Org Geochem 41:675–681 . doi: https://doi.org/10.1016/j.orggeochem.2010.03.014

Velázquez F, De Lorenzo V, Valls M (2006) The m-xylene biodegradation capacity of *Pseudomonas putida* mt-2 is submitted to adaptation to abiotic stresses: evidence from expression profiling of xyl genes. Environ Microbiol 8:591–602 . doi: 10.1111/j.1462-2920.2005.00936.x

Verde LCL, Silva TR e, Dellagnezze BM, Neto EV dos S, Oliveira VM de (2013) Diversity of Hydrocarbon-Related Catabolic Genes in Oil Samples from Potiguar Basin (Rn, Brazil). J Pet Env Biotechnol 4:138 . doi: 10.4172/2157- 7463.1000138

Vila J, Tauler M, Grifoll M (2015) Bacterial PAH degradation in marine and terrestrial habitats. Curr Opin Biotechnol 33:95–102 . doi: https://doi.org/10.1016/j.copbio.2015.01.006

Viñas M, Sabaté J, Espuny MJ, Solanas AM (2005) Bacterial Community Dynamics and Polycyclic Aromatic Hydrocarbon Degradation during Bioremediation of Heavily Creosote-Contaminated Soil. Appl Environ Microbiol 71:7008–7018 . doi: 10.1128/AEM.71.11.7008-7018.2005

von Netzer F, Kuntze K, Vogt C, Richnow HH, Boll M, Lueders T (2016) Functional Gene Markers for Fumarate-Adding and Dearomatizing Key Enzymes in Anaerobic Aromatic Hydrocarbon Degradation in Terrestrial Environments. J Mol Microbiol Biotechnol 26:180–194

Weid V Der, Irene, Korenblum E, Jurelevicius D, Rosado AS, Dino R, SebastIan GV, Seldin L (2008) Molecular Diversity of Bacterial Communities from Subseafloor Rock Samples in a Deep-Water Production Basin in Brazil. J Microbiol Biotechnol 18:5–14

Yang S, Wen X, Zhao L, Shi Y, Jin H (2014) Crude Oil Treatment Leads to Shift of Bacterial Communities in Soils from the Deep Active Layer and Upper Permafrost along the China-Russia Crude Oil Pipeline Route. PLoS One 9:e96552

Zhou HW, Luan TG, Zou F, Tam NFY (2008) Different bacterial groups for biodegradation of three- and four-ring PAHs isolated from a Hong Kong mangrove sediment. J Hazard Mater 152:1179–1185 . doi: https://doi.org/10.1016/j.jhazmat.2007.07.116

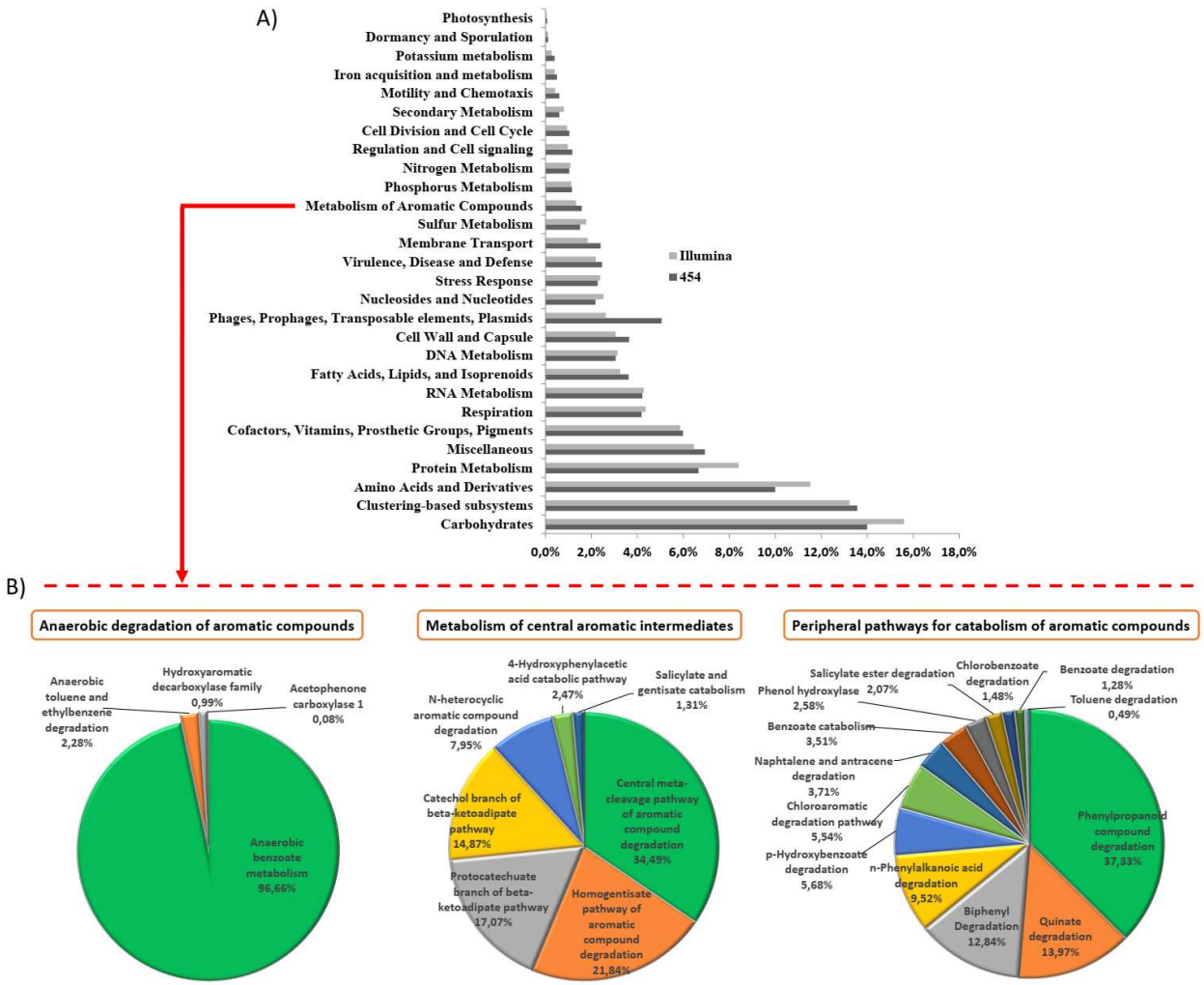


Fig. 1 A) Functional annotation of metagenomic reads from the fosmid library against the COG database. B) Subsystem for the metabolism of aromatic compounds by COG database. Red arrow indicates the target category of this study

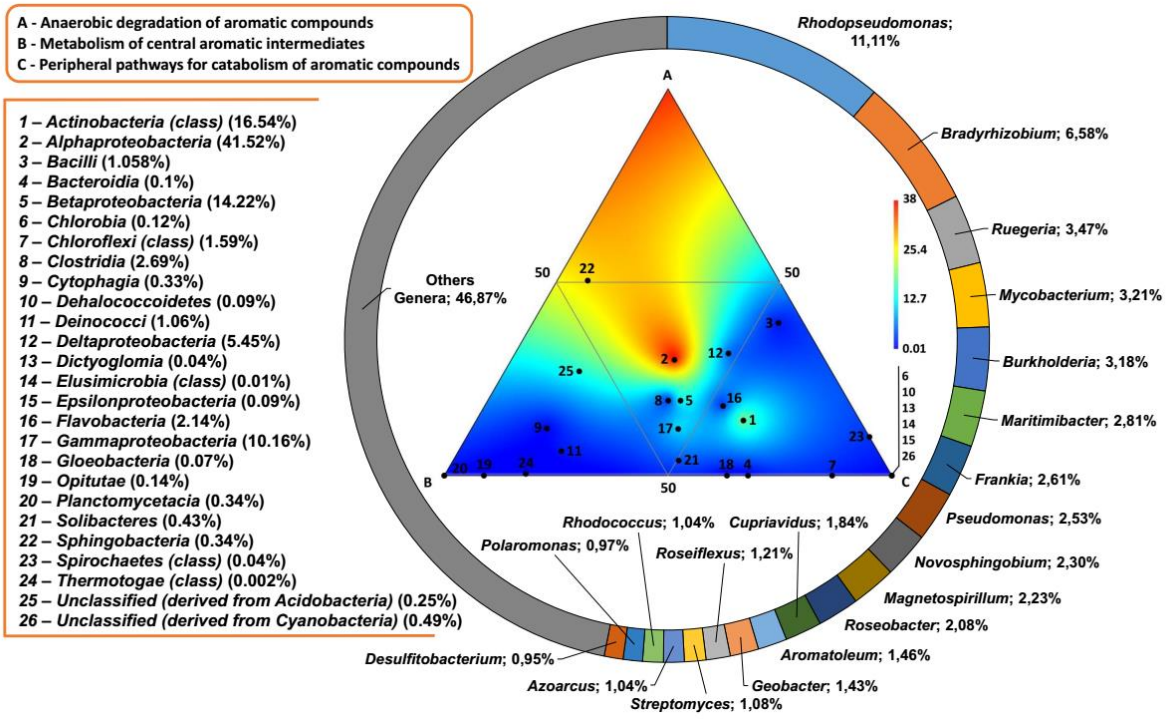


Fig. 2 General taxonomic annotation. On the left, relative abundance of taxonomic classes based on the annotated sequences for the metabolism of aromatic compounds. Donut chart shows the relative abundance of microbial genera. Distribution of taxonomic groups in “Anaerobic degradation of aromatic compounds” (A), “Metabolism of central aromatic intermediates” (B) and “Peripheral pathways for catabolism of aromatic compounds” (C) is visualized in the ternary plot. The location of each point on the graph shows the relative abundance of each taxonomic group among the three functional categories; the color variation on the graph surface represents the relative abundance of taxa

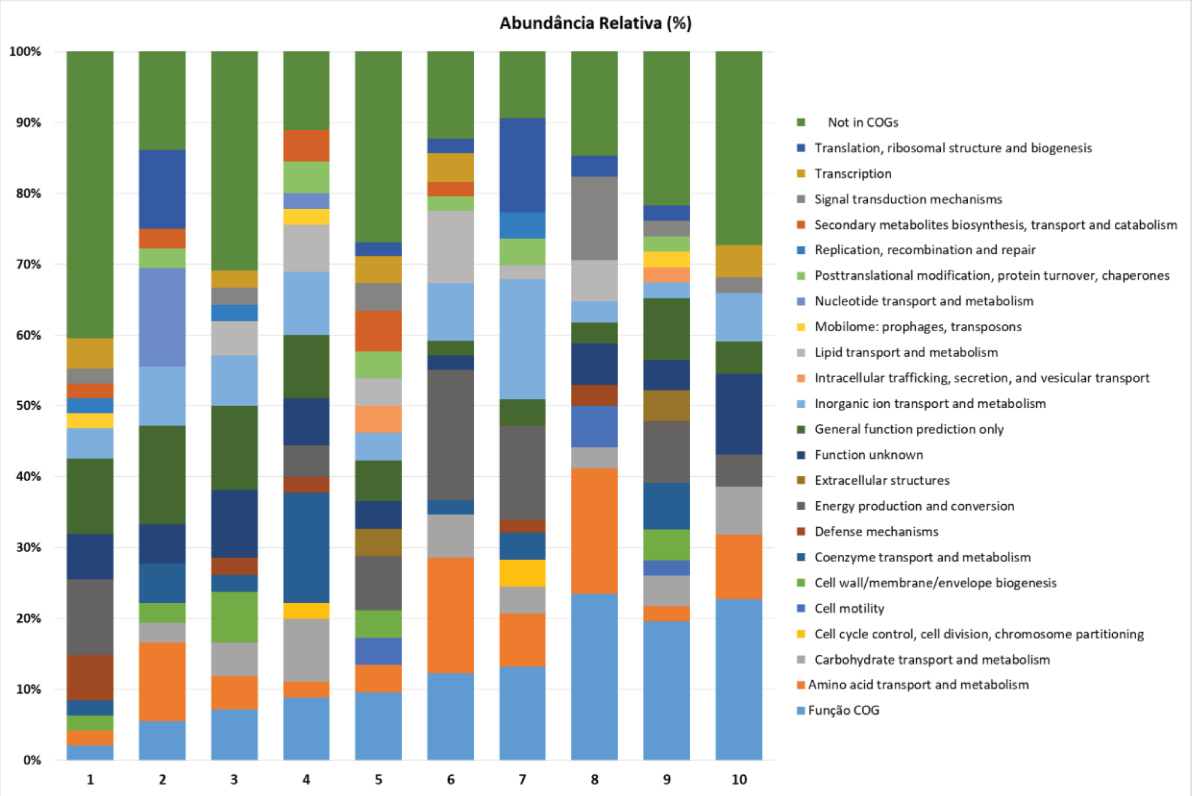


Fig. 3 Relative abundance of ORFs in each one of the ten fosmid clones based on the functional classification against the COG database

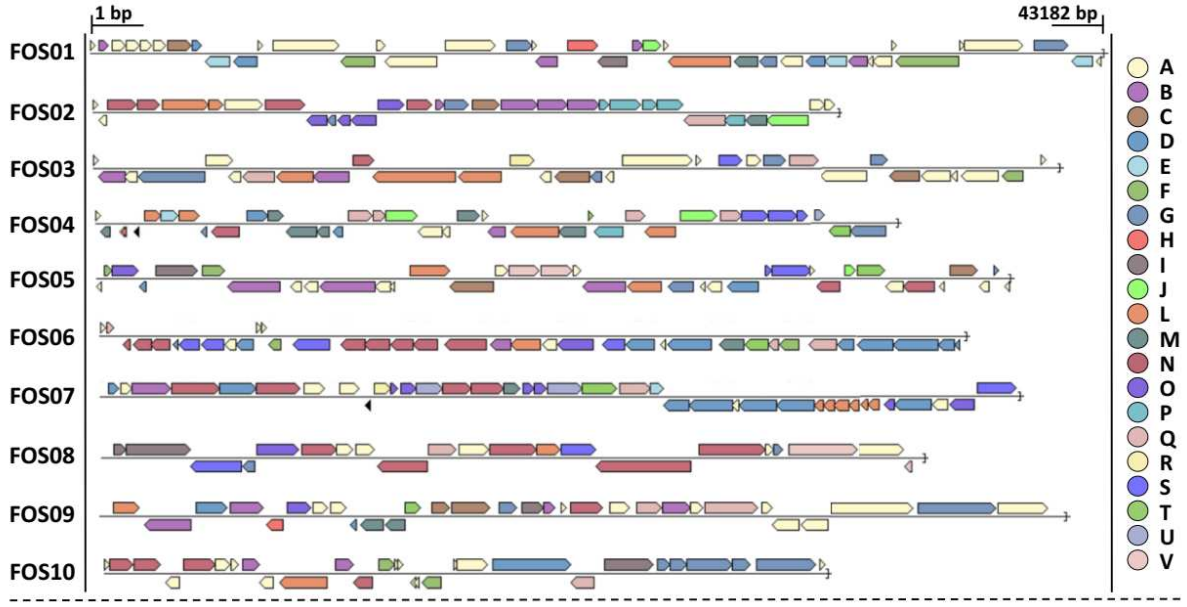
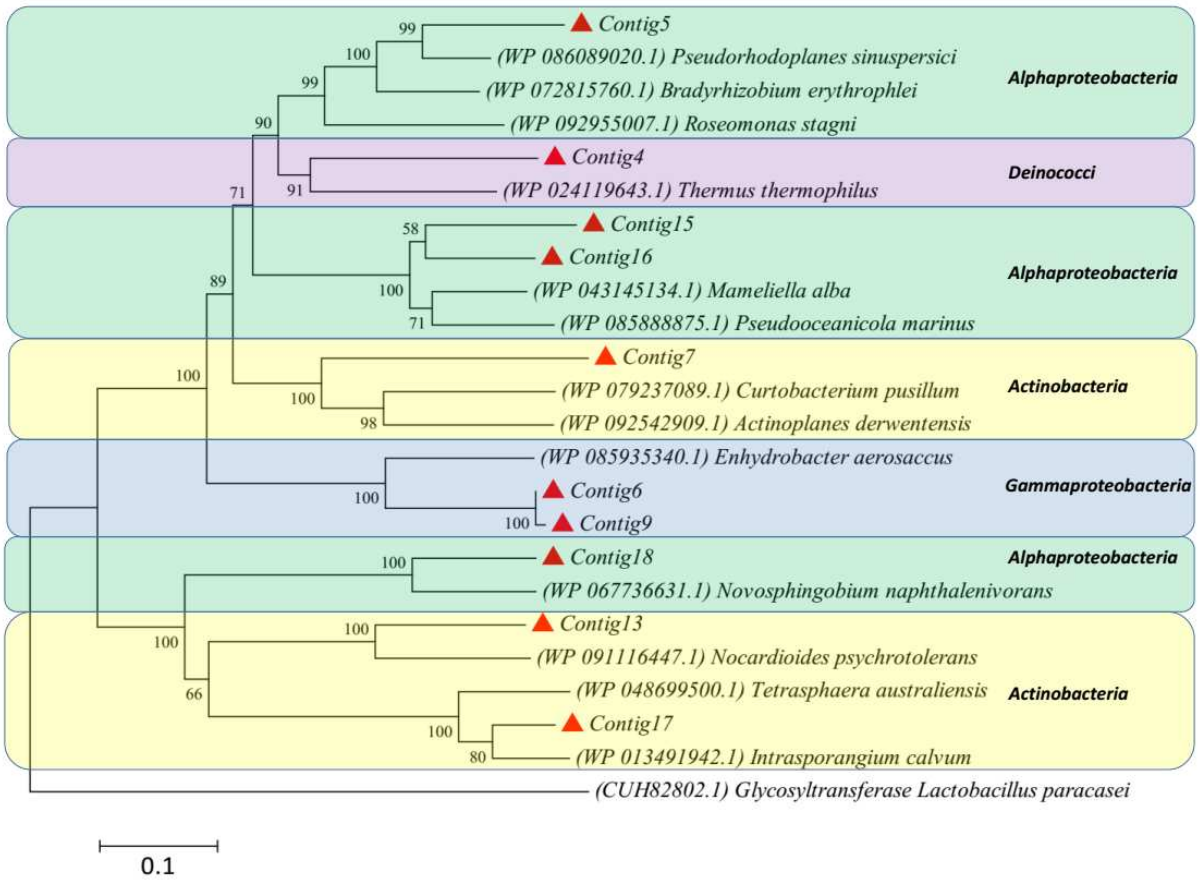
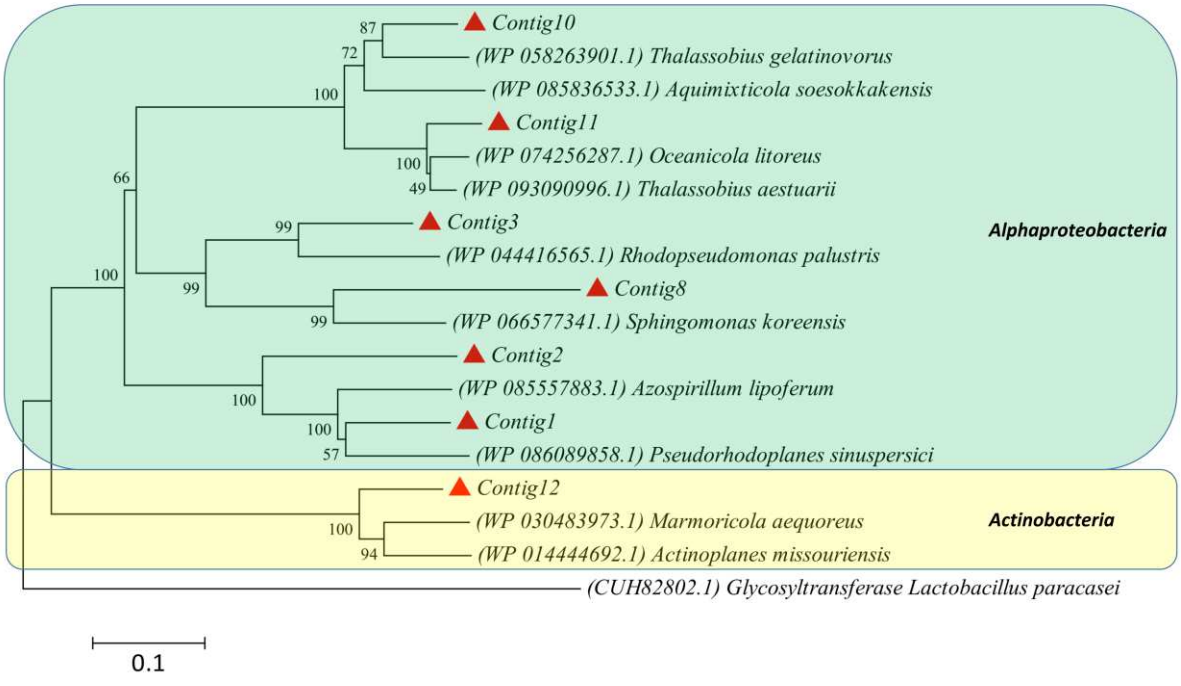


Fig. 4 Schematic representation of the gene organization in the ten fosmid inserts and functional classification against the COG database. Arrows represent open reading frames (ORFs) and letters and colors represent functional categories. Genes of unknown function (A); General function prediction only (B); Cell wall/membrane/envelope biogenesis (C); Energy production and conversion (D); Defense mechanisms (E); Transcription (F); Function unknown (G); Mobilome: prophages, transposons (H); Signal transduction mechanisms (I); Secondary metabolites biosynthesis, transport and catabolism (J); Inorganic ion transport and metabolism (L); Coenzyme transport and metabolism (M); Amino acid transport and metabolism (N); Translation, ribosomal structure and biogenesis (O); Nucleotide transport and metabolism (P); Carbohydrate transport and metabolism (Q); Replication, recombination and repair (R); Lipid transport and metabolism (S); Posttranslational modification, protein turnover, chaperones (T); Cell cycle control, cell division, chromosome partitioning (U); Cell motility (V)

A)


Contig5
(WP 086089020.1) Pseudorhodoplanes sinuspersici
(WP 072815760.1) Bradyrhizobium erythrophlei
(WP 092955007.1) Roseomonas stagni
Alphaproteobacteria
Contig4
(WP 024119643.1) Thermus thermophilus
Deinococci
Contig15
Contig16
(WP 043145134.1) Mameliella alba
(WP 085888875.1) Pseudooceanicola marinus
Alphaproteobacteria
Contig7
(WP 079237089.1) Curtobacterium pusillum
(WP 092542909.1) Actinoplanes derwentensis
Actinobacteria
(WP 085935340.1) Enhydrobacter aerosaccus
Contig6
Contig9
Gammaproteobacteria
Contig18
(WP 067736631.1) Novosphingobium naphthalenivorans
Alphaproteobacteria
Contig13
(WP 091116447.1) Nocardioides psychrotolerans
(WP 048699500.1) Tetrasphaera australiensis
Contig17
(WP 013491942.1) Intrasporangium calvum
Actinobacteria
(CUH82802.1) Glycosyltransferase Lactobacillus paracasei
0.1

B)


Contig10
(WP 058263901.1) Thalassobius gelatinovorus
(WP 085836533.1) Aquimixticola soesokkakensis
Contig11
(WP 074256287.1) Oceanicola litoreus
(WP 093090996.1) Thalassobius aestuarii
Contig3
(WP 044416565.1) Rhodopseudomonas palustris
Contig8
(WP 066577341.1) Sphingomonas koreensis
Contig2
(WP 085557883.1) Azospirillum lipoferum
Contig1
(WP 086089858.1) Pseudorhodoplanes sinuspersici
Alphaproteobacteria
Contig12
(WP 030483973.1) Marmoricola aequoreus
(WP 014444692.1) Actinoplanes missouriensis
Actinobacteria
(CUH82802.1) Glycosyltransferase Lactobacillus paracasei
0.1

C)

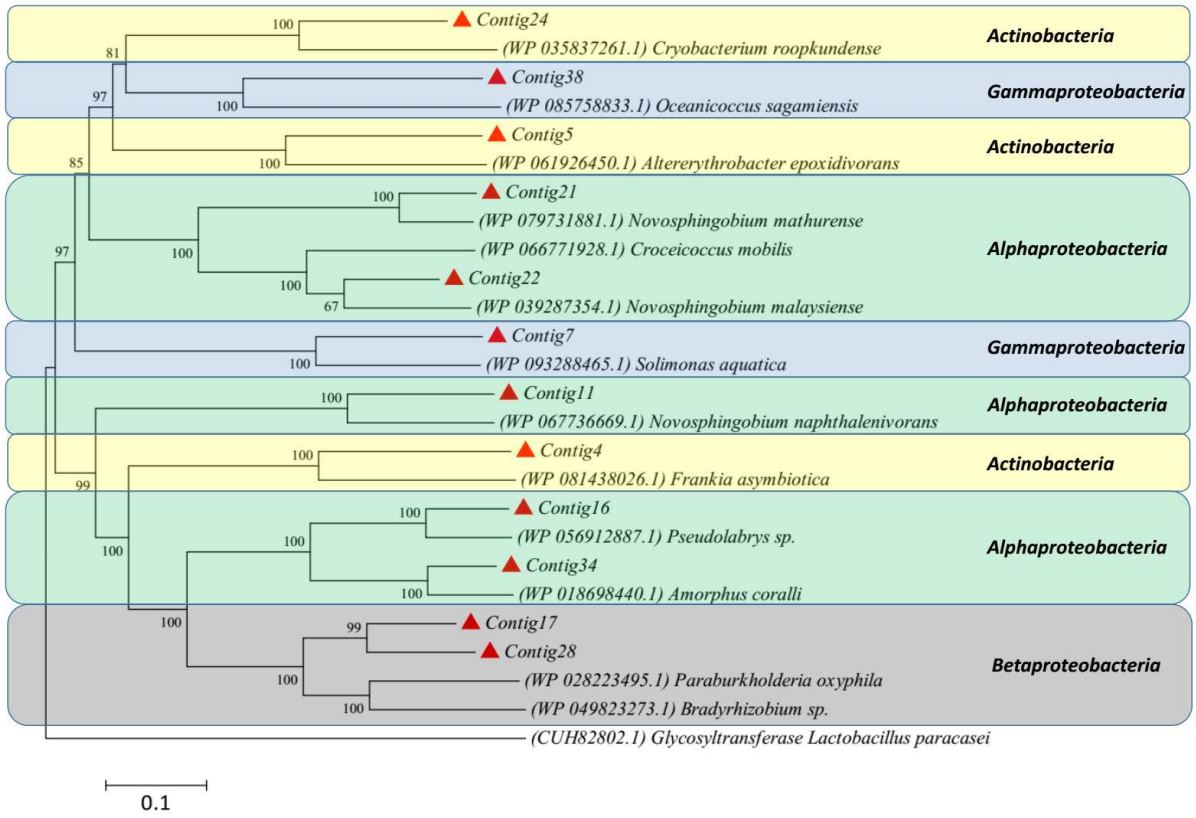


Fig. 5 Phylogenetic trees representing the relationships among the contigs for Catechol dioxygenase (A), Protocatechuate dioxygenase (B) and Aromatic ring-hydroxylating dioxygenase (C) genes recovered from fosmid metagenomic library of mangrove sediment and the most related hits from the GenBank database. Bootstrap values were obtained from 1,000 replicates. The red triangle represents the assembled contigs. The black scale bar represents 10% sequence divergence. The sequence of the glycosyltransferase from *Lactobacillus paracasei* was used as an outgroup

Table 1. General statistical information of the sequencing data processing from metagenomic library by MG-RAST platform.

Properties	Library_Oil_Illumina	Library_Oil_454
Upload: bp count (kb)	12,205,480	428,814
Upload: sequences count	141,295,265	930,817
Upload: mean sequence length (bp)	86± 23	460± 96
Upload: mean GC percent (%)	57± 9	60± 7
Artifcial duplicate reads: sequence count	49,872,747	218,014
Post QC: bp count (kb)	7,729,678	315,309
Post QC: sequences count	88,497,264	650,693
Post QC: mean sequence length (bp)	87± 22	484± 46
Post QC: mean GC percent (%)	58± 8	60± 7
Processed: predicted protein features	64,410,637	728,087
Processed: predicted rRNA features	20,108,996	62,382
Alignment: identifed protein features	9,462,472	410,079
Alignment: identifed rRNA features	5,517	3,201
Annotation: identifed functional categories	7,360,887	337,066

Table 2. Number of sequences assigned to dioxygenases involved in the metabolism of aromatic compounds by the SEED database.

Function	Nº of sequences
Metabolism of central aromatic intermediates	**9782**
Catechol 2,3-dioxygenase (EC 1.13.11.2)	2482
Protocatechuate 4,5-dioxygenase beta chain (EC 1.13.11.8)	2080
Homogentisate 1,2-dioxygenase (EC 1.13.11.5)	1694
Protocatechuate 3,4-dioxygenase beta chain (EC 1.13.11.3)	1620
4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	615
Catechol 1,2-dioxygenase (EC 1.13.11.1)	405
Protocatechuate 4,5-dioxygenase alpha chain (EC 1.13.11.8)	389
3,4-dihydroxyphenylacetate 2,3-dioxygenase	278
Protocatechuate 3,4-dioxygenase alpha chain (EC 1.13.11.3)	114
3,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)	50
Catechol 1,2-dioxygenase 1 (EC 1.13.11.1)	36
1,2-dihydroxynaphthalene dioxygenase	13
Intradiol ring-cleavage dioxygenase (EC 1.13.11.1)	5
1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase	1
Peripheral pathways for catabolism of aromatic compounds	**4993**
Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit	3605
Lignostilbene-alpha,beta-dioxygenase and related enzymes	711
Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)	243
Benzoate 1,2-dioxygenase (EC 1.14.12.10)	214
biphenyl-2,3-diol 1,2-dioxygenase III-related protein	93
2,3-dihydroxybiphenyl 1,2-dioxygenase	71
Ortho-halobenzoate 1,2-dioxygenase alpha-ISP protein OhbB	11
Benzoate 1,2-dioxygenase alpha subunit (EC 1.14.12.10)	10
Ring hydroxylating dioxygenase, alpha subunit (EC 1.14.12.13)	9
Putative phthalate 4,5-dioxygenase oxygenase subunit (OhpA2)	8
2-chlorobenzoate 1,2-dioxygenase beta subunit (EC 1.14.12.13)	6
naphthalene dioxygenase ferredoxin	5
Large subunit naph/bph dioxygenase	3
lignostilbene-alpha,beta-dioxygenase	3
Toluate 1,2-dioxygenase alpha subunit	1
Total	**15078**

Table 3: Structural, functional and taxonomic characterization of the exogenous DNA of the metagenomic clones.

FOS ID	length (pb)	G+C-content (%)	ORFs	COG assignment	Hypothetical proteins	Lineage Percentage	Classification
FOS01	43182	62	40	21	19	65	*Bacteria*
FOS02	30963	68	30	25	5	60	*Actinobacteria* (Class)
FOS03	39017	58	31	18	13	55	*Rhodopirellula* (Genus)
FOS04	33537	64	37	31	6	51	*Rhizobiales* (Order)
FOS05	37537	59	37	23	14	73	*Bacteria*
FOS06	33860	59	37	31	6	57	*Firmicutes* (Phylum)
FOS07	35958	65	38	32	6	55	*Rhizobiales* (Order)
FOS08	32172	60	21	16	5	81	*Bacteria*
FOS09	37944	66	30	20	10	53	*Planctomycetales* (Order)
FOS10	27296	53	29	17	12	59	*Deltaproteobacteria* (Class)

Table 4. GC-MS analysis of hydrocarbon biodegradation by fosmid clones after 16 days of monitoring.

Fosmid clones	Degradation (%)				
	Phenol	**Naphthalene**	**Phenanthrene**	**Pyrene**	**Benzopyrene**
FOS01	98	-	-	-	-
FOS02	-	-	-	-	ND
FOS03	13.99	-	27.32	16.1	13.12
FOS04	-	ND	-	-	-
FOS05	-	-	19.82	ND	ND
FOS06	10.93	-	-	-	-
FOS07	-	-	-	ND	20,3
FOS08	-	-	-	75.00	-
FOS09	-	-	-	10.55	15.14
FOS10	-	71.00	-	-	ND
C1	ND	ND	ND	ND	ND
C2	ND	ND	ND	ND	ND

C1 = without cells
C2 = *E. coli* without fosmid
ND = Not detected
- = unvalued

Journal: Applied Microbiology and Biotechnology

Exploring the genetic potential of a fosmid metagenomic library from an oil-impacted mangrove sediment for metabolism of aromatic compounds

Sanderson Tarciso Pereira de Sousa[1], Lucélia Cabral[1], Gileno Vieira Lacerda Júnior[1], Melline Fontes Noronha[1], Júlia Ronzella Ottoni[1], Adilson Sartoratto[1], Valéria Maia de Oliveira [1]

[1]Research Center for Chemistry, Biology and Agriculture (CPQBA), Institute of Biology, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

Corresponding author. Sanderson Tarciso Pereira de Sousa (sousa.tarciso@gmail.com) Fax: +55 19 2139 2852. *E-mail addresses*: sousa.tarciso@gmail.com (Sanderson Tarciso Pereira de Sousa), luc.g.cabral@gmail.com (Lucélia Cabral), gilenolacerdajr@gmail.com (Gileno Vieira Lacerda Júnior), melfontes@gmail.com (Melline Fontes Noronha), ottoni.julia@gmail.com (Júlia Ronzella Ottoni), adilson@cpqba.unicamp.br (Adilson Sartoratto), vmaia@cpqba.unicamp.br (Valéria Maia de Oliveira).

Table S1. Predicted and annotated ORFs of the fosmids FOS01 to 10 derived from a metagenomic library from oil-impacted mangrove sediment.

ORF	Locus	Length	Putative function	COG ID	COG Function	Pfam ID	Pfam Function	EC Number	Enzyme Function
1	1_99	99	hypothetical protein						
2	368_793	426	Predicted ester cyclase	COG5485	Predicted ester cyclase	pfam07366	SnoaL-like polyketide cyclase		
3	930_1517	588	hypothetical protein						
4	1560_2096	537	hypothetical protein						
5	2129_2620	492	hypothetical protein						
6	2689_3231	543	hypothetical protein						
7	3293_4294	1002	Cytochrome c553	COG2863; COG3264	Cytochrome c553; Small-conductance mechanosensitive channel	pfam00034; pfam13442	Cytochrome c; Cytochrome C oxidase, cbb3-type, subunit III		
8	4321_4746	426	Polyketide cyclase / dehydrase and lipid transport	COG3427	Carbon monoxide dehydrogenase subunit G	pfam10604	Polyketide cyclase / dehydrase and lipid transport		
9	4905_5939	1035	hypothetical protein	COG4257	Streptogramin lyase				
10	6109_7089	981	Cytochrome c, mono- and diheme variants	COG2010; COG2857	Cytochrome c, mono- and diheme variants; Cytochrome c1	pfam00034; pfam13442	Cytochrome c; Cytochrome C oxidase, cbb3-type, subunit III		
11	7128_7359	232	hypothetical protein						
12	7765_10569	2805	hypothetical protein			pfam13620	Carboxypeptidase regulatory-like domain		
13	10675_12087	1413	Superfamily II DNA and RNA helicases	COG0513; COG1278	Superfamily II DNA and RNA helicase; Cold shock protein, CspA family	pfam00270; pfam00271; pfam00313	DEAD/DEAH box helicase; Helicase conserved C-terminal domain; 'Cold-shock' DNA-binding domain	EC:3.6.4.13	RNA helicase.
14	12181_12529	349	hypothetical protein						
15	12554_14722	2169	Chlamydia polymorphic membrane protein (Chlamydia_PMP)			pfam02415	Chlamydia polymorphic membrane protein (Chlamydia_PMP) repeat		
16	15095_17215	2121	Beta-propeller repeat			pfam06739	Beta-propeller repeat		
17	17685_18782	1098	Predicted membrane protein/domain - COG1714	COG1714	Uncharacterized membrane protein YckC, RDD family	pfam06271	RDD family		
18	18779_18931	153	hypothetical protein						

19	18928_19872	945	hypothetical protein	COG3899	Predicted ATPase				
20	20310_21566	1257	Transposase and inactivated derivatives	COG3328	Transposase (or an inactivated derivative)	pfam00872	Transposase, Mutator family		
21	21563_22807	1245	DNA-binding transcriptional activator of the SARP family	COG3629; COG3899	DNA-binding transcriptional activator of the SARP family; Predicted ATPase	pfam03704; pfam13191	Bacterial transcriptional activator domain; AAA ATPase domain		
22	23062_23490	429	Predicted ester cyclase - COG5485	COG5485	Predicted ester cyclase	pfam07366	SnoaL-like polyketide cyclase		
23	23509_24276	768	Acetoacetate decarboxylase	COG4689	Acetoacetate decarboxylase	pfam06314	Acetoacetate decarboxylase (ADC)	EC:4.1.1.4	Acetoacetate decarboxylase.
24	24368_24598	231	hypothetical protein						
25	24595_27213	2619	Cation transport ATPase	COG0474	Magnesium-transporting ATPase (P-type)	pfam00122; pfam00702	E1-E2 ATPase; haloacid dehalogenase-like hydrolase	EC:3.6.3.-	Hydrolases. Acting on acid anhydrides. Acting on acid anhydrides; catalyzing transmembrane movement of substances.
26	27389_28372	984	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) - COG0596	COG0596	Pimeloyl-ACP methyl ester carboxylesterase	pfam12697	Alpha/beta hydrolase family		
27	28443_29165	723	Predicted membrane protein (DUF2154)	COG4758	Predicted membrane protein	pfam09922; pfam17115	Cell wall-active antibiotics response 4TMS YvqF; N-terminal domain of toast_rack, DUF2154		
28	29284_30228	945	hypothetical protein			pfam09922; pfam17115	Cell wall-active antibiotics response 4TMS YvqF; N-terminal domain of toast_rack, DUF2154		
29	30394_31215	822	hypothetical protein	COG1668	ABC-type Na+ efflux pump, permease component	pfam12730	ABC-2 family transporter protein		

30	31212_32171	960	ABC-type multidrug transport system, ATPase component	COG1131	ABC-type multidrug transport system, ATPase component	pfam00005	ABC transporter		
31	32255_33076	822	Predicted protease of the Abi (CAAX) family	COG4449	Predicted protease, Abi (CAAX) family	pfam02517	CAAX protease self-immunity		
32	33073_33303	231	hypothetical protein						
33	33300_34124	825	CAAX amino terminal protease family			pfam02517	CAAX protease self-immunity		
34	34142_34270	129	hypothetical protein						
35	34288_37026	2739	ATP-dependent transcriptional regulator	COG2909	ATP-, maltotriose- and DNA-dependent transcriptional regulator MalT	pfam00196; pfam13191	Bacterial regulatory proteins, luxR family; AAA ATPase domain		
36	37092_37170	79	hypothetical protein						
37	37290_39845	2556	hypothetical protein						
38	40311_41813	1503	ABC-type branched-chain amino acid transport systems, periplasmic component	COG0683; COG3889	ABC-type branched-chain amino acid transport system, periplasmic component; Predicted periplasmic protein	pfam13458	Periplasmic binding protein		
39	41974_42879	906	Restriction endonuclease	COG1715	Restriction endonuclease Mrr	pfam04471; pfam14338	Restriction endonuclease; Mrr N-terminal domain		
40	43023_43180	158	hypothetical protein						

FOS02

ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	1_99	99	hypothetical protein						
2	242_574	333	hypothetical protein						
3	602_1810	1209	ABC-type branched-chain amino acid transport systems, periplasmic component	COG0683	ABC-type branched-chain amino acid transport system, periplasmic component	pfam13458	Periplasmic binding protein		
4	1878_2798	921	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD	COG0175	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD	pfam01507	Phosphoadenosine phosphosulfate reductase family	EC:2.7.7.4	Sulfate adenylyltransferase.

			synthetase and related enzymes		synthetase or related enzyme				
5	2883_4814	1932	GTPases - Sulfate adenylate transferase subunit 1	COG0529; COG2895	Adenylylsulfate kinase or related kinase; Sulfate adenylyltransferase subunit 1, EFTu-like GTPase family	pfam00009; pfam01583	Elongation factor Tu GTP binding domain; Adenylylsulphate kinase	EC:2.7.1.25	Adenylyl-sulfate kinase.
6	4811_5410	600	Adenylylsulfate kinase and related kinases	COG0529	Adenylylsulfate kinase or related kinase	pfam01583	Adenylylsulphate kinase	EC:2.7.1.25	Adenylyl-sulfate kinase.
7	5503_7077	1575	hypothetical protein			pfam13458	Periplasmic binding protein		
8	7176_8831	1656	ABC-type branched-chain amino acid transport systems, periplasmic component	COG0683	ABC-type branched-chain amino acid transport system, periplasmic component				
9	8910_9755	846	Predicted methyltransferases - COG0313	COG0313	16S rRNA C1402 (ribose-2'-O) methylase RsmI	pfam00590	Tetrapyrrole (Corrin/Porphyrin) Methylases	EC:2.1.1.198	16S rRNA (cytidine(1402)-2'-O)-methyltransferase.
10	9752_10081	330	Uncharacterized conserved protein - COG1430	COG1430	Uncharacterized conserved membrane protein, UPF0127 family	pfam02643	Uncharacterized ACR, COG1430		
11	10169_10708	540	Acetyltransferases, including N-acetylases of ribosomal proteins	COG1670	Protein N-acetyltransferase, RimJ/RimL family	pfam13302	Acetyltransferase (GNAT) domain		
12	10722_11798	1077	Predicted GTPase, probable translation factor - COG0012	COG0012	Ribosome-binding ATPase YchF, GTP1/OBG family	pfam01926; pfam06071	50S ribosome-binding GTPase; Protein of unknown function (DUF933)		
13	11860_12939	1080	Predicted GTPases - COG1162	COG1162	Putative ribosome biogenesis GTPase RsgA	pfam03193	Protein of unknown function, DUF258	EC:3.6.1.-	Hydrolases. Acting on acid anhydrides. In phosphorus-containing anhydrides.
14	13042_14109	1068	Histidinol phosphatase and related hydrolases of the PHP family	COG1387	Histidinol phosphatase or	pfam02811	PHP domain		

					related hydrolase of the PHP family				
15	14254_14628	375	Uncharacterized protein conserved in bacteria	COG3011	Predicted thiol-disulfide oxidoreductase YuxK, DCC family	pfam04134	Protein of unknown function, DUF393		
16	14625_15605	981	Protein of unknown function (DUF2419)	COG4024	Uncharacterized protein	pfam10343	Potential Queuosine, Q, salvage protein family		
17	15769_16911	1143	Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein	COG1744	Basic membrane lipoprotein Med, periplasmic binding protein (PBP1-ABC) superfamily	pfam02608	Basic membrane protein		
18	16995_18506	1512	ABC-type uncharacterized transport systems, ATPase components	COG3845	ABC-type uncharacterized transport system, ATPase component	pfam00005	ABC transporter	EC:3.6.3.17	Monosaccharide-transporting ATPase.
19	18503_19756	1254	ABC-type uncharacterized transport system, permease component	COG4603	ABC-type uncharacterized transport system, permease component	pfam02653	Branched-chain amino acid transport system / permease component		
20	19753_21072	1320	Uncharacterized ABC-type transport system, permease component	COG1079	ABC-type uncharacterized transport system, permease component	pfam02653	Branched-chain amino acid transport system / permease component		
21	21069_21473	405	Cytidine deaminase	COG0295	Cytidine deaminase	pfam00383	Cytidine and deoxycytidylate deaminase zinc-binding region	EC:3.5.4.5	Cytidine deaminase.
22	21496_22779	1284	Thymidine phosphorylase	COG0213	Thymidine phosphorylase	pfam00591; pfam02885; pfam07831	Glycosyl transferase family, a/b domain; Glycosyl transferase family, helical bundle domain; Pyrimidine nucleoside phosphorylase C-terminal domain	EC:2.4.2.4	Thymidine phosphorylase.

23	22843_23466	624	Uracil phosphoribosyltransferase	COG0035	Uracil phosphoribosyltransferase	pfam14681	Uracil phosphoribosyltransferase	EC:2.4.2.9	Uracil phosphoribosyltransferase.
24	23468_24556	1089	Adenosine deaminase	COG1816	Adenosine deaminase	pfam00962	Adenosine/AMP deaminase	EC:3.5.4.4	Adenosine deaminase.
25	24587_26290	1704	Phosphomannomutase	COG1109	Phosphomannomutase	pfam00408; pfam02878; pfam02879; pfam02880	Phosphoglucomutase/phosphomannomutase, C-terminal domain; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain II; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III	EC:5.4.2.8	Phosphomannomutase.
26	26271_27104	834	Purine nucleoside phosphorylase	COG0005	Purine nucleoside phosphorylase	pfam01048	Phosphorylase superfamily	EC:2.4.2.1	Purine-nucleoside phosphorylase.
27	27132_28013	882	hypothetical protein	COG0510	Thiamine kinase and related kinases				
28	28024_29754	1731	N-acyl-D-aspartate/D-glutamate deacylase	COG3653	N-acyl-D-aspartate/D-glutamate deacylase	pfam07969	Amidohydrolase family		
29	29880_30515	636	hypothetical protein						
30	30515_30895	381	hypothetical protein						

FOS03

ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	3_77	75	hypothetical protein						
2	220_1299	1080	Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA - COG1058	COG1058	Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA	pfam00994	Probable molybdopterin binding domain	EC:3.5.1.42	Nicotinamide-nucleotide amidase.
3	1265_1765	501	hypothetical protein	COG5020					

4	1820_4492	2673	hypothetical protein	COG0308; COG1738	Aminopeptidase N; Uncharacterized PurR-regulated membrane protein YhhQ, DUF165 family				
5	4557_5603	1047	hypothetical protein						
6	5479_5940	462	hypothetical protein						
7	6007_7320	1314	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	COG0235	Ribulose-5-phosphate 4-epimerase/Fuculose-1-phosphate aldolase	pfam00596	Class II Aldolase and Adducin N-terminal domain	EC:4.1.2.17	L-fuculose-phosphate aldolase.
8	7382_8845	1464	Arylsulfatase A and related enzymes	COG3119	Arylsulfatase A or related enzyme	pfam00884	Sulfatase	EC:3.1.6.-	Hydrolases. Acting on ester bonds. Sulfuric ester hydrolases.
9	8842_10302	1461	Predicted dehydrogenases and related proteins - COG0673	COG0673	Predicted dehydrogenase	pfam01408; pfam02894	Oxidoreductase family, NAD-binding Rossmann fold; Oxidoreductase family, C-terminal alpha/beta domain		
10	10453_11289	837	Histidinol phosphatase and related hydrolases of the PHP family	COG1387	Histidinol phosphatase or related hydrolase of the PHP family	pfam02811	PHP domain	EC:3.1.3.15	Histidinol-phosphatase.
11	11260_14583	3324	Putative silver efflux pump	COG3696	Cu/Ag efflux pump CusA	pfam00873	AcrB/AcrD/AcrF family		
12	14687_16438	1752	hypothetical protein	COG0511; COG0845; COG4531	Biotin carboxyl carrier protein; Multidrug efflux pump subunit AcrA (membrane-fusion protein); ABC-type Zn2+ transport system, periplasmic component/surface adhesin				
13	16777_17727	951	DNA repair proteins	COG2003	DNA repair protein RadC, contains a helix-hairpin-helix DNA-binding motif	pfam04002	RadC-like JAB domain		
14	17972_18415	444	hypothetical protein						

15	18561_19973	1413	hypothetical protein	COG4948	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	pfam13378	Enolase C-terminal domain-like		
16	20040_20480	441	Protein of unknown function (DUF1499)	COG4446	Uncharacterized conserved protein, DUF1499 family	pfam07386	Protein of unknown function (DUF1499)		
17	20635_20922	288	hypothetical protein						
18	21312_24089	2778	hypothetical protein						
19	24257_24472	216	hypothetical protein						
20	25196_26110	915	Sphingosine kinase and enzymes related to eukaryotic diacylglycerol kinase	COG1597	Diacylglycerol kinase family enzyme	pfam00781	Diacylglycerol kinase catalytic domain		
21	26286_26843	558	hypothetical protein						
22	26982_27872	891	hypothetical protein	COG3595	Uncharacterized conserved protein YvlB, contains DUF4097 and DUF4098 domains	pfam13349	Putative adhesin		
23	28021_29187	1167	3-carboxymuconate cyclase	COG2706	6-phosphogluconolactonase, cycloisomerase 2 family	pfam10282	Lactonase, 7-bladed beta-propeller	EC:3.1.1.31	6-phosphogluconolactonase.
24	29356_31245	1890	hypothetical protein						
25	31425_32099	675	hypothetical protein	COG4907	Uncharacterized membrane protein				
26	32210_33412	1203	Periplasmic protein involved in polysaccharide export	COG1596	Periplasmic protein involved in polysaccharide export, contains SLBB domain of the beta-grasp fold	pfam02563; pfam10531	Polysaccharide biosynthesis/export protein; SLBB domain		
27	33487_34680	1194	hypothetical protein						
28	34745_34960	216	hypothetical protein						
29	35119_36666	1548	hypothetical protein						
30	36802_37686	885	Stage II sporulation protein E (SpoIIE)	COG2208	Serine phosphatase RsbU, regulator of sigma subunit	pfam07228	Stage II sporulation protein E (SpoIIE)		

31	38400_38609	210	hypothetical protein						
					FOS04				
ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	3_77	75	hypothetical protein						
2	220_615	396	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	COG0596	Pimeloyl-ACP methyl ester carboxylesterase	pfam00561	alpha/beta hydrolase fold	EC:3.8.1.3	Haloacetate dehalogenase.
3	1010_1111	102	Transposase IS200 like			pfam01797	Transposase IS200 like		
4	1105_1302	198	Transposase and inactivated derivatives	COG1943	REP element-mobilizing transposase RayT	pfam01797	Transposase IS200 like		
6	2059_2751	693	ABC-type tungstate transport system, periplasmic component	COG4662	ABC-type tungstate transport system, periplasmic component				
7	2754_3476	723	ABC-type multidrug transport system, ATPase component	COG1131	ABC-type multidrug transport system, ATPase component	pfam00005	ABC transporter	EC:3.6.3.55	Tungstate-importing ATPase.
8	3515_4351	837	ABC-type tungstate transport system, permease component	COG2998	ABC-type tungstate transport system, permease component	pfam12849	PBP superfamily domain		
9	4430_4672	243	hypothetical protein	COG4866	Uncharacterized protein				
10	4855_6039	1185	Methionine synthase II (cobalamin-independent)	COG0620	Methionine synthase II (cobalamin-independent)	pfam01717	Cobalamin-independent synthase, Catalytic domain	EC:2.1.1.14	5-methyltetrahydropteroyl triglutamate--homocysteine S-methyltransferase.
11	6324_7208	885	Uncharacterized protein required for formate dehydrogenase activity	COG1526	Formate dehydrogenase assembly factor FdhD	pfam02634	FdhD/NarQ family		
12	7208_7849	642	Molybdopterin-guanine dinucleotide biosynthesis protein A	COG0746	Molybdopterin-guanine dinucleotide biosynthesis protein A	pfam12804	MobA-like NTP transferase domain	EC:2.7.7.77	Molybdenum cofactor guanylyltransferase.
13	7971_9242	1272	Molybdopterin biosynthesis enzyme	COG0303	Molybdopterin biosynthesis enzyme	pfam00994; pfam03453; pfam03454	Probable molybdopterin binding domain; MoeA N-	EC:2.10.1.1	Molybdopterin molybdotransferase.

							terminal region (domain I and II); MoeA C-terminal region (domain IV)		
14	9242_9775	534	Molybdopterin-guanine dinucleotide biosynthesis protein	COG1763	Molybdopterin-guanine dinucleotide biosynthesis protein	pfam03205	Molybdopterin guanine dinucleotide synthesis protein B	EC:2.10.1.1	Molybdopterin molybdotransferase.
15	9959_10331	373	hypothetical protein	COG2864	Cytochrome b subunit of formate dehydrogenase				
16	10589_11590	1002	TRAP-type C4-dicarboxylate transport system, periplasmic component	COG1638	TRAP-type C4-dicarboxylate transport system, periplasmic component	pfam03480	Bacterial extracellular solute-binding protein, family 7		
17	11628_12140	513	TRAP-type C4-dicarboxylate transport system, small permease component	COG3090	TRAP-type C4-dicarboxylate transport system, small permease component	pfam04290	Tripartite ATP-independent periplasmic transporters, DctQ component		
18	12144_13457	1314	TRAP-type mannitol/chloroaromatic compound transport system, large permease component	COG4664	TRAP-type mannitol/chloroaromatic compound transport system, large permease component	pfam06808	Tripartite ATP-independent periplasmic transporter, DctM component		
19	13510_14517	1008	hypothetical protein						
20	14514_14822	309	hypothetical protein						
21	15133_16074	942	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) - COG0596	COG0596	Pimeloyl-ACP methyl ester carboxylesterase	pfam00561	alpha/beta hydrolase fold		
22	16170_16409	240	hypothetical protein						
23	16431_17135	705	hypothetical protein	COG2319	WD40 repeat				
24	17374_19365	1992	Molybdopterin biosynthesis enzyme	COG0303; COG1910	Molybdopterin biosynthesis enzyme; Periplasmic molybdate-binding protein/domain	pfam00994; pfam03453; pfam03454; pfam12727	Probable molybdopterin binding domain; MoeA N-terminal region (domain I and II); MoeA		

							C-terminal region (domain IV); PBP superfamily domain		
25	19362_20498	1137	Molybdopterin biosynthesis enzyme	COG0303	Molybdopterin biosynthesis enzyme	pfam03453; pfam03454	MoeA N-terminal region (domain I and II); MoeA C-terminal region (domain IV)		
26	20616_20840	225	Predicted redox protein, regulator of disulfide bond formation - COG0425	COG0425	TusA-related sulfurtransferase	pfam01206	Sulfurtransferase TusA	EC:2.8.1.-	Transferases. Transferring sulfur-containing groups. Sulfurtransferases.
27	20851_22077	1227	Cytosine deaminase and related metal-dependent hydrolases	COG0402	Cytosine/adenosine deaminase or related metal-dependent hydrolase	pfam07969	Amidohydrolase family	EC:3.5.4.1	Cytosine deaminase.
28	22173_22973	801	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	COG3836	2-keto-3-deoxy-L-rhamnonate aldolase RhmA	pfam03328	HpcH/HpaI aldolase/citrate lyase family		
29	22990_24249	1260	NhaP-type Na+/H+ and K+/H+ antiporters	COG0025	NhaP-type Na+/H+ or K+/H+ antiporter	pfam00999	Sodium/hydrogen exchanger family		
30	24474_25970	1497	Aromatic ring hydroxylase	COG2368	Aromatic ring hydroxylase	pfam03241; pfam11794	4-hydroxyphenylacetate 3-hydroxylase C terminal; 4-hydroxyphenylacetate 3-hydroxylase N terminal	EC:4.2.1.120	4-hydroxybutanoyl-CoA dehydratase.
31	26160_27002	843	Citrate lyase beta subunit	COG2301	Citrate lyase beta subunit	pfam03328; pfam15617	HpcH/HpaI aldolase/citrate lyase family; C-C_Bond_Lyase of the TIM-Barrel fold	EC:4.1.3.34	Citryl-CoA lyase.
32	27009_28151	1143	Acyl-CoA dehydrogenases	COG1960	Acyl-CoA dehydrogenase related to the alkylation response protein AidB	pfam00441; pfam02770; pfam02771	Acyl-CoA dehydrogenase, C-terminal domain; Acyl-CoA dehydrogenase, middle domain; Acyl-CoA dehydrogenase, N-terminal domain		

33	28151_29341	1191	Predicted acyl-CoA transferases/carnitine dehydratase - COG1804	COG1804	Crotonobetainyl-CoA:carnitine CoA-transferase CaiB and related acyl-CoA transferases	pfam02515	CoA-transferase family III	EC:2.8.3.16	Formyl-CoA transferase.
34	29338_29790	453	Acyl dehydratase	COG2030	Acyl dehydratase	pfam01575	MaoC like domain		
35	30234_30620	387	hypothetical protein	COG0489	Chromosome partitioning ATPase, Mrp family, contains Fe-S cluster				
36	30827_31699	873	Transglutaminase-like enzymes, putative cysteine proteases	COG1305	Transglutaminase-like enzyme, putative cysteine protease	pfam01841; pfam08379	Transglutaminase-like superfamily; Bacterial transglutaminase-like N-terminal region		
37	31696_33177	1482	Uncharacterized protein conserved in bacteria	COG2307; COG2308	Uncharacterized conserved protein, Alpha-E superfamily; Uncharacterized conserved protein, circularly permuted ATPgrasp superfamily	pfam04168; pfam14403	A predicted alpha-helical domain with a conserved ER motif.; Circularly permuted ATP-grasp type 2		
					FOS05				
ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	1_78	78	hypothetical protein						
2	331_645	315	Factor for inversion stimulation Fis, transcriptional activator	COG2901	DNA-binding protein Fis (factor for inversion stimulation)	pfam02954	Bacterial regulatory protein, Fis family		
3	670_1725	1056	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control - COG0037	COG0037	tRNA(Ile)-lysidine synthase TilS/MesJ	pfam01171	PP-loop family	EC:6.3.4.19	tRNA(Ile)-lysidine synthetase.
4	1792_2079	288	Uncharacterized conserved protein	COG1359	Quinol monooxygenase YgiN	pfam03992	Antibiotic biosynthesis monooxygenase		
5	2470_4182	1713	Serine/threonine protein kinase	COG0515	Serine/threonine protein kinase	pfam00069	Protein kinase domain	EC:2.7.11.1	Non-specific serine/threonine protein kinase.
6	4395_5303	909	Stage II sporulation protein E (SpoIIE)	COG2208	Serine phosphatase RsbU, regulator of sigma subunit	pfam07228	Stage II sporulation protein E (SpoIIE)		

7	5438_7570	2133	hypothetical protein	COG1413	HEAT repeat	pfam13646	HEAT repeats		
8	8008_8451	444	hypothetical protein						
9	8584_9150	567	hypothetical protein						
10	9230_11482	2253	FOG: WD40 repeat	COG2319	WD40 repeat	pfam00400	WD domain, G-beta repeat		
11	11512_12129	618	hypothetical protein						
12	12092_12250	159	hypothetical protein						
13	12947_14572	1626	hypothetical protein	COG1914	Mn2+ and Fe2+ transporters of the NRAMP family				
14	14564_16387	1824	4-amino-4-deoxy-L-arabinose transferase and related glycosyltransferases of PMT family	COG1807	4-amino-4-deoxy-L-arabinose transferase or related glycosyltransferase of PMT family	pfam13231	Dolichyl-phosphate-mannose-protein mannosyltransferase		
15	16452_16964	513	hypothetical protein						
16	17011_18255	1245	Protein of unknown function (DUF1559)	COG2165	Type II secretory pathway, pseudopilin PulG	pfam07596; pfam07963	Protein of unknown function (DUF1559); Prokaryotic N-terminal methylation motif		
17	18357_19622	1266	Type II secretory pathway, pseudopilin PulG	COG2165	Type II secretory pathway, pseudopilin PulG	pfam07596	Protein of unknown function (DUF1559)		
18	19676_19963	288	Protein of unknown function (DUF1559)			pfam07596	Protein of unknown function (DUF1559)		
19	20055_21827	1773	Cytochrome c peroxidase	COG1858; COG3391	Cytochrome c peroxidase; DNA-binding beta-propeller fold protein YncE	pfam02239; pfam03150	Cytochrome D1 heme domain; Di-haem cytochrome c peroxidase		
20	21890_23317	1428	Arylsulfatase A and related enzymes	COG3119	Arylsulfatase A or related enzyme	pfam00884	Sulfatase	EC:3.1.6.-	Hydrolases. Acting on ester bonds. Sulfuric ester hydrolases.
21	23589_24602	1014	Predicted permeases - COG0730	COG0730	Uncharacterized membrane protein YfcA	pfam01925	Sulfite exporter TauE/SafE		
22	24885_25079	195	hypothetical protein						
23	25190_25789	600	hypothetical protein						

24	26029_27282	1254	Dehydrogenases (flavoproteins)	COG0644	Dehydrogenase (flavoprotein)	pfam01494	FAD binding domain		
25	27581_27844	264	hypothetical protein	COG0236	Acyl carrier protein				
26	27847_29442	1596	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	COG0318	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	pfam00501; pfam13193	AMP-binding enzyme; AMP-binding enzyme C-terminal domain		
27	29432_29611	180	hypothetical protein						
28	29608_30609	1002	Phosphotransferase enzyme family	COG4857	5-Methylthioribose kinase, methionine salvage pathway	pfam01636	Phosphotransferase enzyme family	EC:2.7.1.100	S-methyl-5-thioribose kinase.
29	30786_31235	450	Lactoylglutathione lyase and related lyases	COG0346	Catechol 2,3-dioxygenase or other lactoylglutathione lyase family enzyme	pfam00903	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily		
30	31335_32465	1131	hypothetical protein	COG0265	Periplasmic serine protease, S1-C subfamily, contain C-terminal PDZ domain	pfam13180	PDZ domain		
31	32499_33221	723	hypothetical protein						
32	33254_34519	1266	Aspartate/tyrosine/aromatic aminotransferase	COG0436	Aspartate/methionine/tyrosine aminotransferase	pfam00155	Aminotransferase class I and II	EC:2.6.1.1	Aspartate transaminase.
33	34726_34953	228	hypothetical protein						
34	35173_36333	1161	UDP-N-acetylglucosamine 2-epimerase	COG0381	UDP-N-acetylglucosamine 2-epimerase	pfam02350	UDP-N-acetylglucosamine 2-epimerase	EC:5.1.3.14	UDP-N-acetylglucosamine 2-epimerase (non-hydrolyzing).
35	36381_36818	438	hypothetical protein						
36	37011_37214	204	hypothetical protein	COG4747	Uncharacterized conserved protein, contains tandem ACT domains				
37	37434_37536	103	hypothetical protein						

FOS06

ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	45_177	133	hypothetical protein						
2	315_578	264	Phosphoenolpyruvate-protein kinase (PTS	COG1080	Phosphoenolpyruvate-protein kinase (PTS	pfam05524	PEP-utilising enzyme, N-terminal	EC:2.7.3.9	Phosphoenolpyruvate--protein phosphotransferase.

			system EI component in bacteria)		system EI component in bacteria)				
3	925_1236	312	Branched-chain amino acid ABC-type transport system, permease components	COG0559	Branched-chain amino acid ABC-type transport system, permease component	pfam02653	Branched-chain amino acid transport system / permease component		
4	1341_2042	702	ABC-type branched-chain amino acid transport systems, ATPase component	COG0410	ABC-type branched-chain amino acid transport system, ATPase component	pfam00005	ABC transporter		
5	2035_2766	732	ABC-type branched-chain amino acid transport systems, ATPase component	COG0411	ABC-type branched-chain amino acid transport system, ATPase component	pfam00005; pfam12399	ABC transporter; Branched-chain amino acid ATP-binding cassette transporter		
6	2917_3114	198	hypothetical protein	COG2768	Uncharacterized Fe-S center protein	pfam13237	4Fe-4S dicluster domain		
7	3157_3927	771	Acyl CoA:acetate/3-ketoacid CoA transferase, beta subunit	COG2057	Acyl CoA:acetate/3-ketoacid CoA transferase, beta subunit	pfam01144	Coenzyme A transferase	EC:2.8.3.12	Glutaconate CoA-transferase.
8	4018_4917	900	Acyl CoA:acetate/3-ketoacid CoA transferase, alpha subunit	COG1788	Acyl CoA:acetate/3-ketoacid CoA transferase, alpha subunit	pfam01144	Coenzyme A transferase	EC:2.8.3.12	Glutaconate CoA-transferase.
9	4929_5393	465	hypothetical protein						
10	5390_6064	675	Nitroreductase	COG0778	Nitroreductase	pfam00881	Nitroreductase family		
11	6179_6256	78	hypothetical protein						
12	6397_6510	114	hypothetical protein						
13	6651_7133	483	Transcriptional regulators	COG1522	DNA-binding transcriptional regulator, Lrp family	pfam01037; pfam13404	Lrp/AsnC ligand binding domain; AsnC-type helix-turn-helix domain		
14	7605_9047	1443	Carboxylesterase type B	COG2272	Carboxylesterase type B	pfam00135	Carboxylesterase family	EC:3.1.1.-	Hydrolases. Acting on ester bonds. Carboxylic ester hydrolases.
15	9489_10445	957	ABC-type oligopeptide transport system, ATPase component	COG4608	ABC-type oligopeptide transport system, ATPase component	pfam00005; pfam08352	ABC transporter; Oligopeptide/dipeptide transporter, C-terminal region		

16	10429_11403	975	ABC-type dipeptide/oligopeptide/ nickel transport system, ATPase component	COG0444	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	pfam00005; pfam08352	ABC transporter; Oligopeptide/dipeptide transporter, C-terminal region		
17	11447_12334	888	ABC-type dipeptide/oligopeptide/ nickel transport systems, permease components	COG1173	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	pfam00528; pfam12911	Binding-protein-dependent transport system inner membrane component; N-terminal TM domain of oligopeptide transport permease C		
18	12327_13283	957	ABC-type dipeptide/oligopeptide/ nickel transport systems, permease components	COG0601	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	pfam00528	Binding-protein-dependent transport system inner membrane component		
19	13571_15223	1653	ABC-type dipeptide transport system, periplasmic component	COG0747; COG4982	ABC-type transport system, periplasmic component; 3-oxoacyl-ACP reductase domain of yeast-type FAS1	pfam00496	Bacterial extracellular solute-binding proteins, family 5 Middle		
20	15377_16165	789	Zn-dependent hydrolases, including glyoxylases	COG0491	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II	pfam00753	Metallo-beta-lactamase superfamily		
21	16200_17330	1131	Enterochelin esterase and related enzymes	COG2382	Enterochelin esterase or related enzyme	pfam00756	Putative esterase		
22	17409_17975	567	hypothetical protein						
23	18026_19429	1404	hypothetical protein	COG1053; COG1206	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit; Folate-dependent tRNA-U54 methylase TrmFO/GidA	pfam12831	FAD dependent oxidoreductase		

24	19764_20666	903	Esterase/lipase	COG0657	Acetyl esterase/lipase	pfam07859	alpha/beta hydrolase fold		
25	20718_21839	1122	Malate/L-lactate dehydrogenases	COG2055	Malate/lactate/ureidoglycolate dehydrogenase, LDH2 family	pfam02615	Malate/L-lactate dehydrogenase		
26	22051_22275	225	hypothetical protein			pfam14451	Mut7-C ubiquitin		
27	22331_24052	1722	Aldehyde:ferredoxin oxidoreductase	COG2414	Aldehyde:ferredoxin oxidoreductase	pfam01314; pfam02730	Aldehyde ferredoxin oxidoreductase, domains 2 & 3; Aldehyde ferredoxin oxidoreductase, N-terminal domain	EC:1.2.7.5	Aldehyde ferredoxin oxidoreductase.
28	24390_25355	966	Uncharacterized protein, putative amidase	COG1402	Creatinine amidohydrolase/Fe(II)-dependent formamide hydrolase involved in riboflavin and F420 biosynthesis	pfam02633	Creatinine amidohydrolase	EC:3.5.2.10	Creatininase.
29	25370_26287	918	hypothetical protein	COG0826	Collagenase-like protease, PrtC family				
30	26308_26688	381	Cupin domain	COG0662	Mannose-6-phosphate isomerase, cupin superfamily	pfam07883	Cupin domain		
31	26747_27508	762	Transcriptional regulator	COG1414	DNA-binding transcriptional regulator, IclR family	pfam01614; pfam09339	Bacterial transcriptional regulator; IclR helix-turn-helix domain		
32	27908_28963	1056	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)	COG1105	Fructose-1-phosphate kinase or kinase (PfkB)	pfam00294	pfkB family carbohydrate kinase	EC:2.7.1.11	6-phosphofructokinase.
33	29026_29655	630	Archaeal/vacuolar-type H+-ATPase subunit D	COG1394	Archaeal/vacuolar-type H+-ATPase subunit D/Vma8	pfam01813	ATP synthase subunit D		
34	29664_31082	1419	Archaeal/vacuolar-type H+-ATPase subunit B	COG1156	Archaeal/vacuolar-type H+-ATPase subunit B/Vma2	pfam00006; pfam00306; pfam02874	ATP synthase alpha/beta family, nucleotide-binding domain; ATP synthase alpha/beta chain, C		

							terminal domain; ATP synthase alpha/beta family, beta-barrel domain		
35	31103_32872	1770	Archaeal/vacuolar-type H+-ATPase subunit A	COG1155	Archaeal/vacuolar-type H+-ATPase catalytic subunit A/Vma1	pfam00006; pfam00306; pfam02874; pfam16886	ATP synthase alpha/beta family, nucleotide-binding domain; ATP synthase alpha/beta chain, C terminal domain; ATP synthase alpha/beta family, beta-barrel domain; ATPsynthase alpha/beta subunit N-term extension	EC:3.6.3.14	H(+)-transporting two-sector ATPase.
36	32869_33507	639	Archaeal/vacuolar-type H+-ATPase subunit E	COG1390	Archaeal/vacuolar-type H+-ATPase subunit E/Vma4	pfam01991	ATP synthase (E/31 kDa) subunit		
37	33512_33664	153	hypothetical protein	COG1436	Archaeal/vacuolar-type H+-ATPase subunit F/Vma7			EC:4.2.3.4	3-dehydroquinate synthase.
FOS07									
ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	348_725	378	Uncharacterized protein conserved in bacteria	COG3181	Tripartite-type tricarboxylate transporter, receptor component TctC	pfam03401	Tripartite tricarboxylate transporter family receptor		
2	806_1273	468	Tripartite tricarboxylate transporter TctB family			pfam07331	Tripartite tricarboxylate transporter TctB family		
3	1277_2779	1503	Uncharacterized protein conserved in bacteria	COG3333	TctA family transporter	pfam01970	Tripartite tricarboxylate transporter TctA family		
4	2803_4692	1890	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome),	COG0028	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme	pfam00205; pfam02775; pfam02776	Thiamine pyrophosphate enzyme, central domain; Thiamine pyrophosphate enzyme, C-terminal TPP binding domain;	EC:3.7.1.22	3D-(3,5/4)-trihydroxycyclohexane-1,2-dione acylhydrolase (decyclizing).

			glyoxylate carboligase, phosphonopyruvate decarboxylase]				Thiamine pyrophosphate enzyme, N-terminal TPP binding domain		
5	4712_6151	1440	NAD-dependent aldehyde dehydrogenases	COG1012	Acyl-CoA reductase or other NAD-dependent aldehyde dehydrogenase	pfam00171	Aldehyde dehydrogenase family	EC:1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+)).
6	6148_7848	1701	Dihydroxyacid dehydratase/phosphogluconate dehydratase	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	pfam00920	Dehydratase family	EC:4.2.1.9	Dihydroxy-acid dehydratase.
7	8006_8833	828	hypothetical protein			pfam13469	Sulfotransferase family		
8	9408_10190	783	hypothetical protein						
10	10780_11409	630	Ribosomal protein L21	COG0261; COG3743	Ribosomal protein L21; Predicted 5' DNA nuclease, flap endonuclease-1-like, helix-3-turn-helix (H3TH) domain	pfam00829; pfam14520	Ribosomal prokaryotic L21 protein; Helix-hairpin-helix domain		
11	11439_11708	270	Ribosomal protein L27	COG0211	Ribosomal protein L27	pfam01016	Ribosomal L27 protein		
12	11822_12409	588	Acetyltransferases, including N-acetylases of ribosomal proteins	COG1670	Protein N-acetyltransferase, RimJ/RimL family	pfam13302	Acetyltransferase (GNAT) domain		
13	12406_13446	1041	Predicted GTPase - COG0536	COG0536	GTPase involved in cell partioning and DNA repair	pfam01018; pfam01926	GTP1/OBG; 50S ribosome-binding GTPase		
14	13443_14570	1128	Glutamate 5-kinase	COG0263	Glutamate 5-kinase	pfam00696; pfam01472	Amino acid kinase family; PUA domain	EC:2.7.2.11	Glutamate 5-kinase.
15	14567_15850	1284	Gamma-glutamyl phosphate reductase	COG0014	Gamma-glutamyl phosphate reductase	pfam00171	Aldehyde dehydrogenase family	EC:1.2.1.41	Glutamate-5-semialdehyde dehydrogenase.
16	15872_16492	621	Nicotinic acid mononucleotide adenylyltransferase	COG1057	Nicotinic acid mononucleotide adenylyltransferase	pfam01467	Cytidylyltransferase-like	EC:2.7.7.18	Nicotinate-nucleotide adenylyltransferase.
17	16614_17054	441	Uncharacterized homolog of plant Iojap protein	COG0799	Ribosomal silencing factor RsfS, regulates	pfam02410	Ribosomal silencing factor during starvation		

					association of 30S and 50S subunits				
18	17064_17546	483	Uncharacterized conserved protein - COG1576	COG1576	23S rRNA pseudoU1915 N3-methylase RlmH	pfam02590	Predicted SPOUT methyltransferase	EC:2.1.1.177	23S rRNA (pseudouridine(1915)-N(3))-methyltransferase.
19	17572_18933	1362	Membrane-bound metallopeptidase	COG4942	Septal ring factor EnvC, activator of murein hydrolases AmiA and AmiB	pfam01551	Peptidase family M23		
20	18934_20286	1353	Periplasmic protease	COG0793	C-terminal processing protease CtpA/Prc, contains a PDZ domain	pfam03572; pfam13180	Peptidase family S41; PDZ domain	EC:3.4.21.102	C-terminal processing peptidase.
21	20426_21610	1185	Uncharacterized protein conserved in bacteria	COG2861	Uncharacterized conserved protein YibQ, putative polysaccharide deacetylase 2 family	pfam04748	Divergent polysaccharide deacetylase		
22	21607_22131	525	NTP pyrophosphohydrolases including oxidative damage repair enzymes	COG0494	8-oxo-dGTP pyrophosphatase MutT and related house-cleaning NTP pyrophosphohydrolases, NUDIX family	pfam00293	NUDIX domain	EC:3.6.1.-	Hydrolases. Acting on acid anhydrides. In phosphorus-containing anhydrides.
23	22128_23132	1005	NADPH:quinone reductase and related Zn-dependent oxidoreductases	COG0604	NADPH:quinone reductase or related Zn-dependent oxidoreductase	pfam08240; pfam13602	Alcohol dehydrogenase GroES-like domain; Zinc-binding dehydrogenase	EC:1.6.5.5	NADPH:quinone reductase.
24	23150_24841	1692	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	COG0651	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	pfam00361	Proton-conducting membrane transporter		
25	24834_25094	261	hypothetical protein						
26	25108_26571	1464	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit	pfam00361	Proton-conducting membrane transporter		

			Na+/H+ antiporter, MnhA subunit		Na+/H+ antiporter, MnhA subunit				
27	26568_28049	1482	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	COG0651	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	pfam00361	Proton-conducting membrane transporter		
28	28046_28426	381	Multisubunit Na+/H+ antiporter, MnhC subunit	COG1006	Multisubunit Na+/H+ antiporter, MnhC subunit	pfam00420	NADH-ubiquinone/plastoquinone oxidoreductase chain 4L		
29	28426_28842	417	Multisubunit Na+/H+ antiporter, MnhB subunit	COG2111	Multisubunit Na+/H+ antiporter, MnhB subunit	pfam04039	Domain related to MnhB subunit of Na+/H+ antiporter		
30	28839_29414	576	Multisubunit Na+/H+ antiporter, MnhB subunit	COG2111	Multisubunit Na+/H+ antiporter, MnhB subunit	pfam13244	Domain of unknown function (DUF4040)		
31	29411_29830	420	Multisubunit Na+/H+ antiporter, MnhG subunit	COG1320	Multisubunit Na+/H+ antiporter, MnhG subunit	pfam03334	Na+/H+ antiporter subunit		
32	29827_30111	285	Multisubunit Na+/H+ antiporter, MnhF subunit	COG2212	Multisubunit Na+/H+ antiporter, MnhF subunit	pfam04066	Multiple resistance and pH regulation protein F (MrpF / PhaF)		
33	30111_30587	477	Multisubunit Na+/H+ antiporter, MnhE subunit	COG1863	Multisubunit Na+/H+ antiporter, MnhE subunit	pfam01899	Na+/H+ ion antiporter subunit		
34	30762_31190	429	Acetyltransferases	COG0456	Ribosomal protein S18 acetylase RimI and related acetyltransferases	pfam00583	Acetyltransferase (GNAT) family		
35	31187_32626	1440	FAD/FMN-containing dehydrogenases	COG0277	FAD/FMN-containing dehydrogenase	pfam01565; pfam02913	FAD binding domain; FAD linked oxidases, C-terminal domain		
36	32666_33268	603	BA14K-like protein			pfam07886	BA14K-like protein		
37	33355_34296	942	Putative translation factor (SUA5)	COG0009	tRNA A37 threonylcarbamoyladenosine synthetase	pfam01300; pfam03481	Telomere recombination; Putative GTP-binding	EC:2.7.7.87	L-threonylcarbamoyladenylate synthase.

					subunit TsaC/SUA5/YrdC		controlling metal-binding		
38	34433_35956	1524	Acyl-CoA dehydrogenases	COG1960	Acyl-CoA dehydrogenase related to the alkylation response protein AidB	pfam00441; pfam02770; pfam02771	Acyl-CoA dehydrogenase, C-terminal domain; Acyl-CoA dehydrogenase, middle domain; Acyl-CoA dehydrogenase, N-terminal domain	EC:1.3.8.1	Short-chain acyl-CoA dehydrogenase.

FOS08

ORF	Locus	Length	Putative function	COG ID	COG Function	Pfam ID	Pfam Function	EC Number	Enzyme Function
1	489_974	486	GAF domain-containing protein	COG1956	GAF domain-containing protein, putative methionine-R-sulfoxide reductase	pfam01590	GAF domain		
2	988_3483	2496	Signal transduction histidine kinase	COG0642; COG0784; COG2202	Signal transduction histidine kinase; CheY chemotaxis protein or a CheY-like REC (receiver) domain; PAS domain	pfam00072; pfam00512; pfam02518; pfam08447; pfam08448	Response regulator receiver domain; His Kinase A (phospho-acceptor) domain; Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase; PAS fold; PAS fold		
3	3491_5518	2028	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	COG0365	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase	pfam00501; pfam13193; pfam16177	AMP-binding enzyme; AMP-binding enzyme C-terminal domain; Acetyl-coenzyme A synthetase N-terminus	EC:6.2.1.1	Acetate--CoA ligase.
4	5557_6015	459	hypothetical protein	COG2323	Uncharacterized membrane protein YcaP, DUF421 family				
5	6093_7748	1656	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	COG0154	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit or related amidase	pfam01425	Amidase		
6	7856_9232	1377	NADPH-dependent glutamate synthase	COG0493	NADPH-dependent glutamate synthase	pfam07992	Pyridine nucleotide-disulphide oxidoreductase	EC:1.18.1.2	Ferredoxin--NADP(+) reductase.

			beta chain and related oxidoreductases		beta chain or related oxidoreductase				
7	9232_9879	648	Protein of unknown function, DUF642			pfam04862	Protein of unknown function (DUF642)		
8	9996_10700	705	hypothetical protein						
9	10802_12784	1983	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	COG1506	Dipeptidyl aminopeptidase/acylaminoacyl peptidase	pfam00326	Prolyl oligopeptidase family		
10	12838_13881	1044	Zn-dependent alcohol dehydrogenases	COG1064	D-arabinose 1-dehydrogenase, Zn-dependent alcohol dehydrogenase family	pfam00107; pfam08240	Zinc-binding dehydrogenase; Alcohol dehydrogenase GroES-like domain	EC:1.1.1.1	Alcohol dehydrogenase.
11	14019_15188	1170	hypothetical protein						
12	15209_17074	1866	Aminopeptidase N	COG0308	Aminopeptidase N	pfam01433; pfam09127	Peptidase family M1; Leukotriene A4 hydrolase, C-terminal		
13	17076_17993	918	hypothetical protein	COG1235	Phosphoribosyl 1,2-cyclic phosphodiesterase	pfam12706	Beta-lactamase superfamily domain		
14	18024_19385	1362	1-acyl-sn-glycerol-3-phosphate acyltransferase	COG0204	1-acyl-sn-glycerol-3-phosphate acyltransferase	pfam01553	Acyltransferase		
15	19382_23104	3723	Methionine synthase I, cobalamin-binding domain	COG0646; COG1410	Methionine synthase I (cobalamin-dependent), methyltransferase domain; Methionine synthase I, cobalamin-binding domain	pfam00809; pfam02310; pfam02574; pfam02607; pfam02965	Pterin binding enzyme; B12 binding domain; Homocysteine S-methyltransferase; B12 binding domain; Vitamin B12 dependent methionine synthase, activation domain	EC:2.1.1.13	Methionine synthase.
16	23402_26026	2625	Aminopeptidase N	COG0308	Aminopeptidase N	pfam01433; pfam11838	Peptidase family M1; ERAP1-like C-terminal domain	EC:3.4.11.14	Cytosol alanyl aminopeptidase.
17	26039_26335	297	hypothetical protein						
18	26345_26707	363	hypothetical protein	COG3350	Uncharacterized conserved protein, YHS domain				

19	26959_29622	2664	hypothetical protein	COG1843	Flagellar hook assembly protein FlgD	pfam07940; pfam13860	Heparinase II/III-like protein; FlgD Ig-like domain		
20	29669_31474	1806	hypothetical protein			pfam13517	Repeat domain in Vibrio, Colwellia, Bradyrhizobium and Shewanella		
21	31500_31811	312	hypothetical protein	COG1843	Flagellar hook assembly protein FlgD	pfam13860	FlgD Ig-like domain		
FOS09									
ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	521_1534	1014	Sulfite reductase, beta subunit (hemoprotein)	COG0155	Sulfite reductase, beta subunit (hemoprotein)	pfam01077; pfam03460	Nitrite and sulphite reductase 4Fe-4S domain; Nitrite/Sulfite reductase ferredoxin-like half domain	EC:1.8.7.1	Assimilatory sulfite reductase (ferredoxin).
2	1754_3529	1776	hypothetical protein	COG3063; COG4783; COG5010	Tfp pilus assembly protein PilF; Putative Zn-dependent protease, contains TPR repeats; Flp pilus assembly protein TadD, contains TPR repeats	pfam13181; pfam13371; pfam13414	Tetratricopeptide repeat; Tetratricopeptide repeat; TPR repeat		
3	3756_4937	1182	hypothetical protein	COG0348	Polyferredoxin	pfam12801; pfam12837	4Fe-4S binding domain; 4Fe-4S binding domain		
4	5051_6358	1308	Predicted dehydrogenases and related proteins - COG0673	COG0673	Predicted dehydrogenase	pfam01408	Oxidoreductase family, NAD-binding Rossmann fold		
5	6472_7086	615	Transposase and inactivated derivatives	COG1943	REP element-mobilizing transposase RayT	pfam01797	Transposase IS200 like		
6	7259_8197	939	hypothetical protein	COG0275; COG2227	16S rRNA C1402 N4-methylase RsmH; 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase	pfam13489	Methyltransferase domain		

7	8263_8829	567	hypothetical protein						
8	8931_9551	621	hypothetical protein						
9	9688_9951	264	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes	COG0508	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component	pfam00364	Biotin-requiring enzyme	EC:2.3.1.12	Dihydrolipoyllysine-residue acetyltransferase.
10	10087_10971	885	Lipoate synthase	COG0320	Lipoate synthase	pfam04055	Radical SAM superfamily	EC:2.8.1.8	Lipoyl synthase.
11	11079_11804	726	Lipoate-protein ligase B	COG0321	Lipoate-protein ligase B			EC:2.3.1.181	Lipoyl(octanoyl) transferase.
12	11822_12421	600	tmRNA-binding protein	COG0691	tmRNA-binding protein	pfam01668	SmpB protein		
13	12845_13528	684	ABC-type antimicrobial peptide transport system, ATPase component	COG1136	ABC-type lipoprotein export system, ATPase component	pfam00005	ABC transporter		
14	13638_15098	1461	ABC-type transport system, involved in lipoprotein release, permease component	COG4591	ABC-type transport system, involved in lipoprotein release, permease component	pfam02687; pfam12704	FtsX-like permease family; MacB-like periplasmic core domain		
15	15444_16127	684	Uncharacterized protein conserved in bacteria	COG3495	Uncharacterized protein	pfam11736; pfam13828	Protein of unknown function (DUF3299); Domain of unknown function (DUF4190)		
16	16338_17162	825	Uncharacterized homolog of PSP1	COG1774	Cell fate regulator YaaT, PSP1 superfamily (controls sporulation, competence, biofilm development)	pfam04468	PSP1 C-terminal conserved region		
17	17193_17630	438	hypothetical protein	COG1031	Radical SAM superfamily enzyme with C-terminal helix-hairpin-helix motif				
18	17868_18053	186	hypothetical protein						

19	18216_19448	1233	Aspartate/tyrosine/aromatic aminotransferase	COG0436	Aspartate/methionine/tyrosine aminotransferase	pfam00155	Aminotransferase class I and II	EC:2.6.1.83	LL-diaminopimelate aminotransferase.
20	19728_20492	765	hypothetical protein						
21	20796_21737	942	Sugar phosphate isomerases/epimerases	COG1082	Sugar phosphate isomerase/epimerase	pfam01261	Xylose isomerase-like TIM barrel		
22	21786_22847	1062	Predicted oxidoreductases of the aldo/keto reductase family - COG1453	COG1453	Predicted oxidoreductase of the aldo/keto reductase family	pfam00248	Aldo/keto reductase family		
23	22847_23359	513	hypothetical protein			pfam14537	Cytochrome c3		
24	23415_25460	2046	hypothetical protein	COG3345	Alpha-galactosidase				
25	25608_26009	402	hypothetical protein						
26	26017_27051	1035	hypothetical protein						
27	27155_28189	1035	hypothetical protein						
28	28597_31944	3348	hypothetical protein						
29	32118_35273	3156	Predicted bile acid beta-glucosidase	COG4354	Uncharacterized protein, contains GBA2_N and DUF608 domains	pfam04685; pfam12215	Glycosyl-hydrolase family 116, catalytic region; beta-glucosidase 2, glycosyl-hydrolase family 116 N-term		
30	35341_37368	2028	Protein of unknown function (DUF2961)			pfam11175	Protein of unknown function (DUF2961)		
FOS10									
ORF	**Locus**	**Length**	**Putative function**	**COG ID**	**COG Function**	**Pfam ID**	**Pfam Function**	**EC Number**	**Enzyme Function**
1	1_110	110	hypothetical protein						
2	233_1093	861	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	COG2876	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	pfam00793	DAHP synthetase I family	EC:2.5.1.54	3-deoxy-7-phosphoheptulonate synthase.
3	1123_2151	1029	3-dehydroquinate synthetase	COG0337	3-dehydroquinate synthetase	pfam01761	3-dehydroquinate synthase	EC:4.2.3.4	3-dehydroquinate synthase.
4	2339_2857	519	hypothetical protein						
5	3014_4213	1200	Aspartate/tyrosine/aromatic aminotransferase	COG0436	Aspartate/methionine/tyrosine aminotransferase	pfam00155	Aminotransferase class I and II		

6	4204_4785	582	Glycine/sarcosine/betaine reductase selenoprotein B (GRDB)						
7	4804_5094	291	hypothetical protein						
8	5241_5909	669	Zn-dependent hydrolases, including glyoxylases	COG0491	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II				
9	5916_6440	525	hypothetical protein						
10	6673_8445	1773	Citrate transporter	COG0471; COG0569; COG1055	Di- and tricarboxylate transporter; Trk K+ transport system, NAD-binding component; Na+/H+ antiporter NhaD or related arsenite permease	pfam03600	Citrate transporter		
11	8762_9445	684	Predicted Fe-S-cluster oxidoreductase	COG0727	Fe-S-cluster containining protein	pfam03692	Putative zinc- or iron-chelating domain		
12	9473_10165	693	Uncharacterized protein involved in cysteine biosynthesis	COG2981	Uncharacterized protein involved in cysteine biosynthesis	pfam07264	Etoposide-induced protein 2.4 (EI24)		
13	10435_11025	591	Predicted transcriptional regulators - COG1695	COG1695	DNA-binding transcriptional regulator, PadR family	pfam03551	Transcriptional regulator PadR-like family		
14	11028_11118	91	hypothetical protein						
15	11146_11283	138	hypothetical protein						
16	11605_11733	129	hypothetical protein					EC:5.3.2.6	2-hydroxymuconate tautomerase.
17	11737_11838	102	Tautomerase enzyme			pfam01361	Tautomerase enzyme	EC:5.3.2.6	2-hydroxymuconate tautomerase.
18	12063_12767	705	Transcriptional regulators	COG2186	DNA-binding transcriptional regulator, FadR family	pfam00392; pfam07729	Bacterial regulatory proteins, gntR family; FCD domain		
19	13254_13364	111	hypothetical protein						
20	13381_14544	1164	hypothetical protein						

21	14740_17682	2943	FAD/FMN-containing dehydrogenases	COG0247; COG0277	Fe-S oxidoreductase; FAD/FMN-containing dehydrogenase	pfam01565; pfam02754; pfam02913; pfam13183	FAD binding domain; Cysteine-rich domain; FAD linked oxidases, C-terminal domain; 4Fe-4S dicluster domain		
22	17699_18592	894	Uncharacterized proteins, LmbE homologs	COG2120	N-acetylglucosaminyl deacetylase, LmbE family	pfam02585	GlcNAc-PI de-N-acetylase	EC:3.5.1.115	Mycothiol S-conjugate amidase.
23	18998_20818	1821	Predicted membrane GTPase involved in stress response - COG1217	COG1217	Predicted membrane GTPase involved in stress response	pfam00009; pfam00679	Elongation factor Tu GTP binding domain; Elongation factor G C-terminus		
24	20966_21472	507	Uncharacterized paraquat-inducible protein A	COG2995	Uncharacterized paraquat-inducible protein A	pfam04403	Paraquat-inducible protein A		
25	21445_22110	666	Uncharacterized paraquat-inducible protein A	COG2995	Uncharacterized paraquat-inducible protein A	pfam04403	Paraquat-inducible protein A		
26	22107_23837	1731	Paraquat-inducible protein B	COG3008	Paraquat-inducible protein B	pfam02470	MlaD protein		
27	23834_24484	651	Uncharacterized protein conserved in bacteria	COG3009	Uncharacterized lipoprotein YmbA	pfam03886	ABC-type transport auxiliary lipoprotein component		
28	24745_26967	2223	Citrate transporter	COG0471; COG0490; COG3273	Di- and tricarboxylate transporter; K+/H+ antiporter YhaU, regulatory subunit KhtT; Uncharacterized conserved protein, contains PhoU and TrkA_C domains	pfam03600	Citrate transporter		
29	27153_27294	142	hypothetical protein						

Table S2. List of fosmid clones that grew in the presence of aromatic hydrocarbons and exhibited respiratory activity in the MTT assay.

Positive clones					
Plate	**Phenol**	**Naphthalene**	**Phenanthrene**	**Pyrene**	**Benzopyrene**

PL1	2D, 2G				
PL2			11H, 12G, 12H	9D	9D
PL4				11C	
PL9	12C	1H			
PL11	2A, 8E, 10B	3A, 6H			3A, 5A
PL12	1D, 3E, 6F, 7G, 9A, 9H, 10G, 11G	12H	8E		
PL13	1G, 3B, 12E		12B, 12D, 12E	12B, 12E	10C, 1D, 12B, 12E
PL14	4A, 5C, 6G, 7C, 7E, 9H	9E		12C, 12E	12C
PL15	6A, 7C, 1E			10G, 10A	
PL18		1A			
PL19		4G, 4H, 3H			

Activity Scale: blue (+); green (++); red (+++).

Table S3. Taxonomic annotation of *contigs* related to catechol 2,3-dioxygenase, protocatechuate 4,5-dioxygenase and ring-hydroxylating dioxygenases by the genbank database.

Catechol 2,3-dioxygenase					
Contig ID	**Length (pb)**	**Closest bacterial hit**	**Query cover**	**Ident**	**Accession**
4	1600	*Thermus thermophilus*	60%	64%	WP_024119643.1
5	1522	*Pseudorhodoplanes sinuspersici*	62%	82%	WP_086089020.1
6	797	*Enhydrobacter aerosaccus*	65%	81%	WP_085935340.1
7	505	*Curtobacterium pusillum*	98%	58%	WP_079237089.1
9	495	*Enhydrobacter aerosaccus*	88%	72%	WP_085935340.1
13	1090	*Nocardioides psychrotolerans*	95%	72%	WP_091116447.1
15	823	*Mameliella alba*	59%	74%	WP_043145134.1
16	1527	*Mameliella alba*	63%	79%	WP_043145134.1
17	510	*Intrasporangium calvum*	99%	88%	WP_013491942.1

18	714	*Novosphingobium naphthalenivorans*	76%	79%	WP_067736631.1
Protocatechuate 4,5-dioxygenase					
Contig ID	**Length (pb)**	**Closest bacterial hit**	**Query cover**	**Ident**	**Accession**
1	1072	*Pseudorhodoplanes sinuspersici*	69%	79%	WP_086089858.1
2	515	*Azospirillum lipoferum*	99%	65%	WP_085557883.1
3	481	*Rhodopseudomonas palustris*	77%	77%	WP_044416565.1
8	796	*Sphingomonas koreensis*	96%	64%	WP_066577341.1
10	1383	*Thalassobius gelatinovorus*	60%	86%	WP_058263901.1
11	1428	*Thalassobius aestuarii*	59%	93%	WP_093090996.1
12	858	*Marmoricola aequoreus*	73%	82%	WP_030483973.1
Ring-hydroxylating dioxygenases					
Contig ID	**Length (pb)**	**Closest bacterial hit**	**Query cover**	**Ident**	**Accession**
4	1040	*Frankia asymbiotica*	89%	62%	WP_081438026.1
5	1878	*Altererythrobacter mangrovi*	55%	59%	WP_095012085.1
7	1246	*Solimonas aquatica*	94%	65%	WP_093288465.1
11	1194	*Novosphingobium naphthalenivorans*	70%	71%	WP_067736669.1
16	1656	*Pseudolabrys sp.*	78%	83%	WP_056912887.1
17	671	*Paraburkholderia oxyphila*	66%	63%	WP_028223495.1
21	1697	*Novosphingobium mathurense*	77%	85%	WP_013832634.1
22	806	*Novosphingobium malaysiense*	69%	78%	WP_039287354.1
24	1296	*Cryobacterium roopkundense*	50%	65%	WP_035837261.1
28	745	*Bradyrhizobium sp.*	86%	65%	WP_049823273.1
34	1121	*Amorphus coralli*	90%	85%	WP_018698440.1
38	1222	*Oceanicoccus sagamiensis*	77%	51%	WP_085758833.1

CAPÍTULO III

Clonagem, expressão e caracterização de novas dioxigenases a partir de biblioteca metagenômica fosmidial de mangue contaminado

8. INTRODUÇÃO

O bioma manguezal está localizado na zona fronteiriça entre o mar e a terra, e é essencial para a manutenção de uma vasta diversidade biológica. É um bioma caracterizado pela alta salinidade, grandes variações no nível das marés, muita sedimentação e flutuações na taxa de oxigênio no solo (Giri *et al.* 2011). O Brasil possui 7% (20.000 km^2) da distribuição mundial dos manguezais, boa parte ocupados por cidades densamente povoadas ao longo de toda costa marítima (Holguin, Vazquez e Bashan 2001). As atividades antropogênicas têm impactado grandes áreas de manguezal, principalmente através da exploração, transporte e refino do petróleo (Margesin *et al.* 2003; Le Borgne, Paniagua e Vazquez-Duhalt 2008). Na década de 80 ocorreu um grande vazamento de petróleo na região de Bertioga-SP, afetando uma vasta área de manguezal (Coimbra 2006). Seus componentes ficaram impregnados no sedimento úmido, ceifando diversas espécies de plantas, animais e microrganismos. O petróleo é uma matéria prima não-renovável usada, dentre outras coisas, para a produção de combustíveis veiculares. O petróleo é formado por uma mistura de componentes moleculares constituídos principalmente por carbono e hidrogênio (hidrocarbonetos) (Sugiura *et al.* 1997).

Os hidrocarbonetos são substâncias amplamente distribuídas em todo o planeta e são classificados em compostos cíclicos ou alifáticos/saturados, aromáticos e compostos polares (Yender *et al.* 2002). Os hidrocarbonetos aromáticos policíclicos (HAPs) são poluentes ambientais altamente tóxicos e recalcitrantes que apresentam dois ou mais anéis benzênicos fundidos (Cerniglia 1992; Ghosal *et al.* 2016). Geralmente, estes compostos são usados na indústria química para produção de corantes, plásticos, pesticidas, farmacêuticos, conservantes de madeira e resinas, dentre outros (Abdel-Shafy e Mansour 2016). Como consequência, o uso constante dos HAPs na indústria química e de petróleo gera resíduos que vão parar na natureza. As principais formas de eliminação dos HAPs do ambiente são através da volatilização, foto-oxidação, oxidação química, bioacumulação, adsorção a partículas sólidas ou degradação enzimática por microrganismos (Semple, Reid e Fermor 2001; Singh *et al.* 2008; Megharaj *et al.* 2011).

A microbiota indígena desempenha um papel crucial na degradação de HAPs através da ciclagem de carbono, atenuando seus impactos ambientais. Nas bactérias, as principais vias metabólicas de degradação de HAPs é iniciada a partir da di-hidroxilação dos compostos aromáticos, catalisada por dioxigenases que hidroxilam anéis aromáticos (do inglês, *Aromatic Ring-hydroxylating dioxygenases – ARHD*) (Peng *et al.* 2008). As dioxigenases se destacam pela sua importância médica e econômica, através da produção de antioxidantes, pigmentos e *synthons* quirais (componente químico usado na produção de compostos quirais), bem como o seu uso na engenharia genética e biologia sintética para o desenvolvimento de cepas

melhoradas, empregadas em processos de biorremediação (Nolan e O'Connor 2008; Tan e Parales 2016). Estas enzimas são constituídas por quatro componentes: subunidade grande (codificada por *pahAc*), subunidade pequena (*pahAd*), uma redutase (*pahAa*), e uma ferrodoxina (*pahAb*) (Kweon *et al.* 2008). A região da subunidade grande é normalmente usada como marcador genético para caracterização ambiental de bactérias que degradam HAPs (Ding *et al.* 2010). Um estudo recente, que compilou resultados de outros trabalhos, revelou que mais de dez *clusters* gênicos, envolvidos na biodegradação de HAPs já foram identificados, a maioria oriunda de ambientes terrestres (Wang *et al.* 2017). Vários clusters com a mesma estrutura genética foram encontrados no genoma de espécies do mesmo gênero, família ou ordem, o que sugere a presença de um ancestral comum ou eventos de transferência horizontal de genes (DeBruyn, Mead e Sayler 2012). O aparecimento e fixação destes clusters gênicos em comunidades microbianas pode ter ligação com a preservação das espécies, atenuando os efeitos tóxicos e acumulativos dos HAPs, além de usá-los como fontes alternativas de energia molecular.

Devido a interesses econômicos e científicos, pesquisadores têm focado seus esforços na prospecção de enzimas com atividade catalítica sob condições extremas, tais como temperatura, salinidade e pH (Hough e Danson 1999; van den Burg 2003; Schiraldi e De Rosa 2017). Para alcançar bons resultados, pesquisadores tendem a escolher alvos ambientais com características específicas que possam favorecer o estabelecimento de comunidades microbianas com a bioquímica adaptada para condições desejadas. Ambientes marinhos e de manguezal podem ser uma fonte promissora de microrganismos halotolerantes ou halofílicos que sintetizam enzimas com aplicações em bioprocessos sob elevadas concentrações de sais (Thompson *et al.* 2013; Alma'abadi, Gojobori e Mineta 2015). Recentemente, dioxigenases de catecol halotolerantes foram obtidas a partir de espécies de *Marinobacter* isoladas de solos salinos contaminados com óleo (Guo *et al.* 2015). Neste trabalho, duas novas catecol 2,3-dioxigenases foram expressas em sistema heterólogo com sucesso e caracterizadas funcionalmente, apresentando atividade catalítica em meio com até 30% de NaCl e 60°C de temperatura (Guo *et al.* 2015). Baseado na análise do potencial eletrostático da superfície enzimática, os autores também revelaram que a propriedade halotolerante da enzima pode ser devida à presença de resíduos ácidos de aminoácidos na superfície enzimática (Guo *et al.* 2015). Dois anos depois, membros do mesmo grupo de pesquisa publicaram um trabalho relatando a primeira dioxigenase relacionada a *Nah* (*Nah-related dioxygenase*) halotolerante, oriunda de um consórcio bacteriano dominado pelo gênero *Marinobacter* (40,67%), clonada e expressa com sucesso em *E. coli* (Wang *et al*. 2017). Eles observaram também a presença de resíduos

ácidos de aminoácidos na dioxigenase, os quais são características estruturais comuns em enzimas halofílicas e halotolerantes (Wang *et al*. 2017). Lin e Milase (2015) efetuaram a purificação e caracterização de uma catecol 1,2-dioxigenase oriunda da cepa *Acinetobacter* sp. Y64 e transformada em *Escherichia coli* BL21 DE3. Após a clonagem e purificação, a catecol 1,2-dioxigenase não apresentou mudanças em suas propriedades bioquímicas, demonstrando a eficiência do processo (Lin e Milase 2015). Alguns estudos têm voltado sua atenção para elucidação estrutural e funcional de dioxigenases de HAPs com alto peso molecular, motivados pela perspectiva de usá-las na biorremediação (Kweon *et al.* 2010; Wu *et al.* 2014; Li *et al.* 2015). Peng *et al.* (2014) usaram a engenharia genética para desenvolver uma *Arabidopsis thaliana* transgênica contendo genes relacionados ao sistema de dioxigenases de naftaleno, isolados de *Mycobacterium vanbaalenii* PYR-1 e *Pseudomonas putida* G7. Os autores evidenciaram que as plantas transformadas conseguiram super-expressar os genes exógenos, bem como assimilar e degradar mais naftaleno, antraceno e pireno do que as do tipo selvagem (Peng *et al*. 2014). Em outro estudo, Khara *et al.* (2014) realizaram a caracterização funcional de várias oxigenases que hidroxilam anéis aromáticos capazes de transformar etillbenzeno, propilbenzeno, cumeno, p-cimeno, bifenil e alguns hidrocarbonetos aromáticos policíclicos. Em um estudo mais recente, Zeng *et al*. (2017) isolaram genes para dioxigenases que hidroxilam anéis aromáticos a partir de *Mycobacterium* sp. NJS-P e expressaram-nos com sucesso em *E. coli* BL21 (DE3), além disso, a proteína heteróloga foi capaz de oxidar HAPs tais como antraceno, fenantreno, pireno e benzo[a]pireno. Os autores também realizaram uma análise *in silico* de modelagem molecular por homologia, revelando um grande sítio de ligação a prováveis substratos de HAPs com alto peso molecular (Zeng *et al*. 2017).

Há algumas décadas, havia muita dificuldade no isolamento de microrganismos ambientais através de métodos convencionais, o qual limitava os resultados sobre a diversidade microbiana e mitigava a aquisição de informações genéticas preciosas. O advento das tecnologias de sequenciamento de DNA em larga escala, aliado à abordagem metagenômica, ampliou o interesse científico pelos recursos genéticos ainda não cultivados presentes nas comunidades microbianas ambientais (Handelsman *et al*. 1998). O processamento de dados metagenômicos permitiu acessar novos genes com origem desconhecida, derivados de novas espécies microbianas. Os pesquisadores têm como objetivo encontrar novas proteínas com atividade catalítica superior para poder usá-las na medicina, indústria, agricultura, remediação de ambientes contaminados, dentre outras funções (Simon e Daniel 2011; Quince *et al.* 2017).

Devido ao histórico de contaminação por petróleo, associado às condições de salinidade do solo e seu aspecto sedimentar e húmido, a região de manguezal localizada no município de

Bertioga-SP foi escolhida como fonte de recursos genéticos para a descoberta de novas enzimas (ex. dioxigenases) envolvidas na degradação de hidrocarbonetos aromáticos. Neste contexto, partimos do pressuposto de que estas condições podem favorecer a aquisição de genes para enzimas halofílicas/halotolerantes e termoestáveis do nosso interesse. Portanto, após identificar prováveis genes de interesse a partir do metagenoma do manguezal (descrito no capítulo II), nosso objetivo foi realizar a clonagem, expressão, purificação e caracterização funcional e estrutural destes alvos. Os experimentos foram realizados com sucesso até a etapa de expressão heteróloga, quando o tempo de vigência do doutoramento se esgotou. Etapas futuras de caracterização funcional e estrutural das dioxigenases clonadas resultarão no primeiro relato de dioxigenases que hidroxilam anéis aromáticos isoladas de manguezal brasileiro, usando a abordagem metagenômica como ferramenta para bioprospecção. Diante da premissa de que ambientes marinhos e de manguezal são os mais acometidos por impactos causados pela indústria petroquímica, estas enzimas poderiam ser empregadas posteriormente em processos de remediação desses próprios habitats.

9. MATERIAL E MÉTODOS

9.1. Extração e sequenciamento do DNA fosmidial

Os clones fosmidiais que apresentaram melhor crescimento nos ensaios funcionais com MTT, descrito anteriormente no capítulo II desta tese, foram escolhidos para o sequenciamento e caracterização genética do DNA exógeno. Para isso, o isolamento do DNA fosmidial foi realizado com o *kit QIAGEN Large-construct* (Qiagen Sample & Assay Technologies, Germany), seguindo as instruções do fabricante. Antes da extração, 10 clones foram reativados em 2 mL de meio LB contendo Cloranfenicol (12,5 µg/mL) e L-arabinose (0,02%), e incubados a 37°C e 180 rpm por 16 h. Uma alíquota de 500 µL do cultivo do clone foi adicionada em 100 mL do mesmo meio LB em frasco Erlenmeyer de 250 mL. A cultura foi incubada a 37°C e 180 rpm por 16 h. Após o crescimento, as células foram centrifugadas e submetidas à extração do DNA fosmidial.

Todas as amostras de DNA foram devidamente quantificadas em Nanodrop (Thermo Scientific) e enviadas para o sequenciamento em plataforma Illumina MISEQ na empresa MR DNA (Molecular Research LP) nos Estados Unidos da América.

9.2. *Screening in silico* de genes para dioxigenases do metabolismo de HAs

Os dados brutos gerados pelo sequenciamento foram processados utilizando o seguinte *workflow*: primeiro, as sequências *pair-end* foram combinadas com o auxílio do *software Flash* versão 1.2.11; em seguida, as sequências foram trimadas no *software* Trimmomatic versão 0.33 e salvas em fastqc para avaliação da qualidade; a normalização digital dos dados foi feita no *software* Trinity versão 2.0.6; foi feita a limpeza de artefatos de *E. coli* e fosmídeo contra um banco de dados híbrido (Bowtie2 + Banco de *E.colis* + UniVEC); por fim, a montagem dos *contigs* foi realizada no *software* Spades. Após o tratamento das sequências, os *contigs* foram submetidos nas plataformas RAST 2.0 (*The Rapid Annotation using Subsystem Technology*) e IMG/M (*Integrated Microbial Genomes & Microbiomes*) para o mapeamento de ORFs e classificação funcional das sequências, seguindo os parâmetros de análise padrão do servidor. As ORFs de prováveis genes para dioxigenases foram identificadas e separadas.

9.3. Clonagem e expressão heteróloga de possíveis alvos gênicos para dioxigenases

Após os ensaios funcionais, sequenciamento de clones positivos e anotação funcional das ORFs preditas a partir dos insertos fosmidiais, foram escolhidos quatro alvos gênicos que codificam para possíveis dioxigenases (envolvidas nas primeiras etapas das vias de degradação

de HA) e um alvo gênico para epóxido-hidrolase (também pode estar envolvida com processos de degradação de compostos aromáticos) (Tabela 1). Estes alvos foram submetidos a protocolos de clonagem e testes de expressão para posterior purificação e caracterização funcional e estrutural. Inicialmente, foram desenhados *primers* específicos (Tabela 1) para flanquear os alvos gênicos individualmente e enviados para síntese na empresa EXXTEND (Exxtend Inc., São Paulo, Brasil). O DNA fosmidial contendo cada alvo, foi usado como *template* em reações de PCR para amplificação do fragmento de interesse. Valores definidos para um volume final de 25 µL: 30 ng de DNA fosmidial, 1X de tampão de reação (10X), 1,5 mM $MgCl_2$, 0,1 mM de dNTP, 0,2 µM dos *primers* Forward e Reverse (Tabela 1), e 0,05 U/µL de *Taq* DNA polimerase (5 U/µL). Programa da reação de PCR com 32 ciclos: desnaturação inicial 95°C por 5 min, desnaturação subsequente a 95°C por 30 s; anelamento a 59°C por 30 s; extensão a 72°C por 1,5 min e extensão final de 72°C por 7 min. O produto de PCR foi visualizado por eletroforese em gel de agarose (1%) e purificado usando o *kit Illustra GFX PCR Purification* (GE Healthcare Life Sciences), seguindo as instruções do fabricante. Após a purificação, foram feitas as reações de digestão dos produtos de PCR e do vetor de clonagem pET28a(+) (Novagen, USA) usando as enzimas de restrição *Nde*I e *Xho*I (Fermentas, USA) (Figura 2) (Reação Tabela 2). As reações foram incubadas a 37°C por 1 h 20 min, submetidas à eletroforese para diagnóstico de reação, purificadas com o *kit Illustra GFX PCR Purification* e quantificadas em Nanodrop (*Thermo Scientific*). O produto da digestão foi submetido à reação de ligação ao vetor pET28a(+) usando a T4 DNA ligase (Invitrogen, USA), por 16 h a 15°C (Reação Tabela 3). Após a reação de ligação, 0,5 µL de cada construto foi misturado em 40 µL de solução contendo células de *E. coli* BL21(DE3) (Novagen, USA) para eletroporação, seguido pela adição de 960 µL de meio LB e incubação durante 1 h 20 min a 37°C. O mesmo procedimento foi repetido para a transformação em *E. coli* Rosetta (Novagen, USA). Uma alíquota de 200 µL do cultivo foi plaqueada em meio LB sólido contendo canamicina (50 µg/mL), seguido de incubação a 37°C por 16 h. Para confirmar o sucesso da clonagem foram feitas reações de PCR de colônia usando os *primers* T7, que flanqueiam a região do inserto. Os reagentes para amplificação foram calculados para um volume final de 25 µL: 2 µL da solução de células, 1X de tampão (10X), 2,5 mM de $MgCl_2$, 0,2 mM de solução de dNTP, 0,5 µM dos *primers* T7 *Forward* e *Reverse*, e 0,08 U/µL de *Taq* DNA polimerase (5 U/µL). As reações foram submetidas às seguintes condições: desnaturação inicial 95°C por 3 min, desnaturação subsequente a 95°C por 1 min; anelamento a 55°C por 1,5 min; extensão a 72°C por 1,5 min e extensão final de 72°C por 3 min.

Tabela 1. Conjunto de *primers* específicos desenhados para flanquear ORFs alvos de prováveis genes de interesse anotados contra o banco de dados COG.

Amostra	ORFs – COG	Tamanho pb	Identidade (%)	*Primers F e R*
1	Polyferredoxin Domain	1195	57	5' - AACATATGTCGTACTTCGGCCGCCTCTC - 3' 5' - AACTCGAGTTACGGCTGATCTTCCACATCG - 3'
2	Predicted Fe-S-cluster oxidoreductase	697	40	5' - AACATATGGATCTCGAAGGCCGCATTGAG - 3' 5' - AACTCGAGTTACTGCTGGGCTTTTAGCAG - 3'
3	Catechol 2,3-dioxygenase	473	50	5' – CTCGGGAGACACCATATGATTCTGGCAAC - 3' 5' - AACTCGAGCTAGTCTTCGTCGCGC - 3'
4	Predicted Fe-S-cluster oxidoreductase	1545	63	5' - AACATATGCAAGACGTCGAAGCCATGTGG - 3' 5' - AACTCGAGTTCAAGCGGTGGCAC - 3'
5	Epoxide hydrolase	955	67	5' - AACATATGACGCCCGGCCTTGTG - 3' 5' - AACTCGAGTTAGCGCCGCATCAG - 3'

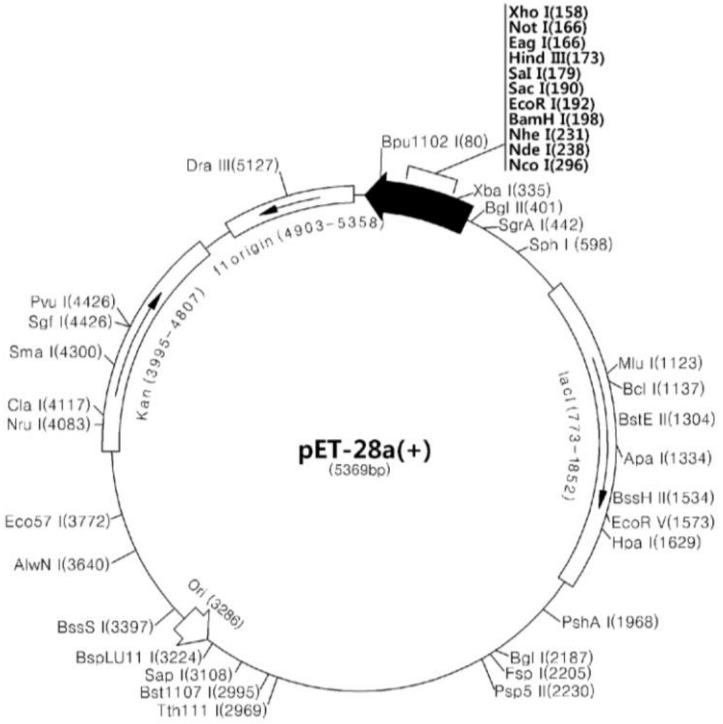


Figura 2. Mapa do vetor de expressão pET-28a(+). Seta preta indica a região com múltiplos sítios de clivagem por endonucleases de restrição, incluindo os palíndromos de reconhecimento para NdeI e XhoI, usados neste estudo para ligação do gene exógeno.

Tabela 2: Definição dos valores calculados de cada reagente para as reações de digestão.

	DNA (µL)	*Tampão (µL)*	*NdeI (µL)*	*XhoI (µL)*	*H_2O (µL)*
Amostra 1	16,4	3	1	1,5	8,1
Amostra 2	16	3	1	1	9
Amostra 3	17,2	3	1	1,5	7,3
Amostra 4	33	4	1	1	0
Amostra 5	15	3	1	1	10
Vetor pET28a(+)	10	2	1	1,5	5,5

Tabela 3: Definição dos valores calculados de cada reagente para as reações de ligação.

	Inserto (µL)	*Vetor pET28a(+) (µL)*	*Tampão (µL)*	*T4 DNA ligase (µL)*	*H_2O (µL)*
Amostra 1	2	4,7	4	1	8,3
Amostra 2	3	6	4	1	6
Amostra 3	2	4,6	4	1	8,4
Amostra 4	3	6	4	1	6
Amostra 5	3	6	4	1	6

Após a confirmação da clonagem, uma colônia de cada amostra foi cultivada em 5 mL de meio LB líquido suplementado com canamicina (50 µg/mL) e incubada a 37°C por cerca de 18 h para preparar um pré-inoculo. Para os testes de expressão, 1 mL de cada pré-inoculo foi adicionado em 50 mL de meio LB com canamicina (50 µg/mL) e IPTG (isopropyl b-D-1-thiogalactopyranoside) a uma concentração final de 0,4 mM. As culturas celulares foram incubadas a 37°C até atingir a fase exponencial (DO_{600nm} entre 0,6 e 0,8). Alíquotas de 500 µL foram coletadas para posterior extração de proteínas da seguinte forma: antes de adicionar o IPTG (OD de 0,8); após 1 hora; 2 horas; 4 horas; e 16 h de cultivo. Todas as alíquotas foram submetidas ao protocolo de extração de proteínas usado anteriormente de Beloti *et al.* (2013). As amostras foram diluídas em tampão fosfato (50 mM, pH 7.2) suplementado com Tween-20 (0,2%), β-mercaptoetanol (12 mM), lisozima (1 mg/mL), e PMSF - *Phenylmethylsulfonyl Fluoride* (1 mM) e incubado em gelo por 30 min. As células foram rompidas por sonicação (Ultrasonic Homogenizer 4710) com 70% de ciclo de trabalho, e a fração solúvel foi recuperada por centrifugação a 27.000 g por 50 min a 4°C. A constatação da expressão de proteínas foi feita usando gel SDS-PAGE com 12,5% de poliacrilamida e sob condições de desnaturação, descrito anteriormente por Laemmli (1970).

10. RESULTADOS E DISCUSSÃO

10.1. Caracterização taxonômica e funcional de genes alvos

As análises de correspondência dos genes alvos no BLASTx foram realizadas antes da clonagem. A amostra 1 apresentou 57% de identidade a uma *4Fe-4S binding protein* oriunda de *Desulfobacter curvatus*. Estudos anteriores isolaram membros deste gênero de sedimento marinho anaeróbico e constataram sua capacidade litoautotrófica de produzir energia a partir de compostos de origem mineral e usar acetado como doador de elétrons (Widdel 1987). O gene alvo da amostra 2 apresentou 40% de identidade com uma proteína hipotética obtida de *Desulfococcus multivorans,* detendo uma região conservada para oxidorredutase com um domínio para ferrodoxina. O gene alvo correspondente a amostra 3 foi anotado como uma catecol 2,3-dioxigenase obtida de *Amycolatopsis saalfeldensis*, com 41% de identidade. O gene da amostra 4 apresentou um domínio correspondente a hidroxilase de anéis aromáticos com 67% de identidade com uma 4-hidroxibutiril-CoA dehidratase oriunda de uma bactéria da ordem *Rhizobiales*. Espécies bacterianas desta ordem são conhecidas por serem fixadoras de nitrogênio, metanotróficas e viverem dentro de nódulos nas raízes de plantas em uma relação simbiótica (Erlacher *et al.* 2015). O alvo gênico da amostra 5 apresentou 61% de identidade com uma epoxido hidrolase contida no genoma de *Caballeronia sordidicola*. Como já foi discutido no capítulo II, esta enzima está envolvida com processos de detoxificação de compostos xenobióticos assim como na degradação de compostos aromáticos (Arand *et al.* 2005). Epóxido hidrolases podem atuar na inativação de HAPs carcinogênicos portadores de grupos epóxidos altamente reativos (Shimada e Fujii-Kuriyama 2003). Uma revisão publicada por Decker, Arand e Cronin (2009) mostrou que o benzo[a]pireno pode ser convertido em Benzo[a]pireno-7,8-epóxido, um composto genotóxico, por uma monoxigenase citocromo P450. Entretanto, este último pode ser detoxificado por uma epóxido hidrolase até o Benzo[a]pireno-7,8-diidrodiol. A atividade catalítica de detoxificação das epóxido hidrolases pode se estender para outros compostos como benzeno, naftaleno, antraceno e outros hidrocarbonetos aromáticos policíclicos (Decker, Arand e Cronin 2009). Estudo anterior constatou um alto nível de expressão de epóxido hidrolases citoplasmáticas e microssomal no fígado, responsável pela detoxificação de epóxidos carcinogênicos (Argiriadi *et al.* 1999). A quebra de HAs com grupos epóxidos genotóxicos é uma excelente estratégia biológica para proteger macromoléculas como DNA, portanto, epóxido hidrolases podem ser usadas em bioprocessos que envolvem a produção de metabólitos intermediários menos tóxicos para vida.

10.2. Expressão de proteínas

Sequências nucleotídicas correspondentes aos genes alvos, de tamanhos esperados, foram recuperadas a partir das reações de PCR do DNA fosmidial. As evidências foram visualizadas por eletroforese em gel de agarose e confirmaram a eficiência dos *primers* desenhados (Figura 3). A eficiência da clonagem em vetor de expressão heteróloga foi confirmada por eletroforese do produto de PCR gerado a partir do DNA plasmidial contendo os genes alvos. O tamanho do fragmento amplificado, e visualizado em gel de agarose, é a soma do tamanho do inserto exógeno com 250 pb de regiões do DNA plasmidial que flanqueiam o sítio de corte (Figura 3).

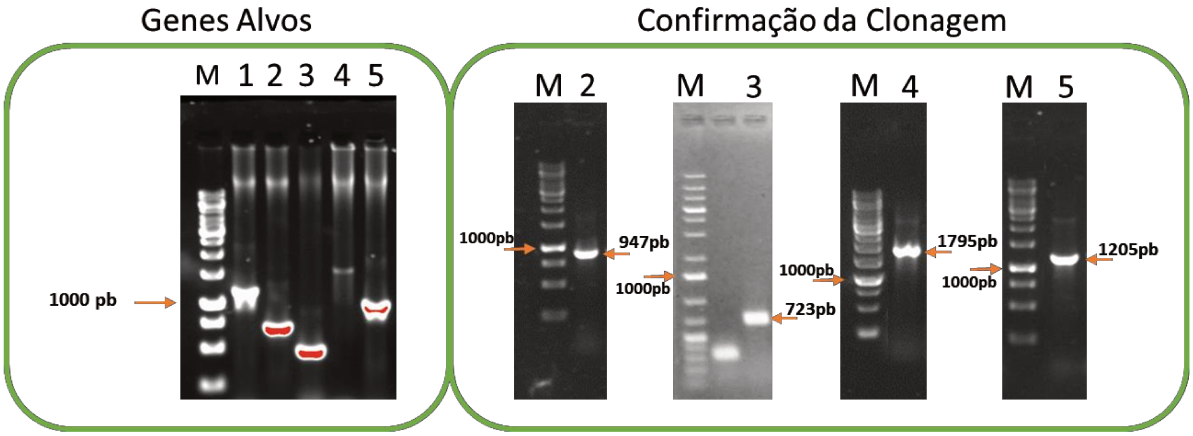


Figura 3. Diagnóstico por eletroforese das etapas de obtenção de genes alvos a partir do DNA fosmidial e da confirmação da clonagem em sistemas de expressão. M = Marcador de Peso Molecular 1kb 1 Kb *Plus DNA Ladder* (Invitrogen); Genes alvos 1 (Polyferredoxin Domain), 2 (Predicted Fe-S-cluster oxidoreductase), 3 (Catechol 2,3-dioxygenase), 4 (Predicted Fe-S-cluster oxidoreductase) e 5 (Epoxide hydrolase).

Após a confirmação da clonagem de 4 dos 5 genes alvos, foram feitos testes de indução para a expressão de proteínas em diferentes tempos. As proteínas extraídas do sobrenadante e do *pellet* da cultura de células centrifugada foram submetidas à eletroforese em gel SDS-PAGE sob condições de desnaturação. Todas as amostras apresentaram sinais de expressão, principalmente aquelas cultivadas por 16 h. Entretanto, apenas a proteína correspondente ao gene alvo 3 foi obtida na fase solúvel (sobrenadante), enquanto as proteínas correspondentes aos genes 2, 4 e 5 ficaram agregadas ao *pellet* insolúvel (Figura 4). A obtenção de proteínas na fase insolúvel se torna um problema para os procedimentos de purificação.

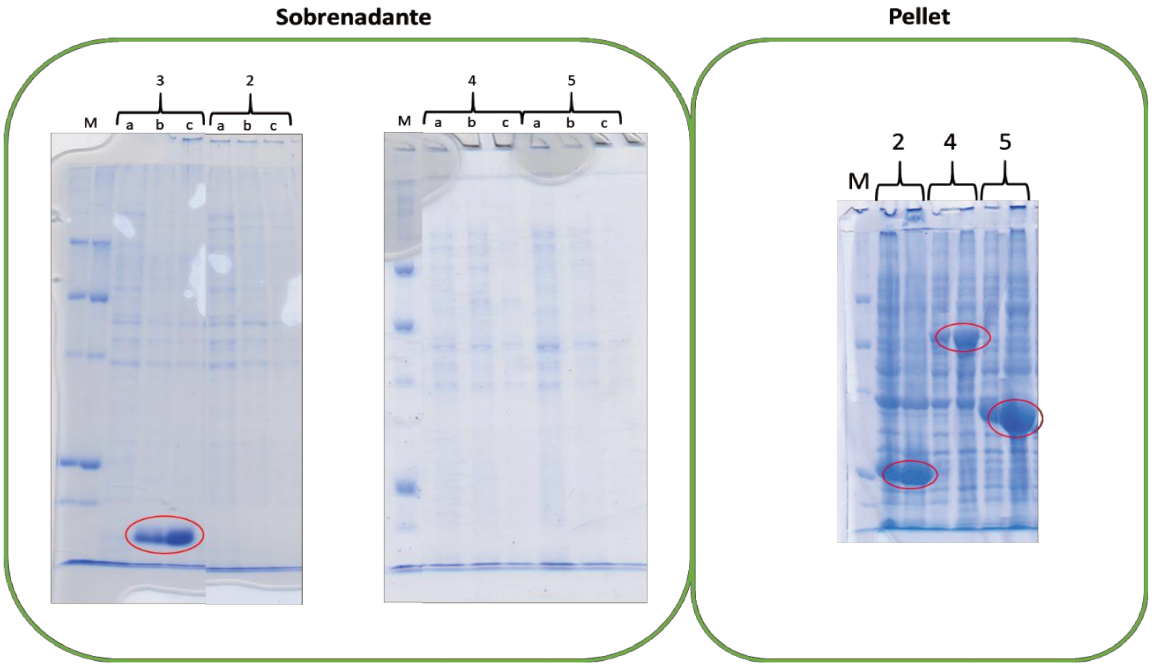


Figura 4. Gel SDS-PAGE das proteínas totais extraídas do sobrenadante e pellet após o teste de expressão. O círculo vermelho evidencia a expressão das proteínas de interesse. M = Marcador de peso molecular; Amostras 2, 3, 4, 5; Tempo de indução: a = 0 h, b = 4 h, c = overnight.

Geralmente, a maioria das proteínas obtidas em sistemas de expressão heteróloga, os quais incluem a *E. coli* como célula hospedeira, é encontrada na fração insolúvel formando corpos de inclusão (Yang *et al.* 2011; Palmer e Wingfield 2012). O grande desafio durante o processo de extração é solubilizar os corpos de inclusão e refazer o enovelamento das proteínas até sua conformação ativa para analisar sua atividade biológica (Palmer e Wingfield 2012). Usualmente, são testados outros tampões de extração de proteínas, alterando concentrações dos diferentes solutos, adicionando outros detergentes (ex. Triton, Tween) ou reagentes que reduzam a polaridade do sobrenadante, ou até mesmo outra estratégia de clonagem, removendo regiões hidrofóbicas da proteína para aumentar sua solubilidade, sem comprometer sua atividade funcional. A lavagem dos corpos de inclusão pode ser feita com a adição de desoxicolato de sódio (2%), Triton X-100 (1%) e ureia (1–2 M), buscando a combinação e a concentração ótima (Burgess 2009). A etapa de lavagem tem o propósito de remover o máximo do material de membrana celular, DNA e proteínas indesejadas. Após a lavagem, a solubilização dos corpos de inclusão pode ser facilitada com a adição e/ou combinação de agentes desnaturantes tais como Hidrocloreto de Guanidina (GuHCl) a 6 M, Ureia (8 M), Sarcosil (0,3%) ou Dodecil sulfato de sódio (SDS) (Burgess 2009). Ao final do processo, na tentativa de refazer o enovelamento nativo, a solução de proteínas desnaturadas pode ser diluída

de três formas: Diluição reversa, Diluição instantânea ou Diluição por gotejamento (Burgess 2009). Alguns tratamentos físicos, tais como temperatura, centrifugação e sonicação, também podem ser alterados para recuperar proteínas alvos na fração solúvel. Resumindo, as condições ideais de extração podem variar de proteína para proteína. A primeira alternativa para o nosso estudo seria diminuir o tempo de indução e a concentração do agente indutor da expressão (IPTG), além de usar um kit de extração de proteínas solúveis e insolúveis chamado ReadyPrep (Bio-Rad), seguindo o protocolo do fabricante. Enfim, todo material genético clonado foi armazenado em ultra freezer -80°C para otimização das etapas futuras de expressão, purificação e caracterização destas novas proteínas.

11. CONCLUSÃO

Nosso trabalho ajudou a confirmar o potencial da abordagem metagenômica para a descoberta de novos genes que codificam para proteínas com prováveis diferenças estruturais e propriedades bioquímicas, presentes na fração não cultivável da microbiota ambiental. O emprego combinado das técnicas de ensaios funcionais, sequenciamento e anotação dos clones com atividade confirmada permitiu identificar prováveis genes alvos (ex. oxigenases) envolvidos no metabolismo de compostos aromáticos.

Os passos iniciais do processo de isolamento, clonagem e expressão dos genes de interesse se mostraram promissores no que concerne a aquisição de proteínas heterólogas. A integração dos resultados parciais deste capítulo aponta o potencial genético da biblioteca metagenômica do sedimento de manguezal impactado com petróleo, demonstrando que este ambiente representa um reservatório de vias metabólicas desconhecidas relacionadas à degradação de hidrocarbonetos do petróleo a ser explorado biotecnologicamente.

12. DISCUSSÃO GERAL

A análise comparativa da diversidade de genes para dioxigenases, obtida através dos dois artigos, permitiu detectar a presença de dioxigenases específicas pelas duas abordagens, tais como fenilpropionato dioxigenase e naftaleno dioxigenase. Estas enzimas desempenham papel fundamental nas vias periféricas de degradação de compostos aromáticos e detoxificação ambiental. A naftaleno dioxigenase participa da etapa inicial de degradação do naftaleno, o qual é convertido em *cis*-Naftaleno dihidrodiol. Estudo realizado por Abo-State *et al.* (2018) mostrou que uma bactéria isolada de águas residuais de uma refinaria de petróleo, identificada por análise do gene RNAr 16S como *Bordetella avium*, apresentou forte atividade de degradação do naftaleno (95%) em meio enriquecido, além da capacidade de utilizá-lo como única fonte de carbono. Bactérias como *Bordetella avium*, com a capacidade de degradar HAs, isoladas a partir de ambientes contaminados, podem ser usadas em processos de biorremediação sem riscos para o ecossistema. Os resultados dos capítulos desta tese também revelaram a presença de sequências homólogas a dioxigenases que hidroxilam anéis aromáticos sem família definida. Em trabalhos com metagenoma, principalmente a partir de ambientes pouco explorados, não é incomum encontrar novos genes para proteínas com assinaturas específicas para domínios funcionais e que não se encaixam em qualquer família enzimática conhecida. Estas descobertas ratificam o valor científico e potencial biotecnológico das comunidades microbianas que ocorrem em ambientes naturais antropizados, como o sedimento de manguezal impactado com petróleo. Em um estudo recente, publicado pelo nosso grupo de pesquisa, foi utilizada a abordagem de sequenciamento *shotgun* para analisar o metagenoma e o metatranscriptoma de manguezais impactados e não impactados pelo homem (Cabral *et al.* 2018). Uma das áreas amostradas foi a mesma usada para os estudos desta tese (área de manguezal impactada com petróleo). Os resultados obtidos por Cabral *et al*. (2018) mostraram o predomínio de genes para diversas enzimas envolvidas na degradação de compostos aromáticos, tais como dioxigenases de Catecol, Protocatecuato e Homogentisato, envolvidas no metabolismo central de HAs, na área de manguezal contaminado com petróleo. Estes resultados corroboram as descobertas descritas nesta tese, demonstrando a capacidade da microbiota do sedimento de manguezal para detoxificação e recuperação ambiental.

As dioxigenases são enzimas que desempenham importante papel funcional em ambientes contaminados com HAs, e dependem de oxigênio molecular para realizar sua atividade catalítica (Jouanneau *et al.* 2011). A ocorrência de genes para dioxigenases em sedimento de manguezal, ambiente predominantemente anaeróbico, é justificada pela ação de

mecanismos naturais de oxigenação tais como a cinética de variação das marés e a difusão de oxigênio pelas raízes de mangue (Alongi 2014). Durante o processo evolutivo, as plantas de mangue fixaram adaptações anatômicas que facilitaram o transporte de oxigênio até as raízes, garantindo seu crescimento e sobrevivência (Pi *et al.* 2009). O trabalho coordenado entre as lenticelas e o aerênquima nas espécies de mangue permite a absorção, transporte e a liberação do oxigênio atmosférico em suas raízes (Nikolausz *et al.* 2008; Srikanth, Lum e Chen 2016). Sendo assim, diante deste pressuposto, a presença de genes para enzimas envolvidas nas vias aeróbicas e anaeróbicas de degradação de compostos aromáticos pode ser explicada pela flutuação diária no suprimento de oxigênio na rizosfera de mangue. Além do mais, a importância deste mecanismo fisiológico pode motivar a realização de futuros estudos para entender como se comporta a comunidade microbiana da rizosfera em função das flutuações da condição redox.

Embora tenha sido usada a mesma amostra ambiental para extrair o DNA metagenômico para a construção das bibliotecas gênica e fosmidial analisadas nos Capítulos I e II, respectivamente, não foi possível acessar genes para benzeno dioxigenase no DNA fosmidial da biblioteca metagenômica, assim como foi observado na biblioteca gênica. Provavelmente, os *primers* usados para o nosso estudo, no primeiro capítulo, favoreceram a amplificação de fragmentos gênicos para dioxigenases que utilizam o benzeno como substrato primário, e que possivelmente está em baixa abundância na microbiota e acabou por não ser representado na biblioteca metagenômica, cuja estratégia não passa por uma etapa de amplificação. Além disso, a maioria das OPFs analisadas no capítulo I mostraram identidade menor do que 60% para dioxigenases conhecidas, depositadas em banco de dados. Estes resultados compactuam com a indefinição enzimática familiar de várias sequências anotadas para dioxigenases que hidroxilam anéis aromáticos presentes na biblioteca metagenômica e mostrado no capítulo II. A baixa percentagem de correspondência das sequências analisadas sugere o importante papel funcional de possíveis novas espécies bacterianas na degradação de hidrocarbonetos aromáticos no ambiente de manguezal.

Os dados da avaliação taxonômica gerados nos Capítulos I e II contribuíram para reafirmar o papel funcional de classes bacterianas como Alphaproteobacteria, Actinobacteria, Betaproteobacteria, Gammaproteobacteria e Deltaproteobacteria na degradação de compostos aromáticos em ambientes contaminados com óleo, em especial, sedimentos de manguezal (Andreote *et al.* 2012; Ghosal *et al.* 2016). Resultados similares foram obtidos no estudo de Cabral *et al*. (2018), o qual revelou os gêneros *Rhodopseudomonas*, *Burkholderia*, *Aromatoleum*, *Pseudomonas* e *Desulfitobacterium* como os principais protagonistas do

metabolismo de compostos aromáticos no sedimento de mangue impactado pelo derramamento de petróleo. Estes mesmos gêneros foram agrupados entre os vinte mais abundantes indicados nos resultados do Capítulo II da presente tese. *Rhodopseudomonas* foi o gênero mais abundante anotado a partir dos genes envolvidos com metabolismo de compostos aromáticos. Estudo anterior já havia revelado sua versatilidade metabólica e capacidade de degradar diferentes compostos aromáticos em ambiente anaeróbico (Larimer *et al.* 2004). Resultados obtidos por outro trabalho revelaram a ocorrência de genes para a α-subunidade de benzoil-CoA redutase a partir de um isolado de *Rhodopseudomonas* (Song e Ward 2005). A benzoil-CoA redutase é uma enzima chave da via anaeróbica periférica do catabolismo de compostos aromáticos, presente em bactérias degradadoras (Carmona *et al*. 2009). O predomínio de sequências para *Rhodopseudomonas* sugere sua forte participação na degradação anaeróbica de HAs presentes nas camadas anóxicas do sedimento contaminado.

Os ensaios de cromatografia demonstraram a eficiência de 3 clones na degradação de fenol (98%), pireno (75%) e naftaleno (71%). Cabral et al. (2018) realizaram uma análise comparativa das concentrações ambientais de vários HAs (disponibilizado por Lima, 2012) presentes em três regiões de manguezal: impactado por petróleo; impactado por efluentes domésticos; e não impactado. De acordo com os resultados, a concentração de todos os HAs analisados foi mais elevada no ambiente impactado pelo derramamento de petróleo, destacando suas consequências a longo prazo (Cabral *et al*. 2018). A ocorrência de clones fosmidiais que mostraram alta eficiência na degradação de pireno e naftaleno na presente tese, pode ser explicada pelo enriquecimento ambiental com estes mesmos HAs. Além disso, foi possível evidenciar a presença de prováveis genes para enzimas envolvidas com a degradação de naftaleno e pireno nos Capítulos I e II. Assim, no geral, a avaliação da biblioteca metagenômica pelas diferentes técnicas adotadas, revelou um potencial genético promissor para o metabolismo de compostos aromáticos e sua aplicação futura no desenvolvimento de novas estratégias para biorremediação de ambientes impactados com o derramamento de petróleo.

Finalmente, as abordagens empregadas nos Capítulos I e II da tese permitiram revelar a diversidade e abundância relativa de genes envolvidos no metabolismo de compostos aromáticos no ambiente de manguezal impactado com petróleo. A abordagem do primeiro capítulo tem a vantagem de investigar mais profundamente a diversidade, riqueza, distribuição e relação genética de genes específicos (no caso, genes de ARHDs) recuperados de diferentes sítios ao redor do mundo. Por outro lado, a abordagem empregada no Capítulo II tem a vantagem de acessar diferentes genes envolvidos em diversas vias metabólicas ao mesmo tempo, permitindo uma compreensão global do potencial funcional da comunidade microbiana

no ambiente analisado. Além disso, o uso da abordagem metagenômica baseada na construção de bibliotecas fosmidiais oferece a vantagem de permitir o isolamento de novos genes e a exploração futura do potencial catalítico de suas respectivas enzimas de interesse.

13. CONCLUSÕES GERAIS

- ✓ O uso de *primers* degenerados desenhados a partir da região conservada do gene α-ARHD permitiu explorar sua diversidade no sedimento de manguezal previamente impactado com petróleo, evidenciando o predomínio de nove OPFs relacionadas à enzima *Benzeno 1, 2-dioxigenase* e uma OPF de ARHD não definida. Além disso, a classificação taxonômica destes genes revelou eles se afiliam predominantemente ao gênero *Pseudomonas*.
- ✓ A comparação baseada no gene ARHD entre os quinze sítios ao redor do mundo contribuiu para elucidarmos a riqueza, distribuição e padrões de distribuição geográfica destes genes, através de similaridade genética compartilhada, bem como os prováveis grupos taxonômicos associados a estes genes funcionais.
- ✓ Os resultados mostraram padrões de endemismo e congruência genética moderada entre os genes ARHD recuperados de sítios ao redor do mundo, presumivelmente modelados por fatores ambientais bióticos e abióticos.
- ✓ A proximidade filogenética entre as sequências de ARHD recuperados de sítios na Antártica e América do Sul sugere eventos de dispersão por correntes marinhas, massas de ar ou organismos hospedeiros.
- ✓ A análise funcional da comunidade microbiana do sedimento de manguezal, a partir da biblioteca metagenômica, evidenciou um balanço equivalente dos genes envolvidos com processos anaeróbicos e aeróbicos relacionados às rotas centrais e periféricas do metabolismo de compostos aromáticos, destacando a prevalência de dioxigenases para catecol, protocatecuato, homogentisato, fenilpropionato, lignostilbene, bifenil e benzoato.
- ✓ De acordo com a classificação taxonômica, a estrutura da comunidade microbiana envolvida com o metabolismo de compostos aromáticos no sedimento de manguezal é formada principalmente por espécies das classes *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* e *Deltaproteobacteria*;

- ✓ Os ensaios de atividade funcional demonstraram o potencial dos clones fosmidiais para a degradação de hidrocarbonetos aromáticos, principalmente na presença de substratos como fenol, pireno e naftaleno.
- ✓ A presença de prováveis genes para dioxigenases nos insertos metagenômicos dos clones sequenciados, aliada à sua homologia com sequencias de grupos microbianos relatados como degradadores de hidrocarbonetos aromáticos, ratificou o potencial genético da biblioteca fosmidial para a prospecção de genes envolvidos no metabolismo de compostos aromáticos.
- ✓ O emprego combinado das técnicas de ensaios funcionais, sequenciamento, anotação, identificação e PCR permitiu clonar e expressar prováveis novos genes para dioxigenases a partir de clones fosmidiais de uma biblioteca metagenômica do sedimento de manguezal impactado com petróleo, demonstrando que este ambiente representa um reservatório de vias metabólicas desconhecidas relacionadas à degradação de hidrocarbonetos do petróleo a ser explorado biotecnologicamente.

14. PERSPECTIVAS FUTURAS

Os genes já clonados em vetor de expressão "PET" serão induzidos para a síntese de suas respectivas proteínas para posterior caracterização molecular e ensaios de atividade enzimática. Para tal, os clones serão inoculados em meio de cultura líquido para indução da expressão heteróloga. As proteínas totais serão extraídas a partir da massa celular centrifugada e da fase líquida do meio de cultura. As proteínas recombinantes, contendo cauda de histidina, serão purificadas em coluna de níquel. Será feito o SDS-PAGE em sistema Omniphor (Laemmli, 1970) para o fracionamento das proteínas com base no peso molecular. A banda diferencial da proteína recombinante será cortada do gel e submetida à digestão tríptica para identificação e caracterização molecular por espectrometria de massas. Para caracterização da atividade enzimática, as proteínas serão purificadas e testadas sob condições variadas de pH, temperatura, detergentes e salinidade. Os produtos de expressão desses clones também serão purificados e analisados quanto aos aspectos estruturais da molécula.

15. REFERÊNCIAS

Abdel-Shafy HI, Mansour MSM. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egypt J Pet* 2015;**25**:107–23.

Abo-State MAM, Riad BY, Bakr AA *et al.* Biodegradation of naphthalene by Bordetella avium isolated from petroleum refinery wastewater in Egypt and its pathway. *J Radiat Res Appl Sci* 2018;**11**:1–9.

Alma'abadi AD, Gojobori T, Mineta K. Marine Metagenome as A Resource for Novel Enzymes. *Genomics Proteomics Bioinformatics* 2015;**13**:290–5.

Alongi DM. Present state and future of the world's mangrove forests. *Environ Conserv* 2002;**29**:331–49.

Alongi DM. Carbon cycling and storage in mangrove forests. *Ann Rev Mar Sci* 2014;**6**:195–219.

Alongi DM, Clough BF, Dixon P *et al.* Nutrient partitioning and storage in arid-zone forests of the mangroves Rhizophora stylosa and Avicennia marina. *Trees* 2003;**17**:51–60.

Alongi DM, Gullaya W, Frank T *et al.* Nutrient capital in different aged forests of the mangrove Rhizophora apiculata. *Bot Mar* 2004;**47**:116.

Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 1995;**59**:143–69.

Andreote FD, Jiménez DJ, Chaves D *et al.* The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One* 2012;**7**:e38600.

Aniszewski E, Peixoto RS, Mota FF *et al.* Bioemulsifier production by Microbacterium sp. strains isolated from mangrove and their application to remove cadmiun and zinc from hazardous industrial residue. *Brazilian J Microbiol* 2010;**41**:235–45.

Arand M, Cronin A, Adamska M *et al.* Epoxide Hydrolases: Structure, Function, Mechanism, and Assay. *Phase II Conjugation Enzymes and Transport Systems*. Vol 400. Academic Press, 2005, 569–88.

Argiriadi MA, Morisseau C, Hammock BD *et al.* Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase. *Proc Natl Acad Sci U S A* 1999;**96**:10637–42.

Barr JG, Fuentes JD, DeLonge MS *et al.* Summertime influences of tidal energy advection on the surface energy balance in a mangrove forest. *Biogeosciences* 2013;**10**:501–11.

Batie CJ, LaHaie E, Ballou DP. Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from Pseudomonas cepacia. *J Biol Chem* 1987;**262**:1510–8.

Beloti LL, Costa BZ, Toledo MAS *et al.* A novel and enantioselective epoxide hydrolase from Aspergillus brasiliensis CCT 1435: Purification and characterization. *Protein Expr Purif* 2013;**91**:175–83.

Bonetti C. Foraminíferos como bioindicadores do gradiente de estresse ecológico em ambientes costeiros poluídos. Estudo aplicado ao sistema estuarino de Santos-São Vicente (SP, Brasil). 2000.

Le Borgne S, Paniagua D, Vazquez-Duhalt R. Biodegradation of Organic Pollutants by Halophilic Bacteria and Archaea. *J Mol Microbiol Biotechnol* 2008;**15**:74–92.

Branton D, Deamer DW, Marziali A *et al.* The potential and challenges of nanopore sequencing. *Nat Biotech* 2008;**26**:1146–53.

Brito EMS, Duran R, Guyoneaud R *et al.* A case study of in situ oil contamination in a mangrove swamp (Rio De Janeiro, Brazil). *Mar Pollut Bull* 2009;**58**:418–23.

Brito EMS, Guyoneaud R, Goni-Urriza M *et al.* Characterization of hydrocarbonoclastic bacterial communities from mangrove sediments in Guanabara Bay, Brazil. *Res Microbiol*

2006;**157**:752–62.

van den Burg B. Extremophiles as a source for novel enzymes. *Curr Opin Microbiol* 2003;**6**:213–8.

Burgess RR. Chapter 17 Refolding Solubilized Inclusion Body Proteins. In: Burgess RR, Deutscher MPBT-M in E (eds.). *Guide to Protein Purification, 2nd Edition*. Vol 463. Academic Press, 2009, 259–82.

Cabral L, Pereira de Sousa ST, Júnior GVL *et al.* Microbial functional responses to long-term anthropogenic impact in mangrove soils. *Ecotoxicol Environ Saf* 2018;**160**, DOI: 10.1016/j.ecoenv.2018.04.050.

Camacho LD, Gevaña DT, Carandang AP *et al.* Tree biomass and carbon stock of a community-managed mangrove forest in Bohol, Philippines. *Forest Sci Technol* 2011;**7**:161–7.

Carmona M, Zamarro MT, Blázquez B *et al.* Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. *Microbiol Mol Biol Rev* 2009;**73**:71–133.

Cerniglia CE. Biodegradation of polycyclic aromatic hydrocarbons. *Biodegradation* 1992;**3**:351–68.

Coimbra MAC. Avaliação dos resultados analíticos de hidrocarbonetos como instrumento jurídico em caso de derrames de petróleo: Estudo de caso. 2006.

DeBruyn JM, Mead TJ, Sayler GS. Horizontal Transfer of PAH Catabolism Genes in Mycobacterium: Evidence from Comparative Genomics and Isolated Pyrene-Degrading Bacteria. *Environ Sci Technol* 2012;**46**:99–106.

Decker M, Arand M, Cronin A. Mammalian epoxide hydrolases in xenobiotic metabolism and signalling. *Arch Toxicol* 2009;**83**:297–318.

Dias ACF, Andreote FD, Dini-Andreote F *et al.* Diversity and biotechnological potential of culturable bacteria from Brazilian mangrove sediment. *World J Microbiol Biotechnol* 2009;**25**:1305–11.

Dias ACF, Andreote FD, Rigonato J *et al.* The bacterial diversity in a Brazilian non-disturbed mangrove sediment. *Antonie Van Leeuwenhoek* 2010;**98**:541–51.

Dias ACF, Dini-Andreote F, Taketani RG *et al.* Archaeal communities in the sediments of three contrasting mangroves. *J Soils Sediments* 2011;**11**:1466–76.

Diaz E. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int Microbiol* 2004;**7**:173–80.

Ding G-C, Heuer H, Zühlke S *et al.* Soil type-dependent responses to phenanthrene as revealed

by determining the diversity and abundance of polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenase genes by using a novel PCR detection system. *Appl Environ Microbiol* 2010;**76**:4765–71.

Donato DC, Kauffman JB, Mackenzie RA *et al.* Whole-island carbon stocks in the tropical Pacific: implications for mangrove conservation and upland restoration. *J Environ Manage* 2012;**97**:89–96.

Duke NC, Meynecke J-O, Dittmann S *et al.* A World Without Mangroves? *Science (80-)* 2007;**317**:41 LP-42.

Erlacher A, Cernava T, Cardinale M *et al.* Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of Lobaria pulmonaria L. *Front Microbiol* 2015;**6**:53.

Ferraro DJ, Gakhar L, Ramaswamy S. Rieske business: structure-function of Rieske non-heme oxygenases. *Biochem Biophys Res Commun* 2005;**338**:175–90.

De Filippo C, Ramazzotti M, Fontana P *et al.* Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Brief Bioinform* 2012;**13**:696–710.

Fuchs G. Anaerobic Metabolism of Aromatic Compounds. *Ann N Y Acad Sci* 2008;**1125**:82–99.

Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds — from one strategy to four. *Nat Rev Microbiol* 2011;**9**:803–16.

Ghosal D, Ghosh S, Dutta TK *et al.* Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *Front Microbiol* 2016;**7**:1369.

Gibson DT, Parales RE. Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr Opin Biotechnol* 2000;**11**:236–43.

Gilman EL, Ellison J, Duke NC *et al.* Threats to mangroves from climate change and adaptation options: A review. *Aquat Bot* 2008;**89**:237–50.

Giovannoni SJ, Britschgi TB, Moyer CL *et al.* Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990;**345**:60–3.

Giri C, Ochieng E, Tieszen LL *et al.* Status and distribution of mangrove forests of the world using earth observation satellite data. *Glob Ecol Biogeogr* 2011;**20**:154–9.

Gkorezis P, Daghio M, Franzetti A *et al.* The Interaction between Plants and Bacteria in the Remediation of Petroleum Hydrocarbons: An Environmental Perspective. *Front Microbiol* 2016;**7**:1836.

Gomes NCM, Borges LR, Paranhos R *et al.* Diversity of ndo genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. *Appl Environ Microbiol* 2007;**73**:7392–9.

Gomes NCM, Flocco CG, Costa R *et al.* Mangrove microniches determine the structural and functional diversity of enriched petroleum hydrocarbon-degrading consortia. *FEMS Microbiol Ecol* 2010;**74**:276–90.

Guo G, Fang T, Wang C *et al.* Isolation and characterization of two novel halotolerant Catechol 2, 3-dioxygenases from a halophilic bacterial consortium. *Sci Rep* 2015;**5**:17603.

Van Hamme JD, Singh A, Ward OP. Recent Advances in Petroleum Microbiology. *Microbiol Mol Biol Rev* 2003;**67**:503–49.

Handelsman J, Rondon MR, Brady SF *et al.* Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 1998;**5**:R245–9.

Hanson CA, Fuhrman JA, Horner-Devine MC *et al.* Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 2012;**10**:1–10.

Haritash AK, Kaushik CP. Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. *J Hazard Mater* 2009;**169**:1–15.

Head IM, Jones DM, Roling WFM. Marine microorganisms make a meal of oil. *Nat Rev Micro* 2006;**4**:173–82.

Heider J, Fuchs G. Anaerobic Metabolism of Aromatic Compounds. *Eur J Biochem* 1997;**243**:577–96.

Heider J, Spormann AM, Beller HR *et al.* Anaerobic bacterial metabolism of hydrocarbons. *FEMS Microbiol Rev* 1998;**22**:459–73.

Holguin G, Vazquez P, Bashan Y. The role of sediment microorganisms in the productivity, conservation, and rehabilitation of mangrove ecosystems: an overview. *Biol Fertil Soils* 2001;**33**:265–78.

Hough DW, Danson MJ. Extremozymes. *Curr Opin Chem Biol* 1999;**3**:39–46.

Hugenholtz P, Goebel BM, Pace NR. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *J Bacteriol* 1998;**180**:4765–74.

Iwai S, Johnson T a, Chai B *et al.* Comparison of the specificities and efficacies of primers for aromatic dioxygenase gene analysis of environmental samples. *Appl Environ Microbiol* 2011;**77**:3551–7.

Jouanneau Y, Martin F, Krivobok S *et al.* Ring-hydroxylating Dioxygenases Involved in PAH

Biodegradation: Structure, Function, and Biodiversity. *Microbial Bioremediation of Non-Metals: Current Research*. 2011, 149–75.

Kathiresan K, Bingham BL. Biology of Mangroves and Magrove Ecosystems. *Adv Mar Biol* 2001;**40**:81–251.

Khara P, Roy M, Chakraborty J *et al.* Functional characterization of diverse ring-hydroxylating oxygenases and induction of complex aromatic catabolic gene clusters in Sphingobium sp. PNB. *FEBS Open Bio* 2014;**4**:290–300.

Kubota K, Koma D, Matsumiya Y *et al.* Phylogenetic analysis of long-chain hydrocarbon-degrading bacteria and evaluation of their hydrocarbon-degradation by the 2,6-DCPIP assay. *Biodegradation* 2008;**19**:749–57.

Kweon O, Kim S-J, Baek S *et al.* A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem* 2008;**9**:11.

Kweon O, Kim S-J, Freeman JP *et al.* Substrate Specificity and Structural Characteristics of the Novel Rieske Nonheme Iron Aromatic Ring-Hydroxylating Oxygenases NidAB and NidA3B3 from Mycobacterium vanbaalenii PYR-1. Lovley DR (ed.). *MBio* 2010;**1**, DOI: 10.1128/mBio.00135-10.

LAEMMLI UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 1970;**227**:680.

Lam KN, Cheng J, Engel K *et al.* Current and future resources for functional metagenomics. *Front Microbiol* 2015;**6**:1196.

Larimer FW, Chain P, Hauser L *et al.* Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nat Biotech* 2004;**22**:55–61.

Levy SE, Myers RM. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* 2016;**17**:95–115.

Lewis M, Pryor R, Wilking L. Fate and effects of anthropogenic chemicals in mangrove ecosystems: A review. *Environ Pollut* 2011;**159**:2328–46.

Li X, Hou L, Liu M *et al.* Abundance and diversity of polycyclic aromatic hydrocarbon degradation bacteria in urban roadside soils in Shanghai. *Appl Microbiol Biotechnol* 2015;**99**:3639–49.

Lima DV. Análise da diversidade, abundância e estrutura funcional da comunidade microbiana de três manguezais do Estado de São Paulo, Brasil. 2012, DOI: 10.11606/T.42.2012.

Lin J, Milase RN. Purification and Characterization of Catechol 1,2-Dioxygenase from

Acinetobacter sp. Y64 Strain and Escherichia coli Transformants. *Protein J* 2015;**34**:421–33.

Lipscomb JD. Mechanism of extradiol aromatic ring-cleaving dioxygenases. *Curr Opin Struct Biol* 2008;**18**:644–9.

Lovley DR. Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Micro* 2003;**1**:35–44.

Lu XY, Zhang T, Fang HHP. Bacteria-mediated PAH degradation in soil and sediment. *Appl Microbiol Biotechnol* 2011;**89**:1357–71.

Lugo AE, Snedaker SC. The Ecology of Mangroves. *Annu Rev Ecol Syst* 1974;**5**:39–64.

Mallick S, Chakraborty J, Dutta TK. Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular-weight polycyclic aromatic hydrocarbons: A review. *Crit Rev Microbiol* 2011;**37**:64–90.

Marcial Gomes NC, Borges LR, Paranhos R *et al.* Exploring the diversity of bacterial communities in sediments of urban mangrove forests. *FEMS Microbiol Ecol* 2008;**66**:96–109.

Margesin R, Labbé D, Schinner F *et al.* Characterization of Hydrocarbon-Degrading Microbial Populations in Contaminated and Pristine Alpine Soils. *Appl Environ Microbiol* 2003;**69**:3085–92.

Margesin R, Schinner F. Biodegradation and bioremediation of hydrocarbons in extreme environments. *Appl Microbiol Biotechnol* 2001;**56**:650–63.

Marshall AG, Rodgers RP. Petroleomics: The Next Grand Challenge for Chemical Analysis. *Acc Chem Res* 2004;**37**:53–9.

Martin F, Malagnoux L, Violet F *et al.* Diversity and catalytic potential of PAH-specific ring-hydroxylating dioxygenases from a hydrocarbon-contaminated soil. *Appl Microbiol Biotechnol* 2013;**97**:5125–35.

Martiny JBH, Bohannan BJM, Brown JH *et al.* Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006;**4**:102–12.

Mcleod E, Chmura GL, Bouillon S *et al.* A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO2. *Front Ecol Environ* 2011;**9**:552–60.

Megharaj M, Ramakrishnan B, Venkateswarlu K *et al.* Bioremediation approaches for organic pollutants: A critical perspective. *Environ Int* 2011;**37**:1362–75.

Meynet P, Head IM, Werner D *et al.* Re-evaluation of dioxygenase gene phylogeny for the development and validation of a quantitative assay for environmental aromatic hydrocarbon degraders. *FEMS Microbiol Ecol* 2015;**91**:1–11.

Nam JW, Nojiri H, Yoshida T *et al.* New classification system for oxygenase components involved in ring-hydroxylating oxygenations. *Biosci Biotechnol Biochem* 2001;**65**:254–63.

Neto FBR. *Estratégias de Sobrevivência de Comunidades Litoraneas Em Regiões Ecologicamente Degradadas: O Caso Da Baixada Santista*. USP-IO: UICN: Fundação Ford, 1989.

Nikolausz M, Kappelmeyer U, Székely A *et al.* Diurnal redox fluctuation and microbial activity in the rhizosphere of wetland plants. *Eur J Soil Biol* 2008;**44**:324–33.

Nolan LC, O'Connor KE. Dioxygenase- and monooxygenase-catalysed synthesis of cis-dihydrodiols, catechols, epoxides and other oxygenated products. *Biotechnol Lett* 2008;**30**:1879–91.

Palmer I, Wingfield PT. Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Curr Protoc protein Sci* 2012;**Chapter 6**:Unit6.3-Unit6.3.

Parales RE, Resnick SM. Aromatic Ring Hydroxylating Dioxygenases. In: Ramos J-L, Levesque RC (eds.). *Pseudomonas: Volume 4 Molecular Biology of Emerging Issues*. Boston, MA: Springer US, 2006, 287–340.

Peixoto R, Chaer GM, Carmo FL *et al.* Bacterial communities reflect the spatial variation in pollutant levels in Brazilian mangrove sediment. *Antonie Van Leeuwenhoek* 2011;**99**:341–54.

Peng R-H, Xiong A-S, Xue Y *et al.* Microbial biodegradation of polyaromatic hydrocarbons. *FEMS Microbiol Rev* 2008;**32**:927–55.

Peng R, Fu X, Tian Y *et al.* Metabolic engineering of Arabidopsis for remediation of different polycyclic aromatic hydrocarbons using a hybrid bacterial dioxygenase complex. *Metab Eng* 2014;**26**:100–10.

Pi N, Tam NFY, Wu Y *et al.* Root anatomy and spatial pattern of radial oxygen loss of eight true mangrove species. *Aquat Bot* 2009;**90**:222–30.

Quince C, Walker AW, Simpson JT *et al.* Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017;**35**:833.

Rappé MS, Giovannoni SJ. The Uncultured Microbial Majority. *Annu Rev Microbiol*

2003;**57**:369–94.

Rengarajan T, Rajendran P, Nandakumar N *et al.* Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. *Asian Pac J Trop Biomed* 2015;**5**:182–9.

Ricklefs RE, Schwarzbach AE, Renner SS. Rate of lineage origin explains the diversity anomaly in the world's mangrove vegetation. *Am Nat* 2006;**168**:805–10.

Santos HF, Carmo FL, Paes JES *et al.* Bioremediation of Mangroves Impacted by Petroleum. *Water, Air, Soil Pollut* 2011;**216**:329–50.

dos Santos HF, Cury JC, do Carmo FL *et al.* Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. *PLoS One* 2011;**6**:e16943.

Schiraldi C, De Rosa M. The production of biocatalysts and biomolecules from extremophiles. *Trends Biotechnol* 2017;**20**:515–21.

Segata N, Boernigen D, Tickle TL *et al.* Computational meta{\textquoteright}omics for microbial community studies. *Mol Syst Biol* 2013;**9**, DOI: 10.1038/msb.2013.22.

Semple KT, Reid BJ, Fermor TR. Impact of composting strategies on the treatment of soils contaminated with organic pollutants. *Environ Pollut* 2001;**112**:269–83.

Seo J-S, Keum Y-S, Li QX. Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health* 2009;**6**:278–309.

Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotech* 2008;**26**:1135–45.

Shimada T, Fujii-Kuriyama Y. Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and1B1. *Cancer Sci* 2003;**95**:1–6.

Siikamäki J, Sanchirico JN, Jardine SL. Global economic potential for reducing carbon dioxide emissions from mangrove loss. *Proc Natl Acad Sci* 2012;**109**:14369–74.

Silva CC, Hayden H, Sawbridge T *et al.* Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. *PLoS One* 2013;**8**:e61811.

Simon C, Daniel R. Metagenomic Analyses: Past and Future Trends. *Appl Environ Microbiol* 2011;**77**:1153–61.

Singh S, Kang SH, Mulchandani A *et al.* Bioremediation: environmental clean-up through pathway engineering. *Curr Opin Biotechnol* 2008;**19**:437–44.

Smits THM, Balada SB, Witholt B *et al.* Functional Analysis of Alkane Hydroxylases from

Gram-Negative and Gram-Positive Bacteria. *J Bacteriol* 2002;**184**:1733–42.

Soclo HH, Garrigues P, Ewald M. Origin of Polycyclic Aromatic Hydrocarbons (PAHs) in Coastal Marine Sediments: Case Studies in Cotonou (Benin) and Aquitaine (France) Areas. *Mar Pollut Bull* 2000;**40**:387–96.

Song B, Ward BB. Genetic Diversity of Benzoyl Coenzyme A Reductase Genes Detected in Denitrifying Isolates and Estuarine Sediment Communities. *Appl Environ Microbiol* 2005;**71**:2036–45.

Spalding M, Kainuma M, Collins L. *World Atlas of Mangroves*. Earthscan, 2010.

Srikanth S, Lum SKY, Chen Z. Mangrove root: adaptations and ecological importance. *Trees* 2016;**30**:451–65.

Sugiura K, Ishihara M, Shimauchi T *et al.* Physicochemical Properties and Biodegradability of Crude Oil. *Environ Sci Technol* 1997;**31**:45–51.

Taketani RG, Franco NO, Rosado AS *et al.* Microbial community response to a simulated hydrocarbon spill in mangrove sediments. *J Microbiol* 2010a;**48**:7–15.

Taketani RG, Yoshiura CA, Dias ACF *et al.* Diversity and identification of methanogenic archaea and sulphate-reducing bacteria in sediments from a pristine tropical mangrove. *Antonie Van Leeuwenhoek* 2010b;**97**:401–11.

Tam NF, Ke L, Wang XH *et al.* Contamination of polycyclic aromatic hydrocarbons in surface sediments of mangrove swamps. *Environ Pollut* 2001;**114**:255–63.

Tam NFY, Wong TWY, Wong YS. A case study on fuel oil contamination in a mangrove swamp in Hong Kong. *Mar Pollut Bull* 2005;**51**:1092–100.

Tan WA, Parales RE. Application of Aromatic Hydrocarbon Dioxygenases. *Green Biocatalysis*. Wiley-Blackwell, 2016, 457–71.

Teeling H, Glöckner FO. Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective. *Brief Bioinform* 2012;**13**:728–42.

Thompson CE, Beys-da-Silva WO, Santi L *et al.* A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. *AMB Express* 2013;**3**:65.

Uchiyama T, Miyazaki K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 2009;**20**:616–22.

Vaillancourt FH, Bolin JT, Eltis LD. The Ins and Outs of Ring-Cleaving Dioxygenases. *Crit Rev Biochem Mol Biol* 2006;**41**:241–67.

Vazquez P, Holguin G, Puente ME *et al.* Phosphate-solubilizing microorganisms associated with the rhizosphere of mangroves in a semiarid coastal lagoon. *Biol Fertil Soils* 2000;**30**:460–8.

Vo-Luong P, Massel S. Energy dissipation in non-uniform mangrove forests of arbitrary depth. *J Mar Syst* 2008;**74**:603–22.

Wang C, Guo G, Huang Y *et al.* Salt Adaptation and Evolutionary Implication of a Nah-related PAHs Dioxygenase cloned from a Halophilic Phenanthrene Degrading Consortium. *Sci Rep* 2017;**7**:12525.

Wang H, Xu J, Zhao W *et al.* Effects and Risk Evaluation of Oil Spillage in the Sea Areas of Changxing Island. *Int J Environ Res Public Health* 2014;**11**:8491–507.

Werlen C, Kohler H-PE, van der Meer JR. The Broad Substrate Chlorobenzene Dioxygenase and cis-Chlorobenzene Dihydrodiol Dehydrogenase of Pseudomonas sp. Strain P51 Are Linked Evolutionarily to the Enzymes for Benzene and Toluene Degradation . *J Biol Chem* 1996;**271**:4009–16.

Widdel F. New types of acetate-oxidizing, sulfate-reducing Desulfobacter species, D. hydrogenophilus sp. nov., D. latus sp. nov., and D. curvatus sp. nov. *Arch Microbiol* 1987;**148**:286–91.

Wu P, Wang YS, Sun FL *et al.* Bacterial polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases in the sediments from the Pearl River estuary, China. *Appl Microbiol Biotechnol* 2014;**98**:875–84.

Xu M, Xiao X, Wang F. Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles* 2008;**12**:255–62.

Yang Z, Zhang L, Zhang Y *et al.* Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PLoS One* 2011;**6**:e22981–e22981.

Yeates C, Holmes AJ, Gillings MR. Novel forms of ring-hydroxylating dioxygenases are widespread in pristine and contaminated soils. *Environ Microbiol* 2000;**2**:644–53.

Yender R, Michel JM, Lord C *et al. Managing Seafood Safety After an Oil Spill*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Office of Response and Restoration, 2002.

Zeng J, Zhu Q, Wu Y *et al.* Characterization of a polycyclic aromatic ring-hydroxylation dioxygenase from Mycobacterium sp. NJS-P. *Chemosphere* 2017;**185**:67–74.

16. APÊNDICES E ANEXOS

16.1. Formulário de aprovação do CNBio

UNICAMP

MANUAL DA QUALIDADE

CENTRO PLURIDISCIPLINAR DE PESQUISAS QUÍMICAS, BIOLÓGICAS E AGRÍCOLAS

Título: FCIbio 01 – Formulário de proposta de projeto envolvendo OGM	Emissão: CIBio	Edição/Revisão: 2/0	nº p.

Esse formulário só tem validade para julgamento de OGM Tipo I. Caso o projeto envolva OGM tipo II, o projeto deverá ser encaminhado para CTNBio via CIBio

(a ser preenchido pela CIBio sob o número CPQBA/CQB 0189/03)
Número do projeto: 411/13
Data de entrada: 12/6/13
Analisado por M.G.S. AUDRICETTA em 17/7/13
"Check list" no laboratório feito em: 17/7/13
Campinas, Situação: ☒ Aprovado ☐ Reprovado Assinatura do Presidente da CIBio

Projeto de Pesquisa

Título: **Investigação de genes envolvidos na biodegradação de hidrocarbonetos aromáticos a partir do metagenoma de mangue impactado com petróleo**	
Data de início do projeto (mês/ano): 07/2013	Data prevista de conclusão (mês/ano): 02/2017

Pesquisador Principal

Nome: Valéria Maia Merzel		
Endereço: DRM-CPQBA/UNICAMP, Rua Alexandre Caselatto, 999. C.P. 6171, CEP: 13081-970		
e-mail: vmaia@cpqba.unicamp.br	Telefone: (19) 2139-2874	Fax: (19) 2139-2852

Objetivo do Projeto

O objetivo do presente projeto é a Identificação e caracterização genética de dioxigenases que degradam hidrocarbonetos aromáticos a partir de clones fosmidiais de uma biblioteca metagenômica de sedimento de manguezal altamente impactado por petróleo. Para tal será feito

uma manipulação de OGMs para a triagem dos clones da biblioteca metagenômica para atividade de degradação de hidrocarbonetos aromáticos através de ensaios funcionais de alto desempenho (HTS- High Throughput Screening). Subsequentemente será feito o sequenciamento dos clones positivos para caracterização estrutural de operons para via de degradação de hidrocarbonetos aromáticos.

Laboratório onde as pesquisas com o OGM serão desenvolvidas

- Lab. de Biologia Molecular, Lab. Microbiologia de Pesquisa, Sala de Equipamentos, Sala de Lavagem e Preparo na Divisão de Recursos Microbianos do CPQBA.

Equipe (Nome e função)

- Dra. Valéria Maia Merzel, orientadora do projeto de doutorado.
- Sanderson Tarciso Pereira de Sousa, aluno de doutorado que desenvolverá o projeto.

Equipamentos utilizados durante o trabalho com contenção do OGMs

- Sequenciador automático ABI 3500 xl (Applied Biosystems)
- Microcentrifuga de placas Eppendorf
- Microcentrifuga para tubos de 1,5 mL Eppendorf
- Freezer -80oC, freezer -20oC
- Geladeiras
- Banho-maria
- Autoclave (Divisão de Microbiologia)
- Shaker de bancada
- Fluxos laminares
- Incubadoras BODs
- Micropipetadores automáticos (diferentes volumes)
- Termociclador automático (PCR)
- Sistemas de eletroforese de DNA

- Fotodocumentador
- Eletroporador

Procedimentos de limpeza, desinfecção, descontaminação e descarte de material/resíduos

- Antes e após a manipulação de OGMs, o fluxo laminar é descontaminado com álcool 70% e luz UV por 15 min.
- Todos os materiais (meios, placas, tubos, ponterias) são descartados após descontaminação em autoclave (ciclo de descontaminação padrão). Soluções contaminadas com DNA são descartadas em frasco com solução de hipoclorito 0,5% e posteriormente os frascos com solução de descarte são autoclavados.
- Material utilizado na manipulação de OGM´s é segregado para descarte em um saco de lixo identificado.

As bancadas são descontaminadas diariamente após término dos experimentos com álcool 70% e hipoclorito 0,5%. Piso do laboratório é limpo com solução de hipoclorito 0,5%.

Termo de Responsabilidade

Título do Projeto: Investigação de genes envolvidos na biodegradação de hidrocarbonetos aromáticos a partir do metagenoma de mangue impactado com petróleo

Eu, ***Valéria Maia Merzel****, Pesquisador(a) responsável pelo projeto, asseguro à CIBio/CPQBA-UNICAMP que:*

- Li as Instruções Normativas da CTNBio, pertinentes para trabalhar com os OGMs acima referidos, que se encontram no site http://www.ctnbio.gov.br/index.php/content/view/11992.html e que concordo com as suas exigências durante a vigência deste projeto.
- A equipe que participa deste projeto também está ciente das referidas Instruções Normativas e é competente para executá-las.
- Comprometo-me a solicitar nova aprovação à CIBio local sempre que ocorra alteração significativa nos objetivos/procedimentos/instalações aqui descritos e a lhe fornecer um relatório anual de andamento do projeto.
- Todas as ações referente a manipulação e descarte do material envolvido no referido projeto será registrado e anexado ao relatório de atividades.
- Tudo que foi declarado é a absoluta expressão da verdade. Estou ciente de que o eventual não cumprimento das Instruções Normativas da CTNBio é de minha total responsabilidade e que estarei sujeito às punições previstas na legislação em vigor.

Data: 12 / 06 / 2013

Assinatura do Pesquisador Principal: 

16.3. Declaração de direitos autorais

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Investigação de genes envolvidos na biodegradação de hidrocarbonetos aromáticos a partir do metagenoma de manguezal impactado com petróleo**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 09/11/2018

Assinatura : Sanderson Tarciso Pereira de Sousa
Nome do(a) autor(a): **Sanderson Tarciso Pereira de Sousa**
RG n.° 0947172149

Assinatura : _____
Nome do(a) orientador(a): **Valéria Maia Merzel**
RG n.° 12.573.419