

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**Prospecção *in silico* de genes envolvidos na degradação de
Hidrocarbonetos Policíclicos Aromáticos**

André Luís Rodrigues de Aquino

Orientador: Prof. Dr. Daniel Guariz Pinheiro

Co-orientadora: Michelli Inácio Gonçalves Funicelli

Trabalho apresentado à Faculdade
de Ciências Agrárias e Veterinárias
- UNESP, Câmpus de Jaboticabal,
para obtenção do título de
Bacharel em Ciências Biológicas.

**Jaboticabal, SP
1º Semestre/2021**

A657p

Aquino, André Luís Rodrigues

Prospecção in silico de genes envolvidos na degradação de Hidrocarbonetos Policíclicos Aromáticos / André Luís Rodrigues Aquino. -- Jaboticabal, 2021

81 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Ciências Biológicas) Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Daniel Guariz Pinheiro

Coorientadora: Michelli Inácio Gonçalves Funicelli

1. Hidrocarbonetos Policíclicos Aromáticos. 2. Biorremediação. 3. Metagenômica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL



DEPARTAMENTO:

TECNOLOGIA

CERTIFICADO DE APROVAÇÃO

TRABALHO DE CONCLUSÃO DE CURSO

TÍTULO: PROSPECÇÃO *IN SILICO* DE GENES ENVOLVIDOS NA DEGRADAÇÃO DE
HIDROCARBONETOS POLICÍCLICOS AROMÁTICOS

ACADÊMICO: ANDRÉ LUIS RODRIGUES DE AQUINO

CURSO: CIÊNCIAS BIOLÓGICAS

ORIENTADOR (ES): Prof. Dr. DANIEL GUARIZ PINHEIRO
Me. MICHELLI INÁCIO GONÇALVES FUNNICELLI

Aprovado e corrigido de acordo com as sugestões da Banca Examinadora

BANCA EXAMINADORA:

| | (Nomes) | (Assinaturas) |
|------------|--|---------------|
| Presidente | Prof. Dr. Daniel Guariz Pinheiro | |
| Membro | Prof. Dr. Alessandro de Mello Varani | |
| Membro | Prof. ^a Dr. ^a Lucia Maria Carareto Alves | |

Jaboticabal 03 / 08 / 2021

Aprovado *ad referendum* do Conselho do Departamento 05 / 08 / 2021
em:

Chefe do Departamento

AGRADECIMENTOS

Agradeço ao Prof. Dr. Daniel Guariz Pinheiro, pelos conselhos, ensinamentos e paciência. É uma pessoa com grandes ideais moralistas, empenho no trabalho e senso de dever que são admiráveis.

Aos meus colegas do Laboratório de Bioinformática que sempre foram muito cordiais e atenciosos para comigo, além de terem me auxiliado em várias ocasiões ao longo do desenvolvimento desse projeto: Michelli Inácio Gonçalves Funicelli, Luis Guillermo Teherán Sierra, Rafael Correia da Silva, Maria Fernanda Campanari, Lucas Amoroso Lopes de Carvalho e Ana Paula Corrêa Moneda, que também de certa forma foram responsáveis pela minha construção pessoal ao longo da minha experiência no estágio e iniciação científica.

À minha família que sempre me deu muito carinho, incentivo e suporte financeiro ao longo do meu curso de graduação, em especial minha mãe Maria Cristina Rodrigues e meu pai Cyro Lindolfo Ninin de Aquino e aos demais familiares que convivem comigo em Sertãozinho.

Aos meus amigos e colegas do Curso de Ciências Biológicas que estiveram comigo ao longo desses 5 anos e compartilharam de vários momentos de alegria, ansiedade, empolgação e dedicação ao longo da jornada na graduação.

A Pró-reitoria de Pesquisa (PROPe) da Universidade Estadual Paulista “Júlio e Mesquita Filho” (UNESP) e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pela bolsa concedida durante o desenvolvimento deste trabalho (edital do PIBIC 2018/2019).

SUMÁRIO

| | |
|---|----|
| RESUMO..... | v |
| ABSTRACT..... | vi |
| 1. INTRODUÇÃO..... | 7 |
| 2. OBJETIVOS..... | 9 |
| 3. REVISÃO DE LITERATURA..... | 10 |
| 3.1 HIDROCARBONETOS POLICÍCLICOS AROMÁTICOS..... | 10 |
| 3.2 MICRO-ORGANISMOS DEGRADADORES DE HPAs..... | 13 |
| 3.3 FAMÍLIAS DE PROTEÍNAS RELACIONADAS À dHPAs..... | 15 |
| 3.4 METAGENÔMICA..... | 19 |
| 3.4.1 ESTUDOS METAGENÔMICOS..... | 19 |
| 3.4.2 ANÁLISE DE DADOS DE SEQUENCIAMENTO <i>SHOTGUN</i> DE METAGENOMAS..... | 22 |
| 3.4.3 ANÁLISE DE DIVERSIDADE EM METAGENOMAS..... | 25 |
| 3.4.4 RESUMO..... | 25 |
| 3.5 BUSCAS POR BIOMOLÉCULAS DE INTERESSE..... | 29 |
| 3.6 ABORDAGEM METAGENÔMICA APLICADA À ANÁLISE DE XENOBIÓTICOS..... | 31 |
| 4. MATERIAL E MÉTODOS..... | 33 |
| 4.1 IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS PROTEÍNAS dHPAs..... | 33 |
| 4.2 ANÁLISES DE METAGENOMA ASSOCIADO À dHPAs..... | 35 |
| 4.2.1 OBTENÇÕES DOS DADOS METAGENÔMICOS..... | 35 |
| 4.2.2 PRÉ-PROCESSAMENTO DOS DADOS..... | 36 |
| 4.2.3 MONTAGEM <i>DE NOVO</i> DO METAGENOMA E PREDIÇÃO GÊNICA..... | 36 |

| | |
|--|----|
| 4.2.4 PROSPECÇÃO DAS PROTEÍNAS dHPAS..... | 37 |
| 4.2.5 IMPLEMENTAÇÃO DA AUTOMATIZAÇÃO DOS PROCESSOS DE ANÁLISE..... | 39 |
| 5. RESULTADOS E DISCUSSÃO..... | 41 |
| 5.1 PROTEÍNAS dHPAS..... | 41 |
| 5.1.1 IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS PROTEÍNAS dHPAS..... | 41 |
| 5.1.2 ÁRVORES DE SIMILARIDADE..... | 43 |
| 5.1.3 ANÁLISE TAXONÔMICA DAS PROTEÍNAS DO UNIREF..... | 43 |
| 5.2 METAGENOMA DE ROCK BAY..... | 47 |
| 5.2.1 PRÉ-PROCESSAMENTO..... | 47 |
| 5.2.2 ÍNDICES DE MONTAGEM E PREDIÇÃO GÊNICA..... | 48 |
| 5.2.3 ANÁLISE TAXONÔMICA DO METAGENOMA..... | 52 |
| 5.3.1 PROSPECÇÃO DE NOVAS PROTEÍNAS dHPAS..... | 58 |
| 5.3.2 CORRESPONDÊNCIA TAXONÔMICA EM RELAÇÃO ÀS PROTEÍNAS dHPAS | 61 |
| 6. CONCLUSÃO..... | 65 |
| 7. REFERÊNCIAS BIBLIOGRÁFICAS..... | 66 |

RESUMO

Os Hidrocarbonetos Policíclicos Aromáticos (HPAs) representam uma classe de moléculas consideradas poluentes para o meio ambiente, dessa forma é de grande interesse a busca e o aperfeiçoamento de técnicas que visem a sua degradação, com destaque para a técnica de biorremediação. Neste trabalho realizamos um processo de identificação de novas proteínas degradadoras de HPAs (dHPAs) a partir de 1.298 proteínas distribuídas em 12 famílias de proteínas previamente descritas. O processo envolveu a busca de sequências homólogas e construção de modelos estatísticos para a classificação em subfamílias de dHPAs, o qual resultou em 10.489 dHPAs. A partir daí realizamos uma prospecção em dados metagenômicos obtidos de sequenciamento de DNA extraído a partir de amostras enriquecidas provenientes de sedimentos em Rock Bay (Victoria Harbour, British Columbia, Canadá), onde há contaminação por derivados de petróleo entre eles HPAs, devido principalmente a atividades industriais na região. Foram utilizadas duas estratégias para a prospecção de dHPAs a partir da predição gênica nos metagenomas montados: uma considerando buscas por similaridade de sequências no conjunto completo de sequências identificadas como dHPAs, a outra, é a partir da correspondência com os modelos estatísticos para as subfamílias. Ambas as estratégias permitiram a categorização em família/subfamília dessas proteínas preditas nos metagenomas, por fim, foram identificados os micro-organismos mais abundantes no metagenoma contendo as proteínas dHPAs. Os resultados obtidos na predição gênica nesse metagenoma foram de 1.053.057 ORFs (*Open Reading Frames*), sendo que 2.954 correspondem a dHPAs. A partir daí detectamos que a maioria das bactérias com essas ORFs são pertencentes ao gênero *Pseudomonas*. Esse resultado possui grande suporte de outros trabalhos onde foram feitas identificações taxonômicas de micro-organismos em ambientes contaminados por xenobióticos, dentre eles HPAs. Dessa forma, indicamos que a ferramenta é útil para a prospecção de proteínas dHPAs. Além disso, não há limitação para ser empregada em buscas para outras famílias gênicas.

Palavras-chave: Hidrocarbonetos policíclicos aromáticos, biorremediação, metagenômica

ABSTRACT

Polycyclic Aromatic Hydrocarbons (PAHs) represent a class of molecules considered pollutants for the environment, therefore the search and improvement of techniques aimed at their degradation is of great interest, focusing on bioremediation technology. In this work, we performed a process of identification of new PAH degrading proteins (dHPAs) from 1,298 proteins distributed in 12 families of previously described enzymes. The process involved searching for homologous sequences and constructing statistical models to classify dHPAs into subfamilies, which resulted in 10,489 dHPAs distributed in 606 subfamilies. From there, we conducted prospecting of metagenomic data obtained from DNA sequencing of enriched samples from sediments in Rock Bay (Victoria Harbour, British Columbia, Canada), where there is contamination by petroleum derivatives, including PAHs, mainly due to industrial activities in the region. Two strategies were used to search for dHPAs based on gene predictions in assembled metagenomes: one considers the search for sequence similarity in the full set of sequences identified as dHPAs, and the other is based on the match with the statistical models for the previously identified subfamilies. Both strategies allowed the categorization in family/subfamily of these proteins predicted in the metagenomes, finally the most abundant micro-organisms in the metagenome containing the proteins dHPAs were identified. The gene prediction results in this metagenome were 1,053,057 ORFs (Open Reading Frames), of which 2,954 correspond to dHPAs. From this, we could see that most bacteria with these ORFs belong to the genus *Pseudomonas*. This result is highly supported by other work in which taxonomic identifications of micro-organisms in environments contaminated by xenobiotics, including HPAs, have been made. Thus, we suggest that the tool is useful for the detection of dHPAs. Moreover, there is no limitation to its use in searches for other gene families.

Key words: Polycyclic aromatic hydrocarbons, bioremediation, metagenome

1. INTRODUÇÃO

Os xenobióticos representam um abrangente grupo de compostos químicos que são estranhos para os ambientes naturais ou os organismos, não fazendo parte de sua rede metabólica usual, podendo ser produzidos por eventos naturais ou por ação antropogênica. Eles são enquadrados dentro de muitas categorias, entre elas: pesticidas agrícolas, inseticidas, plásticos, produtos de limpeza, fármacos e poluentes ambientais, sendo que o efeito e biotransformação desses diferentes xenobióticos em um sistema biológico dependem da constituição molecular do composto e a composição enzimática presente no organismo que foi exposto a essas moléculas, assim este organismo terá sua cadeia metabólica agindo com maior ou menor eficiência na degradação do xenobiótico dependendo das proteínas e cofatores que vão interagir com ele no sistema biológico afetado (GODHEJA, J. et al., 2016).

Entre as diferentes classes de xenobióticos se tem um importante grupo que é considerado um grande poluente ambiental, trata-se dos Hidrocarbonetos Policíclicos Aromáticos (HPAs) que são uma classe de moléculas constituída por um arranjo de átomos de carbono e hidrogênio estando fundidos em dois ou mais anéis benzênicos. Esse grupo de moléculas é altamente mutagênico, carcinogênico e instável, sendo um conhecido contaminante de ecossistemas e de organismos que quando entram em contato com esses compostos sofrem uma bioacumulação dessas moléculas em seus órgãos e tecidos, assim não só o sistema fisiológico fica comprometido como também a integridade das reações metabólicas normais que esses organismos realizam. Normalmente a origem dos HPAs está intimamente relacionada com a queima incompleta de matéria orgânica que pode ocorrer de maneira natural ou por conta de fatores antropogênicos como, por exemplo, queima de carvão mineral, derramamento de petróleo e derivados em áreas ambientais, entre outros (ABDEL-SHAFY; MANSOUR, 2016; DUAN et al., 2015; SUMAN; SINHA; TARAFDAR, 2016).

Por conta desses fatores, existe um enorme interesse na busca e aperfeiçoamento de técnicas que visem à eliminação completa ou parcial de tais substâncias nocivas aos diferentes ecossistemas e aos organismos vivos, entre essas técnicas está a biorremediação que consiste na utilização da maquinaria

enzimática de micro-organismos específicos, sejam eles bactérias, fungos ou protozoários, que são adaptados para a incorporação e degradação de moléculas tóxicas como no caso dos HPAs. Considerando-se que existe uma enorme gama de diferentes proteínas que estão agrupadas em famílias (ALKB, BCOAR, BSSA, CYP153, EXDO, FERRE, INDO, MACR, MCIS, RHDO, RHMO e TDLH), que exercem um efeito eficiente na degradação dessas moléculas dependendo da sua composição (diferentes radicais e funções incorporadas nesses diferentes xenobióticos) e que boa parte dessas sequências proteicas ainda não foi explorada, se abre uma grande perspectiva para o tratamento e recuperação de locais contaminados por esses compostos tóxicos nos diferentes ambientes sejam eles: solos, rios, planícies ou outros habitats (DZIOŃEK; WOJCIESZYŃSKA; GUZIK, 2016; JACQUES et al., 2007; VILCHEZ-VARGAS et al., 2013).

Levando isso em conta, o processo de biorremediação é favorecido por meio de técnicas que envolvem a microbiologia clássica, ou seja, o cultivo de micro-organismos *in vitro* que tenham uma ação específica na degradação de xenobióticos específicos. No entanto, a microbiologia clássica é capaz de representar aproximadamente um pouco mais de 1% da verdadeira diversidade de bactérias e fungos encontrados no ambiente que é capaz de promover uma ação direcionada contra esses compostos nocivos (entre eles os HPAs), dessa forma a pesquisa fica limitada e restrita aos micro-organismos que podem ser cultivados (HANDELSMAN, 2004). Por conta disso, atualmente são realizados estudos metagenômicos, que utilizam dados de sequenciamento de DNA em larga-escala e constitui uma abordagem independente de cultivo para explorar a capacidade gênica dos micro-organismos que habitam determinado nicho ecológico. Dessa forma, por meio de análises que *in silico*, ou seja, extraíndo informação a partir do processamento dos dados sequenciados, há a possibilidade de explorar as diferentes enzimas que podem interagir com xenobióticos tóxicos em um ambiente contaminado. Este processamento de dados é realizado utilizando diferentes ferramentas de Bioinformática para realizar possíveis previsões de genes envolvidos na degradação dessas moléculas e assim aumentar o arcabouço de informações que podem contribuir para a biorremediação de áreas contaminadas por HPAs (BAMFORTH et al., 2005; BERINI et al., 2017; ROUMPEKA et al., 2017).

2. OBJETIVOS

A fim de contribuir com a anotação funcional dos genes responsáveis pela codificação de proteínas relacionadas com a dHPAs propomos a construção de uma base de dados baseada em Modelos Ocultos de Markov (*Hidden Markov Models*, ou HMMs) a partir de alinhamentos múltiplos de proteínas previamente anotadas como membros da via de degradação desses compostos. Utilizando os modelos desenvolvidos, pretendemos prospectar essas proteínas a partir de dados metagenômicos de sedimento contaminado com HPAs em uma área industrial na Baía de Rock Bay localizada no Canadá (MEIER; PATERSON; LAMBERT, 2016), explorando dessa forma o potencial genético para a degradação pelos micro-organismos identificados nesse ambiente.

Para a construção dessa base de dados, buscamos identificar proteínas homólogas a aquelas já reconhecidas como membros de famílias de dHPAs, a partir de uma base de dados abrangente de proteínas de procariotos, o UniRef (SUZEK et al., 2007). A partir dessa predição de homólogos serão construídos os HMMs para as subfamílias relacionadas a cada uma das famílias de dHPAs. Após a obtenção dos modelos, para demonstrar a aplicabilidade, realizaremos análises de metagenomas em ambientes que sofreram alterações antrópicas e foram contaminados com esse tipo de composto químico.

Essas análises incluíram Montagem de *contigs* genômicos a partir das sequências disponíveis no *Sequence Read Archive* (SRA); Anotação taxonômica dos *contigs* e predição gênica; Realização de buscas partindo tanto das sequências identificadas e classificadas quanto dos HMMs para prospecção de novas proteínas a partir da predição gênica no metagenoma. Além disso, pretendemos avaliar a abundância dessas proteínas nos ambientes contaminados. Por fim, pretendemos disponibilizar as sequências proteicas classificadas em famílias e subfamílias e os HMMs obtidos para serem compartilhados na comunidade científica.

A motivação deste trabalho foi a identificação de novas proteínas e micro-organismos, os quais poderão ser avaliados em trabalhos futuros para possivelmente serem utilizados em estratégias de biorremediação.

3. REVISÃO DE LITERATURA

3.1 HIDROCARBONETOS POLICÍCLICOS AROMÁTICOS

Os Hidrocarbonetos Policíclicos Aromáticos (HPAs) são compostos orgânicos, formados por dois ou mais anéis benzênicos fundidos em sua estrutura, estando arranjados de diversas formas e podendo apresentar diferentes tipos de radicais e funções orgânicas, bem como também, podem possuir heteroátomos integrados em sua estrutura, tais como flúor (F), nitrogênio (N), enxofre (S) e oxigênio (O) (ABDEL-SHAFY; MANSOUR, 2016; DUAN et al., 2015).

A origem dos HPAs está relacionada com a combustão incompleta da matéria orgânica que pode ocorrer de maneira natural como, por exemplo, vulcanismo e queimadas espontâneas em florestas. No entanto a maior parte desses compostos é produzida principalmente por fontes antropogênicas, tais como, combustão incompleta de biomassas, acidentes de contaminação por meio de combustíveis fósseis, incineração de resíduos domésticos e industriais e emissão de gases provenientes dos motores de veículos (PATEL et al., 2020).

Sendo assim o resultado da contaminação de ecossistemas por esses compostos está amplamente relacionado às ações humanas, com destaque ao derramamento de petróleo, diesel e de processos industriais como liquefação e gaseificação do carvão durante a produção de coque. Esses fatores vêm contribuindo de maneira constante e significativa para o aumento da produção de HPAs nos ecossistemas desde o início do século XX (GEIER et al., 2018).

Outro fator preocupante é que esses contaminantes possuem um grau considerável de estabilidade em sua conformação estrutural, que é diretamente proporcional ao número de anéis aromáticos inseridos em sua molécula, sendo assim quanto maior esse valor, maior será a sua permanência no meio ambiente e dificuldade para a sua eliminação (XUE; WARSHAWSKY, 2005).

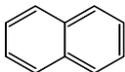
Os HPAs, assim como os pesticidas organoclorados também são agrupados na categoria de POPs (Poluentes Orgânicos Persistentes), representando as principais classes de moléculas prejudiciais ao meio ambiente dentro desta categoria, essas moléculas compartilham grupos funcionais e semelhanças estruturais entre si, entre elas os anéis aromáticos (YUAN et al., 2017).

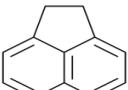
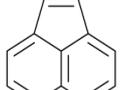
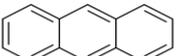
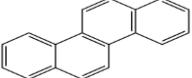
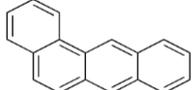
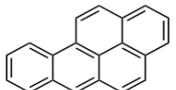
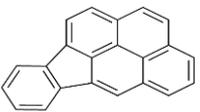
Entre os principais pesticidas organoclorados relacionados quimicamente com os HPAs estão o dicloro-difenil-tricloroetano (DDT), pentaclorobenzeno (PeCB) e hexaclorocicloexano (HCH). O DDT é um dos pesticidas mais estudados, tendo sido utilizado a princípio em larga escala, no entanto caiu em desuso devido as suas propriedades carcinogênicas, tendo o seu uso limitado em vários países, ainda assim este composto é utilizado atualmente na indústria de solventes (KIM; KABIR; JAHAN, 2017; NICOLOPOULOU-STAMATI et al., 2016).

Além disso, algumas moléculas de HPAs são consideradas como poluentes prioritários, de acordo com a Agência de Proteção Ambiental dos Estados Unidos (United States Environmental Protection Agency – U.S. EPA). Dentre as mais de 100 diferentes moléculas HPAs foram classificadas 16 potencialmente carcinogênicas e com características mutagênicas (SUN; LITTLEJOHN; DAVID GIBSON, 1998).

No caso da contaminação ambiental os principais HPAs encontrados são: Naftaleno, Antraceno, Criseno, Coroneno, Fluoranteno, Acenaftileno, Fenantreno, Fluoreno, Pireno e Benzo(alfa)pireno, dentre estes, o Benzo(alfa)pireno é o mais conhecido e descrito na literatura, caracterizado por seu alto grau de persistência e difícil remediação (KUPPUSAMY et al., 2017). As principais informações sobre a categorização desses compostos e seu risco na saúde, podem ser visualizadas na Tabela 1 adaptada de (PAZ et al., 2017).

Tabela 1: Representação estrutural dos principais compostos da classe de HPAs que são contaminantes do solo tidos como prioritários de acordo com a Agência de Proteção Ambiental Americana e sua classificação segundo a Agência Internacional para Pesquisa do Câncer.

| Composto | Estrutura | Classificação IARC | Nº de Anéis |
|-----------|---|-----------------------------|-------------|
| Naftaleno |  | Possivelmente carcinogênico | 2 |

| | | | |
|--------------------------|---|-----------------------------|---|
| Acenafteno |  | Não avaliado | 3 |
| Acenaftileno |  | Não avaliado | 3 |
| Antraceno |  | Não classificado | 3 |
| Criseno |  | Possivelmente carcinogênico | 4 |
| Benzo (a) antraceno |  | Possivelmente carcinogênico | 4 |
| Benzo (a) pireno |  | Carcinogênico | 5 |
| Indeno (1,2,3-cd) pireno |  | Possivelmente carcinogênico | 6 |

Fonte: (PAZ et al., 2017)

Para além dos contaminantes classificados na **Tabela 1**, um dos grupos mais tóxicos e carcinogênicos derivados dessa classe de moléculas são os policíclicos aromáticos nitrados (Nitro-HPAs ou NHPAs), que são originados a partir de reações químicas entre os HPAs e óxidos de nitrogênio (Nox) encontrados principalmente em uma fração dos gases presentes na atmosfera por conta da poluição industrial ou a partir de interações entre os HPAs previamente provenientes do ambiente por conta de uma contaminação anterior e que podem ficar expostos a uma nova contaminação naquele local por produtos nitrogenados derivados de processos industriais, como por exemplo alguns fertilizantes e pesticidas (BANDOWE; MEUSEL, 2017).

Conseqüentemente pode ocorrer a formação de uma enorme gama de moléculas com um enorme potencial mutagênico e bioacumulador tanto para o ecossistema, quanto em relação aos seres vivos que entrem em contato com estes agentes. Entre os NHPAs mais conhecidos e estudados está o nitropireno que se encontra em uma faixa de concentração superior aos demais hidrocarbonetos policíclicos aromáticos nitrados, sendo assim considerado um importante marcador para a presença dos demais no local da contaminação (SABINO et al., 2015).

Dessa maneira há um enorme interesse na busca e aperfeiçoamento de métodos que visam à degradação completa ou parcial destes poluentes e dessa forma recuperar ao longo do tempo os ecossistemas que sofreram grandes perdas de sua biodiversidade original por conta da ação tóxica dessas moléculas (SINGH; VARSHNEY; AGARWAL, 2016).

3.2 MICRO-ORGANISMOS DEGRADADORES DE HPAs

Os HPAs sofrem alterações em suas estruturas perante os agentes causadores de oxidação ou fotodegradação, este último se baseia na ação dos raios ultravioleta (UV), que interferem nas ligações de fosfato entre as cadeias das moléculas, assim alterando sua conformação (MARQUÈS et al., 2016).

Esses processos são formas abióticas de degradação (que não envolvem o uso de micro-organismos vivos) apresentando pouca eficiência devido à formação de moléculas intermediárias indesejáveis ao ecossistema, dessa forma essas técnicas abióticas eliminam esses compostos tóxicos em quantidades não significativas para o meio ambiente (CARUSO; ALABURDA, 2008).

Entre os agentes abióticos relacionados à dHPAs estão: minerais rochosos, como por exemplo, Hematita, Goethita e Magnetita que são capazes de fazer a decomposição de compostos aromáticos com baixa massa molecular, entre ele o Naftaleno, Antraceno e o Fenantreno, dessa maneira diminuindo a concentração desses contaminantes no ambiente através de reações simples de oxidação (KULIK et al., 2006).

No entanto, isoladamente esses processos não são suficientes para reduzir de forma eficaz a concentração dos HPAs presentes no ambiente, visto que essa classe de moléculas possui em alguns casos uma variada gama de funções e

radicais associados que não podem ser oxidados ou degradados apenas por esses minérios, sendo assim um processo que tem se mostrado mais eficaz nesse sentido seria a biorremediação (ABATENH et al., 2017).

A biorremediação consiste na utilização dos metabólitos primários ou secundários produzidos por micro-organismos para a degradação de xenobióticos nocivos presentes no ambiente através de uma sequência de processos catalíticos. Os micro-organismos capazes de atuar nestas moléculas são bactérias ou fungos, além disso, outros fatores estão relacionados com a decomposição desses resíduos, tais como: composição das moléculas de HPAs, diversidade da comunidade microbiana disponível no solo e sua interação, além da disponibilidade de oxigênio do meio, concentração de pH, temperatura do ambiente e substratos associados, por exemplo ao solo (DZIOŃEK; WOJCIESZYŃSKA; GUZIK, 2016).

Dentre os micro-organismos relacionados a dHPAs, estão as bactérias, as principais responsáveis por este processo. Nesse contexto, os gêneros bacterianos que mais se destacam são: *Pseudomonas*, *Mycobacterium*, *Ochrobactrum*, *Ralstonia*, *Comamonas*, *Rhodococcus*, *Sphingomonas*, *Staphylococcus* e *Acinetobacter* (GHOSAL et al., 2016). Destas as bactérias do gênero *Pseudomonas* são amplamente descritas na literatura por atuarem na degradação de xenobióticos, tendo demonstrado a capacidade de degradar várias classes de compostos com grupos aromáticos, dentre esses, pesticidas e HPAs. (WASI; TABREZ; AHMAD, 2013).

No caso do processo de biorremediação realizada por fungos se têm a produção de enzimas ligninolíticas que são responsáveis pela degradação total ou parcial de vários poluentes orgânicos incluindo os HPAs, isso se dá pela similaridade existente entre ambas as classes de moléculas já que são compostos altamente apolares e formados por vários anéis aromáticos (DREVINSKAS et al., 2016). Entre os principais fungos classificados como dHPAs estão às espécies: *Phanerochaete laevis*, *Phanerochaete chrysosporium*, *Ramaria sp.*, *Trametes versicolor*, *Cunninghamella elegans*, *Bjerkandera sp.*, *Pleurotus ostreatus*, *Naematoloma frowardii*, *Aspergillus niger* e *Syncephalastrum racemosum* (KADRI et al., 2017).

Além disso, a biorremediação se subdivide em três categorias básicas *in situ*, *ad situ* e *ex situ* sendo essas duas últimas categorias mais utilizadas em casos de

contaminação por HPAs, pesticidas e poluentes orgânicos clorados que se aderem em solos e sedimentos nos diferentes ecossistemas. Essas 2 últimas categorias são realizadas através do uso de biorreatores que consistem em aparelhos de isolamento para a amostra ambiental contaminada junto com os micro-organismos nativos daquele local, sendo que no biorreator acontece um grande número de reações químicas catalisadas por enzimas principalmente (ROBLES-GONZÁLEZ; FAVA; POGGI-VARALDO, 2008).

Esses biorreatores disponibilizam a criação de um ambiente com condições favoráveis para uma otimização da performance das enzimas responsáveis pela degradação dos xenobióticos, assim o ambiente é provido de uma solução aquosa com pH nas melhores faixas para as reações de catalisação, temperatura ideal, e enriquecimento com nutrientes inorgânicos a base principalmente de sais de nitrogênio e fósforo que seriam utilizados como cofatores enzimáticos na eliminação dos poluentes em questão, além de condições aeróbias favoráveis, dessa forma promovendo uma grande redução na quantidade de HPAs e outros poluentes em comparação com condições normais na natureza (ROBLES-GONZÁLEZ; FAVA; POGGI-VARALDO, 2008).

Sendo que para a maioria dos micro-organismos adaptados para metabolizar e degradar compostos aromáticos, a obtenção de energia e matéria prima para a manutenção de suas estruturas ocorre por meio da quebra e incorporação de carbono das moléculas de HPAs, assim os micro-organismos com essa atividade específica para a degradação de compostos aromáticos e derivados do petróleo representam cerca de 1 a 10% da microbiota de ambientes contaminados por esses resíduos enquanto que em ambientes livres de contaminação esses micro-organismos representam cerca de 0,1% ou menos da microbiota do ambiente (VARJANI, 2017).

3.3 FAMÍLIAS DE PROTEÍNAS RELACIONADAS À dHPAs

Em relação aos principais grupos de enzimas relacionadas com a dHPAs estão: as *oxigenases*, *peroxidases*, *esterases* e suas variações de acordo com o radical associado a proteína. Estas participam de várias vias metabólicas, atuando diretamente nas etapas de degradação, sendo essas vias incluídas dentro da

categoria de degradação de várias moléculas HPAs entre outras classes de poluentes orgânicos persistentes (PAGÉ; YERGEAU; GREER, 2015).

As principais proteínas relacionadas à dHPAs foram descritas no trabalho de (VILCHEZ-VARGAS et al., 2013) e foram distribuídas nas seguintes famílias: ALKB (*Alkane hydroxylases*), BCOAR (*Benzoyl coenzyme A reductases*), BSSA (*Benzylsuccinate syntases*), CYP153 (*Cytochrome P450 alkane hydroxylases*), EXDO (*Extradiol dioxygenases*), FERRE (*Ferredoxins*), INDO (*Intradiol dioxygenases*), MACR (*Maleylacetate reductases*), MCIS (*Muconate cycloisomerases*), RHDO (*Ring hydroxylate dioxygenases*) e RHMO (*Ring hydroxylate monooxygenases*).

Dessas famílias um dos grupos de enzimas mais conhecidas dessa via são as enzimas do complexo do citocromo P450 *monooxygenases*, que engloba a família de proteínas CYP153 (SAZYKIN et al., 2019), essa classe de enzimas é responsável pela clivagem e degradação de alguns alcanos e outros compostos cíclicos aromáticos como: limoneno, ciclohexano e estireno. Esse grupo de enzimas além de ser encontrado em bactérias, também é presente em grande parte dos vertebrados inclusive mamíferos, tendo sua participação na clivagem de xenobióticos assimilados pelo organismo, sendo o fígado o principal órgão onde esses compostos tóxicos são metabolizados (FIORENTINI et al., 2018).

Outra família é a ALKB, representada pelas enzimas *alkane hydroxylases*, que são as principais enzimas relacionadas com a degradação de alcanos. A via de degradação desses compostos é complexa, envolvendo várias etapas e com moléculas intermediárias originando alguns subprodutos, as proteínas AlkB possuem outras enzimas e moléculas que agem como cofatores associados na rota de degradação, entre elas as principais moléculas são: AlkG, AlkT e NADPH (PICCOLO et al., 2011).

Assim para o funcionamento das enzimas *alkane hydroxylases* (AlkB), elas requerem duas proteínas, a rubredoxina (AlkG) e redredoxina *redutase* (AlkT) para que ocorram as reações relacionadas a dHPAs e a interação entre essas proteínas se dá pela transferência de elétrons entre AlkG e AlkT através da molécula NADPH que realiza o carregamento dos elétrons (LIU et al., 2017). Além disso, a família ALKB é codificada pelo gene *almA*, este gene está presente principalmente nas

bactérias dos gêneros: *Alcanivorax*, *Marinobacter*, *Acinetobacter* e *Parvibaculum* (SHAO; WANG, 2013).

Em relação à família INDO (*Intradiol dioxygenases*), elas possuem enzimas dos grupos: *catechol dioxygenases*, *hidroxiquinol dioxygenases* e *protocatechuate dioxygenases*, sendo responsáveis pela clivagem de HPAs que possuem as funções de catecol, pirogalol, hidroxiquinol e ácido protocatecuico (GUZIK; HUPERT-KOCUREK; WOJCIESZYŃSKA, 2013). Nesta família se tem destaque a enzima catecol 1,2-dioxigenase (C12O) que é a responsável pela clivagem da molécula de catecol na posição -1,2 (orto) do anel aromático, sendo que essa família de enzimas foi identificada principalmente em bactérias dos gêneros: *Acinetobacter*, *Arthrobacter*, *Corynebacterium*, *Pseudomonas*, *Rhodococcus*, *Stenotrophomonas* e *Streptomyces* (XU et al., 2017).

Tratando-se da família EXDO (*Extradiol dioxygenases*), essa compartilha grupos de enzimas com a família INDO com a diferença que durante os eventos de degradação em cascata dos HPAs, a clivagem do anel aromático ocorre na posição -1,3 (meta) entre os carbonos e utiliza como cofatores para esta degradação os íons de ferro (Fe^{2+}) e magnésio (Mg^{2+}) (KUATSJAH et al., 2017). Essa família de enzimas é encontrada principalmente em bactérias das espécies: *Pseudomonas putida*, *Pseudomonas pseudoalcaligenes*, *Cupriavidus necator*, *Rhodococcus rhodochrous*, *Sphingomonas paucimobilis* e *Burkholderia xenovorans* (TERRÓN-GONZÁLEZ et al., 2016).

No caso da família FERRE (*Ferredoxins*) elas são um grupo de proteínas responsáveis pela transferência de elétrons em um grande número de reações de oxido-redução nas diferentes etapas de dHPAs, sendo neste caso aqueles HPAs que possuem um baixo peso molecular, ou seja, moléculas de cadeia pequena em comparação à média de tamanho e peso dos demais compostos aromáticos como por exemplo grupos de n-alcenos e atuando em interações de metabolização com outras famílias de proteínas dHPAs como por exemplo a família BCOAR em etapas de degradação do benzeno e benzoato (MECKENSTOCK et al., 2016; MIGNÉE et al., 2017). Além disso, esse grupo de proteínas também atua em conjunto com a família CYP153, se envolvendo nas reações de degradação de grupos de alcenos, toluenos e catecois, sendo que as Ferredoxinas estão presentes principalmente em bactérias do gênero *Pseudomonas* (YANG et al., 2018).

Outra família é representada pelas BCOAR (*Benzoyl coenzyme A reductases*) que possuem como principal característica a atividade catalítica da molécula de Benzoil-CoA que é um dos subprodutos intermediários da degradação do Benzoato, que é caracterizado como uma molécula representada por um grupo benzeno com uma ligação dupla e simples para 1 átomo de oxigênio e outro átomo de enxofre respectivamente, essa molécula está presente em alguns HPAs, e a reação de degradação é realizada principalmente por bactérias do gênero *Rhodopseudomonas*, *Azoarcus* e *Thauera* (SCHMID; RENÉ; BOLL, 2015).

Já a família das enzimas BSSA (*Benzylsuccinate syntases*) são responsáveis pela degradação do tolueno e derivados, que são formados basicamente por anéis benzênicos com grupamentos metil em sua estrutura, neste caso, a função da enzima é facilitar uma reação de adição entre o tolueno e uma molécula de fumarato, criando assim o benzilsuccinato, uma molécula mais hidrossolúvel, portanto mais facilmente assimilável para o meio ambiente e que pode ser eliminada mais facilmente por outros organismos. Essas enzimas são encontradas principalmente em bactérias do gênero: *Thauera*, *Aromatoleum*, *Azoarcus*, *Magnetospirillum* e *Georgfuchsia* (HEIDER et al., 2016; SEYHAN et al., 2016).

Em relação às enzimas pertencentes à família MCIS (*Muconate cycloisomerases*), elas estando associadas com outras enzimas das famílias BSSA e BCOAR, e são responsáveis pela clivagem da cascata catalítica na via de dHPAs, principalmente do Benzoato e HPAs com função Fenol atuando também na degradação de compostos derivados do muconato que são moléculas com função de aldeídos intermediários da via de metabolização do Tolueno. As MCIS também interagem com enzimas das famílias EXDO e INDO na decomposição de catecol e protococatecol (CABRAL et al., 2018).

No caso da família MACR (*Maleylacetate reductases*), ela têm como função em fungos e leveduras atuar na degradação de certos compostos aromáticos como: Fenol, Tirosina e derivados do Benzoato, além disso, em bactérias essa família participa principalmente na rota de dHPAs, com destaque para a decomposição do beta-resorcilato (*2,4-dihydroxybenzoate*) e gama-resorcilato (*2,6-dihydroxybenzoate*), esta classe de enzimas é auxiliada pela molécula de NADH, que participa como cofator na reação de redução (FUJII et al., 2016).

Em relação às famílias RHMO (*Ring hydroxylate monooxygenases*) e RHDO (*Ring hydroxylate dioxygenases*) ambas pertencem à classe de proteínas *oxigenases* de hidroxilação do anel aromático de grande parte dos HPAs que possuem baixo peso molecular como Fenantreno, Pireno, Fluoranteno e Naftaleno, dessa maneira produzindo moléculas de cis-diidrodióis como produto final da reação, que é um composto menos nocivo. Essas enzimas são encontradas principalmente em algumas bactérias das espécies: *Rhodococcus sp.* e *Mycobacterium sp.* (KIM et al., 2018; MUSUMECI et al., 2019; PENG et al., 2018).

Por último a família TDLH (*trans-Dienelactone hydrolases*) são enzimas dependentes de íons de zinco, responsáveis pela dHPAs com derivados de clorosalicilatos, atuando especificamente sobre a molécula de 4-cloro muconolactona convertendo ela em maleilacetato, que é uma forma mais hidrossolúvel da molécula, essas enzimas são presentes principalmente nas bactérias da espécie *Pseudomonas reineke* (CÂMARA et al., 2008).

Considerando-se o enorme potencial dos micro-organismos para a degradação dos xenobióticos, estudos têm sido realizados para o aperfeiçoamento das técnicas de biorremediação que por meio de análises dos metagenomas de ambientes contaminados por HPAs, tornou-se possível a identificação de novos genes responsáveis por codificar enzimas diretamente ligadas com a dHPAs de grande permanência no meio ambiente e nos seres vivos. Por conta disso, pesquisas nesse campo vêm apresentando grandes resultados e uma enorme quantidade de novas informações (TECHTMANN; HAZEN, 2016; THOMAS et al., 2016).

3.4 METAGENÔMICA

3.4.1 ESTUDOS METAGENÔMICOS

Anteriormente, o acesso à diversidade microbiana apresentava grandes desafios, pelo fato de a maioria dos micro-organismos existentes no meio ambiente não poderem ser cultivados *in vitro* utilizando abordagens convencionais, sendo assim o estudo ficava restrito as populações de bactérias e fungos que representavam uma parcela bem menor da verdadeira comunidade microbiana que

compõe os diferentes ecossistemas. Foi só com o advento da metagenômica que houve um aumento na prospecção de novos biocatalisadores ou biomoléculas produzidas por micro-organismos não cultiváveis *in vitro*, dessa maneira contribuindo para maiores informações que não poderiam ser acessadas pelos métodos de microbiologia clássica tradicional (BERINI et al., 2017).

O termo metagenômica foi introduzido por HANDELSMAN et al. (1998), propondo um acesso a composição molecular de micro-organismos desconhecidos presentes no solo, como uma fronteira para obtenção de produtos de origem biológica. A técnica se baseia na prospecção de atividades biológicas, a partir da obtenção do DNA da amostra ambiental (*environmental DNA*, ou eDNA), ou seja, DNA de uma mistura de diferentes populações de micro-organismos em um determinado ambiente, assim sendo possível por meio da análise do eDNA obter informações da diversidade e riqueza de espécies da comunidade microbiana que está sendo avaliada no estudo em questão (HANDELSMAN, 2004).

Os estudos metagenômicos basicamente estão subdivididos em duas categorias: metagenômica estrutural e metagenômica funcional. Na primeira abordagem o objetivo é estudar a estrutura da população microbiana não cultivada, a fim de reconstruir a complexa rede metabólica estabelecida entre os membros (populações de bactérias e fungos) da comunidade, isso ocorre a partir do sequenciamento, montagem, anotação e demais análises complementares realizadas a partir do metagenoma em que a comunidade está inserida. Enquanto que a metagenômica funcional tem por objetivo a partir de bibliotecas construídas, a busca por clones que expressem uma característica de interesse e isso pode ser realizado pela subclonagem e sequenciamento do fragmento metagenômico em um vetor de expressão e posterior anotação gênica com base nas ORFs (*Open Read Frames*) que potencialmente expressam a proteína ou a biomolécula de interesse e assim essa molécula ou proteína pode ser posteriormente isolada e utilizada de acordo com suas aplicações (ALVES et al., 2018).

Tais análises começaram a ser realizadas mais rapidamente e de maneira mais eficiente com o surgimento do Sequenciamento de Nova Geração (*Next Generation Sequencing*, ou NGS), onde a partir de 2005 a quantidade de genomas sequenciados aumentou de maneira significativa e as facilidades atualmente encontradas para se fazer o sequenciamento auxiliaram na melhora na aquisição

de dados e no aumento da performance obtida por esses métodos (SCHOLZ; LO; CHAIN, 2012).

Além disso, diversas ferramentas de Bioinformática foram desenvolvidas e estão sendo constantemente aperfeiçoadas para auxiliarem na reconstrução dos genomas, os *Metagenome-Assembled Genomes* (MAGs), na predição de genes, nos estudos de composição taxonômica e funcional com o auxílio de análises estatísticas, nos estudos de relações filogenéticas, entre outros relacionados ao metagenoma. Tais estudos têm por objetivo auxiliar na elucidação de certos aspectos da comunidade microbiana que atualmente não pode ser facilmente cultivada, permitindo obter conhecimentos a fim de compreender mais a complexa rede metabólica de interação entre as diferentes populações de micro-organismos do ambiente a ser analisado (ROUMPEKA et al., 2017), e com isso poder desenvolver novos produtos de interesses biotecnológicos, como por exemplo relacionados à biorremediação.

Em relação as etapas de análise do estudo de metagenômica, elas se têm início com a aquisição de material coletado no ambiente de interesse. Assim as amostras devem representar uma parcela das comunidades de micro-organismos do local em que elas foram retiradas, sendo uma das maiores dificuldades ser estimar a quantidade de amostras ideais para se obter um alto grau de confiabilidade. Por conta disso são usadas, as curvas de rarefação que são responsáveis por relacionar o a proporção de espécies no ambiente como uma função do número de indivíduos existentes na amostra (WOOLEY; GODZIK; FRIEDBERG, 2010).

Outro fator importante a ser levando em consideração é a distribuição da comunidade microbiana que deve ser analisada antes da realização do sequenciamento. Sendo que a comunidade microbiana é estimada em função de sua diversidade. A diversidade é avaliada de acordo com o número de espécies em uma determinada comunidade, ou seja, a riqueza que ela representa, como também da abundância relativa dos seus componentes, também conhecida como equitabilidade (KUNIN et al., 2008).

Esses fatores entram na categoria de análise preliminar de metadados, onde além do conhecimento sobre a composição biológica da comunidade, são necessárias outras informações como, por exemplo: condições geográficas da

amostra do metagenoma coletado, temperatura, pH do ambiente coletado, interferência antrópica no local, métodos de extração de DNA utilizado para o sequenciamento, preparo das bibliotecas de sequenciamento, e o sequenciamento em si, entre outras informações. Assim considerando esses fatores, melhores serão as inferências sobre as características ambientais, além de servir como um arcabouço de informação para a interpretação dos dados que serão gerados durante as próximas análises (FELCZYKOWSKA et al., 2015).

Quanto ao sequenciamento do metagenoma, esse procedimento ocorre principalmente por meio de duas abordagens, uma envolvendo o sequenciamento de forma dirigida e a outra por meio da abordagem do sequenciamento *shotgun* de DNA. A metodologia mais comum consiste na utilização do sequenciamento de forma dirigida, que consiste na amplificação por PCR de um ou mais genes marcadores, que demonstrem elevada variabilidade de sequência nos grupos de organismos em estudo, sendo que o tamanho dos fragmentos da PCR deve ser adequado ao comprimento das leituras (*reads*) da plataforma de sequenciamento, para que dessa forma possam ser sequenciados na totalidade evitando perda de informação biológica. Além disso, as leituras dos fragmentos sequenciados são posteriormente analisadas utilizando algoritmos estatísticos e bases de dados específicas de sequências, como por exemplo do gene 16S rRNA (no caso de procariotos) (HODKINSON; GRICE, 2014).

3.4.2 ANÁLISE DE DADOS DE SEQUENCIAMENTO *SHOTGUN* DE METAGENOMAS

No caso da metodologia de sequenciamento do tipo *shotgun*, ela se baseia em fragmentar as moléculas de DNA em cadeias de pequeno comprimento e sequenciá-las sem qualquer tipo de seleção previa. No entanto, devido a esse método de fragmentação, a análise bioinformática dos dados gerados pelo sequenciamento é muito mais complexa do que a utilizada pelo sequenciamento de forma dirigida. Entre outros fatores por conta de que o genoma de cada micro-organismo é sequenciado a partir de múltiplos fragmentos distintos e dessa forma se faz necessário a aplicação de algoritmos informatizados específicos de forma a obterem-se sequências parcialmente sobrepostas e contíguas, que permitam

reconstruir a sequência genômica de cada espécie presente na amostra que está fragmentada (HODKINSON; GRICE, 2014).

A partir do sequenciamento se têm as etapas de avaliação da qualidade das sequências geradas e um pré-processamento delas seguidas por uma nova avaliação de qualidade. Muitas vezes isso é chamado de análise primária e deve ser utilizada antes que as *reads* passem para uma nova fase de análise metagenômica. A fase de pré-processamento deve iniciar com uma avaliação global dos dados obtidos, incluindo a qualidade das bases sequenciadas, a proporção de cada base ao longo do comprimento das *reads*, a porcentagem do conteúdo guanina-citosina (GC) e a presença de sequências de adaptadores que representam impurezas que atrapalharão na próxima etapa de montagem do metagenoma (SCHMIEDER; EDWARDS, 2011).

Entre as principais abordagens no processo de análise de dados de sequenciamento *shotgun* de amostras metagenômicas se encontra o processo de montagem do metagenoma que ocorre posteriormente ao sequenciamento, o qual é realizado normalmente por um sequenciador de alto rendimento, ou também chamado nova geração (*Next Generation Sequencing*, ou NGS). A técnica de montagem se baseia na sobreposição das leituras curtas de fragmentos de DNA e um mesmo *locus* genômico. Esse processo é realizado para a obtenção de sequências mais longas, denominadas *contigs*, com o propósito de reconstruir o genoma individual de cada organismo presente em um metagenoma ou representar o repertório gênico de uma dada comunidade (RODRIGUE et al., 2010).

De maneira geral o processo de montagem pode ser feito de duas formas: por meio de alinhamentos utilizando referências (*reference-based assembly*) ou sem referências (*de novo assembly*). A montagem com referências pode ser realizada através do alinhamento dos fragmentos sequenciados contra a sequência de DNA de micro-organismos, frequentemente dos que podem ser cultivados em laboratório. Dessa maneira, a montagem por referência é bastante limitada, pois os genomas de referência somente representam uma pequena porção da diversidade biológica de micro-organismos encontrados nos ambientes naturais.

Assim, é comum fazer a montagem de um metagenoma sem a utilização de referências, possibilitando o conhecimento das sequências de uma parcela maior de micro-organismos presentes na amostra coletada (NIELSEN et al., 2014). Essa

estratégia também é conhecida como montagem *de novo*. Nessa estratégia, as leituras curtas obtidas das plataformas NGS são divididas em subsequências menores de tamanho k (*k-mers*), as quais serão combinadas de acordo com a identidade prefixo-sufixo e organizadas em uma estrutura de grafos, os quais serão analisados em busca de um caminho que levará à reconstrução dos genomas. Por exemplo, supondo que uma leitura seja composta pela sequência GTATTAG, na utilização de 3-*mers* serão obtidos os *k-mers* GTA, TAT, ATT, TTA e TAG. Em seguida, é criado um grafo (*De Bruijn graph*) com as sobreposições entre todos os *k-mers* distintos obtidos das *reads* e os *contigs* são gerados através dos caminhos percorridos a partir desse grafo de sobreposições e assim as sequências que compõem o metagenoma vão sendo formadas por meio dessa lógica (BREITWIESER; LU; SALZBERG, 2019).

A etapa seguinte a esse processo é a anotação genômica, a qual envolve a predição e anotação de genes, e a designação taxonômica.

A predição gênica é um processo que envolve o reconhecimento dos *loci* gênicos dentro dos *contigs*. Existem muitos programas de predição gênica que permitem o reconhecimento de padrões correspondentes às fases de leitura abertas nas sequências genômicas, também conhecidas como ORFs (*Open Read Frame*), que são regiões na sequência demarcadas por um códon iniciador que se prolonga até o surgimento de um códon de parada (*stop codon*). Dessa forma, as ORFs correspondem a sequências de trinças de nucleotídeos codificadoras de aminoácidos. Nesse sentido, a busca e a quantificação dessas regiões contribui com o arcabouço de informações para supor a predição de possíveis genes funcionais e diferentes proteínas que podem ser expressas por eles (HOFF; STANKE, 2015).

No processo de anotação gênica, a atribuição de funções se dá por meio da identificação de padrões nas sequências correspondentes a outras sequências gênicas com funções já determinadas. Dessa forma, nessa etapa de anotação dos genes são realizadas buscas por similaridade das proteínas preditas em bancos de sequências responsáveis por sintetizar proteínas funcionais previamente conhecidas (WANG; CHEN; LI, 2004). Essas proteínas nos bancos de dados são consideradas referências no processo de anotação funcional.

A designação taxonômica pode ser descrita como o processo de atribuição de uma unidade taxonômica definida a uma sequência, seja ela um *contig* ou uma *read*, ou então pode ser realizada em um agrupamento dessas sequências. Esse agrupamento acontece, de forma ideal, a partir das sequências correspondentes aos fragmentos de DNA pertencentes a determinado táxon de um mesmo grupo taxonômico (*bin*). Esse processo de agrupamento, chamado de *binning*, é realizado considerando as características das sequências além das correspondentes abundâncias nas amostras (SEDLAR; KUPKOVA; PROVAZNIK, 2017). A seleção do método utilizado para efetuar a designação taxonômica é um aspecto muito importante, sendo que atualmente existe um número muito elevado de métodos e programas disponíveis para este fim. No entanto, no meio da aparente diversidade, é possível categorizar os métodos de classificação taxonômica em 4 principais tipos, que incluem os métodos de semelhança, composição, híbridos e de genes marcadores (MANDE; MOHAMMED; GHOSH, 2012).

3.4.3 ANÁLISE DE DIVERSIDADE EM METAGENOMAS

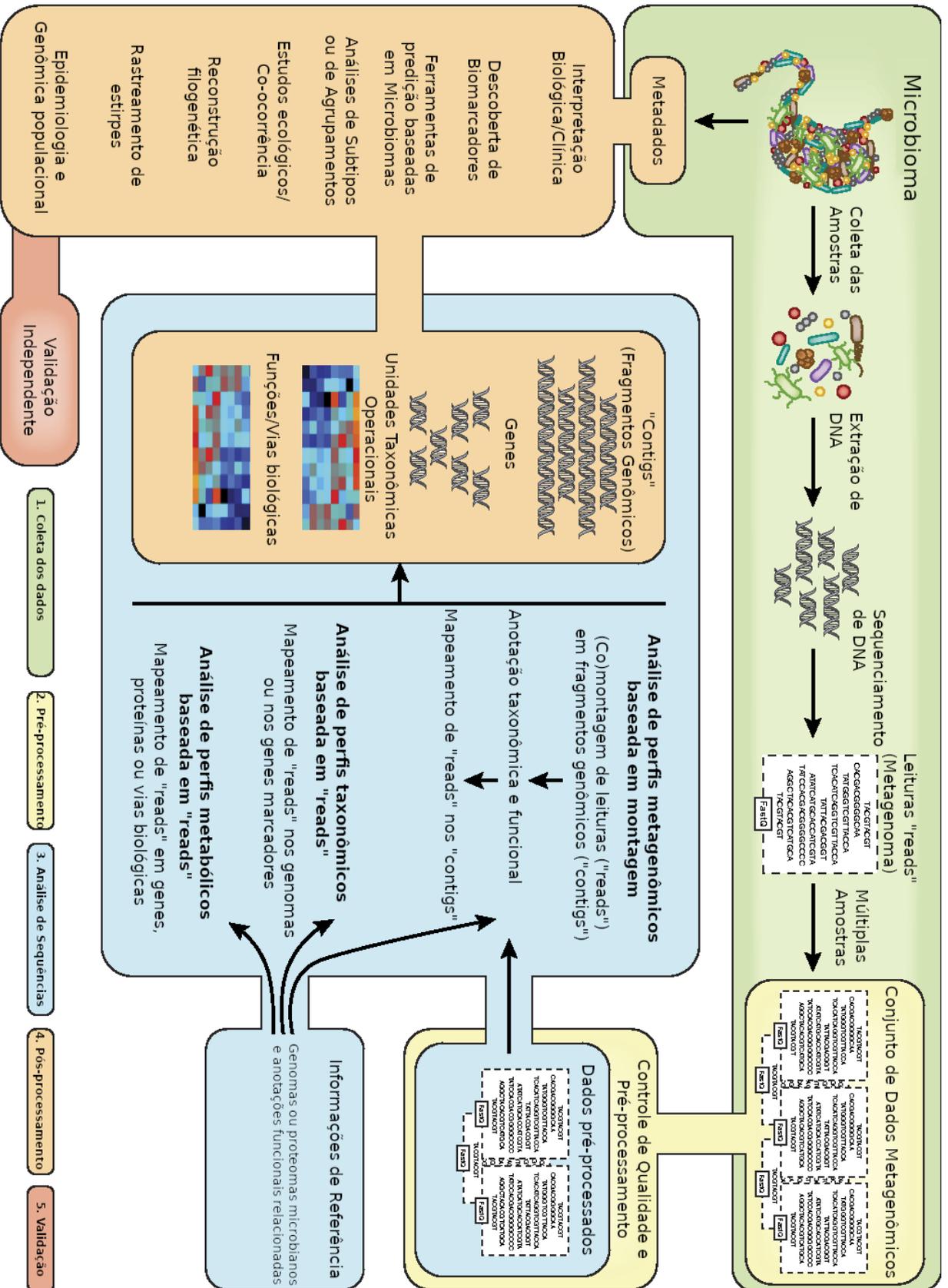
A partir dos dados de designação taxonômica é possível construir uma relação de abundância e riqueza de micro-organismos que constituem a comunidade do metagenoma que está sendo analisada, porém, são várias limitações para se obter uma designação taxonômica bastante fidedigna com a realidade. Uma delas é que as bases de dados dos genomas de referência são atualmente incompletas e altamente tendenciosas (HUGENHOLTZ, 2002).

3.4.4 RESUMO

Levando-se em conta a estrutura geral do estudo e análises de metagenomas é possível visualizar o fluxo de etapas desses processos descritos na **Figura 1**.

Figura 1: Representação resumida de um fluxo de trabalho com as abordagens usuais em estudos metagenômicos (QUINCE et al., 2017). Etapas: (1) Desenho do estudo e protocolo experimental; (2): pré-processamento dos dados,

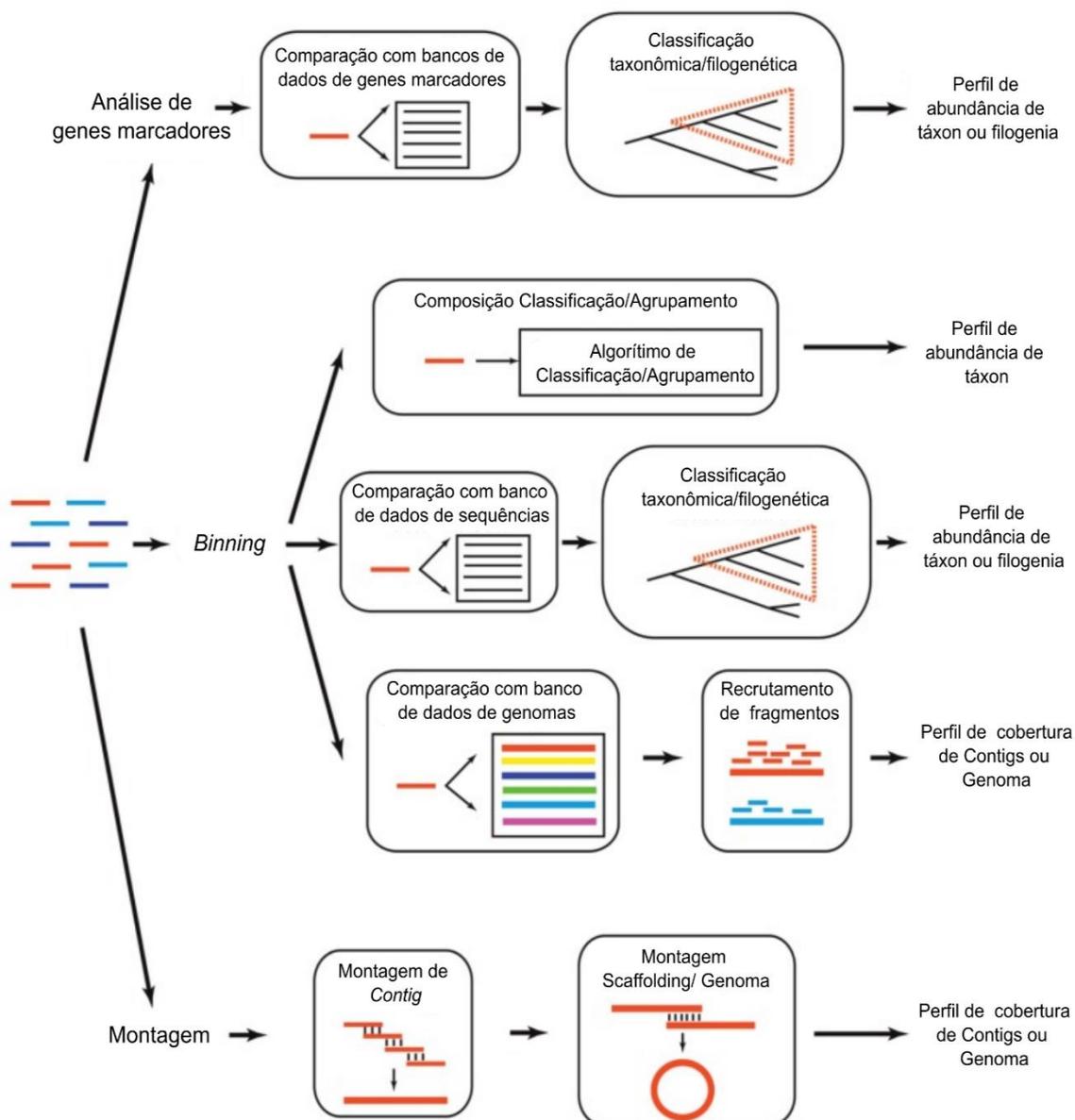
incluindo o controle de qualidade, remoção de adaptadores e regiões de baixa qualidade, além da remoção de sequências provenientes de DNA que não seja microbiano (ex. hospedeiro quando for o caso); (3): Análise de sequências, com uma combinação de estratégias baseadas em leituras (*reads*) ou montagem; (4): Pós-processamento, onde são empregadas técnicas de análise multivariada (ex.: Análise de Coordenadas Principais – PCoA) para auxiliar na exploração e interpretação dos dados; (5): Validação, das conclusões obtidas a partir da análise exploratória, o que poderá direcionar estudos dirigidos, onde análises subsequentes serão essenciais.



Fonte: Adaptado de (QUINCE et al., 2017).

A seguir, em mais detalhes na **Figura 2**, o fluxo de trabalho da etapa de análises das sequências metagenômicas, apresentando três estratégias analíticas para a obtenção de perfis genômicos, taxonômicos ou de linhagens filogenéticas em uma comunidade microbiana (SHARPTON, 2014). A análise de genes marcadores, que consiste em comparar cada *read* com um banco de dados referência de informações taxonômicas ou filogenéticas, usando algoritmo de classificação a fim de reconhecer a *read* como pertencente a um gene marcador homólogo, possibilitando a anotação baseada na similaridade. A estratégia utilizando a segregação de sequências baseada em características semelhantes (*binning*), pode considerar 3 metodologias: (1) composicional, o qual utiliza a composição de nucleotídeos semelhante para determinar os grupos taxonômicos; (2) por similaridade, o qual classifica as sequências em grupos baseando-se em similaridade a genes ou proteínas previamente reconhecidas para determinado grupo taxonômico; (3) por recrutamento de fragmentos, em que as sequências são alinhadas a sequências genômicas a fim de que a estimativa de cobertura semelhante, permita definir os agrupamentos. Finalmente, as sequências podem ser submetidas a um processo de montagem a fim de criar sequências contíguas, (*contigs*) que podem ser reorganizados em *scaffolds* ou *supercontigs* ou até mesmo genomas completos.

Figura 2: Representação resumida de um fluxo de trabalho para a análise de dados metagenômicos utilizando sequenciamento *shotgun*.



Fonte: Adaptado de (SHARPTON, 2014).

3.5 BUSCAS POR BIOMOLÉCULAS DE INTERESSE

Além desses processos, outras análises podem ser avaliadas e posteriormente integradas no fluxo de estudo metagenômico, dentre elas a busca e investigação de biomoléculas que podem estar presentes na maquinaria enzimática dos micro-organismos presentes na comunidade estudada na amostra de metagenoma coletada.

No caso de famílias de proteínas, existem muitos programas que por meio do alinhamento contra banco de dados, podem fazer uma busca por membros da mesma família com base na similaridade dos domínios proteicos entre as sequências comparadas, um desses programas é o FlowerPower (KRISHNAMURTHY; BROWN; SJÖLANDER, 2007), que possui um nível de acurácia maior em comparação com as ferramentas BLAST e PSI-BLAST para a recuperação de sequências homólogas para um mesmo domínio proteico.

O FlowerPower possui um algoritmo iterativo de agrupamento, que faz uso do método de buscas por similaridade utilizado pelo PSI-BLAST, ou seja, ele parte de um alinhamento de sequências similares e construção de um perfil PSSM (*Position-Specific Scoring Matrix*) e o utiliza para recuperar outras proteínas similares que são consideradas na atualização e aperfeiçoamento do perfil PSSM para então realizar um novo ciclo de buscas até atingir convergência. Dessa forma, o processo visa encontrar proteínas de uma mesma família, as quais potencialmente são ortólogas. Depois o FlowerPower utiliza o método de SCI-PHY (*Subfamily Classification In Phylogenomics*), para predizer subfamílias com base no alinhamento de múltiplas sequências recuperadas em cada estágio e dispostas em uma árvore hierárquica baseada em um algoritmo aglomerativo conforme a similaridades. Para particionar as sequências em subfamílias, o SCI-PHY corta a árvore em um ponto determinado utilizando uma combinação de métodos de teoria da informação e modelos de mistura de Dirichlet (SJÖLANDER et al., 1996). Essa classificação em subfamílias deve corresponder intimamente a clados filogenéticos e subtipos identificados por espécie (SJÖLANDER, 1998). Ao final do processo, a partir dos alinhamentos dentro de uma mesma subfamília são gerados os HMMs que podem ser utilizados para buscar em bancos de sequências, aquelas que são similares.

Dessa forma, é possível a partir desses modelos encontrar proteínas homólogas, mesmo as mais distantes e detectar novos membros de cada família e

apesar de não ser possível afirmar categoricamente que genes ortólogos assumem funções equivalentes, temos que assumir essa conjectura dos ortólogos, considerando as evidências de que eles tendem a reter a função do gene ancestral (ALTENHOFF et al., 2012).

Por fim com base nos resultados obtidos em cada uma das etapas é possível fazer buscas e comparações em diferentes metagenomas, para dessa forma se obter informações sobre a capacidade gênica das populações de micro-organismos levando-se em conta as diferentes condições ambientais. Assim a comparação de amostras em diferentes ambientes podem indicar como fatores do ecossistema em que as comunidades estão inclusas podem moldar a distribuição desses micro-organismos de acordo com as características ambientais, por exemplo, favorecendo assim um grupo de micro-organismos em detrimento de outro por conta da presença de um xenobiótico tóxico que prejudicam as bactérias sensíveis a eles e em contraste favorecem aquelas bactérias que possuem enzimas dirigidas a sua degradação (POURSAT et al., 2019; ZHANG et al., 2019).

3.6 ABORDAGEM METAGENÔMICA APLICADA À ANÁLISE DE XENOBIÓTICOS

Entre os estudos realizados sobre o tema de xenobióticos, particularmente sobre contaminações de HPAs se tem como exemplo o trabalho de revisão bibliográfica realizado por Ufarté et al.. (2015), que reportou diferentes resultados relacionados com a prospecção de enzimas degradadoras de poluentes orgânicos (HPAs, plásticos, piretróides, entre outros) obtidos por metagenômica funcional.

Também há o exemplo do trabalho realizado por Duarte et al. (2017), onde foi analisado um solo altamente contaminado por HPAs e metais pesados, que passaram por um processo de remediação através da escavação do solo (para promover oxigenação) e biodegradação, em um período de 12 horas, com o propósito de fazer uma triagem para avaliar o catabolismo aeróbio dos micro-organismos do solo. Nesse trabalho foi obtido um total de 422.750 clones de fosmídeos que foram rastreados em busca de atividades de clivagem de anéis aromáticos usando 2,3-di-hidroxibifenil como substrato. Assim após o sequenciamento mais de 200 genes foram identificados como codificadores de *extradiol dioxigenase* de três superfamílias diferentes. Além disso, genes chaves

adicionais de vias metabólicas, relacionadas a compostos aromáticos foram identificados, incluindo vários que codificam as enzimas *oxigenases*. Os resultados obtidos nesse estudo, indicaram uma complexa rede de interação na microbiota local do estudo, que compreende organismos semelhantes ao *Immundisolibacter cernigliae* TR3.2 e *Rugosibacter aromaticivorans* Ca6.

Outro trabalho com metagenoma de solo foi realizado por Bao et al. (2017), que focou em uma abordagem na metagenômica estrutural. Nesse trabalho, os autores tiveram o objetivo de analisar o metagenoma do solo de uma das áreas de perfuração no campo petrolífero de Tianjin Dagang, no leste da China. A análise desses metagenomas identificou predominância para classes γ -*Proteobacteria* e α -*Proteobacteria*, que são grupos chave na dHPAs. Na análise funcional foram encontrados grupos de enzimas super-representadas e vias envolvidas na degradação de um amplo conjunto de xenobióticos, e HPAs, incluindo tolueno, xileno, clorobenzoato, aminobenzoato, DDT, metilnaftaleno e bisfenol.

Outro trabalho relevante é o de Zhao et al. (2020), onde foi realizado um estudo metagenômico comparativo da resposta microbiana diante de vazamentos naturais de hidrocarbonetos no Golfo do México. Os sedimentos subterrâneos (10–15 cm abaixo do fundo do mar) foram coletados de quatro locais de vazamento por gás natural e fluidos de hidrocarbonetos (D27, D33, D72 e D75), também de dois locais de referência (D21 e D30) onde não houve contaminação e 1 local de transição entre a área contaminada (D24). Os conjuntos de dados de sequenciamento de metagenoma resultantes foram analisados com abordagens baseadas em genes e baseadas em genomas. As análises baseadas no gene 16S rRNA sugerem que as amostras de vazamento são distintas das referências tanto pelo conteúdo fracionário do rRNA 16S quanto pela filogenia, sendo que bactérias redutoras de sulfato estão presentes em ambos os tipos de amostras, com maiores abundâncias relativas das amostras de vazamento do que nas referências. Genes para fixação de nitrogênio foram encontrados predominantemente nos locais de vazamento, enquanto os locais de referência mostraram um sinal dominante para a oxidação anaeróbia de amônio. Este estudo sugere que locais de contaminação de hidrocarbonetos podem mudar muito a estrutura da comunidade microbiana, estimulando a fixação de nitrogênio, alterando inerentemente o metabolismo do

nitrogênio em comparação com os dos sedimentos de referência (ZHAO et al., 2020).

Da mesma maneira, outro importante estudo relacionado à contaminação de uma área sedimentar por HPAs e outros derivados do petróleo é o de Meier; Paterson; Lambert (2016), onde foram analisadas amostras de sedimentos na Baía de Rock Bay (Victoria Harbour, British Columbia, Canada) que é uma área industrial, onde ocasionalmente há acidentes e derramamento de combustíveis fósseis, ocasionando a contaminação no local. Dentre os principais resíduos poluentes orgânicos persistentes encontrados nesse local, estão os HPAs. Este estudo combinou uma triagem fenotípica conhecida como expressão do gene induzida por substrato (SIGEX) com o sequenciamento *shotgun* das amostras de metagenoma. Nesse trabalho foram construídas 3 bibliotecas para o sequenciamento de alto rendimento. Como resultado, os clones recuperados continham sequências com vários graus de semelhança com genes (ou genes parciais) envolvidos no metabolismo de compostos aromático. Os resultados obtidos demonstraram haver ampla aplicabilidade na identificação de subconjuntos biologicamente relevantes de metagenomas (incluindo sequências novas e conhecidas de genes responsáveis pela dHPAs) (MEIER; PATERSON; LAMBERT, 2016).

Portanto, a abordagem utilizando o estudo da metagenômica com o suporte das ferramentas e programas de bioinformática na análise dos dados, abre grandes perspectivas para o futuro das técnicas de biorremediação de HPAs, as quais tiveram um grande aumento na eficiência de seus resultados nos últimos anos, dessa maneira sendo uma área ainda emergente e promissora para o auxílio na recuperação dos ecossistemas contaminados por HPAs (SHAHSAVARI et al., 2019).

4. MATERIAL E MÉTODOS

4.1 IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS PROTEÍNAS dHPAs

Foi realizada a busca por sequências de proteínas relacionadas à dHPAs, tendo sido encontradas 1.302 sequências proteicas distribuídas entre 12 famílias

(VILCHEZ-VARGAS et al., 2013) descritas e disponibilizadas em um trabalho de análise de genes microbianos relacionados a degradação de compostos aromáticos.

As proteínas foram comparadas ao banco de dados UniRef100 (SUZEK et al., 2007) somente para procariotos (excluindo eucariotos) utilizando o programa DIAMOND (BUCHFINK; XIE; HUSON, 2015) no modo blastp para encontrar as melhores correspondências a fim de recuperar as proteínas não-redundantes compatíveis que estão dentre as que possuem os melhores escores de alinhamento, ou seja, até 1% menores do que o escore máximo e uma cobertura mínima da sequência consultada (*query coverage*) e da sequência do banco de dados (*subject coverage*) de 99%, além de *e-value* mínimo de 1e-05 e com incremento na sensibilidade de busca (*--more-sensitive*). Das 1.302 proteínas, foram detectadas correspondências (*hit matches*) para 1.298. Ou seja, das 1.302, três proteínas (AAT02410.1, ACG41239.1 e XP_958382.1) eram de eucariotos (*Brassica napus*, *Zea mays* e *Neurospora crassa* OR74A), a outra (AAL76201.1), de *Comamonas sp. JS765*, foi deixada de lado por ter um tamanho discrepante (132 aa) dos tamanhos das outras (104 aa) que compõem o agrupamento UniRef100 correspondente (UniRef100_A1WAE5).

As sequências recuperadas para cada família foram alinhadas de um modo global utilizando a ferramenta T-Coffee (DI TOMMASO et al., 2011), responsável por realizar alinhamentos múltiplos de sequências e foi configurada para a execução no modo *m-coffee*, que combina a saída de múltiplos alinhadores.

Posteriormente, os resultados dos alinhamentos múltiplos das sequências de cada família proteica foram submetidas ao programa FlowerPower (KRISHNAMURTHY; BROWN; SJÖLANDER, 2007), que realizará uma busca por sequências homólogas no banco de dados UniRef100 com base na similaridade entre os perfis PSSM (*Position-Specific Scoring Matrix*) identificados de modo iterativo usando PSI-BLAST, para agrupar e classificar as sequências proteicas identificadas em subfamílias, de acordo com a similaridade, por fim, o programa gera HMMs para cada subfamília.

Além disso, realizamos uma etapa de calibração dos HMMs para a obtenção de *E-values* mais apropriados (WISTRAND; SONNHAMMER, 2005). A calibração envolve um ajuste dos parâmetros da distribuição de valores de escores extremos

(*extreme value distribution* - EVD) a partir da pontuação de sequências geradas de forma aleatória, esses parâmetros são utilizados para o ajuste dos *E-values* (*Expect value*), ou seja, do valor que descreve o número de correspondências esperadas ocorrerem ao acaso quando realizadas buscas em bancos de dados de tamanhos particulares (100000) considerando um determinado tipo de alinhamento (global).

Com essa finalidade, utilizamos o programa *hmmscore*, da ferramenta SAM-T2K (KARPLUS et al., 1997), para a calibração dos modelos com os parâmetros: -*calibrate 100000 -sw 0*). Após a etapa do FlowerPower as sequências referentes a cada família e subfamília foram utilizadas para a criação de árvores *Neighbor joining* (NJ) com base na similaridade das sequências em contraste com cada uma das subfamílias de proteínas. Depois a representação esquemática e ajuste na descrição das árvores foram geradas através do programa FigTree¹ v1.4.2.

A fonte da anotação taxonômica das famílias de proteínas dHPAs identificadas pelo sfFinder foi o banco de dados UniProt (UniRef100) (SUZEK et al., 2007), considerando o resultado do script *addTreeInfo.pl*, desenvolvido neste trabalho, e descrito na seção 4.2.5.

4.2 ANÁLISES DE METAGENOMA ASSOCIADO À dHPAs

4.2.1 OBTENÇÕES DOS DADOS METAGENÔMICOS

Os dados das três bibliotecas metagenômicas que são do tipo *paired-end* (2x100 pb) foram obtidos do repositório público Sequence Read Archive (SRA) (LEINONEN; SUGAWARA; SHUMWAY, 2011). Sendo que os conjuntos de dados são provenientes do sequenciamento de amostras de sedimento coletadas da região de Rock Bay (Victoria Harbour, British Columbia, Canadá).

Essas amostras correspondem ao estudo de metagenoma de uma área industrial próxima a uma zona costeira, que sofreu contaminação de seu meio ambiente por produtos derivados de petróleo, que geram várias moléculas nocivas ao ecossistema, entre elas os HPAs, além disso, essas amostras passaram por uma etapa de enriquecimento em um biorreator, antes de serem sequenciadas (MEIER; PATERSON; LAMBERT, 2016).

¹ <http://tree.bio.ed.ac.uk/software/figtree/>

Os grupos de sequências utilizadas neste trabalho, são do tipo *paired-end* (2x100 pb) e foram geradas por um sequenciador de nova geração, no caso foi utilizada a plataforma Illumina-HiSeq 2000 e as 3 corridas resultantes desse sequenciamento foram disponibilizadas no repositório SRA, pelos seguintes números de acesso identificadores: SRR2901746, SRR2902415 e SRR2903329.

4.2.2 PRÉ-PROCESSAMENTO DOS DADOS

Os arquivos referentes às bibliotecas recuperadas do SRA foram convertidos para o formato fastq utilizando a ferramenta fastq-dump (SRA-Toolkit) (versão 2.8.1) (HOWISON, 2013). Posteriormente a qualidade das *reads* foram avaliadas com o programa FastQC (versão 0.11.5) (BROWN; PIRRUNG; MCCUE, 2017) e em seguida foram processadas com o programa Atropos (versão 1.1.19) (DIDION; MARTIN; COLLINS, 2017), que realizou a remoção de adaptadores que representam artefatos e devem ser removidos para não interferirem nas próximas etapas de análise. Nesse caso, as *reads* foram processadas em duas etapas pelo Atropos, a primeira começando por uma abordagem baseada no algoritmo de correspondência entre as *reads* do inserto, posteriormente foi usada uma abordagem para detecção de correspondência com adaptadores para as sequências que não foram devidamente removidas na primeira etapa. Em seguida, as *reads* que foram filtradas passaram pelo programa PRINSEQ-lite v0.20.4 (SCHMIEDER; EDWARDS, 2011), responsável por fazer um controle de qualidade por meio da detecção e filtragem de regiões de baixa qualidade e complexidade nas *reads*, sendo que depois a qualidade foi novamente avaliada pelo programa FastQC após o processo de filtragem.

4.2.3 MONTAGEM DE NOVO DO METAGENOMA E PREDIÇÃO GÊNICA

Em seguida os conjuntos de *reads* do metagenoma de Rock Bay passaram pela etapa de montagem, utilizando para isso as ferramentas MEGAHIT (LI et al., 2015), IDBA_ud (PENG et al., 2012) e MetaSPAdes (NURK et al., 2017). Os *contigs* gerados foram avaliados segundo um índice de alinhamento de *reads* através do programa Bowtie2 (LANGMEAD; SALZBERG, 2012). A partir desse alinhamento

os *contigs* que obtiveram um melhor suporte de *reads* foram combinados em uma montagem única usando o programa MeGAMerge (v.1.1). Esse programa faz uma fusão dos resultados de diferentes montadores com o objetivo de gerar *contigs* de melhor qualidade, no caso ele fez a combinação para os resultados do MEGAHIT e do MetaSPAdes (SCHOLZ; LO; CHAIN, 2014), ou seja, os que tiveram melhores indicadores de qualidade de montagem.

Os índices de qualidade para os resultados de todas as montagens foram avaliados a partir da análise como programa MetaQuast (v4.4) (MIKHEENKO; SAVELIEV; GUREVICH, 2016), o qual revela os números, tamanhos e correspondências dos *contigs* gerados por cada montador e exibe um perfil de qualidade da montagem.

A partir da montagem consenso (referência) foram preditos os genes codificadores de proteínas incluídos nos *contigs* usando um *pipeline* de anotação para genomas procarióticos, o programa Prokka (v1.11) (SEEMANN, 2014). A execução com o Prokka foi realizada com o parâmetro '--meta' para predição em metagenomas. Esse programa identificou as *Open Reading Frames* (ORFs) correspondentes aos genes preditos. As predições foram selecionadas considerando um tamanho mínimo de 86 aminoácidos, a fim de evitar um número excessivo de possíveis falsas predições.

Assim como para o caso das proteínas dHPAs foi feita uma análise taxonômica, porém, neste caso baseada nas sequências dos *contigs* utilizando o programa Kraken (v.2.0.7-beta) (ONDOV; BERGMAN; PHILLIPPY, 2011). Este programa é responsável por fazer a identificação taxonômica baseada em *k-mers* extraídos de um banco de dados de genomas. Nós personalizamos esse banco, compreendendo as sequências do NCBI-nt, além das sequências de genomas completos e incompletos de bactérias, arqueias e vírus.

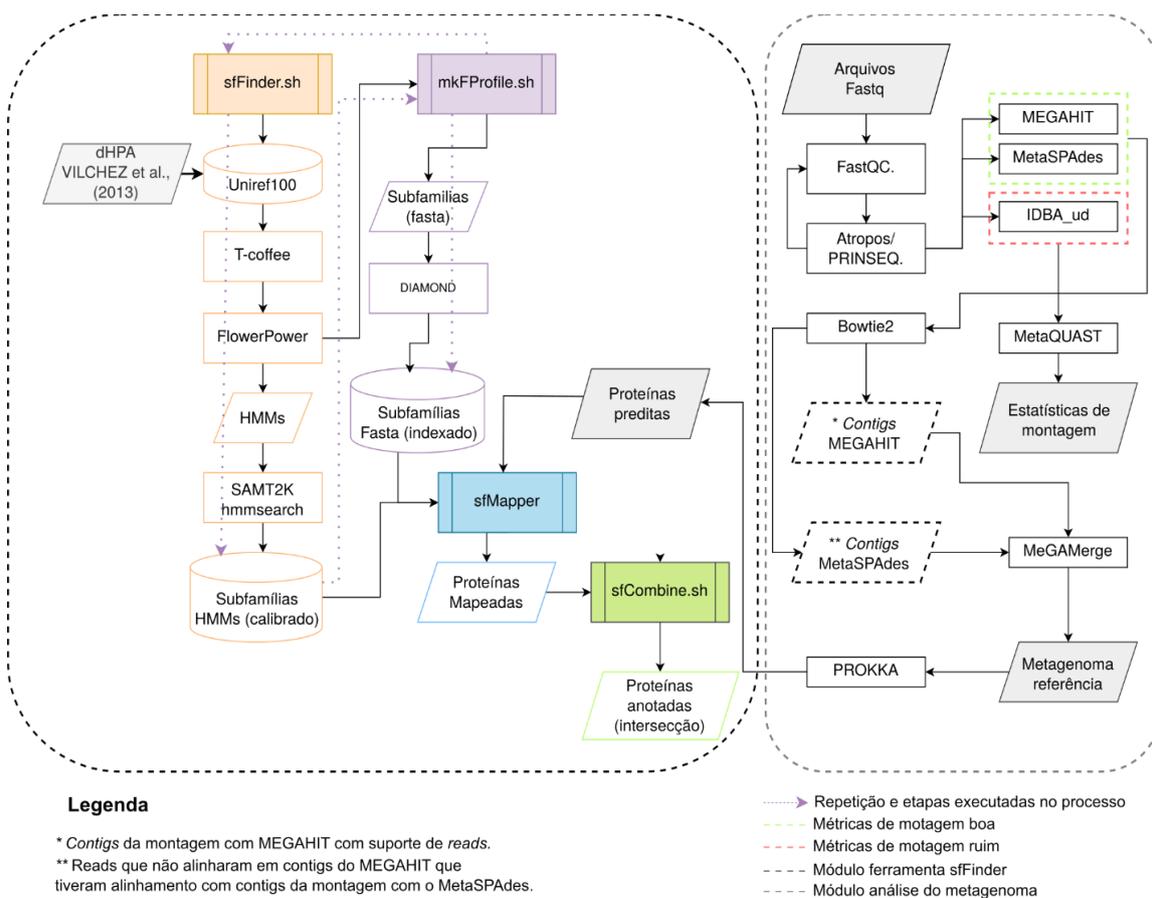
4.2.4 PROSPECÇÃO DAS PROTEÍNAS dHPAS

Foram consideradas duas estratégias para a prospecção de novas proteínas no metagenoma. Uma baseada no alinhamento das proteínas preditas no

metagenoma pelo programa Prokka, contra todas as sequências de subfamílias identificadas na etapa anterior pelo sfFinder, utilizando para isso o programa de alinhamento DIAMOND (BUCHFINK; XIE; HUSON, 2015) no modo blastp com os demais parâmetros: (*E-Value*) `--evaluate 10-5` (porcentagem de identidade mínima para reportar um alinhamento) `--id 30` (porcentagem de cobertura mínima da sequência interrogada) `--query-cover 90` e (modo de busca mais sensível) `--more-sensitive`. A outra estratégia é utilizando uma correspondência com os HMMs obtidos com o FlowerPower e previamente calibrados. Essa estratégia utiliza o programa *hmmsearch* da ferramenta SAM-T2K (KARPLUS et al., 1997).

O fluxograma das etapas de processamento dos dados pode ser visualizado pela **Figura 3**.

Figura 3: Fluxograma ilustrando as principais etapas utilizadas na metodologia de processamentos de dados.



Fonte: Próprio Autor

4.2.5 AUTOMATIZAÇÃO DOS PROCESSOS DE ANÁLISE

A fim de automatizar o processo de busca por proteínas compatíveis com uma determinada família até o reconhecimento das subfamílias e a construção dos modelos estatísticos, foram desenvolvidos scripts e disponibilizados no repositório GitHub com o nome de **sfFinder**². A ferramenta é dividida em módulos implementados nos seguintes Perl ou Shell *scripts*:

getTreeInfo.pl: Perl *script* auxiliar que acrescenta informação taxonômica ao resultado do FlowerPower a partir do identificador NCBI Taxon ID (*tax_id*) contido na descrição de cada sequência do UniRef100. Como parâmetros da linha de comando recebe-os na seguinte ordem: (1) o arquivo resultante da execução do FlowerPower para cada família (*last.tree*) no formato Newick; (2) o arquivo texto com os identificadores do banco de dados UniRef100 e as respectivas descrições; (3) o arquivo resultante da anotação que irá conter o ID UniRef100 da proteína, a descrição, a subfamília identificada pelo FlowerPower, a descrição taxonômica e o NCBI Taxon ID; (4) outro arquivo resultante da análise com FlowerPower (*last.subfam*) contendo o relacionamento de cada sequência com a subfamília; (5) o arquivo *fasta* com as sequências correspondentes à família considerada no alinhamento múltiplo que foi submetido como entrada ao FlowerPower (servirá para indicar na descrição com um asterisco (“*”) quais proteínas fizeram parte do alinhamento múltiplo na entrada do FlowerPower); (6) o arquivo *fasta* com a sequência reconhecida como semente (*seed*) no FlowerPower (servirá para indicar na descrição com um asterisco extra (“**”) qual sequência foi selecionada como *seed*). Este script faz parte de outro repositório GitHub, o *bioinfoutilities*³.

addTreeInfo.pl: Perl *script* auxiliar que acrescenta informação taxonômica às sequências proteicas consultadas (*queries*) pelo **sfMapper.sh**. Essa informação taxonômica dependerá da identificação de uma família/subfamília, no caso de haver mais de um táxon nesta subfamília, a informação taxonômica será obtida a partir do programa **GetLCA.pl**, da ferramenta Krona (v 2.7.1) (ONDOV; BERGMAN;

² <https://github.com/bioinfo-fcav/sfFinder>

³ <https://github.com/dgpinheiro/bioinfoutilities>

PHILLIPPY, 2011), que irá obter o menor ancestral comum (*Lowest Common Ancestor*, ou LCA) entre essas *taxa*.

getProteinByID.pl: Perl *script* auxiliar que recebe como entrada um arquivo contendo uma lista de registros, contendo nome da família e o ID das proteínas e um destino para os arquivos que serão gerados com as sequências no formato fasta correspondentes a cada família. Este *script* foi utilizado para recuperar as sequências no formato fasta das 1.302 proteínas (VILCHEZ-VARGAS et al., 2013);

sfFinder.sh: Shell *script* para automatizar o processo de encontrar as sequências similares, possivelmente homólogas, além da classificação delas em subfamílias, e, finalmente, obtenção dos respectivos modelos estatísticos calibrados para o banco de dados UniRef100. Dessa forma, o sfFinder.sh automatiza a recuperação inicial das proteínas de uma família no banco de dados UniRef100 utilizando DIAMOND blastp, obtém um alinhamento múltiplo das sequências dessa família com o T-Coffee, e executa a análise com o FlowerPower para a obtenção das sequências classificadas em subfamílias e os HMMs, além da calibração dos modelos. Este *script* recebe como entrada o nome de uma família de proteínas, o arquivo com as sequências no formato fasta correspondente à essa família, e, finalmente, um diretório onde serão gerados os resultados;

mkFProfiles.sh: Shell *script* que executa automaticamente o **getProteinByID.pl** para obtenção das sequências das famílias e, em seguida, já executa o *script* **sfFinder.sh** para cada família, gerando os perfis de cada subfamília no diretório especificado como o segundo parâmetro da linha de comando. Ao final do processo, invoca a execução do *script* **getTreeInfo.pl** a fim de acrescentar as informações taxonômicas;

sfMapper.sh: Shell *script* que realiza a classificação de sequências proteicas (e.g. sequências proteicas obtidas a partir de predições gênicas em metagenomas) em família e subfamília considerando as sequências e os modelos gerados com o **sfFinder.sh**. Com essa ferramenta é possível realizar a classificação utilizando alinhamento de sequências (**diamond**) ou correspondências com os modelos estatísticos, os HMMs (**hmmscore**). O *script* recebe como entrada o modo de busca (**diamond/hmmscore**), o arquivo contendo as sequências proteicas desconhecidas, o diretório contendo os resultados do **mkFProfiles.sh** (ou seja, das

execuções do **sfFinder.sh** para cada família), um diretório para armazenamento dos resultados da prospecção nesse conjunto de proteínas interrogadas.

sfCombine.sh: Shell *script* que permite combinar os resultados das duas abordagens propostas, ou seja, utilizando alinhamento de sequências (**diamond**) ou correspondência com os HMMs (**hmmscore**).

Os módulos do **sfFinder** podem ser utilizados também para análises considerando outras famílias de proteínas que não sejam necessariamente relacionadas às dHPAs.

5. RESULTADOS E DISCUSSÃO

5.1 PROTEÍNAS dHPAS

5.1.1 IDENTIFICAÇÃO E CLASSIFICAÇÃO DAS PROTEÍNAS dHPAS

As sequências de proteínas previamente descritas como dPHAs caracterizadas por Vilchez et. al, (2013) contabilizam um total de 1302 sequências, distribuídas em 12 famílias. No entanto, após o processo inicial de busca de sequências similares no banco de dados Uniref100, sequências que não eram provenientes de bactérias e arqueias foram desconsideradas da contagem, assim como uma sequência pertencente à família FERRE, proveniente de bactéria, foi descartada pelo fato de que seu tamanho ser superior às das sequências de referências. A distribuição de sequências por famílias está descrita na **Tabela 2**.

A busca exaustiva realizada pelas diversas iterações do programa sfFinder para recuperação de sequências homólogas, permitiu identificar um total de 10.489 sequências relacionadas às 12 categorias de famílias de dHPAs.

Este total de sequências recuperadas, apresentaram um valor que é 8,08 vezes superior em relação ao que foi descrito anteriormente, o que reflete o poder de análise dessa aplicação utilizando o sfFinder para expandir as buscas por sequências de proteínas similares potencialmente ortólogas e que possuem os mesmos domínios. Um total de 606 subdivisões (subfamílias) foram identificadas a partir das 12 famílias, descritas na **Tabela 2**.

Tabela 2: Descrição do número de sequências de proteínas identificadas para cada uma das famílias e distribuição das subfamílias realizadas pelo programa sfFinder.

| Famílias | Nº inicial de proteínas | Nº de proteínas após a busca pelo sfFinder | Nº de subfamílias identificadas pelo sfFinder | Porcentagem de ganho |
|-----------------|--------------------------------|---|--|-----------------------------|
| ALKB | 246 | 1.204 | 79 | 389% |
| BCOAR | 84 | 218 | 9 | 159% |
| BSSA | 17 | 263 | 13 | 1447% |
| CYP153 | 66 | 513 | 29 | 677% |
| EXDO | 224 | 1.009 | 70 | 350% |
| FERRE | 39 | 746 | 36 | 1812% |
| INDO | 235 | 1.635 | 119 | 595% |
| MACR | 50 | 605 | 66 | 1110% |
| MCIS | 55 | 878 | 21 | 1496% |
| RHDO | 238 | 1.881 | 85 | 690% |
| RHMO | 43 | 700 | 16 | 1527% |

| | | | | |
|-------|-------|--------|-----|--------|
| TDLH | 1 | 837 | 63 | 83600% |
| TOTAL | 1.298 | 10.489 | 606 | 708% |

Fonte: Próprio Autor

O teste realizado considerando o alinhamento das sequências de dHPAs obtidas no trabalho de Vilchez com as sequências de subfamílias recuperadas, com base no método usando o alinhador DIAMOND, foi possível recuperar todas as sequências (1.298) do Vilchez e no caso do método usando os modelos HMMs foi possível recuperar 1.295 sequências devido a um corte que foi limitante para recuperar as sequências de HMMs, dessa maneira 3 sequências foram filtradas representando 0,0023% do total.

5.1.2 ÁRVORES DE SIMILARIDADE

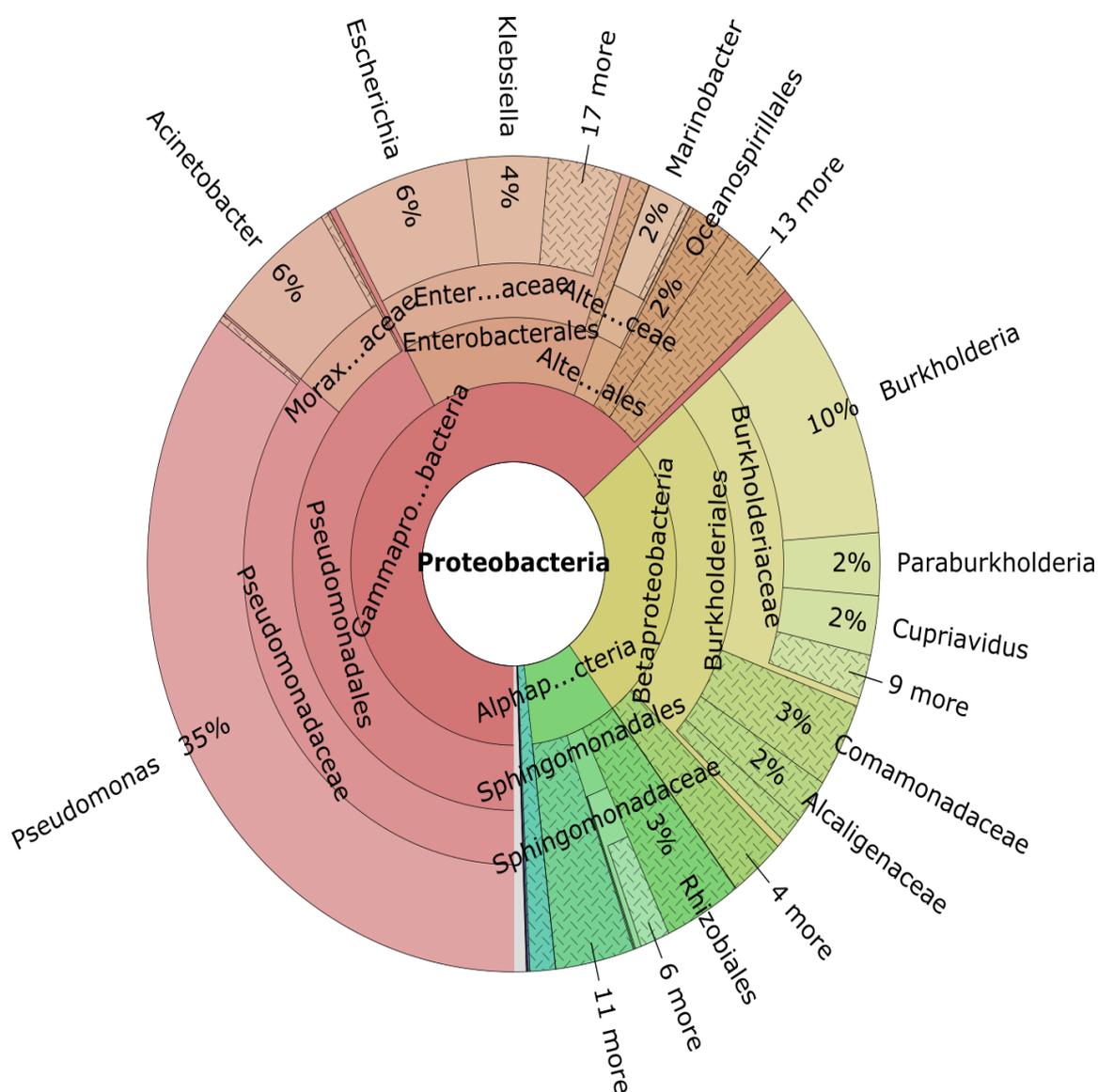
As árvores de distância *Neighbor joining* (NJ) com base na similaridade das sequências em contraste com cada uma das subfamílias geradas pelo FlowerPower foram disponibilizadas a um repositório de dados público o Zenodo (PETERS et al., 2017), onde a visualização das relações de similaridade das diferentes classes de subfamílias de proteínas relacionadas a dHPAs para cada uma das 12 famílias está no link: <https://doi.org/10.5281/zenodo.4299043>.

5.1.3 ANÁLISE TAXONÔMICA DAS PROTEÍNAS DO UNIREF

Em relação à análise taxonômica realizada pelo Kraken das sequências de proteínas que foram identificadas pelo FlowerPower, 99,6% das sequências estão classificadas dentro do domínio *Bacteria*, 0,02% correspondem ao domínio *Archaea* e 0,3% das sequências não tiveram uma classificação taxonômica encontrada. Considerando as sequências relacionadas ao domínio *Bacteria*, 75% das sequências pertencem ao filo *Proteobacteria*, 19% das sequências pertencem ao filo *Terrabacteria* e o restante não apresentaram assinatura taxonômica em nível de filo. (**Figura 4**)

Dentro do filo *Proteobacteria* foram identificadas 64% de abundância para a classe *Gammaproteobacteria*, 26% de abundância para a classe *Betaproteobacteria*, 9% de abundância para a classe *Alphaproteobacteria* e 1% para a classe *Deltaproteobacteria*. Assim, com base no panorama das sequências homólogas às famílias de proteínas dHPAs, o filo *Proteobacteria*, é um dos principais filos envolvido na degradação desse grupo de xenobióticos.

Figura 4: Identificação taxonômica das proteínas dHPAs a partir das subfamílias de proteínas identificadas pelo sfFinder



Fonte: Próprio Autor

Dentro da principal classe em termos de maior abundância de táxons a classe *Gammaproteobacteria* possui destaque para as ordens: *Pseudomonadales* com abundância de 66% dentro da classe, a ordem *Enterobacterales* com 22%, a ordem *Alteromonadales* com 3% e a ordem *Oceanospirillales* com 3%.

Considerando-se as principais espécies encontradas dentro da classe *Gammaproteobacteria*, se têm destaque os táxons: *Pseudomonas aeruginosa* com 20% de abundância, *Pseudomonas putida* com 3%, *Pseudomonas fluorescens* com 2%, *Pseudomonas stutzeri* com 2% e *Acinetobacter baumannii* com 2%. Essa distribuição pode ser visualizada na **Figura 5**.

A espécie *Pseudomonas aeruginosa* (**Figura 5**) é descrita no trabalho de Mahto; Das (2020) como sendo uma bactéria Gram Negativa em forma de bastonete responsável pela degradação de uma grande gama de xenobióticos entre eles várias moléculas de HPAs. No trabalho foi feito o isolamento da bactéria em uma área contaminada por HPAs na Costa de Odisha, na Índia. Nesse trabalho mencionado, após o isolamento da bactéria foi feito o sequenciamento do micro-organismo, seguido pela montagem *Denovo* e anotação funcional. Pelas análises de anotação foi observada a presença de 145 genes envolvidos na biodegradação e metabolismo dos xenobióticos em várias vias de meta-clivagem, entre os principais HPAs tóxicos envolvidos na degradação das enzimas dessa bactéria estão: Fenantreno, Naftaleno, Benzoato, Aminobenzoato, Fluorbenzoato, Tolueno, Xileno e Estireno (MAHTO; DAS, 2020).

Em relação a espécie *Acinetobacter baumannii* (**Figura 5**), esta não possui um histórico abrangente de estudos na comprovação da grande eficácia na dHPAs, diferentemente para espécies pertencentes aos seguintes gêneros: *Pseudomonas*, *Rhodococcus*, *Mycobacterium*, *Cycloclasticus*, *Bacillus* e *Sphingomonas* que têm um longo histórico de artigos demonstrando resultados satisfatórios em relação a esses micro-organismos na biorremediação de áreas contaminadas por HPAs. No entanto, em um recente estudo realizado por Gupta et al. (2020), constatou que a espécie *Acinetobacter baumannii*, têm a capacidade de degradação para concentrações intermediárias de Pireno entre outros HPAs variando em sua maioria de concentrações baixas a intermediárias e em alguns casos altas quando o composto aromático era a única fonte de carbono no ambiente, dessa forma

Em seguida os resultados provenientes das classificações taxonômicas foram posteriormente disponibilizados no repositório de banco de dados do Zenodo pelo link: <https://doi.org/10.5281/zenodo.4484298>.

5.2 METAGENOMA DE ROCK BAY

5.2.1 PRÉ-PROCESSAMENTO

A partir de um total inicial de 625.509.142 pares de *reads* (*paired-end reads*), (em média 208 milhões de pares de *reads* por biblioteca distribuídos entre as três bibliotecas: SRR2901746, SRR2902415 e SRR2903329 do tipo *paired-end* (2x100 pb) restaram 611.521.525 pares de sequências após o procedimento de remoção de sequências referentes a adaptadores ou de baixa qualidade. Essa etapa é importante, pois posteriormente na montagem do genoma, a inclusão de *reads* de baixa qualidade levará à geração de falsos *k-mers* o que, por sua vez, aumenta a complexidade do processo de montagem subsequente. Dessa forma, podemos aumentar a qualidade e a confiabilidade nos resultados, com ganhos simultâneos em termos de tempo de execução e recursos computacionais necessários (FABBRO et al., 2013). Os resultados obtidos nas três bibliotecas para cada etapa do pré-processamento estão descritos na **Tabela 3**.

Tabela 3: Resultados das etapas de pré-processamento das *reads* de cada biblioteca. Os números nesta tabela correspondem aos pares de *reads* de cada clone de fragmentos oriundos do sequenciamento.

| Pré-processamento | | SRR2901746 | SRR2902415 | SRR2903329 |
|-------------------|--------------|-------------|-------------|-------------|
| | brutas | 203.397.410 | 204.247.656 | 217.864.076 |
| Atropos | retidas | 2.694.063 | 2.874.288 | 8.279.200 |
| | selecionadas | 200.703.347 | 201.373.368 | 209.584.876 |

| | | | | |
|---------|--------------|-------------|-------------|-------------|
| PrinSeq | retidas | 34.182 | 33.036 | 71.555 |
| | selecionadas | 200.668.820 | 201.340.001 | 209.512.704 |

Fonte: Próprio Autor

5.2.2 ÍNDICES DE MONTAGEM E PREDIÇÃO GÊNICA

Avaliação de montagens consiste em uma etapa crucial e é um desafio sem o uso de referências (VAN DER WALT et al., 2017). O índice de mapeamento de *reads* na montagem é comumente usado para se avaliar a qualidade da montagem.

Considerando o critério de mapeamento de *reads*, o programa MEGAHIT teve o melhor desempenho (**Tabela 4**). Com base nos índices de alinhamento das sequências, a montagem realizada pelo programa MEGAHIT foi aquela que obteve o maior número de fragmentos alinhados, seguidos pelo, IDBA_ud e MetaSPAdes.

Em relação ao critério de porcentagem de fragmentos alinhados entre os montadores, o MetaSPAdes foi aquele que obteve os maiores valores seguidos do MEGAHIT e do IDBA_ud, além disso em relação ao número de fragmentos alinhados de forma única, o MetaSPAdes foi o montador com maior destaque seguidos pelo MEGAHIT e IDBA_ud.

Deste modo, todos os *contigs* que tiveram alinhamento com o MEGAHIT e as *reads* que não tiveram alinhamento foram usados para recuperar *contigs* da montagem com o MetaSPAdes. Ambos os conjuntos de *contigs* com suporte de alinhamento, foram usados para gerar uma montagem referência com a ferramenta MeGAMerge que combinou os resultados do MEGAHIT e MetaSPAdes em uma única montagem.

Tabela 4: Índices das sequências mapeadas pelos diferentes montadores referentes ao metagenoma do solo contaminado por HPAs.

| Índices de alinhamento | MEGAHIT | MetaSPAdes | IDBA_ud | MeGAMerge |
|-------------------------------------|-------------|-------------|-------------|-------------|
| Total de fragmentos alinhados | 606.174.466 | 604.740.175 | 605.198.160 | 605.339.143 |
| Total de fragmentos pareados | 606.047.539 | 604.615.834 | 605.072.298 | 605.213.259 |
| Alinhamentos únicos | 563.583.800 | 574.624.153 | 536.979.688 | 577.125.387 |
| Fragmentos <i>single-end</i> | 126.927 | 124.917 | 125.862 | 125.884 |
| Porcentagem de fragmentos alinhados | 92,97% | 95,02% | 88,73% | 95,34% |

Fonte: Próprio Autor

A utilização da ferramenta MetaQUAST, possibilitou a comparação entre cada um dos montadores, avaliando-se o desempenho geral de cada um, considerando diversas métricas além dos valores de alinhamento, como por exemplo: valores de N50, N75, L50, L75, número total de *contigs*, maior *contig*, número de *contigs* de acordo com a escala de tamanho, número de *mismatches*, todos descritos na **Tabela 5**.

Os parâmetros N50 e N75 são relacionados ao tamanho do *contig* na porção de 50% e 75% da montagem respectivamente, eles são utilizados para avaliar se o montador gerou muitos *contigs* de tamanhos grandes. Valores elevados mostram que o montador utilizado obteve um desempenho superior na construção dos *contigs* com o objetivo de obter sequências completas das moléculas originais. Nesse sentido, os montadores que obtiveram melhor desempenho em ordem decrescente foram: MetaSPAdes, MeGAMerge, MEGAHIT e IDBA_ud.

Já os parâmetros L50 e L75 estão relacionados ao número de *contigs* gerados pelos montadores ao longo de 50% e 75% da montagem, assim quanto

menor esse valor, menor o número de *contigs* pequenos gerados pelo montador, demonstrando assim um bom resultado. Os valores N e L, portanto, permitem uma avaliação quanto ao nível de fragmentação da montagem, nesse sentido os montadores que obtiveram melhor desempenho em ordem decrescente foram: MetaSPAdes, MEGAHIT e IDBA_ud.

Em relação aos valores de maiores *contigs* o MetaSPAdes foi o montador individual que teve o maior tamanho para o *contig*, seguido pela MEGAHIT e IDBA_ud e no quesito de comprimento total dos *contigs* o IDBA_ud foi o montador que obteve a melhor performance seguido do MetaSPAdes e IDBA_ud.

De acordo, os resultados de a avaliação e comparação entre as montagens realizada pelo programa MetaQUAST, ficou claro que a fusão das montagens do MEGAHIT e MetaSPAdes por meio do MeGAMerge, apresentou ganhos em relação ao tamanho do maior *contig* e no comprimento total do *contig* (**Tabela 5**), fator importante para aferição da qualidade da montagem do metagenoma.

Tabela 5: Resultados obtidos entre os programas de montadores genômicos que foram comparados no MetaQUAST

| Referências estatísticas | Megamerge | Idba_ud | Megahit | Metaspades |
|------------------------------|-----------|-----------|-----------|------------|
| <i>contigs</i> | 424.253 | 462.019 | 393.120 | 323.254 |
| <i>contigs</i> (>= 0 bp) | 1.039.549 | 1.347.371 | 1.965.126 | 661.186 |
| <i>contigs</i> (>= 1000 bp) | 199.575 | 203.672 | 177.955 | 157.705 |
| <i>contigs</i> (>= 5000 bp) | 32.749 | 26.858 | 26.056 | 30.799 |
| <i>contigs</i> (>= 10000 bp) | 13.345 | 10.602 | 9.894 | 13.549 |
| <i>contigs</i> (>= 25000 bp) | 3.400 | 2.742 | 2.259 | 3.754 |

| | | | | |
|--|---------------|---------------|---------------|-------------|
| <i>contigs</i> (\geq 50000 bp) | 971 | 794 | 572 | 1.060 |
| Maior <i>contig</i> | 608.596 | 515.762 | 515.769 | 536.579 |
| Total <i>length</i> | 937.501.575 | 881.168.184 | 776.015.982 | 821.313.067 |
| Total <i>length</i> (\geq 0 bp) | 1.130.562.421 | 1.134.546.364 | 1.135.929.994 | 936.345.085 |
| Total <i>length</i> (\geq 1000 bp) | 782.648.711 | 703.159.941 | 627.909.400 | 706.507.641 |
| Total <i>length</i> (\geq 5000 bp) | 447.320.942 | 362.726.614 | 328.301.210 | 450.679.875 |
| Total <i>length</i> (\geq 10000 bp) | 313.515.212 | 250.958.370 | 217.217.573 | 330.750.894 |
| Total <i>length</i> (\geq 25000 bp) | 164.798.834 | 133.909.935 | 103.707.334 | 183.054.645 |
| Total <i>length</i> (\geq 50000 bp) | 82.602.592 | 67.225.905 | 47.055.963 | 91.169.902 |
| N50 | 4.493 | 3.264 | 3.554 | 6.315 |
| N75 | 1.467 | 1.190 | 1.260 | 1.775 |
| L50 | 37.278 | 46.364 | 40.301 | 23.666 |
| L75 | 133.485 | 164.858 | 136.949 | 88.721 |
| N's <i>Mismatches</i> | 404.170 | 70.904 | 0 | 392.818 |
| N's per 100 kbp | 43.11 | 8.05 | 0 | 47.83 |

Fonte: Próprio Autor

A de predição gênica realizada pelo Prokka a partir da montagem referência obtida pelo MeGAMerge, apresentou um total de 1.053.057 ORFs, sendo que destas 554.271 ORFs foram classificadas como hipotéticas ou desconhecidas. Após a etapa de filtragem com base no tamanho de 86 aminoácidos, foram retidas 816.315 ORFs dentre as quais 498.786 apresentaram uma função conhecida com base nos bancos de dados pré-configurados do programa Prokka.

5.2.3 ANÁLISE TAXONÔMICA DO METAGENOMA

A identificação taxonômica dos *contigs* no metagenoma de Rock Bay, realizada pelo programa Kraken, obteve a classificação de um total de 89% dos *contigs* para o domínio *Bacteria*, em relação aos domínios *Eukaryota* e *Archaea*, estes tiveram uma atribuição taxonômica de apenas 6% e 0,07% dos *contigs* do metagenoma, respectivamente, sendo que 2% dos *contigs* não tiveram correspondência na classificação.

Em relação a camada taxonômica mais abaixo, se têm destaque o filo *Proteobacteria* que foi o mais abundante, apresentando 81% dos *contigs* classificados para o seu domínio. Também fica em evidencia o grupo das *Terrabacteria* apresentando 14% dos *contigs* classificados para o domínio, além disso, 2% dos *contigs* não apresentou classificação para o domínio e os outros 3% dos *contigs* ficaram distribuídos entre 19 pequenos grupos (**Figura 6**).

Dentro do filo *Proteobacteria* foram identificadas 51% de abundância para a classe *Gammaproteobacteria*, 28% de abundância para a classe *Betaproteobacteria*, 19% de abundância para a classe *Alphaproteobacteria*, 0,8% de abundância para a classe *Deltaproteobacteria* e 0,03% de abundância para a classe *Epsilonproteobacteria*.

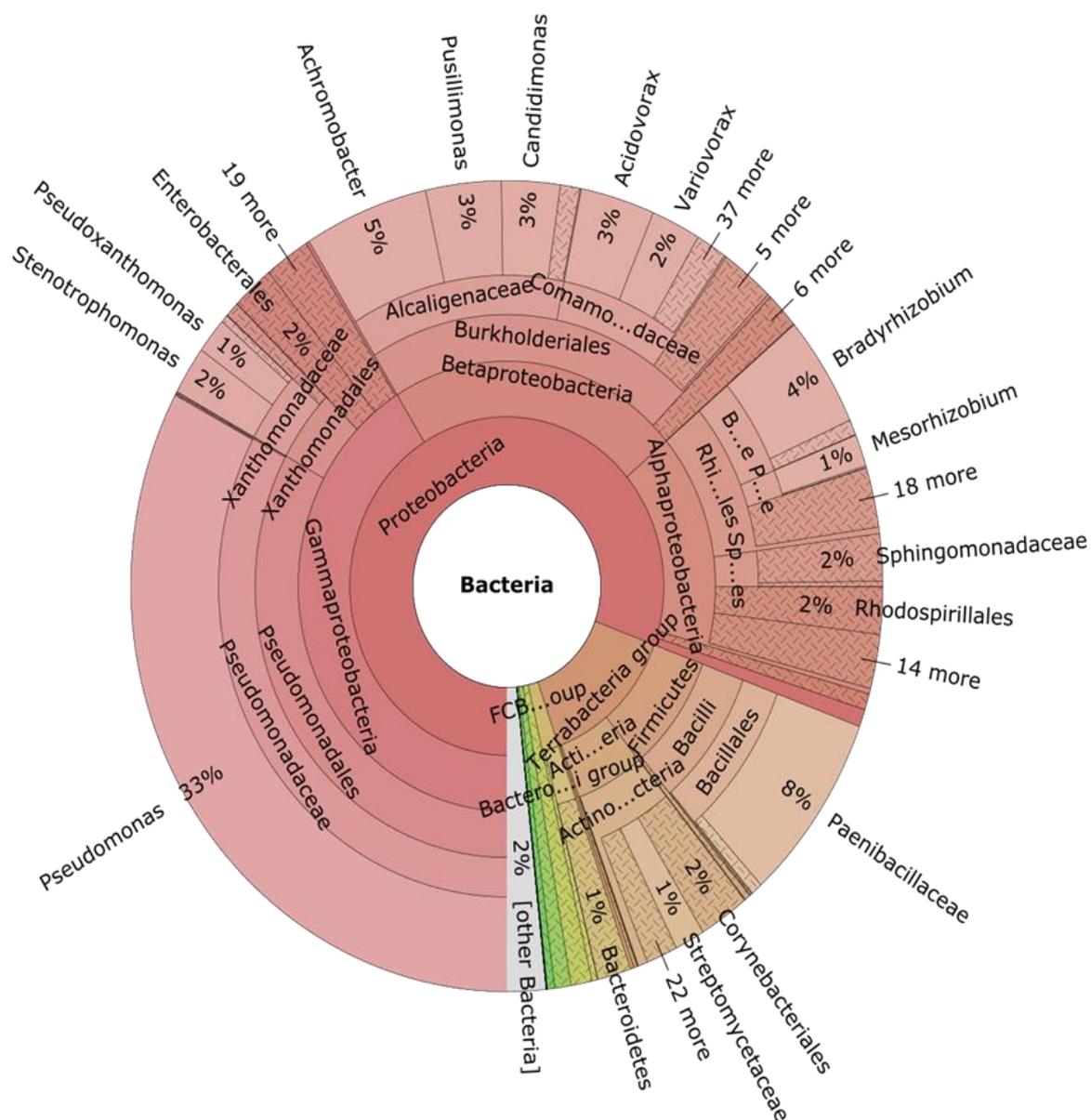
Considerando os resultados encontrados no trabalho de Tan et al. (2015), onde foram coletadas amostras do metagenoma de três diferentes tipos de solos (NAPDC, SCADC e TOLDC) que sofreram por contaminação de derivados do petróleo, entre eles HPAs. Foram identificados principalmente micro-organismos da classe *Deltaproteobacteria* com 23% de abundância em relação ao domínio *Bacteria*, tendo representado um grande destaque entre as demais 4 classes

encontradas. Dentre estas, as que tiveram maior representatividade em ordem decrescente foram: *Gammaproteobacteria*, *Alphaproteobacteria*, *Betaproteobacteria* e *Epsilonproteobacteria* (TAN et al., 2015).

No caso dos resultados obtidos pela classificação taxonômica com o Kraken a partir dos *contigs* do metagenoma de Rock Bay, foi identificado um valor de 0,6% de abundância para a classe *Deltaproteobacteria* em relação ao domínio *Bacteria*, valor pouco significativo. Sendo as *Deltaproteobacteria* são descritas na literatura como importantes integrantes da via de degradação do Benzoato (BERGMANN et al., 2011) que é uma molécula relativamente comum em uma parcela dos HPAs, além disso, a BCOAR (SCHMID; RENÉ; BOLL, 2015) e MACR (FUJII et al., 2016) são as principais famílias de proteínas que participam de sua via catalítica assim sendo de importante relevância para a biorremediação desses xenobióticos.

A distribuição dos táxons inferiores está representada na **Figura 6**.

Figura 6: Classificação e distribuição taxonômica dos *contigs* do metagenoma, anotados com o Kraken.



Fonte: Próprio Autor

No metagenoma de Rock Bay destaca-se a classe *Gammaproteobacteria*, onde os táxons que tiveram maior abundância foram: a ordem *Pseudomonadales* que obteve cerca de 80%, sendo que dentro desta ordem os gêneros *Pseudomonas* teve grande representatividade, em relação a ordem *Xanthomonadales* esta teve 10% (**Figura 7**).

Em nível de táxons de espécie se tem destaque: *Pseudomonas fluorescens* com 5% de abundância dentro do Domínio, *Pseudomonas frederiksbergensis* com

3% de abundância, *Pseudomonas putida* com 1% de abundância, *Immundisolibacter cernigliae* com 1%, *Candidimonas bauzanensis* com 3%, *Pusillimonas* sp. JR1/69-3-13 com 2%, *Acidovorax carolinensis* com 1%, *Bradyrhizobium erythrophlei* com 1%, *Rhodospirillales bacterium* URHD0017 com 1% e *Paenibacillus ginsengarvi* com 1%.

Comparando-se os resultados taxonômicos obtidos no metagenoma com outro estudo de Mukherjee et al. (2017) onde foi realizada uma análise taxonômica a partir do metagenoma de 12 localidades diferentes que estavam contaminadas por HPAs e distribuídas geograficamente tanto no Oriente quanto no Ocidente, foi observado que em relação à parcela microbiana representada pelas bactérias, o filo *Proteobacteria* foi aquele que representou uma maior abundância para todas as amostras analisadas, com destaque também para as ordens *Xanthomonadales* e *Pseudomonadales* que tiveram uma abundância variando entre 15% a 25% em pelo menos 3 das 12 amostras avaliadas, demonstrando um valor relativamente significativo para esses grupos de bactérias que também pertencem a classe *Gammaproteobacteria*, dando suporte para os resultados do metagenoma de Rock Bay, onde a Classe *Gammaproteobacteria* foi a mais abundante e a ordem *Pseudomonadales* teve uma alta representatividade (80%) dentro dela (MUKHERJEE et al., 2017).

Analisando também esses resultados com os dados de um trabalho de um experimento na Índia realizado por Shah et al. (2013), onde foi estudado o metagenoma de uma comunidade microbiana proveniente de solo contaminado por xenobióticos, dentre eles os HPAs e outros poluentes derivados de petróleo. O estudo obteve resultados similares na proporção de abundância para o filo *Proteobacteria*, principalmente a classe *Gammaproteobacteria* e a ordem *Pseudomonadales*. Sendo que este grupo de micro-organismos é especializado principalmente na via de degradação das moléculas de alcanos, um dos hidrocarbonetos de cadeia mais simples. Essas bactérias são detentoras principalmente das famílias de proteínas: ALKB (PICCOLO et al., 2011) e CYP153 (SAZYKIN et al., 2019), atuando na metabolização e degradação destes compostos (SHAH et al., 2013).

Em relação à classificação taxonômica no metagenoma de Rock Bay, a Ordem *Enterobacterales* teve cerca de 4% e a espécie *Immundisolibacter cernigliae*

teve um grande destaque em seu nível taxonômico tendo cerca de 3% de abundância dentro de *Gammaproteobacteria*. Sendo que a espécie *Immundisolibacter cernigliae*, em geral é descrita em alguns trabalhos pela capacidade de dHPAs pertencentes a via de metabolização do Pireno e entre eles está um dos HPAs mais prejudiciais ao meio ambiente o Benzo(a)pireno amplamente descrito na literatura e em trabalhos de Zeng et al. (2019), como um HPA altamente carcinogênico e mutagênico, dessa forma a presença da espécie de bactéria em questão é importante no ambiente para a metabolização desse perigoso contaminante (ZENG et al., 2019).

No metagenoma de Rock Bay a espécie que foi encontrada em maior abundância foi a *Pseudomonas fluorescens* (**Figura 7**), pertencente a classe *Gammaproteobacteria*, no trabalho de Sangkharak et al (2020) ela é descrita pela alta versatilidade na degradação de compostos aromáticos entre eles: Fenantreno, Pireno e Fluoranteno que foram os principais HPAs contaminantes no caso das amostras analisadas.

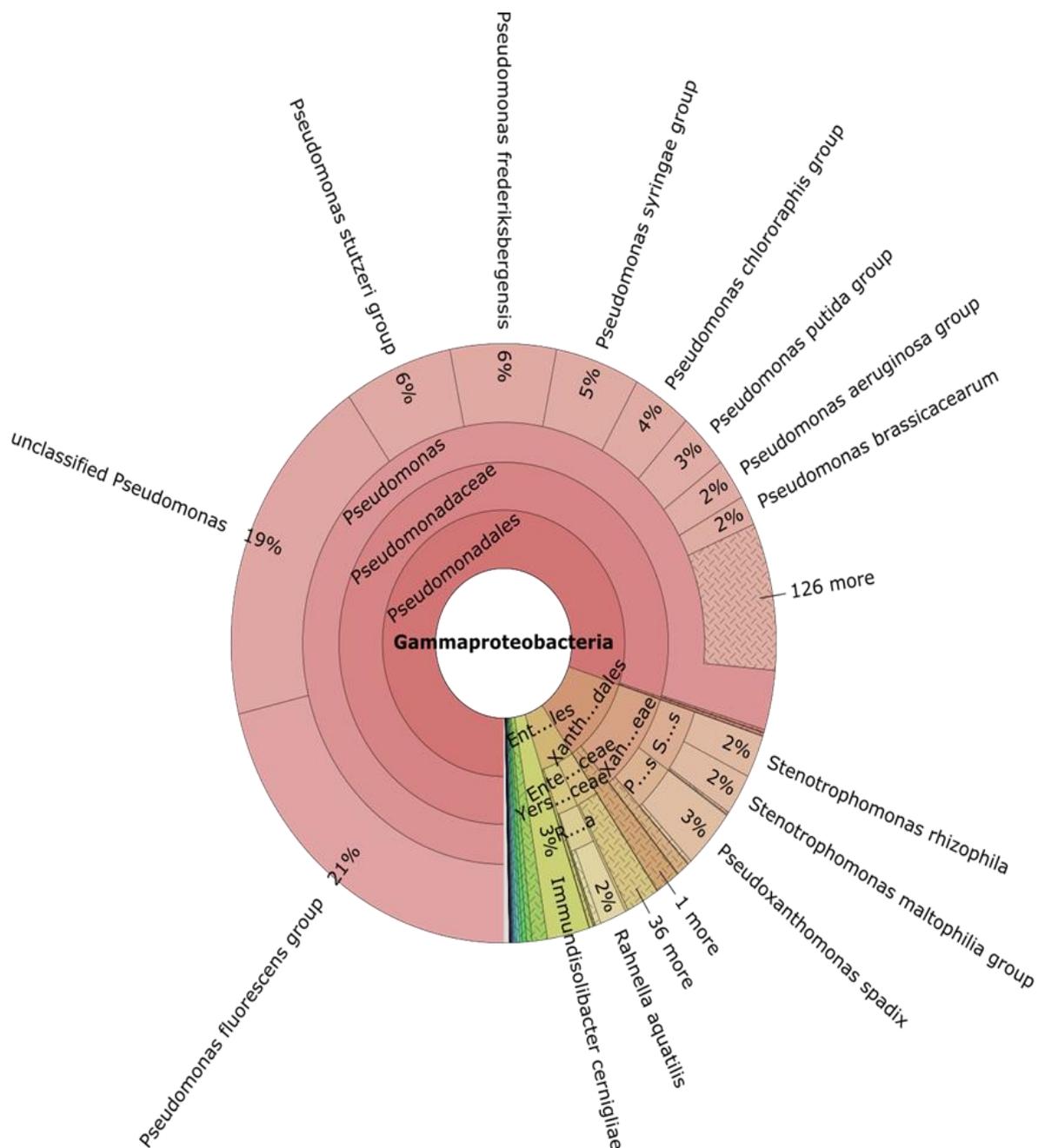
Nesse estudo de Sangkharak et al (2020) o gênero *Pseudomonas* obteve uma eficiência próxima de 100% na degradação dos 3 HPAs (Fenantreno, Pireno e Fluoranteno) mesmo quando esses estavam em alta concentração, demonstrando a incrível capacidade e eficiência dessas bactérias na biorremediação de compostos aromáticos. Além disso, foram observados ação enzimática de orto-clivagem e meta-clivagem na degradação desses compostos, sendo que tal método é empregado principalmente pelas famílias de proteínas INDO (XU et al., 2017) e EXDO (TERRÓN-GONZÁLEZ et al., 2016) respectivamente (SANGKHARAK et al., 2020).

Dessa maneira pode-se sugerir que a presença em grande número dessas bactérias descritas com a função dHPAs no meio ambiente contaminado podem sinalizar que comunidades com essa distribuição específica dos táxons de suas populações proporcionem uma maior adaptabilidade, resistência e capacidade de incorporação dos HPAs para sobreviverem àquela região com presença majoritária de Hidrocarbonetos Aromáticos entre outros xenobióticos. Provavelmente, possuem uma maquinaria enzimática que as permite retirar dessas moléculas carbono e outros subprodutos para sua manutenção celular e aquisição de energia, para dessa forma permanecerem em um ambiente desfavorável e se reproduzirem

nessas áreas, apesar da toxicidade dos compostos presentes nela (SMUŁEK et al., 2019).

A distribuição dos táxons inferiores a *Gammaproteobacteria* está representada na **Figura 7**.

Figura 7: Classificação taxonômica das proteínas dHPAs no metagenoma anotados com o Kraken a partir do nível de classe *Gammaproteobacteria*.



Fonte: Próprio Autor

Em seguida os resultados provenientes das classificações taxonômicas foram posteriormente disponibilizados no repositório de banco de dados do Zenodo pelo link: <https://doi.org/10.5281/zenodo.4484298>

5.3.1 PROSPECÇÃO DE NOVAS PROTEÍNAS dHPAS

A prospecção de novas proteínas de dHPAs no metagenoma utilizando a abordagem de alinhamento com o DIAMOND identificou 2.261 proteínas no metagenoma que corresponde a 21,55% do número de proteínas geradas pelo sfFinder. Enquanto que para a abordagem utilizando o *hmmscore* foram reconhecidas 2.284 proteínas no metagenoma, o que corresponde a 21,77% do número de proteínas geradas pelo sfFinder.

Em relação a análise baseada em subfamílias, o alinhamento com o DIAMOND recuperou um total de 274 (45,21%) subfamílias diferentes e com o programa *hmmscore* foi possível recuperar 237 subfamílias (39,1%) (**Tabela 6**). A distribuição do número de ORFs preditas e total de subfamílias identificadas no metagenoma por família estão descritas na **Tabela 6**.

Ao todo 693 sequências tiveram anotação exclusiva para a busca com HMMs e 670 para a busca com o DIAMOND. Em relação à proporção e quantidade de proteínas encontradas para cada família, o DIAMOND identificou mais sequências relacionadas às famílias ALKB, EXDO e RHDO que representa 59,7% das sequências encontradas para todas as famílias, enquanto que para o método do *hmmscore* foram identificadas mais sequências relacionadas às famílias RHDO, MACR e MCIS que representa 57,8% das sequências encontradas para todas as famílias (**Tabela 6**).

Tabela 6: Proteínas alinhadas pelo DIAMOND e pelo *hmmscore* que posteriormente foram preditas no metagenoma.

| Famílias | Nº de proteínas preditas (DIAMOND) | Nº de subfamílias preditas (DIAMOND) | Nº de proteínas preditas (HMMs) | Nº de subfamílias preditas (HMMs) |
|----------|------------------------------------|--------------------------------------|---------------------------------|-----------------------------------|
|----------|------------------------------------|--------------------------------------|---------------------------------|-----------------------------------|

| | | | | |
|--------|-------|-----|-------|-----|
| ALKB | 315 | 20 | 53 | 14 |
| BCOAR | 1 | 1 | 5 | 2 |
| BSSA | 12 | 2 | 11 | 3 |
| CYP153 | 123 | 13 | 212 | 14 |
| EXDO | 338 | 49 | 255 | 35 |
| FERRE | 189 | 19 | 99 | 14 |
| INDO | 224 | 44 | 221 | 38 |
| MACR | 176 | 30 | 262 | 22 |
| MCIS | 89 | 13 | 331 | 16 |
| RHDO | 697 | 65 | 728 | 62 |
| RHMO | 24 | 5 | 28 | 5 |
| TDLH | 73 | 13 | 79 | 12 |
| TOTAL | 2.261 | 274 | 2.284 | 237 |

Fonte: Próprio Autor

A comparação entre as duas abordagens de alinhamentos para identificação das subfamílias, teve uma intersecção de 1.591 ORFs na proporção de distribuição descritos na **Tabela 7**, observando os resultados as principais famílias com mais sequências identificadas foram: RHDO, EXDO e INDO que juntas contabilizam 64,88% das sequências para as 12 famílias de proteínas.

Tabela 7: Proporção de sequências para as 12 famílias de enzimas prospectadas na intersecção utilizando o DIAMOND e o *hmmscore*.

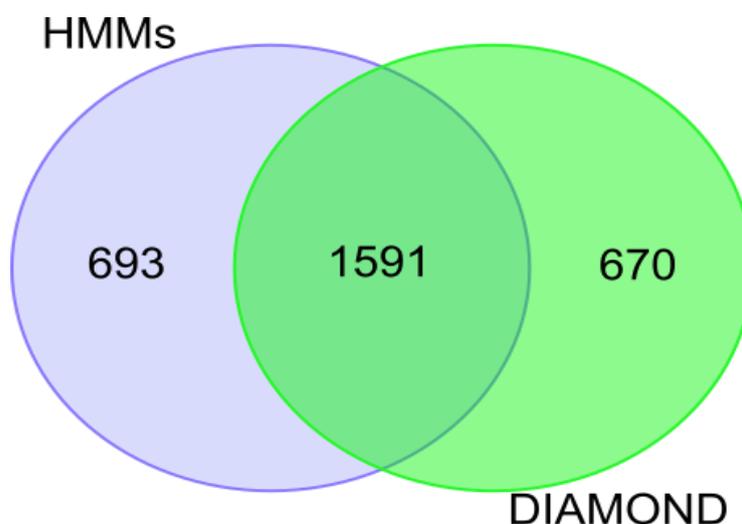
| Família | Nº de proteínas | Representatividade |
|----------------|------------------------|---------------------------|
| ALKB | 52 | 3,26% |
| BCOAR | 1 | 0,06% |
| BSSA | 5 | 0,31% |
| CYP153 | 96 | 6,04% |
| EXDO | 235 | 14,78% |
| FERRE | 87 | 5,46% |
| INDO | 211 | 13,26% |
| MACR | 146 | 9,18% |
| MCIS | 78 | 4,90% |
| RHDO | 586 | 36,84% |
| RHMO | 23 | 1,45% |
| TDLH | 71 | 4,46% |
| Total | 1.591 | 100% |

Fonte: Próprio Autor

A abordagem usando o programa *hmmscore*, obteve um valor de 1,02% de ORFs exclusivas a mais anotadas para as subfamílias do que o método empregado

pela ferramenta DIAMOND cujos números são de 693 e 670 ORFs para as duas ferramentas respectivamente. Essa relação pode ser observada na **Figura 8**.

Figura 8: Diagrama de Venn representando as quantidades e distribuição de proteínas das famílias de proteínas associadas à dHPAs preditas no metagenoma de Rock Bay utilizando os programas *hmmscore* (HMMs) e DIAMOND.



Fonte: Próprio Autor

No caso da abordagem complementar utilizando o DIAMOND e HMMs juntos, esta combinação permitiu uma identificação de um total de 2.954 proteínas relacionadas à dHPAs, o que representa um ganho mínimo em termos de número de classificações de ao menos 29,33% no poder de classificação dessas proteínas, comparado com os 2 métodos isoladamente.

As sequências foram disponibilizadas no Zenodo pelo link DOI: <https://doi.org/10.5281/zenodo.5076605>.

5.3.2 CORRESPONDÊNCIA TAXONÔMICA EM RELAÇÃO ÀS PROTEÍNAS dHPAS

A análise de correspondência taxonômica entre as 1.591 proteínas dHPAs da intersecção utilizando as 2 abordagens, pelo método do DIAMOND e o

hmmscore pode ser visualizada na **Tabela 8**, a partir dela pode se constatar que o gênero *Pseudomonas* foi o que encontramos maior correspondência para as famílias: CYP153, FERRE, INDO, MACR, MCIS, RHMO, RHDO e TDLH. Em relação às outras famílias, a ALKB teve maior correspondência com o gênero *Immundisolibacter*, a família BCOAR teve maior correspondência com o gênero *Geobacter*, a família BSSA teve uma maior correspondência com o gênero *Rahnella* e a família EXDO com o gênero *Paenibacillus*.

Na literatura o gênero *Pseudomonas* é descrito como um grupo de bactérias bacilares com enorme capacidade de dHPAs devido a sua grande versatilidade na obtenção de carbono a partir de moléculas extremamente tóxicas para a maioria dos micro-organismos. Dentre os HPAs que são clivados pelas *Pseudomonas* está: Fenantreno, Pireno, Fluoranteno, entre outros HPAs que variam seu peso molecular de baixo a elevado (CHEBBI et al., 2017).

O gênero *Immundisolibacter* é descrito como um grupo de bactérias com destaque na degradação de grandes compostos de cadeias polares. Além de possuírem algumas espécies como a *Immundisolibacter cernigliae* que possui a capacidade de dHPAs e outros xenobióticos tóxicos a natureza, entre os HPAs mais degradados por essa espécie estão o Pireno, Fenantreno e Aminobenzeno principalmente (CORTESELLI et al., 2017).

Em relação ao gênero *Geobacter* ele é caracterizado como um grupo de bactérias de respiração anaeróbica que possui capacidades de degradação de vários compostos orgânicos, entre eles HPAs como o Fenantreno, atuando principalmente na transferência de elétrons durante a etapa de biorremediação de áreas contaminadas por derivados de petróleo, já que esse grupo tem grande atividade oxidativa para a degradação dessa categoria de xenobióticos (ZHANG et al., 2021).

No caso do gênero *Rahnella*, este é caracterizado como um grupo de bactérias em forma de bastonete, com abundância prevalente em efluentes de água doce, atuando na fixação de nitrogênio e degradação de alguns poluentes orgânicos persistentes, dentre eles, moléculas de HPAs de baixo a intermediário peso molecular (SARKAR et al., 2017).

Considerando-se o gênero *Paenibacillus*, ele é descrito como um grupo de bactérias que possuem grande relevância na fixação de nitrogênio e solubilização

de fosfato em condições ideais. Além de serem capazes de secretar muitas enzimas que degradam uma grande variedade de grupos biomoleculares, entre eles: polissacarídeos e alguns poluentes orgânicos, tendo uma performance de atuação moderada na degradação de dibenzofurano que é um HPA de peso molecular intermediário e sendo altamente mutagênico (THI et al., 2019).

Tabela 8: Táxons a nível de gênero que tiveram maior correspondência entre as 12 famílias de proteínas dHPAs

| Táxon associado | Família | Nº de proteínas |
|--------------------------|----------------|------------------------|
| <i>Immundisolibacter</i> | ALKB | 7 |
| <i>Geobacter</i> | BCOAR | 1 |
| <i>Rahnella</i> | BSSA | 2 |
| <i>Pseudomonas</i> | CYP153 | 99 |
| <i>Paenibacillus</i> | EXDO | 20 |
| <i>Pseudomonas</i> | FERRE | 19 |
| <i>Pseudomonas</i> | INDO | 105 |
| <i>Pseudomonas</i> | MACR | 61 |
| <i>Pseudomonas</i> | MCIS | 27 |
| <i>Pseudomonas</i> | RHDO | 132 |
| <i>Pseudomonas</i> | RHMO | 13 |
| <i>Pseudomonas</i> | TDLH | 19 |

Fonte: Próprio Autor

A anotação de táxons para o nível de classe das proteínas dHPAs da intersecção entre os resultados do DIAMOND e do *hmmscore*, resultou em uma associação de 42,8% para a classe *Gammaproteobacteria*, 23,7% para *Betaproteobacteria*, 13,8% para *Alphaproteobacteria*, 7,1% para *Bacilli*, 3,8% para *Actinomycetia* e 0,2% para *Deltaproteobacteria*, entre as classes mais significativas. Além disso, a anotação dos gêneros que tiveram maior frequência para o total das 12 famílias de proteínas dHPAs pode ser visualizada na **Tabela 9**, onde os 6 principais gêneros com maior frequência entre as proteínas dHPAs são respectivamente: *Pseudomonas*, *Immundisolibacter*, *Paenibacillus*, *Achromobacter*, *Bradyrhizobium* e *Pusillimonas*.

Em relação ao gênero *Achromobacter*, ele é descrito como um grupo de bactérias com grande potencial de biorremediação de áreas contaminadas por derramamento de petróleo e derivados entre eles HPAs e outros hidrocarbonetos, também é um grupo associado a degradação de bifenilo que é um dos HPAs mais simples encontrados na natureza (HONG et al., 2017).

No caso das bactérias do gênero *Bradyrhizobium*, elas são caracterizadas principalmente por seu papel na fixação de nitrogênio, no entanto por possuírem enzimas da família ALKB, elas possuem a capacidade de dHPAs de cadeias mais simples como por exemplo: n-alcanos, ciclohexanos e estirenos de peso molecular intermediário (CHICIUDEAN et al., 2019).

Por fim as bactérias do gênero *Pusillimonas*, são relacionadas a dHPAs de n-alcanos e HPAs de baixo peso molecular, sob concentrações destes compostos no meio ambiente que variam de baixas a intermediárias, dependendo da temperatura e pH do ambiente, dessa forma esse gênero possui atividade de degradação moderada para esses compostos aromáticos (WANG et al., 2020).

Tabela 9: Gêneros relacionados às proteínas dHPAs que tiveram maior Representatividade na intersecção entre as 2 abordagens (DIAMOND e *hmmscore*).

| Gênero | Nº de proteínas | Representatividade |
|--------|-----------------|--------------------|
|--------|-----------------|--------------------|

| | | |
|--------------------------|-----|-------|
| <i>Pseudomonas</i> | 490 | 30,7% |
| <i>Immundisolibacter</i> | 120 | 7,5% |
| <i>Paenibacillus</i> | 101 | 6,3% |
| <i>Achromobacter</i> | 99 | 6,2% |
| <i>Bradyrhizobium</i> | 96 | 6% |
| <i>Pusillimonas</i> | 71 | 4,4% |

Fonte: Próprio Autor

A associação entre os *taxa* e a intersecção de proteínas obtidas pelo DIAMOND e *hmmscore* foi disponibilizada no repositório Zenodo pelo link: <https://doi.org/10.5281/zenodo.5158607>.

6. CONCLUSÃO

A combinação das montagens com a utilização do MeGAMerge se mostrou mais eficaz para uma montagem com maior integridade do metagenoma na construção dos *contigs*, em comparação às montagens individuais feitas pelo MEGAHIT, IDBA_ud e MetaSPAdes. Além disso, houve uma ampliação no número de proteínas previamente associadas à dHPAs e uma categorização em subfamílias das famílias de proteínas por meio da aplicação da ferramenta sfFinder. A construção de modelos estatísticos (HMMs) utilizando esse novo conjunto de sequências foi realizado com sucesso e tanto os modelos quanto às sequências foram aplicados na abordagem de prospecção do metagenoma, além disso, esses dados das sequências de proteínas com potencial para a dHPAs foram

compartilhados em um repositório público para que possam ser utilizadas em projetos futuros que visam o uso dessas proteínas no desenvolvimento de estratégias de biorremediação. Dessa forma, obtivemos sucesso na predição de regiões codificadoras de proteínas (ORFs) com função de dHPAs, uma vez que obtivemos uma correspondência significativa entre as proteínas encontradas pelo sfFinder e as regiões codificadoras destas no metagenomas. Assim como também foi observado que as abordagens propostas, considerando o uso em conjunto das buscas por similaridade (DIAMOND) ou por correspondência de HMMs (*hmmscore*) para a identificação de proteínas dHPAs no metagenoma tiveram uma resposta mais ampla comparada à estratégia trivial. Por fim, a anotação taxonômica das proteínas prospectadas a partir do resultado da predição gênica do metagenoma, revelou uma proporção maior de bactérias da classe *Gammaproteobacteria*, com destaque aos gêneros: *Pseudomonas*, *Immunisolibacter*, *Paenibacillus*, *Achromobacter*, *Bradyrhizobium* e *Pusillimonas*. Esses gêneros de bactérias possuem na literatura científica evidências dessa capacidade dHPAs. Dessa forma, isso revela que a abordagem de prospecção proposta está retornando resultados coerentes com o que foi esperado dela. No entanto, ressaltamos a necessidade de ensaios *in vitro* para aferir a capacidade dHPAs das proteínas prospectadas no metagenoma.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABATENH, Endeshaw et al. The role of microorganisms in bioremediation-A review. **Open Journal of Environmental Biology**, v. 2, n. 1, p. 038-046, 2017.

ABDEL-SHAFY, H. I.; MANSOUR, M. S. M. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. **Egyptian Journal of Petroleum**, v. 25, n. 1, p. 107–123, 1 mar. 2016.

ALTENHOFF, A. M. et al. Resolving the Ortholog Conjecture: Orthologs Tend to Be Weakly, but Significantly, More Similar in Function than Paralogs. **PLOS Computational Biology**, v. 8, n. 5, p. e1002514, 17 maio 2012.

ALVES, Luana de Fátima et al. Metagenomic approaches for understanding new concepts in microbial science. **International journal of genomics**, v. 2018, 2018.

BAMFORTH, S. M.; SINGLETON, I. Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. **Journal of Chemical Technology & Biotechnology**, v. 80, n. 7, p. 723–736, 2005.

BANDOWE, B. A. M.; MEUSEL, H. Nitrated polycyclic aromatic hydrocarbons (nitro-PAHs) in the environment – A review. **Science of The Total Environment**, v. 581–582, p. 237–257, 1 mar. 2017.

BAO, Y.-J. et al. High-throughput metagenomic analysis of petroleum-contaminated soil microbiome reveals the versatility in xenobiotic aromatics metabolism. **Journal of Environmental Sciences**, v. 56, p. 25–35, 1 jun. 2017.

BERGMANN, F. et al. Genomic insights into the metabolic potential of the polycyclic aromatic hydrocarbon degrading sulfate-reducing Deltaproteobacterium N47. **Environmental Microbiology**, v. 13, n. 5, p. 1125–1137, 2011.

BERINI, F. et al. Metagenomics: novel enzymes from non-culturable microbes. **FEMS Microbiology Letters**, v. 364, n. 21, 15 nov. 2017.

BREITWIESER, F. P.; LU, J.; SALZBERG, S. L. A review of methods and databases for metagenomic classification and assembly. **Briefings in Bioinformatics**, 8 fev. 2019.

BROWN, J.; PIRRUNG, M.; MCCUE, L. A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. **Bioinformatics**, v. 33, n. 19, p. 3137–3139, 1 out. 2017.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using IAMOND. **Nature Methods**, v. 12, n. 1, p. 59–60, jan. 2015.

CABRAL, L. et al. Microbial functional responses to long-term anthropogenic impact in mangrove soils. **Ecotoxicology and Environmental Safety**, v. 160, p. 231–239, 30 set. 2018.

CÁMARA, B. et al. trans-Dienelactone hydrolase from *Pseudomonas reinekei* MT1, a novel zinc-dependent hydrolase. **Biochemical and Biophysical Research Communications**, v. 376, n. 2, p. 423–428, 14 nov. 2008.

CARUSO, M. S. F.; ALABURDA, J. Polycyclic aromatic hydrocarbons - benzo(a)pyrene: a review. **Revista do Instituto Adolfo Lutz (Impresso)**, v. 67, n. 1, p. 1–27, abr. 2008.

CHEBBI, Alif et al. Polycyclic aromatic hydrocarbon degradation and biosurfactant production by a newly isolated *Pseudomonas* sp. strain from used motor oil-contaminated soil. **International Biodeterioration & Biodegradation**, v. 122, p. 128-140, 2017

CHICIUDEAN, Iulia et al. Jet A-1 Bacterial contamination: A case study of cultivable bacteria diversity, alkane degradation and biofilm formation. **Polish Journal of Environmental Studies**, v. 28, n. 6, p. 4139-4146, 2019..

CORTESELLI, Elizabeth M.; AITKEN, Michael D.; SINGLETON, David R. Description of *Immundisolibacter cernigliae* gen. nov., sp. nov., a high-molecular-weight polycyclic aromatic hydrocarbon-degrading bacterium within the class Gammaproteobacteria, and proposal of *Immundisolibacterales* ord. nov. and *Immundisolibacteraceae* fam. nov. **International journal of systematic and evolutionary microbiology**, v. 67, n. 4, p. 925, 2017.

DI TOMMASO, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. **Nucleic Acids Research**, v. 39, n. suppl_2, p. W13–W17, 1 jul. 2011.

DIDION, J. P.; MARTIN, M.; COLLINS, F. S. Atropos: specific, sensitive, and speedy

trimming of sequencing reads. **PeerJ**, v. 5, p. e3720, 2017.

DREVINSKAS, T. et al. Downscaling the in vitro test of fungal bioremediation of polycyclic aromatic hydrocarbons: methodological approach. **Analytical and Bioanalytical Chemistry**, v. 408, n. 4, p. 1043–1053, 1 fev. 2016.

DUAN, L. et al. Managing long-term polycyclic aromatic hydrocarbon contaminated soils: a risk-based approach. **Environmental Science and Pollution Research**, v. 22, n. 12, p. 8927–8941, 1 jun. 2015.

DUARTE, M. et al. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of in situ bioremediation. **Environmental Microbiology**, v. 19, n. 8, p. 2992–3011, 2017.

DZIOŃEK, A.; WOJCIESZYŃSKA, D.; GUZIK, U. Natural carriers in bioremediation: A review. **Electronic Journal of Biotechnology**, v. 19, n. 5, p. 28–36, set. 2016.

FABBRO, C. D. et al. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. **PLOS ONE**, v. 8, n. 12, p. e85024, 23 dez. 2013.

FELCZYKOWSKA, A. et al. Sampling, metadata and DNA extraction - important steps in metagenomic studies. **Acta Biochimica Polonica**, v. 62, n. 1, 13 fev. 2015.

FIORENTINI, F. et al. The Extreme Structural Plasticity in the CYP153 Subfamily of P450s Directs Development of Designer Hydroxylases. **Biochemistry**, v. 57, n. 48, p. 6701–6714, 4 dez. 2018.

FUJII, T. et al. The crystal structure of maleylacetate reductase from *Rhizobium* sp. strain MTP-10005 provides insights into the reaction mechanism of enzymes in its original family. **Proteins: Structure, Function, and Bioinformatics**, v. 84, n. 8, p. 1029–1042, 2016.

GEIER, Mitra C. et al. Comparative developmental toxicity of a comprehensive suite of polycyclic aromatic hydrocarbons. **Archives of toxicology**, v. 92, n. 2, p. 571-586, 2018.

GHOSAL, D. et al. Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. **Frontiers in Microbiology**, v. 7, p. 1369, 2016.

GODHEJA, J. et al. Xenobiotic compounds present in soil and water: a review on remediation strategies. **J. Environ. Anal. Toxicol**, v. 6, n. 392, p. 2161-0525.1000392, 2016.

GUPTA, B. et al. Enhanced pyrene degradation by a biosurfactant producing *Acinetobacter baumannii* BJ5: Growth kinetics, toxicity and substrate inhibition studies. **Environmental Technology & Innovation**, v. 19, p. 100804, 1 ago. 2020.

GUZIK, U.; HUPERT-KOCUREK, K.; WOJCIESZYŃSKA, D. Intradiol Dioxygenases — The Key Enzymes in Xenobiotics Degradation. **Biodegradation of Hazardous and Special Products**, 14 jun. 2013.

HANDELSMAN, J. Metagenomics: application of genomics to uncultured microorganisms. **Microbiology and molecular biology reviews: MMBR**, v. 68, n. 4, p. 669–685, dez. 2004.

HEIDER, J. et al. Structure and Function of Benzylsuccinate Synthase and Related Fumarate-Adding Glycyl Radical Enzymes. **Journal of Molecular Microbiology and Biotechnology**, v. 26, n. 1–3, p. 29–44, 2016.

HODKINSON, B. P.; GRICE, E. A. Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. **Advances in Wound Care**, v. 4, n. 1, p. 50–58, 28 jul. 2014.

HOFF, K.; STANKE, M. Current methods for automated annotation of protein-coding genes. **Current Opinion in Insect Science**, Insect genomics * Development and regulation. v. 7, p. 8–14, 1 fev. 2015.

HONG, Yue-Hui et al. Genome sequencing reveals the potential of *Achromobacter* sp. HZ01 for bioremediation. **Frontiers in microbiology**, v. 8, p. 1507, 2017.

HOWISON, M. High-Throughput Compression of FASTQ Data with SeqDB. **IEEE/ACM Trans. Comput. Biol. Bioinformatics**, v. 10, n. 1, p. 213–218, jan. 2013.

HUGENHOLTZ, P. Exploring prokaryotic diversity in the genomic era. **Genome Biology**, v. 3, n. 2, p. reviews0003.1, 29 jan. 2002.

IARC **Cancer Base** No. 11. Lyon, IARC, 2013. 2016.

JACQUES, R. J. S. et al. Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons. **Ciência Rural**, v. 37, n. 4, p. 1192–1201, ago. 2007.

KADRI, T. et al. Biodegradation of polycyclic aromatic hydrocarbons (PAHs) by fungal enzymes: A review. **Journal of Environmental Sciences**, v. 51, p. 52–74, 1 jan. 2017.

KARPLUS, K. et al. Predicting protein structure using hidden Markov models. **Proteins**, v. Suppl 1, p. 134–139, 1997.

KIM, D.-W. et al. Comparative genomic analysis of pyrene-degrading *Mycobacterium* species: Genomic islands and ring-hydroxylating dioxygenases involved in pyrene degradation. **Journal of Microbiology**, v. 56, n. 11, p. 798–804, 1 nov. 2018.

KIM, K.-H.; KABIR, E.; JAHAN, S. A. Exposure to pesticides and the associated

human health effects. **Science of The Total Environment**, v. 575, p. 525–535, 1 jan. 2017.

KRISHNAMURTHY, N.; BROWN, D.; SJÖLANDER, K. FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function. **BMC Evolutionary Biology**, v. 7, n. 1, p. S12, 8 fev. 2007.

KUATSJAH, E. et al. Characterization of an extradiol dioxygenase involved in the catabolism of lignin-derived biphenyl. **FEBS Letters**, v. 591, n. 7, p. 1001–1009, 2017.

KULIK, N. et al. Degradation of polycyclic aromatic hydrocarbons by combined chemical pre-oxidation and bioremediation in creosote contaminated soil. **Journal of Environmental Management**, v. 78, n. 4, p. 382–391, 1 mar. 2006.

KUNIN, V. et al. A bioinformatician's guide to metagenomics. **Microbiology and molecular biology reviews: MMBR**, v. 72, n. 4, p. 557–578, Table of Contents, dez. 2008.

KUPPUSAMY, S. et al. Remediation approaches for polycyclic aromatic hydrocarbons (PAHs) contaminated soils: Technological constraints, emerging trends and future directions. **Chemosphere**, v. 168, p. 944–968, 1 fev. 2017.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 4 mar. 2012.

LEINONEN, R.; SUGAWARA, H.; SHUMWAY, M. The Sequence Read Archive. **Nucleic Acids Research**, v. 39, n. suppl_1, p. D19–D21, 1 jan. 2011.

LI, D. et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. **Bioinformatics (Oxford, England)**, v. 31, n. 10, p. 1674–1676, 15 maio 2015.

LIU, Q. et al. Aerobic degradation of crude oil by microorganisms in soils from four geographic regions of China. **Scientific Reports**, v. 7, n. 1, p. 14856, 1 nov. 2017.

MAHTO, K. U.; DAS, S. Whole genome characterization and phenanthrene catabolic pathway of a biofilm forming marine bacterium *Pseudomonas aeruginosa* PFL-P1. **Ecotoxicology and Environmental Safety**, v. 206, p. 111087, 15 dez. 2020.

MANDE, S. S.; MOHAMMED, M. H.; GHOSH, T. S. Classification of metagenomic sequences: methods and challenges. **Briefings in Bioinformatics**, v. 13, n. 6, p. 669–681, 1 nov. 2012.

MARQUÈS, M. et al. Photodegradation of polycyclic aromatic hydrocarbons in soils under a climate change base scenario. **Chemosphere**, v. 148, p. 495–503, 1 abr. 2016.

MECKENSTOCK, R. U. et al. Anaerobic Degradation of Benzene and Polycyclic Aromatic Hydrocarbons. **Journal of Molecular Microbiology and Biotechnology**, v. 26, n. 1–3, p. 92–118, 2016.

MEIER, M. J.; PATERSON, E. S.; LAMBERT, I. B. Use of Substrate-Induced Gene Expression in Metagenomic Analysis of an Aromatic Hydrocarbon-Contaminated Soil. **Applied and Environmental Microbiology**, v. 82, n. 3, p. 897–909, 01 2016.

MIGNÉE, C. et al. Gallium ferredoxin as a tool to study the effects of ferredoxin binding to photosystem I without ferredoxin reduction. **Photosynthesis Research**, v. 134, n. 3, p. 251–263, 1 dez. 2017.

MIKHEENKO, A.; SAVELIEV, V.; GUREVICH, A. MetaQUAST: evaluation of metagenome assemblies. **Bioinformatics**, v. 32, n. 7, p. 1088–1090, 1 abr. 2016.

MUKHERJEE, A. et al. Bioinformatic Approaches Including Predictive Metagenomic

Profiling Reveal Characteristics of Bacterial Response to Petroleum Hydrocarbon Contamination in Diverse Environments. **Scientific Reports**, v. 7, n. 1, p. 1–22, 24 abr. 2017.

MUSUMECI, M. A. et al. Substrate specificities of aromatic ring-hydroxylating oxygenases of an uncultured gammaproteobacterium from chronically-polluted subantarctic sediments. **International Biodeterioration & Biodegradation**, v. 137, p. 127–136, 1 fev. 2019.

NICOLOPOULOU-STAMATI, P. et al. Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture. **Frontiers in Public Health**, v. 4, 2016.

NIELSEN, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. **Nature Biotechnology**, v. 32, n. 8, p. 822–828, ago. 2014.

NURK, S. et al. metaSPAdes: a new versatile metagenomic assembler. **Genome Research**, v. 27, n. 5, p. 824–834, 2017.

ONDOV, B. D.; BERGMAN, N. H.; PHILLIPPY, A. M. Interactive metagenomic visualization in a Web browser. **BMC Bioinformatics**, v. 12, n. 1, p. 385, 30 set. 2011.

PAGÉ, A. P.; YERGEAU, É.; GREER, C. W. *Salix purpurea* Stimulates the Expression of Specific Bacterial Xenobiotic Degradation Genes in a Soil Contaminated with Hydrocarbons. **PLOS ONE**, v. 10, n. 7, p. e0132062, 10 jul. 2015.

PATEL, Avani Bharatkumar et al. Polycyclic aromatic hydrocarbons: sources, toxicity and remediation approaches. **Frontiers in Microbiology**, v. 11, p. 2675, 2020.

PAZ, A. P. S. DA et al. Presença de hidrocarbonetos policíclicos aromáticos em produtos alimentícios e a sua relação com o método de cocção e a natureza do alimento. **Brazilian Journal of Food Technology**, v. 20, 2017.

PENG, T. et al. Identification of A Ring-Hydroxylating Dioxygenases Capable of Anthracene and Benz[a]anthracene Oxidization from *Rhodococcus* sp. P14. **Journal of Molecular Microbiology and Biotechnology**, v. 28, n. 4, p. 183–189, 2018.

PENG, Y. et al. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics (Oxford, England)**, v. 28, n. 11, p. 1420–1428, 1 jun. 2012.

PETERS, I. et al. Zenodo in the Spotlight of Traditional and New Metrics. **Frontiers in Research Metrics and Analytics**, v. 2, 2017.

PICCOLO, L. L. et al. Involvement of an Alkane Hydroxylase System of *Gordonia* sp. Strain SoCg in Degradation of Solid n-Alkanes. **Applied and Environmental Microbiology**, v. 77, n. 4, p. 1204–1213, 15 fev. 2011.

POURSAT, Baptiste AJ et al. Implications of microbial adaptation for the assessment of environmental persistence of chemicals. **Critical Reviews in Environmental Science and Technology**, v. 49, n. 23, p. 2220-2255, 2019.

QUINCE, C. et al. Shotgun metagenomics, from sampling to analysis. **Nature Biotechnology**, v. 35, n. 9, p. 833–844, set. 2017.

ROBLES-GONZÁLEZ, I. V.; FAVA, F.; POGGI-VARALDO, H. M. A review on slurry bioreactors for bioremediation of soils and sediments. **Microbial Cell Factories**, v. 7, n. 1, p. 5, 29 fev. 2008.

RODRIGUE, S. et al. Unlocking Short Read Sequencing for Metagenomics. **PLOS ONE**, v. 5, n. 7, p. e11840, 28 jul. 2010.

ROUMPEKA, D. D. et al. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. **Frontiers in Genetics**, v. 8, 2017.

SABINO, F. et al. HPA e Nitro-HPA em Ambiente Semifechado Impactado por Emissão da Combustão de Diesel/Biodiesel (B5). **Orbital - The Electronic Journal of Chemistry**, v. 7, 30 set. 2015.

SANGKHARAK, K. et al. The Degradation of Phenanthrene, Pyrene, and Fluoranthene and Its Conversion into Medium-Chain-Length Polyhydroxyalkanoate by Novel Polycyclic Aromatic Hydrocarbon-Degrading Bacteria. **Current Microbiology**, v. 77, n. 6, p. 897–909, 1 jun. 2020.

SARKAR, Poulomi et al. Enrichment and characterization of hydrocarbon-degrading bacteria from petroleum refinery waste as potent bioaugmentation agent for in situ bioremediation. **Bioresource technology**, v. 242, p. 15-27, 2017.

SAZYKIN, I. et al. Cyclohexane, naphthalene, and diesel fuel increase oxidative stress, CYP153, sodA, and recA gene expression in *Rhodococcus erythropolis*. **MicrobiologyOpen**, v. 0, n. 0, p. e855, [s.d.].

SCHMID, G.; RENÉ, S. B.; BOLL, M. Enzymes of the benzoyl-coenzyme A degradation pathway in the hyperthermophilic archaeon *Ferroglobus placidus*. **Environmental Microbiology**, v. 17, n. 9, p. 3289–3300, 2015.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics (Oxford, England)**, v. 27, n. 6, p. 863–864, 15 mar. 2011.

SCHOLZ, M. B.; LO, C.-C.; CHAIN, P. S. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. **Current Opinion in Biotechnology**, Analytical biotechnology. v. 23, n. 1, p. 9–15, 1 fev.

2012.

SCHOLZ, M.; LO, C.-C.; CHAIN, P. S. G. Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. **Scientific Reports**, v. 4, p. 6480, 1 out. 2014.

SEDLAR, Karel; KUPKOVA, Kristyna; PROVAZNIK, Ivo. Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. **Computational and Structural Biotechnology Journal**, v. 15, p. 48-55, 2017.

SEEMANN, T. Prokka: rapid prokaryotic genome annotation. **Bioinformatics (Oxford, England)**, v. 30, n. 14, p. 2068–2069, 15 jul. 2014.

SEYHAN, D. et al. Elucidating the Stereochemistry of Enzymatic Benzylsuccinate Synthesis with Chirally Labeled Toluene. **Angewandte Chemie International Edition**, v. 55, n. 38, p. 11664–11667, 2016.

SHAH, V. et al. Taxonomic Profiling and Metagenome Analysis of a Microbial Community from a Habitat Contaminated with Industrial Discharges. **Microbial Ecology**, v. 66, n. 3, p. 533–550, 1 out. 2013.

SHAHSAVARI, E. et al. Biological Degradation of Polycyclic Aromatic Compounds (PAHs) in Soil: a Current Perspective. **Current Pollution Reports**, v. 5, n. 3, p. 84–92, 1 set. 2019.

SHAO, Z.; WANG, W. Enzymes and genes involved in aerobic alkane degradation. **Frontiers in Microbiology**, v. 4, 2013.

SHARPTON, T. J. An introduction to the analysis of shotgun metagenomic data. **Frontiers in Plant Science**, v. 5, 2014.

SINGH, L.; VARSHNEY, J. G.; AGARWAL, T. Polycyclic aromatic hydrocarbons'

formation and occurrence in processed food. **Food Chemistry**, v. 199, p. 768–781, 15 maio 2016.

SJÖLANDER, K. et al. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. **Computer applications in the biosciences: CABIOS**, v. 12, n. 4, p. 327–345, ago. 1996.

SJÖLANDER, K. Phylogenetic inference in protein superfamilies: analysis of SH2 domains. **Proceedings. International Conference on Intelligent Systems for Molecular Biology**, v. 6, p. 165–174, 1998.

SMUŁEK, W. et al. Increased biological removal of 1-chloronaphthalene as a result of exposure: A study of bacterial adaptation strategies. **Ecotoxicology and Environmental Safety**, v. 185, p. 109707, 15 dez. 2019.

SUMAN, S.; SINHA, A.; TARAFDAR, A. Polycyclic aromatic hydrocarbons (PAHs) concentration levels, pattern, source identification and soil toxicity assessment in urban traffic soil of Dhanbad, India. **Science of The Total Environment**, v. 545–546, p. 353–360, 1 mar. 2016.

SUN, F.; LITTLEJOHN, D.; DAVID GIBSON, M. Ultrasonication extraction and solid phase extraction clean-up for determination of US EPA 16 priority pollutant polycyclic aromatic hydrocarbons in soils by reversed-phase liquid chromatography with ultraviolet absorption detection Presented in part at FACSS, the 23rd Annual Conference of the Federation of Analytical Chemistry and Spectroscopy Societies, Kansas City, MO, USA, 29 September–4 October 1996.1. **Analytica Chimica Acta**, v. 364, n. 1, p. 1–11, 25 maio 1998.

SUZEK, B. E. et al. UniRef: comprehensive and non-redundant UniProt reference clusters. **Bioinformatics**, v. 23, n. 10, p. 1282–1288, 15 maio 2007.

TAN, B. et al. Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. **The**

ISME Journal, v. 9, n. 9, p. 2028–2045, set. 2015.

TECHTMANN, S. M.; HAZEN, T. C. Metagenomic applications in environmental monitoring and bioremediation. **Journal of Industrial Microbiology & Biotechnology**, v. 43, n. 10, p. 1345–1354, 1 out. 2016.

TERRÓN-GONZÁLEZ, L. et al. Functional Metagenomics of a Biostimulated Petroleum-Contaminated Soil Reveals an Extraordinary Diversity of Extradiol Dioxygenases. **Applied and Environmental Microbiology**, v. 82, n. 8, p. 2467–2478, 15 abr. 2016.

THI, Tran Vu Ngoc et al. Isolation and characterization of a moderate thermophilic *Paenibacillus naphthalenovorans* strain 4B1 capable of degrading dibenzofuran from dioxin-contaminated soil in Vietnam. **Journal of bioscience and bioengineering**, v. 128, n. 5, p. 571-577, 2019.

THOMAS, F. et al. Isolation and substrate screening of polycyclic aromatic hydrocarbon degrading bacteria from soil with long history of contamination. **International Biodeterioration & Biodegradation**, v. 107, p. 1–9, 1 fev. 2016.

UFARTÉ, L. et al. Metagenomics for the discovery of pollutant degrading enzymes. **Biotechnology Advances**, v. 33, n. 8, p. 1845–1854, 1 dez. 2015 .U. S. EPA Guidelines for neurotoxicity risk assessment. **Federal Register**, v. 63, n. 93, p. 26926-26954, 1998.

VAN DER WALT, A. J. et al. Assembling metagenomes, one community at a time. **BMC Genomics**, v. 18, n. 1, p. 521, 10 jul. 2017.

VARJANI, S. J. Microbial degradation of petroleum hydrocarbons. **Bioresource Technology**, v. 223, p. 277–286, 1 jan. 2017.

VILCHEZ-VARGAS, R. et al. Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel

internally calibrated microarray system. **Environmental Microbiology**, v. 15, n. 4, p. 1016–1039, abr. 2013.

WANG, Z.; CHEN, Y.; LI, Y. A Brief Review of Computational Gene Prediction Methods. **Genomics, Proteomics & Bioinformatics**, v. 2, n. 4, p. 216–221, 1 nov. 2004.

WANG, Wanpeng et al. The oxidation of hydrocarbons by diverse heterotrophic and mixotrophic bacteria that inhabit deep-sea hydrothermal ecosystems. **The ISME journal**, v. 14, n. 8, p. 1994-2006, 2020.

WASI, S.; TABREZ, S.; AHMAD, M. Use of *Pseudomonas* spp. for the bioremediation of environmental pollutants: a review. **Environmental Monitoring and Assessment**, v. 185, n. 10, p. 8147–8155, 1 out. 2013.

WISTRAND, M.; SONNHAMMER, E. L. Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. **BMC Bioinformatics**, v. 6, n. 1, p. 99, 15 abr. 2005.

WOOD, D. E.; SALZBERG, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. **Genome Biology**, v. 15, n. 3, p. R46, 2014.

WOOLEY, J. C.; GODZIK, A.; FRIEDBERG, I. A Primer on Metagenomics. **PLOS Computational Biology**, v. 6, n. 2, p. e1000667, 26 fev. 2010.

XU, B. et al. Genetic diversity of catechol 1,2-dioxygenase in the fecal microbial metagenome. **Journal of Basic Microbiology**, v. 57, n. 10, p. 883–895, 2017.

XUE, W.; WARSHAWSKY, D. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review. **Toxicology and Applied Pharmacology**, v. 206, n. 1, p. 73–93, 1 ago. 2005.

YANG, J. W. et al. Evaluation of aromatic hydrocarbon decomposition catalyzed by

the dioxygenase system and substitution of ferredoxin and ferredoxin reductase. **Environmental Science and Pollution Research**, 23 set. 2018.

YUAN, H. et al. Historical records and sources of polycyclic aromatic hydrocarbons (PAHs) and organochlorine pesticides (OCPs) in sediment from a representative plateau lake, China. **Chemosphere**, v. 173, p. 78–88, 1 abr. 2017.

ZENG, J. et al. Isolation of diverse pyrene-degrading bacteria via introducing readily utilized phenanthrene. **Chemosphere**, v. 222, p. 534–540, 1 maio 2019.

ZHANG, Le et al. Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters: A review. **Renewable and Sustainable Energy Reviews**, v. 100, p. 110-126, 2019.

ZHANG, Zuotao et al. Investigation of anaerobic biodegradation of phenanthrene by a sulfate-dependent *Geobacter sulfurreducens* strain PheS2. **Journal of Hazardous Materials**, v. 409, p. 124522, 2021.

ZHAO, R. et al. Metagenomic views of microbial dynamics influenced by hydrocarbon seepage in sediments of the Gulf of Mexico. **Scientific Reports**, v. 10, n. 1, p. 5772, 1 abr. 2020.