

BRUNO KAROLSKI

METAGENÔMICA COMPARATIVA E PERFIL
METABÓLICO *IN SILICO* DE SOLOS NO
MUNICÍPIO DE CUBATÃO, SP

Tese apresentada ao Programa de
Pós-Graduação Interunidades em
Biotecnologia USP/Instituto
Butantan/IPT, para obtenção do
Título de Doutor em Biotecnologia

São Paulo
2013

BRUNO KAROLSKI

METAGENÔMICA COMPARATIVA E
PERFIL METABÓLICO *IN SILICO* DE SOLOS
NO MUNICÍPIO DE CUBATÃO, SP

Tese apresentada ao Programa de
Pós-Graduação Interunidades em
Biotecnologia USP/Instituto
Butantan/IPT, para obtenção do
Título de Doutor em Biotecnologia

Área de concentração: Biotecnologia

Orientadora: Marie-Anne Van Sluys

Versão original

São Paulo
2013

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

reprodução não autorizada pelo autor

Karolski, Bruno.

Metagenômica comparativa e perfil metabólico in silico de solos no município de Cubatão, SP / Bruno Karolski. -- São Paulo, 2013.

Orientador: Profa. Dra. Marie Anne van Sluys.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan. Área de concentração: Biotecnologia. Linha de pesquisa: Microbiologia ambiental.

Versão do título para o inglês: Comparative metagenomics and in silico metabolic soil profiling in Cubatão County, SP.

1. BTEX 2. Microbiologia ambiental 3. Compostos aromáticos
4. Metagenômica 5. Cubatão I. Sluys, Profa. Dra. Marie-Anne van II.
Universidade de São Paulo. Instituto de Ciências Biomédicas.
Programa de Pós-Graduação Interunidades em Biotecnologia
USP/IPT/Instituto Butantan III. Título.

ICB/SBIB046/2013¹

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Bruno Karolski.

Título da Tese: Metagenômica comparativa e perfil metabólico in silico de solos no município de Cubatão, SP.

Orientador(a): Profa. Dra. Marie Anne van Sluys.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a/...../....., considerou

Aprovado(a)

Reprovado(a)

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:

À minha esposa Nathália Fernandes de Azevedo
pela paciência, incentivo e cumplicidade
incondicional.

AGRADECIMENTOS

À professora Marie-Anne Van Sluys por ter mais uma vez acreditado em meu trabalho e ter aceitado o desafio de mais uma orientação além de todos os investimentos realizados nas diferentes formas.

A Elen Aquino Perpétuo por todo o apoio e disponibilidade sempre que se fez necessário, assim como todo o incentivo e confiança.

Ao professor Cláudio Oller pelo voto de confiança e oferecimento dos meios para que esse trabalho fosse realizado.

Aos meus familiares pelo suporte, apoio e pelo investimento realizado em minha educação.

Aos amigos Jonas, Guilherme, Andrés e Tatiana pelas conversas, troca de ideias e convivência no dia a dia.

Às grandes amigas Louise, Ingrid e Marcela pelo apoio e auxílio a todos os trabalhos realizados no CEPEMA.

À Vanessa e à Lígia pela ajuda com a revisão da tese.

À toda equipe do GaTE Lab pela convivência e trocas de experiências.

À equipe do CEPEMA pela receptividade e amizade.

Aos funcionários do Instituto de biociências.

À equipe de apoio às coletas de solo.

À FINEP e ao CAPES pelo financiamento e bolsas cedidas.

RESUMO

KAROLSKI, B. **Metagenômica comparativa e perfil metabólico *in silico* de solos do município de Cubatão, SP.** 2013. 203 f. Tese (Doutorado em Biotecnologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2013.

Cubatão, o maior polo industrial da América Latina também já foi uma das cidades mais poluídas do mundo. Os 30 anos de intensa atividade industrial vêm pressionando o meio ambiente com substâncias tóxicas e afetando gravemente a saúde da população. Dentre as substâncias contaminantes mais importantes da região estão os derivados de petróleo como benzeno, tolueno, etilbenzeno e xilenos. Conhecidos como BTEX, eles são produzidos e utilizados em larga escala e a contaminação ocorre frequentemente através de vazamentos. Nos solos, devido à sua solubilidade em água, essas substâncias podem se espalhar por longas distâncias a partir do ponto afetado contaminando locais afastados. Já foi comprovada a capacidade de procariotos de sobreviver em locais contaminados com BTEX e até utilizá-los como fonte de carbono. Os procariotos adaptados catabolizam os contaminantes transformando-os em substâncias menos tóxicas e até mesmo eliminando-os do ambiente, capacidade de grande interesse econômico e ambiental. Nessa linha, nossa proposta visa o estudo das comunidades microbianas de solos afetados e não afetados por BTEX. Para isso foi utilizada a metagenômica como abordagem de estudo identificando-se diferenças qualitativas e quantitativas nas estruturas microbianas de três diferentes locais do município de Cubatão, sendo um deles afetado diretamente por BTEX. Pelo método utilizado e aqui desenvolvido, foi possível identificar um panorama metabólico geral identificando-se genes relevantes e o potencial de degradação de hidrocarbonetos aromáticos de procariotos conhecidos e desconhecidos, revelando melhor o potencial metabólico dos procariotos dos solos identificados. Os resultados apresentados podem contribuir para um melhor entendimento da dinâmica *in situ* de uma comunidade microbiana afetada por BTEX assim como melhorar o conhecimento sobre a comunidade microbiana de um local altamente impactado como Cubatão.

Palavras-chave: BTEX. Microbiologia. Compostos aromáticos. Metagenômica. Cubatão.

ABSTRACT

KAROLSKI, B. **Comparative metagenomics and metabolic soil profiling in Cubatão County, SP**. 2013. 203 p. Ph. D. thesis (Biotechnology) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2013.

Cubatão is the largest industrial site in Latin America and was in the past one of the most polluted cities in the world. Thirty years of intense industrial activity has damaged the environment with toxic substances and has severely affected the inhabitants' health. Among the contaminants found in the region, the petroleum derivatives benzene, toluene, ethylbenzene and xylenes are the most important. Known collectively as BTEX, they are produced and used at a large scale and environment contamination frequently occurs. Because it is highly soluble in water, when in soil BTEX can spread long distances from the original contamination site, thus affecting large areas. Some microorganisms are known to live in contaminated environments and use contaminants such as BTEX as a unique carbon source for energy production. They catabolize contaminants into less dangerous products or even eliminate them from environment, a feature which has great commercial and environmental interest. We therefore compared the microbial communities in soils which were affected and un-affected by BTEX contamination. For this, we used a metagenomics approach and developed a comparison method to identify microorganisms and degradation potential of soils studied. We found qualitative and quantitative differences in microbial structures from three different sites in Cubatão County, one of which is contaminated with BTEX. We constructed a metabolic overview identifying important genes, degradation potential and microorganisms related to BTEX degradation. The results presented here could contribute to understanding the *in situ* dynamics of a BTEX affected microbial community as well as improving our knowledge of the microbial community of Cubatão, a highly environmentally impacted place.

Keywords: BTEX. Environmental microbiology. Aromatic compounds. Metagenomics. Cubatão.

LISTA DE FIGURAS

Figura 1 - Representação das etapas de um estudo metagenômico	29
Figura 2: Esquema de diferentes modos de contaminação do ambiente por BTEX.....	33
Figura 3 – Vias de degradação anaeróbia de BTEX	38
Figura 4 – Representação da degradação aeróbia de BTEX	39
Figura 5 - Imagem aérea dos pontos de coleta	45
Figura 6 – Região de desenho dos iniciadores Mox.....	47
Figura 7 – Análise de BTEX nos solos dos locais contaminado e Perequê por meio de técnica GC-MS	61
Figura 8 – Fotodocumentação da extração de DNA total de solo	62
Figura 9 – Distribuição percentual das sequências em filios para os domínios Bacteria e Archaea.....	64
Figura 10 – Identificação geral das sequências metagenômicas pelo MG-RAST com base no GenBank.....	66
Figura 11 – Distribuição das sequências metagenômicas classificadas em filios por diferentes métodos comparativos	69
Figura 12 – Distribuição de sequências em filios para os três locais estudados pelo método BLASTN.....	71
Figura 13 – Distribuição das sequências em subgrupos da classe Actinobacteria	72
Figura 14 – Distribuição das sequências dos filios Acidobacteria e Firmicutes em famílias.....	73
Figura 15 – Distribuição de sequências do filo Proteobacteria em famílias.....	75
Figura 16 – Classificação dos grupos mais representados em gêneros utilizando BLASTN	77
Figura 17 – Identificação dos genes das vias de degradação de xilenos e tolueno nos metagenomas.....	80
Figura 18 – Identificação dos operons de degradação de tolueno e xilenos ...	82
Figura 19 – Identificação dos domínios conservados PFAM para operons selecionados	85
Figura 20 – Contabilização e identificação de domínios conservados em diferentes metagenomas.....	87

Figura 21 – Identificação taxonômica das sequências selecionadas com similaridade aos domínios conservados de degradação de BTEX.....	89
Figura 22 – Alinhamento das sequências similares aos respectivos HMMs ...	91
Figura 23 – Identificação dos domínios conservados nos genomas dos organismos mais abundantes.....	95
Figura 24 – Amplificação do domínio tolueno-4-monoxigenase em <i>Bradyrhizobium japonicum</i> SEMIA 566.....	96
Figura 25 – Crescimento de <i>Bradyrhizobium japonicum</i> SEMIA 566 na presença de xilenos e tolueno.	97
Figura 26 – Teste da capacidade de degradação de xilenos e tolueno por <i>Bradyrhizobium japonicum</i> SEMIA 566.....	98
Figura 27 – Imagem dos arredores do local Cotia-Pará	112

LISTA DE TABELAS

Tabela 1 – Estimativa da abundância de procariotos em diferentes ambientes	18
Tabela 2 – Características dos hidrocarbonetos do grupo BTEX	32
Tabela 3 – Níveis máximos para a potabilidade da água com relação às concentrações de BTEX (CONAMA 396/2008).....	35
Tabela 4 – Relação de áreas contaminadas de Cubatão e seus respectivos contaminantes	42
Tabela 5 – Experimento de sobrevivência na presença de xilenos e tolueno por <i>Bradyrhizobium japonicum</i> SEMIA 566	59
Tabela 6 – Total de sequências analisadas pelo método Sanger.....	63
Tabela 7 – Contabilidade das sequências geradas pelo sequenciamento GS FLX	65
Tabela 8 – Contabilidade das sequências identificadas pelos diferentes métodos de análises taxonômicas	68
Tabela 9 – Identificação dos domínios conservados de modificação de anéis aromáticos.....	86
Tabela 10 – Identificação da atividade de degradação de gêneros encontrados no estudo metagenômico	121

LISTA DE ABREVIATURAS E SIGLAS

AFLP – Amplified Fragment Length Polymorphism

ANI – Average Nucleotide Identity

DNA – Ácido desoxirribonucleico

DO – Densidade óptica

EIA – US Energy Information Administration

KEGG – Kyoto Encyclopedia of Genes and Genomes

MALDI-TOF - Matrix Assisted Laser Desorption Ionization –Time Of Flight

MLSA – Multi-Locus Sequencing Analysis

MLST – Multi-Locus Sequencing Typing

MUM – Maximal Unique Matches

PCB – Bifenil Policlorados

PCR-RFLP – Polymerase Chain Reaction – Restriction Fragment Length Polymorphism

RDP – Ribosomal Database Project

RFLP – *Restriction Fragment Length Polymorphism*

RNA – Ácido ribonucleico

rRNA – Gene que codifica para o ribossomo

SSU rRNA – Sequência do gene do ribossomo que codifica a subunidade menor

LISTA DE SÍMBOLOS

T_m – Temperatura de dissociação

$^{\circ}\text{C}$ – Graus Celsius

mg/L – Miligramas por litro

g/L – Gramas por litro

$^{\circ}$ – Graus

' – Minutos

" – Segundos

mM – Milimolar

V – Volts

mL – Mililitros

X – Vezes

% – Porcentagem

μL – Microlitro

U/ μL – Unidades por microlitro

$\mu\text{g/mL}$ – Microgramas por mililitro

g – Gramas

pheS – Gene que codifica a fenilalanil – tRNA sintase

rpoA – Gene que codifica a subunidade α da RNA polimerase

pb – Pares de base

ng – nanogramas

rpm – Rotações por minuto

ppm – Partes por milhão

μm – Micrômetros

LISTA DE FÓRMULAS

Fórmula Molecular	Nome Químico
MgCl ₂	Cloreto de magnésio
CH ₃ COOK	Acetato de potássio
MnCl ₂	Cloreto de manganês
KCl	Cloreto de potássio
CaCl ₂	Cloreto de cálcio
NaOH	Hidróxido de sódio
Na-MOPS	MOPS sal sódio
X-gal	5-bromo-4-cloro-3-indol-β-D-galactosidase
C ₄ H ₁₁ NO ₃	(Tris) 2-amino-2-hidroximetil-propano-1,3-diol
C ₁₀ H ₁₆ N ₂ O ₈	(EDTA) Ácido etilenodiamina tetracético
C ₁₂ H ₂₅ SO ₄	(SDS) Dodecil sulfato de sódio
C ₂ H ₄ O ₂	Ácido acético glacial
NH ₄ Cl	Cloreto de amônio
Na ₂ HPO ₄	Fosfato dissódico
KH ₂ PO ₄	Fosfato monopotássico
MgSO ₄	Sulfato de magnésio
NaCl	Cloreto de sódio
FeSO ₄	Sulfato de ferro
MgSO ₄ .7H ₂ O	Sulfato de magnésio heptahidratado
CaCl ₂ .2H ₂ O	Cloreto de cálcio dihidratado

SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	Importância dos organismos procariotos	17
1.2	Histórico da classificação em procariotos	18
1.3	A presente identificação e classificação microbiana	22
1.4	Metagenômica	26
1.5	BTEX	31
1.5.1	<i>BTEX e contaminação ambiental</i>	31
1.5.2	<i>Degradação microbiana de BTEX</i>	35
1.6	Cubatão.....	39
2	OBJETIVOS	43
3	MATERIAL E MÉTODOS	44
3.1	Locais de coleta	44
3.2	Procedimento de coleta e processamento das amostras	45
3.3	Amplificações genômicas	46
3.3.1	<i>Iniciadores</i>	46
3.3.2	<i>Reação de polimerização em cadeia</i>	48
3.4	Eletroforese	48
3.5	Clonagem de fragmentos	49
3.5.1	<i>Ligação</i>	49
3.5.2	<i>Bactérias competentes</i>	49
3.5.3	<i>Transformação por choque térmico</i>	50
3.5.4	<i>Extração de DNA plasmidial em placa de 96 poços</i>	51
3.6	Sequenciamento automático de nucleotídeos por método Sanger	52
3.6.1	<i>Reação de sequenciamento para método Sanger</i>	52
3.6.2	<i>Reação de precipitação para sequenciamento por método Sanger</i>	52
3.7	Sequenciamento GS FLX Titanium 454.....	53
3.8	Bioinformática	53
3.8.1	<i>Análise das sequências obtidas por sequenciamento pelo método Sanger</i>	53

3.8.2	Análise das sequências obtidas por sequenciamento de alta performance	53
3.8.2.1	Classificação taxonômica	54
3.8.2.2	Classificação funcional do metabolismo de BTEX	56
3.8.2.3	Operons e domínios PFAM	57
3.9	Degradação de BTEX por <i>Bradyrhizobium japonicum</i>	58
4	RESULTADOS	61
4.1	Análise de solo	61
4.2	Identificação de procariotos por amplificação do gene SSU rRNA	62
4.3	Identificação de procariotos por sequenciamento direto de DNA do solo	65
4.3.1	Panorama geral da diversidade de procariotos dos locais estudados ..	65
4.3.2	Comparação de diferentes métodos para refinamento da identificação taxonômica de sequências	68
4.3.3	Identificação metabólica de sequências	78
4.3.4	Degradação de contaminantes por <i>Bradyrhizobium japonicum</i>	94
5	DISCUSSÃO	99
5.1	Locais analisados e processamento do solo	99
5.2	Identificação taxonômica por diferentes métodos	102
5.3	Aprofundamento e comparação da classificação taxonômica	104
5.4	Identificação de subgrupos e comparação metagenômica	105
5.4.1	Filo <i>Actinobacteria</i>	106
5.4.2	Filos <i>Acidobacteria</i> e <i>Firmicutes</i>	106
5.4.3	Filo <i>Proteobacteria</i>	107
5.4.3.1	Classe <i>Alphaproteobacteria</i> e família <i>Bradyrhizobiaceae</i>	108
5.4.3.2	Classe <i>Gammaproteobacteria</i> e família <i>Enterobacteriaceae</i>	110
5.4.3.3	Classes <i>Betaproteobacteria</i> e <i>Deltaproteobacteria</i>	113
5.5	Identificação metabólica de degradação de benzeno e tolueno	114
5.6	Degradação de substâncias aromáticas por <i>Bradyrhizobium</i>	122
5.7	Considerações finais da discussão	125
6	CONCLUSÕES	127
	REFERÊNCIAS	129

APÊNDICE – Versão preliminar do artigo para publicação	143
ANEXO – Análise físico química dos solos contaminado e Perequê	161

1 INTRODUÇÃO

1.1 Importância dos organismos procariotos

Os micro-organismos procarióticos, grupo de seres vivos microscópicos, unicelulares e desprovidos de membrana nuclear, são a primeira e mais abundante forma de vida do planeta. Segundo Whitman e colaboradores, estima-se que no planeta existam entre $4,5 \times 10^{30}$ e $6,4 \times 10^{30}$ procariotos acumulando uma massa de carbono estimada entre 350 e 550 Pg (1 Pg= 1×10^{15} gramas) (WHITMAN; COLEMAN; WIEBE, 1998). Esta forma de vida está espalhada pelos mais diversos ambientes no planeta habitando tanto locais onde há outras formas de vida quanto locais extremos como ambientes de gelo eterno localizados na Antártica (RIVKINA et al., 2000) e fontes hidrotermais nas profundezas dos oceanos com temperatura aproximada de 350 °C (JANNASCH; MOTTTL, 1985). A grande variabilidade metabólica encontrada nos procariotos é uma das características que confere ao grupo a capacidade de sobreviver nos mais variados ambientes. A ampla gama de substâncias orgânicas e inorgânicas utilizadas para a produção de energia confere a capacidade de colonização e modificação de diferentes habitats. A atividade procariótica também é essencial para diversos ciclos biogeoquímicos como os ciclos do nitrogênio, oxigênio, carbono e enxofre.

Além da ubiquidade, o que por si só confere ao grupo importância ímpar, os procariotos são considerados o grupo mais diverso do planeta estando muito acima de qualquer outro grupo de seres vivos. O número de espécies de procariotos estimada ainda é controversa, mas está na ordem de 10^5 a 10^7 (WHITMAN; COLEMAN; WIEBE, 1998). Whitman e colaboradores estimaram também a abundância dos procariotos em diferentes ambientes, mostrada na Tabela 1

Tabela 1 – Estimativa da abundância de procariotos em diferentes ambientes

Ambiente	Número de células procarióticas x 10²⁸
<i>Ambientes aquáticos</i>	12
<i>Oceano profundo</i>	355
<i>Solo (superfície)</i>	26
<i>Subsolo</i>	25-250
<i>Total</i>	415-640

Fonte: (WHITMAN; COLEMAN; WIEBE, 1998).

1.2 Histórico da classificação em procariotos

Apesar de ser o maior grupo de seres vivos existentes, os procariotos não eram conhecidos até o século XVIII. Serem microscópicos, extremamente simples e não disporem de informações fósseis são fatores que contribuíram para esse anonimato. Inicialmente os organismos procariotos foram tratados como formas pleiotrópicas do mesmo organismo. A primeira classificação dos procariotos aconteceu nesse século a partir de observações morfológicas das células que levavam em consideração os diferentes formatos (ROSSELLO-MORA; AMANN, 2001).

O grande avanço no estudo dos organismos procariotos aconteceu com o advento da cultura *in vitro* no século XIX (ROSSELLO-MORA; AMANN, 2001). Por meio de uma cultura pura, a morfologia tanto das células quanto das colônias pôde ser melhor estudada. Quando comparadas, diferenças suficientes foram observadas para o início de uma classificação. Os primórdios da identificação microbiana foram baseados em observações fenotípicas como morfologia, temperatura e pH de crescimento, (KAMPFER; GLAESER, 2012) e apesar de serem observações básicas, foram o suficiente para a criação de seis gêneros procarióticos dentro do reino *Plantae* em 1872 (SCHLEIFER, 2009).

No final do século XIX e início do século XX, muitas publicações com a descrição de diversos gêneros bacterianos apareceram. Estas publicações descreviam ecologia, morfologia e bioquímica dos organismos identificados. Em 1909 houve a primeira tentativa de uma relação genealógica entre os procariotos. Outras formas de classificação surgiram até que em 1923 foi publicada a primeira edição do Bergey's Manual a qual continha a primeira chave de identificação de procariotos (ROSSELLO-MORA; AMANN, 2001). Esta edição e edições subsequentes deste manual tentaram

unificar a classificação bacteriana, que por falta de uma regra geral, contato entre os pesquisadores e divergências na nomenclatura, muitas vezes acabava por identificar com nomes diferentes um mesmo procarioto. Exemplo disso foi o caso da *Pseudomonas stutzeri* que também era classificado como *Bacillus denitrificans* II em 1895 (VAN NIEL; ALLEN, 1952).

O crescente uso de características morfológicas, fisiológicas e bioquímicas para a identificação dos procariotos levou ao surgimento da taxonomia numérica, um tipo de classificação baseada em um grande número de características tanto morfológicas quanto metabólicas. Possuía um caráter qualitativo e gerava uma quantidade muito grande de dados comparáveis. Características como pigmentação, local da inserção de flagelo, utilização de fontes de carbono, utilização de aminoácidos e suscetibilidade a antibióticos eram considerados para essa classificação. Cada uma dessas características apresentavam diversos subitens chegando a dezenas de critérios de classificação (MOORE et al., 2010; VALDERRAMA et al., 1991). A taxonomia numérica tinha como objetivo agrupar bactérias que apresentavam características semelhantes possibilitando uma análise taxonômica mas não filogenética dos procariotos. O uso da taxonomia numérica foi facilitada pelo desenvolvimento computacional e possibilidade de análises multivariadas, o que permitiu a comparação do grande número de dados (ROSSELLO-MORA; AMANN, 2001; SCHLEIFER, 2009).

Na década de 1960 com o aumento do conhecimento das propriedades do DNA e desenvolvimento de técnicas de biologia molecular as características genéticas passaram também a ser utilizadas como critérios de classificação juntamente com as características morfológicas, o que abriu novas portas na classificação dos procariotos (GUPTA; GRIFFITHS, 2002).

O método clássico de comparação do material genético entre diferentes organismos é a comparação do conteúdo G+C presente no genoma. Essa comparação consiste em uma medida indireta deste conteúdo por meio da desnaturação do material genético. Genomas ricos em G+C tendem a desnaturar a temperaturas mais altas, enquanto genomas pobres desnaturam a temperaturas mais baixas (MARMUR; DOTY, 1962). Portanto, diferentes pontos de desnaturação de genomas poderiam ser utilizados como informação para diferenciação de espécies. Com a evolução das observações a respeito dos conteúdos G+C em diferentes genomas, foi estabelecido que dois procariotos que compartilham 90% de conteúdo G+C podem ser classificados dentro do mesmo gênero e se o compartilhamento for maior que 95%, eles podem ser

considerados da mesma espécie (ROSSELLO-MORA; AMANN, 2001). Apesar de ser uma característica utilizada até hoje na composição do conjunto de dados para identificação procariótica, o conteúdo G+C não é suficientemente robusto para sustentar uma classificação mais precisa (MOORE et al., 2010).

O refinamento da técnica de conteúdo G+C se deu pelo desenvolvimento de uma nova técnica: a hibridação de ácidos nucleicos, DNA-DNA e DNA-RNA. Esta técnica é baseada na capacidade de desnaturação e renaturação dos ácidos nucleicos. Quando aquecidos a temperaturas por volta de 100 °C, as moléculas de DNA têm sua dupla-fita separada e, em condições controladas, podem ter suas fitas religadas. Quando as moléculas de DNA de dois organismos diferentes são misturadas, desnaturadas e em seguida renaturadas, pode ocorrer a formação de moléculas híbridas, com maior ou menor complementariedade, dependendo do quão similar as sequências de DNA dos organismos são.

Duas taxas são calculadas com base na homologia das moléculas de DNA e na temperatura de renaturação da técnica de hibridação de ácidos nucleicos. A RBR, que reflete a taxa de moléculas híbridas formadas, e a ΔT_m , que corresponde à variação entre a temperatura de renaturação da dupla-fita híbrida e da dupla-fita não híbrida (ROSSELLO-MORA; AMANN, 2001). Para que um procarioto pertença à mesma espécie de acordo com a técnica de hibridação de ácidos nucleicos, é preciso que ele apresente pelo menos 70% de RBR e ΔT_m de até 5 °C (WAYNE, 1988). Um problema do método é que ele depende da escolha de um organismo de referência, que precisa ser próximo filogeneticamente do organismo a ser identificado. A técnica de hibridação DNA-DNA ainda é válida e é atualmente uma das melhores para identificação de espécies. Ela pode teoricamente ser aplicada a qualquer procarioto e é um método de identificação de espécies não conhecidas. (GORIS et al., 2007; ROSSELLO-MORA; AMANN, 2001).

Com o avanço das técnicas de biologia molecular e seu uso para a determinação das relações filogenéticas entre os procariotos passou-se a uma busca constante por características genéticas que auxiliassem no refinamento da classificação dos procariotos. A descoberta dos genes que codificam para as subunidades ribossômicas foi uma importante ferramenta para esse fim.

O ribossomo é a organela celular envolvida na tradução das informações do DNA em proteínas. É constituído por três subunidades de diferentes tamanhos, o que pode ser evidenciado por ultracentrifugação. Nos seres procarióticos as subunidades

encontradas são a 5S, 16S e 23S codificadas por genes específicos de tamanhos aproximados de 120, 1650 e 3300 nucleotídeos respectivamente (ROSSELLO-MORA; AMANN, 2001).

A primeira abordagem para a classificação dos procariotos a partir de estudos do gene que codifica para a subunidade 16S ou SSU rRNA foi realizada por Woese e colaboradores na década de 1970. Eles também iniciaram o banco de dados de sequências desse gene em diferentes organismos. Este gene se mostrou muito propício à classificação procariótica, pois é encontrado em todos os seres vivos, apresenta considerável quantidade de informação genética e é conservado entre as espécies (SCHLEIFER, 2009). Por não ser propenso a transferências laterais e apresentar variação em sua sequência de organismo para organismo, ele se mostrou efetivo também para o cálculo de distâncias evolutivas. As variações encontradas nesses genes refletem a própria evolução do organismo sendo pouco afetado por pressões seletivas (WOESE, 1987).

A classificação de um organismo baseada no gene SSU rRNA ocorre a partir da comparação da sequência de nucleotídeos desse gene com a sequência de outros organismos. Sequências mais similares identificam organismos mais próximos evolutivamente. Para que dois organismos pertençam à mesma espécie deve haver 97% de identidade entre as sequências, e para que pertençam ao mesmo gênero, 95% de identidade (TINDALL et al., 2010). Essas porcentagens foram definidas baseadas na técnica de hibridação de DNA-DNA. Foi observado que organismos da mesma espécie, que compartilhavam valor igual ou maior do que 70% na técnica de hibridação, apresentavam valores de 97% de similaridade nas sequências de seus SSU rRNAs.

A facilidade de sequenciamento e rapidez na identificação dos procariotos determinaram a técnica da comparação da sequência de genes SSU rRNA como um dos melhores e mais abrangentes métodos de classificação procariótica. No entanto, uma das críticas recorrentemente feita a esse método é a sua baixa para identificação de espécies (GUPTA; GRIFFITHS, 2002; ROSSELLO-MORA; AMANN, 2001; SCHLEIFER, 2009; TINDALL et al., 2010). Na época em que as árvores filogenéticas baseadas na sequência do SSU rRNA foram iniciadas, existiam poucas ramificações e a separação entre os diferentes grupos era melhor visualizada. Com cada vez mais sequências ribossomais adicionadas aos bancos de dados, os ramos das árvores filogenéticas foram sendo preenchidos e a classificação dos organismos foi ficando cada vez mais difícil e imprecisa (GUPTA; GRIFFITHS, 2002). Começaram a surgir

casos de espécies diferentes com sequências de genes SSUrRNA com semelhanças maiores do que 97%, causando imprecisões nas classificações, como demonstrado pelo método de hibridação DNA-DNA, que apresenta maior resolução para definição de espécies procarióticas. Fox e colaboradores mostraram claramente a imprecisão do método de SSU rRNA quando compararam as espécies *Bacillus psychrophilus* e *Bacillus globisporus*. O resultado para o método SSU rRNA foi de 99,8% de identidade entre as duas espécies, sugerindo que esses dois organismos poderiam ser agrupados como uma mesma espécie. Já o método de hibridação de DNA mostrou uma taxa de hibridação de apenas 50%, o que significa que esses dois organismos claramente não pertencem à mesma espécie (FOX; WISOTZKEY; JURTSUK, 1992).

Gupta e Griffiths (2002) ainda questionam a classificação de 90% a 95% das bactérias em somente quatro grupos, como indicado pelas análises do gene SSU rRNA, e indicam que esse resultado deve ser reflexo da baixa resolução desse método.

Outro fato que pode ser decisivo para a identificação de espécies, pelo método de SSU rRNA, é a variação das diferentes cópias do gene no mesmo genoma. As diferentes cópias ribossomais podem ter variações de 1% a 2%, e às vezes até mais, o que pode dificultar uma classificação mais precisa do procarioto (SCHLEIFER, 2009).

Os procariotos compõem um grupo muito diverso e sua classificação ainda é controversa. O avanço das técnicas de biologia molecular, estão possibilitando um maior conhecimento desses organismos, porém em muitos casos, esse conhecimento está colocando à prova métodos de classificação que outrora cumpriam bem seu papel. A cada novo aporte de informação nos bancos de dados, novas falhas são descobertas, assim, o desenvolvimento de técnicas mais precisas é necessário.

1.3 A presente identificação e classificação microbiana

Para todos os fins, desde a pesquisa básica até a aplicada, a classificação dos organismos procarióticos é de extrema importância para a comunicação entre as partes. Para esta finalidade, o sistema de classificação e a nomenclatura de Linneu foram adotados para identificação e filogenia desses organismos, nos mesmos moldes estabelecidos para outros grupos de seres vivos (MOORE et al., 2010). A nomenclatura se mostra adequada para a identificação de procariotos, mas é necessária a definição de melhores critérios e parâmetros para que se estabeleça a filogenia dos procariotos.

O conceito de espécie dentro dos organismos procariotos é bastante questionável. Antes do surgimento das técnicas moleculares, por volta da década de 1950, a identificação de espécies era realizada com base em observações morfológicas de organismos isolados, o que culminava em uma classificação que continha poucas informações, e com uma visão subjetiva (GEVERS et al., 2005; VAN NIEL; ALLEN, 1952). Nesta época, surgia um conceito de espécie que vigorou por muito tempo, até aproximadamente o ano 2000, e que ainda tem influência na classificação procariótica. O conceito criado por Gordon e Mihm em 1962 definia espécie procariótica como *“um grupo de linhagens isoladas recentemente, mantidas in vitro por diferentes períodos de tempo; e suas variantes, as quais possuem em comum um conjunto de características correlacionáveis que as separam de outras linhagens”* (GORDON; MIHM, 1962). Este conceito de espécie, devido à época em que foi criado, refere-se somente a características morfológicas, não levando em conta todo o ramo da classificação microbiana baseada nos dados de DNA.

Rosseló-Morá e Amann, em 2000, reformularam o conceito de espécie procariótica incluindo também os métodos de comparação de material genético como métodos válidos para determinação de espécies. Os autores sugeriram como espécie um grupo de linhagens que compartilham alto grau de similaridade com relação a diversas características. Essa definição foi aceita pelo comitê de bacteriologia, sendo publicada da seguinte forma: *“espécie é uma categoria que engloba (preferencialmente) um grupo coerente, do ponto de vista genômico, de indivíduos isolados/linhagens que compartilham um alto grau de similaridade em (muitas) características independentes, comparativamente testadas sob condições altamente padronizadas”* (ROSSELLO-MORA; AMANN, 2001; STACKEBRANDT et al., 2002). O documento, apesar de não instituir um método único para identificação de espécies, sugere que observações de caráter comparativo sejam feitas a características independentes. Esta comparação engloba características tanto genóticas quanto fenotípicas, e o cruzamento destes dados para a determinação de espécies. Este método multivariável de classificação procariótica foi denominado de taxonomia polifásica.

A taxonomia polifásica, sendo válidos também os nomes identificação polifásica e classificação polifásica, foi primeiramente proposta por Colwell em 1968 (VANDAMME et al., 1996). Ela define um sistema de classificação baseado em informações genóticas e fenotípicas como forma de identificação procariótica.

Com relação a análises fenotípicas podemos destacar o uso da tecnologia de MALDI-TOF ou espectrometria de massas. Essa tecnologia utiliza cultura isolada, lisado celular ou extratos bacterianos diretamente e permite o estudo do material genético e proteínas celulares tanto isoladas quanto em células (LAY, 2001).

Esse método aplicado a células em suspensão ou mesmo em colônias gera um perfil espectral exclusivo de cada tipo de procarioto, o qual pode ser utilizado para identificação. Possui a vantagem de ser bastante rápido para análises, pois não requer, necessariamente, nenhuma técnica complementar de preparação da amostra, característica bastante vantajosa em identificação de bactérias patogênicas. Para esse tipo de bactéria, a técnica funciona relativamente bem, já que se dispõe de um banco de dados bastante informativo (CARBONNELLE et al., 2012), podendo-se chegar a índices de identificação de 96% para quase 700 espécies (KHOT et al., 2012). Para bactérias não patogênicas, a identificação por MALDI-TOF não é precisa devido à falta de informações dos espectros em bancos de dados e às próprias variações encontradas nas culturas, dificultando a identificação (LAY, 2001).

A necessidade de cultura é outro fator que dificulta a identificação de procariotos. Sem cultura não se tem um espectro, e sem o espectro não é possível a identificação. Assim, faz-se necessário o desenvolvimento de novas técnicas de cultura ou a descoberta de características que permitam a identificação sem a necessidade do cultivo.

Outro método de análise fenotípica é a classificação quimiotáxica. Esta classificação consiste no estudo e identificação de componentes estruturais da célula como parede celular, membrana plasmática e citoplasma. Dentro deste escopo, estruturas específicas são exploradas como lipopolissacarídeos, poliaminas, lipídeos, peptidoglicanos e benzoquinonas para a diferenciação e identificação de procariotos (TINDALL et al., 2010). A simples identificação de presença ou ausência de determinada estrutura pode já ser informativa para a classificação de um organismo. Um exemplo disso são as ubiquinonas que são restritas às classes Alphaproteobacteria, Betaproteobacteria e Gammaproteobacteria, e a simples constatação de sua presença já restringe a identificação do procarioto a essas três classes (TINDALL et al., 2010).

A quimiotaxonomia pode fornecer maiores informações e ser mais precisa do que testes fisiológicos e bioquímicos, porém sofre com a falta de resolução para níveis taxonômicos mais específicos. Divergências de resultados entre laboratórios também

podem ocorrer devido ao comprometimento da reprodutibilidade relacionado a diferenças de equipamentos e protocolos (MOORE et al., 2010).

Completando a taxonomia polifásica, os métodos genotípicos de identificação existem em maior número e são os mais utilizados para a classificação procariótica. Muitos desses métodos são baseados em reação de polimerização em cadeia, hidrólise do material genético por enzimas de restrição e sequenciamento de DNA. Estes métodos evidenciam diferenças e similaridades do material genético, o que ajuda na identificação de procariotos e no estabelecimento das relações filogenéticas entre eles.

As técnicas como AFLP, PFGE e PCR-RFLP são baseadas na digestão dos genomas por enzimas de restrição e na amplificação dos fragmentos formados evidenciando diferenças entre os genomas que podem ser utilizadas para classificação (VANDAMME et al., 1996).

Outros métodos de análise genômica como MLSA e MLST se baseiam em genes como *rpoA* e *pheS*, conhecidamente conservados e com valor para diferenciações entre espécies. As duas técnicas evidenciam diferenças existentes nesses genes usadas na identificação e construção de relações filogenéticas entre os organismos, já que, como o gene SSU rRNA, esses genes são menos propensos a transferências laterais e a pressões seletivas. Para realização desta comparação, iniciadores específicos, técnicas de reação de polimerização em cadeia e sequenciamento de nucleotídeos são utilizados. Este método abrange muito bem níveis mais específicos de identificação, porém, perde resolução para níveis mais amplos como filos e classes (MAIDEN et al., 1998; NASER et al., 2005).

Com a diminuição de custo de sequenciamento e incremento da capacidade bioinformática, os projetos de sequenciamento de genomas completos estão em plena ascensão. Em 2010, estavam depositados no banco de dados Genome OnLine Database (GOLD); <<http://www.genomesonline.org>>, 2400 genomas completos (MOORE et al., 2010). Atualmente, neste mesmo banco se encontram 3788 genomas completos depositados, o que representa um aumento de 50% em apenas 2 anos. Número bastante expressivo desde a primeira publicação do banco de dados há mais de 10 anos, com apenas 21 genomas (KYRPIDES, 1999).

O sequenciamento de genomas completos é uma das técnicas mais promissoras para classificação de procariotos (GUPTA; GRIFFITHS, 2002; MOORE et al., 2010; WAYNE, 1988). Com o genoma completo, comparações totais do material

genético entre diferentes organismos se tornam possíveis. Análises entre as sequências do genoma dos organismos poderão ser realizadas *in silico*, com alta padronização e qualidade. Duas técnicas que surgiram baseadas no sequenciamento completo de genomas são as técnicas ANI (KONSTANTINIDIS; TIEDJE, 2007) e MUM (DELOGER; EL KAROUI; PETIT, 2009). Ambas se baseiam em comparações nucleotídicas entre regiões em comum dos genomas e se mostraram bastante eficientes na construção de relações filogenéticas.

Não há dúvidas de que entre as técnicas genotípicas de identificação, a mais utilizada é a técnica de análise da sequência do gene SSU rRNA. Apesar das críticas, ainda é uma das melhores ferramentas de comparação que existem atualmente e uma das poucas com a capacidade de fornecer dados para construção de filogenia. Um fato que possibilita esta capacidade é o enorme banco de dados de sequências do SSU rRNA que existe atualmente. Os bancos de dados RDP (<http://rdp.cme.msu.edu/>) e SILVA (<http://www.arb-silva.de/>) possuem respectivamente 2.578.902 e 3.587.011 sequências desse gene depositadas.

Os estudos genotípicos dos procariotos estão se voltando cada vez mais a análises computacionais e a comparações em bancos de dados *in silico*. Técnicas antes realizadas em laboratório e que despendiam grande quantidade de tempo podem ser feitas atualmente em minutos e com uma quantidade de dados muito maior. Esse avanço, tanto no número de dados quanto na capacidade de processamento das informações, levou ao aparecimento de técnicas de estudos de comunidades procarióticas inteiras sem a necessidade de cultivo. Atualmente podemos ter acesso tanto à composição quanto ao tipo de metabolismo dessas comunidades, uma porta de estudos para um grande potencial biotecnológico.

1.4 Metagenômica

Como já mencionado no item 1.1, o grupo dos procariotos é o de maior diversidade e abundância do planeta. Mesmo com todos os avanços tecnológicos, o nosso conhecimento desse grupo de organismos ainda é muito pequeno. A necessidade de cultivo desses organismos para estudo dificulta ainda mais nosso aprofundamento no conhecimento. Estima-se que seja possível o cultivo de apenas 1% de todos os procariotos (RAJENDHRAN; GUNASEKARAN, 2008; SCHLOSS; HANDELSMAN, 2003; STREIT; RA, 2004;). Na tentativa de superar essa dificuldade e

poder aprofundar o conhecimento desse grupo, técnicas moleculares de estudo sem a necessidade de cultivo de organismo foram desenvolvidas. Extração de material genético, clonagem, sequenciamento e expressão gênica foram aplicados diretamente a comunidades microbianas do ambiente, o que resultou no aparecimento de um novo tipo de estudo chamado de metagenômica.

A metagenômica pode ser definida como uma análise funcional e genômica de uma comunidade microbiana do ambiente sem a necessidade de cultivo. Esse nome foi utilizado oficialmente na primeira conferência sobre o assunto (Metagenomics 2003) realizada na cidade de Darmstadt na Alemanha (REISENFELD; SCHLOSS; HANDELSMAN, 2004).

O estudo metagenômico está baseado inteiramente no estudo do material genético de uma comunidade microbiana, e o isolamento de ácidos nucleicos é o primeiro passo. Este isolamento pode ser realizado diretamente no ambiente a partir de uma lise celular total, ou indiretamente a partir de separação das células presentes na amostra que então segue para a lise celular (RAJENDHRAN; GUNASEKARAN, 2008). O material genético é purificado e recuperado, e, a partir deste momento, empregado para o estudo, que pode ser dirigido a diversos caminhos de acordo com o interesse da pesquisa. Em uma abordagem funcional, o material genético extraído do ambiente é fragmentado e clonado em vetores de expressão e transfectado em organismos receptores. Estes organismos expressando os clones são então testados para a função de interesse, e os clones positivos são selecionados para um estudo mais aprofundado de sua sequência nucleotídica (HESS et al., 2011; REISENFELD; GOODMAN; HANDELSMAN, 2004). Este tipo de abordagem é bastante laboriosa e apresenta baixo rendimento.

Um exemplo desse tipo de abordagem metagenômica é trabalho realizado por Riesenfeld e colaboradores (2004), cujo foco eram genes de resistência a antibióticos. Nas 4,1 gigabases (10^9 bases) clonadas, foram encontrados apenas 10 clones que se mostraram resistentes aos antibióticos testados. Para otimizar o trabalho de busca quando o foco é a abordagem funcional, um passo de enriquecimento de material pode ser realizado anteriormente à extração de DNA.

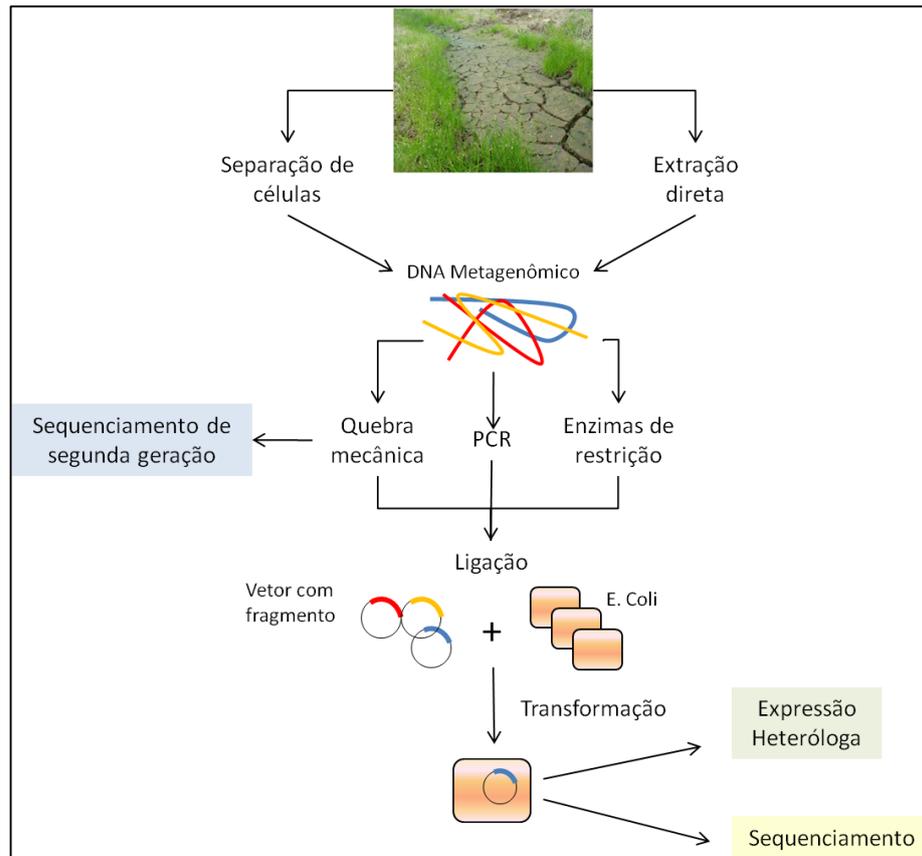
Outro exemplo foi o estudo de Hess e colaboradores (2011), cujo interesse estava em enzimas de degradação de biomassa de origem vegetal em rúmen bovino. Para o estudo, uma massa vegetal foi introduzida no rúmen e apenas procariotos que

estavam aderidos e processando o material vegetal foram selecionados para a análise metagenômica .

A outra abordagem metagenômica é a baseada em sequência ou sequência dirigida (SCHLOSS; HANDELSMAN, 2003). Ela consiste no uso de iniciadores conhecidos e na amplificação direta do DNA metagenômico. É comum a aplicação dessa abordagem para identificação microbiana no ambiente a partir do uso de iniciadores específicos para o gene SSU rRNA. Com isso, um grande número de sequências deste gene em diversos organismos pode ser amplificado e a comparação dessas sequências, além de permitir a identificação dos organismos que constituem a comunidade, pode ser utilizada para estabelecer relações filogenéticas e identificar sequências que não são identificadas nos bancos de dados, ou seja, possíveis novas espécies, revelando uma diversidade que não poderia ser alcançada somente por meio de comparações com o banco de dados (BRUCE et al., 2010).

A abordagem da sequência dirigida ainda permite a recuperação de genes que codificam enzimas de interesse. Através de comparações de sequências conhecidas, regiões conservadas podem ser identificadas e utilizadas como molde para iniciadores. Genes presentes no metagenoma que possuem a mesma função ou uma função parecida podem ser recuperados por amplificação, clonados e ter sua atividade avaliada (COURTOIS et al., 2003). A Figura 1 esquematiza as principais etapas de um estudo metagenômico.

Figura 1 - Representação das etapas de um estudo metagenômico



Uma grande diversidade de procaríotos foi descoberta com o avanço dos estudos metagenômicos. No entanto, a capacidade de geração de informações era ainda muito limitada e demorada. As análises metagenômicas envolvem sempre uma etapa de sequenciamento de nucleotídeos. O sequenciamento utilizado para os estudos em questão era o sequenciamento de nucleotídeos pelo método Sanger (SANGER; COULSON, 1975), que é dependente de clonagem e gera uma quantidade de dados que para a metagenômica se mostra insuficiente.

Em 1998 Ronaghi e colaboradores desenvolveram uma técnica de sequenciamento de nucleotídeos que revolucionou a metagenômica (RONAGHI; UHLEN; NYREN, 1998). Este sequenciamento baseado em emissão de luz recebeu o nome de pirosequenciamento. A técnica foi a base para o desenvolvimento da terceira geração de sequenciadores e aumentou exponencialmente a quantidade de dados gerada. Enquanto o sequenciamento pelo método Sanger pode gerar por corrida cerca de 1000 bases, as tecnologias baseadas no pirosequenciamento podem chegar à ordem de 10^9 a 10^{12} , dependendo da plataforma utilizada (DANIEL, 2005).

Além de aumentar a capacidade de informação gerada, o sequenciamento de terceira geração promoveu novas abordagens de análises metagenômicas, alcançadas pela possibilidade de sequenciamento de ácidos nucleicos sem a necessidade de clonagem.

Uma das novas abordagens que puderam ser desenvolvidas graças à técnica de pirosequenciamento foi o sequenciamento ambiental direto em larga escala. Esta técnica consiste no sequenciamento direto do material genético encontrado em diferentes habitats. Por não passar por clonagem, não é necessária a preparação de bibliotecas genômicas em vetores ou desenvolvimento de iniciadores para estudo de genes. Este método pode gerar dados que representam toda a comunidade microbiana, possibilitando a comparação entre comunidades de diferentes locais (QUAISER et al., 2011) ou de diferentes datas (MARCHETTI et al., 2012).

O primeiro trabalho utilizando essa abordagem metagenômica foi publicado em 2004 e teve como local de estudo o mar de sargaço (VENTER et al., 2004). Venter e colaboradores foram os pioneiros a aplicar a tecnologia de pirosequenciamento diretamente em amostras ambientais. Pela primeira vez pôde-se mostrar tão detalhada e profundamente uma comunidade microbiana aquática abordando-se diferentes características. Várias faces da metagenômica, reveladas a partir do sequenciamento em larga escala, se apresentam neste trabalho. Identificação microbiana, identificação metabólica, montagem de genomas e identificação e comparação de genes de interesse puderam ser abordadas em um único estudo, mostrando todo o potencial inovador e promissor da técnica.

A partir deste primeiro trabalho, muitos outros ambientes foram sendo explorados como os tratos digestórios de animais, ambientes pelágicos e solos. Habitats específicos como fezes (GILL et al., 2006) e pele humana (GRICE; SEGRE, 2011) também já foram abordados, e esses trabalhos além de gerar conhecimento sobre procariotos em geral, possuem também um caráter terapêutico, pois ao se entender melhor a dinâmica microbiana envolvida, melhorias no combate a infecções, previsão de doenças e tratamentos mais específicos e eficazes podem ser desenvolvidos.

Entre toda a diversidade microbiana existente, os procariotos ambientais são ainda uma grande incógnita para a ciência e, por esse motivo, muitos trabalhos de metagenômica focam comunidades microbianas ambientais na tentativa de entender quem são e o que fazem. Estes estudos não se restringem apenas a ambientes

inóspitos, desabitados e desprovidos de contato com seres humanos, mas também incluem ambientes que sofrem com atividades antropocêntricas e acabam sendo contaminados. Nesses ambientes, comunidades microbianas conseguem sobreviver e muitas vezes conseguem até utilizar o contaminante do local como fonte de carbono produzindo energia para manutenção e multiplicação celular (VOGEL, 1996). O potencial de degradação de contaminantes por procariotos é um dos focos de estudos metagenômicos graças ao seu grande potencial biotecnológico e de aplicabilidade, podendo se tornar uma ferramenta biotecnológica.

1.5 BTEX

1.5.1 BTEX e contaminação ambiental

Os xenobióticos são conhecidos por causarem impactos ambientais modificando a homeostase do ambiente e influenciando diretamente a interação entre os seres vivos e o meio onde vivem. Além do impacto ambiental causado pelo aporte de um elemento estranho ao ambiente, muitas vezes os xenobióticos são tóxicos, podendo ser os responsáveis direto pela mortalidade de seres vivos, ocasionando queda tanto no número de indivíduos quanto no número de espécies de um local afetado. O petróleo e seus derivados se posicionam em local de destaque entre os xenobióticos de maior importância com relação a impactos ambientais.

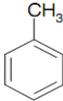
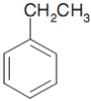
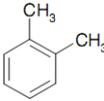
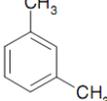
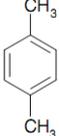
O petróleo é um recurso natural amplamente explorado e utilizado ao redor do mundo, sendo a principal fonte energética do planeta. Segundo a EIA, em 2011 foram produzidos em média mais de 87 bilhões de barris de petróleo por dia no mundo e os maiores consumidores desta fonte energética são, em ordem, Estados Unidos, China, Japão, Índia, Rússia e Brasil (EIA, 2013).

O petróleo raramente é utilizado em sua forma natural. Seu uso se dá após seu refino, que é o processo de separação de seus componentes, uma diversidade de substâncias orgânicas conhecidas como hidrocarbonetos. Uma vez separados, os componentes do petróleo são utilizados nas mais diversas produções industriais, sendo utilizados como pavimentação, solventes, combustíveis, lubrificantes, plásticos, borrachas e fibras sintéticas.

Entre os produtos derivados do petróleo, um grupo de substâncias tem chamado a atenção por suas características físico-químicas, toxicidade, taxa de produção e impacto no ambiente: o grupo denominado de BTEX. A sigla BTEX se refere ao

conjunto que agrupa benzeno, tolueno, etilbenzeno e xilenos. Suas estruturas moleculares e características químicas se encontram na Tabela 2.

Tabela 2 – Características dos hidrocarbonetos do grupo BTEX

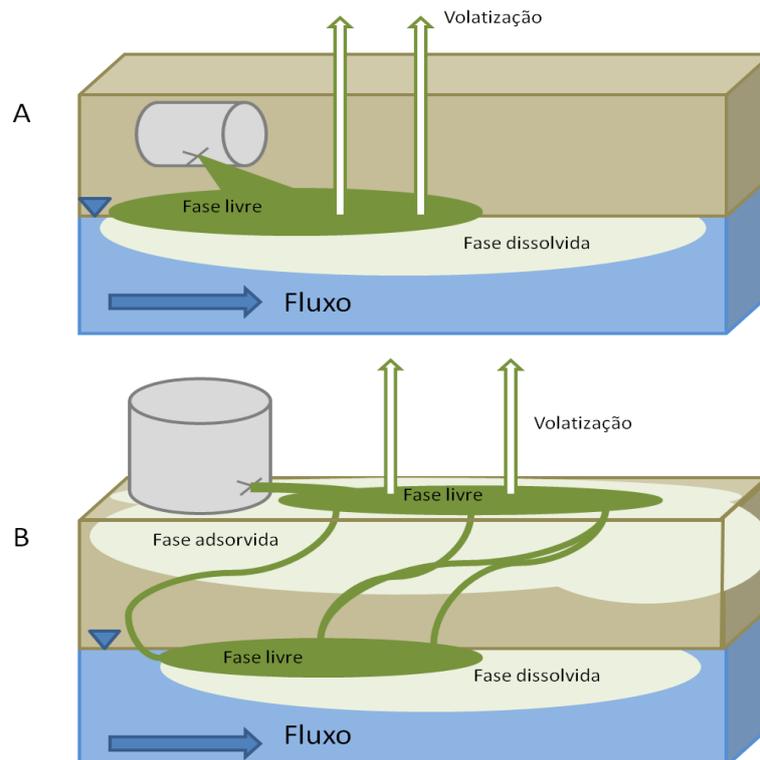
						
	Benzeno	Tolueno	Etilbenzeno	<i>orto</i> -Xileno	<i>meta</i> -Xileno	<i>para</i> -Xileno
Fórmula molecular	C ₆ H ₆	C ₇ H ₈	C ₈ H ₁₀	C ₈ H ₁₀	C ₈ H ₁₀	C ₈ H ₁₀
Peso molecular	78,11	92,1	106,2	106,2	106,2	106,2
Temperatura de fusão (°C)	5,5	-95	-95	-25	-48	13
Temperatura de ebulição (°C)	80,1	110,8	136,2	144,4	139,0	138,4
Solubilidade em água a 20 °C (mg/L)	1780	515	152	175	200	198

Fonte: (Baseado em AGTEREN; KEUNING; JANSSEN, 1998; WEELINK; EEKERT, 2010).

Estes hidrocarbonetos derivados do petróleo são produzidos em quantidades equivalentes a 10³ toneladas por ano e são utilizados largamente em processos industriais como solventes e matérias-primas para produção de pesticidas, plásticos e fibras sintéticas (JINDROVA et al., 2002). Esse grupo de hidrocarbonetos também reúne componentes da gasolina, que pode conter de 10% a 40% destes hidrocarbonetos aromáticos em sua composição (FINOTTI; CAICEDO; RODRIGUEZ, 2001). Devido à importância desses hidrocarbonetos em processos industriais e ao fato de estarem presente em produtos difundidos mundialmente, contaminações ambientais em diversas partes do mundo envolvendo BTEX são comuns (ALFREIDER; VOGHT, 2007; BALDWIN et al., 2009; CAVALCA; DELL'AMICO; ANDREONI, 2004; MAZZEO et al., 2010). As contaminações com BTEX ocorrem principalmente por vazamentos, sejam dos componentes puros de tanques de armazenamentos ou encanamentos de indústrias, ou de gasolina de postos de combustíveis. Os postos de combustíveis têm grande contribuição nas contaminações com BTEX. Esse fato se deve à grande

quantidade desse tipo de local de distribuição e ao fato de a gasolina estar amplamente relacionada à vida cotidiana (FINOTTI; CAICEDO; RODRIGUEZ, 2001). Um recente estudo realizado pela Secretaria do Meio Ambiente do Estado de São Paulo revelou que 79,5% das contaminações do Estado de São Paulo tinham como origem postos de combustíveis (SMA, 2012). A Figura 2 exemplifica dois tipos de vazamentos que ocorrem no ambiente.

Figura 2 - Esquema de diferentes modos de contaminação do ambiente por BTEX



A) Vazamento que ocorre geralmente em tanques de postos de combustível onde os tanques de armazenamento são subterrâneos. Esse tipo de contaminação afeta o subsolo e a camada saturada (azul). Existe a formação de uma fase livre e de uma fase dissolvida que percorre grandes distâncias, podendo contaminar outros locais.

B) Exemplo de vazamento de um tanque de armazenamento sobre o solo. Quando ocorre este tipo de vazamento, existe contaminação superficial do solo, formando uma fase livre e uma fase adsorvida. A fase livre superficial pode penetrar no subsolo e formar outra fase livre e uma fase dissolvida na camada saturada do subsolo. Em ambos os casos ocorre contaminação do solo, água e ar através da volatilização do BTEX.

Pela sua solubilidade relativamente baixa em água, o BTEX quando liberado no ambiente pode ser carregado e alcançar longas distâncias. Pode contaminar o lençol freático e comprometer reservas de água de uso humano.

Não só a água é alvo de contaminação por BTEX, o ar também pode ser afetado. Por serem hidrocarbonetos voláteis e por terem baixa tendência de adsorção em matéria orgânica, quando derramados em solo, grande parte evapora atingindo a

atmosfera e pode retornar ao solo carregado pelas águas da chuva (FINOTTI; CAICEDO; RODRIGUEZ, 2001).

No ambiente, não se tem muitas informações das consequências diretas de contaminação por BTEX. Muito mais conhecidas são contaminações por petróleo que causam mortalidade de peixes, aves, mamíferos e podem contaminar amplas áreas pelágicas e costeiras. Contudo, alguns estudos comprovam a toxicidade de BTEX em seres vivos. Já foi demonstrada a toxicidade de BTEX no desenvolvimento de plantas utilizando como modelo as espécies *Sorghum bicolor* e *Cucumis sativum* (AN, 2004). A toxicidade também foi testada em alga verde *Selenastrum capricornutum* (HERMAN; INNIS; MAYFIELD, 1991). Nos estudos citados, em concentrações elevadas, o crescimento das plantas ficou comprometido, e no caso da alga verde foi comprovado danos à integridade da membrana celular causada pela exposição ao BTEX.

Em seres humanos, a toxicidade do BTEX já foi estudada (COSTA et al., 2006; GÉRIN et al., 1998; TRAVIS et al., 1994). A contaminação pode ocorrer através da pele, ingestão ou inalação. Altas concentrações de BTEX podem ocasionar irritação da pele, depressão do sistema nervoso central e problemas respiratórios. A exposição por períodos prolongados pode causar, além dos problemas já citados, problemas hepáticos e renais. O benzeno em particular é uma substância conhecidamente carcinogênica e exposições prolongadas a ele podem causar leucemia (TRAVIS et al., 1994). Doenças de coração, do sistema nervoso central e anemia aplástica também já foram relacionadas à exposição de benzeno (GUIMARÃES, 2005).

Sendo xilenos, tolueno e etilbenzeno substâncias tóxicas e benzeno, além de tóxico, conhecidamente carcinogênico, regulamentações para controle destas substâncias no ambiente existem tanto para atividades industriais que utilizam o BTEX quanto para monitoramento ambiental. A CONAMA, resolução 396 de 2008, determina valores máximos para a potabilidade da água no Brasil com relação à presença de BTEX (Tabela 3).

Tabela 3 – Níveis máximos para a potabilidade da água com relação às concentrações de BTEX (CONAMA 396/2008)

<i>Hidrocarboneto</i>	<i>Valor máximo tolerado µg/L</i>
<i>Benzeno</i>	5
<i>Etilbenzeno</i>	200
<i>Tolueno</i>	170
<i>Xileno</i>	300

A presença de BTEX em praticamente todo o planeta aumenta as chances de contaminação acidental do ambiente colocando em risco a saúde de seres humanos. Suas propriedades químicas permitem uma rápida movimentação pelo solo e água se espalhando rapidamente por grandes áreas. Devido ao seu caráter tóxico, ações de contenção e recuperação ambiental precisam ser tomadas com agilidade em caso de acidentes. Medidas de retirada de BTEX do ambiente existem baseadas em princípios químicos, físicos e biológicos, neste último caso, utilizando-se a capacidade metabólica de procariotos para catabolizar e eliminar a contaminação do ambiente.

1.5.2 Degradação microbiana de BTEX

Seja em solo ou em ambiente aquático, o BTEX quando atinge o meio ambiente causa problemas não apenas no local onde ocorre o derramamento, mas também em locais distantes, podendo se espalhar rapidamente. Devido ao seu potencial tóxico, medidas de retirada do BTEX do ambiente precisam ser tomadas com urgência impedindo o afastamento do contaminante. Estas medidas de retirada de contaminantes do ambiente são denominadas remediação.

A remediação ambiental pode ser aplicada a substâncias nocivas como pesticidas e metais pesados, mas, em sua grande maioria, a remediação é utilizada em casos de contaminação com petróleo e seus derivados. Diversas técnicas de remediação ambiental estão disponíveis atualmente para remediação de solo, água ou ambos. Estas técnicas estão baseadas em princípios físicos, como a retirada do contaminante através de sucção ou escavação; princípios químicos, como oxidação do contaminante transformando-o em um produto menos tóxico; e princípios

biológicos, através da modificação do contaminante por intermédio de ação metabólica de seres vivos (KHAN; HUSAIN; HEJAZI, 2004).

Os meios biológicos de remediação são conhecidos como biorremediação. Esse método utiliza a capacidade dos seres vivos de degradar contaminantes (GENOVESE et al., 2008) transformando-os em produtos de menor toxicidade, ou eliminando-os totalmente do ambiente. Este tipo de remediação vem ganhando cada vez mais espaço entre as técnicas de remediação ambiental. Por não utilizar produtos químicos nem recursos de maquinário e tecnologias que requerem altos investimentos, torna-se uma alternativa viável e vantajosa para o tratamento de locais contaminados (VIDALI, 2001). Em contrapartida, este método é bastante demorado comparado a outros métodos físicos e químicos, e, dependendo da concentração, os seres vivos se tornam incapazes de catabolizar o contaminante. Para um melhor resultado, a utilização de diferentes técnicas é frequente. A biorremediação, utilizada por último, consegue diminuir a concentração do contaminante a níveis dentro do limite tolerável (KHAN; HUSAIN; HEJAZI, 2004).

A biodegradação de BTEX é conhecida desde 1908 quando pela primeira vez Stormer observou a capacidade de crescimento do *Bacillus hexabovorum* em meio de cultura contendo tolueno e xilenos (MAZZEO et al., 2010). Posteriormente, foram descobertas muitas bactérias, de diversos locais, que também apresentavam a capacidade de degradação de BTEX. Já foi constatada a presença destas bactérias tanto na parte superficial de solos contaminados (HENDRICKX et al., 2006) quanto no subsolo (HERRMANN et al., 2008) e em águas subterrâneas (CAVALCA; DELL'AMICO; ANDREONI, 2004). Entre os gêneros bacterianos conhecidos por metabolizar BTEX, o gênero *Pseudomonas* é o de maior representatividade. Este gênero é conhecido por utilizar BTEX como única fonte de carbono, processando estas moléculas por respiração e transformando-as em dióxido de carbono e água (BYRNE; KUKOR; OLSEN, 1995; JOHNSON; OLSEN, 1995). Outros três gêneros já estudados e que possuem a capacidade de degradação de hidrocarbonetos aromáticos são *Comamonas* (TERAMOTO et al., 1999), *Xhantobacter* (ZHOU et al., 1999) e *Burkholderia* (SHIELDS et al., 1995). A capacidade de degradação de BTEX dessas bactérias foi comprovada por estudos das bactérias isoladas em cultura pura.

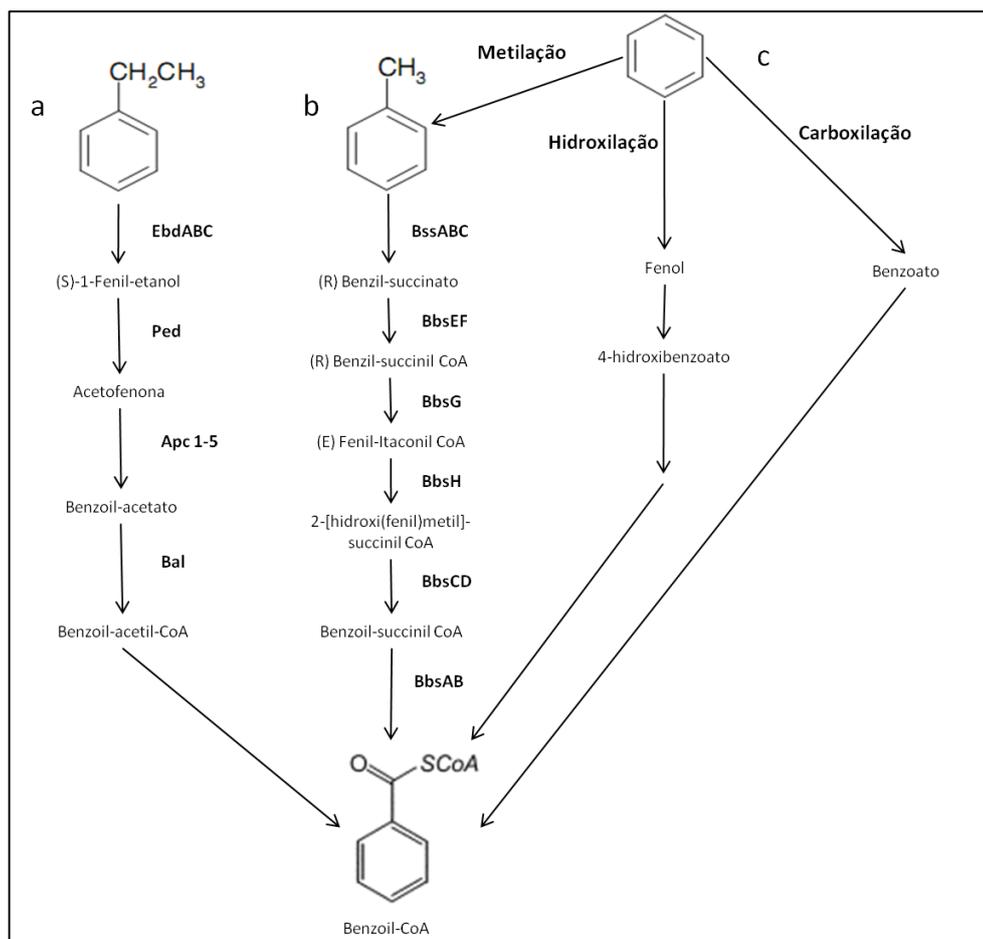
No ambiente, geralmente são encontrados vários gêneros bacterianos presentes em locais contaminados. Cavalca e colaboradores, além do gênero *Pseudomonas*, encontraram também os gêneros *Mycobacterium*, *Azoarcus* e *Bradyrhizobium*

(CAVALCA; DELL'AMICO; ANDREONI, 2004). Os gêneros *Desulfomicrobium*, *Acetobacterium*, *Desulfovibrio*, *Espirochaeta*, *Desulfocapsa* e *Acetobater* foram identificados em experimentos relacionados à contaminação com benzeno (HERRMANN et al., 2008). Amplificações genômicas utilizando iniciadores conservados para o gene SSU rRNA revelaram sequências que quando comparadas aos bancos de dados compartilham pouca identidade com sequências lá depositadas. Isso aponta uma diversidade microbiana desconhecida com potencial de degradação de compostos aromáticos.

Nesses trabalhos, o que chama a atenção é que a maioria dos gêneros identificados não são conhecidos por degradar compostos aromáticos, mas eles frequentemente aparecem em locais contaminados juntamente com gêneros bacterianos que tem capacidade de degradação. Na literatura, existem relatos de que essa diversidade de micro-organismos em locais contaminados com BTEX, e até mesmo outros hidrocarbonetos aromáticos, é reflexo de um consórcio microbiano que se estabelece e promove a degradação (ALFREIDER; VOGHT, 2007; ANDREOZZI et al., 2006; GENOVESE et al., 2008). Em um consórcio microbiano, uma dinâmica metabólica é estabelecida entre diferentes micro-organismos. A fonte de carbono de um organismo constituinte do consórcio pode ser o produto de outro, o que aumenta as chances de sucesso no caso de uma remediação e pode levar a elevados índices de degradação do contaminante (GENOVESE et al., 2008).

Para o BTEX, a biodegradação pode ocorrer aeróbia ou anaerobiamente. Quando o ambiente é anóxico, o metabolismo microbiano é anaeróbio e a degradação de BTEX ocorre pela redução de sulfatos e nitratos presentes no solo (DO et al., 2008). Os hidrocarbonetos aromáticos são modificados estruturalmente em benzoil-CoA e então degradados (WEELINK; EEKERT, 2010). Essa é uma degradação mais lenta e menos eficaz em comparação à degradação aeróbia, mas não menos importante, uma vez que no solo, o BTEX pode penetrar até grandes profundidades e alcançar a zona saturada com baixa ou nenhuma concentração de oxigênio. A Figura 3 representa exemplos de degradação anaeróbia que ocorrem principalmente em grandes profundidades dos solos e zonas saturadas.

Figura 3 – Vias de degradação anaeróbia de BTEX



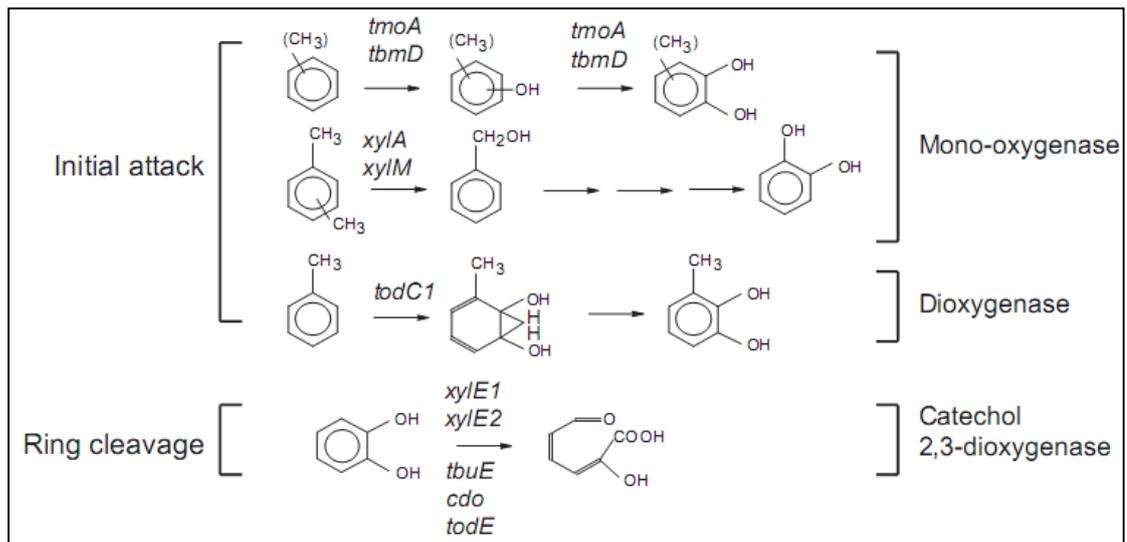
a) Degradação de etilbenzeno: *EbdABC*, etilbenzeno desidrogenase; *Ped*, (S)-1-fenil-etanol desidrogenase, *Apc1-5*, acetofenona carboxilase, *Bal*, benzoil-acetato-CoA ligase (KÜHNER et al., 2005). **b)** Degradação anaeróbia de tolueno: *BssABC*, benzilsuccinato sintase; *BbsEF*, succinil-CoA(R)-benzilsuccinato CoA transferase; *BbsG*, (R)Benzil-succinil CoA-desidrogenase; *BbsH*, Fenil-itaconil-CoA hidratase; *BbsCD*, 2[hidroxi(fenil)metil]succinil-CoA desidrogenase; *BbsAB*, benzoil-succinil-CoA tiolase (KUBE et al., 2004). **c)** Três possíveis vias de degradação de benzeno. Por metilação, o benzeno é transformado em tolueno e segue na via de degradação. A hidroxilação transforma o benzeno em fenol e segue na via de degradação de fenol. Pela carboxilação, o benzeno se transforma em benzoato e, ao final, em benzoil-CoA

Fonte: (WEELINK; EEKERT, 2010).

A degradação das três formas de xilenos não está ainda bem elucidada, mas existem evidências de que a degradação seja muito parecida com a cascata de degradação de tolueno. A ativação de xilenos é feita pela adição de fumarato formando metil-benzil-succinato antes de prosseguir por essa cascata de degradação (FOGHT, 2008; KRIEGER et al., 1999; WEELINK; EEKERT, 2010).

As vias de degradação aeróbias de BTEX são vias bastante eficazes e mais velozes, utilizando o oxigênio comoceptor final de elétrons. Essa via possui três enzimas principais que modificam a estrutura do anel aromático (Figura 4).

Figura 4 – Representação da degradação aeróbia de BTEX



A figura representa os passos iniciais de degradação de substâncias monoaromáticas. Os genes *tmoA*, *tbmD*, *XylA*, *XylM*, *todC1*, *xylE1*, *xylE2*, *tbuE*, *cdo* e *todE* codificam para monooxigenases, dioxigenases e catecol 2,3-dioxigenase. Essas três enzimas são indicadores de degradação de BTEX por metabolismo microbiano

Fonte: (HENDRICKX et al., 2006).

A biorremediação vem se tornando uma opção cada vez mais viável no tratamento de áreas contaminadas. Através do estudo da composição microbiana dessas áreas, consórcios puderam ser elucidados assim como a dinâmica de degradação desses consórcios. Evidências ainda mostram que organismos totalmente desconhecidos podem ter papel importante no processo de remediação ambiental, indicando um grande potencial a ser explorado. Atenção especial deve ser empregada principalmente em áreas com histórico de contaminação, onde, por uma pressão ambiental ocasionada pela ação antrópica, micro-organismos com metabolismos especializados podem ter sido selecionados e podem revelar novas vias de degradação ou até mesmos servir como agentes biorremediadores.

1.6 Cubatão

O município de Cubatão (latitude 23° 50 ' e 23 ° 55 ' ; longitude 46 ° 30 ' a oeste do meridiano de Greenwich) está localizado a 57 quilômetros do centro de São Paulo. A área onde se situa o município está localizada no complexo da Serra do Mar na região de transição entre o planalto paulista e a baixada santista. O clima é tropical úmido, mas com ocorrências de microclimas formados pelas variações do relevo (AUGUSTO FILHO et al., 1988; TORRES; BRAGA JR; BORGES, 2002). As escarpas

presentes são responsáveis pelos índices pluviométricos anuais que variam de 2.500 mm nas partes mais baixas chegando a 4.000 mm nas partes mais altas da encosta (IPT, 1985). Segundo a CETESB, a umidade relativa do ar varia de 70% a 90%.

Com relação à cobertura vegetal, a região de Cubatão se encontra dentro do complexo climático de Mata Atlântica, caracterizado por formações florestais definidas como ombrófila densa, ombrófila mista, estacional semidecidual, estacional decidual e ombrófila aberta, além da presença de manguezais e campos de altitude. Essa região é considerada uma das regiões mais ricas em biodiversidade do mundo. O domínio tem 1.300.000 km² de extensão, que são distribuídos quase em sua totalidade em uma faixa litorânea voltada ao Oceano Atlântico, estendendo-se do sul ao nordeste do país. Apenas 7% de sua extensão está conservada, fato que pode ser atribuído à ocupação humana. Aproximadamente 120 milhões de pessoas habitam o complexo da Mata Atlântica (MMA, 2013; WWF, 2013). A presença de pessoas não é a única causa da degradação do bioma. A região é intensamente urbanizada e inclui grandes cidades do Brasil, como São Paulo e Rio de Janeiro. Grandes polos industriais também se localizam na região, utilizando-se de recursos hídricos, pluviais e territoriais. Na região são depositados todos os rejeitos de produção, principal motivo da degradação do bioma.

O município de Cubatão é conhecido por ser o principal polo industrial do Brasil e um dos maiores da América Latina. Sua industrialização ocorreu no início do século XX com o desenvolvimento de três empresas pioneiras. Uma do ramo de couros, outra de produtos químicos e a última do ramo de papel. Foi criada também uma usina hidrelétrica que garantia o aporte de energia na região (FERREIRA; TORRES; BORGES, 2008).

Em 1930 houve incentivos para a instalação de indústrias de base no Brasil e em 1948, com o primeiro plano econômico integrado do país, que visava saúde, alimentação, transporte e energia (plano SALTE), o grande desenvolvimento industrial de Cubatão ocorreu. A primeira grande empresa atraída pelo plano econômico a se instalar na região foi a Refinaria Presidente Bernardes, em 1953, época do surgimento do monopólio do petróleo e da fundação da Petrobrás. Esta empresa foi o chamariz para que outras viessem se estabelecer em Cubatão, e entre 1953 e 1997 foram instaladas 23 grandes empresas ali. As empresas instaladas em Cubatão, em sua maioria, são indústrias químicas voltadas ao refinamento e processamento de petróleo, produção de fertilizantes, químicos industriais e siderurgia. Estas empresas

encontraram em Cubatão uma perfeita ligação entre aporte de matérias-primas e distribuição de seus produtos. Estão localizadas perto do porto de Santos, usufruem de acesso a estradas de ferro e proximidade com a capital do estado (FERREIRA; TORRES; BORGES, 2008).

A grande concentração de indústrias no local, a manipulação de compostos químicos com grande potencial poluente e a baixa ou nenhuma preocupação ambiental, levaram a cidade de Cubatão a ser chamada de Vale da Morte na década de 1980. Nesse período, a ONU chegou a classificar Cubatão como a pior cidade do mundo para se viver. Moradores intoxicados pelos rejeitos industriais que eram lançados em água, solo e ar, sem tratamento, e recém-nascidos apresentando casos de anencefalia faziam parte do cenário da cidade na época. Entre os poluentes que contaminavam a cidade, estavam metais particulados (363 mil toneladas por ano depositadas no meio ambiente), hidrocarbonetos e óxidos de enxofre e amônia, que causavam chuva ácida e degradavam toda a cobertura vegetal que se encontrava nas encostas (OLIVETTE, 2008).

Esforços realizados pelos órgãos governamentais do estado e do município, além das empresas da região, fizeram com que a poluição de Cubatão fosse diminuída por meio do controle de emissão de poluentes. Esse programa alcançou taxas de despoluição da ordem de 98,9% em redução de material particulado, 95,79% em emissão de hidrocarbonetos e redução significativa de óxidos e fluoretos (OLIVETTE, 2008).

Os esforços na redução da poluição em Cubatão trouxeram grandes melhorias à cidade, ambiente e população, o que valeu o título de cidade símbolo da recuperação ambiental, dado pela ONU durante a ECO 92 na cidade do Rio de Janeiro (CNO, 2012).

Realmente existe uma queda real da taxa de contaminação da região de Cubatão atualmente, mas o acúmulo de 30 anos de poluição continua atingindo o município. Metais pesados ainda são encontrados no ambiente fazendo parte de águas e sedimentos marinhos e pluviais. O resultado é o acúmulo destas substâncias em organismos vivos como peixes, crustáceos e moluscos consumidos pela população (CETESB, 2001). De acordo com o relatório de áreas contaminadas e reabilitadas do estado de São Paulo, muitas áreas do município de Cubatão ainda se apresentam contaminadas com hidrocarbonetos aromáticos. Entre elas incluem-se áreas consideradas residenciais além de áreas industriais (CETESB, 2011). A Tabela 4 mostra diferentes áreas contaminadas da região de Cubatão.

Tabela 4 – Relação de áreas contaminadas de Cubatão e seus respectivos contaminantes

<i>Cubatão - Local</i>	<i>Resíduo Industrial / Substância tóxica no solo</i>
<i>Centro</i>	Hidrocarbonetos aromáticos.
<i>Vila. Couto</i>	Metais pesados, resíduos organoclorados, aromáticos e halogenados.
<i>Piaçaguera</i>	Hidrocarbonetos policíclicos aromáticos, fenóis, metais pesados, compostos nitrogenados, fluoretos, fósforo, sulfatos, cloreto de vinila, 1,2 dicloroetano.
<i>Industrial</i>	Resíduos organoclorados, solventes halogenados, fluoretos, fósforo, sulfatos e mercúrio.
<i>Ponte Preta</i>	Hidrocarbonetos policíclicos aromáticos, fenóis, metais pesados, policloretos de bifenila, derivados de petróleo.

Fonte: Modificado de Guimarães (2005).

A região de Cubatão, de acordo com seu histórico, é uma região de destaque dentro do problema de contaminações ambientais. Sua história industrial e sua localização fazem com que o local seja único no mundo. A pressão ambiental exercida pelos contaminantes durante aproximadamente 30 anos pode ter resultado em modificações no seres vivos, adaptados aos contaminantes.

Os procariotos dentro deste panorama, merecem atenção especial. Variações bióticas e abióticas no ambiente podem ter feito com que os procariotos fossem selecionados de modo a utilizar estes contaminantes em seus metabolismos. Um estudo microbiano realizado por meio de técnicas de identificação e metagenômica pode trazer maior conhecimento sobre a comunidade microbiana desses locais. Este conhecimento pode gerar atualizações nos métodos de biorremediação e até mesmo o desenvolvimento de novas técnicas, possibilidades que justificam o investimento em pesquisas que podem trazer benefícios ao ambiente e às populações afetadas.

2 OBJETIVOS

Objetivo Geral:

Uso de metagenômica para avaliar a diversidade de procariotos em solos da cidade de Cubatão e o seu potencial na degradação de hidrocarbonetos aromáticos em solos contaminados.

Objetivos específicos:

1. Identificação e comparação da diversidade de procariotos em solos de três locais da cidade de Cubatão utilizando-se as técnicas de sequenciamento clássico e de alta performance.
2. Avaliação da técnica de sequenciamento de alta performance na identificação de procariotos com potencial de degradação de BTEX.
3. Identificação do potencial metabólico de degradação de substâncias aromáticas em solos do município de Cubatão.
4. Elucidação de um possível consórcio microbiano envolvido na degradação de anéis aromáticos.

3 MATERIAL E MÉTODOS

3.1 Locais de coleta

Parque Estadual Cotia-Pará

O parque Estadual Cotia-Pará se encontra às margens da rodovia Anchieta entre os quilômetros 55 e 56 dentro do município de Cubatão, no Estado de São Paulo. Nele está localizada a Secretaria do Meio Ambiente e o núcleo de educação ambiental da cidade. Conta com viveiro de mudas, zoológico e reabilitação de animais. A vegetação é de remanescentes de Mata Atlântica e possui clima quente e úmido. Posição da coleta: S 23 ° 54 ' 14,20 "; W 46 ° 25 ' 47,10 "

Parque estadual do Perequê

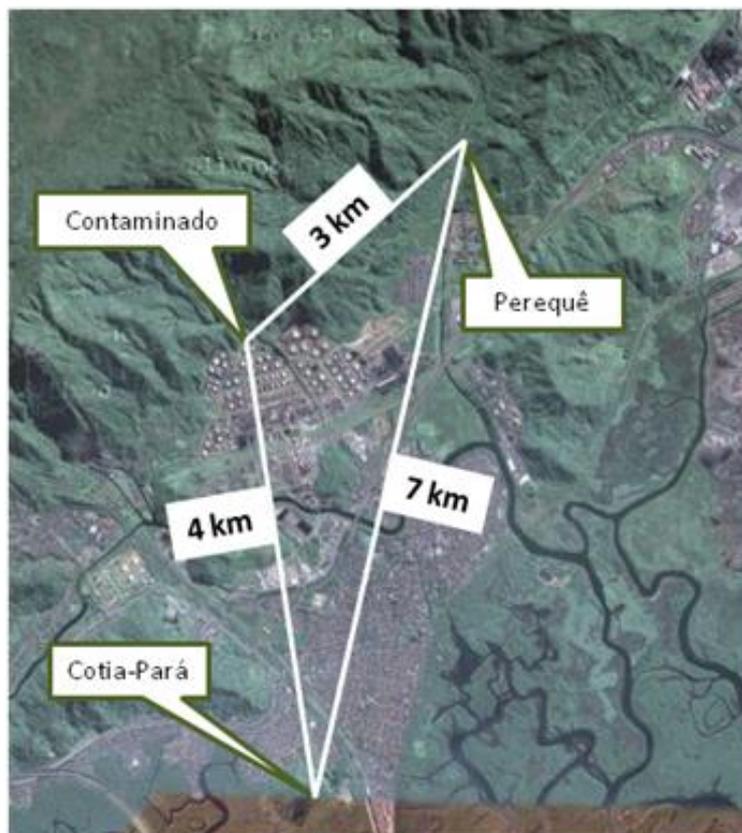
O parque se encontra na estrada de Piaçaguera, uma estrada secundária localizada no quilômetro quatro da rodovia Cônego Domênico Rangoni. Sua fauna e flora são, em toda a sua extensão, provenientes de Mata Atlântica. Possui clima quente e úmido e está localizado a aproximadamente cinco quilômetros do centro de Cubatão. Posição do local de coleta: S 23 ° 50 ' 48,3 "; W 46 ° 24 ' 57,1 "

Local contaminado

A região contaminada está localizada às margens da rodovia Caminho do Mar, rodovia secundária à rodovia Cônego Domênico Rangoni. A área sofreu um derramamento de um tanque de combustível contendo gasolina em 2008. O combustível foi drenado da área do dique do tanque e o solo ainda está sob monitoramento. Posição do local de coleta: S 23 ° 51 ' 57,0 "; W 46 ° 26 ' 15,1 "

A Figura 5 mostra a exata localização das coletas de solo assim como as distâncias entre os pontos.

Figura 5 - Imagem aérea dos pontos de coleta



Imagery ©2013 TerraMetrics, Map data 2013 Google, Inav/Geosistemas SRL.

Os pontos de coletas estão localizados nos vértices do triângulo formado entre eles. As distâncias aproximadas foram calculadas com base na localização por GPS de cada uma das amostras.

3.2 Procedimento de coleta e processamento das amostras

Aproximadamente 100 gramas de solo foram coletados de todas as áreas escolhidas para o projeto. O solo foi perfurado com o auxílio de uma pá de jardim até uma profundidade aproximada de 20 centímetros. Logo após a retirada desta camada superficial, amostras de solo foram coletadas com o auxílio de uma espátula de alumínio previamente esterilizada. As amostras de solo foram alocadas em tubos de polipropileno estéreis e imediatamente congeladas em nitrogênio líquido para manutenção de características físico-químicas e microbiológicas. Estas amostras foram estocadas a -80 °C até seu processamento.

Alíquotas do solo coletado (0,5 g a 2 g) foram utilizadas para extração de ácidos nucleicos. Para a extração de DNA foi utilizado o kit E.Z.N.A Soil DNA (Omega Bio-Tek, Norcross, GA, Estados Unidos) de acordo com as indicações do fabricante. Após a extração, a concentração e qualidade do material extraído foi medida em

espectrofotômetro Nanodrop ND-1000 (Thermo-Scientific, Wilmington, GA, Estados Unidos) e visualizado em gel de agarose 0,8%. Os ácidos nucleicos foram estocados a -80 °C para uso posterior.

Para análise dos contaminantes presentes no solo coletado, uma alíquota de 2 g de solo foi acondicionada em um vial de 20 ml. A análise foi iniciada com uma pré-extração, na qual o vial permaneceu em agitação a 300 rpm, 60 °C por 2 minutos alternando-se 10 segundos de agitação e 4 segundos em descanso. A extração foi realizada colocando-se a fase gasosa do vial em contato com uma fibra que possui afinidade a hidrocarbonetos (Black 75 µm Carboxen/PMDS) em uma agitação de 250 rpm por 10 minutos a 60 °C. Após a extração, a amostra foi dessorvida, ficando no injetor do equipamento por 90 segundos. O processamento de análise se iniciou a 60°C por 1 minuto. A variação da temperatura se deu em intervalos de 5 °C por minuto até que fosse atingida a temperatura de 150 °C. Nessa temperatura a amostra ficou por 21 minutos. Decorrido este tempo, uma nova variação de temperatura de intervalos de 10°C por minuto até 250 °C foi realizada. Nesta temperatura, a amostra ficou por 2 minutos totalizando 33 minutos o tempo da análise.

Os materiais utilizados foram: cromatógrafo a gás 450 GC (Varian, Walnut Creek, CA, Estados Unidos) acoplado a um espectrofotômetro de massa 240 MS (Varian, Walnut Creek, CA, Estados Unidos), detector Ion trap. Injetor split 10, coluna VF 5MS. O gás de arraste utilizado foi o hélio com fluxo de 1 mL por minuto e a temperatura do injetor foi ajustada para 250 °C.

Além da análise dos contaminantes, foi realizado também um estudo analítico do solo para levantamento de íons presentes, capacidade de troca iônica, micronutrientes, macronutrientes e granulometria. Todos os parâmetros analisados foram executados pela empresa Bioagri Ambiental, a qual segue padrões internacionais de controle de qualidade. Os resultados estão mostrados no ANEXO A.

3.3 Amplificações genômicas

3.3.1 Iniciadores

Os iniciadores utilizados foram:

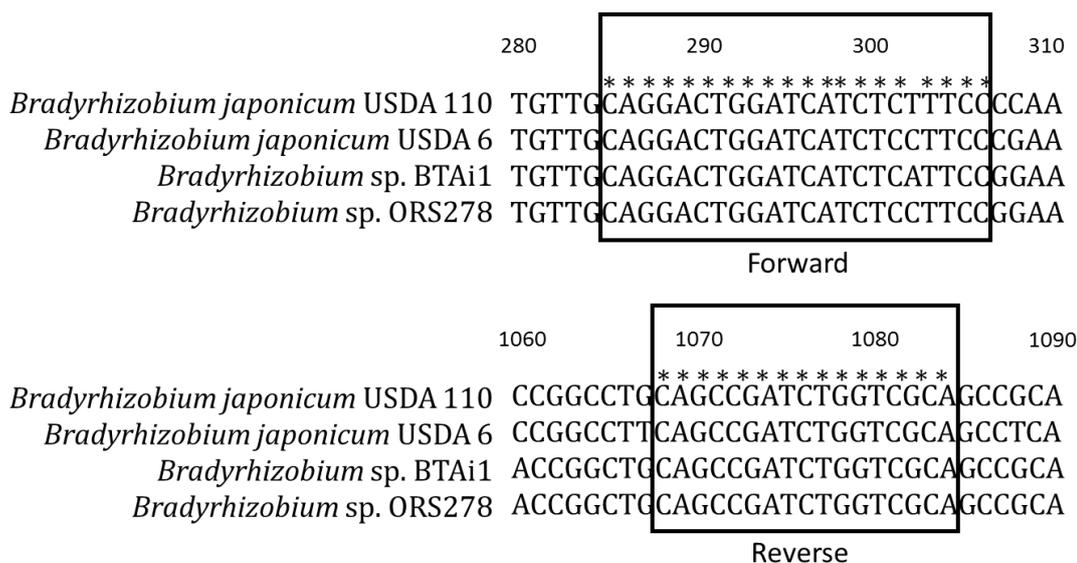
27-F	5' AGAGTTTGATCCTGGCTCAG 3'
1492-R	5' ACGGYTACCTTGTTACGACTT 3'

Arch21-F 5' TTCCGGTTGATCCYGCCGGA 3'
 Mox - F 5' CAGGACTGGATCATCTCCTTCC 3'
 Mox - R 5' TGCGACCAGATCGGCTG 3'

Para as amplificações das SSU rRNA de bactérias foram utilizados o par de iniciadores 27-F e 1492-R, para arqueias, o par Arch21-F e 1492-R. O iniciador 1492-R é um iniciador universal utilizado para amplificação da subunidade ribossomal menor tanto em bactérias como em arqueias.

Os iniciadores Mox foram desenvolvidos para identificar a capacidade de degradação de substâncias aromáticas pela cepa *Bradyrhizobium japonicum* SEMIA 566. Os iniciadores foram desenhados sobre regiões genômicas que correspondiam ao domínio conservado tolueno-4-monoxigenase contendo aproximadamente 1179 pares de base entre as bases de número 4.068.831 e 4.069.949 para o genoma completo de *Bradyrhizobium japonicum* USDA 110. As regiões foram alinhadas e duas sub-regiões com alta conservação foram escolhidas para desenho dos iniciadores (Figura 6). Os genomas utilizados para desenho dos iniciadores foram: *Bradyrhizobium japonicum* USDA 110 DNA (BA000040.2); *Bradyrhizobium japonicum* USDA 6 (AP012206.1); *Bradyrhizobium* sp. BTAi1 (CP000494.1) e *Bradyrhizobium* sp. ORS278 (CU234118.1). O produto de amplificação previsto foi de 800 bases.

Figura 6 – Região de desenho dos iniciadores Mox



Regiões escolhidas para desenvolvimento dos iniciadores Mox para o gênero *Bradyrhizobium*. As caixas indicam as regiões alvo dos iniciadores. As espécies utilizadas estão ao lado de cada uma das seqüências. Os asteriscos indicam que a base abaixo é conservada nos quatro genomas.

3.3.2 Reação de polimerização em cadeia

A reação de PCR foi realizada para amplificação das regiões conservadas da subunidade menor do gene que codifica o ribossomo. Para esse fim, foi utilizada a DNA polimerase GoTaq (Promega Corporation, Madison, WI, Estados Unidos) de acordo com as especificações do fabricante. As reações foram feitas em termociclador PTC-100 (MJ Research Inc., Waltham, MA, Estados Unidos) Os ciclos e concentrações estão descritos abaixo.

Reação de amplificação (concentrações finais)

Tampão	1 X
MgCl ₂	2,5 mM
dNTPs	0,2 mM
Iniciadores	0,15 µM
DNA Polimerase	0,04 U/µL

Ciclos da reação de amplificação:

Desnaturação		94 °C	3 minutos
	Desnaturação	94 °C	30 segundos
Ciclo (35 vezes)	Anelamento	53 °C	30 segundos
	Polimerização	72 °C	2 minutos
Extensão final		72 °C	5 minutos
Manutenção		4 °C	Infinito

3.4 Eletroforese

Os resultados de amplificação foram submetidos à eletroforese em gel de agarose para verificação de eficiência de amplificação e tamanho do produto amplificado. Os experimentos foram realizados em gel de agarose 0,8% confeccionado com TAE diluído para 0,5 X. A tensão utilizada foi de 90 V por 30 minutos para géis de 25 mL e 110 V por uma hora para géis de 100 mL em cuba de eletroforese Horizon 11.14 (Biometra, Goettingen, Alemanha).

Após a migração dos fragmentos, o gel foi corado em solução de 0,5 µg/L de brometo de etídeo por 1 minuto e deixado em repouso em água por 10 minutos para retirar excessos de brometo de etídeo. Para visualização da migração, o gel foi fotografado por um fotodocumentador L-Pix (Loccus biotecnologia, Cotia, SP, Brasil) e os fragmentos foram analisados.

TAE 50X	Tris	2 M
	Ácido acético glacial	2 M
	EDTA	50 mM

3.5 Clonagem de fragmentos

3.5.1 Ligaçã

Os fragmentos amplificados foram ligados ao vetor pGEM-T Easy (Promega Corporation, Madison, WI, Estados Unidos) na proporçã molar inserto:vetor de 3:1. As condições e o protocolo utilizados foram definidos pelo fabricante. Apõs a ligaçã dos fragmentos, parte da reaçã (2 µL) foi transfectada em bactõrias competentes.

3.5.2 Bactõrias competentes

Uma colõnia de bactõria *E. coli* DH10b foi inoculada em 2 mL de meio lquido LB e incubada por 2 horas a 37 °C sob agitaçã. Apõs este perõdo, uma alõquota de 1 mL foi adicionada à 50 mL de LB em um erlenmeyer. A cultura foi incubada por um perõdo de duas a trõs horas atõ que uma densidade õptica de 0,5 a 0,6 fosse atingida. Em seguida, foram adicionados 500 µL de uma soluçã estõril de MgCl₂ 1 M mantendo o erlenmeyer em gelo.

A cultura foi transferida para um tubo cõnico de 50 mL previamente gelado e mantida no gelo por 15 minutos. Foi realizada uma centrifugaçã a 3000 rpm por 20 minutos a 4 °C em centrõfuga refrigerada Hettich Universal 30RF (Hettich, Tuttlingen Alemanha) e rotor angular 1414 sendo o sobrenadante descartado. O sedimento foi ressuspendido em 15 mL de tampã RFI gelado e mantido em gelo por 10 minutos.

A suspensã bacteriana foi centrifugada a 3000 rpm por 20 minutos a 4 °C e ressuspendida em 2 ml de tampã RFII em gelo. As bactõrias ressuspendidas foram alõquotadas (50 µL) em tubos de centrõfuga previamente resfriados.

Tampã RFI	CH ₃ COOK (pH=6,9)	30 mM
	MnCl ₂	50 mM
	KCl	100 mM
	CaCl ₂	10 mM
	Glicerol	15%

Ajustar pH em 5,8 com ácidu acõtico e passar por filtro Millipore de 22 µm.

Tampã RFII	Na-MOPS (pH=7)	10 mM
-------------------	----------------	-------

CaCl ₂	75 mM
KCl	10 mM
Glicerol	15%

Ajustar o pH em 6,6 com NaOH e passar por filtro Millipore de 22 µm

Meio LB (Luria-Bertani)

	Gramas por litro
Triptona	10 g
Extrato de levedura	5 g
NaCl	10 g

O meio foi autoclavado por 15 minutos a 120 °C. Os antibióticos foram adicionados com o meio de cultura ainda morno

3.5.3 Transformação por choque térmico

Alíquotas de 50 µL de bactérias competentes foram descongeladas em gelo por 15 minutos. Após estarem completamente descongeladas, foram adicionados 2 µL de ligação e a mistura foi deixada em gelo por 30 minutos. Depois de decorrido este tempo, o tubo foi submetido a banho úmido a 42 °C por 90 segundos e imediatamente colocado em gelo por 120 segundos. Foram adicionados 1 mL de meio SOC e a cultura foi incubada por 30 minutos a 37 °C. A cultura foi então centrifugada a 4400 rpm em temperatura ambiente. Com as células já centrifugadas, foi retirado do tubo 800 µL do sobrenadante sendo as células ressuspensas em volume final de 200 µL. Para a seleção, 100 µL de células ressuspensas foram plaqueadas em placas de Petri de 140 x 15 mm contendo 25 mL de meio LB acrescidos de 100 µg/mL de ampicilina e 40 µg/mL de X-Gal.

As placas inoculadas foram colocadas em estufa a 37 °C por aproximadamente 16 horas até que as colônias das células transformadas pudessem se desenvolver. Após este período, as colônias foram retiradas da placa de seleção com o auxílio de palitos de madeira autoclavados e colocadas para crescimento em placas de cultura de 96 poços contendo 1 mL de meio LB acrescido de 100 µg/mL de ampicilina. As placas foram incubadas por aproximadamente 16 horas em plataforma de agitação orbital a 300 rpm a 37 °C. Após este período, a cultura foi submetida à extração de DNA plasmidial.

Meio SOC		
	Triptona	20 g/L
	Extrato de levedura	5 g/L
	NaCl	8,5 mM

KCl	2,5 mM
*MgCl ₂	10 mM
*MgSO ₄	10 mM
*Glicose	20 mM

O pH foi ajustado para 7 com NaOH 1N e a solução foi esterilizada em autoclave por 15 minutos a 120 °C e estocado a 4 °C.

* Soluções estéreis. Adicionados após a esterilização em ambiente estéril.

3.5.4 Extração de DNA plasmidial em placa de 96 poços

Após o crescimento das colônias selecionadas, 40 µL da cultura foi utilizada para estoque sendo que, para esse fim, foram adicionados à cultura 40 µL de glicerol 50% autoclavado. O estoque foi preparado em placas de 96 poços para criogenia e guardadas a -80 °C

O restante da cultura foi centrifugado a 4000 rpm por 10 minutos para precipitação de células. O sobrenadante foi descartado e a placa invertida por 1 minuto em papel toalha para retirar excessos de sobrenadante. Foram adicionados 240 µL de Solução I, a placa selada e as bactérias ressuspendidas em agitador elétrico com adaptador para placas de 96 poços até que não fosse visto nenhum precipitado nos poços. A placa foi centrifugada a 4000 rpm por 10 minutos. Novamente a placa foi colocada invertida sobre papel toalha para retirada de excessos. Ao precipitado, foram adicionados 80 µL de Solução I, a placa foi selada com adesivo e as células foram ressuspendidas por meio de agitação até que nenhum precipitado fosse visualizado.

A cultura foi transferida para uma nova placa de 96 poços de fundo “U” já contendo 5 µl de RNase H (10 mg/mL). Na sequência, 60 µL de Solução II foram adicionados a cada poço, sendo a placa novamente selada e invertida 10 vezes para homogeneização. A placa foi incubada em temperatura ambiente por 10 minutos e o conteúdo levado ao fundo através de uma rápida centrifugação. Em seguida foram adicionados 60 µL de Solução III gelada e a placa foi invertida 10 vezes. Novamente a placa foi incubada a temperatura ambiente e sofreu rápida centrifugação, sendo colocada em estufa a 90 °C sem adesivo por 30 minutos. Decorrido este passo, a placa foi novamente selada e imediatamente colocada em gelo por 10 minutos sendo em seguida centrifugada por 8 minutos a 4000 rpm.

Uma placa de filtro foi presa a uma placa de fundo “V” por fita adesiva e todo o conteúdo da placa centrifugada foi coletado com auxílio de micropipeta multicanal e repassado para os filtros. A placa então foi centrifugada por 6 minutos a 4000 rpm e a

placa de filtro foi descartada. Logo em seguida, foram adicionados 110 μL de isopropanol 100% e a placa foi selada e invertida 20 vezes para homogeneização, sendo então centrifugada por 45 minutos a 4000 rpm. O sobrenadante foi descartado e uma lavagem foi feita com adição de 200 μL de etanol 70% gelado sendo a placa centrifugada por mais 5 minutos. O sobrenadante foi descartado e a placa incubada a 37 °C por 15 minutos para secagem. Decorrido o tempo de incubação, as amostras foram ressuspensas em 40 μL de água ultrafiltrada autoclavada e a placa foi armazenada a -20 °C.

Solução I

Glicose	50 mM
Tris	25 mM
EDTA	10 mM

A solução foi autoclavada durante 15 minutos a 120°C estocada a 4°C.

Solução II

NaOH	0,2 N
SDS	34 mM

A solução foi preparada na hora e utilizada à temperatura ambiente.

Solução III

CH ₃ COOK	300 mM
C ₂ H ₄ O ₂	11,5% (volume/volume)

A solução foi autoclavada por 15 minutos a 120 °C e estocada a 4 °C

3.6 Sequenciamento automático de nucleotídeos por método Sanger

3.6.1 Reação de sequenciamento para método Sanger

A reação de sequenciamento foi realizada utilizando-se o Big Dye Terminator Cycle Sequencing Kit (Life Technologies, Grand Island, NY, Estados Unidos) de acordo com as especificações do fabricante. Para cada reação foram utilizados 100 ng de DNA e 5 pmol de iniciadores. O iniciador utilizado para essa reação foi o T7 (5' TAATACGACTCACTATAGGG 3'). A reação foi feita em termociclador PTC-100.

3.6.2 Reação de precipitação para sequenciamento por método Sanger

Para este procedimento, foram adicionados à reação de sequenciamento 80 μL de isopropanol 65%. A placa foi agitada durante 10 segundos e incubada à temperatura ambiente no escuro por 20 minutos. Após este período, a placa foi centrifugada a 4000 rpm durante 45 minutos, sendo o sobrenadante descartado e a placa deixada invertida sobre papel toalha para retirada de excessos. Foram adicionados 150 μL de etanol 60% gelado e novamente a placa foi centrifugada a 4000 rpm por 10 minutos. O

sobrenadante foi descartado e a placa foi submetida a uma rápida centrifugação posicionada de modo invertido sobre papel toalha até que as rotações atingissem o máximo de 300 rpm. A placa foi secada no escuro durante 1 hora e guardada em papel alumínio a -20 °C.

Para a entrada em sequenciador, as amostras foram ressuspensas em 10 µL de formamida e desnaturadas a 95 °C por 5 minutos. O sequenciamento foi realizado no equipamento ABI 3730 (Life Technologies, Grand Island, NY, Estados Unidos).

3.7 Sequenciamento GS FLX Titanium 454

A plataforma de pirosequenciamento utilizada foi a plataforma GS FLX 454 (Roche, Mannheim, Alemanha). A preparação do material foi realizada segundo todas as indicações e materiais mencionados pelo fabricante.

Para a preparação do material genômico de solo, o protocolo seguido foi o mesmo descrito no item 3.2 sendo submetido ao protocolo de sequenciamento GS FLX.

3.8 Bioinformática

3.8.1 Análise das sequências obtidas por sequenciamento pelo método Sanger

As sequências provindas das amplificações das SSU rRNAs foram utilizadas para comparação no banco de dados do RDP (COLE et al., 2009). Primeiramente, as sequências foram tratadas para retirada do vetor nas extremidades 3' e 5' para que somente a parte amplificada fosse classificada. Com o vetor retirado, as sequências foram analisadas para a qualidade de sequenciamento. Somente as bases sequenciadas que obtiveram qualidade *phred* 20 ou maior foram selecionadas, sendo cortadas as extremidades que não possuíam essa qualidade.

Após o tratamento, as sequências foram classificadas de acordo com o programa de classificação do RDP utilizando-se os parâmetros definidos pelo programa, assim como o seu banco interno de sequências ribossomais.

3.8.2 Análise das sequências obtidas por sequenciamento de alta performance

O arquivo de saída do equipamento da Roche foi utilizado para submissão ao programa MG-RAST (MEYER et al., 2008). Ao mesmo tempo, arquivos multifasta foram criados utilizando-se a ferramenta SSFtools do conjunto de ferramentas para GS FLX.

3.8.2.1 Classificação taxonômica

A classificação das sequências produzidas pelo método de sequenciamento GS FLX foram submetidas primeiramente ao programa MG-RAST para análise taxonômica geral das sequências geradas.

As sequências, processadas pelo “*pipeline*” do programa, foram primeiramente filtradas pelo tamanho. O programa somente seleciona sequências para passos seguintes se estas possuírem no mínimo 75 bases de boa qualidade, ou seja, qualidade phred mínima de 15 e corte de parte da sequência se esta apresentar pelo menos 5 bases com qualidade menor do que o limite. Após esta seleção ocorre a retirada de sequências repetidas. Sequências que apresentam as primeiras 50 bases idênticas são agrupadas e a maior sequência é selecionada para prosseguir na análise.

Após a seleção por qualidade, as sequências foram submetidas a uma predição de código e leitura aberta pelo programa FragGeneScan (RHO; TANG; YE, 2010). Após esta predição, as sequências foram anotadas utilizando o programa BLAT (KENT, 2002). Diferentes bancos de dados inclusive os utilizados neste trabalho como o M5NR e o de subunidades ribossomais menores do RDP (COLE et al., 2007) estão disponíveis no programa para comparação. Os resultados foram visualizados e posteriormente exportados para organização dos dados.

Para a identificação taxonômica geral das sequências foram utilizados os limiares de *e-value* de $1e-10$ ou menores e pelo menos 60% de similaridade. As sequências que cumpriram estes limiares foram utilizadas para comparação.

Pelo programa MG-RAST também foi realizada uma primeira análise de identificação por SSU rRNA para refinamento das identificações. Para identificação no banco de dados do RDP, as sequências que possuíram no mínimo 70% de similaridade com sequências do banco de dados e com um valor de *e-value* de $1e-10$ ou menor foram contabilizadas.

Outro método utilizado para refinamento de identificação dos procariotos foi realizado através da comparação de genes conservados com base no projeto AMPHORA (WU; EISEN, 2008) e pelo programa MetaPhyler (LIU et al., 2011). O projeto AMPHORA foi um esforço feito por Wu e Eisen (2008) na identificação de genes conservados tendo como base genomas fechados disponibilizados nos bancos de dados genômicos. Através de comparações feitas, os autores chegaram a um grupo de 31 genes conservados e que possuíam valor taxonômico para procariotos. O

programa MetaPhyler utiliza os genes conservados em sua base de dados para identificação de sequências.

Para essa classificação, as sequências em multifasta foram verificadas quanto a duplicidade de sequências. As sequências duplicadas foram retiradas através do programa replicates (GOMEZ-ALVAREZ; TEAL; SCHMIDT, 2009). Para que uma sequência fosse considerada uma réplica, ela deveria começar com pelo menos 5 bases idênticas a uma ou mais sequências do mesmo metagenoma e possuir pelo menos 95% de identidade, podendo variar de tamanho. As sequências que apresentavam esses limites foram agrupadas e a maior sequência do grupo foi utilizada para a classificação.

Após a limpeza das sequências duplicadas, as sequências foram traduzidas nos 6 diferentes códigos abertos de leitura pelo programa Fastatranslate versão 2.20.5 que faz parte de um conjunto de ferramentas chamado fastutils desenvolvida pelo Ensembl Project (FLICEK et al., 2012). O metagenoma traduzido foi então processado pelo programa MetaPhyler versão 1.25 e a saída do programa foi utilizada para composição dos resultados.

O terceiro e último método de classificação taxonômica para refinamento das identificações foi realizado através de comparação do programa BLAST versão 2.2.27 (ALTSCHUL et al., 1997). As sequências previamente selecionadas para qualidade e livre de réplicas foram comparadas ao banco de dados de genomas procarióticos fechados do NCBI versão 194 (GEER et al., 2010) através do programa BLASTN utilizando-se um *e-value* mínimo de $1e-10$, saída tabelar, e anotando-se somente os 10 primeiros resultados para cada uma das sequências. Somente as sequências metagenômicas que apresentaram cobertura maior ou igual a 300 bases e identidade maior ou igual a 80% às sequências do banco de dados foram utilizadas para a classificação taxonômica.

A classificação taxonômica foi feita utilizando-se o programa MEGAN versão 4.61.5 (HUSON et al., 2007). As sequências que cumpriram os limites de cobertura e identidade tiveram seus arquivos fastas recuperados e foram novamente comparadas ao banco de dados de genomas procarióticos fechados do NCBI. Para isso, foi utilizado o programa BLASTN com os parâmetros já citados acima, porém, utilizando-se a saída padrão do programa e não a saída tabelar. A saída do programa BLASTN foi carregada no programa MEGAN sendo processada utilizando-se os parâmetros padrão e valores normalizados para 100.000 sequências. Através do MEGAN, vários níveis taxonômicos

atribuídos às sequências puderam ser visualizados e utilizados para comparação entre os metagenomas.

3.8.2.2 *Classificação funcional do metabolismo de BTEX*

A classificação funcional das sequências metagenômicas foi baseada nos mapas metabólicos do KEGG versão 65 (KANEHISA; GOTO, 2000). Todos os passos metabólicos dos mapas de degradação de xilenos e tolueno tiveram seu gene identificado no banco de dados UniprotKB/Swissprot versão 2012_1 (APWEILER et al., 2004). As sequências foram selecionadas através do número EC e priorizadas de acordo com o fato de terem ou não sido revisadas no banco de dados. A primeira opção de sequências incluía aquelas que se apresentavam revisadas. Caso o gene a ser estudado não apresentasse sequências revisadas, as sequências não revisadas foram escolhidas. As sequências recuperadas foram comparadas aos metagenomas através de tBLASTN com os parâmetros de *e-value* 1e-10 e com anotação dos 10 melhores hits.

Após a comparação entre as sequências recuperadas do UniprotKB/Swissprot e os metagenomas, o resultado foi analisado de modo qualitativo e anotado dentro dos mapas metabólicos. Foram anotadas a presença ou ausência dos diversos passos metabólicos para as vias de degradação de xilenos e tolueno. Para que um passo metabólico fosse anotado como presente, a sequência recuperada do banco de dados deveria apresentar pelo menos 60% de similaridade com alguma sequência metagenômica. Quando as sequências encontradas cumpriam os limites determinados, caixas cheias eram marcadas no passo metabólico do mapa do KEGG. Passos metabólicos que não apresentaram genes anotados no banco de dados foram marcados com caixas na cor cinza. Genes que quando comparados aos metagenomas não apresentaram similaridade suficiente foram deixados em branco.

Para identificação da presença ou ausência dos passos metabólicos iniciais das vias de degradação, operons responsáveis pelos passos iniciais de transformação dos contaminantes foram identificados no banco de dados Bionemo versão 6 (CARBAJOSA et al., 2009). Suas sequências foram recuperadas do NCBI e um banco de dados de operons de degradação de tolueno e xilenos foi montado com sequências dos seguintes números de acesso: AB024741.1; AJ344068.1; L40033.1; U04052.1; U15298.1; M65106.1; J04996.1; GQ884177.1; AF319657.1; AJ005663.3 e

AJ012090.1. Esses operons foram comparados aos metagenomas por BLASTN e sequências dos metagenomas que apresentaram identidade de no mínimo 80%, cobertura de 100 aminoácidos e *e-value* de 1e-10 ou inferior foram contabilizadas. Desse modo, sequências com identidade aos operons comparados foram identificadas qualitativamente no início dos mapas metabólicos de degradação de xilenos e tolueno e os metagenomas foram comparados para presença ou ausência dos operons de degradação.

3.8.2.3 Operons e domínios PFAM

Foram explorados também operons de degradação de BTEX e foi realizada uma comparação entre operons de degradação conhecidos e os metagenomas estudados. Para tal finalidade, catorze operons já conhecidos por sua atividade de degradação de BTEX foram identificados no banco de dados Bionemo com as seguintes identificações no banco de dados do NCBI: AB024741.1; AJ344068.1; L40033.1; U04052.1; AF019635; U15298.1; M65106.1; J04996.1; GQ884177.1; AF319657.1; AJ005663.3 e AJ012090.1; AB120956.1 e AB120955.1. A estrutura dos genes dentro dos operons foi representada e para cada gene foram identificados domínios conservados PFAM versão 26 (PUNTA et al., 2012), que podem ser definidos como as regiões funcionais que compõem as enzimas. Ao todo foram identificados dezoito domínios PFAM pertencentes aos genes que constituem os operons de interesse. Destes dezoito, seis domínios foram escolhidos por serem domínios de modificação direta de estruturas de anéis aromáticos. A identificação dos domínios conservados utilizados no banco de dados do PFAM é: PF06099; PF02332; PF04663; PF06234; PF00848; PF00866.

Após a seleção dos domínios conservados de degradação, as sequências alinhadas dos domínios foram recuperadas do banco de dados do PFAM e um estudo de HMMs (Hidden Markov Model) foi realizado através do programa HMMER versão 3.0 (FINN; CLEMENTS; EDDY, 2011). Este estudo consistiu na construção dos HMMs dos domínios escolhidos e comparação destes HMMs com as sequências traduzidas dos três metagenomas gerados neste estudo e mais quatro identificados no banco de dados do NCBI com as seguintes identificações: SRX038779; SRX000607; SRR49666 e AAFX00000000. Apenas as sequências que apresentaram *e-value* menor ou igual a 1e-10 foram contabilizadas. Os resultados da contabilização de sequências foram normalizados para um total de 500.000 sequências.

Como complementação dos estudos de domínios conservados, as sequências metagenômicas que apresentaram similaridade com os domínios PFAM tiveram sua sequência de nucleotídeos recuperada e identificadas taxonomicamente.

A classificação seguiu a mesma maneira do item 3.8.2.1. As sequências foram comparadas por BLASTN às sequências do banco de dados do NCBI utilizando-se um *e-value* de $1e-10$ e apenas sequências que tiveram cobertura mínima de 50 aminoácidos foram selecionadas. A comparação entre as sequências selecionadas e o banco de dados foi processada pelo MEGAN utilizando-se os parâmetros padrão. A visualização das comparações entre os diferentes metagenomas foi realizada através da interface do programa.

Foi realizado um estudo de verificação de domínios conservados PFAM de degradação de BTEX nos genomas dos organismos mais representados. Para isso, os HMMs construídos para os domínios PFAM selecionados foram utilizados para comparação nos genomas traduzidos. Os genomas utilizados foram os genomas fechados depositados no banco de dados do NCBI. Todos os genomas fechados pertencentes ao gênero identificado nos metagenomas foram utilizados para a comparação. Essa comparação se deu de forma qualitativa e um domínio foi considerado como presente se encontrado no genoma com um *e-value* de $1e-10$ ou menor.

3.9 Degradação de BTEX por *Bradyrhizobium japonicum*

Para o experimento de degradação de xilenos e tolueno por *Bradyrhizobium japonicum* SEMIA 566 uma colônia crescida em placa com meio YEM foi inoculada em 2 ml de meio YEM em tubo de ensaio e incubado a 28 °C a 200 rpm por 72 horas em câmara de agitação orbital MA830 (Marconi, Piracicaba, SP, Brasil). Após o período mencionado, a cultura foi transferida para erlenmeyers de 250 mL contendo 50 ml de meio YEM que foram novamente incubados à temperatura de 28 °C a 200 rpm por mais 72 horas. Após o período de crescimento, as culturas foram transferidas para tubos cônicos de 50 mL e as células foram precipitadas a 5000 rpm durante 15 minutos. Com as células precipitadas, o sobrenadante foi descartado e as células foram ressuspensas em 3 mL de meio M2, e 100 µL foram colocados em cada um dos frascos *headspace* contendo 5 mL de meio M2. Para o experimento de degradação foi utilizada uma concentração de 1000 ppm para cada um dos contaminantes em frascos

separados. Para controle de crescimento foram utilizados o meio M2 complementado com manitol a uma concentração final de 0,4%. A Tabela 5 demonstra o experimento realizado.

Tabela 5 – Experimento de sobrevivência na presença de xilenos e tolueno por *Bradyrhizobium japonicum* SEMIA 566

<i>Desenho experimental</i>	
<i>Controle positivo de crescimento</i>	Meio M2 + manitol + inóculo
<i>Controle negativo de crescimento</i>	Meio M2 + inóculo
<i>Controle de degradação de tolueno</i>	Meio M2 + tolueno
<i>Controle de degradação de xilenos</i>	Meio M2 + xilenos
<i>Ensaio tolueno (tolerância)</i>	Meio M2 + manitol + tolueno + inóculo
<i>Ensaio xilenos (tolerância)</i>	Meio M2 + manitol + xilenos + inóculo
<i>Ensaio tolueno (degradação)</i>	Meio M2 + tolueno + inóculo
<i>Ensaio xilenos (degradação)</i>	Meio M2 + xilenos + inóculo

Os ensaios foram realizados em quintuplicata em frascos de *headspace*, sendo sacrificados logo após as leituras. Cada leitura foi realizada em um espaçamento de 24 horas. A leitura T0 logo após a preparação do experimento e a leitura T5 com um período de 48 horas em relação à leitura T4.

A verificação de degradação de tolueno e xilenos foi realizada em cromatógrafo a gás 450 GC (Varian, Middelburg, EA, Holanda) acoplado a um espectrofotômetro de massa 240 MS (Varian) seguindo-se o mesmo protocolo do item 3.2.

Para verificação de crescimento das células, 1 mL da cultura de cada uma das preparações foi submetida à leitura de densidade óptica (DO) a um comprimento de onda de 600 nm em espectrofotômetro UV-2450 (Shimadzu, Kyoto, Japão), sendo o aparelho calibrado com meio M2 apenas.

Meio YEM	Gramas por litro
Extrato de levedura	1,0 g
Manitol	10,0 g
KH ₂ PO ₄	0,5 g
MgSO ₄	0,2 g
NaCl	0,1 g

O meio foi autoclavado a 121 °C por 15 minutos e estocado a 4 °C.

M2 Sais 10X	Gramas por litro
NH ₄ Cl	5,0 g
Na ₂ HPO ₄	20,9 g
KH ₂ PO ₄	10,6 g

A solução M2 sais 10X foi autoclavada a 121 °C durante 15 minutos e estocada a 4 °C.

Soluções

	Concentração
FeSO ₄	1 mM
MgSO ₄ .7H ₂ O	1 M
CaCl ₂ .2H ₂ O	1 M
*Manitol	10 g/L

As soluções foram autoclavadas a 121 °C durante 15 minutos e estocada a 4 °C.

*Solução esterilizada por filtração em filtro 24 µm.

Meio M2	Quantidade para 1 litro
M2 sais 10X	10,0 mL
FeSO ₄	10,0 mL
MgSO ₄ .7H ₂ O	1,0 mL
CaCl ₂ .2H ₂ O	0,5 mL
Manitol	20 mL

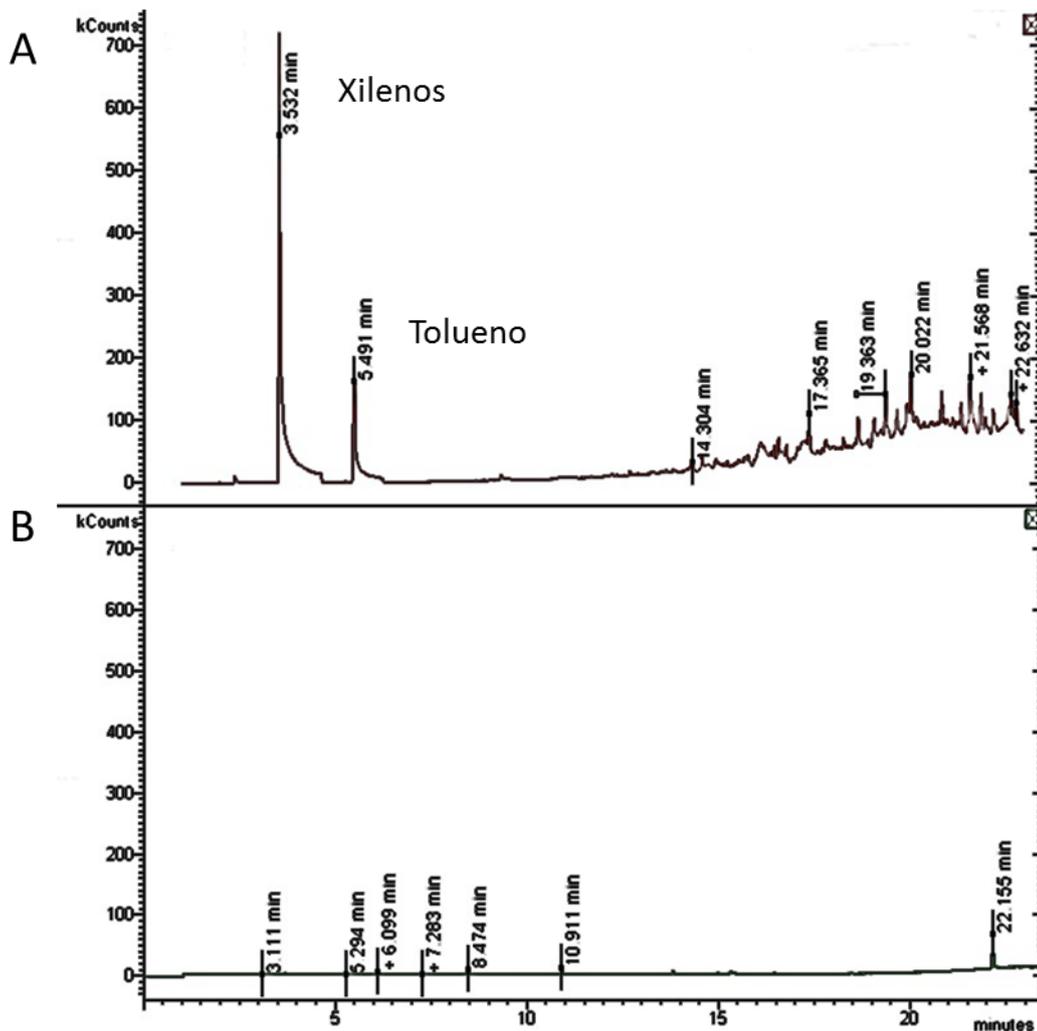
As soluções foram adicionadas em fluxo laminar para a composição do meio de cultura.

4 RESULTADOS

4.1 Análise de solo

A análise de solo se deu através de análise físico-química para um levantamento das características gerais do solo como granulometria, íons presentes, quantidade de material orgânico (Anexo A) e uma análise voltada a contaminantes orgânicos com foco em BTEX. O resultado da análise de contaminantes orgânicos está mostrado na Figura 7.

Figura 7 – Análise de BTEX nos solos dos locais contaminado e Perequê por meio de técnica GC-MS



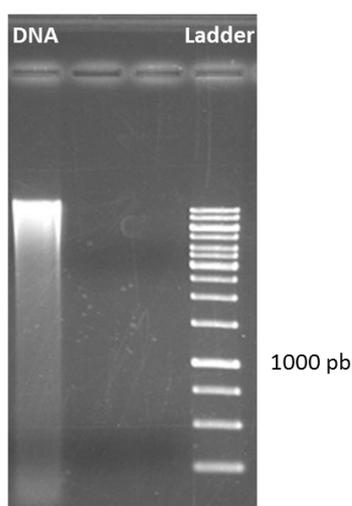
Análise de contaminação de solo por BTEX utilizando-se tecnologia de camada gasosa após extração padrão de hidrocarbonetos. **A)** Solo contaminado. **B)** Solo do Perequê. Os picos apresentados aos 3 e 5 minutos aproximadamente se referem à presença de xileno e tolueno respectivamente. Os outros picos menos proeminentes representam hidrocarbonetos presentes no solo não relacionados à contaminação estudada.

Os dois tipos de extração utilizados revelaram contaminação do solo. A análise de *headspace* do solo se mostrou mais sensível à contaminação mostrando picos maiores nos tempos de retenção do xileno e tolueno. As análises de *headspace* líquido e sólido foram realizadas com a mesma amostra de solo. Apesar de a análise ser voltada para os quatro componentes de BTEX incluindo os 3 tipos de xilenos, não foi possível verificar a presença de benzeno e etilbenzeno nas amostras.

4.2 Identificação de procariotos por amplificação do gene SSU rRNA

Este primeiro sequenciamento realizado no projeto foi feito a partir da extração direta do DNA presente no solo. O resultado da extração está representado na Figura 8.

Figura 8 – Fotodocumentação da extração de DNA total de solo



Resultado da extração de DNA total de solo. Na canaleta esquerda foram aplicados 5 microlitros de material genético extraído do solo do local contaminado. Na canaleta apresentada à direita está o marcador molecular.

Para todas as amostras de solo coletadas foi padronizada a extração de material genômico. O perfil do material para migração em gel de agarose se manteve o mesmo para os três locais analisados, apresentando maioria do material próximo à marcação de 10.000 pb do marcador molecular. O rendimento de material genômico apresentou variação de concentração entre 2 μ g e 5 μ g por grama de solo. As menores concentrações foram apresentadas pelo solo contaminado.

O material genético extraído foi processado e deu origem às sequências metagenômicas que foram analisadas pelo banco de dados do RDP. No total foram

geradas 762 sequências já selecionadas por qualidade e tendo o vetor retirado. Dessas 762 sequências, 470 foram identificadas como sendo provenientes do domínio Bacteria e 292 sequências provindas do domínio Archaea. A Tabela 6 mostra o número de sequências utilizadas para cada um dos metagenomas e seus respectivos domínios.

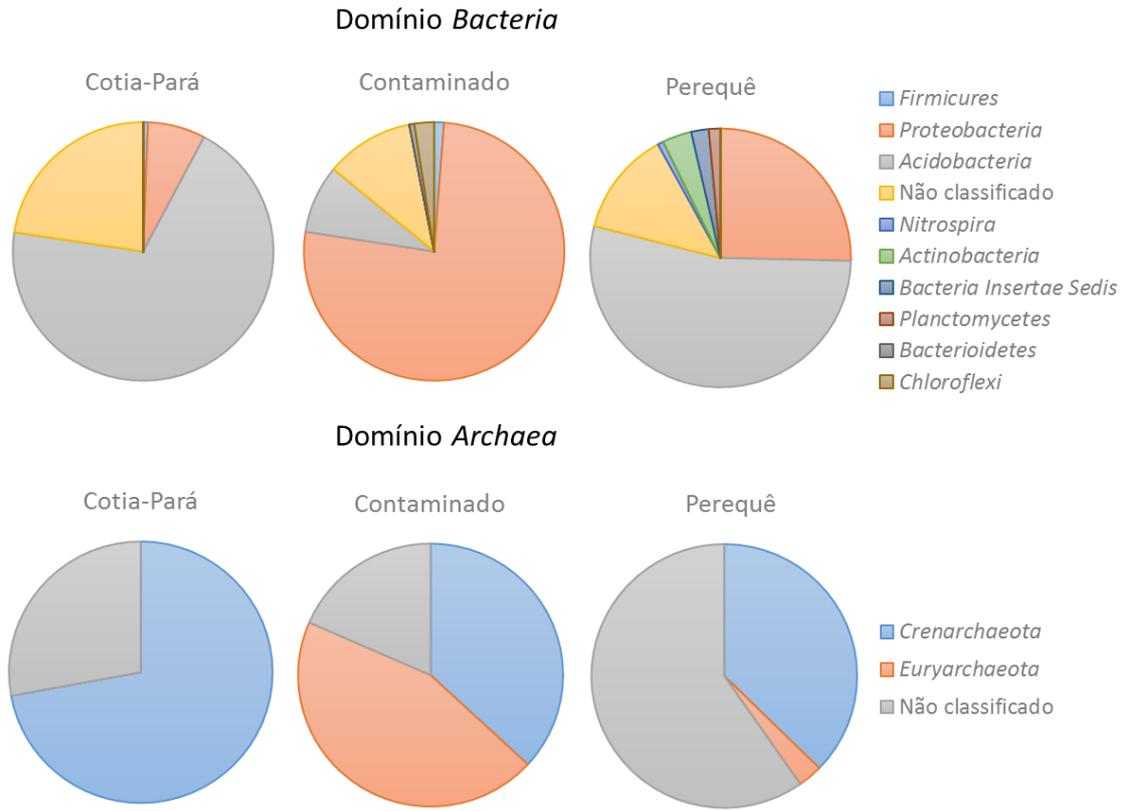
Tabela 6 – Total de sequências analisadas pelo método Sanger

<i>Domínios</i>	<i>Locais</i>		
	Cotia-Pará	Perequê	Contaminado
Bacteria	168	138	164
Archaea	122	67	103
Total	290	203	267

A tabela dispõe as sequências analisadas levando-se em consideração os locais de coleta e os domínios Bacteria e Archaea.

A distribuição das sequências de acordo com os filos identificados está mostrada na Figura 9.

Figura 9 – Distribuição percentual das sequências em filios para os domínios Bacteria e Archaea



As sequências foram identificadas no banco de dados do RDP e distribuídas percentualmente para cada um dos metagenomas.

As análises das sequências geradas pelo método Sanger indicam diferenças na composição da estrutura procariótica, como pode ser visto na figura acima. Para o domínio Bacteria os filios mais representados foram Proteobacteria e Acidobacteria. O local contaminado foi aquele em que o filo Proteobacteria mais estava representado enquanto o filo Acidobacteria se mostrou mais abundante em Cotia-Pará e em Perequê. Este último local foi o que apresentou maior diversidade comparado aos outros dois locais.

Diferenças também puderam ser vistas para o domínio Archaea. O filo Crenarchaeota foi o mais abundante para o Cotia-Pará, o filo Euryarchaeota foi o mais abundante no local contaminado e o Perequê apresentou como grupo mais expressivo arqueias que não puderam ser identificadas.

4.3 Identificação de procariotos por sequenciamento direto de DNA do solo

Foram conseguidas no total 1.710.216 sequências para os três metagenomas. A contabilização das sequências está mostrada na Tabela 7.

Tabela 7 – Contabilidade das sequências geradas pelo sequenciamento GS FLX

	<i>Cotia-Pará</i>	<i>Perequê</i>	<i>Contaminado</i>
<i>Pares de bases</i>	260.105.297	305.140.693	311.978.477
<i>Número de sequências</i>	533.697	578.478	598.041
<i>Tamanho médio</i>	487	527	521

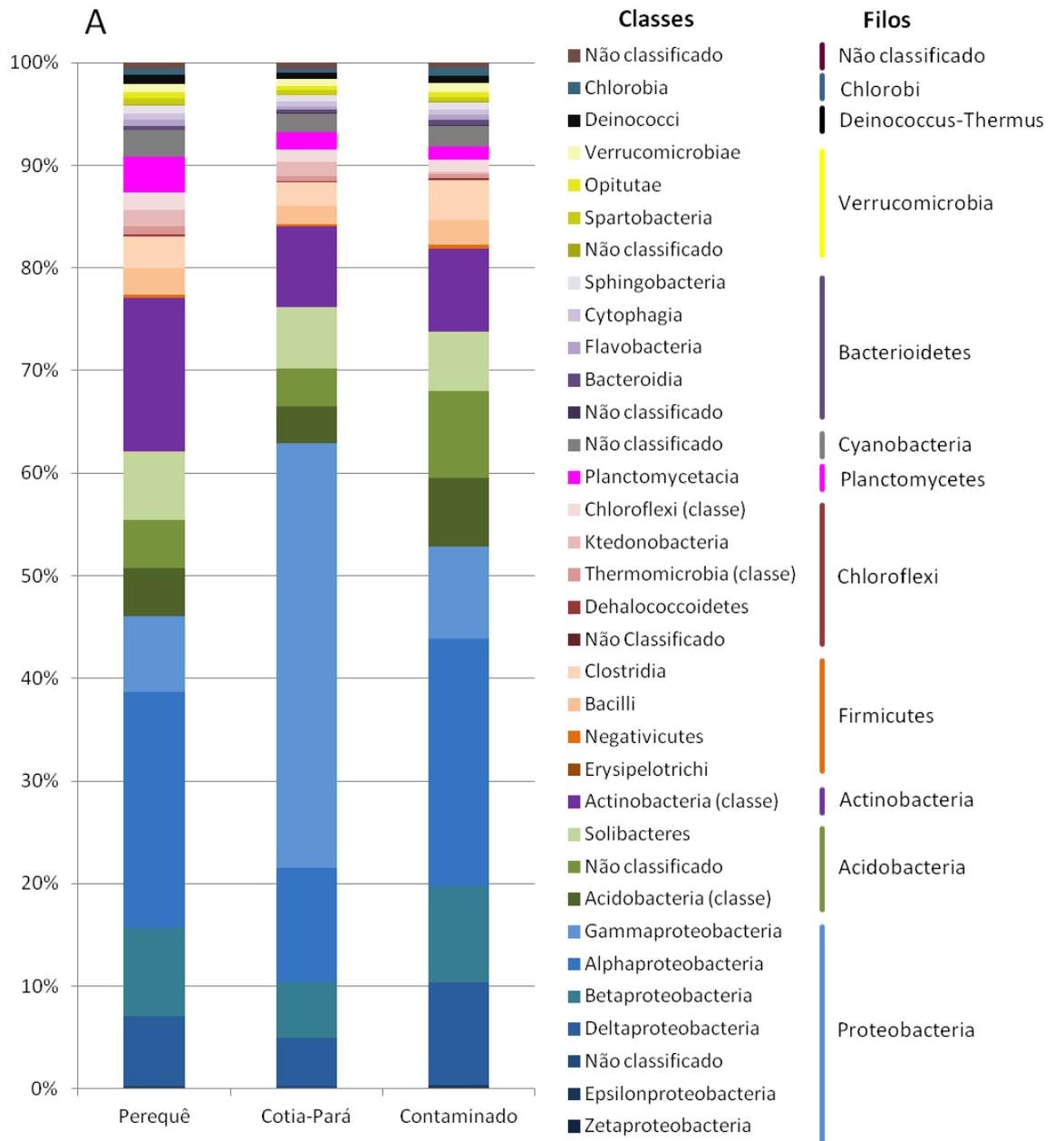
As informações apresentadas na tabela foram recuperadas do servidor do MG-RAST onde foram processadas. A contabilidade foi realizada pelo próprio programa.

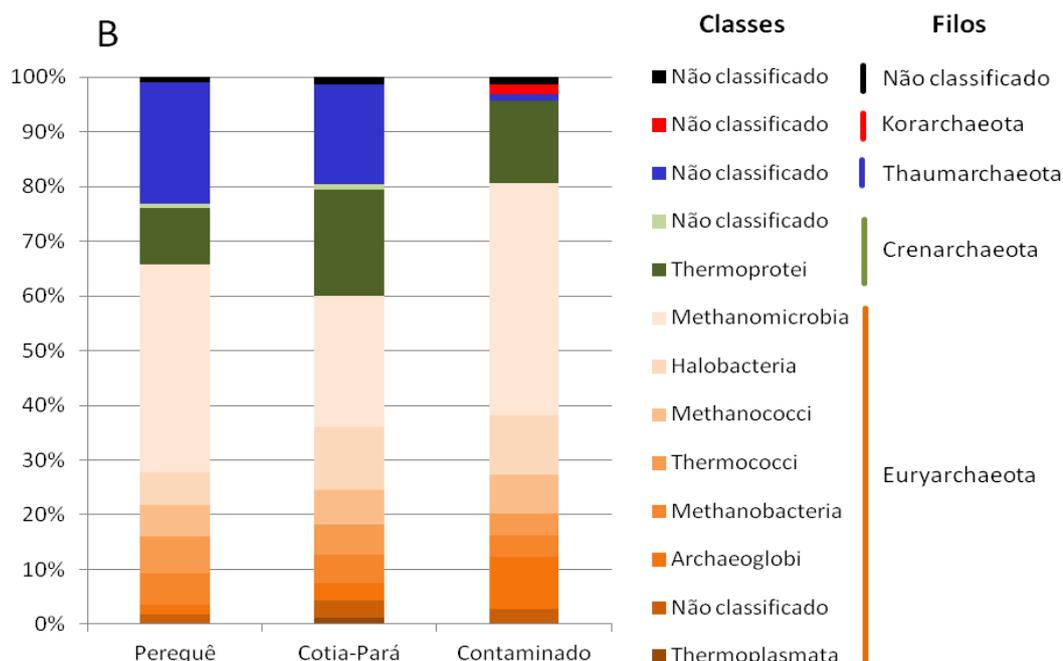
O número de sequências apresentadas na tabela se refere ao número bruto de sequências providas do sequenciador 454. Essas sequências apenas passaram por seleção de qualidade de sequenciamento pelo próprio sistema. Posteriormente foram dirigidas para análises secundárias de qualidade e identificação dentro de cada objetivo.

4.3.1 Panorama geral da diversidade de procariotos dos locais estudados

Com a finalidade de se identificar e conhecer melhor a diversidade procariótica dos solos, assim como avaliar a sensibilidade do método de identificação metagenômica aliado ao sequenciamento de alta performance, uma identificação geral foi proposta. Através deste procedimento foram identificados os filos e classes presentes nos diferentes locais estudados e a comparação entre eles nos mostrou a diversidade geral de procariotos de solos da região (Figura 10).

Figura 10 – Identificação geral das sequências metagenômicas pelo MG-RAST com base no GenBank





Identificação das sequências metagenômicas pelo programa MG-RAST com base no banco de dados do M5NR. O gráfico representa a distribuição percentual dos filos e classes encontrados para cada um dos três diferentes solos estudados. Somente os filos cuja soma dos três locais ultrapassou mil sequências estão mostrados na figura. O gráfico **A** representa a distribuição das sequências identificadas dentro do domínio Bacteria. O gráfico **B** representa a distribuição das sequências identificadas dentro do domínio Archaea.

Esta abordagem conseguiu identificar 371.284, 498.998 e 405.847 sequências para os locais Perequê, Cotia-Pará e contaminado respectivamente. Dentro do domínio Bacteria foram identificados 27 filos e um agrupamento que reuniu sequências que não puderam ser identificadas. Nenhum local apresentou filos. Quanto à distribuição das sequências, os filos Proteobacteria, Actinobacteria e Acidobacteria foram responsáveis sozinhos por aproximadamente 80% das sequências e esta distribuição foi vista para todos os locais analisados. A maior diferença encontrada ficou por conta do filo Proteobacteria para o local contaminado, responsável por mais de 60% das sequências deste local. A classe Gammaproteobacteria, a mais abundante do local contaminado, foi destaque. Apesar desta classe ter sido identificada nos outros sítios estudados, sua abundância não foi tão expressiva se comparada ao solo contaminado, característica exclusiva deste local.

Bem menos expressivo (por volta de 1% das sequências), o domínio Archaea apresentou 5 filos e um agrupamento que reuniu sequências não identificadas. As principais diferenças vistas ficaram por conta do filo Thaumarchaeota, que apresentou abundância relativa de apenas 1% das sequências de Archaea do local contaminado enquanto nos locais Cotia-Pará e Perequê apresentaram respectivamente 18% e 20%

das sequências identificadas. O filo Korarchaeota, apesar da baixa abundância, só foi encontrado no local contaminado.

4.3.2 Comparação de diferentes métodos para refinamento da identificação taxonômica de sequências

Para refinamento da identificação metagenômica e já vislumbrando uma possível relação entre os procariotos e os seus respectivos locais, foram realizadas no total 3 tipos de identificações taxonômicas mais estridentes para classificação das sequências. Essas identificações, como já mencionado no item 3.8.2.1, tiveram bases conceituais diferentes, utilizaram diferentes bancos de dados e produziram diferentes quantidades de sequências identificadas (Tabela 8).

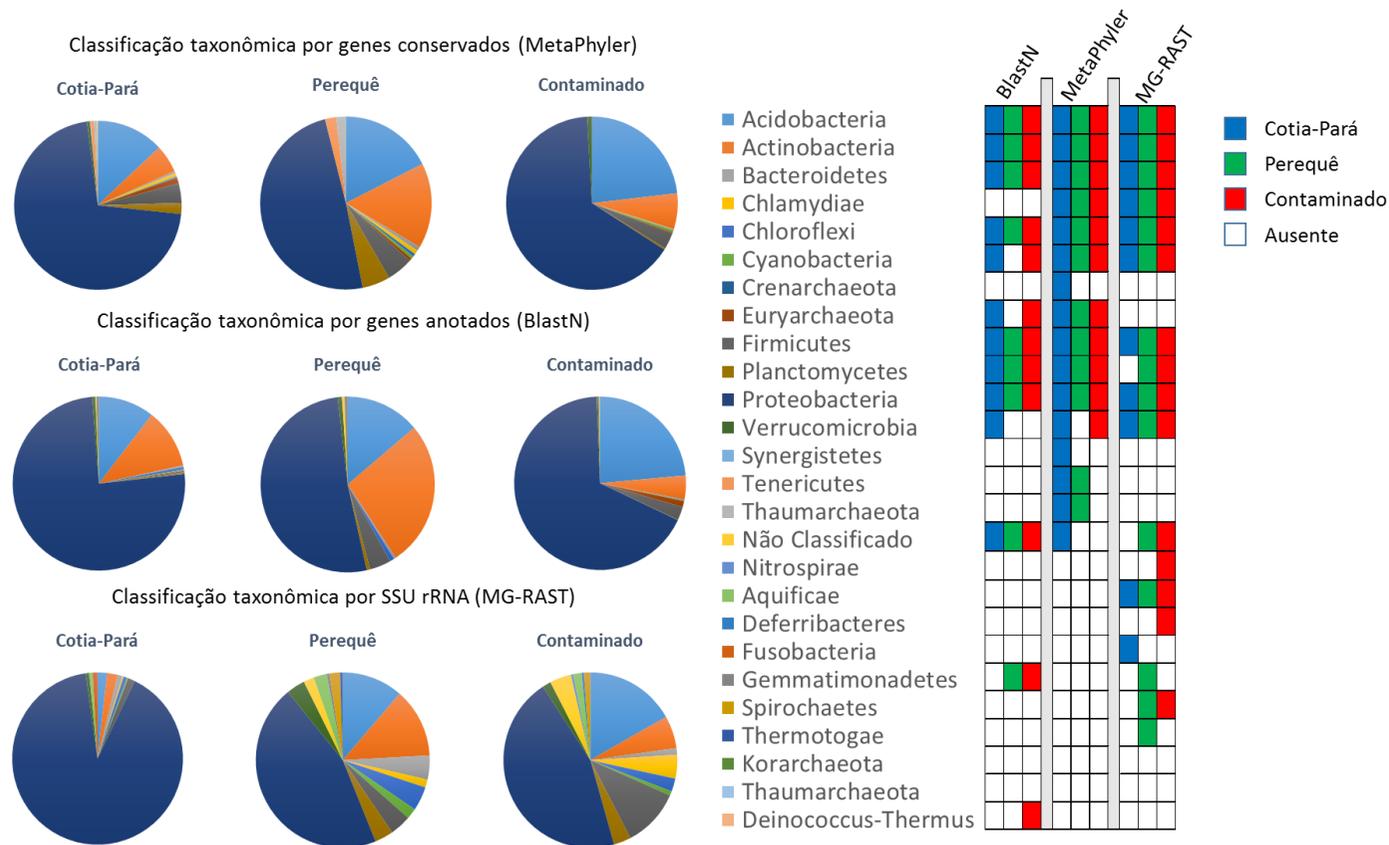
Tabela 8 – Contabilidade das sequências identificadas pelos diferentes métodos de análises taxonômicas

Local	SSU rRNA	Genes conservados	BLASTN
<i>Cotia-Pará</i>	469	990	3760
<i>Perequê</i>	242	382	1365
<i>Contaminado</i>	274	1002	7214

Método de SSU rRNA realizado pelo programa MG-RAST através da submissão das sequências no servidor. Método de genes conservados realizado pelo programa MetaPhyler. Método de identificação por BlastN realizado pela comparação com banco de dados de genomas procarióticos do NCBI.

As sequências identificadas pelos diferentes métodos de classificação taxonômica foram agrupadas por filos. Isto foi feito com o objetivo de uma primeira comparação entre grupos de organismos encontrados nos diferentes metagenomas, e também como uma comparação dos resultados obtidos pelos diferentes métodos utilizados para uma identificação mais precisa. A distribuição das sequências está representada na Figura 11.

Figura 11 – Distribuição das sequências metagenômicas classificadas em filios por diferentes métodos comparativos



Os gráficos refletem a porcentagem de sequências encontradas para cada filo identificado nos diferentes metagenomas.

À direita se encontra uma tabela de cores representando os filios encontrados em cada um dos metagenomas de acordo com os diferentes métodos de classificação abordados. BLASTN pela comparação ao banco de dados do NCBI, MetaPhyler pela comparação ao banco de genes conservados e MG-RAST pela comparação ao banco de dados do RDP. Os quadros em branco representam a ausência do respectivo filo.

Os diferentes métodos de classificação resultaram em modos distintos de interpretar a composição de procariotos dos diferentes metagenomas. Com relação ao número de filios, para todos os metagenomas estudados, foram encontrados 12 filios pela comparação por BLASTN, 15 filios pela comparação por genes conservados e 18 pela comparação por MG-RAST, levando-se em consideração os três metagenomas estudados. O método de comparação que mais identificou sequências foi o método por BLASTN, porém, o método MG-RAST, pelo qual menos sequências foram identificadas, revelou uma maior diversidade dos locais estudados.

Filios muito bem representados como Acidobacteria e Proteobacteria foram detectados pelas três técnicas utilizadas, porém, as técnicas apresentaram sensibilidades diferentes. O filo Tenericutes, por exemplo, só foi identificado pela técnica de comparação ao banco de genes conservados, e o filo Aquificae só foi identificado pela técnica MG-RAST. Por essa técnica, o local Perequê foi o de maior diversidade de filios, enquanto que pelas técnicas de BLASTN e MetaPhyler, o local Cotia-Pará foi o mais diverso. A tabela da Figura 11 mostra detalhadamente os filios encontrados pelas técnicas utilizadas.

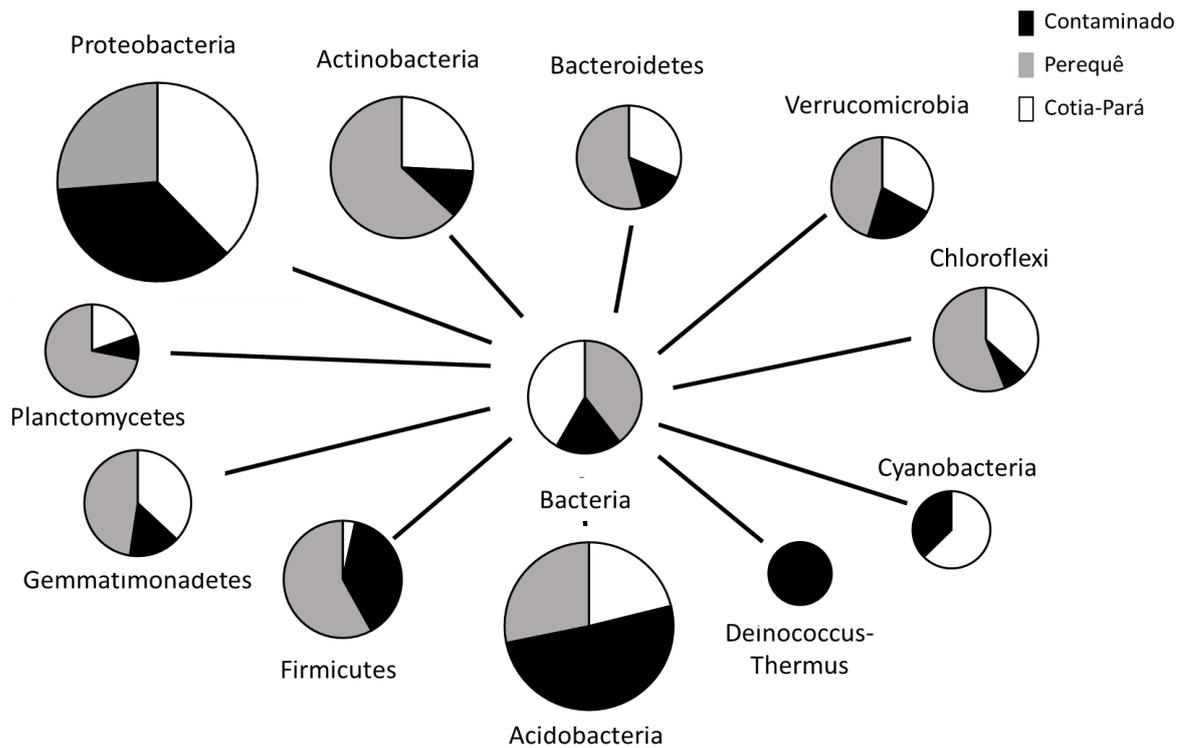
Outro aspecto que apresentou variação relativa ao tipo de técnica de identificação utilizada foi a porcentagem de cada filo encontrada em cada metagenoma. Os filios mais representados como Proteobacteria, Acidobacteria e Actinobacteria foram confirmados como tal pelas três técnicas, porém, diferenças na quantidade identificada variaram. O filo Proteobacteria foi o melhor exemplo. Pelas técnicas MetaPhyler e BLASTN, esse grupo apresentou uma representatividade por volta de 70% para o local Cotia-Pará; pela técnica do MG-RAST, essa representatividade chegou a 91%. Para o filo Actinobacteria, a comparação realizada pelo MG-RAST apresentou 13% de sequências identificadas para o local Perequê, enquanto as outras técnicas revelaram uma porcentagem acima de 20%.

Para todas as técnicas foi notada uma variação na quantidade de filios quanto ao número de sequências distribuídas entre eles. As técnicas BLASTN e MetaPhyler foram as que apresentaram resultados mais próximos, tanto com relação ao número de filios encontrados quanto à distribuição das sequências dentro deles. Para continuidade de identificações taxonômicas, a técnica de identificação por BLASTN foi selecionada. Devido à coerência entre as técnicas de BLASTN e MetaPhyler, e também à quantidade de sequências utilizadas para identificação, essa técnica foi escolhida

para uma identificação mais detalhada da composição procariótica dos locais estudados.

A distribuição em filios encontrada para as sequências identificadas através de BLASTN está mostrada na Figura 12.

Figura 12 – Distribuição de sequências em filios para os três locais estudados pelo método BLASTN



Distribuição das sequências identificadas por BLASTN em filios dentro do domínio Bacteria. Cada círculo representa um filio encontrado e a variação de tamanho do círculo é proporcional à quantidade de sequências identificadas em cada um deles. Cada cor representa um metagenoma estudado. A distribuição mostrada das sequências está normalizada permitindo relações tanto entre filios quanto entre metagenomas. Os resultados foram baseados no programa MEGAN.

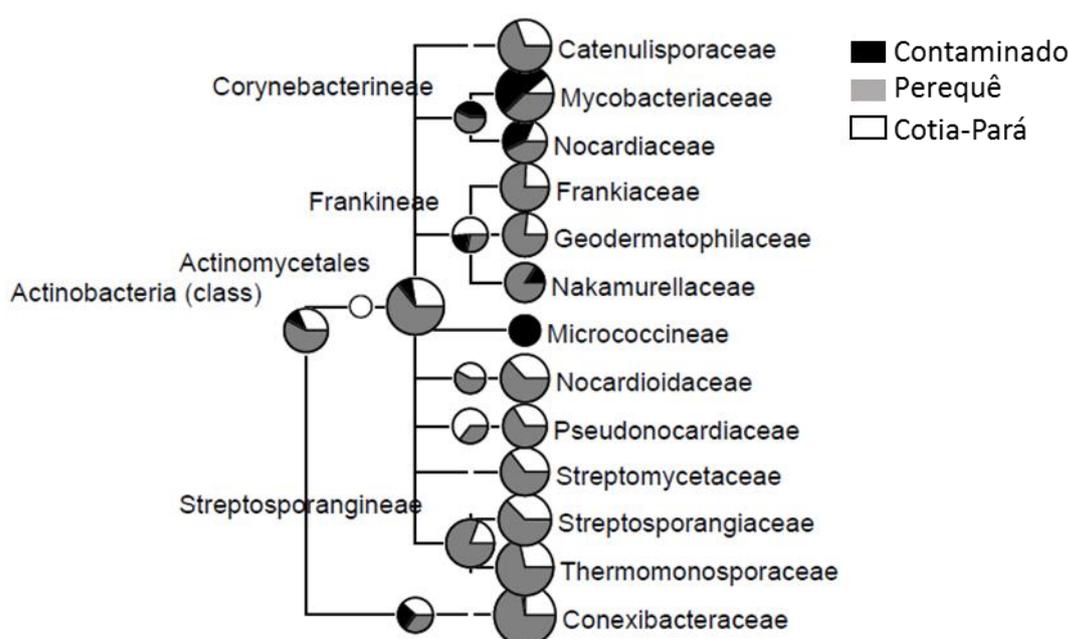
Por essa abordagem de classificação, onze filios bacterianos foram encontrados para os três metagenomas estudados. A maioria dos filios foi encontrada em todos os locais estudados exceto os filios Cyanobacteria e Deinococcus-Thermus. O filio Cyanobacteria foi encontrado apenas nos locais Cotia-Pará e contaminado, e o filio Deinococcus-Thermus foi identificado apenas no local contaminado. Pela figura podemos notar sequências que não puderam ser identificadas de acordo com os limiares utilizados. Estas ficaram agrupadas na base, no nó central da figura (Bacteria) agrupando 3% das sequências para o local contaminado e 6% para os locais Cotia-

Pará e Perequê. O tamanho dos círculos reflete a quantidade de sequências encontradas para cada um dos filios. Os filios Proteobacteria, Acidobacteria, Actinobacteria e Firmicutes foram os filios com maior número de sequências atribuídas.

Das sequências identificadas, o local Perequê apresentou a maior abundância de sequências em sete dos onze filios encontrados, porém, o local com maior diversidade segundo a análise por BLASTN foi o local contaminado.

Os quatro maiores grupos foram escolhidos para uma análise mais detalhada dos subgrupos que os compõem. A distribuição de sequências dentro da classe Actinobacteria está representada na Figura 13.

Figura 13 – Distribuição das sequências em subgrupos da classe Actinobacteria



A figura representa os subgrupos da classe Actinobacteria encontrados para cada um dos metagenomas. As sequências foram normalizadas e organizadas pelo programa MEGAN. O tamanho dos círculos refletem a quantidade de sequências em cada um deles. As sequências que não puderam ser classificadas em famílias foram incluídas em níveis anteriores.

A classe Actinobacteria agrupou 4%, 24% e 9% das sequências classificadas para os locais contaminado, Perequê e Cotia-Pará, respectivamente. Ao todo foram identificadas 13 famílias nos metagenomas estudados. Das 13 famílias, apenas 5 foram identificadas no local contaminado (Figura 13) e a família Micrococcineae foi exclusiva.

A família Mycobacteriaceae apresentou maior representatividade no local contaminado comparativamente aos outros locais estudados.

Os locais Cotia-Pará e Perequê apresentaram as mesmas famílias com exceção da família Nakamurellaceae, que só foi identificada no Perequê e local contaminado. O Perequê, de modo geral, apresentou maior quantidade relativa de organismos da classe Actinobacteria, assim como a maior diversidade.

Os filios Acidobacteria e Firmicutes também foram bastante representados nos metagenomas analisados e suas famílias estão indicadas na Figura 14.

Figura 14 – Distribuição das seqüências dos filios Acidobacteria e Firmicutes em famílias

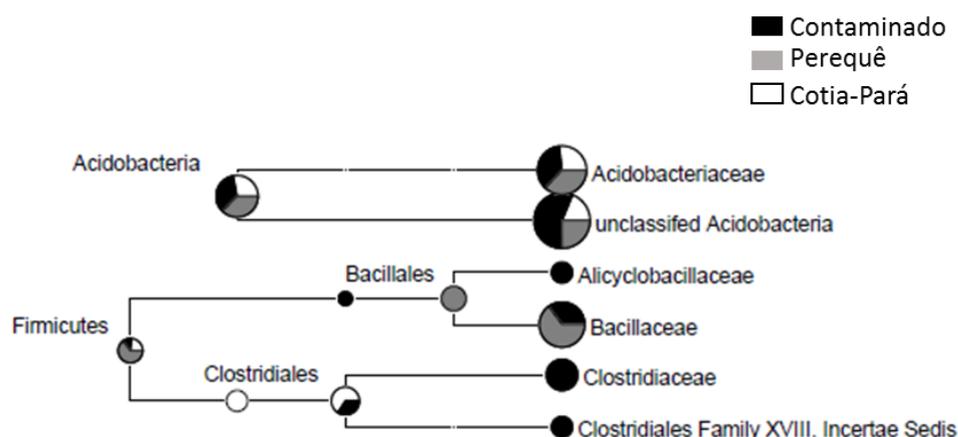


Figura representando a divisão em famílias classificadas dentro dos filios Firmicutes e Acidobacteria. Os tamanhos dos círculos representam a quantidade de seqüências atribuídas a cada um deles. A classificação e organização das seqüências foi feita pelo programa MEGAN.

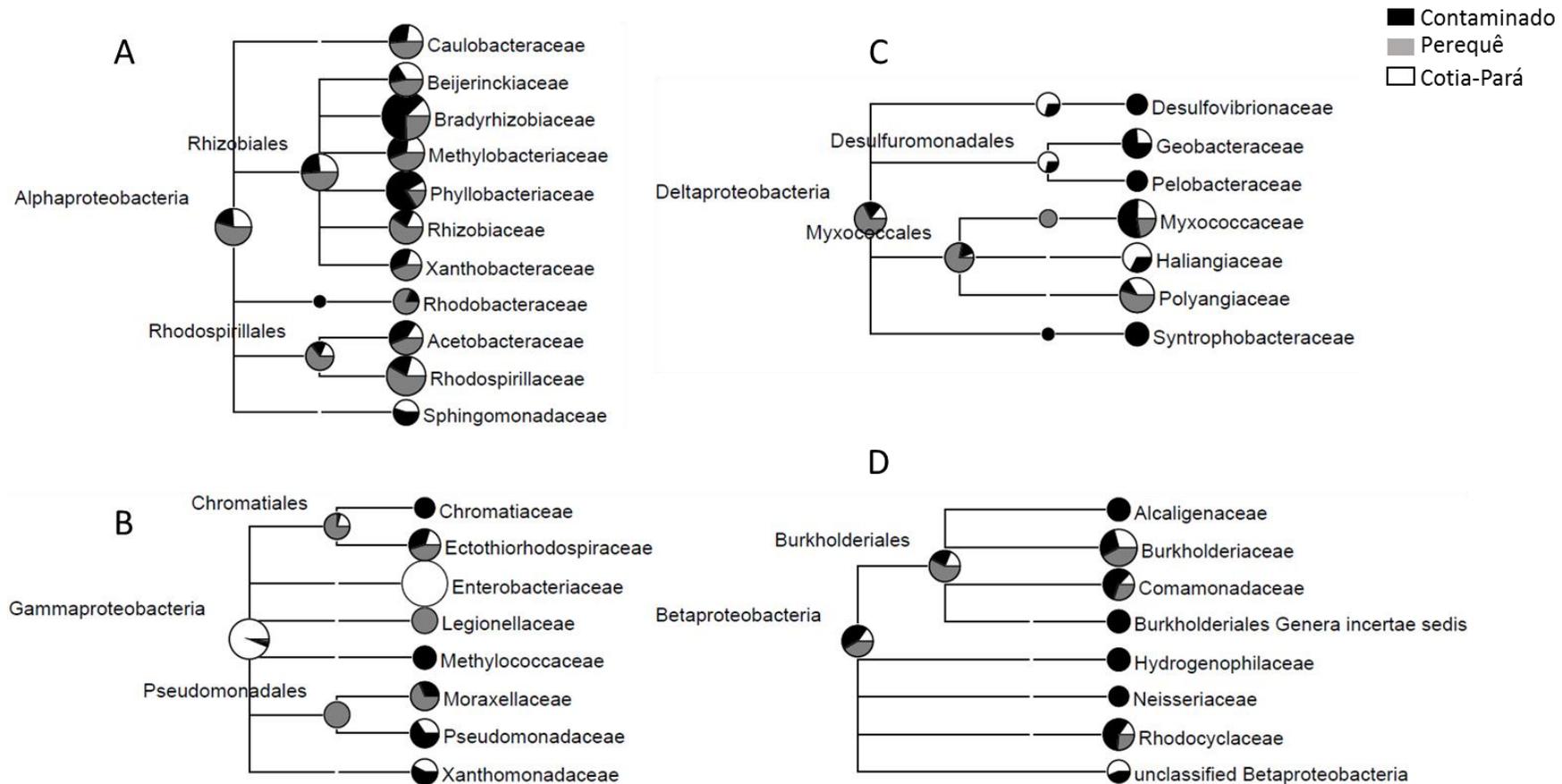
O filo Acidobacteria foi mais representativo no local contaminado com 22% das seqüências identificadas; para os locais Perequê e Cotia-Pará foram identificadas 12% e 9% do total de seqüências para cada um dos respectivos locais. Este filo foi constituído de apenas uma família (Acidobacteriaceae) e um agrupamento de organismos não classificados dentro do filo. O grupo não classificado foi o mais numeroso dentro do filo Acidobacteria para todos os locais analisados, porém, uma quantidade cerca de duas vezes maior foi encontrada para o local contaminado. O filo Acidobacteria apresentou abundância relativa equivalente nos três locais com uma quantidade um pouco maior no Perequê. Esse filo foi representado nos três locais analisados somente pelo gênero *Acidobacterium*.

O filo Firmicutes se mostrou presente nos três locais analisados, com maior abundância para o local Perequê. Este filo agrupou 2%, 3% e 0,1% das seqüências identificadas para os locais contaminado, Perequê e Cotia-Pará respectivamente.

Quatro famílias foram identificadas e as quatro estavam presentes no local contaminado, aquele que teve maior diversidade de famílias dentro do filo Firmicutes. Das quatro famílias identificadas, três foram exclusivas deste local. A família Bacillaceae foi a mais abundante e o local que mais apresentou essa família foi o Perequê. Como a figura mostra, o filo Firmicutes não foi muito representado no local Cotia-Pará, não sendo possível identificar nenhuma família. Existiram sequências que fizeram parte da classe Clostridiales mas que não puderam ser melhor identificadas.

O último e mais abundante filo entre todos os filios identificados foi o filo Proteobacteria. Com 64%, 46% e 67% do total de sequências identificadas para os locais contaminado, Cotia-Pará e Perequê. Este filo apresentou grandes diferenças quando uma análise mais específica foi proposta (Figura 15).

Figura 15 – Distribuição de seqüências do filo Proteobacteria em famílias



Distribuição das seqüências atribuídas ao filo Proteobacteria em famílias. O tamanho dos círculos representa o número de seqüências atribuídas para cada um dentro do mesmo agrupamento. As seqüências foram normalizadas e classificadas pelo programa MEGAN e distribuídas dentro das quatro classes do filo Proteobacteria representadas acima pelas quatro árvores da figura.

Dentro do filo Proteobacteria a classificação das sequências destacou diferenças entre os locais analisados, relativas tanto a abundâncias quanto a diversidades encontradas. Na classe Alphaproteobacteria (Figura 15A), onze famílias foram encontradas para os três locais analisados. Nenhuma dessas famílias foi exclusiva de determinado local, mas houveram locais que não apresentaram algumas famílias como Rhodobacteraceae e Sphingomonadaceae. Das famílias que foram encontradas em todos os locais, a maior abundância relativa foi encontrada no Perequê (com exceção de Bradyrhizobiaceae e Phyllobacteriaceae). A família Bradyrhizobiaceae se apresentou com uma abundância muito maior no local contaminado apresentando 2546 sequências, o que corresponde a 35% de todas as sequências identificadas para o local. Uma análise mais aprofundada deste grupo é mostrada mais adiante.

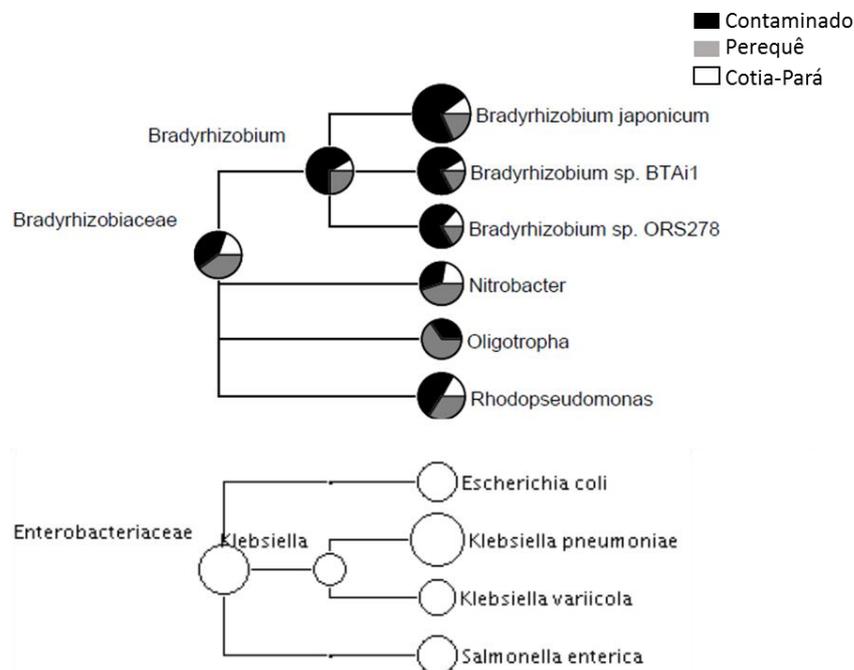
A classe Gammaproteobacteria (Figura 15B) apresentou oito famílias para os locais estudados. A maior diversidade desse grupo foi identificada para o local contaminado, o qual apresentou seis famílias. Entre essas, as famílias Chromatiaceae e Methylococcaceae foram exclusivas. Os locais Perequê e Cotia-Pará apresentaram uma família exclusiva cada. A família Enterobacteriaceae identificada para o local Cotia-Pará foi um dos maiores agrupamentos encontrados com 1293 sequências, o que representa 34% das sequências identificadas para este local. Este grupo foi melhor identificado e será mostrado mais adiante. Para esse local, pudemos observar uma grande quantidade de sequências que não puderam ser identificadas, ficando posicionadas em nós mais basais.

A classe Deltaproteobacteria (Figura 15C) teve grande diversidade identificada no local contaminado. Quanto à abundância relativa, esta família apresentou equivalência nos locais estudados com 4%, 4% e 3% para os locais contaminado, Perequê e Cotia-Pará respectivamente. Todas as sete famílias identificadas foram encontradas nesse local. Destas, três foram exclusivas do local contaminado. Dados parecidos foram encontrados para a classe Betaproteobacteria (Figura 15D), com 4%, 3% e 2% das sequências identificadas para os locais contaminado, Perequê e Cotia-Pará, respectivamente. Esta classe apresentou nove famílias identificadas e todas elas foram encontradas no local contaminado, sendo quatro exclusivas dele. Este grupo não foi muito representado nos outros lugares estudados, merecendo destaque apenas a família Burkholderiaceae, pela quantidade relativa encontrada no local Perequê.

De modo geral, o filo Proteobacteria foi bastante representado no local contaminado, apresentando muitos grupos exclusivos, assim como quantidade de sequências identificadas. Podemos destacar dentro desse grupo a quantidade de sequências que não puderam ter uma classificação mais detalhada no Perequê e que ficaram agrupadas nos nós mais basais. Quanto ao local Cotia-Pará, apenas o grupo Enterobacteriaceae se destacou, reunindo a maior quantidade de sequências.

Os grupos mais representados para os diferentes locais foram estudados de forma um pouco mais detalhada e uma tentativa de identificação mais específica foi realizada. A Figura 16 apresenta esta classificação.

Figura 16 – Classificação dos grupos mais representados em gêneros utilizando BLASTN



Identificação mais aprofundada das sequências atribuídas às famílias Bradyrhizobiaceae (A) e Enterobacteriaceae (B). As sequências foram classificadas e organizadas através do programa MEGAN. Os tamanhos dos círculos representam a quantidade de sequências agrupadas em cada um deles dentro de cada árvore.

A família Bradyrhizobiaceae foi a que mais apresentou sequências identificadas para o local contaminado. Quatro gêneros bacterianos foram encontrados para essa família e o gênero *Bradyrhizobium* foi o mais representativo deles. O número de sequências atribuídas ao gênero *Bradyrhizobium* para o local contaminado foi de 2073, o que representa 29% do total de sequências identificadas para o local contaminado.

O segundo grupo mais abundante foi o filo Acidobacteria, o qual agrupou 1627 sequências. Apenas o gênero *Bradyrhizobium* apresentou mais sequências do que o filo Acidobacteria inteiro. Dentro do gênero, três espécies puderam ser identificadas (Figura 16A), e a espécie *Bradyrhizobium japonicum* foi responsável por 19% das sequências identificadas para o local contaminado. Apesar de não ser exclusivo do local contaminado, a representatividade do gênero *Bradyrhizobium* fez com que esse fosse o grupo-chave de estudo do local. Esta posição alcançada mereceu testes metabólicos para elucidação dos motivos da presença desse micro-organismo no local contaminado.

A família Enterobacteriaceae (Figura 16B) foi um grupo bastante interessante que se destacou no local Cotia-Pará. Além de se apresentar em abundância com um total de 1293 sequências, ele foi exclusivo do local, motivos que levaram a uma identificação mais específica das sequências. No total foram encontrados três gêneros identificados como *Escherichia*, *Salmonella* e *Klebsiella*, este último o mais abundante agrupando 19% das sequências identificadas para o local contaminado. Devido às características metabólicas e habitats onde os gêneros identificados geralmente são encontrados, foram relacionados e identificados possíveis motivos para a presença desses gêneros, que serão mostrados mais adiante.

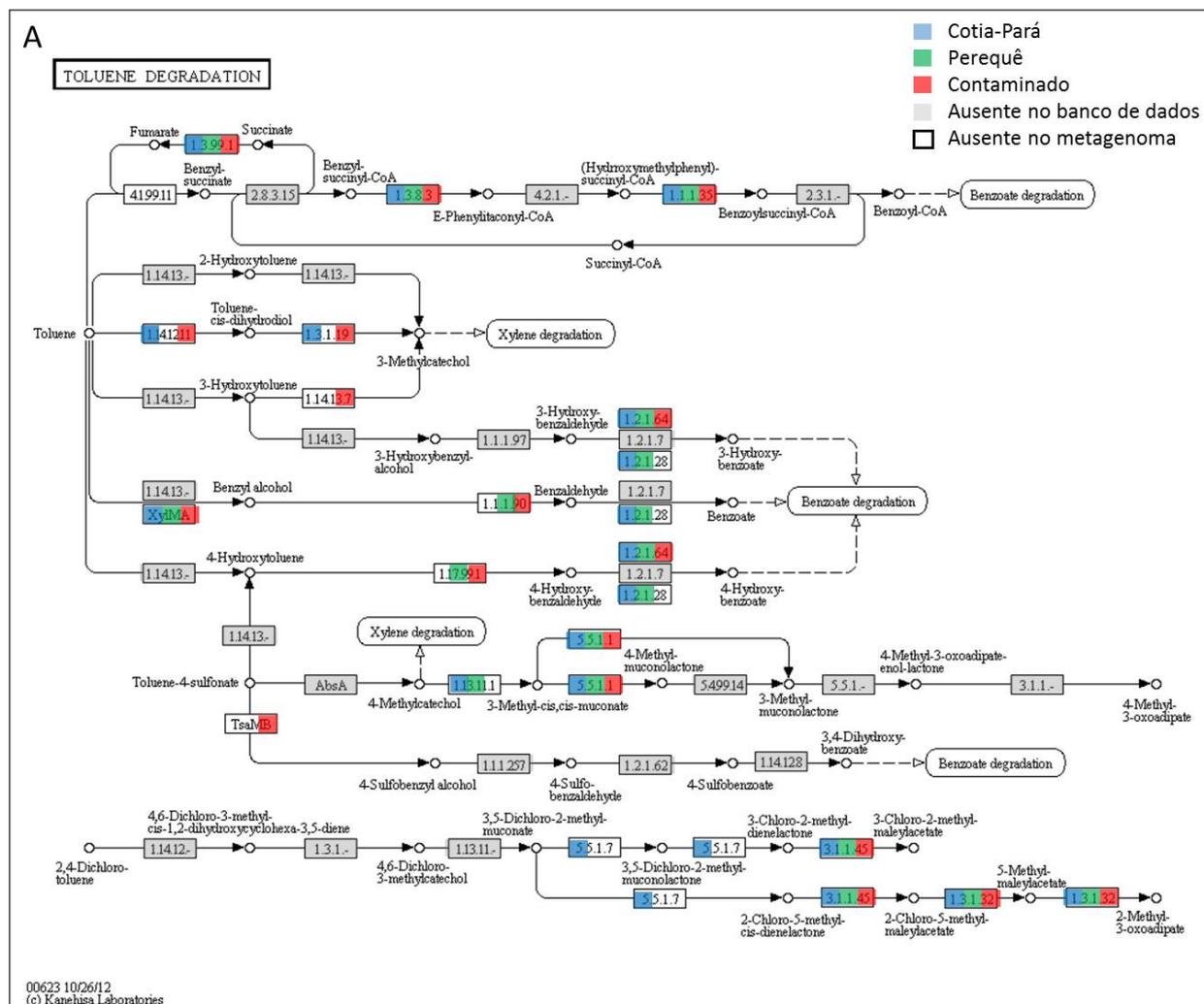
4.3.3 Identificação metabólica de sequências.

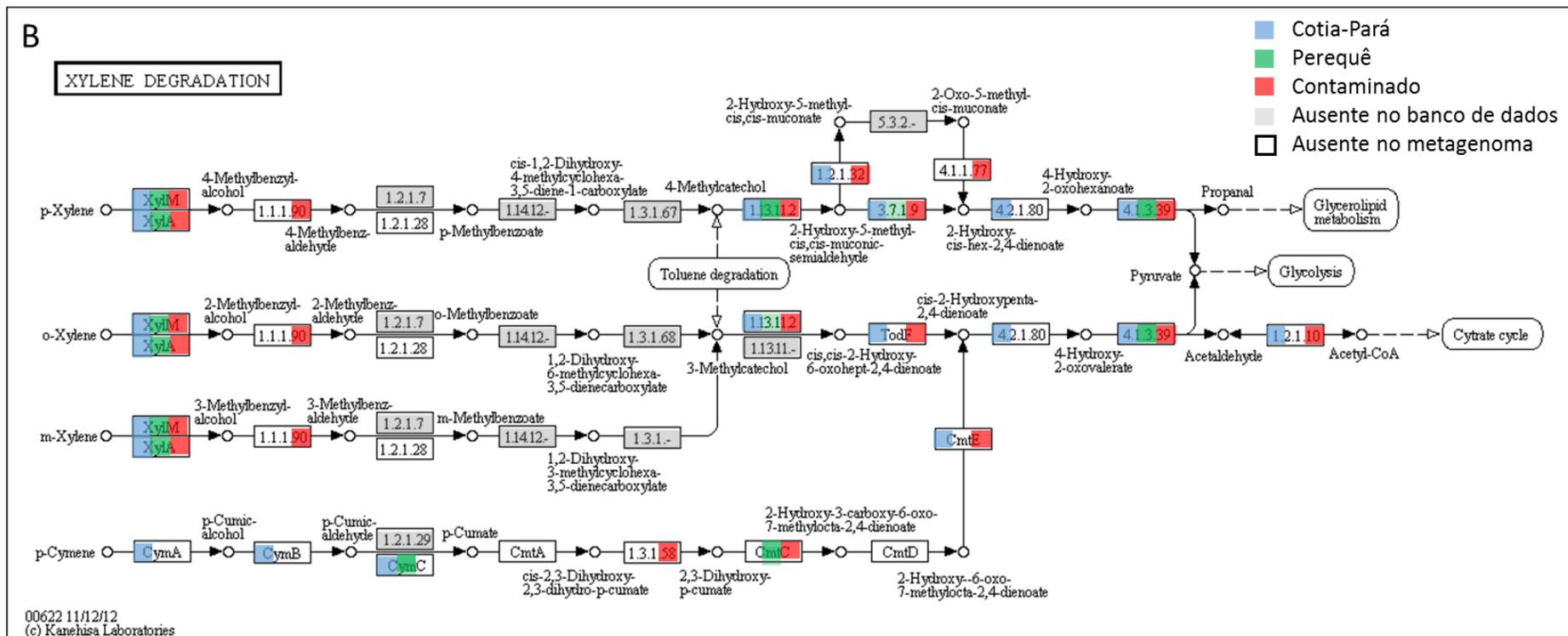
Além da possibilidade de identificação taxonômica de procariotos, a técnica metagenômica utilizada permitiu também uma análise focada em genes presentes no ambiente. Através da identificação desses genes, inferências metabólicas puderam ser realizadas baseadas nos genes encontrados. Devido ao município de Cubatão possuir um histórico de contaminação por compostos aromáticos de petróleo e por uma das áreas estudadas apresentar contaminação com hidrocarbonetos aromáticos, foram estudados e comparados metabolismos de degradação de substâncias aromáticas. O foco foi dado em diferenças quanto à presença de genes de degradação e quantidade de sequências relacionadas aos genes nos locais estudados.

Nossas análises realizadas no local contaminado revelaram a presença dos hidrocarbonetos xilenos e tolueno. Devido a este fato, os metabolismos de degradação desses hidrocarbonetos foram utilizados para comparações entre os metagenomas dos locais estudados. Inicialmente todos os genes das vias de degradação dos

hidrocarbonetos foram comparados aos metagenomas para uma identificação geral. Os genes encontrados foram marcados nos metabolismos para verificação da presença ou ausência dos genes nos locais. A Figura 17 mostra os genes encontrados dos metabolismos de xilenos e tolueno nos locais estudados.

Figura 17 – Identificação dos genes das vias de degradação de xilenos e tolueno nos metagenomas



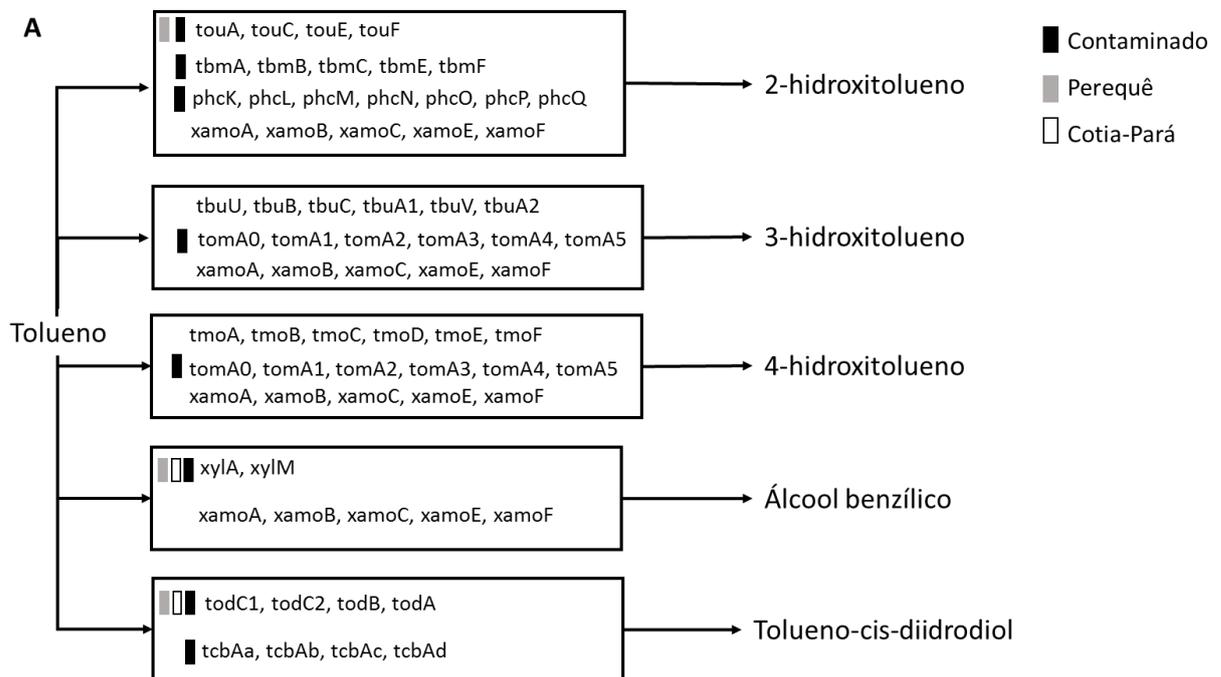


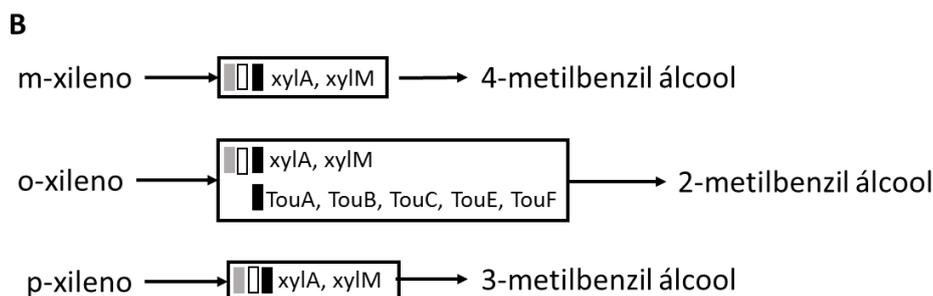
Representação dos genes presentes nas vias de degradação de tolueno (Figura 17A) e xilenos (Figura 17B) baseado no metabolismo encontrado no banco de dados do KEGG. Sequências relativas aos genes que realizam os passos metabólicos de degradação foram contabilizadas e os passos metabólicos marcados no próprio mapa. Cada metagenoma foi marcado com uma cor correspondente e os genes que não possuíam sequências nos bancos de dados foram marcados em cinza. Sequências que não apresentaram similaridade suficiente com genes presentes na via de degradação ou que não estavam presentes no metagenoma não foram marcados.

Pelas análises realizadas, os metagenomas apresentaram sequências com similaridade aos genes presentes nas vias de degradação dos dois hidrocarbonetos. Nos dois metabolismos, um maior número de sequências com similaridade aos genes das vias foi encontrado nos metagenomas dos locais Cotia-Pará e contaminado. No local Perequê, apesar de apresentar sequências correspondentes a estes metabolismos, a diversidade de genes foi um pouco menor. Os passos iniciais das vias de degradação pelo mapa metabólico do KEGG não estão muito definidos e um estudo para distinguir sequências que poderiam estar relacionadas ao começo das vias foi realizado.

Os genes foram estudados com base no banco de dados do Bionemo. Os operons responsáveis pela degradação de tolueno e xilenos foram recuperados e comparados aos metagenomas estudados. A Figura 18 identifica os operons estudados, seus genes e os locais onde foram encontrados.

Figura 18 – Identificação dos operons de degradação de tolueno e xilenos





Identificação dos operons responsáveis pela degradação de tolueno (Figura 18A) e xilenos (Figura 18B). A figura indica os operons e seus respectivos caminhos para a degradação de cada um dos hidrocarbonetos. As caixas coloridas na frente de cada operon representam o local onde cada um foi identificado.

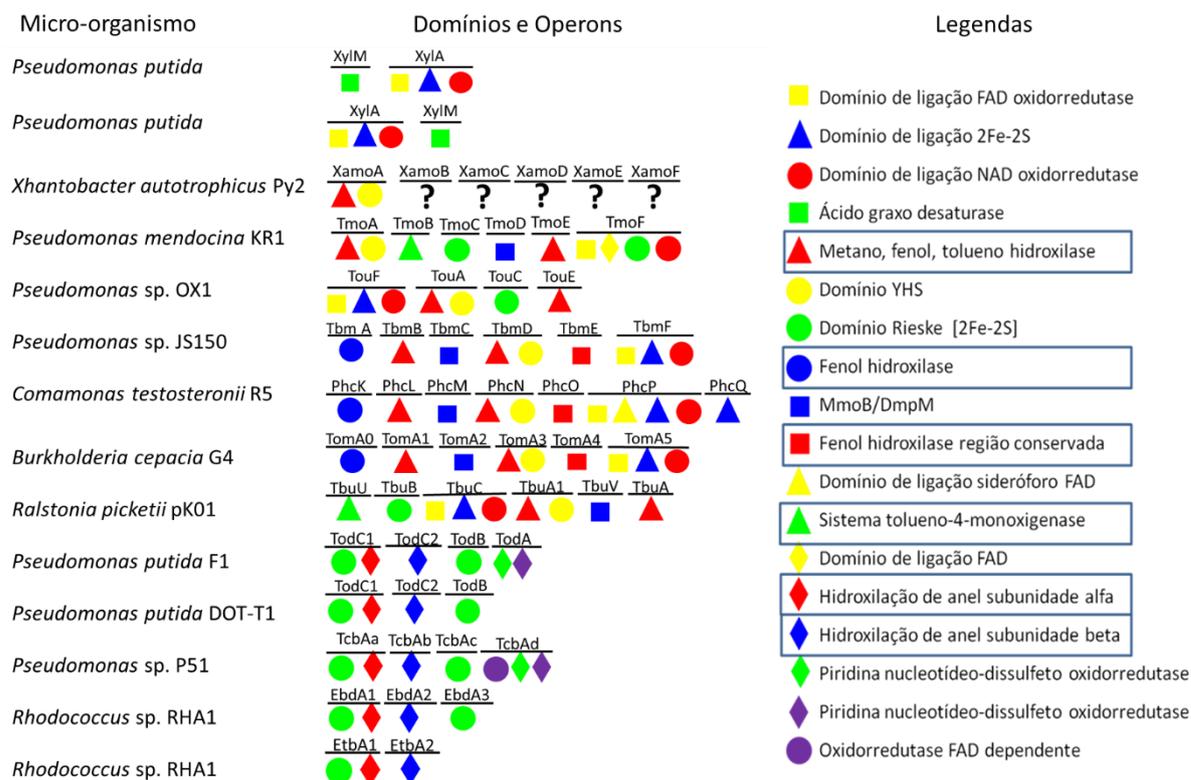
A Figura 18 apresenta de forma mais detalhada o início das vias de degradação de tolueno e xilenos. Ela relaciona os operons responsáveis pela degradação e a direção que os contaminantes podem seguir quando são utilizados metabolicamente por procaríotos. Como apresentado primeiramente na Figura 17A, o passo metabólico 1.14.13.- é realizado pelos operons mostrados na Figura 18A, responsáveis pelo primeiro processamento de tolueno além de passos metabólicos seguintes.

Fica claro pela observação da figura acima que a maioria dos operons de degradação selecionados do banco de dados do Bionemo foram identificados no local contaminado. Dos onze operons identificados para degradação de xilenos e tolueno, sete foram encontrados no local contaminado e apenas dois e três foram encontrados nos locais Cotia-Pará e Perequê, respectivamente. Quanto ao número de sequências encontradas, uma grande diferença foi observada para os locais analisados. Para o local Cotia-Pará e contaminado foram contabilizadas 53 e 56 sequências respectivamente, enquanto, para o local Perequê, apenas 3 sequências foram contabilizadas.

As análises realizadas envolvendo os operons selecionados foram baseadas até aqui em comparações de identidade entre as sequências dos operons e os metagenomas. Para um avanço no estudo, foi utilizado o conceito de regiões conservadas funcionais entre os genes (PFAM). Isto levou a uma melhor análise dos operons e um método distinto de comparação que poderia trazer informações mais aprofundadas.

Para esse estudo, uma identificação dos domínios conservados dos genes que compõem os operons foi realizada (Figura 19). Não apenas os operons de degradação

de toluenos e xilenos entraram na análise, outros operons de degradação de anéis aromáticos foram utilizados para melhor identificação dos domínios de degradação nos metagenomas estudados.

Figura 19 – Identificação dos domínios conservados PFAM para operons selecionados

Representação gráfica dos genes contidos nos operons de degradação estudados. Na coluna micro-organismo estão identificados os procarionotos cujos operons estudados (coluna domínios e operons) foram identificados. Nas representações dos operons foram identificados os genes de cada um deles e os domínios conservados presentes em cada um dos genes. À direita se encontra a legenda para cada um dos domínios encontrados (coluna legendas). Os domínios que possuem caixas em seu entorno, na coluna legendas, foram os domínios selecionados para estudo.

Catorze operons foram selecionados no total sendo conhecidos pela degradação do grupo BTEX e fenol. Esses operons apresentaram dezoito domínios conservados presentes nos genes que os formam. Foi observado que muitos domínios de degradação se repetiram em alguns operons como foi o caso dos operons ácido graxo desaturase e metano, fenol, tolueno hidroxilase.

Alguns operons se mostraram bastante similares quanto à presença de domínios conservados como os operons encontrados em *Burkholderia cepacia* G4 e *Comamonas testosteronii* R5. Até o quinto gene de seus operons, os domínios conservados foram os mesmos, diferenciando-se quanto ao tipo e número de domínio conservado somente adiante desse gene. Outro exemplo de similaridade de operons foi o domínio encontrado em *Pseudomonas putida*, existindo dois operons que possuem os mesmos genes e domínios conservados, porém, apresentados em diferentes ordens.

Muitos dos domínios conservados identificados para os operons estudados são domínios comuns, que se encontram em diversos outros operons, com funções completamente distintas. Para otimizar a verificação de domínios de degradação nos metagenomas estudados, os domínios que modificavam diretamente os anéis aromáticos foram escolhidos como mostrado na Figura 19. Esses domínios foram identificados (Tabela 9) e utilizados para comparação por HMMs nos metagenomas.

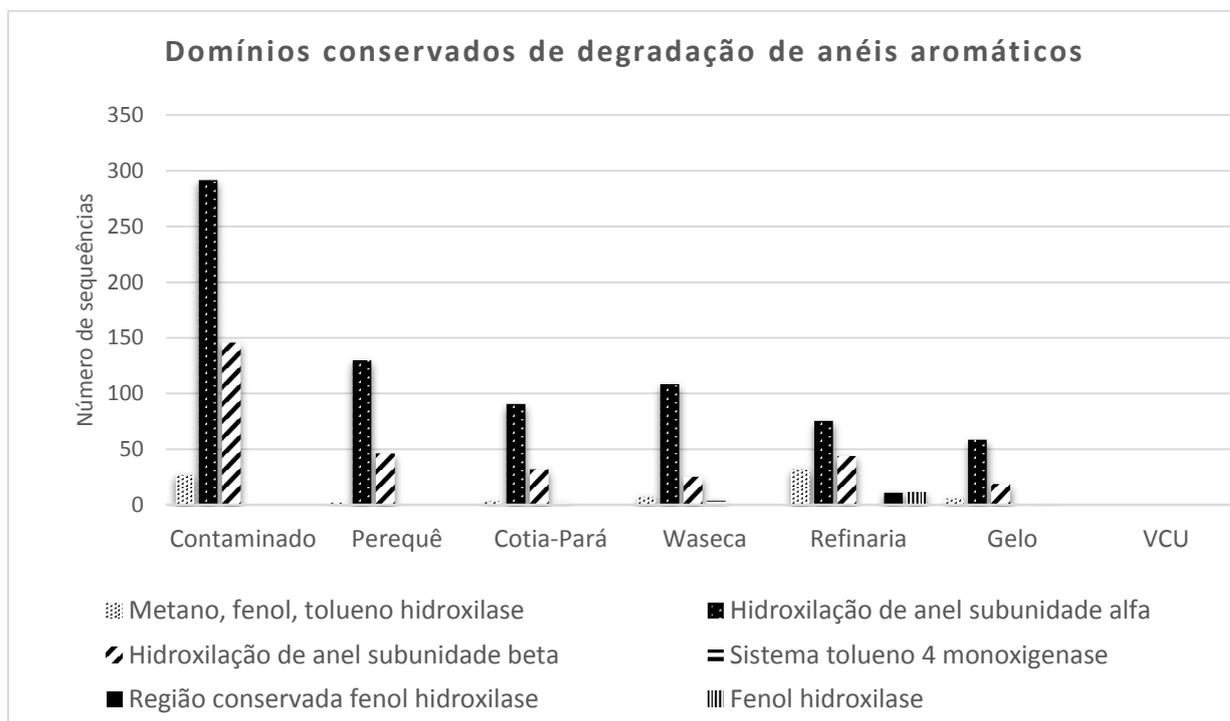
Tabela 9 – Identificação dos domínios conservados de modificação de anéis aromáticos

Domínios	Função
<i>Metano, fenol, tolueno hidroxilase</i>	Cataboliza fenol e derivados metilados
<i>Fenol hidroxilase</i>	Conversão de fenol em catecol
<i>Subunidade alfa e beta de hidroxilação de anel</i>	Constituem dioxigenases. Transformam estruturas de anéis fechados em cis-diol não aromáticos
<i>Fenol hidroxilase região conservada</i>	Região conservada entre monoxigenases Catalisa a reação fenol – catecol.
<i>Sistema tolueno-4-monoxigenase</i>	Hidroxilação do tolueno para formação de p-cresol

As informações das funções dos domínios conservados foram retiradas do banco de dados PFAM.

A comparação dos domínios conservados se deu por HMMs e apenas sequências que possuíram *e-value* menor ou igual a $1e-10$ foram contabilizadas. Como método de comparação adicional, três metagenomas distintos foram também utilizados para identificação dos domínios conservados escolhidos (Figura 20).

Figura 20 – Contabilização e identificação de domínios conservados em diferentes metagenomas



O gráfico indica a distribuição de sequências com similaridade aos domínios conservados identificados entre os metagenomas. Os metagenomas Refinaria (número de acesso SRX038779), Gelo (número de acesso SRX000607), Waseca (número de acesso AAFX00000000) e VCU (número de acesso SRR049666) são metagenomas depositados no NCBI escolhidos para comparação. Os resultados foram normalizados para um total de 500.000 sequências por metagenoma.

Para uma comparação entre metagenomas, foram escolhidos quatro metagenomas externos de variados ambientes. Os metagenomas batizados de Refinaria (SILVA et al., 2012), Gelo (SIMON et al., 2009), Waseca (TRINGE et al., 2005) e VCU, um estudo da microbiota presente no ambiente vaginal de seres humanos cujo estudo está depositado no NCBI.

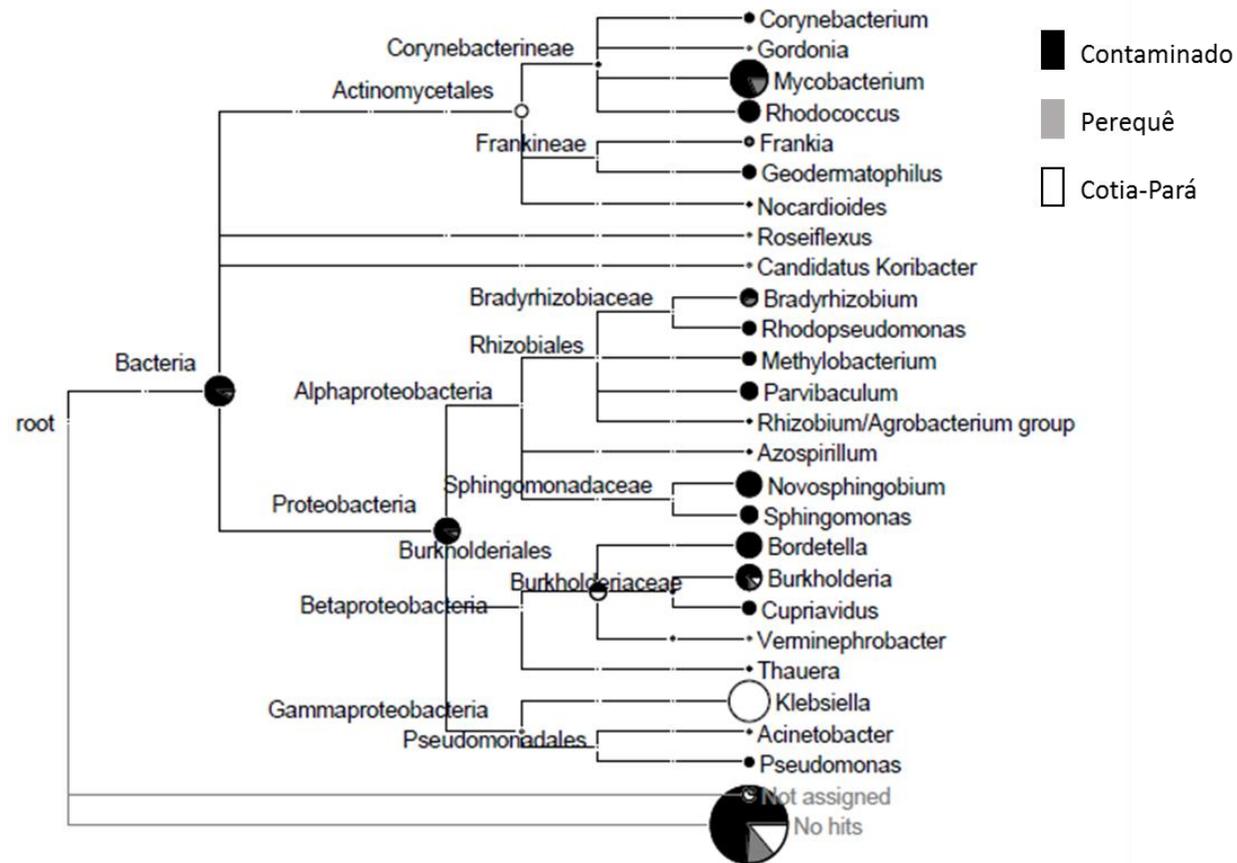
O gráfico revela que o local contaminado apresentou o maior número de sequências relacionadas aos domínios conservados de degradação selecionados. A maioria das sequências ficou relacionada aos domínios de hidroxilação de anel alfa e beta. Esses dois domínios foram os mais representados em todos os metagenomas estudados exceto o metagenoma VCU, que não apresentou sequências relacionadas a nenhum dos domínios selecionados.

O domínio metano, fenol, tolueno hidroxilase foi identificado na maioria dos metagenomas, e os que apresentaram maior número de sequências relacionadas foram o metagenoma Refinaria e o local contaminado. No metagenoma Refinaria, os

domínios região conservada fenol hidroxilase e fenol hidroxilase foram encontrados em quantidades maiores que os outros metagenomas, e o domínio fenol hidroxilase foi encontrado apenas nesse metagenoma. O local Cotia-Pará apresentou o domínio sistema tolueno-4-monoxigenase, sendo o único a apresentar esse domínio.

As sequências que apresentaram similaridade com os domínios conservados selecionados para os metagenomas de nosso trabalho foram separadas e comparadas novamente aos genomas bacterianos. Através deste passo, foi realizada uma tentativa de identificação dos possíveis procariotos relacionados às sequências selecionadas. Os resultados apresentados foram organizados no programa MEGAN e os gêneros bacterianos foram identificados (Figura 21).

Figura 21 – Identificação taxonômica das sequências selecionadas com similaridade aos domínios conservados de degradação de BTEX

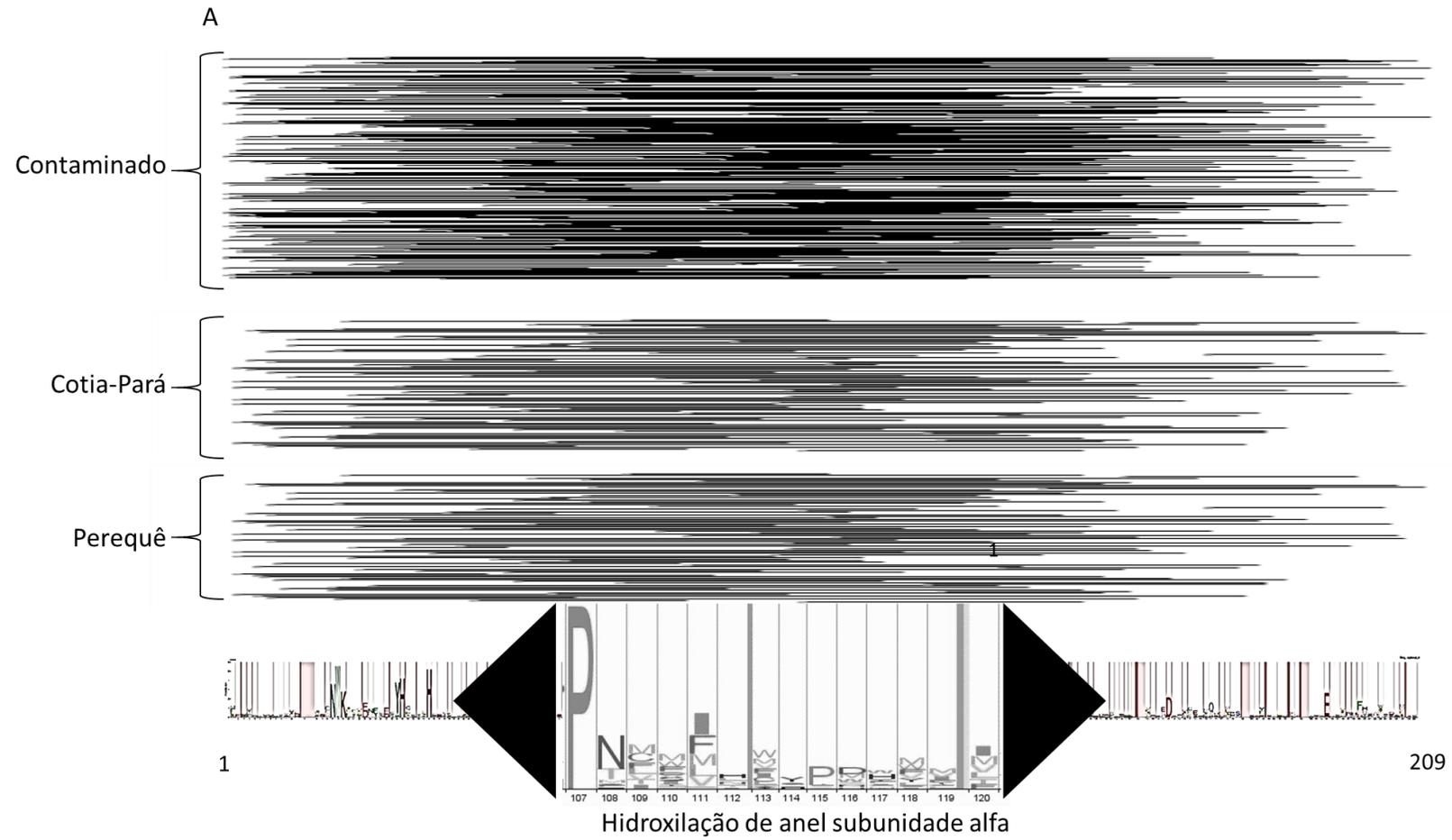


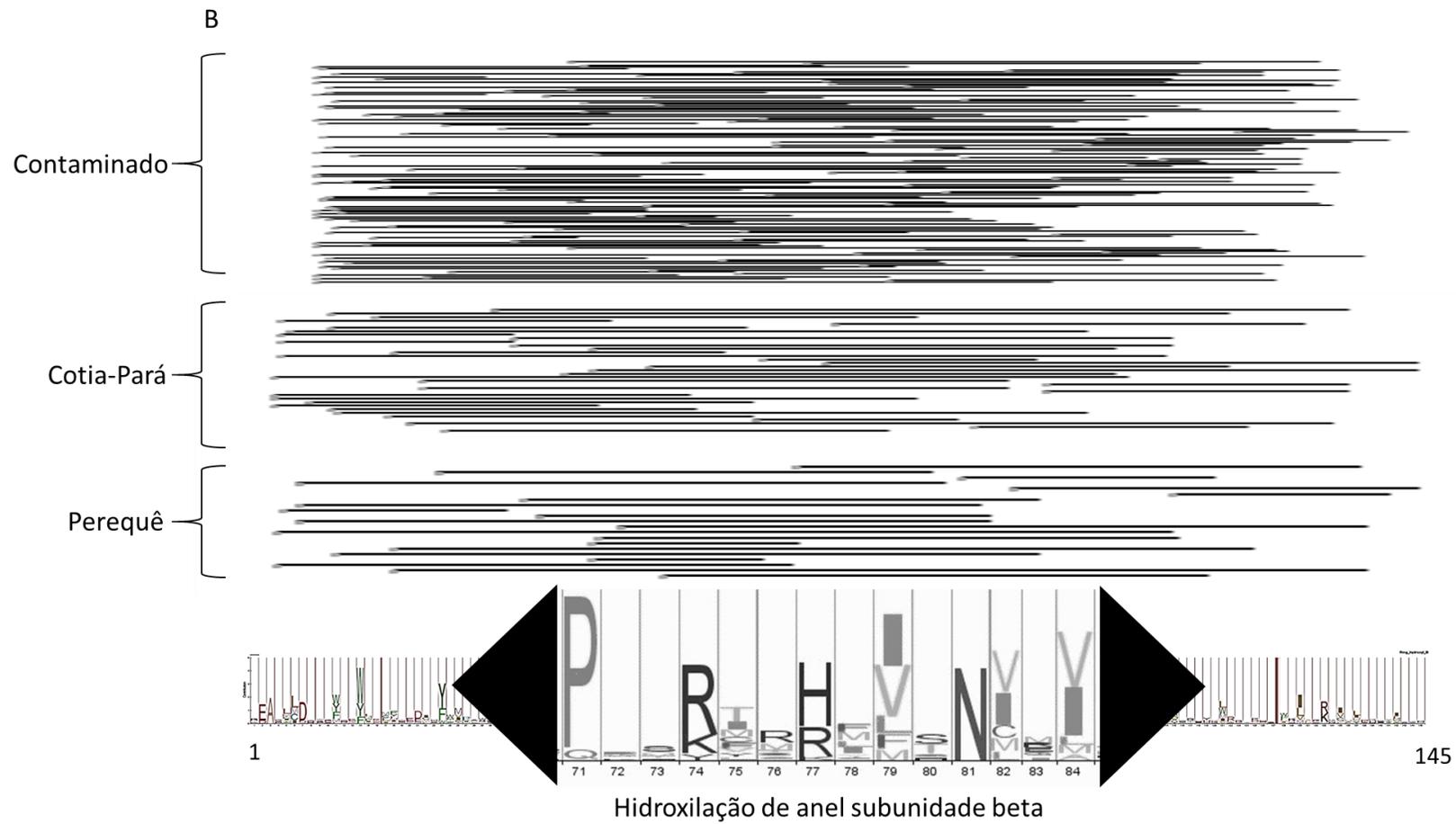
A comparação entre as sequências selecionadas foi realizada pelo programa MEGAN. Os tamanhos dos círculos representam a quantidade de sequências atribuídas a cada um. Os metagenomas Cotia-Pará, Perequê e local contaminado estão sendo mostrados na mesma árvore até o nível de gênero. O nó *unclassified sequences* apresenta sequências que tiveram similaridade com sequências ainda não identificadas no banco do NCBI. O nó *not assigned* apresenta sequências que não puderam ser classificadas e o nó *no hits* reúne sequências que não apresentaram similaridade alguma com sequências do banco.

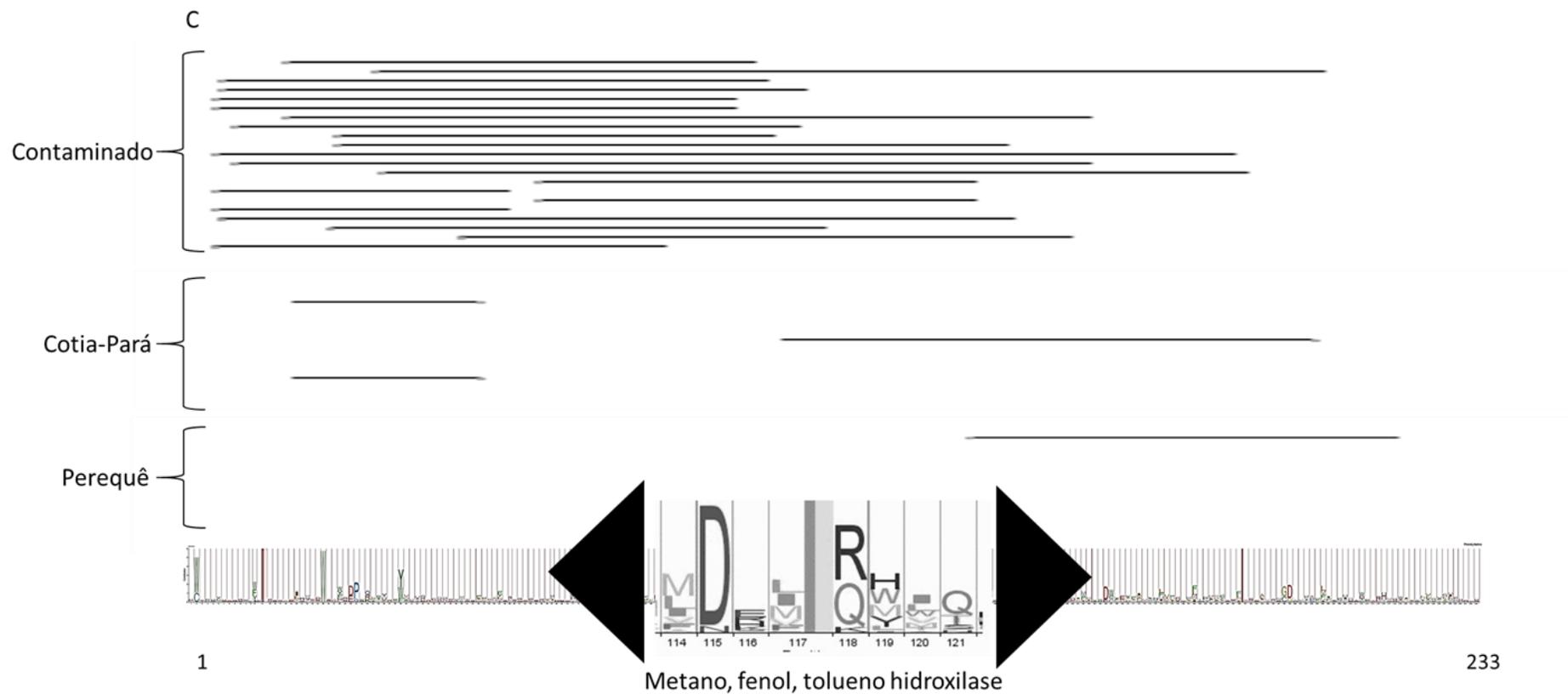
No total foram identificadas 375 sequências relacionadas à degradação de BTEX no local contaminado, 123 no Cotia-Pará e 85 no Perequê. Além de possuir a maior quantidade de sequências, o local contaminado foi o mais diverso em relação aos gêneros identificados. Dos 19 gêneros encontrados, 18 foram exclusivos do local contaminado e apenas o gênero *Klebsiella* não foi encontrado, sendo este exclusivo do local Cotia-Pará. Muitas sequências não puderam ser classificadas de forma mais específica e ficaram em nós mais basais como o caso dos nós Bacteria e Proteobacteria. Outro nó que apresentou muitas sequências e que se faz necessário citar é o nó *No hits* apresentando 287, 74 e 71 sequências respectivamente para os locais contaminado, Cotia-Pará e Perequê. Este nó representa sequências que não apresentaram similaridade, de acordo com nossos limiares adotados, com nenhuma sequência no banco de dados.

Para verificação da distribuição das sequências em relação aos HMMs construídos, as sequências que apresentaram similaridade com os HMMs hidroxilação de anel subunidade alfa e beta e metano, fenol, tolueno hidroxilase foram alinhadas aos respectivos HMMs (Figura 22). Foram representados apenas os domínios presentes nos três locais estudados.

Figura 22 – Alinhamento das sequências similares aos respectivos HMMs







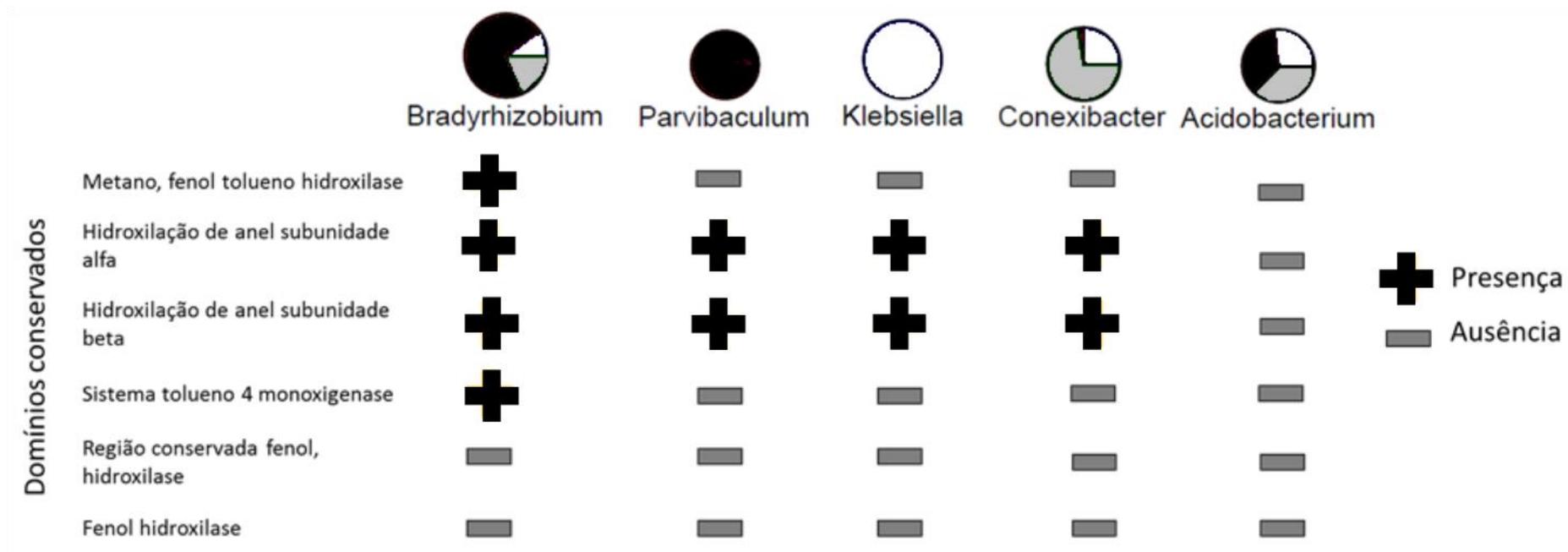
Distribuição das sequências com similaridade para os domínios conservados. A barra maior representa o logo do domínio com seu tamanho em pares de base. Cada um dos traços menores representa uma sequência alinhada respeitando o começo e o final de seu alinhamento em comparação ao domínio conservado. Ao lado estão os nomes dos locais onde as sequências foram encontradas. **A**:sequências alinhadas para o domínio hidroxilação de anel subunidade alfa; **B**: sequências alinhadas para o domínio hidroxilação de anel subunidade beta; **C**: Sequências alinhadas para o domínio metano, fenol, tolueno hidroxilase Distribuição das sequências com similaridade para os domínios conservados. A barra maior representa graficamente o domínio com seu tamanho em pares de base. A região central representa um aumento no tamanho na região para melhor visualização do domínio. Cada um dos traços menores representa uma sequência alinhada respeitando o começo e o final de seu alinhamento em comparação ao domínio conservado. Ao lado estão os nomes dos locais onde as sequências foram encontradas

4.3.4 *Degradação de contaminantes por Bradyrhizobium japonicum*

O estudo de domínios conservados e a comparação com os metagenomas revelaram a presença e quantidade de sequências relacionadas aos domínios de degradação. Essas sequências, através de comparações nos bancos de dados do NCBI, puderam ser identificadas, e os possíveis gêneros aos quais as sequências pertencem foram elucidados.

A identificação dos domínios conservados nos levou a mais uma comparação. Foi verificada a presença dos domínios conservados nos micro-organismos mais abundantes dos locais estudados. Essa verificação foi feita buscando uma possível capacidade de degradação de anéis aromáticos pelos organismos mais abundantes (Figura 23).

Figura 23 – Identificação dos domínios conservados nos genomas dos organismos mais abundantes

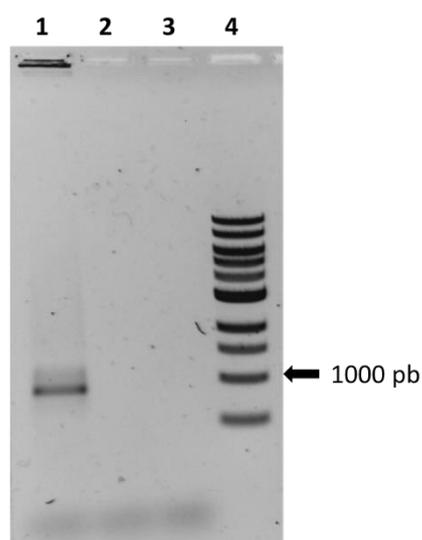


O gráfico mostra os gêneros mais abundantes encontrados em cada um dos metagenomas e a quantidade de sequências encontradas para cada um dos gêneros selecionados. A tabela logo abaixo do gráfico indica a presença e ausência dos domínios conservados para cada um dos gêneros.

A figura revela a presença dos domínios conservados para degradação de BTEX em alguns gêneros mais abundantes dos locais estudados. Focando-se o local contaminado, os dois gêneros mais representados apresentaram domínios de degradação para BTEX. O gênero *Parvibaculum* apresentou apenas os domínios de hidroxilação de anel subunidade alfa e beta, já o gênero *Bradyrhizobium*, que foi o mais representado em número de sequências tanto do local contaminado quanto dos outros metagenomas, apresentou quatro domínios conservados. Gêneros abundantes em outros metagenomas como *Klebsiella* e *Conexibacter* também apresentaram domínios de degradação para anéis aromáticos, porém uma diversidade menor do que o gênero *Bradyrhizobium*.

A presença desses domínios conservados identificados em genomas do gênero *Bradyrhizobium* levou ao desenvolvimento do experimento de degradação de anéis aromáticos. A espécie *Bradyrhizobium japonicum* SEMIA 566 foi primeiramente testada quanto à presença do domínio conservado de degradação tolueno-4-monoxigenase. Para isso foi realizada uma amplificação deste domínio no genoma da cepa escolhida e pela visualização da amplificação em gel de agarose pudemos identificar uma banda (Figura 24).

Figura 24 – Amplificação do domínio tolueno-4-monoxigenase em *Bradyrhizobium japonicum* SEMIA 566.

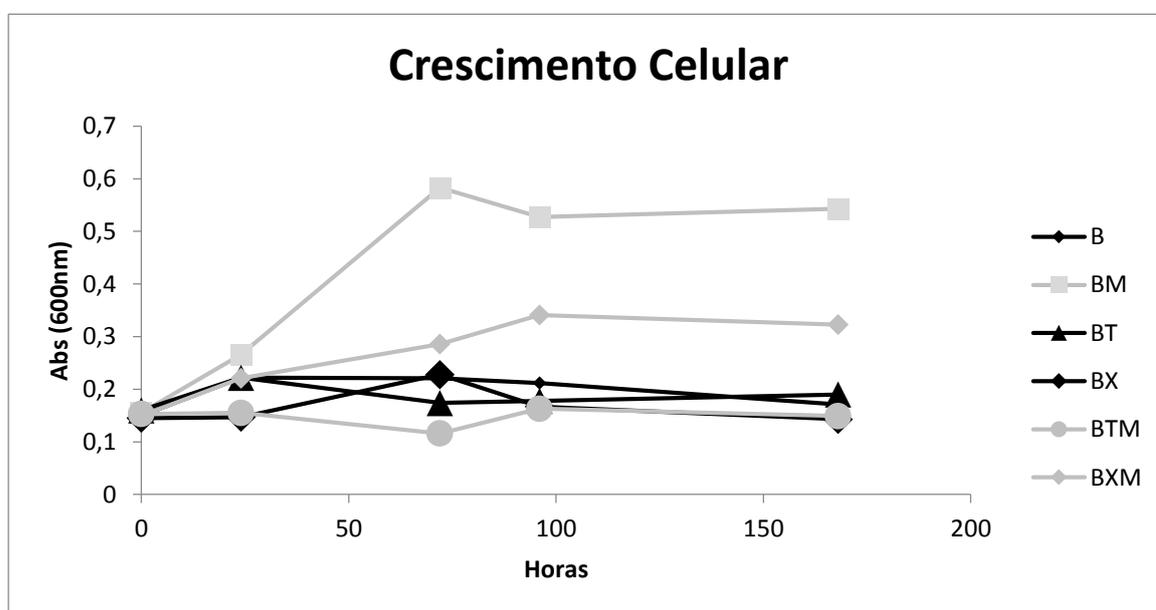


Visualização em gel de agarose da amplificação do domínio conservado tolueno-4-monoxigenase em *Bradyrhizobium japonicum* SEMIA 566. **1** – *Bradyrhizobium japonicum* SEMIA 566; **2** – *Escherichia coli* DH10b; **3** – controle negativo de amplificação; **4** – marcador molecular com marcação da altura de 1000 pb

A banda identificada no gel de agarose foi a banda esperada para a amplificação, a qual foi predita *in silico* contendo 800 pb. Somente o *Bradyrhizobium* apresentou amplificação, não sendo possível identificar nenhuma banda na amplificação realizada para *E. coli*.

Após a verificação da presença do domínio conservado, foi desenvolvido o experimento de crescimento e degradação de tolueno e xilenos por *Bradyrhizobium*. Tanto o crescimento em meio contaminado quanto a capacidade de degradação foram medidos. O crescimento na presença do contaminante está apresentado na Figura 25.

Figura 25 – Crescimento de *Bradyrhizobium japonicum* SEMIA 566 na presença de xilenos e tolueno



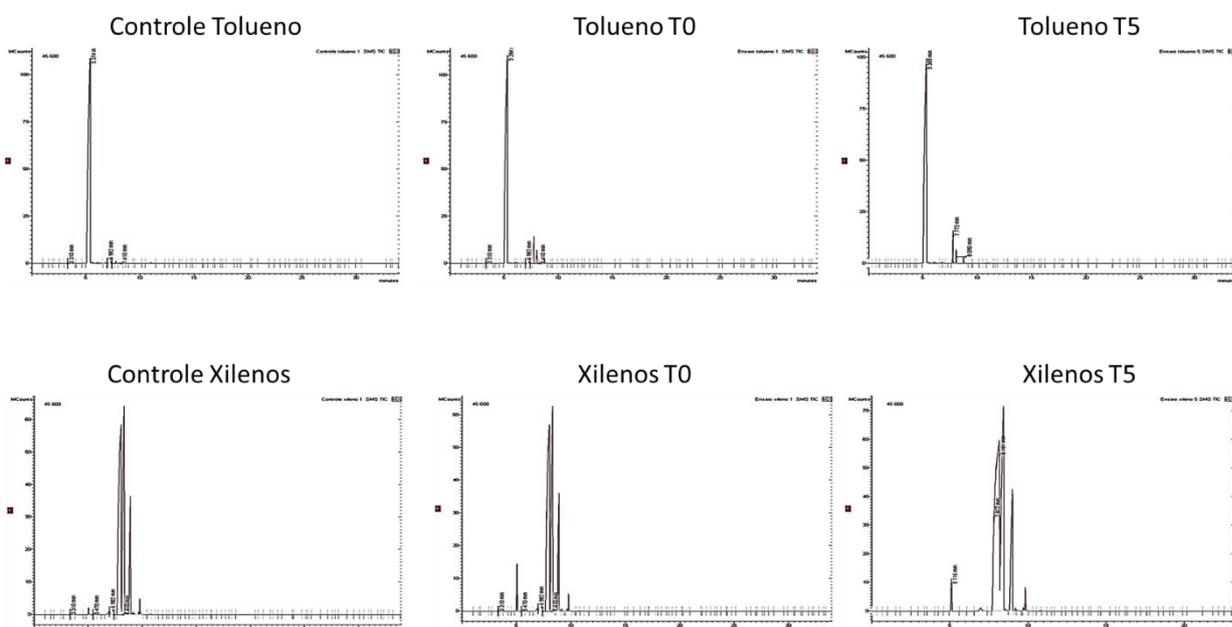
Resultado de crescimento celular de *Bradyrhizobium japonicum* SEMIA 566 na presença de tolueno e xilenos. **B** - controle negativo de crescimento; **BM** – Controle positivo de crescimento; **BT** – tolueno como única fonte de carbono; **BX** – xilenos como única fonte de carbono; **BTM** – tolueno e manitol como fontes de carbono; **BXM** – xilenos e manitol como fontes de carbono.

O experimento de crescimento em diferentes composições de meio de cultura revelou a capacidade de crescimento de *Bradyrhizobium* na presença de contaminantes. Pelo gráfico podemos notar o crescimento da cepa utilizada na presença de xilenos tendo manitol também como fonte de carbono. A curva de crescimento tendo xileno como única fonte de carbono evoluiu durante o tempo juntamente com a curva de controle negativo de crescimento (B), indicando que o micro-organismo não foi capaz de crescer nestas condições. Todas as curvas envolvendo tolueno (BT e BTM) não apresentaram aumento de DO em relação ao

tempo, ficando bastante próximas às medidas do controle negativo. Podemos concluir que, na presença de tolueno, não houve crescimento de *Bradyrhizobium* nem mesmo com a presença de manitol no meio, fonte de carbono utilizada pelo micro-organismo para produção de energia e crescimento como indicado pelas DO medidas para a curva BM.

Além do crescimento celular, a degradação do contaminante no meio de cultura também foi medida. A Figura 26 mostra os resultados obtidos.

Figura 26 – Teste da capacidade de degradação de xilenos e tolueno por *Bradyrhizobium japonicum* SEMIA 566



Verificação de degradação de tolueno e xilenos por *Bradyrhizobium japonicum* SEMIA 566 em meio de cultura M2 complementados por manitol. Os gráficos representam as medições feitas em cromatógrafo a gás acoplado à espectrometria de massas. As medições foram realizadas para cada uma das réplicas somando um total de 5 medições. Nos gráficos estão representados apenas a última medição dos controles (controle tolueno e controle xilenos) e a primeira e última medições das preparações contendo o micro-organismo com tolueno (Tolueno T0 e T5) e xilenos (Xilenos T0 e T5).

As medições realizadas mostraram as mesmas medidas tanto para os controles de degradação quanto para os ensaios realizados. Estas concentrações se repetiram para todos os ensaios feitos desde a primeira medição (T0) até a última (T5). Os diferentes picos encontrados para os gráficos de medição para xilenos representam as diferentes formas orto, meta e para encontradas.

5 DISCUSSÃO

5.1 Locais analisados e processamento do solo

Como já mencionado, a contaminação por BTEX é um problema conhecido que pode causar danos ambientais, assim como afetar diretamente a saúde de seres humanos. No município de Cubatão, um dos locais mais industrializados da América Latina, com a presença de muitas indústrias que processam o petróleo e que utilizam seus subprodutos, o BTEX é uma presença constante. Como o presente projeto visa à aplicação de técnicas de sequenciamento de nova geração em estudos metagenômicos de áreas contaminadas com BTEX, a contaminação mostrou um fator-chave para a seleção de determinada área.

A área escolhida apresentava histórico de contaminação por combustível. Um derramamento de gasolina de um tanque de armazenamento ocorreu aproximadamente dez meses antes da coleta e, para verificação da presença de contaminantes, o solo foi testado para presença de BTEX. As análises de hidrocarbonetos aromáticos da área escolhida revelaram a presença de tolueno e xilenos no solo coletado. Apesar de o teste realizado ser direcionado para os quatro componentes do BTEX, somente tolueno e xilenos foram encontrados de acordo com as análises. Este fato pode ter acontecido por um viés intrínseco da técnica de detecção utilizada, pois existe uma competição de hidrocarbonetos pela fibra de detecção que pode ter atrapalhado no processo de identificação. Outra explicação possível é sustentada pelas frações de benzeno e etilbenzeno na gasolina, menores do que tolueno e xilenos (CHRISTENSEN; ELTON, 1996), e que por isso podem ter sido eliminados do local mais rapidamente.

Os locais Cotia-Pará e Perequê foram escolhidos como controle. Esses locais não apresentavam histórico de contaminação por hidrocarbonetos aromáticos e por estarem próximos à área contaminada, possíveis diferenças em relação à temperatura, à umidade e a outras características ambientais que poderiam influenciar na composição microbiológica do solo seriam minimizadas. Como controle de contaminação, foi escolhido o local Perequê por se localizar ao sopé da Serra do Mar e por estar menos propício a ações antrópicas diárias. Sua análise com relação à presença de BTEX foi negativa, confirmando nossas expectativas. Apesar de não ser possível a identificação de BTEX no local Perequê, este local e o Cotia-Pará estão

localizados nos arredores de áreas com atividade industrial. Por aproximadamente trinta anos eles vêm sofrendo pressões ambientais causadas pela poluição presente no município. Nossa escolha como solos de controle também levou em consideração esse fato, pois um solo de uma região distante não refletiria essa situação. Um solo distante poderia apresentar uma microbiota completamente diferente impossibilitando qualquer comparação entre os metagenomas.

Nossa amostragem de solo se deu pela coleta em um ponto em cada uma das três regiões escolhidas. De cada ponto foi retirada uma amostra para extração de DNA. No caso das amostras dos locais Perequê e contaminado, uma parte foi diretamente acondicionada em *vial* para análise de contaminações por BTEX e outra parte foi acondicionada em frascos plásticos para análises físico-químicas.

As técnicas de coleta variam nos diferentes trabalhos, por exemplo, existem casos em que a coleta é feita em diversos pontos de um mesmo local e casos em que é coletado uma maior quantidade de material de um único ponto para armazenamento (MCGRATH et al., 2008; URICH et al., 2008). Nossa abordagem de estudo metagenômico envolveu uma quantidade pequena de solo de um único ponto de cada local (1 g de solo por local analisado). As amostras foram processadas diretamente. Mesmo que fossem coletadas amostras de diversos pontos e posteriormente homogeneizadas, a pequena quantidade de solo coletado ainda não permitiria que não houvesse uma superestimativa dos micro-organismos do local. Como o solo utilizado se apresentava em fragmentos que muitas vezes ultrapassavam a massa necessária, partes desses fragmentos foram utilizados. Um processamento pós-coleta, como peneiramento desse solo poderia ser realizado, mas isso requereria uma maior manipulação do solo e modificação de suas características físicas e quiçá biológicas, modificando todas as suas características. Essa manipulação em nosso protocolo de amostragem não era desejável e por isso o solo foi congelado imediatamente em nitrogênio líquido, o que permite uma preservação mais eficiente das características dos locais amostrados.

Outro ponto que foi levado em consideração com respeito à coleta foi o tamanho do local de amostragem. Excluindo-se o local contaminado que era uma área bem delimitada, os outros pontos de coleta faziam parte de parques ecológicos com muitos quilômetros de extensão, e assim, a escolha do ponto de coleta foi importante para se evitar uma amostra tendenciosa. Solos são os ambientes muito ricos com relação a micro-habitats, podendo apresentar milhares de espécies e uma abundância celular

considerável. Diferenças de poucos centímetros podem significar grandes mudanças físico-químicas e biológicas como presença de diferentes elementos químicos, raízes, fungos e modificações realizadas por organismos macroscópicos; e podem abrigar comunidades microbiológicas totalmente diferentes (PAUL, 2007). Essa quantidade de seres microscópicos diferentes e distribuição não equivalente impossibilita a representação de forma totalmente fiel à abundância e diversidade microbiana de um local em um estudo metagenômico. Apesar dessas adversidades, nossa abordagem metagenômica e o método de coleta aplicado foram capazes de captar diferenças entre os locais com relação aos grupos procarióticos e com relação aos metabolismos estudados. Foi gerada uma grande quantidade de dados da qual se pôde inferir uma estrutura procariótica para cada um dos metagenomas e apontar diferenças que podem estar relacionadas com as características particulares de cada local.

As amostras foram retiradas de vinte centímetros de profundidade aproximadamente, região do solo que possui um aporte de oxigênio e que possui ainda uma camada que evitaria, no caso de presença de contaminantes voláteis, a rápida passagem destes para o ar. No local contaminado, os micro-organismos dessa região do solo teriam contato com o contaminante e teriam disponibilidade de oxigênio, duas variáveis que levariam a uma degradação aeróbia, conhecidamente mais rápida, eficaz e com maior valor de biorremediação do que uma degradação anaeróbia (HENDRICKX et al., 2006). Por agregar um conjunto de fatores único do ponto de vista de localização, impacto, permanência dos contaminantes, modo de contaminação e tempo de ocorrência, o estudo metabólico voltado à degradação aeróbia dos contaminantes poderia trazer novas informações, que poderiam servir até para uma atualização das técnicas de biorremediação atuais.

A extração de material genético do solo se deu através do processamento de 1 g de solo para cada um dos locais estudados. Essa quantidade apresentou um rendimento de aproximadamente oito microgramas de DNA para os locais Perequê e Cotia-Pará e de quatro microgramas para o local contaminado. O uso de kit de extração padroniza a técnica de obtenção do material genético, assim, a menor quantidade obtida de DNA para o local contaminado já pode ser considerada um reflexo direto da contaminação pela toxicidade das substâncias, que pode ter ocasionado uma queda na abundância de micro-organismos do solo. Também pode ser uma influência indireta ocasionada pelo nível de interferência antrópica e falta de cobertura vegetal, modificação das características do solo em relação aos outros locais estudados.

A extração de material genético de solos não é trivial. Diversas técnicas de extração foram desenvolvidas e problemas como degradação, contaminação e inibição de amplificação por PCR são bastante relatados (DONG et al., 2006; LAKAY; BOTHA; PRIOR, 2007; MILLER, 2001). Para verificar a qualidade da extração e capacidade de manipulação do DNA extraído por meio de técnicas de biologia molecular, submetemos a extração do material genético a ampliações e sequenciamento em um estudo piloto. Este estudo, além de comprovar a qualidade do material genético, foi a primeira abordagem de identificação taxonômica de micro-organismos nos locais selecionados. O DNA extraído, apresentou alto peso molecular e baixa degradação, como evidenciado pela corrida em gel de agarose. Também mostrou-se apto a técnicas moleculares, podendo seguir para o desenvolvimento de bibliotecas para pirosequenciamento.

5.2 Identificação taxonômica por diferentes métodos

A identificação taxonômica dos micro-organismos foi realizada em duas etapas. A primeira abordagem, envolveu um estudo piloto de amplificação dos genes SSU rRNAs. As sequências geradas puderam ser identificadas e definiram de modo geral a composição da comunidade procariótica de cada local analisado, mostrando que havia diferenças entre elas. A outra abordagem envolveu o sequenciamento direto do DNA extraído, que por não necessitar de etapas de amplificação não sofreu seleção pelo uso de iniciadores. Isso gerou uma coleção de sequências provindas diretamente dos genomas presentes no solo contendo diversos tipos de genes, inclusive os genes SSU rRNAs.

O estudo piloto que utilizou o método Sanger para sequenciamento e o RDP como banco de dados de identificação de sequências mostrou como principal resultado a diferença de abundância entre os filos Acidobacteria e Proteobacteria nos locais estudados. Estes dois filos são os mais conhecidos e abundantes em solos e são encontrados recorrentemente em estudos publicados (BRUCE et al., 2010; JANSSEN, 2006; URICH et al., 2008). O local contaminado apresentou uma abundância maior do filo Proteobacteria enquanto os locais Cotia-Pará e Perequê apresentaram uma abundância maior do filo Acidobacteria. Relacionamos a dominância do filo Proteobacteria no local contaminado à contaminação ali encontrada e propusemos a hipótese de que esse filo poderia ter tido uma vantagem adaptativa com a modificação

do ambiente aumentando seu número de forma acentuada. Essa interpretação foi sustentada pelo fato de o filo Proteobacteria possuir gêneros bastante conhecidos por habitarem locais contaminados (ALFREIDER; VOGHT, 2007) e degradarem hidrocarbonetos aromáticos como os gêneros *Rhodococcus* (TANCSICS et al., 2008), *Pseudomonas* e *Burkholderia* (JINDROVA et al., 2002).

Com as análises por pirosequenciamento da segunda abordagem, a quantidade de dados cresceu acentuadamente. Por utilizar outro procedimento metodológico e possibilitar análises em diferentes bancos de dados, o método forneceu mais dados para complementar a identificação taxonômica do estudo piloto.

Comparando-se os filios identificados, tanto o método de pirosequenciamento quanto o método Sanger concordaram quanto aos mais abundantes, mas estes também variaram quanto à representatividade individual no metagenoma, dependendo da técnica de sequenciamento utilizada. O resultado mais marcante foi a dominância de sequências identificadas para o filo Acidobacteria nos locais não contaminados identificadas pelo sequenciamento do estudo piloto, que não se repetiu com a identificação pelo sequenciamento de alta performance.

Já é de conhecimento que a amplificação por reação de polimerização em cadeia pode introduzir erros que comprometeriam os resultados. Entre esses erros está a formação de quimeras (HAAS et al., 2011), que gera sequências falsas, vistas como sequências pertencentes a novas espécies no momento da classificação. Outro erro está relacionado à especificidade dos iniciadores que pode influenciar de modo seletivo um estudo de genes SSU rRNA. (KANAGAWA, 2003; POLZ; CAVANAUGH, 1998; SUZUKI; GIOVANNONI, 1996). Os iniciadores para a amplificação de genes SSU rRNAs são baseados em regiões conservadas de micro-organismos conhecidos. Mesmo o gene SSU rRNA sendo conservado, pode apresentar diferenças, principalmente tratando-se de amostras ambientais, nas quais a maioria dos micro-organismos são desconhecidos. Estes iniciadores, quando utilizados na amplificação de um metagenoma, irão apresentar afinidades maiores ou menores com os micro-organismos que constituem uma comunidade (MORALES; HOLBEN, 2009). A cada ciclo de amplificação, as sequências com maior afinidade serão mais facilmente amplificadas crescendo exponencialmente e se tornando mais representativas no metagenoma. Em uma clonagem para o método Sanger ou em um sequenciamento direto por pirosequenciamento, essas sequências serão identificadas como abundantes na comunidade, distorcendo a realidade.

Essa distorção acarretada pelo método de amplificação por PCR pode ter acontecido para o estudo piloto nos locais não contaminados, onde o filo Acidobacteria foi encontrado em número muito maior. A discordância entre os dois métodos com relação à abundância percentual dos filos mais representados também pode ser reflexo da introdução de erro pelo processo de PCR.

Os resultados do estudo piloto, em parte, concordaram com os resultados do pirosequenciamento. Apesar de não ser o método mais adequado para quantificação, o método de sequenciamento por Sanger é, em nossa opinião, adequado para identificação taxonômica qualitativa, uma vez que todos os filos mais representados pelo método de sequenciamento clássico foram confirmados pelo sequenciamento de alta performance.

5.3 Aprofundamento e comparação da classificação taxonômica

A identificação taxonômica dos procariotos foi realizada de maneiras diferentes visando-se uma identificação mais apurada, relacionando os organismos encontrados com o solo de onde provêm. Para se verificar diferenças entre grupos presentes nos solos, as sequências foram submetidas a três tipos de identificação taxonômica. Primeiro pela comparação ao banco de dados de nucleotídeos do NCBI, depois pela comparação de genes conservados e por último pela comparação com o banco de dados do RDP. As três abordagens foram realizadas pelo fato de que, apesar do gene SSU rRNA ainda ser o mais utilizado para identificação taxonômica, outros métodos de comparação vem se apresentando bastante confiáveis para a classificação. Eles podem ser considerados uma alternativa para identificações mais complexas, como as abordagens metagenômicas (LIU et al., 2011; VENTER et al., 2004).

Nosso objetivo era visualizar possíveis diferenças de classificação com relação aos métodos empregados e identificar o melhor método para as análises utilizando-se maiores estringências de seleção de sequências.

Apesar de nosso método não envolver as tendências que podem surgir com a técnica de PCR, as sequências com similaridade a sequências do banco de dados ribossomais são sequências parciais de um gene ribossomal inteiro. Esse gene é conhecido por suas regiões variáveis e não variáveis (PETROSINO et al., 2009), que auxiliam na identificação. Porém, quando uma sequência parcial possui identidade com sequências ribossomais do banco de dados e esta representa um fragmento de uma

região conservada, sua classificação muitas vezes é equivocada, alterando de forma tanto qualitativa quanto quantitativa a identificação de uma estrutura microbiana.

O método de comparação por genes conservados se mostra mais sólido na identificação microbiana, pois utiliza não um, mas trinta e um genes conservados entre todos os genomas bacterianos fechados do NCBI, mais sessenta genomas ainda não fechados (LIU et al., 2011). Estes genes são genes “housekeeping” conservados e com valor taxonômico. Uma maior quantidade de genes para classificação taxonômica gera uma maior acurácia e robustez na identificação de sequências quando comparada ao método de classificação por SSU rRNA. No nosso caso, esse método resultou em um número maior de sequências que puderam ser identificadas.

O maior número de sequências utilizadas para a classificação taxonômica se deu para o método de comparação com o banco de dados nr do NCBI. Pela natureza do método de pirosequenciamento, fragmentos dos genomas dos micro-organismos presentes no solo puderam ser anotados. Esse método permite rastrear, através da sequência, organismos no banco de dados que possuem sequências similares, e assim relacionar as sequências metagenômicas aos organismos resultando em uma identificação taxonômica.

Dentre os três métodos, a comparação por BLASTN e o método de comparação pelo programa MetaPhyler apresentaram resultados bem semelhantes. Tanto a identificação de filos quanto as porcentagens encontradas dos filos em seus metagenomas foram parecidas. O método de classificação pelo MG-RAST que utiliza o banco de dados de genes SSU rRNA foi o método destoante. A equivalência na classificação entre os métodos MetaPhyler e BLASTN, e pela maior quantidade de sequências que puderam ser identificadas, o método de comparação por BLASTN foi utilizado para construção da estrutura procariótica dos metagenomas.

5.4 Identificação de subgrupos e comparação metagenômica

Para a comparação entre metagenomas, foram estudados os subgrupos dos quatro filos mais representados.

Entre os filos encontrados, o filo *Deinococcus-Thermus* foi exclusivo do local contaminado. Apesar de não ser um filo expressivo em número de sequências, ele pode estar relacionado à contaminação no local, já que não foi encontrado em nenhum

outro local estudado e possui capacidade de degradação de anéis aromáticos (PEREZ-PANTOJA et al., 2010).

5.4.1 Filo *Actinobacteria*

Os subgrupos do filo *Actinobacteria* se mostraram relativamente mais representados nos locais Cotia-Pará e Perequê, com destaque para esse último local, onde se obteve a maior representatividade na maioria dos subgrupos identificados (Figura 13). O local contaminado não apresentou nem diversidade nem abundância desse filo, apresentando uma maior quantidade de sequências para a família *Mycobacteriaceae* apenas. Esta família possui um potencial metabólico para degradação de anéis aromáticos já conhecidos (PEREZ-PANTOJA et al., 2010). Membros desta família possuem catecol dioxigenases, enzimas-chave na degradação de substâncias aromáticas. Essa família ainda tem capacidade de degradação de substâncias poliaromáticas e foi encontrada em locais contaminados com BTEX (XIN; WU; LIN, 2013). Esta capacidade de degradação de anéis aromáticos pode ter proporcionado uma vantagem adaptativa para essa família quando o local sofreu a contaminação. As outras famílias podem não ter suportado a contaminação do local e sofrido forte queda em seus números, tanto que de algumas delas não foram encontrados representantes no local contaminado, apesar de serem encontradas nos locais não contaminados.

Apesar de estar em pequeno número, a família *Micrococcineae* também foi encontrada exclusivamente no local contaminado. Não existem relatos de que esta família tenha a capacidade de degradação de hidrocarbonetos aromáticos ou de que já tenha sido encontrada em locais contaminados.

5.4.2 Filos *Acidobacteria* e *Firmicutes*

O filo *Acidobacteria* é um filo abundante, encontrado em diversos ambientes (LEE; KA; CHO, 2008; KIELAK et al., 2009). Até 2003, apenas três espécies representavam o filo (QUAISER et al., 2003), atualmente outras espécies foram isoladas e identificadas (DEDYSH et al., 2012; KOCH et al., 2008). Contudo, este grupo ainda é uma incógnita, já que muitas sequências são agrupadas neste filo, porém, não podem ser submetidas a uma melhor classificação. Em nosso trabalho, o filo *Acidobacteria* se apresentou em segundo lugar em abundância com representantes

em todos os locais estudados. Todas as sequências identificadas deste filo foram atribuídas aos gêneros *Acidobacterium*, porém, a maioria das sequências do filo não foram identificadas. Isto ocorreu porque grande número dessas sequências não identificadas foram relacionadas ao gênero *Koribacter*, que ainda não possui uma posição taxonômica definida.

O filo Acidobacteria também é conhecido por habitar locais contaminados (SMITH; KIRISITS; REIBLE, 2012) e existem indícios de que apresentam dioxigenases, sendo capazes de degradar substâncias aromáticas (PEREZ-PANTOJA et al., 2010). A presença do gênero *Koribacter* em maior número no local contaminado pode indicar que ele possui uma vantagem adaptativa, apesar de não haver relatos deste gênero habitando solos contaminados com hidrocarbonetos aromáticos.

O filo Firmicutes apesar de ser mais abundante na amostra do local Perequê, apresentou maior diversidade na amostra do local contaminado, com três famílias exclusivas deste local. A presença dessas famílias exclusivas pode estar relacionada à contaminação, já que este filo é conhecido pela capacidade de degradação de aromáticos (LI et al., 2012). A família Clostridiaceae, a mais abundante no local contaminado, já foi identificada como um dos agentes de degradação aeróbia de BTEX (ALFREIDER; VOGHT, 2007) e também de poliaromáticos de forma anaeróbia (FOGHT, 2008). A ordem Bacillales, a mais abundante encontrada para o filo Firmicutes, é conhecida pela atividade de degradação de anéis aromáticos e é munida de catecol 2,3-dioxigenase (PEREZ-PANTOJA et al., 2010). Apesar desse filo estar em menor abundância no local contaminado do que no Perequê, a contaminação parece ter selecionado uma maior diversidade de espécies, principalmente as que conseguem utilizar o contaminante como fonte de carbono.

Pouca foi a expressão deste filo para o local Cotia-Pará, e as sequências que puderam ser atribuídas à ele, não puderam ser classificadas em famílias, pois não tiveram identidade suficiente.

5.4.3 Filo Proteobacteria

O filo Proteobacteria, conhecido por ser abundante e apresentar ampla distribuição, foi o filo mais abundante nos locais estudados, como mostrado na Figura 12 (BRUCE et al., 2010; NAVARRO-NOYA et al., 2010; ROESCH et al., 2007; URICH et al., 2008). Juntamente com o filo Acidobacteria, são os filios mais abundantes em

diferentes metagenomas, e por muitas vezes se alternam como os mais representados. Em locais contaminados com hidrocarbonetos aromáticos, a presença do filo Proteobacteria é bastante comum (ALFREIDER; VOGHT, 2007; GENOVESE et al., 2008). Isto se deve ao grande número de espécies que possuem metabolismo de degradação para estas substâncias. Apesar de terem sido representados em todos os locais estudados, diferenças com relação aos subgrupos e suas abundâncias relativas foram encontradas. Para entender melhor essas diferenças, uma identificação mais aprofundada para esse filo foi realizada, na qual se observou a quantidade de sequências e subgrupos identificados com relação aos locais estudados.

5.4.3.1 Classe Alphaproteobacteria e família Bradyrhizobiaceae

A classe Alphaproteobacteria foi a que se destacou no local Perequê (Figura 15) em relação à quantidade de sequências encontradas. Existem relatos desta classe ser encontrada em maior quantidade em solos de locais não contaminados e relativamente preservados (BRUCE et al., 2010; URICH et al., 2008) e, em nosso trabalho, esta observação foi confirmada. O local Cotia-Pará, que não estava contaminado com hidrocarbonetos, não apresentou diversidade e abundância desse grupo, porém, existem evidências que serão explicadas mais à frente que podem estar relacionadas à pobreza desta classe no local.

A família Phyllobacteriaceae, que foi quase que totalmente constituída pelo gênero *Parvibaculum*, se destacou no local contaminado e foi encontrada nos três locais analisados. A sua ocorrência em solos é comum, o que pode explicar a presença deste gênero em todos os locais estudados, porém a maior quantidade em solo contaminado pode ser explicada por uma vantagem adaptativa fornecida pela contaminação. Não há indícios de que o gênero *Parvibaculum* esteja relacionado à degradação de BTEX, porém, ele apresenta potencial de degradação de PCB (MACEDO; TIMMIS; ABRAHAM, 2007), um composto aromático. Já foi demonstrado que alguns micro-organismos, como os do gênero *Pseudomonas*, podem degradar diferentes composições estruturais aromáticas (FRANKLIN et al., 1981; FURUKAWA; MIYAZAKI, 1986). Esta plasticidade na degradação de aromáticos pode ter sido o motivo da maior quantidade de sequências do gênero *Parvibaculum* no local contaminado, insinuando uma atividade de degradação por esse grupo.

Outra família dentro da classe Alphaproteobacteria que se destacou foi a Bradyrhizobiaceae. Além de ser encontrada nos três locais estudados, sua abundância no local contaminado foi a maior entre todos os subgrupos estudados (Figura 15). Esta família mereceu atenção especial e foi melhor identificada, chegando até gêneros (Figura 16).

A classificação em níveis mais específicos revelou que a maioria das sequências desta família foi identificada dentro do gênero *Bradyrhizobium*. Este gênero é conhecido por habitar a região de rizosfera do solo e está vinculado a solos com cobertura vegetal. São de vida livre, porém podem se associar a plantas, principalmente a leguminosas, provocando o aparecimento de nódulos nas raízes (DENARIE; DEBELLE; ROSENBERG, 1992), onde podem fixar nitrogênio.

Apesar de não ser comum o vínculo de *Bradyrhizobium* a solos contaminados, bactérias desse gênero já foram encontradas habitando locais contaminados por hidrocarbonetos aromáticos (CAVALCA; DELL'AMICO; ANDREONI, 2004). Este gênero bacteriano nunca foi atribuído à degradação de hidrocarbonetos diretamente, porém já foi demonstrada sua capacidade de degradação do catecol (HUSSAIN; TEWFIK; HANDI, 1974), um intermediário da degradação do BTEX e de outros hidrocarbonetos aromáticos e também já foi confirmada a presença do gene que codifica para a benzeno/tolueno monoxigenase (PEREZ-PANTOJA et al., 2010), enzima relacionada à degradação de benzeno e tolueno. Cavalca e colaboradores (2004) ainda apresentaram a hipótese de que a transferência lateral de genes de degradação, possibilitou a presença dos micro-organismos desse gênero em locais contaminados. Foi realizado um teste de crescimento em meio contaminado com resultado positivo, porém não foi realizado um estudo de degradação mais aprofundado utilizando-se o gênero.

Posteriormente, em novo estudo de levantamento de micro-organismos em locais contaminados, foi relatada novamente a presença deste gênero bacteriano (SINGLETON et al., 2012). Este último relato foi bastante interessante pois mostra a evolução da estrutura de micro-organismos em uma coluna de degradação de hidrocarbonetos aromáticos. Juntamente com o aumento dos micro-organismos que são capazes de realizar a degradação, notou-se um aumento da quantidade de micro-organismos classificados como *Bradyrhizobium*. Este experimento mostra que o ambiente contaminado, de alguma forma foi propício ao desenvolvimento do grupo,

porém esse crescimento não foi atribuído à utilização do contaminante como fonte de energia nem foi insinuada uma possível degradação pelo gênero.

Em nosso trabalho, das sequências analisadas as que apresentaram maior identidade com o banco de dados foram as do gênero *Bradyrhizobium*, o que dá confiabilidade à classificação proposta. Isto significa que mesmo aumentando-se a estringência das classificações, este gênero continuou com uma abundância muito alta. Como tentativa de relacionar a presença do gênero *Bradyrhizobium* no local contaminado, foram buscados domínios de degradação de compostos aromáticos nos genomas fechados deste gênero depositados no NCBI. Nossa busca encontrou quatro dos seis domínios neste gênero (Figura 23). Os relatos da literatura e nossas observações de quantidade de sequências e domínios de degradação encontrados nos levaram a realizar testes de crescimento e degradação de substâncias aromáticas para o gênero *Bradyrhizobium*.

Dentro da família Bradyrhizobiaceae ainda, outro gênero de destaque foi o *Rhodopseudomonas*. Esse gênero possui potencial de degradação de anéis aromáticos, porém de forma anaeróbia (GERISCHER, 2002). A maior quantidade de sequências pertencentes a esse gênero pode indicar uma vantagem adaptativa ao local contaminado, e mesmo a poucos centímetros de profundidade podem existir condições para sustentar metabolismos anaeróbios.

5.4.3.2 Classe Gammaproteobacteria e família Enterobacteriaceae

A classe Gammaproteobacteria foi uma das mais representadas no geral (Figura 15). As famílias Methylococcaceae e Chromatiaceae foram encontradas apenas no local contaminado e já foi comprovada sua capacidade de degradação de aromáticos. A família Methylococcaceae possui micro-organismos munidos de monoxigenases não específicas, ou seja, podem utilizar uma diversidade de substratos para o metabolismo, entre os quais estão incluídos o BTEX (SHUKLA; UPADHYAY; DUBEY, 2012). Já a família Chromatiaceae não é conhecida por se mostrar presente em locais contaminados, porém possui um gene para degradação de homogentisato (PEREZ-PANTOJA et al., 2010), intermediário da degradação de estireno que também é uma substância aromática.

As famílias Pseudomonadaceae e Xanthomonadaceae foram encontradas no local contaminado e em Cotia-Pará. A família Pseudomonadaceae já é conhecida por

ser um grupo que promove biorremediação de locais contaminados com compostos orgânicos. Já a família Xanthomonadaceae não possui relatos de atividades de degradação de anéis aromáticos especificamente.

Além de ser conhecido por degradação de aromáticos, o gênero *Pseudomonas* da família Pseudomonadaceae também possui espécies conhecidas por serem patogênicas para seres humanos, como *Pseudomonas aeruginosa* (STOVER et al., 2000). A família Xanthomonadaceae também possui membros que podem ser patogênicos (MANFREDI et al., 1998).

Um dos grupos mais representados no nosso estudo foi a família Enterobacteriaceae, encontrada só na amostra do local Cotia-Pará. Esse grupo não só se mostrou exclusivo como foi o mais abundante deste local. Foram identificados para esta família os gêneros *Escherichia*, *Salmonella* e *Klebsiella* (Figura 16), todos conhecidos por causarem doenças em seres humanos (SACK, 1975; GLYNN et al., 1998; SHON; BAJWA; RUSSO, 2013). Estes organismos são conhecidos por serem oportunistas e por causarem infecções em pacientes com imunidade comprometida como pessoas com HIV, diabéticas ou em recuperação de cirurgia, causando infecções hospitalares. Apesar de estarem relacionados a ambientes hospitalares e resíduos em sistemas de coleta e armazenamento de esgoto, existem relatos de presença de micro-organismos da família Enterobacteriaceae em solos (BERG; EBERL; HARTMANN, 2005) e até mesmo associados a plantas (HOLDEN; PRITCHARD; TOTH, 2009). Esses resultados juntamente com o fato de o local de coleta, apesar de ser localizado dentro de um parque ecológico, estar a aproximadamente 300 metros de distância de uma ocupação imobiliária de casas simples e aparentemente sem estrutura de saneamento básico adequado (Figura 27), nos levaram à hipótese de contaminação do solo por resíduos residenciais.

espécies com resistências a antibióticos outrora utilizados e funcionais (PATTERSON et al., 2000) e até mesmo modificações fisiológicas (SHON; BAJWA; RUSSO, 2013). Se o meio ambiente pode proporcionar um compartilhamento de material genético entre diferentes micro-organismos, o local estudado pode ser considerado um “laboratório a céu aberto”. A diversidade e abundância de micro-organismos patogênicos, a presença de outros organismos, o fato do ambiente proporcionar uma troca de material genético e a proximidade com seres humanos formam o cenário perfeito para o aparecimento de cepas patogênicas resistentes. Isto pode levar ao surgimento e alastramento de infecções que podem se tornar extremamente perigosas.

5.4.3.3 Classes *Betaproteobacteria* e *Deltaproteobacteria*

De modo geral, essas duas classes foram bastante expressivas no local contaminado (Figura 15) e a ocorrência de muitos grupos exclusivos pode estar vinculada à contaminação do local.

Dentro da classe *Deltaproteobacteria*, as famílias *Desulfovibrionaceae* (ALFREIDER; VOGHT, 2007), *Geobacteraceae* (BOTTON et al., 2007), *Myxococcaceae* (ACOSTA-GONZALEZ; ROSSELLO-MORA; MARQUES, 2013) e *Syntrophobacteraceae* (ALFREIDER; VOGT, 2012) foram exclusivas ou bastante abundantes no local contaminado. Existem relatos do vínculo dessas famílias à degradação de substâncias aromáticas. Essas bactérias podem não degradar diretamente as substâncias aromáticas, como é o caso do gênero *Geobacter*, porém, possuem recursos genéticos que permitem utilizar produtos de outros micro-organismos que processam diretamente substâncias aromáticas. Elas aparentemente formam consórcios, o que promove e otimiza uma degradação mais completa.

Pertencente à família *Syntrophobacteraceae*, o gênero *Syntrophobacter* identificado em nosso estudo foi relacionado à degradação anaeróbia de substâncias aromáticas. Isso é mais uma evidência da presença de metabolismo de degradação anaeróbia no local contaminado apesar da coleta ter sido realizada a aproximadamente vinte centímetros de profundidade.

A família *Polyangiaceae* foi mais abundante dos locais Cotia-Pará e Perequê. A presença dessa família em diversos solos foi observada, porém a maior abundância se deu em locais florestados (SHANGE et al., 2012), o que foi apoiado pelas observações

feitas no nosso trabalho, uma vez que as áreas em que observamos maior abundância dessa família também são áreas com cobertura vegetal intensa.

A classe Betaproteobacteria, do mesmo modo que a Deltaproteobacteria, mostrou diversidade e abundância no local contaminado (Figura 15). Todas as famílias encontradas dentro da ordem Burkholderiales, da qual faz parte o gênero *Burkholderia*, já bastante conhecido por sua atividade comprovada de degradação de BTEX (SHIELDS et al., 1995), já foram encontradas em locais contaminados com BTEX (SUN; CUPPLES, 2012). Esse gênero não apenas se encontra em locais contaminados. Aparentemente ele está presente em diversos tipos de solos, incluindo cultiváveis (SALLES; VAN VEEN; VAN ELSAS, 2004) o que pode explicar a abundância encontrada no Perequê.

As famílias Hydrogenophilaceae (LARENTIS; HOERMANN; LUEDERS, 2012), Neisseriaceae e Rhodocyclaceae foram encontradas em locais contaminados (TANCSICS et al., 2012). Apesar de sua função de degradação não estar comprovada, elas estão associadas a comunidades microbianas que são encontradas em locais contaminados e que promovem degradação de substâncias aromáticas. Isto, novamente, pode indicar a presença dessas famílias em consórcios que promovem a degradação. Apesar de não serem vinculadas diretamente à degradação, essas famílias podem estar participando indiretamente, captando produtos de diferentes micro-organismos responsáveis pela manipulação direta do contaminante.

5.5 Identificação metabólica de degradação de benzeno e tolueno

Em nosso projeto, um dos objetivos era a identificação de genes das vias metabólicas de degradação de BTEX. Como nem todas as substâncias do grupo BTEX foram encontradas no ambiente contaminado, foram estudados os metabolismos de degradação de xilenos e tolueno, os dois contaminantes encontrados. O estudo foi baseado no metabolismo de degradação do KEGG e foram identificadas sequências com similaridade aos genes da via metabólica.

Segundo nosso estudo, todos os locais apresentaram genes das vias metabólicas (Figura 17), e, teoricamente, todos os locais poderiam ser capazes de lidar com os contaminantes estudados. Pelo mapa metabólico geral, grandes diferenças entre os locais estudados não puderam ser identificadas. O início das vias metabólicas pelos mapas do KEGG não se mostra muito definido e genes responsáveis pelo

processamento inicial dos aromáticos ainda não estão atribuídos. Este é um passo metabólico importante para a identificação da atividade de degradação, uma vez que no início das vias estão posicionadas as oxigenases, enzimas responsáveis pela modificação e abertura de anéis aromáticos (HENDRICKX et al., 2006). As observações feitas pelo mapa metabólico geral não foram muito conclusivas e, pelos resultados alcançados, uma comparação mais aprofundada foi realizada com o intuito de identificar possíveis diferenças de potencial de degradação entre os locais estudados.

O aprofundamento do estudo dos primeiros passos de degradação dos contaminantes se deu pela identificação de operons de degradação, recuperação das sequências nucleotídicas e comparação nos metagenomas. Por esta comparação pudemos diferenciar o potencial de degradação dos diferentes locais. Foi observada uma maior diversidade de operons de degradação para o local contaminado (Figura 18) e foram identificados diferentes caminhos metabólicos que os contaminantes podem seguir até sua degradação, o que está vinculado diretamente à diversidade de micro-organismos capazes de degradar anéis aromáticos. Isto pode indicar uma seleção de micro-organismos que aconteceu no local devido à contaminação e que possui a capacidade de degradação do contaminante.

Podemos identificar operons de degradação nos locais não contaminados também. Isto pode estar relacionado ao potencial metabólico de degradação de anéis aromáticos em locais não expostos a contaminantes, como já reportado em outros estudos (MACEDO; TIMMIS; ABRAHAM, 2007; XIE et al., 2011). Foi notada também uma diferença no número de sequências identificadas para os operons de degradação. O local Perequê apresentou apenas três sequências relacionadas aos operons, número bem baixo se comparado aos locais contaminado e Cotia-Pará, com 53 e 56 sequências respectivamente. O local Perequê não possui contato direto com contaminantes, não havendo uma seleção de micro-organismos como no local contaminado. Com a contaminação, pode ter ocorrido uma seleção dos micro-organismos capacitados à degradação, e essa vantagem adaptativa pode ter levado a um aumento populacional, resultando em um maior número de sequências relacionadas à degradação. No local Cotia-Pará, foi encontrado um número ainda maior de sequências relacionadas a operons de degradação, porém, com baixa diversidade. O operon que se apresentou mais abundante continha os genes *xyIA* e *xyIM*, genes de degradação de xilenos. Estes genes fazem parte de um operon

encontrado no plasmídeo pWWO, primeiramente identificado em *Pseudomonas putida* (FRANKLIN et al., 1981). Este plasmídeo pode ser passado para outros organismos por conjugação em solos se espalhando pelo local (RAMOS-GONZALEZ; DUQUE; RAMOS, 1991). Nesse local principalmente foi encontrada uma grande quantidade de micro-organismos dos gêneros *Klebsiella*, *Salmonella* e *Escherichia*, gêneros conhecidos por realizarem conjugação e carregarem plasmídios (AHMER; TRAN; HEFFRON, 1999; SANDEGREN et al., 2012). Juntando-se neste cenário o histórico do município de Cubatão com poluentes contendo anéis aromáticos, pode-se sugerir que os genes de degradação *xyIA* e *xyIM*, assim como seu operon estejam espalhados pelos procaríotos da região, presentes em diferentes espécies bacterianas que habitam o local e que possuem capacidade de degradação de hidrocarbonetos aromáticos. Outro fato que corrobora esta hipótese é a presença de grande número de sequências relacionadas ao plasmídeo pWWO no local contaminado também, onde cerca de 34% das sequências identificadas apresentaram identidade com esse plasmídeo. Neste local existe histórico de contaminação e foi encontrado vestígios de contaminação na ocasião da coleta. Apesar de não ter sido identificada a presença de bactérias dos gêneros encontrados no local Cotia-Pará, foi encontrado um número bastante expressivo de sequências provindas do gênero *Bradyrhizobium*. Este gênero, apesar de não estar relacionado à degradação diretamente, como já citado, já foi relatado integrando consórcios microbianos de degradação de substâncias aromáticas. Também já foi verificado que micro-organismos do gênero *Bradyrhizobium* são capazes de adquirir o plasmídeo pWWO, porém não se mostraram bons exemplos de aquisição da capacidade de degradação devido a modificações estruturais ocorridas ao plasmídeo dentro das bactérias (HOLLOWELLA et al., 1999). Apesar desta adversidade, o gênero *Bradyrhizobium* se mostrou apto à aquisição de plasmídeos do ambiente. Esta capacidade de aquisição de plasmídeos pode sugerir que em nosso estudo este gênero encontrado no local contaminado pode estar munido de um plasmídeo, aparentemente o plasmídeo pWWO e pode estar utilizando o operon de degradação para sobrevivência no local. Isto traria uma vantagem adaptativa, refletindo diretamente em uma maior abundância. Pelas nossas observações, não podemos afirmar que o micro-organismo em questão seja do gênero *Bradyrhizobium*, mas os altos níveis de estringência utilizados para as comparações sugerem uma grande proximidade ao gênero.

Para continuidade da identificação das capacidades de degradação dos solos estudados, foi realizado um experimento de domínios conservados entre os operons de degradação, e estes domínios foram identificados nos metagenomas.

Os micro-organismos que comprovadamente conseguem degradar BTEX foram identificados dentro do banco de dados do Bionemo, assim como seus operons, os quais foram utilizados para aprofundamento dos passos metabólicos iniciais das vias de degradação. Através de uma representação gráfica dos operons de degradação de BTEX (Figura 19), pudemos identificar domínios conservados comuns entre esses operons. Estes domínios, identificados a partir do banco de dados do PFAM, são regiões funcionais, conservadas, presentes em diferentes proteínas que exercem funções semelhantes. Uma busca desses domínios nos metagenomas nos forneceria um valor quantitativo para que o potencial de degradação de BTEX dos metagenomas fosse avaliado. Esta comparação ainda adicionaria um caráter inovador, já que o modo como foram comparados, por HMMs, é baseado em uma comparação probabilística. Esta comparação não utiliza palavras somente como uma comparação por BLAST, ela leva em consideração a probabilidade de se encontrar os aminoácidos em determinadas posições, o que possibilita a identificação de domínios funcionais em sequências desconhecidas dos bancos de dados.

O resultado de nossa análise identificou uma quantidade maior de sequências com similaridade aos domínios conservados no local contaminado, confirmando as observações feitas anteriormente. O avanço se deu na quantidade de sequências identificadas que foram identificadas revelando uma quantidade três vezes maior para o local contaminado do que para o local Cotia-Pará, o segundo em número absoluto de sequências. Os dois domínios mais representados foram as hidroxilases de anel subunidades alfa e beta que são conhecidas por degradar diversos tipos de substâncias apresentando anéis aromáticos (GIBSON; PARALES, 2000) e que poderiam também degradar xilenos e tolueno, os dois compostos aromáticos que foram encontrados no local. A maior abundância desses domínios indica uma comunidade procariótica adaptada à contaminação e possível capacidade de biorremediação.

Para embasar nossas análises, o método de identificação da capacidade de degradação de anéis aromáticos pelos domínios conservados foi aplicado a outros quatro metagenomas depositados no banco de dados do NCBI. Nossos resultados indicaram potencial de degradação em todos os metagenomas exceto o metagenoma VCU. Este metagenoma contém material genético de comunidade microbiana muito

específica que habita o sistema genital feminino e, por isso, não está habituada a contato com substâncias aromáticas. A ausência de domínios conservados verificado em nosso estudo para esse metagenoma mostrou que nosso método pode identificar domínios conservados que realmente atuam somente na degradação de anéis aromáticos, funcionando como um controle da sensibilidade do método.

O metagenoma gelo apresentou a menor abundância de domínios conservados de degradação. Este metagenoma é originário de comunidade microbiana de uma geleira da qual não existem relatos de qualquer contaminação no local. A identificação de domínios conservados neste metagenoma indicam um potencial de degradação intrínseco como já citado anteriormente. O potencial de degradação já foi reportado para solos congelados. Algumas espécies bacterianas resistentes e capazes de catabolizar anéis aromáticos conseguem sobreviver e degradar o contaminante quando ocorre uma contaminação (VAZQUEZ et al., 2009).

Para o metagenoma refinaria, a quantidade de sequências relacionadas à degradação de anéis aromáticos foi menor do que as encontradas para solo. Por se tratar de um ambiente conhecidamente contaminado, nossa expectativa seria uma maior abundância relativa dessas sequências para esse metagenoma. Porém, este metagenoma foi originado de efluentes contaminados de uma refinaria de petróleo e o ambiente aquático pode diminuir a disponibilidade de oxigênio inibindo a capacidade de degradação aeróbia de anéis aromáticos. Sem oxigênio suficiente, a quantidade de domínios conservados de degradação poderia ser reduzida pois poucos procaríotos com capacidade de degradação estariam ativos. Mesmo com uma abundância menor, a maior diversidade de domínios conservados foi identificada para este metagenoma. Grande quantidade de compostos fenólicos foi encontrada no ambiente do metagenoma estudado (SILVA et al., 2012). A diversidade de compostos encontrada está diretamente relacionada à diversidade de domínios de degradação encontrados, já que diferentes procaríotos poderiam estar degradando diferentes compostos.

Foram encontrados também domínios de degradação para os solos Cotia-Pará, Perequê e para o metagenoma Waseca apesar desses solos não estarem contaminados. Novamente foi encontrado um potencial de degradação intrínseco para esses solos através do estudo de domínios conservados de degradação de anéis aromáticos. Embora não saibamos classificar a quantidade encontrada como alta ou baixa por falta de comparação na literatura, sabemos que os locais estudados neste trabalho se localizam próximos a áreas contaminadas e focos de contaminação.

Indiretamente poderia ocorrer o contato de contaminantes com os solos estudados influenciando na composição de micro-organismos dos locais através de maior número de procaríotos com capacidade de catabolismo de anéis aromáticos. Os metagenomas dos locais Perequê e Cotia-Pará apresentaram abundâncias bastante próximas ao metagenoma Waseca. Este por se tratar de um solo destinado à agricultura, a probabilidade de contaminação com hidrocarbonetos aromáticos é baixa, e comparado aos solos Perequê e Cotia-Pará indicaria uma baixa capacidade de degradação intrínseca encontrada nesses últimos. Porém, os solos agráveis são por muitas vezes afetados com defensivos agrícolas que possuem em suas composições anéis aromáticos que podem ser biodegradados (MEGADI et al., 2010). Este contato faria com que a abundância de procaríotos capazes de catabolizar anéis aromáticos aumentasse influenciando na abundância de domínios conservados de degradação.

As sequências dos metagenomas que apresentaram similaridade com domínios conservados de degradação de hidrocarbonetos aromáticos foram comparadas ao banco de dados do NCBI. Esta comparação teve como objetivo a identificação taxonômica das sequências. Com a identificação, uma estrutura microbiana de degradação de anéis aromáticos poderia ser construída. Esta comparação indicaria os gêneros com potencial de degradação de hidrocarbonetos aromáticos do local.

Como demonstrado pela identificação de domínios conservados, o local contaminado se destacou em abundância e diversidade de gêneros microbianos relacionados à degradação de anéis aromáticos (Figura 21). De forma geral, o potencial de degradação dos locais Perequê e Cotia-Pará ficou em nós mais basais. Estes nós agruparam sequências que não apresentaram similaridade suficiente a nenhum dos genomas microbianos do banco de dados para que fossem alocadas de forma mais definida na árvore. Isto indica que o potencial de degradação identificado para estes dois locais provém de micro-organismos desconhecidos ou de sequências com similaridade a domínios conservados, que podem estar presentes em genomas de organismos conhecidos, porém com mutações que impede uma classificação mais precisa. O nó *No Hits* reuniu sequências que possuem domínios de degradação de anéis aromáticos, pois não possuíram similaridade a nenhum dos genomas depositados no banco. Esse grupo indica a presença de organismos completamente desconhecidos que habitam os locais estudados, e com potencial de degradação de anéis aromáticos.

Dentro dos gêneros identificados, muitos deles já são conhecidos por degradarem substâncias aromáticas, e nosso método foi sensível para identificá-los nos locais estudados. Outros gêneros, também encontrados em nosso estudo, já foram reportados habitando locais contaminados ou se desenvolvendo em comunidades enriquecidas que promovem a degradação (Tabela 10).

Tabela 10 – Identificação da atividade de degradação de gêneros encontrados no estudo metagenômico

Gênero	Atividade de degradação	Referência
<i>Gordonia</i>	Comprovada	(ARENSKOTTER; BROKER; STEINBUCHEL, 2004)
<i>Mycobacterium</i>	Comprovada	(XIN; WU; LIN, 2013)
<i>Rhodococcus</i>	Comprovada	(TANCSICS et al., 2008)
<i>Frankia</i>	Potencial	(YERGEAU et al., 2012)
<i>Geodermatophilus</i>	Potencial	(EL AZHARI; BRU; MARTIN-LAURENT, 2008)
<i>Roseiflexus</i>	_____	_____
<i>Koribacter</i>	_____	_____
<i>Bradyrhizobium</i>	Potencial	(SINGLETON et al., 2012)
<i>Rhodopseudomonas</i>	Comprovada	(GERISCHER, 2002)
<i>Methylobacterium</i>	Potencial	(KAO et al., 2008)
<i>Parvibaculum</i>	Potencial	(MACEDO; TIMMIS; ABRAHAM, 2007)
<i>Sphingomonas</i>	Comprovada	(ZYLSTRA; KIM, 1997)
<i>Bordetella</i>	Potencial	(ERIKSSON et al., 2003)
<i>Burkholderia</i>	Comprovada	(SHIELDS et al., 1995)
<i>Cupriavidus</i>	Comprovada	(FISCHER et al., 2008)
<i>Verminephrobacter</i>	_____	_____
<i>Thauera</i>	Comprovada	(SHINODA et al., 2004)
<i>Klebsiella</i>	Comprovada	(CHAMKHA et al., 2011)
<i>Acinetobacter</i>	Potencial	(MAZZEO et al., 2010)
<i>Pseudomonas</i>	Comprovada	(FRANKLIN et al., 1981)

A tabela mostra os gêneros encontrados nos metagenomas estudados que foram identificados como potenciais degradadores de substâncias aromáticas. A coluna **Atividade de degradação** indica se o gênero possui a atividade de degradação comprovada (identificado como **Comprovada**) ou se apenas foi encontrado em comunidades responsáveis por degradação (identificado como **Potencial**). A presença de um traço nesta coluna representa que o gênero em questão não possui informações relacionadas à degradação de substâncias aromáticas.

Três gêneros encontrados em nosso estudo não foram relacionados a nenhum tipo de degradação de substâncias aromáticas. Esses gêneros, *Roseiflexus*, *Koribacter* e *Verminephrobacter*, apresentaram domínios conservados de degradação e, pelas nossas observações, possuem potencial de degradação.

Para visualização da distribuição das sequências nos domínios conservados e verificação de tendências neste método desenvolvido, as sequências foram alinhadas aos seus respectivos domínios. A busca nesta análise seria por regiões dos domínios que acumulassem maior quantidade de sequências o que indicaria uma tendência tanto por origem de sequenciamento quanto pela representatividade no metagenoma. A análise mostrou uma distribuição aleatória das sequências relacionadas aos domínios conservados não sendo possível a percepção de regiões mais representadas. Este resultado indica que realmente existiu uma aleatoriedade na amostra, descartando duplicações causadas pelo sequenciamento. Isto aponta uma melhor representatividade do cenário real, o que confere maior credibilidade às análises.

As sequências, apesar de apresentarem similaridade aos domínios conservados, quando comparadas entre si, não apresentam homologia. Isto impossibilita um alinhamento múltiplo de sequências e maiores comparações. Do mesmo jeito que elas não apresentam homologia entre si, muitas delas não apresentam qualquer relação de identidade quando comparadas aos bancos de dados. Estas sequências representam 76%, 65% e 83% das sequências que apresentaram similaridade com os domínios conservados para os locais contaminado, Cotia-Pará e Perequê, respectivamente. Isto indica que as sequências podem ter origem em organismos completamente desconhecidos, mas por terem relações a domínios conservados apresentam enorme potencial de degradação de anéis aromáticos. Este método de análise desenvolvido aqui neste trabalho se mostrou uma poderosa ferramenta para verificação do potencial de degradação de anéis aromáticos *in situ*, podendo ser adaptado a outros metabolismos de interesse.

5.6 Degradação de substâncias aromáticas por *Bradyrhizobium*

Tentativas de isolamento de bactérias do local contaminado foram realizadas, porém sem sucesso, portanto, não pudemos isolar e realizar os estudos de degradação de anéis aromáticos com a cepa do gênero *Bradyrhizobium* que foi encontrada pelo sequenciamento no local contaminado. Por isso, os testes de crescimento e

degradação de xilenos e tolueno foram realizados com a cepa *Bradyrhizobium japonicum* SEMIA 566. Esta cepa foi largamente utilizada no Brasil para fins de inoculação em plantações de soja entre 1966 e 1978 (BODDEY; HUNGRIA, 1997). Pelo fato desta cepa não possuir o genoma depositado no banco de dados do GenBank, uma busca de genes conservados de degradação de anéis aromáticos através de comparações genômicas não foi possível. Comparações moleculares entre quatro genomas do gênero *Bradyrhizobium* foram realizadas *in silico* e mostraram que o domínio tolueno-4-monoxigenase, domínio que se apresentou mais abundante no local contaminado, estava presente de forma conservada nos quatro genomas estudados. Para verificar a presença deste domínio em nossa cepa de estudo, um par de iniciadores foi desenvolvido. A presença do domínio foi identificada, o que nos confirmou a semelhança genética com as cepas depositadas nos bancos de dados e nos respaldou no desenvolvimento do experimento de degradação de xilenos e tolueno.

Os testes de degradação revelaram uma tolerância da cepa SEMIA 566 a xilenos. Pelo gráfico de crescimento (Figura 25), podemos ver um aumento da DO com relação ao tempo no meio de cultura complementada com manitol, significando um aumento no número de células deste experimento. Apesar do aumento do número de células, o micro-organismo não utilizou o contaminante como fonte de carbono (Figura 26), o que indica somente o consumo de manitol para o crescimento, não mostrando nenhum indício da capacidade metabólica em utilizar xilenos como fonte de energia. Essa afirmação pode ser verificada pelo comportamento do crescimento da cepa em meio com xilenos e não complementado com manitol em que a curva ficou bastante próxima ao controle negativo de crescimento.

Para o experimento envolvendo tolueno, não foi notado nenhum crescimento tanto em meio complementado quanto em não complementado com manitol. Pelos nossos resultados, a cepa é mais sensível a tolueno, não conseguindo crescer na presença dessa substância aromática nessas concentrações.

As concentrações utilizadas para nossos experimentos foram bastante altas e não são encontradas facilmente em ambientes contaminados. A concentração calculada de 1000 ppm utilizada também está bastante acima da solubilidade em água para xilenos e tolueno, que é de aproximadamente 500 ppm e 200 ppm respectivamente. Estas concentrações foram utilizadas pelo fato da alta volatilidade dos contaminantes e pela facilidade de escape em sistemas fechados. Nosso objetivo

foi garantir sempre a máxima concentração de contaminante no meio que em caso de degradação pudesse ser facilmente capturado e, se existisse algum escape, que a concentração dentro do sistema ainda fosse grande o bastante para garantir o contato do micro-organismo com o contaminante. Por um lado, esta alta concentração reitera a capacidade de sobrevivência de *Bradyrhizobium* na presença de xilenos, explicando em parte a grande presença deste micro-organismo no local contaminado. Por outro, pode ter suprimido a capacidade de sobrevivência do micro-organismo a concentrações menores de tolueno e mesmo utilizar estes contaminantes como fonte de carbono, já que o domínio conservado encontrado em nossa cepa foi um domínio para degradação de tolueno.

O domínio tolueno-4-monoxigenase, de acordo com o levantamento feito (Figura 18), faz parte de operons de degradação de diferentes substâncias aromáticas, e a presença deste domínio no genoma de *Bradyrhizobium* pode estar ligado a um operon. A presença de um número mínimo de genes contendo diferentes domínios conservados é necessária para que o organismo tenha a capacidade de degradação de tolueno (YEN et al., 1991). Nossa cepa escolhida, apesar de apresentar o domínio tolueno-4-monoxigenase, pode não apresentar o operon completo ou pelo menos funcional, o que explicaria a falta de capacidade de uso de tolueno como fonte de carbono. Caso este operon estivesse completo e funcional, pelas concentrações utilizadas em nossos experimentos, a atividade de degradação pode ter sido suprimida pela toxicidade de tolueno resultando na morte das células.

Não há relato na literatura sobre o crescimento de *Bradyrhizobium* na presença de xilenos e tolueno nem sobre o uso dessas substâncias como fontes de carbono. O que já foi relatado é a tolerância desse gênero a locais contaminados com uma possível função no consórcio de degradação, como já mencionado anteriormente neste trabalho. Esta tolerância foi identificada pelos nossos experimentos para xilenos, que também demonstrou a falta de capacidade catabólica desse gênero no processamento direto dos contaminantes. Os resultados indicam que a presença e o sucesso deste gênero em locais contaminados com BTEX podem não estar relacionados ao processamento direto dos contaminantes, mas sim a outros fatores e funções que este gênero possa ter no consórcio. Para comprovar nossos resultados, novos experimentos de degradação devem ser feitos utilizando-se variações de concentração de contaminantes para se verificar o nível crítico de toxicidade e deste modo utilizar

concentrações menores a fim de se verificar mais precisamente a capacidade de degradação de tolueno e xilenos.

5.7 Considerações finais da discussão

Todos os resultados usados para a identificação taxonômica e metabólica foram baseados em sequenciamento de DNA diretamente do solo. Esta técnica de análise metagenômica não apresenta erros gerados por PCR nem depende de iniciadores, que podem gerar uma pré-seleção na identificação da diversidade procariótica, causada pela afinidade dos iniciadores. Os resultados apresentados refletem de forma mais fiel a quantidade de micro-organismos e genes identificados e se mostram uma forma mais apropriada para o estudo de abundância de comunidades microbianas. Nossas comparações foram realizadas tendo como base os bancos de dados disponíveis, como NCBI, KEGG, Uniprot e Bionemo. Estes bancos de dados estão baseados em micro-organismos conhecidos, e, por isso, muitas sequências metagenômicas não puderam ser identificadas. Isto evidencia a grande biodiversidade de micro-organismos de solo, inclusive no local contaminado, porém nos estudos metagenômicos, todas as abordagens estão baseadas em bancos de dados, o que limita a identificação de novas espécies, e apenas com o aumento das informações dos bancos de dados uma melhor interpretação pode ser feita. Mesmo com essas limitações, as comparações puderam ser realizadas e mostraram que o método de sequenciamento direto de DNA é uma técnica adequada para comparações de comunidades microbianas ambientais.

No local contaminado foram encontradas 375 sequências relacionadas à degradação. Sabemos pelos nossos resultados que nem todos os micro-organismos encontrados no local contaminado possuem a capacidade de degradar os contaminantes, porém para uma comunidade microbiana adaptada, este número encontrado mereceu um pouco de reflexão.

O gene SSU rRNA está presente em todos os micro-organismos e a probabilidade de se encontrar sequências desse gene é bastante alta em uma análise metagenômica. Em nosso estudo encontramos para o local contaminado 274 sequências ao todo, abaixo da quantidade de sequências identificadas relacionadas à degradação de aromáticos. Assim, o número identificado destas sequências é bastante alto e indica grande abundância de micro-organismos capazes de catabolizarem os hidrocarbonetos aromáticos. Apesar da maioria das sequências ficarem sem

identificação, a abordagem realizada pelo estudo de domínios conservados se mostrou bastante propícia para a identificação do potencial de degradação da comunidade procariótica. Na literatura não foi encontrada nenhuma relação envolvendo número de sequências e atividade específica, e esta relação está sendo estabelecida pela primeira vez neste trabalho.

A abordagem de identificação de domínios conservados pôde extrapolar a tradicional comparação de sequências. Assim sendo, apresentou maior profundidade na identificação do potencial de degradação e revelou melhor a característica da comunidade microbiana. Este método se revela mais adequado para esse fim comparado ao método de amplificação de genes de degradação. Pelo fato de não precisar de amplificações por PCR ele reflete melhor a realidade quando uma quantificação é desejada. Outra vantagem é que, se uma busca por genes de degradação fosse realizada nos locais não contaminados, um potencial de degradação iria ser identificado com muito mais intensidade utilizando-se uma abordagem por PCR, o que não diferenciaria os locais, como foi visto pelo método de comparação de domínios conservados. Isto levaria a uma interpretação errada da realidade e não distinguiria diferentes locais apresentando diferentes potenciais. Este método poderia ser aplicado a outras funções cuja identificação e quantificação de determinado potencial fosse aspirada.

6 CONCLUSÕES

A abordagem metagenômica de sequenciamento direto de DNA ambiental dos solos selecionados se mostrou uma boa ferramenta para acesso às informações *in situ* pretendidas neste trabalho. Conseguimos identificar as comunidades procarióticas dos três locais analisados encontrando diferenças entre elas. Pudemos relacionar grupos de micro-organismos com as características dos locais. Exemplos disso foram a grande diversidade de micro-organismos com potencial de degradação de hidrocarbonetos aromáticos no local contaminado e a presença de representantes da família Enterobacteriaceae no local Cotia-Pará, com provável contaminação por rejeitos residenciais. Com estas comparações pudemos entender melhor a composição procariótica da região e identificar a diversidade e abundância de uma comunidade procariótica de solo após evento de contaminação. O uso da tecnologia de pirosequenciamento certamente nos forneceu uma quantidade de dados muito superior comparado ao método clássico de sequenciamento. Isto refletiu na profundidade das análises, possibilitando a partir de um único sequenciamento construir um panorama mais completo, unindo resultados de identificação taxonômica e metabólica.

As diferenças entre os locais analisados também foram observadas sob o ponto de vista metabólico. Apesar da identificação de genes e domínios de degradação de anéis aromáticos em todos os locais analisados, o método utilizado neste estudo foi refinado o bastante para diferenciar qualitativamente e quantitativamente os diferentes locais. Para o local contaminado, foi encontrada maior diversidade de genes e domínios conservados de degradação, o que reflete uma comunidade microbiana mais adaptada à contaminação por anéis aromáticos e com potencial de degradação, que pode estar desempenhando um papel de biorremediação intrínseca no local. Não apenas os genes foram identificados, mas também um perfil taxonômico, sugerindo um consórcio microbiano de degradação de hidrocarbonetos aromáticos para os locais estudados e principalmente para o local contaminado.

Este trabalho aprofunda um pouco mais o conhecimento de comunidades procarióticas ambientais, área de estudo que apresenta grande potencial biotecnológico e que carece de maior quantidade de estudos. Mais precisamente, aprofunda o conhecimento sobre procariotos de áreas contaminadas com BTEX. Apesar de já serem conhecidos exemplares capazes de degradar essas substâncias

aromáticas, pouco se conhece sobre a dinâmica de uma comunidade *in situ* de locais contaminados. Este conhecimento pode trazer adaptações ao modo de aplicação de biorremediações, assim como desenvolver novos métodos.

Um dos aspectos mais importantes deste trabalho foi o local estudado. Cubatão é um município bastante afetado pela poluição que influencia o meio ambiente e os seres vivos que neles vivem. Os trinta anos de pressão ambiental causada por contaminantes fazem da região única no mundo, principalmente sob o aspecto da biorremediação. Apesar de toda esta importância, poucos trabalhos na região foram desenvolvidos, e este foi pioneiro ao utilizar a metagenômica para identificar e comparar micro-organismos da região, focando identificar potenciais biorremediadores e sugerindo um panorama metabólico e estrutural das comunidades de solo ali presentes.

REFERÊNCIAS*

- ACOSTA-GONZALEZ, A.; ROSSELLO-MORA, R.; MARQUES, S. Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the Prestige oil spill. **Environ. Microbiol.**, v. 15, n. 1, p. 77-92, Jan 2013.
- AGTEREN, M. H. V.; KEUNING, S.; JANSSEN, D. B. **Handbook on biodegradation and biological treatment of hazardous organic compounds**. Netherlands: Kluwer Academic Publishers, 1998.
- AHMER, B. M.; TRAN, M.; HEFFRON, F. The virulence plasmid of Salmonella typhimurium is self-transmissible. **J. Bacteriol.**, v. 181, n. 4, p. 1364-1368, Feb 1999.
- ALFREIDER, A.; VOGHT, C. Bacterial diversity and aerobic biodegradation potential in a BTEX-contaminated aquifer. **Water Air Soil Pollut.**, v. 183, p. 415-426, 2007.
- ALFREIDER, A.; VOGT, C. Genetic evidence for bacterial chemolithoautotrophy based on the reductive tricarboxylic acid cycle in groundwater systems. **Microbes Environ.**, v. 27, n. 2, p. 209-214, 2012.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.**, v. 25, n. 17, p. 3389-3402, Sep 1997.
- AN, Y. J. Toxicity of benzene, toluene, ethylbenzene, and xylene (BTEX) mixtures to Sorghum bicolor and Cucumis sativus. **Bull. Environ. Contam. Toxicol.**, v. 72, n. 5, p. 1006-1011, May 2004.
- ANDREOZZI, R.; CESARO, R.; MAROTTA, R.; PIROZZI, F. Evaluation of biodegradation kinetic constants for aromatic compounds by means of aerobic batch experiments. **Chemosphere**, v. 62, n. 9, p. 1431-1436, Mar 2006.
- APWEILER, R.; BAIROCH, A.; WU, C. H.; BARKER, W. C.; BOECKMANN, B.; FERRO, S.; GASTEIGER, E.; HUANG, H.; LOPEZ, R.; MAGRANE, M.; MARTIN, M. J.; NATALE, D. A.; O'DONOVAN, C.; REDASCHI, N.; YEH, L. S. UniProt: the Universal Protein knowledgebase. **Nucleic Acids Res.**, v. 32, p. D115-119, Jan 2004.
- ARENSKOTTER, M.; BROKER, D.; STEINBUCHER, A. Biology of the metabolically diverse genus Gordonia. **Appl. Environ. Microbiol.**, v. 70, n. 6, p. 3195-3204, Jun 2004.
- AUGUSTO FILHO, O.; MACEDO, E. S.; CERRI, L. E. S.; OGURA, A. T. Carta Geotécnica da Serra do Mar nas Folhas de Santos e Riacho Grande. **Seminário de integração técnica sobre a poluição da Serra do Mar**, p. 69-75, 1988.
- BALDWIN, B. R.; NAKATSU, C. H.; NEBE, J.; WICKHAM, G. S.; PARKS, C.; NIES, L. Enumeration of aromatic oxygenase genes to evaluate biodegradation during multi-phase extraction at a gasoline-contaminated site. **J. Hazard. Mater.**, v. 163, n. 2-3, p. 524-530, Apr 30 2009.
- BERG, G.; EBERL, L.; HARTMANN, A. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. **Environ. Microbiol.**, v. 7, n. 11, p. 1673-1685, Nov 2005.
- BODDEY, H. L.; HUNGRIA, M. Phenotypic grouping of Brazilian Bradyrhizobium strains which nodulate soybean. **Biol. Fertil. Soils**, n. 25, p. 407-405, 1997.

*De acordo com:

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

BOTTON, S.; VAN HARMELEN, M.; BRASTER, M.; PARSONS, J. R.; ROLING, W. F. Dominance of Geobacteraceae in BTX-degrading enrichments from an iron-reducing aquifer. **FEMS Microbiol. Ecol.**, v. 62, n. 1, p. 118-130, Oct 2007.

BRUCE, T.; MARTINEZ, I. B.; MAIA NETO, O.; VICENTE, A. C.; KRUGER, R. H.; THOMPSON, F. L. Bacterial community diversity in the Brazilian Atlantic forest soils. **Microb. Ecol.**, v. 60, n. 4, p. 840-849, Nov 2010.

BYRNE, A. M.; KUKOR, J. J.; OLSEN, R. H. Sequence analysis of the gene cluster encoding toluene-3-monooxygenase from *Pseudomonas pickettii* PKO1. **Gene**, v. 154, n. 1, p. 65-70, Feb 27 1995.

CARBAJOSA, G.; TRIGO, A.; VALENCIA, A.; CASES, I. Bionemo: molecular information on biodegradation metabolism. **Nucleic Acids Res.**, v. 37, p. 598-602, Jan 2009.

CARBONNELLE, E.; GROHS, P.; JACQUIER, H.; DAY, N.; TENZA, S.; DEWAILLY, A.; VISSOUARN, O.; ROTTMAN, M.; HERRMANN, J. L.; PODGLAJEN, I.; RASKINE, L. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. **Journal of microbiological methods**, v. 89, n. 2, p. 133-136, May 2012.

CAVALCA, L.; DELL'AMICO, E.; ANDREONI, V. Intrinsic bioremediability of an aromatic hydrocarbon-polluted groundwater: diversity of bacterial population and toluene monooxygenase genes. **Appl. Microbiol. Biotechnol.**, v. 64, n. 4, p. 576-587, May 2004.

CHAMKHA, M.; TRABELSI, Y.; MNIF, S.; SAYADI, S. Isolation and characterization of *Klebsiella oxytoca* strain degrading crude oil from a Tunisian off-shore oil field. **J. Basic Microbiol.**, v. 51, n. 6, p. 580-589, Dec 2011.

CHRISTENSEN, J. S.; ELTON, J. Soil and Groundwater pollution from BTEX. 1996. Disponível em: <<http://www.webapps.cee.vt.edu/ewr/environmental/teach/gwprimer/btex/btex.html>>. Acesso em: 07 fev. 2013.

COLE, J. R.; CHAI, B.; FARRIS, R. J.; WANG, Q.; KULAM-SYED-MOHIDEEN, A. S.; MCGARRELL, D. M.; BANDELA, A. M.; CARDENAS, E.; GARRITY, G. M.; TIEDJE, J. M. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. **Nucleic Acids Res.**, v. 35, p. 169-172, Jan 2007.

COLE, J. R.; WANG, Q.; CARDENAS, E.; FISH, J.; CHAI, B.; FARRIS, R. J.; KULAM-SYED-MOHIDEEN, A. S.; MCGARRELL, D. M.; MARSH, T.; GARRITY, G. M.; TIEDJE, J. M. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. **Nucleic Acids Res.**, v. 37, p. 141-145, Jan 2009.

Comitê Nacional de Organização Rio + 20. Cubatão mostra como é possível aliar crescimento com sustentabilidade. 2012. Disponível em: <http://www.rio20.gov.br/sala_de_imprensa/noticias-nacionais1/cubatao-mostra-como-e-possivel-aliar-crescimento-com-sustentabilidade.html>. Acesso em: 14/01/2013.

Companhia de Tecnologia de Saneamento Ambiental. Sistema estuarino de Santos e São Vicente. AMBIENTE, S. D. E. D. M. Disponível em: <www.acpo.org.br>. PROCOP - Programa de controle de poluição 2001.

_____. **Áreas contaminadas e reabilitadas no estado de São Paulo**. CETESB. Disponível em: <www.cetesb.sp.gov.br>. 2011

COSTA, C.; DE PASQUALE, R.; SILVARI, V.; BARBARO, M.; CATANIA, S. In vitro evaluation of oxidative damage from organic solvent vapours on human skin. **Toxicology in Vitro**. p. 324-331 2006.

COURTOIS, S.; CAPPELLANO, C. M.; BALL, M.; FRANCOU, F. X.; NORMAND, P.; HELYNCK, G.; MARTINEZ, A.; KOLVEK, S. J.; HOPKE, J.; OSBURNE, M. S.; AUGUST, P. R.; NALIN, R.; GUERINEAU, M.; JEANNIN, P.; SIMONET, P.; PERNODET, J. L. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. **Applied and environmental microbiology**, v. 69, n. 1, p. 49-55, Jan 2003.

DANIEL, R. The metagenomics of soil. **Nature reviews. Microbiology**, v. 3, n. 6, p. 470-478, Jun 2005.

DEDYSH, S. N.; KULICHEVSKAYA, I. S.; SERKEBAEVA, Y. M.; MITYAEVA, M. A.; SOROKIN, V. V.; SUZINA, N. E.; RIJPSTRA, W. I.; DAMSTE, J. S. *Bryocella elongata* gen. nov., sp. nov., a member of subdivision 1 of the Acidobacteria isolated from a methanotrophic enrichment culture, and emended description of *Edaphobacter aggregans* Koch et al. 2008. **Int. J. Syst. Evol. Microbiol.**, v. 62, n. 3, p. 654-664, Mar 2012.

DELOGER, M.; EL KAROUI, M.; PETIT, M. A. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. **Journal of bacteriology**, v. 191, n. 1, p. 91-99, Jan 2009.

DENARIE, J.; DEBELLE, F.; ROSENBERG, C. Signaling and host range variation in nodulation. **Annu. Rev. Microbiol.**, v. 46, p. 497-531, 1992.

DO, J.; LIU, X.; HU, Z.; DENG, D. Anaerobic BTEX biodegradation linked to nitrate and sulfate reduction. **Journal of Hazardous Materials**. n. 151, p. 720-729, 2008.

DONG, D.; YAN, A.; LIU, H.; ZHANG, X.; XU, Y. Removal of humic substances from soil DNA using aluminium sulfate. **J. Microbiol. Methods**, v. 66, n. 2, p. 217-222, Aug 2006.

Energy Information Administration. Top World Oil Consumers, 2011. Disponível em: <<http://www.eia.gov/countries/index.cfm?topL=con>>, 2013. Acesso em: 13/03.

EL AZHARI, N.; BRU, D.; MARTIN-LAURENT, F. Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR. **European Journal of Soil Science**, v. 59, n. 4, p. 665-673, 2008.

ERIKSSON, M.; SODERSTEN, E.; YU, Z.; DALHAMMAR, G.; MOHN, W. W. Degradation of polycyclic aromatic hydrocarbons at low temperature under aerobic and nitrate-reducing conditions in enrichment cultures from northern soils. **Appl. Environ. Microbiol.**, v. 69, n. 1, p. 275-284, Jan 2003.

FERREIRA, C. C.; TORRES, F. R.; BORGES, W. R. **Cubatão Caminhos da História**. Cubatão: Edição do Autor, 2008.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic acids research**, v. 39, n. Web Server issue, p. 29-37, Jul 2011.

FINOTTI, A. R.; CAICEDO, N. O. L.; RODRIGUEZ, M. T. R. Contaminações Subterrâneas com Combustíveis Derivados de Petróleo: Toxicidade e a Legislação Brasileira. **Revista Brasileira de Recursos Hídricos**, v. 6, n. 2, p. 29-46, 2001.

FISCHER, A.; HERKLOTZ, I.; HERRMANN, S.; THULLNER, M.; WEELINK, S. A.; STAMS, A. J.; SCHLOMANN, M.; RICHNOW, H. H.; VOGT, C. Combined carbon and hydrogen isotope fractionation investigations for elucidating benzene biodegradation pathways. **Environ. Sci. Technol.**, v. 42, n. 12, p. 4356-4363, Jun 15 2008.

FLICEK, P.; AMODE, M. R.; BARRELL, D.; BEAL, K.; BRENT, S.; CARVALHO-SILVA, D.; CLAPHAM, P.; COATES, G.; FAIRLEY, S.; FITZGERALD, S.; GIL, L.; GORDON, L.; HENDRIX, M.; HOURLIER, T.; JOHNSON, N.; KAHARI, A. K.; KEEFE, D.; KEENAN, S.; KINSELLA, R.; KOMOROWSKA, M.; KOSCIELNY, G.; KULESHA, E.; LARSSON, P.; LONGDEN, I.; MCLAREN, W.; MUFFATO, M.; OVERDUIN, B.; PIGNATELLI, M.; PRITCHARD, B.; RIAT, H. S.; RITCHIE, G. R.; RUFFIER, M.; SCHUSTER, M.; SOBRAL, D.; TANG, Y. A.; TAYLOR, K.; TREVANION, S.; VANDROVCOVA, J.; WHITE, S.; WILSON, M.; WILDER, S. P.; AKEN, B. L.; BIRNEY, E.; CUNNINGHAM, F.; DUNHAM, I.; DURBIN, R.; FERNANDEZ-SUAREZ, X. M.; HARROW, J.; HERRERO, J.; HUBBARD, T. J.; PARKER, A.; PROCTOR, G.; SPUDICH, G.; VOGEL, J.; YATES, A.; ZADISSA, A.; SEARLE, S. M. Ensembl 2012. **Nucleic Acids Res.**, v. 40, p. D84-90, Jan 2012.

FOGHT, J. Anaerobic Biodegradation of Aromatic Hydrocarbons: Pathways and Prospects. **Journal of Molecular Microbiology and Biotechnology**: 93-120 p. 2008.

FOX, G. E.; WISOTZKEY, J. D.; JURTSCHUK, P., JR. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. **International journal of systematic bacteriology**, v. 42, n. 1, p. 166-170, Jan 1992.

FRANKLIN, F. C.; BAGDASARIAN, M.; BAGDASARIAN, M. M.; TIMMIS, K. N. Molecular and functional analysis of the TOL plasmid pWVO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. **Proc. Natl. Acad. Sci. U. S. A.**, v. 78, n. 12, p. 7458-7462, Dec 1981.

FURUKAWA, K.; MIYAZAKI, T. Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. **J. Bacteriol.**, v. 166, n. 2, p. 392-398, May 1986.

GEER, L. Y.; MARCHLER-BAUER, A.; GEER, R. C.; HAN, L.; HE, J.; HE, S.; LIU, C.; SHI, W.; BRYANT, S. H. The NCBI BioSystems database. **Nucleic Acids Res.**, v. 38, p. D492-496, Jan 2010.

GENOVESE, M.; DENARO, R.; CAPPELLO, S.; DI MARCO, G.; LA SPADA, G.; GIULIANO, L.; GENOVESE, L.; YAKIMOV, M. M. Bioremediation of benzene, toluene, ethylbenzene, xylenes-contaminated soil: a biopile pilot experiment. **J. Appl. Microbiol.**, v. 105, n. 5, p. 1694-1702, Nov 2008.

GÉRIN, M.; SIEMIATYCKI, J.; DÉSY, M.; KREWSKI, D. Associations between several sites of cancer and occupational exposure to benzene, toluene, xylene, and styrene: Results of a case-control study in Montreal. **American Journal of Industrial Medicine**: 144–156 p. 1998.

GERISCHER, U. Specific and global regulation of genes associated with the degradation of aromatic compounds in bacteria. **J. Mol. Microbiol. Biotechnol.**, v. 4, n. 2, p. 111-121, Mar 2002.

GEVERS, D.; COHAN, F. M.; LAWRENCE, J. G.; SPRATT, B. G.; COENYE, T.; FEIL, E. J.; STACKEBRANDT, E.; VAN DE PEER, Y.; VANDAMME, P.; THOMPSON, F. L.; SWINGS, J. Opinion: Re-evaluating prokaryotic species. **Nature reviews. Microbiology**, v. 3, n. 9, p. 733-739, Sep 2005.

GIBSON, D. T.; PARALES, R. E. Aromatic hydrocarbon dioxygenases in environmental biotechnology. **Curr. Opin. Biotechnol.**, v. 11, n. 3, p. 236-243, Jun 2000.

GILL, S. R.; POP, M.; DEBOY, R. T.; ECKBURG, P. B.; TURNBAUGH, P. J.; SAMUEL, B. S.; GORDON, J. I.; RELMAN, D. A.; FRASER-LIGGETT, C. M.; NELSON, K. E. Metagenomic analysis of the human distal gut microbiome. **Science**, v. 312, n. 5778, p. 1355-1359, Jun 2 2006.

GLYNN, M. K.; BOPP, C.; DEWITT, W.; DABNEY, P.; MOKHTAR, M.; ANGULO, F. J. Emergence of multidrug-resistant *Salmonella enterica* serotype typhimurium DT104 infections in the United States. **N. Engl. J. Med.**, v. 338, n. 19, p. 1333-1338, May 7 1998.

GOMEZ-ALVAREZ, V.; TEAL, T. K.; SCHMIDT, T. M. Systematic artifacts in metagenomes from complex microbial communities. **ISME J.**, v. 3, n. 11, p. 1314-1317, Nov 2009.

GORDON, R. E.; MIHM, J. M. The type species of the genus *Nocardia*. **Journal of general microbiology**, v. 27, p. 1-10, Jan 1962.

GORIS, J.; KONSTANTINIDIS, K. T.; KLAPPENBACH, J. A.; COENYE, T.; VANDAMME, P.; TIEDJE, J. M. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. **International journal of systematic and evolutionary microbiology**, v. 57, n. 1, p. 81-91, Jan 2007.

GRICE, E. A.; SEGRE, J. A. The skin microbiome. **Nat. Rev. Microbiol.**, v. 9, n. 4, p. 244-253, Apr 2011.

GUIMARÃES, J. R. P. F. **Resíduos industriais na Baixada Santista**. Disponível em: <http://www.acpo.org.br/biblioteca/08_residuos/residuos_bx_santista.pdf>. ACPO 2005.

GUPTA, R. S.; GRIFFITHS, E. Critical issues in bacterial phylogeny. **Theoretical population biology**, v. 61, n. 4, p. 423-434, Jun 2002.

HAAS, B. J.; GEVERS, D.; EARL, A. M.; FELDGARDEN, M.; WARD, D. V.; GIANNOUKOS, G.; CIULLA, D.; TABBAA, D.; HIGHLANDER, S. K.; SODERGREN, E.; METHE, B.; DESANTIS, T. Z.; PETROSINO, J. F.; KNIGHT, R.; BIRREN, B. W. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. **Genome Res.**, v. 21, n. 3, p. 494-504, Mar 2011.

HENDRICKX, B.; JUNCA, H.; VOSAHLOVA, J.; LINDNER, A.; RUEGG, I.; BUCHELI-WITSCHER, M.; FABER, F.; EGLI, T.; MAU, M.; SCHLOMANN, M.; BRENNEROVA, M.; BRENNER, V.; PIEPER, D. H.; TOP, E. M.; DEJONGHE, W.; BASTIAENS, L.; SPRINGAEL, D. Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. **J. Microbiol. Methods.**, v. 64, n. 2, p. 250-265, Feb 2006.

HERMAN, D. C.; INNISS, W. E.; MAYFIELD, C. I. Toxicity testing of aromatic hydrocarbons utilizing a measure of their impact on the membrane integrity of the green alga *Selenastrum capricornutum*. **Bull. Environ. Contam. Toxicol.**, v. 47, n. 6, p. 874-881, Dec 1991.

HERRMANN, S.; KLEINSTEUBER, S.; NEU, T. R.; RICHNOW, H. H.; VOGT, C. Enrichment of anaerobic benzene-degrading microorganisms by in situ microcosms. **FEMS Microbiol. Ecol.**, v. 63, n. 1, p. 94-106, Jan 2008.

HESS, M.; SCZYRBA, A.; EGAN, R.; KIM, T. W.; CHOKHAWALA, H.; SCHROTH, G.; LUO, S.; CLARK, D. S.; CHEN, F.; ZHANG, T.; MACKIE, R. I.; PENNACCHIO, L. A.; TRINGE, S. G.;

- VISEL, A.; WOYKE, T.; WANG, Z.; RUBIN, E. M. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. **Science**, v. 331, n. 6016, p. 463-467, Jan 28 2011.
- HOLDEN, N.; PRITCHARD, L.; TOTH, I. Colonization outwith the colon: plants as an alternative environmental reservoir for human pathogenic enterobacteria. **FEMS Microbiol. Rev.**, v. 33, n. 4, p. 689-703, Jul 2009.
- HOLLOWELLA, G. P.; KUYKENDALLB, L. D.; GILLETTEC, W. K.; HASHEMD, F. M.; HOUA, L.; TATEME, H. E.; DUTTA, S. K. Genetic transfer and expression of plasmid RP4::TOL in *Sinorhizobium meliloti*, *Bradyrhizobium japonicum* and *B. elkanii*. **Soil Biology and Biochemistry**, v. 31, n. 13, p. 1811-1819, 1999.
- HUSON, D. H.; AUCH, A. F.; QI, J.; SCHUSTER, S. C. MEGAN analysis of metagenomic data. **Genome Res.**, v. 17, n. 3, p. 377-386, Mar 2007.
- HUSSAIN, Y. A.; TEWFIK, M. S.; HANDI, Y. A. DEGRADATION OF CERTAIN AROMATIC COMPOUNDS BY RHIZOBIA. **Soil Biol. Biochem.**, v. 6, p. 377-381, 1974.
- IPT. Elaboração de subsídios técnicos para um plano de emergência para a área de Cubatão. São Paulo: Instituto de Pesquisas do Estado de São Paulo 1985.
- JANNASCH, H. W.; MOTTL, M. J. Geomicrobiology of deep-sea hydrothermal vents. **Science**, v. 229, n. 4715, p. 717-725, Aug 23 1985.
- JANSSEN, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. **Appl. Environ. Microbiol.**, v. 72, n. 3, p. 1719-1728, Mar 2006.
- JINDROVA, E.; CHOCOVA, M.; DEMNEROVA, K.; BRENNER, V. Bacterial aerobic degradation of benzene, toluene, ethylbenzene and xylene. **Folia Microbiol. (Praha)**, v. 47, n. 2, p. 83-93, 2002.
- JOHNSON, G. R.; OLSEN, R. H. Nucleotide sequence analysis of genes encoding a toluene/benzene-2-monooxygenase from *Pseudomonas* sp. strain JS150. **Appl. Environ. Microbiol.**, v. 61, n. 9, p. 3336-3346, Sep 1995.
- KAMPFER, P.; GLAESER, S. P. Prokaryotic taxonomy in the sequencing era--the polyphasic approach revisited. **Environmental microbiology**, v. 14, n. 2, p. 291-317, Feb 2012.
- KANAGAWA, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). **J. Biosci. Bioeng.**, v. 96, n. 4, p. 317-323, 2003.
- KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic Acids Res.**, v. 28, n. 1, p. 27-30, Jan 1 2000.
- KAO, C. M.; CHEN, C. Y.; CHEN, S. C.; CHIEN, H. Y.; CHEN, Y. L. Application of in situ biosparging to remediate a petroleum-hydrocarbon spill site: field and microbial evaluation. **Chemosphere**, v. 70, n. 8, p. 1492-1499, Feb 2008.
- KENT, W. J. BLAT--the BLAST-like alignment tool. **Genome Res.**, v. 12, n. 4, p. 656-664, Apr 2002.
- KHAN, F. I.; HUSAIN, T.; HEJAZI, R. An overview and analysis of site remediation technologies. **J. Environ. Manage.**, v. 71, n. 2, p. 95-122, Jun 2004.

KHOT, P. D.; COUTURIER, M. R.; WILSON, A.; CROFT, A.; FISHER, M. A. Optimization of MALDI-TOF Mass Spectrometry Analysis for Bacterial Identification. **Journal of clinical microbiology**, Sep 19 2012.

KIELAK, A.; PIJL, A. S.; VAN VEEN, J. A.; KOWALCHUK, G. A. Phylogenetic diversity of Acidobacteria in a former agricultural soil. **ISME J.**, v. 3, n. 3, p. 378-382, Mar 2009.

KNOTHE, H.; SHAH, P.; KRCMERY, V.; ANTAL, M.; MITSUHASHI, S. Transferable resistance to cefotaxime, ceftazidime, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. **Infection**, v. 11, n. 6, p. 315-317, Nov-Dec 1983.

KOCH, I. H.; GICH, F.; DUNFIELD, P. F.; OVERMANN, J. *Edaphobacter modestus* gen. nov., sp. nov., and *Edaphobacter aggregans* sp. nov., acidobacteria isolated from alpine and forest soils. **Int. J. Syst. Evol. Microbiol.**, v. 58, n. 5, p. 1114-1122, May 2008.

KONSTANTINIDIS, K. T.; TIEDJE, J. M. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. **Current opinion in microbiology**, v. 10, n. 5, p. 504-509, Oct 2007.

KRIEGER, C.; BELLER, H.; REINHARD, M.; SPORMANN, M. Initial Reactions in Anaerobic Oxidation of m-Xylene by the Denitrifying Bacterium *Azoarcus* sp. Strain T. **Journal of Bacteriology**. 181: 6403-6410 p. 1999.

KUBE, M.; HEIDER, J.; J, A.; HUFNAGEL, P.; KÜHNER, S.; BECK, A.; REINHARDT, R.; RABUS, R. Genes involved in the anaerobic degradation of toluene in a denitrifying bacterium, strain EbN1. **Arch. Microbiol.** 182-194 p. 2004.

KÜHNER, S.; WÖHLBRAND, L.; FRITZ, I.; WRUCK, W.; HULTSCHIG, C.; HUFNAGEL, P.; KUBE, M.; REINHARD, R.; RABUS, R. Substrate-Dependent Regulation of Anaerobic Degradation Pathways for Toluene and Ethylbenzene in a Denitrifying Bacterium, Strain EbN1. **Journal of Bacteriology**. 187: 1493-1503 p. 2005.

KYRPIDES, N. C. Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. **Bioinformatics**, v. 15, n. 9, p. 773-774, Sep 1999.

LAKAY, F. M.; BOTHA, A.; PRIOR, B. A. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. **J. Appl. Microbiol.**, v. 102, n. 1, p. 265-273, Jan 2007.

LARENTIS, M.; HOERMANN, K.; LUEDERS, T. Fine-scale degrader community profiling over an aerobic/anaerobic redox gradient in a toluene-contaminated aquifer. **Environmental Microbiology Reports**, p. 1-10, 2012.

LAY, J. O., JR. MALDI-TOF mass spectrometry of bacteria. **Mass spectrometry reviews**, v. 20, n. 4, p. 172-194, Jul-Aug 2001.

LEE, S. H.; KA, J. O.; CHO, J. C. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. **FEMS Microbiol. Lett.**, v. 285, n. 2, p. 263-269, Aug 2008.

LI, H.; ZHANG, Q.; WANG, X. L.; MA, X. Y.; LIN, K. F.; LIU, Y. D.; GU, J. D.; LU, S. G.; SHI, L.; LU, Q.; SHEN, T. T. Biodegradation of benzene homologues in contaminated sediment of the East China Sea. **Bioresour. Technol.**, v. 124, p. 129-136, Nov 2012.

LIU, B.; GIBBONS, T.; GHODSI, M.; TREANGEN, T.; POP, M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. **BMC Genomics**, v. 12 Suppl 2, p. S4, 2011.

MACEDO, A. J.; TIMMIS, K. N.; ABRAHAM, W. R. Widespread capacity to metabolize polychlorinated biphenyls by diverse microbial communities in soils with no significant exposure to PCB contamination. **Environ. Microbiol.**, v. 9, n. 8, p. 1890-1897, Aug 2007.

MAIDEN, M. C.; BYGRAVES, J. A.; FEIL, E.; MORELLI, G.; RUSSELL, J. E.; URWIN, R.; ZHANG, Q.; ZHOU, J.; ZURTH, K.; CAUGANT, D. A.; FEAVERS, I. M.; ACHTMAN, M.; SPRATT, B. G. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 6, p. 3140-3145, Mar 17 1998.

MANFREDI, R.; NANETTI, A.; FERRI, M.; CHIODO, F. Xanthomonas maltophilia: an emerging pathogen in patients with HIV disease. **Int. J. STD. AIDS.**, v. 9, n. 4, p. 201-207, Apr 1998.

MARCHETTI, A.; SCHRUTH, D. M.; DURKIN, C. A.; PARKER, M. S.; KODNER, R. B.; BERTHIAUME, C. T.; MORALES, R.; ALLEN, A. E.; ARMBRUST, E. V. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. **Proc. Natl. Acad. Sci. U. S. A.**, v. 109, n. 6, p. 317-325, Feb 7 2012.

MARMUR, J.; DOTY, P. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. **Journal of molecular biology**, v. 5, p. 109-118, Jul 1962.

MAZZEO, D. E.; LEVY, C. E.; DE ANGELIS DDE, F.; MARIN-MORALES, M. A. BTEX biodegradation by bacteria from effluents of petroleum refinery. **Sci. Total Environ.**, v. 408, n. 20, p. 4334-4340, Sep 15 2010.

MCGRATH, K. C.; THOMAS-HALL, S. R.; CHENG, C. T.; LEO, L.; ALEXA, A.; SCHMIDT, S.; SCHENK, P. M. Isolation and analysis of mRNA from environmental microbial communities. **J. Microbiol. Methods**, v. 75, n. 2, p. 172-176, Oct 2008.

MEGADI, V. B.; TALLUR, P. N.; MULLA, S. I.; NINNEKAR, H. Z. Bacterial degradation of fungicide captan. **J. Agric. Food. Chem.**, v. 58, n. 24, p. 12863-12868, Dec 22 2010.

MEYER, F.; PAARMANN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E. M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R. A. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v. 9, p. 386, 2008.

MILLER, D. N. Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. **J. Microbiol. Methods.**, v. 44, n. 1, p. 49-58, Feb 1 2001.

Ministério do Meio Ambiente. Mata Atlântica. Disponível em: <<http://www.mma.gov.br/biomas/mata-atlantica>>, Acesso em: 15/02/2013.

MOORE, E. R.; MIHAYLOVA, S. A.; VANDAMME, P.; KRICHEVSKY, M. I.; DIJKSHOORN, L. Microbial systematics and taxonomy: relevance for a microbial commons. **Research in microbiology**, v. 161, n. 6, p. 430-438, Jul-Aug 2010.

MORALES, S. E.; HOLBEN, W. E. Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. **Appl. Environ. Microbiol.**, v. 75, n. 9, p. 2677-2683, May 2009.

NASER, S. M.; THOMPSON, F. L.; HOSTE, B.; GEVERS, D.; DAWYNDT, P.; VANCANNEYT, M.; SWINGS, J. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. **Microbiology**, v. 151, n. 7, p. 2141-2150, Jul 2005.

NAVARRO-NOYA, Y. E.; JAN-ROBLERO, J.; GONZALEZ-CHAVEZ MDEL, C.; HERNANDEZ-GAMA, R.; HERNANDEZ-RODRIGUEZ, C. Bacterial communities associated with the rhizosphere of pioneer plants (*Bahia xylopada* and *Viguiera linearis*) growing on heavy metals-contaminated soils. **Antonie Van Leeuwenhoek**, v. 97, n. 4, p. 335-349, May 2010.

OLIVETTE, C. Programa de Controle de Poluição em Cubatão completa 25 anos. Disponível em: <www.ambiente.sp.gov.br/acontece/noticias/programa-de-controle-de-poluicao-em-cubatao-completa-25-anos/>, 2008. Acesso em: 13/01/2013.

PATTERSON, J. E.; HARDIN, T. C.; KELLY, C. A.; GARCIA, R. C.; JORGENSEN, J. H. Association of antibiotic utilization measures and control of multiple-drug resistance in *Klebsiella pneumoniae*. **Infect. Control. Hosp. Epidemiol.**, v. 21, n. 7, p. 455-458, Jul 2000.

PAUL, E. A. Soil, microbiology and biochemistry. 3rd Edition. UK: Academic Press, 2007. 337.

PEREZ-PANTOJA, D.; DONOSO, R.; JUNCA, H.; GONZÁLEZ, B.; PIEPER, D. Phylogenomics of Aerobic Bacterial Degradation of Aromatics. In: TIMIS, K. N. (Ed.). **Handbook of Hydrocarbon and Lipid Microbiology**: Springer-Verlag GmbH, 2010.

PETROSINO, J. F.; HIGHLANDER, S.; LUNA, R. A.; GIBBS, R. A.; VERSALOVIC, J. Metagenomic pyrosequencing and microbial identification. **Clin. Chem.**, v. 55, n. 5, p. 856-866, May 2009.

POLZ, M. F.; CAVANAUGH, C. M. Bias in template-to-product ratios in multitemplate PCR. **Appl. Environ. Microbiol.**, v. 64, n. 10, p. 3724-3730, Oct 1998.

PUNTA, M.; COGGILL, P. C.; EBERHARDT, R. Y.; MISTRY, J.; TATE, J.; BOURSNEILL, C.; PANG, N.; FORSLUND, K.; CERIC, G.; CLEMENTS, J.; HEGER, A.; HOLM, L.; SONNHAMMER, E. L.; EDDY, S. R.; BATEMAN, A.; FINN, R. D. The Pfam protein families database. **Nucleic Acids Res.**, v. 40, p. 290-301, Jan 2012.

QUAISER, A.; OCHSENREITER, T.; LANZ, C.; SCHUSTER, S. C.; TREUSCH, A. H.; ECK, J.; SCHLEPER, C. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. **Mol. Microbiol.**, v. 50, n. 2, p. 563-575, Oct 2003.

QUAISER, A.; ZIVANOVIC, Y.; MOREIRA, D.; LOPEZ-GARCIA, P. Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. **ISME J.**, v. 5, n. 2, p. 285-304, Feb 2011.

RAJENDHRAN, J.; GUNASEKARAN, P. Strategies for accessing soil metagenome for desired applications. **Biotechnology advances**, v. 26, n. 6, p. 576-590, Nov-Dec 2008.

RAMOS-GONZALEZ, M. I.; DUQUE, E.; RAMOS, J. L. Conjugational transfer of recombinant DNA in cultures and in soils: host range of *Pseudomonas putida* TOL plasmids. **Appl. Environ. Microbiol.**, v. 57, n. 10, p. 3020-3027, Oct 1991.

REISENFELD, C.; SCHLOSS, P.; HANDELSMAN, J. Metagenomics: Genomic Analysis of Microbial Communities. **Annu. Rev. Genet.**, p. 525-552, 2004.

- RHO, M.; TANG, H.; YE, Y. FragGeneScan: predicting genes in short and error-prone reads. **Nucleic Acids Res.**, v. 38, n. 20, p. 191, Nov 2010.
- RIESENFELD, C. S.; GOODMAN, R. M.; HANDELSMAN, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. **Environmental microbiology**, v. 6, n. 9, p. 981-989, Sep 2004.
- RIVKINA, E. M.; FRIEDMANN, E. I.; MCKAY, C. P.; GILICHINSKY, D. A. Metabolic activity of permafrost bacteria below the freezing point. **Applied and environmental microbiology**, v. 66, n. 8, p. 3230-3233, Aug 2000.
- ROESCH, L. F.; FULTHORPE, R. R.; RIVA, A.; CASELLA, G.; HADWIN, A. K.; KENT, A. D.; DAROUB, S. H.; CAMARGO, F. A.; FARMERIE, W. G.; TRIPLETT, E. W. Pyrosequencing enumerates and contrasts soil microbial diversity. **ISME J.**, v. 1, n. 4, p. 283-290, Aug 2007.
- RONAGHI, M.; UHLEN, M.; NYREN, P. A sequencing method based on real-time pyrophosphate. **Science**, v. 281, n. 5375, p. 363, 365, Jul 17 1998.
- ROSSELLO-MORA, R.; AMANN, R. The species concept for prokaryotes. **FEMS microbiology reviews**, v. 25, n. 1, p. 39-67, Jan 2001.
- SACK, R. B. Human diarrheal disease caused by enterotoxigenic Escherichia coli. **Annu. Rev. Microbiol.**, v. 29, p. 333-353, 1975.
- SALLES, J. F.; VAN VEEN, J. A.; VAN ELSAS, J. D. Multivariate analyses of Burkholderia species in soil: effect of crop and land use history. **Appl. Environ. Microbiol.**, v. 70, n. 7, p. 4012-4020, Jul 2004.
- SANDEGREN, L.; LINKEVICIUS, M.; LYTSY, B.; MELHUS, A.; ANDERSSON, D. I. Transfer of an Escherichia coli ST131 multiresistance cassette has created a Klebsiella pneumoniae-specific plasmid associated with a major nosocomial outbreak. **J. Antimicrob. Chemother.**, v. 67, n. 1, p. 74-83, Jan 2012.
- SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal of molecular biology**, v. 94, n. 3, p. 441-448, May 25 1975.
- SCHLEIFER, K. H. Classification of Bacteria and Archaea: past, present and future. **Systematic and applied microbiology**, v. 32, n. 8, p. 533-542, Dec 2009.
- SCHLOSS, P. D.; HANDELSMAN, J. Biotechnological prospects from metagenomics. **Current opinion in biotechnology**, v. 14, n. 3, p. 303-310, Jun 2003.
- SHANGE, R. S.; ANKUMAH, R. O.; IBEKWE, A. M.; ZABAWA, R.; DOWD, S. E. Distinct soil bacterial communities revealed under a diversely managed agroecosystem. **PLoS One**, v. 7, n. 7, p. 1-11, 2012.
- SHIELDS, M. S.; REAGIN, M. J.; GERGER, R. R.; CAMPBELL, R.; SOMERVILLE, C. TOM, a new aromatic degradative plasmid from Burkholderia (Pseudomonas) cepacia G4. **Appl. Environ. Microbiol.**, v. 61, n. 4, p. 1352-1356, Apr 1995.
- SHINODA, Y.; SAKAI, Y.; UENISHI, H.; UCHIHASHI, Y.; HIRAISHI, A.; YUKAWA, H.; YURIMOTO, H.; KATO, N. Aerobic and anaerobic toluene degradation by a newly isolated

denitrifying bacterium, *Thauera* sp. strain DNT-1. **Appl. Environ. Microbiol.**, v. 70, n. 3, p. 1385-1392, Mar 2004.

SHON, A. S.; BAJWA, R. P.; RUSSO, T. A. Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: A new and dangerous breed. **Virulence**, v. 4, n. 2, Jan 9 2013.

SHUKLA, A. K.; UPADHYAY, S. N.; DUBEY, S. K. Current trends in trichloroethylene biodegradation: a review. **Crit. Rev. Biotechnol.**, Oct 11 2012.

SILVA, C. C.; HAYDEN, H.; SAWBRIDGE, T.; MELE, P.; KRUGER, R. H.; RODRIGUES, M. V.; COSTA, G. G.; VIDAL, R. O.; SOUSA, M. P.; TORRES, A. P.; SANTIAGO, V. M.; OLIVEIRA, V. M. Phylogenetic and functional diversity of metagenomic libraries of phenol degrading sludge from petroleum refinery wastewater treatment system. **AMB Express**, v. 2, n. 1, p. 18, 2012.

SIMON, C.; WIEZER, A.; STRITTMATTER, A. W.; DANIEL, R. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. **Appl. Environ. Microbiol.**, v. 75, n. 23, p. 7519-7526, Dec 2009.

SINGLETON, D. R.; JONES, M. D.; RICHARDSON, S. D.; AITKEN, M. D. Pyrosequence analyses of bacterial communities during simulated in situ bioremediation of polycyclic aromatic hydrocarbon-contaminated soil. **Appl. Microbiol. Biotechnol.**, Nov 7 2012.

SMA. Meio Ambiente Paulista: Relatório de Qualidade Ambiental 2012. AMBIENTAL, S. D. M. A. C. D. P. São Paulo: SMA / CPLA: n. 252, 2012.

SMITH, A. M.; KIRISITS, M. J.; REIBLE, D. D. Assessment of potential anaerobic biotransformation of organic pollutants in sediment caps. **N. Biotechnol.**, v. 30, n. 1, p. 80-87, Nov 15 2012.

STACKEBRANDT, E.; FREDERIKSEN, W.; GARRITY, G. M.; GRIMONT, P. A.; KAMPFER, P.; MAIDEN, M. C.; NESME, X.; ROSSELLO-MORA, R.; SWINGS, J.; TRUPER, H. G.; VAUTERIN, L.; WARD, A. C.; WHITMAN, W. B. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. **International journal of systematic and evolutionary microbiology**, v. 52, n. 3, p. 1043-1047, May 2002.

STOVER, C. K.; PHAM, X. Q.; ERWIN, A. L.; MIZOGUCHI, S. D.; WARRENER, P.; HICKEY, M. J.; BRINKMAN, F. S.; HUFNAGLE, W. O.; KOWALIK, D. J.; LAGROU, M.; GARBER, R. L.; GOLTRY, L.; TOLENTINO, E.; WESTBROCK-WADMAN, S.; YUAN, Y.; BRODY, L. L.; COULTER, S. N.; FOLGER, K. R.; KAS, A.; LARBIG, K.; LIM, R.; SMITH, K.; SPENCER, D.; WONG, G. K.; WU, Z.; PAULSEN, I. T.; REIZER, J.; SAIER, M. H.; HANCOCK, R. E.; LORY, S.; OLSON, M. V. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. **Nature**, v. 406, n. 6799, p. 959-964, Aug 31 2000.

STREIT, W.; RA, S. Metagenomics – the key to the uncultured microbes. **Current Opinion in Microbiology**. n.7, p. 492-498, 2004.

SUN, W.; CUPPLES, A. M. Diversity of five anaerobic toluene-degrading microbial communities investigated using stable isotope probing. **Appl. Environ. Microbiol.**, v. 78, n. 4, p. 972-980, Feb 2012.

SUZUKI, M. T.; GIOVANNONI, S. J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. **Appl. Environ. Microbiol.**, v. 62, n. 2, p. 625-630, Feb 1996.

TANCSICS, A.; SZOBOSZLAY, S.; KRISZT, B.; KUKOLYA, J.; BAKA, E.; MARIALIGETI, K.; REVESZ, S. Applicability of the functional gene catechol 1,2-dioxygenase as a biomarker in the detection of BTEX-degrading *Rhodococcus* species. **J. Appl. Microbiol.**, v. 105, n. 4, p. 1026-1033, Oct 2008.

TANCSICS, A.; SZOBOSZLAY, S.; SZABO, I.; FARKAS, M.; KOVACS, B.; KUKOLYA, J.; MAYER, Z.; KRISZT, B. Quantification of subfamily I.2.C catechol 2,3-dioxygenase mRNA transcripts in groundwater samples of an oxygen-limited BTEX-contaminated site. **Environ. Sci. Technol.**, v. 46, n. 1, p. 232-240, Jan 3 2012.

TERAMOTO, M.; FUTAMATA, H.; HARAYAMA, S.; WATANABE, K. Characterization of a high-affinity phenol hydroxylase from *Comamonas testosteroni* R5 by gene cloning, and expression in *Pseudomonas aeruginosa* PAO1c. **Mol. Gen. Genet.**, v. 262, n. 3, p. 552-558, Oct 1999.

TINDALL, B. J.; ROSSELLO-MORA, R.; BUSSE, H. J.; LUDWIG, W.; KAMPFER, P. Notes on the characterization of prokaryote strains for taxonomic purposes. **International journal of systematic and evolutionary microbiology**, v. 60, n. 1, p. 249-266, Jan 2010.

TORRES, F. R.; BRAGA JR, J. C.; BORGES, W. R. O que você precisa saber sobre Cubatão. São Paulo: Design & Print e Arquivo Histórico Municipal de Cubatão, 2002.

TRAVIS, L.; CHIN-YANG, L.; ZHANG, Z.; LI, D.-G.; YIN, S.-N.; CHAO, W.-H.; LI, G.-L.; DOSEMECI, M.; BLOT, W.; FRAUMENI, J.; JR; HAYES, R.; LINET, M. Hematopoietic Malignancies and Related Disorders Among Benzene-Exposed Workers in China. **Leukemia and Lymphoma**. n. 14, p. 91-102, 1994.

TRINGE, S. G.; VON MERING, C.; KOBAYASHI, A.; SALAMOV, A. A.; CHEN, K.; CHANG, H. W.; PODAR, M.; SHORT, J. M.; MATHUR, E. J.; DETTER, J. C.; BORK, P.; HUGENHOLTZ, P.; RUBIN, E. M. Comparative metagenomics of microbial communities. **Science**, v. 308, n. 5721, p. 554-557, Apr 22 2005.

URICH, T.; LANZEN, A.; QI, J.; HUSON, D. H.; SCHLEPER, C.; SCHUSTER, S. C. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. **PLoS One**, v. 3, n. 6, p. e2527, 2008.

VALDERRAMA, M. J.; PRADO, B.; DEL MORAL, A.; RIOS, R.; RAMOS-CORMENZANA, A.; CAMPOS, V. Numerical taxonomy of moderately halophilic gram-positive cocci isolated from the Salar de Atacama (Chile). **Microbiologia**, v. 7, n. 1, p. 35-41, Jun 1991.

VAN NIEL, C. B.; ALLEN, M. B. A note on *Pseudomonas stutzeri*. **Journal of bacteriology**, v. 64, n. 3, p. 413-422, Sep 1952.

VANDAMME, P.; POT, B.; GILLIS, M.; DE VOS, P.; KERSTERS, K.; SWINGS, J. Polyphasic taxonomy, a consensus approach to bacterial systematics. **Microbiological reviews**, v. 60, n. 2, p. 407-438, Jun 1996.

VAZQUEZ, S.; NOGALES, B.; RUBERTO, L.; HERNANDEZ, E.; CHRISTIE-OLEZA, J.; LO BALBO, A.; BOSCH, R.; LALUCAT, J.; MAC CORMACK, W. Bacterial community dynamics during bioremediation of diesel oil-contaminated Antarctic soil. **Microb. Ecol.**, v. 57, n. 4, p. 598-610, May 2009.

VENTER, J. C.; REMINGTON, K.; HEIDELBERG, J. F.; HALPERN, A. L.; RUSCH, D.; EISEN, J. A.; WU, D.; PAULSEN, I.; NELSON, K. E.; NELSON, W.; FOUTS, D. E.; LEVY, S.; KNAP, A. H.; LOMAS, M. W.; NEALSON, K.; WHITE, O.; PETERSON, J.; HOFFMAN, J.; PARSONS, R.; BADEN-TILLSON, H.; PFANNKOCHE, C.; ROGERS, Y. H.; SMITH, H. O. Environmental

genome shotgun sequencing of the Sargasso Sea. **Science**, v. 304, n. 5667, p. 66-74, Apr 2 2004.

VIDALI, M. Bioremediation. An overview. **Pure Appl. Chem.**, v. 73, n. 7, p. 1163-1172, 2001.

VOGEL, T. M. Bioaugmentation as a soil bioremediation approach. **Curr. Opin. Biotechnol.**, v. 7, n. 3, p. 311-316, Jun 1996.

WAYNE, L. G. International Committee on Systematic Bacteriology: announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. **Zentralblatt für Bakteriologie, Mikrobiologie, und Hygiene. Series A, Medical microbiology, infectious diseases, virology, parasitology**, v. 268, n. 4, p. 433-434, Jun 1988.

WEELINK, S. A. B.; EEKERT, M. H. A.; M., S. A. J. Degradation of BTEX by anaerobic bacteria: physiology and application. **Rev. Environ. Sci. Biotechnol.** p. 359-385 2010.

WHITMAN, W. B.; COLEMAN, D. C.; WIEBE, W. J. Prokaryotes: the unseen majority. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 12, p. 6578-6583, Jun 9 1998.

WOESE, C. R. Bacterial evolution. **Microbiological reviews**, v. 51, n. 2, p. 221-271, Jun 1987.

World Widelife Fund. Mata Atlântica. Disponível em: <http://www.wwf.org.br/natureza_brasileira/areas_prioritarias/mata_atlantica/>, 2013. Acesso em: 15/02/2013.

WU, M.; EISEN, J. A. A simple, fast, and accurate method of phylogenomic inference. **Genome Biol**, v. 9, n. 10, p. R151, 2008.

XIE, S.; SUN, W.; LUO, C.; CUPPLES, A. M. Novel aerobic benzene degrading microorganisms identified in three soils by stable isotope probing. **Biodegradation**, v. 22, n. 1, p. 71-81, Feb 2011.

XIN, B. P.; WU, C. H.; LIN, C. W. Bioaugmented remediation of high concentration BTEX-contaminated groundwater by permeable reactive barrier with immobilized bead. **J. Hazard. Mater.**, v. 244-245, p. 765-772, Jan 15 2013.

YEN, K. M.; KARL, M. R.; BLATT, L. M.; SIMON, M. J.; WINTER, R. B.; FAUSSET, P. R.; LU, H. S.; HARCOURT, A. A.; CHEN, K. K. Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. **J. Bacteriol.**, v. 173, n. 17, p. 5315-5327, Sep 1991.

YERGEAU, E.; SANSCHAGRIN, S.; BEAUMIER, D.; GREER, C. W. Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high arctic soils. **PLoS One**, v. 7, n. 1, p. e30058, 2012.

ZHOU, N. Y.; JENKINS, A.; CHAN KWO CHION, C. K.; LEAK, D. J. The alkene monooxygenase from *Xanthobacter* strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol. **Appl. Environ. Microbiol.**, v. 65, n. 4, p. 1589-1595, Apr 1999.

ZYLSTRA, G. J.; KIM, E. Aromatic hydrocarbon degradation by *Sphingomonas yanoikuyae* B1. **Journal of Industrial Microbiology and Biotechnology**, v. 19, n. 5-6, p. 408-414, 1997.

ANEXO – Análise físico química dos solos contaminado e Perequê

Análise técnica dos solos dos locais Perequê e contaminado nesta ordem, realizada pela empresa Bioagri, identificando as características físico-químicas de cada um deles.

BOLETIM DE ANÁLISE Nº 60416/2011-0
Processo Comercial Nº 3727/2011-1

DADOS REFERENTES AO CLIENTE

Empresa solicitante:	Bruno Karolski
Endereço:	Rua do Matao, 277 - sala 153 Cidade Universitaria - São Paulo-SP - CEP: 05.508-090 .
Nome do Solicitante:	Bruno Karolski

DADOS REFERENTES A AMOSTRA

Identificação do Cliente:	Solo		
Amostra Rotulada como:	Solo		
Coletor:	Interessado	Data da coleta:	22/3/2011
Data da entrada no laboratório:	22/03/2011 14:03:00	Data de Elaboração do BA:	11/04/2011

RESULTADOS ANALÍTICOS DA AMOSTRA

Parâmetros	Unidade	LQ	Resultados analíticos
Porcentagem de Sólidos	% p/p	0,05	82,9
Índice de Fenóis	mg/kg	0,06	1,5
Prata	mg/kg	0,4	< 0,4
Alumínio	mg/kg	0,4	10600
Arsênio	mg/kg	0,4	< 0,4
Bário	mg/kg	0,4	37
Berílio	mg/kg	0,4	< 0,4
Bismuto	mg/kg	0,4	< 0,4
Cádmio	mg/kg	0,043	< 0,043
Cobalto	mg/kg	0,4	5,2
Cromo	mg/kg	0,4	24
Lítio	mg/kg	0,4	11
Molibdênio	mg/kg	0,4	< 0,4
Sódio	mg/kg	21	77
Níquel	mg/kg	0,4	6,7
Chumbo	mg/kg	0,4	46
Antimônio	mg/kg	0,4	< 0,4
Selênio	mg/kg	0,4	< 0,4
Estanho	mg/kg	0,4	< 0,4
Estrôncio	mg/kg	0,4	1,7
Titânio	mg/kg	0,4	560
Tálio	mg/kg	0,4	< 0,4
Urânio	mg/kg	0,4	< 0,4
Vanádio	mg/kg	0,4	26
Dureza Calculada	mg/kg	141,4	7776

BTEX

Parâmetros	Unidade	LQ	Resultados analíticos
Benzeno	mg/kg	0,006	< 0,006
Tolueno	mg/kg	0,006	< 0,006
Etilbenzeno	mg/kg	0,006	< 0,006
o-Xileno	mg/kg	0,006	< 0,006
m,p-Xilenos	mg/kg	0,012	< 0,012

PAH

Parâmetros	Unidade	LQ	Resultados analíticos
Benzo(a)pireno	mg/kg	0,24	< 0,24
Benzo(b)fluoranteno	mg/kg	0,24	< 0,24
Benzo(k)fluoranteno	mg/kg	0,24	< 0,24
Criseno	mg/kg	0,24	< 0,24
Acenaftileno	mg/kg	0,24	< 0,24
Fluoreno	mg/kg	0,24	< 0,24
Antraceno	mg/kg	0,24	< 0,24
Benzo(g,h,i)perileno	mg/kg	0,24	< 0,24
Fenantreno	mg/kg	0,24	< 0,24
Dibenzo(a,h)antraceno	mg/kg	0,24	< 0,24
Indeno(1,2,3,cd)pireno	mg/kg	0,24	< 0,24
Pireno	mg/kg	0,24	< 0,24

PAH

Parâmetros	Unidade	LQ	Resultados analíticos
Acenafteno	mg/kg	0,24	< 0,24
Fluoranteno	mg/kg	0,24	< 0,24
Naftaleno	mg/kg	0,24	< 0,24
Benzo(a)antraceno	mg/kg	0,24	< 0,24

CTC - Capacidade de Troca Catiônica

Parâmetros	Unidade	LQ	Resultados analíticos
pH em CaCl ₂		---	4,3
Matéria Orgânica	g/dm ³	---	8
Fósforo Resina	mg/dm ³	---	1
Potássio	mmol/dm ³		0,6
Cálcio	mmol/dm ³		16
Magnésio	mmol/dm ³		4
H+Al	mmol/dm ³	---	31
Al	mmol/dm ³	---	4
Soma de Bases	mmol/dm ³	---	21
CTC	mmol/dm ³	---	52
Saturação de Bases	V%	---	40
Saturação Al	m%	---	16

Micronutrientes

Parâmetros	Unidade	LQ	Resultados analíticos
Cobre	mg/kg	0,4	12
Ferro	mg/kg	0,4	8668
Zinco	mg/kg	0,4	24
Manganês	mg/kg	0,4	153
Boro	mg/kg	0,4	81

Granulometria

Parâmetros	Unidade	LQ	Resultados analíticos
Argila (< 0,002 mm)	g/kg	0,01	0
Silte (0,053 a 0,002 mm)	g/kg	0,01	305
Areia Total (2,00 a 0,053 mm)	g/kg	0,01	695
Areia Grossa (2,00 a 0,210 mm)	g/kg	0,01	460
Areia Fina (0,210 a 0,053 mm)	g/kg	0,01	235

Macronutrientes - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Nitrogênio Total Kjeldahl	mg/kg	2,7	271
Enxofre	mg/kg	60	78
Fósforo	mg/kg	0,4	60

CONTROLE DE QUALIDADE DO LABORATÓRIO**Controle de Qualidade - TPH Voláteis + BTEX - Solo****60738/2011-0 - Branco de Análise - TPH Voláteis + TEXB - Solo**

Parâmetros	Unidade	LQ	Resultados analíticos
Benzeno	mg/kg	0,005	< 0,005
Tolueno	mg/kg	0,005	< 0,005
Etilbenzeno	mg/kg	0,005	< 0,005
m,p-Xilenos	mg/kg	0,01	< 0,01
o-Xileno	mg/kg	0,005	< 0,005

Ensaio de Recuperação

Parâmetros	Quantidade Adicionada	Unidade	Resultado da Recuperação (%)	Faixa Aceitável de Recuperação (%)
60739/2011-0 - LCS - TPH Voláteis + BTEX - Solo				
Benzeno	0,05	mg/kg	70	60 -140
Tolueno	0,05	mg/kg	73	60 -140
Etilbenzeno	0,05	mg/kg	74	60 -140
m,p-Xilenos	0,1	mg/kg	75	60 -140
o-Xileno	0,05	mg/kg	72	60 -140
Surrogates				
60738/2011-0 - Branco de Análise - TPH Voláteis + TEXB - Solo				
Tolueno-d8	0,05	%	75	60 -140
p-Bromofluorbenzeno	0,05	%	83	60 -140
60739/2011-0 - LCS - TPH Voláteis + BTEX - Solo				
Tolueno-d8	0,05	%	80	60 -140
p-Bromofluorbenzeno	0,05	%	91	60 -140
60416/2011-0 - Solo				
Tolueno-d8	0,05	%	81	60 -140

p-Bromofluorbenzeno	0,05	%	89	60 - 140
---------------------	------	---	----	----------

Controle de Qualidade - PAH - Amostras Sólidas

61573/2011-0 - Branco de Análise - PAH - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Benzo(a)pireno	mg/kg	0,01	< 0,01
Benzo(b)fluoranteno	mg/kg	0,01	< 0,01
Benzo(k)fluoranteno	mg/kg	0,01	< 0,01
Criseno	mg/kg	0,01	< 0,01
Acenaftileno	mg/kg	0,01	< 0,01
Fluoreno	mg/kg	0,01	< 0,01
Antraceno	mg/kg	0,01	< 0,01
Benzo(g,h,i)perileno	mg/kg	0,01	< 0,01
Fenantreno	mg/kg	0,01	< 0,01
Dibenzo(a,h)antraceno	mg/kg	0,01	< 0,01
Indeno(1,2,3,cd)pireno	mg/kg	0,01	< 0,01
Pireno	mg/kg	0,01	< 0,01
Acenafteno	mg/kg	0,01	< 0,01
Fluoranteno	mg/kg	0,01	< 0,01
Naftaleno	mg/kg	0,01	< 0,01
Benzo(a)antraceno	mg/kg	0,01	< 0,01

Ensaio de Recuperação

Parâmetros	Quantidade Adicionada	Unidade	Resultado da Recuperação (%)	Faixa Aceitável de Recuperação (%)
61574/2011-0 - LCS - PAH - Solo				
Fenantreno	0,2	mg/kg	57	40 - 120
Naftaleno	0,2	mg/kg	45	40 - 120
Fluoranteno	0,2	mg/kg	66	40 - 120
Criseno	0,2	mg/kg	49	40 - 120
Pireno	0,2	mg/kg	64	40 - 120
Surrogates				
61573/2011-0 - Branco de Análise - PAH - Solo				
2-Fluorbifenil	0,2	%	50	40 - 120
Terfenil d14	0,2	%	66	40 - 120
61574/2011-0 - LCS - PAH - Solo				
2-Fluorbifenil	0,2	%	51	40 - 120
Terfenil d14	0,2	%	72	40 - 120
60416/2011-0 - Solo				
Terfenil d14	0,24	%	Amostra Diluída	40 - 120
2-Fluorbifenil	0,24	%	Amostra Diluída	40 - 120

Controle de Qualidade - Metais - Solo

57802/2011-0 - Branco de Análise - Metais - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Prata	mg/kg	0,5	< 0,5
Prata	mg/kg	0,5	< 0,5
Alumínio	mg/kg	0,5	< 0,5
Alumínio	mg/kg	0,5	11
Arsênio	mg/kg	0,5	< 0,5
Arsênio	mg/kg	0,5	< 0,5
Bário	mg/kg	0,5	< 0,5
Bário	mg/kg	0,5	< 0,5
Berílio	mg/kg	0,5	< 0,5
Berílio	mg/kg	0,5	< 0,5
Bismuto	mg/kg	0,5	< 0,5
Bismuto	mg/kg	0,5	< 0,5
Cálcio	mg/kg	25	< 25
Cálcio	mg/kg	25	< 25
Cádmio	mg/kg	0,05	< 0,05
Cádmio	mg/kg	0,05	< 0,05
Cobalto	mg/kg	0,5	< 0,5
Cobalto	mg/kg	0,5	< 0,5
Cromo	mg/kg	0,5	< 0,5
Cromo	mg/kg	0,5	< 0,5
Cobre	mg/kg	0,5	< 0,5
Cobre	mg/kg	0,5	< 0,5
Ferro	mg/kg	0,5	< 0,5
Ferro	mg/kg	0,5	< 0,5
Potássio	mg/kg	25	44
Potássio	mg/kg	25	< 25
Lítio	mg/kg	0,5	< 0,5
Lítio	mg/kg	0,5	< 0,5
Magnésio	mg/kg	25	< 25
Magnésio	mg/kg	25	< 25
Manganês	mg/kg	0,5	< 0,5
Manganês	mg/kg	0,5	< 0,5
Molibdênio	mg/kg	0,5	< 0,5
Molibdênio	mg/kg	0,5	< 0,5
Sódio	mg/kg	25	< 25

57802/2011-0 - Branco de Análise - Metais - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Sódio	mg/kg	25	< 25
Níquel	mg/kg	0,5	< 0,5
Níquel	mg/kg	0,5	< 0,5
Chumbo	mg/kg	0,5	< 0,5
Chumbo	mg/kg	0,5	< 0,5
Antimônio	mg/kg	0,5	< 0,5
Antimônio	mg/kg	0,5	2,5
Selênio	mg/kg	0,5	< 0,5
Selênio	mg/kg	0,5	< 0,5
Estanho	mg/kg	0,5	< 0,5
Estanho	mg/kg	0,5	< 0,5
Estrôncio	mg/kg	0,5	< 0,5
Estrôncio	mg/kg	0,5	< 0,5
Titânio	mg/kg	0,5	< 0,5
Titânio	mg/kg	0,5	< 0,5
Tálio	mg/kg	0,5	1,9
Tálio	mg/kg	0,5	< 0,5
Urânio	mg/kg	0,5	< 0,5
Urânio	mg/kg	0,5	< 0,5
Vanádio	mg/kg	0,5	< 0,5
Vanádio	mg/kg	0,5	< 0,5
Zinco	mg/kg	0,5	< 0,5
Zinco	mg/kg	0,5	< 0,5

Ensaio de Recuperação

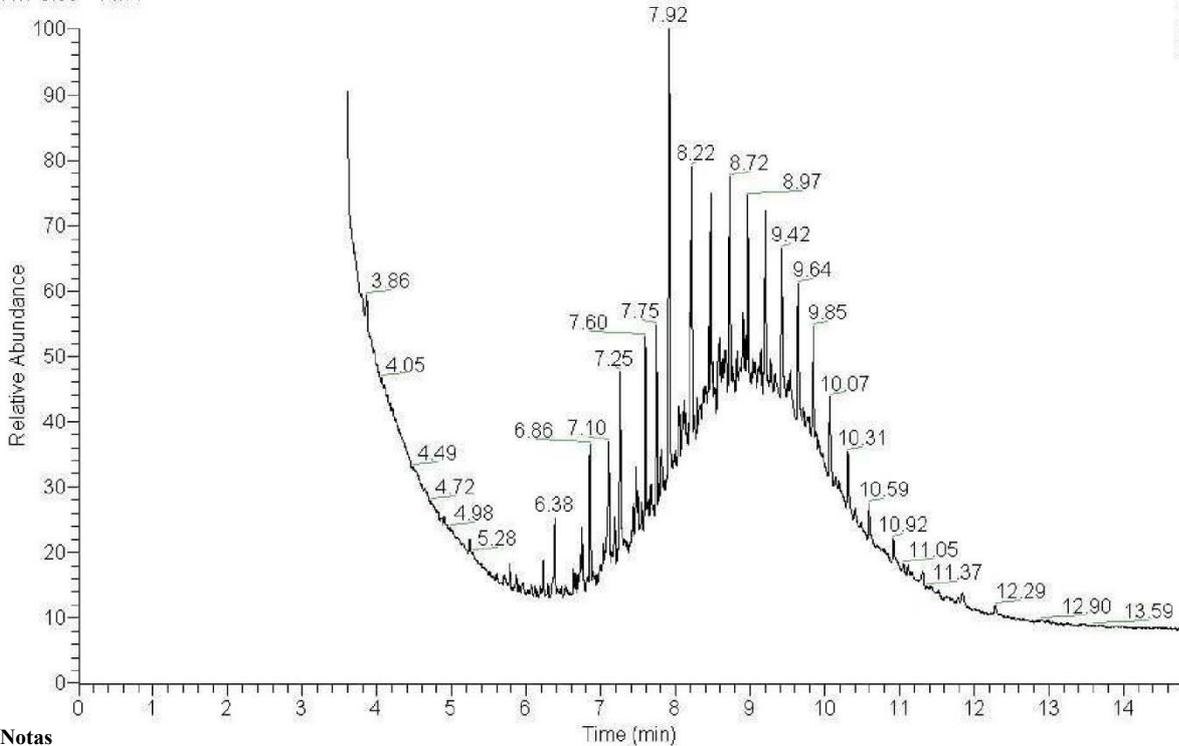
Parâmetros	Quantidade Adicionada	Unidade	Resultado da Recuperação (%)	Faixa Aceitável de Recuperação (%)
57803/2011-0 - Solo Padrão - RTC - CRM023				
Alumínio		mg/kg		
Alumínio	8470	mg/kg	100	70 - 130
Arsênio	380	mg/kg	100	95 - 105
Arsênio		mg/kg		
Bário		mg/kg		
Bário	75,5	mg/kg	100	85 - 115
Cádmio	0,92	mg/kg	98	50 - 150
Cádmio		mg/kg		
Cálcio		mg/kg		
Cálcio	5420	mg/kg	100	85 - 115
Cromo		mg/kg		
Cromo	31	mg/kg	100	80 - 120
Cobalto		mg/kg		
Cobalto	4,68	mg/kg	99	85 - 115
Cobre	8,9	mg/kg	96	80 - 120
Cobre		mg/kg		
Ferro	10700	mg/kg	100	85 - 115
Ferro		mg/kg		
Chumbo	213	mg/kg	100	90 - 110
Chumbo		mg/kg		
Magnésio		mg/kg		
Magnésio	3060	mg/kg	101	85 - 115
Manganês	206	mg/kg	98	90 - 110
Manganês		mg/kg		
Níquel		mg/kg		
Níquel	11	mg/kg	100	85 - 115
Potássio	2230	mg/kg	100	65 - 135
Potássio		mg/kg		
Selênio		mg/kg		
Selênio	105	mg/kg	100	90 - 110
Tálio	111	mg/kg	100	90 - 110
Tálio		mg/kg		
Vanádio		mg/kg		
Vanádio	21,7	mg/kg	99	70 - 130
Zinco	93,8	mg/kg	100	85 - 115
Zinco		mg/kg		
Surrogates				
57802/2011-0 - Branco de Análise - Metais - Solo				
Ítrio	50	%	90	70-130
Ítrio	50	%	90	70-130
57803/2011-0 - Solo Padrão - RTC - CRM023				
Ítrio	50	%	86	70-130
Ítrio		%		
60416/2011-0 - Solo				
Ítrio	42,8	%	97	70-130

Página 4 de 6 / B.A.: 60416/2011-0

Bioagri Ambiental - Piracicaba - SP. Rua Aljovil Martini, 201 - Dois Córregos - Piracicaba - SP - CEP: 13420-750 - Fone: (19) 3417.4700 - Fax: (19) 3417.4711 - vendas@bioagriambiental.com.br

PAH - Solo

RT: 0.00 - 14.77



Notas

LQ = Limite de Quantificação.

Abrangência

O(s) resultado(s) se referem somente à(s) amostra(s) analisada(s).

Este Boletim de Análise só pode ser reproduzido por inteiro e sem nenhuma alteração.

Data de realização das análises

A Bioagri Ambiental garante que todas as análises oramf executadas dentro do prazo de validade de cada parâmetro segundo o Guia de Coleta e Preservação de Amostra da Bioagri Ambiental, quando todo o trâmite analítico (coleta e análise) é de esponsabilider da Bioagri Ambiental. Quando a coleta é de responsabilidade do interessado, caso haja algum desvio, o cliente é previamenteconsultado sobre a disposição das amostras e a continuidade do processo analítico.

Todas estas datas constam nos dados brutos das análises e estão à disposição para serem solicitadas a qualquer momento pelo interessado.

Plano de Amostragem

Plano de amostragem de responsabilidade do interessado.

Trabalhos Subcontratados

As análises dos parâmetros pH em CaCl₂, Matéria Orgânica, Fósforo Resina, Potássio, Cálcio, Magnésio, H+Al, Al, Soma de Bases, CTC, Saturação de Bases, Saturação Al foram executadas e m laboratório subcontratado autorizados pela GQL – Bioagri Ambiental.

Referências Metodológicas

Índice de Fenóis: USEPA SW 846 - 9065 - Phenolics (Spectrophotometric, Manual 4-AAP with Distillation)

Nitrogênio Total Kjeldahl: POP PA 113 / SMEWW 4500

Granulometria: ISO 13320-1/1999 - Análise de Partícula por Difração a Laser

SVOC's: POP PA 076, 096 / USEPA SW 846 - 8270D e 351°C, SMWW 6410B

Metais (ICP-OES): POP PA 035 / SMWW 3120 B, USEPA 6010

VOC : POP PA 075 / USEPA SW 846 826°C, 5021A

Sólidos Totais: POP PA 009 / SMWW 2540B

Revisores

Débora Fernandes da Silva

Rogério Caldorin

Nereida Aparecida Bongiorno

André Alex Colletti



Chave de Validação: a6°C3300e0b263af44bbc58bedd7a97 f


Juliana Bombasaro
Coordenadora de Projeto
CRQ 04469985 - 4ª Região

Página 6 de 6 / B.A.: 60416/2011-0

Bioagri Ambiental - Piracicaba - SP, Rua Aljovil Martini, 201 - Dois Córregos - Piracicaba - SP - CEP: 13420-750 - Fone: (19) 3417.4700 - Fax: (19) 3417.4711 - vendas@bioagriambiental.com.br

BOLETIM DE ANÁLISE N° 17657/2011-0
Processo Comercial N° 1054/2011-2

DADOS REFERENTES AO CLIENTE

Empresa solicitante:	Bruno Karolski
Endereço:	Rua do Matao, 277 - Cidade Universitaria - São Paulo - SP - CEP: 05.508-090 .
Nome do Solicitante:	Bruno Karolski

DADOS REFERENTES A AMOSTRA

Identificação do Cliente:	Solo NC		
Amostra Rotulada como:	Solo		
Coletor:	Interessado	Data da coleta:	26/1/2011
Data da entrada no laboratório:	27/01/2011 13:53:00	Data de Elaboração do BA:	09/02/2011

RESULTADOS ANALÍTICOS DA AMOSTRA

Parâmetros	Unidade	LQ	Resultados analíticos
Porcentagem de Sólidos	% p/p	0,05	67,3
Índice de Fenóis	mg/kg	0,07	< 0,07
Prata	mg/kg	0,6	< 0,6
Alumínio	mg/kg	0,6	21752
Arsênio	mg/kg	0,6	15
Bário	mg/kg	0,6	50
Berílio	mg/kg	0,6	< 0,6
Bismuto	mg/kg	0,6	< 0,6
Cádmio	mg/kg	0,061	< 0,061
Cobalto	mg/kg	0,6	8,3
Cromo	mg/kg	0,6	35
Lítio	mg/kg	0,6	15
Molibdênio	mg/kg	0,6	< 0,6
Sódio	mg/kg	30	88
Níquel	mg/kg	0,6	19
Chumbo	mg/kg	0,6	26
Antimônio	mg/kg	0,6	< 0,6
Selênio	mg/kg	0,6	< 0,6
Estanho	mg/kg	0,6	< 0,6
Estrôncio	mg/kg	0,6	2,4
Titânio	mg/kg	0,6	1026
Tálio	mg/kg	0,6	< 0,6
Urânio	mg/kg	0,6	< 0,6
Vanádio	mg/kg	0,6	61
Dureza Calculada	mg/kg	201,4	14838

BTEX

Parâmetros	Unidade	LQ	Resultados analíticos
Benzeno	mg/kg	0,007	< 0,007
Tolueno	mg/kg	0,007	< 0,007
Etilbenzeno	mg/kg	0,007	< 0,007
o-Xileno	mg/kg	0,007	< 0,007
m,p-Xilenos	mg/kg	0,015	< 0,015

PAH

Parâmetros	Unidade	LQ	Resultados analíticos
Benzo(a)pireno	mg/kg	0,01	< 0,01
Benzo(b)fluoranteno	mg/kg	0,01	< 0,01
Benzo(k)fluoranteno	mg/kg	0,01	< 0,01
Criseno	mg/kg	0,01	< 0,01
Acenaftileno	mg/kg	0,01	< 0,01
Fluoreno	mg/kg	0,01	< 0,01
Antraceno	mg/kg	0,01	< 0,01
Benzo(g,h,i)perileno	mg/kg	0,01	< 0,01
Fenantreno	mg/kg	0,01	< 0,01
Dibenzo(a,h)antraceno	mg/kg	0,01	< 0,01
Indeno(1,2,3,cd)pireno	mg/kg	0,01	< 0,01

PAH

Parâmetros	Unidade	LQ	Resultados analíticos
Pireno	mg/kg	0,01	< 0,01
Acenafteno	mg/kg	0,01	< 0,01
Fluoranteno	mg/kg	0,01	< 0,01
Naftaleno	mg/kg	0,01	< 0,01
Benzo(a)antraceno	mg/kg	0,01	< 0,01

CTC - Capacidade de Troca Catiônica

Parâmetros	Unidade	LQ	Resultados analíticos
pH em CaCl2		---	3,8
Matéria Orgânica	g/dm ³	---	25
Fósforo Resina	mg/dm ³	---	8
Potássio	mmol/dm ³		0,5
Cálcio	mmol/dm ³		7
Magnésio	mmol/dm ³		4
H+Al	mmol/dm ³	---	135
Al	mmol/dm ³	---	16
Soma de Bases	mmol/dm ³	---	12
CTC	mmol/dm ³	---	147
Saturação de Bases	V%	---	8
Saturação Al	m%	---	58

Micronutrientes

Parâmetros	Unidade	LQ	Resultados analíticos
Cobre	mg/kg	0,6	19
Ferro	mg/kg	0,6	18418
Zinco	mg/kg	0,6	52
Manganês	mg/kg	0,6	251
Boro	mg/kg	0,6	26

Granulometria

Parâmetros	Unidade	LQ	Resultados analíticos
Argila (< 0,002 mm)	g/kg	0,01	21
Silte (0,053 a 0,002 mm)	g/kg	0,01	345
Areia Total (2,00 a 0,053 mm)	g/kg	0,01	634
Areia Grossa (2,00 a 0,210 mm)	g/kg	0,01	185
Areia Fina (0,210 a 0,053 mm)	g/kg	0,01	449

Macronutrientes - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Nitrogênio Total Kjeldahl	mg/kg	3,1	905
Enxofre	mg/kg	61	365
Fósforo	mg/kg	0,6	388

CONTROLE DE QUALIDADE DO LABORATÓRIO**Controle de Qualidade - Metais - Solo****9560/2011-0 - Branco de Análise - Metais - Solo**

Parâmetros	Unidade	LQ	Resultados analíticos
Prata	mg/kg	0,5	< 0,5
Alumínio	mg/kg	0,5	< 0,5
Arsênio	mg/kg	0,5	< 0,5
Bário	mg/kg	0,5	< 0,5
Berílio	mg/kg	0,5	< 0,5
Bismuto	mg/kg	0,5	< 0,5
Cálcio	mg/kg	25	< 25
Cádmio	mg/kg	0,05	< 0,05
Cobalto	mg/kg	0,5	< 0,5
Cromo	mg/kg	0,5	< 0,5
Cobre	mg/kg	0,5	< 0,5
Ferro	mg/kg	0,5	< 0,5
Potássio	mg/kg	25	< 25
Lítio	mg/kg	0,5	< 0,5
Magnésio	mg/kg	25	< 25
Manganês	mg/kg	0,5	< 0,5
Molibdênio	mg/kg	0,5	< 0,5
Sódio	mg/kg	25	< 25
Níquel	mg/kg	0,5	< 0,5
Chumbo	mg/kg	0,5	< 0,5
Antimônio	mg/kg	0,5	< 0,5
Selênio	mg/kg	0,5	< 0,5
Estanho	mg/kg	0,5	< 0,5
Estrôncio	mg/kg	0,5	< 0,5

9560/2011-0 - Branco de Análise - Metais - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Titânio	mg/kg	0,5	< 0,5
Tálio	mg/kg	0,5	< 0,5
Urânio	mg/kg	0,5	< 0,5
Vanádio	mg/kg	0,5	< 0,5
Zinco	mg/kg	0,5	< 0,5

Ensaio de Recuperação

Parâmetros	Quantidade Adicionada	Unidade	Resultado da Recuperação (%)	Faixa Aceitável de Recuperação (%)
9561/2011-0 - Solo Padrão - RTC - CRM023				
Alumínio	8470	mg/kg	72	70 - 130
Arsênio	380	mg/kg	99	95 - 105
Bário	75,5	mg/kg	86	85 - 115
Cádmio	0,92	mg/kg	114	50 - 150
Cálcio	5420	mg/kg	86	85 - 115
Cromo	31	mg/kg	88	80 - 120
Cobalto	4,68	mg/kg	93	85 - 115
Cobre	8,9	mg/kg	95	80 - 120
Ferro	10700	mg/kg	108	85 - 115
Chumbo	213	mg/kg	103	90 - 110
Magnésio	3060	mg/kg	108	85 - 115
Manganês	206	mg/kg	94	90 - 110
Níquel	11	mg/kg	111	85 - 115
Potássio	2230	mg/kg	79	65 - 135
Selênio	105	mg/kg	107	90 - 110
Tálio	111	mg/kg	104	90 - 110
Vanádio	21,7	mg/kg	77	70 - 130
Zinco	93,8	mg/kg	95	85 - 115

Surrogates
9560/2011-0 - Branco de Análise - Metais - Solo

Ítrio	50	%	99	70-130
-------	----	---	----	--------

9561/2011-0 - Solo Padrão - RTC - CRM023

Ítrio	62,5	%	93	70-130
-------	------	---	----	--------

17657/2011-0 - Solo NC

Ítrio	60,9	%	87	70-130
-------	------	---	----	--------

Controle de Qualidade - TPH Voláteis + BTEX - Solo
19850/2011-0 - Branco de Análise - TPH Voláteis +TEXB - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Benzeno	mg/kg	0,005	< 0,005
Tolueno	mg/kg	0,005	< 0,005
Etilbenzeno	mg/kg	0,005	< 0,005
m,p-Xilenos	mg/kg	0,01	< 0,01
o-Xileno	mg/kg	0,005	< 0,005

Ensaio de Recuperação

Parâmetros	Quantidade Adicionada	Unidade	Resultado da Recuperação (%)	Faixa Aceitável de Recuperação (%)
19851/2011-0 - LCS - TPH Voláteis + BTEX - Solo				
Benzeno	0,05	mg/kg	111	60 -140
Tolueno	0,05	mg/kg	92	60 -140
Etilbenzeno	0,05	mg/kg	97	60 -140
m,p-Xilenos	0,1	mg/kg	97	60 -140
o-Xileno	0,05	mg/kg	99	60 -140
Surrogates				
19850/2011-0 - Branco de Análise - TPH Voláteis +TEXB - Solo				
Tolueno-d8	0,05	%	107	60 -140
p-Bromofluorbenzeno	0,05	%	83	60 -140
19851/2011-0 - LCS - TPH Voláteis + BTEX - Solo				
Tolueno-d8	0,05	%	97	60 -140
p-Bromofluorbenzeno	0,05	%	89	60 -140
17657/2011-0 - Solo NC				
Tolueno-d8	0,05	%	86	60 -140
p-Bromofluorbenzeno	0,05	%	74	60 -140

Controle de Qualidade - PAH - Amostras Sólidas
19914/2011-0 - Branco de Análise - PAH - Solo

Parâmetros	Unidade	LQ	Resultados analíticos
Benzo(a)pireno	mg/kg	0,01	< 0,01
Benzo(b)fluoranteno	mg/kg	0,01	< 0,01
Benzo(k)fluoranteno	mg/kg	0,01	< 0,01
Criseno	mg/kg	0,01	< 0,01
Acenaftileno	mg/kg	0,01	< 0,01

Fluoreno	mg/kg	0,01	< 0,01
Antraceno	mg/kg	0,01	< 0,01
Benzo(g,h,i)perileno	mg/kg	0,01	< 0,01

Página 3 de 5 / B.A.: 17657/2011-0

Bioagri Ambiental - Piracicaba - SP. Rua Aljovil Martini, 201 - Dois Córregos - Piracicaba - SP - CEP: 13420-750 - Fone: (19) 3417.4700 - Fax: (19) 3417.4711 - vendas@bioagriambiental.com.br

APÊNDICE – VERSÃO PRELIMINAR DO ARTIGO PARA PUBLICAÇÃO

Comparative metagenomics to identify aromatic degradation potential in soils

Bruno Karolski^{§1,2}, Elen Aquino Perpetuo², Guilherme Marcello Quiroga Cruz¹, Marie-Anne Van Sluys¹

¹GaTE lab, Departamento de Botânica, Universidade de São Paulo, Rua do Matão, 277, sala 153, cidade universitária, 05508-090, São Paulo, Brasil

²Centro de Capacitação em Pesquisas do Meio Ambiente, Rod. Cônego Domênico Rangoni km 270, Cubatão, São Paulo, Brasil

Email addresses:

BK: brukarolski@gmail.com

EAP: elen@cepema.usp.br

GMQC: guilhermemarcello@yahoo.com.br

MAVS: mavsluys@ib.usp.br

Abstract

Background

Benzene, toluene, ethyl-benzene and xylenes are aromatic hydrocarbon derived from petroleum. Known collectively as BTEX, they are produced and used at a large scale and contamination frequently occurs. Because it is highly soluble in water, when in soil BTEX can spread long distances from the original contamination site, thus affecting large areas. Some microorganisms are known to live in contaminated environments and use contaminants such as BTEX as a unique carbon source for energy production. They catabolize contaminants into less dangerous products or even eliminate them from environment, a feature which has great commercial and environmental interests.

Results

We therefore compared the microbial communities in soils which were affected and un-affected by BTEX contamination. To this end, we used a metagenomics approach and developed a comparison method to identify microorganisms and degradation potential of soils studied. We found qualitative and quantitative differences in microbial structures from three different sites in Cubatão County, one of which is contaminated with BTEX. We constructed a metabolic overview identifying important genes, degradation potential and microorganisms related to BTEX degradation

Conclusions

Through our study, we identified different microorganisms and exclusive groups capable to degrade aromatic hydrocarbons confirming the adaptation of the microbial community to the contamination. We also find sequences with aromatic degradation potential belonged to unknown microorganisms through our conserved degradation domains approach. This method can be expanded to others metabolic functions of interest to evaluate environmental microbial communities *in situ*.

Background

Benzene, toluene, ethyl-benzene and xylenes are collectively known as BTEX. Approximately 10^3 tons of this group of aromatic hydrocarbons is produced per year to produce pesticides, plastics and synthetic fibers [1]. Gasoline is the most important product, where all four aromatic hydrocarbons are found [2]. The high volumes production, the worldwide manipulation and the close contact with human activities increase the risk of environmental contamination. The contamination can occur by leakage from diverse sources, such as storage tanks, transportation pipes, and gas stations. Due to the highly solubility in water BTEX can spread easily into environment, contaminating the soil, water and air [3-6]. The toxicity of BTEX in human beings has been well studied. Exposure to high levels of the different components of BTEX causes various symptoms including skin and eye irritation, drowsiness, dizziness, headaches tremors, confusion, unconsciousness and even death [7-10]. Benzene is the most important contaminant because, apart from its high toxicity, long-term exposure can cause leukemia [8, 11].

Despite the toxicity of BTEX, there are some microorganisms that can use it as metabolic precursors to produce energy [12]. This metabolic capacity is being increasingly examined because of the potential to transform an environmental contaminant into less harmful products or even removing the contaminant from environment, a capacity which has great commercial and environmental interest. Soil is one of the richest environments in diversity and abundance of microbial organisms. Studies suggest that one gram of soil can contain 10^9 cells of more than 2,000 different species [13, 14]. In spite of the high number of microorganisms in the soil, estimates suggest that only 1% are cultivable [14-16], which is a limiting factor in diversity

studies. To overcome this limitation, a state of the art technique was developed to explore microbial communities *in situ* without the need to isolate and cultivate microorganisms. This technique is called metagenomics and made possible identifications, genes recovery and metabolism elucidation from a wide range of different microbial communities [17-21]. With the rise of high throughput sequencing, the metagenomic studies evolve to a powerful identification tool due to the increase number of sequences reaching higher coverage of environmental microbial communities. The first environmental high throughput sequencing study was published in 2004 [22]. For the first time, microbial identification, genome drafts, metabolism and gene comparison were included in a single study, highlighting the potential of a metagenomic study which included high throughput sequencing.

Using this state-of-art metagenomic approach, we compared microbial communities from soil at three sites in Cubatão County in São Paulo State, which was the most polluted place in the world due to high industrial activity. Using this approach, we were able to examine the microbial diversity in soils using genome shotgun sequences and comparisons to prokaryotic genome database to find differences between communities that may be related to BTEX contamination. In addition to examining microbial diversity, we also surveyed, by developing a conserved domains based method, the metabolic potential of the degradation of aromatic hydrocarbons in soil. We were also able to elucidate the taxonomic structure, at the genus, of the microbial community associated with aromatic hydrocarbon degradation activity level.

Results

Soils analysis and sequencing

Soils analysis and sequencing

The presence of BTEX was tested for in soil taken from sites C and S1. Toluene and xylenes were identified in site C (Fig.1). No aromatic hydrocarbon contamination was identified at Site S1.

A total of 1,710,216 sequences was generated from all sites (table 1). After a comparison against the NCBI prokaryotic genomes using the chosen cut-off values, 12,339 sequences were identified and used for further taxonomic classification.

Microorganism identification

In order to identify microorganisms from the metagenomic sequences, we used the sequences as queries against the NCBI prokaryotic genome database. A total of 11 phyla were identified from all 3 sites (Fig. 2). In terms of number of phyla, site C was the richest site with all phyla present and site S1 had higher relative abundance of sequences identified for the majority. Some sequences, 3% for site C and 6% for sites S1 and S2, could not be identified even in phyla level were therefore grouped in the Bacteria node. Proteobacteria, Acidobacteria and Actinobacteria were the most abundant phyla identified and therefore further analysed.

Twenty-two%, 12% and 9% of the sequences from sites C, S1 and S2 respectively were identified as from microorganisms of the phylum Acidobacteria. Despite the large number of sequences, only microorganisms from the genera *Acidobacterium* and *Candidatus Koribacter* were identified, suggesting a very low diversity of sequences from this phylum.

Four%, 24% and 9% of the sequences from sites C, S1 and S2 respectively were identified as from microorganisms of the phylum Actinobacteria. Further analysis identified a total of 13 families from all the three sites (Fig. 3). Site C had the lowest

abundance and diversity with only five families identified. Within site C belonged the highest abundance was found in the family Micobacteriaceae, 2.8% of these sequences were classified in the *Mycobacterium* genus. Although the number of sequences identified for family Micrococcineae (0,01%) is low, site C was the only site at which this family was identified. All known Actinobacteria families were identified at sites S1 and S2, site S1 was especially rich in sequences from this phylum.

The phylum Proteobacteria was the most abundant phylum identified, with 64%, 46% and 67% of sequences identified from sites C, S1 and S2 respectively. Figure 4 shows the classification in family level for this phylum. A total of 11 families were identified from the Class Alphaproteobacteria (figure 4 A) for all metagenomes. For almost all families, site S1 had a higher relative abundance, except for the families Bradyrhizobiaceae and Phyllobacteriaceae, which were more abundant at site C. For the family Bradyrhizobiaceae alone 2,546 sequences identified for site C, corresponding to 35% of all sequences identified for this site. At the genus level sequences from the following genera were identified: *Bradyrhizobium*, *Nitrobacter*, *Oligotropha* and *Rhodopseudomonas*. The largest number of sequences were from the *Bradyrhizobium* genus, with 2,073 sequences, representing 28% of all sequences for site C. 19% of these were related to *Bradyrhizobium japonicum*.

Sequences from eight families of the Class Gammaproteobacteria were identified. Site C showed the higher diversity for this family (Fig. 4b). Families Chromatiaceae, Methylococcaceae and Enterobacteriaceae, this last family was the most abundant group for the site with 34% of all identified sequences. Enterobacteriaceae family sequences fell in three genera, *Escherichia*, *Klebsiella* and *Salmonella*. The largest number of sequences, 19%, was from the Genus *Klebsiella* at Site C.

4%, 3% and 2% of sequences, from sites C, S1 and S2 respectively, were identified as belonging to Class Betaproteobacteria (figure 4C). Although found in a low percentage of sequences, seven families were identified from this class, four of which were found exclusively at site C. Similarly, seven families were identified from class Deltaproteobacteria, of which three were exclusively from site C. 4%, 4% and 3% of total sequences were Identified for sites C, S1 and S2 respectively.

METABOLIC IDENTIFICATION

Metabolic identification through KEEG's metabolic pathways

For an overall analysis of the xylene and toluene degradation potential of the three metagenomes, we first compared the metagenomic sequences from all sites with sequences associated with xylene and toluene degradation pathways from the UNIPROT and KEGG databases. Genes associated with xylene and toluene degradation pathways were identified from all 3 metagenomes analysed (see additional files). In general, site S1 contained the lowest diversity of genes associated with the degradation pathways, while sites C and S2 contained almost the same genes, which were identified from throughout the pathways. Using this approach we could not identify many differences regarding toluene and xylenes potential between the sites studied. This first analysis showed that genes associated with degradation were present at all 3sites, suggesting na intrinsic degradation potential.

Initial aromatic ring attack

In order to compare the potential of the 3 sites at the initial stages of xylene and toluene degradation, the operon sequences associated with the first steps were retrieved from NCBI database and were compared with the metagenomic sequences (figure 5). Site C had a higher diversity of toluene and xylenes degradation operons compared to sites S1 and S2. Of a total of eleven operons identified, eight were found at site C. For sites S1

and S2, on the other hand, only two and three operons were identified, respectively. In terms of the total number of sequences related to degradation operons, only three sequence were identified from site S1 whereas 53 and 56 sequences were identified from sites C and S2 respectively. These results suggests that microorganisms at sites S2 and C may have a higher potential to catabolize xylenes and toluene at the initial steps.

BTEX degradation PFAM domains

In order to further our understanding of sequences related to aromatic ring modification, the concept of functional conserved regions (PFAM) was used. The operons selected contained eighteen unique PFAM domains of which six were chosen for comparison because of their aromatic ring modification capacity. The domains were compared against the metagenomes from sites C, S1 and S2, as well as 4 other metagenomes, Waseca, Refinery, Ice and VCU, from the NCBI database. All metagenomes included sequences related to the degradation domains identified except for the VCU metagenome for which no sequences could be found (figure 6). Figure 6 shows that site C sample contained the largest number of sequences. A total of 375 sequences related to PFAM domains associated with degradation processes were found for this site. Site S2 had the second highest number of sequences, 123. Alpha and beta subunit ring hydroxylation were the most highly represented domains in all metagenomes examined, except for VCU metagenome. Methane, phenol, toluene hydroxylase degrading domain was identified in almost all metagenomes. The richest metagenome for this domain were Refinery and site C from this study. The refinery metagenome had the highest number of different domains of all metagenomes studied, with five of six domains, in addition, it was the metagenome with the phenol hydroxylase domain. Site C had the largest number of sequences related to degradation

domains, which suggests that at Site C there is a microbial community adapted to aromatic ring catabolism.

Taxonomic identification of metabolic relevant sequences

To identify the microorganisms which may be adapted to aromatic ring degradation, sequences from the 3 sites related to the degradation domains identified were used as queries against the NCBI prokaryotic genomes (figure 7). A total of 375, 85 and 123 sequences were found for sites C, S1 and S2 respectively. Site C had the largest number of sequences and it was also the most diverse in genus identified with aromatic degradation capacity. Of the 19 genera identified from all three metagenomes, 18 were found at site C. The only genus not found at site C was *Klebsiella*, which was found exclusively at site S2. The node that gathered the majority of sequences was No hits node with 287, 71 and 74 sequences for sites C, S1 and S2 respectively (Fig. 6). Sequences at this node showed no identity to any genome from the NCBI database. We were able to identify microorganisms with the potential for aromatic ring degradation potential at all three sites analysed.

Discussion

BTEX soil analysis

Although the soil analyses were designed to identify all BTEX components, only xylenes and toluene were identified in sample of Site C. Site C, known to be a contaminated site, was initially contaminated in 2008. We were unable to find any more information about the evolution of contamination plume between the contamination event in 2008 and our sampling in 2010. The BTEX analysis was performed to confirm that the site remained contaminated checking aromatic hydrocarbons presence in soil at the time of sampling. We could identify traces of

toluene and xylenes contaminating site C indicating the contamination in the site and the contact with the microbial community. The sequences used for taxonomic identification were selected using cut-off values. They were compared with prokaryotic genome databases in order to identify the microorganisms present in the soil samples and determine if there were any differences between the sites. Groups of microorganisms related to contamination were identified and the structure of the microbial community in contaminated soil was elucidated.

Taxonomic identification

Our work aimed to compare three sites with varying degrees of contamination, by using DNA sequencing from samples obtained directly from the soil. Differences in microbial composition was identified in the three sites. The Acidobacterium phylum is the second most abundant group and is present in diverse habitats and apparently in high numbers [23, 24]. Although the close relation of this group with non-contaminated sites, it was found also in contaminated sites and has representatives that code for dioxygenases that can catabolize aromatic rings [25, 26]. Sequences from the genus *Koribacter* was found in high abundance at site C. Microorganisms from this genus are not known to degrade aromatic substances, but our findings suggests that *Koribacter* microorganisms may have adapted to the presence of contaminants.

Sequences from the phylum Actinobacteria had higher diversity and abundance for sites S1 and S2, the non-contaminated sites. Despite this, we also identified two Actinobacteria families that was most abundant at site C. Sequences from the genus *Mycobacterium* of the Actinobacteria family, Micobacteriaceae, was the most abundant at site C. This genus is known to live in contaminated sites and to degrade BTEX as well as more complex aromatic substances [27]. Sequences from the Family

Micrococcineae, were found in lower numbers but exclusively at site C. There is no known relationship between this family and BTEX or other types of contaminants and no reports of the degradation capacity of this family. Despite this, our findings suggests an adaptive advantage of this family to the environment at site C and it may be a family with an previously unidentified potential to degrade aromatic hydrocarbons.

The largest number of sequences overall was identified from the Phylum Proteobacteria. The Alphaproteobacteria family (Phylum Proteobacteria) were more abundant at site S1 and is known to be found/common in? non-contaminated sites [28, 29]. Our results confirms these findings, but we also identified families at site C, among them family Phylobacteriaceae. The higher abundance suggests an adaptation to contamination at site C. Within the family Phylobacteriaceae, the genus *Parvibaculum* have been reported to be able to deal with PCBs (polychlorinated biphenyl), coding for oxygenases that PCBS. Because oxygenases are known to be promiscuous in terms of their substrates, this genus may also use the xylenes and toluene at site C to produce energy, as a way to to survive and thrive in this site [30, 31]. Without a doubt, the most important result was the finding of the sequences related to the family Bradyrhizobiaceae, of which the genus *Bradyrhizobium* was the most highly represented. This genus is found in the soil [32]. Although this genus is not known to directly degrade aromatic compounds, it has been found in contaminated sites [4]. More recently, a study on the evolution of a microbial community in contaminated soil showed that there was an increase in the number of sequences identified as *Bradyrhizobium* [33].

Within the class Gammaproteobacteria sequences from the families Methylococcaceae and Chromatiaceae were found only in site C. Microorganisms from the family Methylococcaceae are known to carry genes that code for unspecific monooxygenases

that can degrade some types of aromatic rings [34] while microorganisms from the family Chromatiacea are not known for their capacity to degrade aromatic substances. The fact that sequences from both families were found exclusively at site C suggests a possible relation with aromatic contaminants. Site S2, an uncontaminated site, had a great abundance of sequences identified from the Enterobacteriaceae family. Among the genera identified were *Escherichia*, *Salmonella* and *Klebsiella*, microorganisms from which are known to cause diseases in humans [35-37]. These genera are already known to live in soils and even associated with plants [38, 39]. Sequences from the Enterobacteriaceae family were found exclusively at site S2, they were found in high numbers, the highest number of sequences for S2, and also one of the highest numbers for any group. The sampling site was close to a poor community with no adequate sanitary treatment. Contaminations due to leaks from residential wastes may be contaminating the soil, changing the microbial community. This could explain the high amount of sequences identified from Enterobacteriaceae family in our results. This contamination of soil, besides creating diseases in the inhabitants, could also create highly resistant pathogenic microorganisms and diverse genotypes [37, 40, 41].

In general, sequences from class Betaproteobacteria were found in higher relative abundance at site C. Within this class, sequences from the genus *Burkholderia*, family Burkholderiales, were the most interesting. Microorganisms from this genus are found in soils and are known for their ability to degrade aromatic substances [42, 43]. We could not identify significant differences in abundance between contaminated and non-contaminated sites, apparently therefore this genus is wide spread in the region and due to its ability to catabolize aromatic substances, it can compose the degradation potential of soils in region. Several families were found exclusively, or in high abundance at Site C. These were families Hydrogeniphilaceae, Neisseraceae and

Rhodocyclaceae (class x) and families Desulfovibrionaceae, Geobacteriaceae, Myxococcaceae and Syntrophobacteriaceae (class Deltaproteobacteria). Microorganisms from these families have not been reported to degrade aromatic substances, but they have been found in contaminated sites in association with microorganisms known to degrade aromatic substances [3, 44-48].

Metabolic identification of toluene and xylenes degradation through sequence comparison

Our experimental design allowed us to identify genes related to aromatic degradation by comparing metagenomic sequences with several databases. Our soil analysis showed that xylenes and toluene were present; our study therefore began with the identification of genes associated with the degradation pathways of these aromatic substances. It has been previously shown that environments that have not been known to be contaminated have microorganisms with degradation potential [31, 49]. Our study was consistent with these results. Based on metabolic pathways on the KEGG database, we were able to identify genes associated with different stages of degradation (see additional files). The most important enzymes involved in aromatic degradation is the oxygenases. These enzymes are responsible for initial steps of aerobic degradation, breaking the molecular structure of aromatic rings and making them available for oxidation and energy production [50]. To evaluate the capacity and diversity of initial degradation steps of aromatic ring degradation of soils, operons coding for oxygenases were compared with metagenomes. Our results showed that several operons associated with the degradation process are present in all soils studied. The highest diversity of degradation operons was found in site C, which is the site contaminated with toluene and xylenes. This diversity at site C indicates a microbial community more adapted to aromatic ring contaminants with an abundance of species that may be able to directly catabolise toluene and xylenes. In terms of the sequences identified, sites C and S2 had

more sequences related to degradation operons than site S1. About 35% of the metagenomic sequences related to operons at sites S2 and C had high identity to plasmid pWWO. This plasmid was first identified in *Pseudomonas putida* and carries the genes *xyIA* and *xyIM*, responsible for the degradation of toluene and xylenes [51]. This plasmid is known to be shared by conjugation between soil microorganisms, conferring the ability to catabolize aromatic rings [52, 53]. The high pollution levels of aromatic hydrocarbons in Cubatão County, the finding of high number of sequences related to pWWO plasmid, and the presence of the plasmid in several species known to survive aromatic hydrocarbon contamination, we speculate that the plasmid is being transferred between microorganisms through the region. Although the S2 site is not known to be contaminated with aromatic hydrocarbons, it is near industrial and contaminated areas and BTEX residues could indirectly affect the site studied, explaining the pWWO plasmid sequences found at this site. In addition, at Site 2 high numbers of sequences from the genera *Escherichia*, *Klebsiella* and *Salmonella* were found. Microorganisms from these genera carry plasmids, are able to conjugate and could be both a receptor and a reservoir of the pWWO plasmid [54, 55]. This could explain the higher percentage of this plasmid found in site S2. The most abundant sequences found for site C, as site S2, was related to pWWO plasmid. This site is contaminated and the presence of degradation genes was expected. What was not expected was the massive presence of sequences from the genus *Bradyrhizobium*, which has no proven aromatic hydrocarbon degradation capacities. However, microorganisms from this genus have been found in increased abundance in soils contaminated with BTEX [4, 33] and it has been previously shown they are able to conjugate and incorporate the plasmid pWWO, but are unable to maintain the plasmid within the cell *in vitro* conditions [56]. We hypothesize that the contamination of the

site has exerted strong selective pressure on local microorganisms and that surviving *Bradyrhizobium* species at site C may be able to maintain the pWWO plasmid..

Study of conserved domains indicating BTEX degradation potential

In an attempt to deepen our study of the BTEX degradation potential in soils, we compared the PFAM domains present in genes composing the operons. We identified 18 PFAM domains, six of which are known to directly modify aromatic rings found in hydrocarbons. These domains are responsible for attacking the aromatic ring and are the active sites of the oxygenase enzymes. Since these domains are shared between operons that degraded different kinds of aromatic ring hydrocarbons and are conserved among these operons, a study focused on the conserved PFAM domains was developed to identify and compare the degradation potential of different soils studied. The conserved domains comparison through HMMs take into consideration the positioning of amino acids in the sequencing and not a simple word comparison like BLAST. This feature could identify sequences with degrading potential even in sequences not present in database deepening the degradation potential identification. The study revealed conserved domains in the three sites studied following the previous results found and the site with higher sequences identified was site C also corroborating our previous results. This method revealed three times more conserved domains compared to site S1, the second in degradation domains abundance. The two most abundant domains found was the ring hydroxylation subunits alpha and beta, the two subunits of dioxygenases which are known to degrade several types of aromatic rings [57]. The higher number of these sequences for site C indicates a microbial community adapted to aromatic hydrocarbons contaminants. To support our analysis, the method developed to identify aromatic hydrocarbon degradation potential was used with four metagenomes from public databases. Our approach identified degradation potential in

all the metagenomes except for the VCU metagenome, which worked as our negative control. The VCU contains genetic material from microorganisms living in human beings females' vagina. This microbial community is a very specific and specialized group of microorganisms that have never been in contact with any kind of aromatic hydrocarbon. Due to the habitat that these microorganisms live, the presence of degradation domains were not expected. The absence of degradation domains [58] showed for this metagenome means that our method can identify specific domains responsible only for aromatic hydrocarbon degradation, which proved the metabolic identification capacity by conserved degradation domains.

The Ice metagenome [59] had the lowest abundance of degradation domains. This metagenome comes from a glacier locate in Germany and no evidence of aromatic hydrocarbon contamination was reported by the authors. Assuming the contamination free for the glacier, the domains found indicate the intrinsic degradation capacity already reported for uncontaminated sites. This degradation potential was reported in frozen soil studies when contamination occurs. Some kind of resistant and degradation active microorganisms can survive promoting an intrinsic bioremediation [58]. Refinery metagenome also presented low numbers of aromatic degradation domains. Although the refinery metagenome had contact with aromatic hydrocarbons, this metagenome belongs to a microbial community from a contaminated wastewater and due to the water environment the oxygen can be a limiting factor making difficult the uptake of oxygen and the development of aerobic aromatic hydrocarbon degradation microorganisms [60]. The wastewater was contaminated with several kinds of aromatic hydrocarbons and the higher diversity of domains found for this metagenome indicate several kinds of metabolisms to deal with different contaminants. In uncontaminated soils, we also could find degradation domains. Although we cannot say if the numbers

found are high or low, we know that our metagenomes S1 and S2 are situated in a high-polluted area and contamination from other places could affected the sites studied increasing the number of microorganisms capable to deal with aromatic substances increasing the basal degradation potential. Similar reasoning can be applied for Waseca metagenoma. This metagenome has its source in the upper layer of agriculture soil [61]. This kind of soil many times is affected by pesticides, which have aromatic rings in its composition and can be degraded by microorganisms[62]. Put together these considerations, we believe that degradation domains found for soils studied can be higher than a regular soil due to more contact to aromatic xenobiotics.

Identification of sequences related to degradation domains.

Sequences with high similarity to PFAM conserved degradation domains were compared with prokaryotes genomes from the NCBI database. This was in order to identify the genera containing degradation domains, and to identify a microbial community structure with aromatic degradation potential. When used as a query in a BLAST against the NCBI database, the majority of sequences did not show any identity to any sequences and were grouped in No hits node in the taxonomy structure (Fig. 6). This result suggests that the majority of sequences identified belonged to unknown microorganisms but had degradation potential. For those sequences that were identified in the database a taxonomy regarding the degradation potential was constructed. The genera identified as degradation domain holders were checked in literature regarding degradation activity to confirm ours results (Table 2). Through this confirmation, we could measure the accuracy of the degradation potential. We could identify 20 genera comparing sequences with database which 10 had its aromatic degradation proved, 7 were found in contamination sites marked as potential degrader and 3 were not related to any contamination. These 3 genera holds the degradation domains and could be

candidates to a deep investigation regarding aromatic degradation. The search for degradation activity in genera identified using conserved domains showed a high accuracy of the method. The major genera found was linked with aromatic xenobiotics what endorse the others unidentified sequences to belong to microorganisms capable to degrade aromatic hydrocarbons.

Conclusions

The direct DNA sequencing evaluation done for three soils located at one of the most polluted region in the word was able to point differences both in microorganisms' diversity and in metabolism. Different genera and relative abundance showed for soils studied indicated microorganisms capable to deal with aromatic contamination mainly in site C, which was contaminated with toluene and xylenes. We could identify exclusive groups capable to degrade aromatic hydrocarbons and confirm the adaptation of the microbial community to the contamination developing the study *in situ*.

Due to our sequencing approach, we could evaluate the metabolism of toluene and xylenes identifying the genes related to each degradation pathway. We showed an intrinsic degradation potential for all sites studied and a higher diversity for site C, which had the contamination history. In an attempt to explore deeply the presence of degradation capacities of soils microbial community, we developed a study based on PFAM conserved domains and HMMs search. However, through this method, we could find sequences with aromatic degradation potential even if the sequences belonged to unknown microorganisms. We could disengage the simple sequence comparison to identify degradation metabolism with the search for active conserved domains, what resulted in more sequences identified and a better resolution of the degradation potential when compared with different soils. Furthermore, we were able to identify the sequences in prokaryotic database suggesting a taxonomic structure with

the identification of microorganisms with potential and proved degradation activity. This result showed the accuracy of the method developed to identify microorganisms related to degradation, which supports the relation of unidentified sequences with unknown microorganisms with aromatic degradation potential. Although this method cannot identify unknown microorganisms, it proved to be a good method to evaluate the degradation potential of aromatic hydrocarbons focusing environmental microbial communities what can be expanded to others metabolic functions of interest in an *in situ* approach.

Methods

Sampling and DNA extraction

Samples were taken from an area around Cubatão city in the São Paulo state of Brazil at three points; S1 (GPS location 23°54'14.2''S; 46°25'47.1''W), S2 (GPS location 23°50'48.3'' S; 46°24'57.1''W) and C (GPS location 23°51'57.0''S; 46°26'15.1''W). Approximately 100 grams of soil was collected at each site and immediately frozen in liquid nitrogen. Samples were then transferred to -80°C until used. Total DNA was extracted from aliquots of the frozen samples (0.5 to 2.0 grams) using the E.Z.N.A. Soil DNA kit (Omega Bio-Tek) according to the manufacturer's manual. After extraction, sample concentration was measured in a NanoDrop ND-1000 spectrophotometer (Thermo Scientific Inc.) and sample integrity visualized by 0.8% agarose gel electrophoresis. The extracted DNA was allocated and stored at -80°C until sequencing.

Soil analysis

To assess contamination, 2 grams from sites S1 and C were transferred to vials and pre-extracted using aromatic hydrocarbon and 300 rpm orbital agitation at 60°C for 2

minutes, followed by alternating 10 seconds of agitation and 4 seconds resting. Samples were extracted from the air phase using a Black 75 μm Carboxen/PMDS fiber and 250-rpm orbital agitation for 10 minutes at 60°C. The sample was desorbed in a gas chromatographer (Varian 450 GC) coupled with a mass spectrophotometer (Varian 240 S) ion trap detector. The sample was loaded, left standing for 90 seconds and then processed at 60°C for 1 minute. The temperature was then increased 5°C per minute until 150°C and kept at this temperature for 21 minutes. Finally, the temperature was increased at 10°C intervals until it reached 250°C and then kept at this temperature for 2 minutes. Helium was used as the carrier gas at 1 ml per minute flow rate and a VF 5MS column was used for the sample analysis.

Sequencing and microbial identification

Sequencing was performed using the GS FLX (Roche) technology. DNA preparation and library construction were carried out according to manufacturer's instruction. Each micro titer plate was divided in two gaskets and one gasket was used for each sample. Duplicated sequences were removed through the replicates program [63]. Sequences sharing more than 5 identical initial bases with other sequence and greater than 95% identity were considered replicates and clustered. The longest sequence from each cluster was used for classification. We compared unique sequences against prokaryotic genomes from NCBI [64] using the BLASTN [65] program. against prokaryotic genomes. Sequences used for forward analysis presented an e-value 1e-10 or less, 300 bp minimum length and at least 80% identity. Sequences were identified in prokaryotes genomes database of NCBI with an e-value 1e-10 or less, 300 bp minimum length and at least 80% identity. The BLASTN output including the selected sequences were classified using the MEGAN [66] program with default parameters.

Metabolic identification of BTEX degradation

Functional identification was based on the Uniprot/Swissprot [67] and KEGG database [68]. All sequences with related to the xylene and toluene degradation pathways, based on EC numbers, were retrieved from the Uniprot/Swissprot database. Uniprot/Swissprot protein sequences were compared to our conceptually translated unique sequences using the BLASTP program [65]. Sequences with greater than 60% similarity to Uniprot/Swissprot sequences related to xylene and toluene degradation pathways were also identified in the KEGG metabolic pathway database.

Initial toluene and xylenes degradation steps analysis was based on the BIONEMO database. Operons associated with the beginning of each degradation pathway were found in BIONEMO database [69]. The corresponding sequence of all the operons associated with the initial steps of toluene and xylenes degradation were retrieved from the NCBI database. The identification numbers of sequences used for this analysis are AB024741.1; AJ344068.1; L40033.1; U04052.1; AF019635; U15298.1; M65106.1; J04996.1; GQ884177.1; AF319657.1; AJ005663.3 e AJ012090.1. These operons were compared against our metagenomes through BLASTN and metagenomic sequences with at least 80% identity, 100 amino acids length an e-value equal or below $1e-10$ were used for identification.

In order to identify conserved domains associated with BTEX degradation, fourteen operons associated with BTEX degradation were identified in the BIONEMO database. All operons previously identified as well as operons AB120956.1 and AB120955.1 with ethyl-benzene degradation capacity were used. Eighteen unique Pfam [70] domains were found in these operons of which six were chosen for comparison against metagenomes because they are associated with the modification of aromatic rings which are found in BTEX compounds. The Pfam domains used were phenol hydroxylase subunit (PF06099); methane, phenol, toluene hydroxylase (PF02332); phenol

hydroxylase conserved region (PF04663); toluene-4-monooxygenase system protein B (TmoB) (PF06234); ring hydroxylating alpha subunit (catalytic domain) (PF00848) and ring hydroxylating beta subunit (PF00866). Each Pfam domain was compared against the translated metagenomes through HMMs search using HMMER [71] and sequences with an e-value equal or lower than $1e-10$ were count. The comparison included metagenomic sequences from sites C, S1 and S2 as well as an additional four metagenomes. They were Waseca (AAFX00000000); Refinery (SRX038779; Ice (SRX000607); and VCU (SRR49666). These metagenomes were chosen because they had produced using a similar approach to ours and were sequenced using the GS FLX platform, like our sequences.

Sequences with an evalue equal or lower than $1e-10$ similarity with the domains chosen were used for further taxonomic classification. They were used as queries in a BLASTN against the prokaryotic sequence database at NCBI. Sequences with an e-value equal or less than $1e-10$ and at least 50 amino acids long were used for taxonomic identification of sequences using the software MEGAN [66].

Authors' contributions

Please see sample text in the instructions for authors.

Acknowledgements

Text for this section.

References

1. Jindrova E, Chocova M, Demnerova K, Brenner V: Bacterial aerobic degradation of benzene, toluene, ethylbenzene and xylene. *Folia microbiologica* 2002, 47(2):83-93.
2. Weelink SAB, van Eekert MHA, Stams AJM: Degradation of BTEX by anaerobic bacteria: physiology and application. *Rev Environ Sci Biotechnol* 2010, 9:359-385.
3. Alfreider A, Voght C: Bacterial Diversity and Aerobic Biodegradation Potential in a BTEX-Contaminated Aquifer. *Water Air Soil Pollut* 2007, 183:415-426.
4. Cavalca L, Dell'Amico E, Andreoni V: Intrinsic bioremediability of an aromatic hydrocarbon-polluted groundwater: diversity of bacterial population and toluene monooxygenase genes. *Applied microbiology and biotechnology* 2004, 64(4):576-587.
5. Baldwin BR, Nakatsu CH, Nebe J, Wickham GS, Parks C, Nies L: Enumeration of aromatic oxygenase genes to evaluate biodegradation during multi-phase extraction at a gasoline-contaminated site. *Journal of hazardous materials* 2009, 163(2-3):524-530.
6. Mazzeo DE, Levy CE, de Angelis Dde F, Marin-Morales MA: BTEX biodegradation by bacteria from effluents of petroleum refinery. *The Science of the total environment* 2010, 408(20):4334-4340.
7. (ATSDR) AfTSAaDR: Toxicological Profile for Ethylbenzene (Draft for Public Comment). In. Edited by Services USDoPHaH. Atlanta, GA: Public Health Service; 2007.
8. (ATSDR) AfTSAaDR: Toxicological Profile for Benzene (Update). In. Edited by Services USDoPHaH. Atlanta, GA: Public Health Service; 2007.
9. (ATSDR) AfTSAaDR: Toxicological Profile for Toluene. In. Edited by Services USDoHaH. Atlanta, GA: Public Health Service; 2007.
10. (ATSDR) AfTSAaDR: Toxicological Profile for Xylene (Draft for Public Comment). In. Edited by Services DoPHaH. Atlanta, GA: Public Health Service; 2005.
11. Travis LB, Li CY, Zhang ZN, Li DG, Yin SN, Chow WH, Li GL, Dosemeci M, Blot W, Fraumeni JF, Jr. *et al*: Hematopoietic malignancies and related disorders among benzene-exposed workers in China. *Leukemia & lymphoma* 1994, 14(1-2):91-102.
12. Genovese M, Denaro R, Cappello S, Di Marco G, La Spada G, Giuliano L, Genovese L, Yakimov MM: Bioremediation of benzene, toluene,

- ethylbenzene, xylenes-contaminated soil: a biopile pilot experiment. *Journal of applied microbiology* 2008, 105(5):1694-1702.
13. Daniel R: The metagenomics of soil. *Nat Rev Microbiol* 2005, 3(6):470-478.
 14. Schloss PD, Handelsman J: Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 2003, 14(3):303-310.
 15. Streit WR, Schmitz RA: Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol* 2004, 7(5):492-498.
 16. Rajendhran J, Gunasekaran P: Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv* 2008, 26(6):576-590.
 17. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: Metagenomic analysis of the human distal gut microbiome. *Science* 2006, 312(5778):1355-1359.
 18. Hendrickx B, Junca H, Vosahlova J, Lindner A, Ruegg I, Bucheli-Witschel M, Faber F, Egli T, Mau M, Schlomann M *et al*: Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *J Microbiol Methods* 2006, 64(2):250-265.
 19. Bohorquez LC, Delgado-Serrano L, Lopez G, Osorio-Forero C, Klepac-Ceraj V, Kolter R, Junca H, Baena S, Zambrano MM: In-depth Characterization via Complementing Culture-Independent Approaches of the Microbial Community in an Acidic Hot Spring of the Colombian Andes. *Microb Ecol* 2012, 63(1):103-115.
 20. Bruce T, Martinez IB, Maia Neto O, Vicente AC, Kruger RH, Thompson FL: Bacterial community diversity in the Brazilian Atlantic forest soils. *Microb Ecol* 2010, 60(4):840-849.
 21. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T *et al*: Metagenomic discovery of biomass-

- degrading genes and genomes from cow rumen. *Science* 2011, 331(6016):463-467.
22. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W *et al*: Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004, 304(5667):66-74.
 23. Lee SH, Ka JO, Cho JC: Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. *FEMS microbiology letters* 2008, 285(2):263-269.
 24. Kielak A, Pijl AS, van Veen JA, Kowalchuk GA: Phylogenetic diversity of Acidobacteria in a former agricultural soil. *The ISME journal* 2009, 3(3):378-382.
 25. Smith AM, Kirisits MJ, Reible DD: Assessment of potential anaerobic biotransformation of organic pollutants in sediment caps. *New biotechnology* 2012, 30(1):80-87.
 26. Perez-Pantoja D, Donoso R, Junca H, González B, Pieper D: Phylogenomics of Aerobic Bacterial Degradation of Aromatics. In: *Handbook of Hydrocarbon and Lipid Microbiology*. Edited by Timis KN: Springer-Verlag GmbH; 2010.
 27. Xin BP, Wu CH, Lin CW: Bioaugmented remediation of high concentration BTEX-contaminated groundwater by permeable reactive barrier with immobilized bead. *Journal of hazardous materials* 2013, 244-245:765-772.
 28. Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC: Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PloS one* 2008, 3(6):e2527.
 29. Bruce T, Martinez IB, Maia Neto O, Vicente AC, Kruger RH, Thompson FL: Bacterial community diversity in the Brazilian Atlantic forest soils. *Microbial ecology* 2010, 60(4):840-849.
 30. Franklin FC, Bagdasarian M, Bagdasarian MM, Timmis KN: Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. *Proceedings of the National Academy of Sciences of the United States of America* 1981, 78(12):7458-7462.
 31. Macedo AJ, Timmis KN, Abraham WR: Widespread capacity to metabolize polychlorinated biphenyls by diverse microbial communities

- in soils with no significant exposure to PCB contamination. *Environmental microbiology* 2007, 9(8):1890-1897.
32. Denarie J, Debelle F, Rosenberg C: Signaling and host range variation in nodulation. *Annual review of microbiology* 1992, 46:497-531.
 33. Singleton DR, Jones MD, Richardson SD, Aitken MD: Pyrosequence analyses of bacterial communities during simulated in situ bioremediation of polycyclic aromatic hydrocarbon-contaminated soil. *Applied microbiology and biotechnology* 2012.
 34. Shukla AK, Upadhyay SN, Dubey SK: Current trends in trichloroethylene biodegradation: a review. *Critical reviews in biotechnology* 2012.
 35. Sack RB: Human diarrheal disease caused by enterotoxigenic *Escherichia coli*. *Annual review of microbiology* 1975, 29:333-353.
 36. Glynn MK, Bopp C, Dewitt W, Dabney P, Mokhtar M, Angulo FJ: Emergence of multidrug-resistant *Salmonella enterica* serotype typhimurium DT104 infections in the United States. *The New England journal of medicine* 1998, 338(19):1333-1338.
 37. Shon AS, Bajwa RP, Russo TA: Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: A new and dangerous breed. *Virulence* 2013, 4(2).
 38. Berg G, Eberl L, Hartmann A: The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environmental microbiology* 2005, 7(11):1673-1685.
 39. Holden N, Pritchard L, Toth I: Colonization outwith the colon: plants as an alternative environmental reservoir for human pathogenic enterobacteria. *FEMS microbiology reviews* 2009, 33(4):689-703.
 40. Knothe H, Shah P, Krcmery V, Antal M, Mitsuhashi S: Transferable resistance to cefotaxime, ceftazidime, ceftiofur and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. *Infection* 1983, 11(6):315-317.
 41. Patterson JE, Hardin TC, Kelly CA, Garcia RC, Jorgensen JH: Association of antibiotic utilization measures and control of multiple-drug resistance in *Klebsiella pneumoniae*. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2000, 21(7):455-458.
 42. Shields MS, Reagin MJ, Gerger RR, Campbell R, Somerville C: TOM, a new aromatic degradative plasmid from *Burkholderia* (*Pseudomonas*)

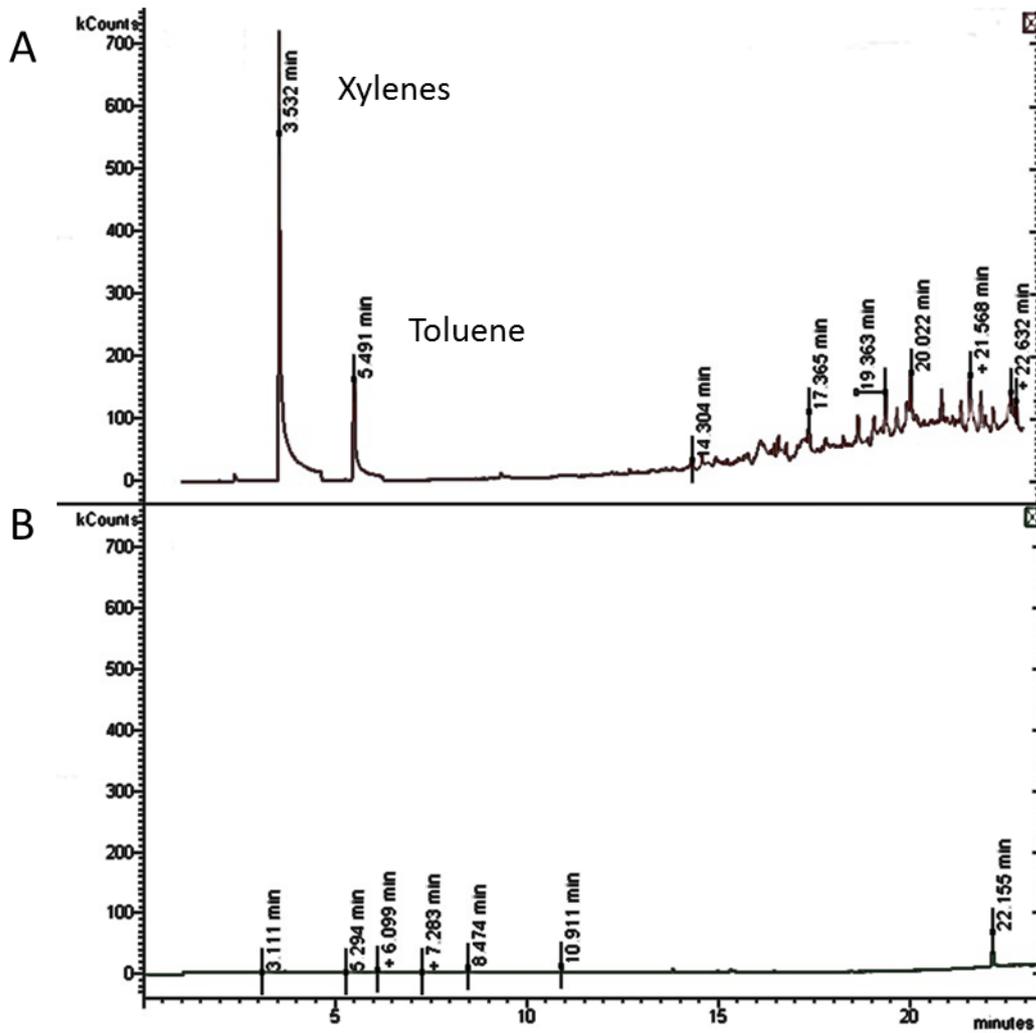
- cepacia G4. *Applied and environmental microbiology* 1995, 61(4):1352-1356.
43. Salles JF, van Veen JA, van Elsas JD: Multivariate analyses of Burkholderia species in soil: effect of crop and land use history. *Applied and environmental microbiology* 2004, 70(7):4012-4020.
 44. Larentis M, Hoermann K, Lueders T: Fine-scale degrader community profiling over an aerobic/anaerobic redox gradient in a toluene-contaminated aquifer. *Environmental Microbiology Reports* 2012:1-10.
 45. Tancsics A, Szoboszlay S, Kriszt B, Kukolya J, Baka E, Marialigeti K, Revesz S: Applicability of the functional gene catechol 1,2-dioxygenase as a biomarker in the detection of BTEX-degrading Rhodococcus species. *Journal of applied microbiology* 2008, 105(4):1026-1033.
 46. Alfreider A, Vogt C: Genetic evidence for bacterial chemolithoautotrophy based on the reductive tricarboxylic acid cycle in groundwater systems. *Microbes and environments / JSME* 2012, 27(2):209-214.
 47. Botton S, van Harmelen M, Braster M, Parsons JR, Roling WF: Dominance of Geobacteraceae in BTX-degrading enrichments from an iron-reducing aquifer. *FEMS microbiology ecology* 2007, 62(1):118-130.
 48. Acosta-Gonzalez A, Rossello-Mora R, Marques S: Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the Prestige oil spill. *Environmental microbiology* 2013, 15(1):77-92.
 49. Xie S, Sun W, Luo C, Cupples AM: Novel aerobic benzene degrading microorganisms identified in three soils by stable isotope probing. *Biodegradation* 2011, 22(1):71-81.
 50. Hendrickx B, Junca H, Vosahlova J, Lindner A, Ruegg I, Bucheli-Witschel M, Faber F, Egli T, Mau M, Schlomann M *et al*: Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *Journal of microbiological methods* 2006, 64(2):250-265.
 51. Worsey MJ, Williams PA: Metabolism of toluene and xylenes by Pseudomonas (putida (arvilla) mt-2: evidence for a new function of the TOL plasmid. *Journal of bacteriology* 1975, 124(1):7-13.
 52. Ramos-Gonzalez MI, Duque E, Ramos JL: Conjugational transfer of recombinant DNA in cultures and in soils: host range of Pseudomonas

- putida TOL plasmids. *Applied and environmental microbiology* 1991, 57(10):3020-3027.
53. Top EM, Springael D: The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Current opinion in biotechnology* 2003, 14(3):262-269.
 54. Ahmer BM, Tran M, Heffron F: The virulence plasmid of *Salmonella typhimurium* is self-transmissible. *Journal of bacteriology* 1999, 181(4):1364-1368.
 55. Sandegren L, Linkevicius M, Lytsy B, Melhus A, Andersson DI: Transfer of an *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-specific plasmid associated with a major nosocomial outbreak. *The Journal of antimicrobial chemotherapy* 2012, 67(1):74-83.
 56. Hollowella GP, Kuykendallb LD, Gillettec WK, Hashemd FM, Houa L, Tateme HE, Dutta SK: Genetic transfer and expression of plasmid RP4::TOL in *Sinorhizobium meliloti*, *Bradyrhizobium japonicum* and *B. elkanii*. *Soil Biology and Biochemistry* 1999, 31(13):1811-1819.
 57. Gibson DT, Parales RE: Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Current opinion in biotechnology* 2000, 11(3):236-243.
 58. Vazquez S, Nogales B, Ruberto L, Hernandez E, Christie-Oleza J, Lo Balbo A, Bosch R, Lalucat J, Mac Cormack W: Bacterial community dynamics during bioremediation of diesel oil-contaminated Antarctic soil. *Microbial ecology* 2009, 57(4):598-610.
 59. Simon C, Wiezer A, Strittmatter AW, Daniel R: Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Applied and environmental microbiology* 2009, 75(23):7519-7526.
 60. Silva CC, Hayden H, Sawbridge T, Mele P, Kruger RH, Rodrigues MV, Costa GG, Vidal RO, Sousa MP, Torres AP *et al*: Phylogenetic and functional diversity of metagenomic libraries of phenol degrading sludge from petroleum refinery wastewater treatment system. *AMB Express* 2012, 2(1):18.
 61. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC *et al*: Comparative

- metagenomics of microbial communities. *Science* 2005, 308(5721):554-557.
62. Megadi VB, Tallur PN, Mulla SI, Ninnekar HZ: Bacterial degradation of fungicide captan. *Journal of agricultural and food chemistry* 2010, 58(24):12863-12868.
 63. Gomez-Alvarez V, Teal TK, Schmidt TM: Systematic artifacts in metagenomes from complex microbial communities. *The ISME journal* 2009, 3(11):1314-1317.
 64. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH: The NCBI BioSystems database. *Nucleic acids research* 2010, 38(Database issue):D492-496.
 65. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 1997, 25(17): 3389-3402.
 66. Huson DH, Auch AF, Qi J, Schuster SC: MEGAN analysis of metagenomic data. *Genome research* 2007, 17(3):377-386.
 67. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic acids research* 2012, 40(Database issue):D71-75.
 68. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000, 28(1):27-30.
 69. Carbajosa G, Trigo A, Valencia A, Cases I: Bionemo: molecular information on biodegradation metabolism. *Nucleic acids research* 2009, 37(Database issue):D598-602.
 70. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: The Pfam protein families database. *Nucleic acids research* 2012, 40(Database issue):D290-301.
 71. Finn RD, Clements J, Eddy SR: HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 2011, 39(Web Server issue):W29-37.
 72. Arenskotter M, Broker D, Steinbuchel A: Biology of the metabolically diverse genus *Gordonia*. *Applied and environmental microbiology* 2004, 70(6):3195-3204.
 73. Yergeau E, Sanschagrín S, Beaumier D, Greer CW: Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high arctic soils. *PloS one* 2012, 7(1):e30058.
 74. El Azhari N, Bru D, Martin-Laurent F: Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR. *European Journal of Soil Science* 2008, 59(4):665-673.

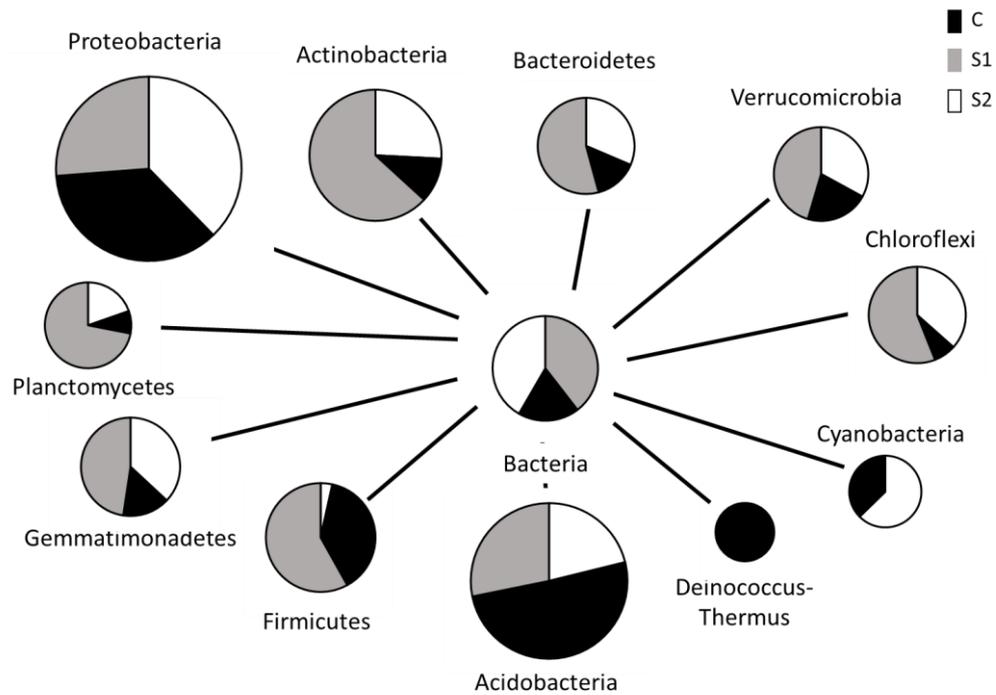
75. Gerischer U: Specific and global regulation of genes associated with the degradation of aromatic compounds in bacteria. *Journal of molecular microbiology and biotechnology* 2002, 4(2):111-121.
76. Kao CM, Chen CY, Chen SC, Chien HY, Chen YL: Application of in situ biosparging to remediate a petroleum-hydrocarbon spill site: field and microbial evaluation. *Chemosphere* 2008, 70(8):1492-1499.
77. Zylstra GJ, Kim E: Aromatic hydrocarbon degradation by *Sphingomonas yanoikuyae* B1. *Journal of Industrial Microbiology and Biotechnology* 1997, 19(5-6):408-414.
78. Eriksson M, Sodersten E, Yu Z, Dalhammar G, Mohn WW: Degradation of polycyclic aromatic hydrocarbons at low temperature under aerobic and nitrate-reducing conditions in enrichment cultures from northern soils. *Applied and environmental microbiology* 2003, 69(1):275-284.
79. Fischer A, Herklotz I, Herrmann S, Thullner M, Weelink SA, Stams AJ, Schlomann M, Richnow HH, Vogt C: Combined carbon and hydrogen isotope fractionation investigations for elucidating benzene biodegradation pathways. *Environmental science & technology* 2008, 42(12):4356-4363.
80. Shinoda Y, Sakai Y, Uenishi H, Uchihashi Y, Hiraishi A, Yukawa H, Yurimoto H, Kato N: Aerobic and anaerobic toluene degradation by a newly isolated denitrifying bacterium, *Thauera* sp. strain DNT-1. *Applied and environmental microbiology* 2004, 70(3):1385-1392.
81. Chamkha M, Trabelsi Y, Mnif S, Sayadi S: Isolation and characterization of *Klebsiella oxytoca* strain degrading crude oil from a Tunisian off-shore oil field. *Journal of basic microbiology* 2011, 51(6):580-589.

Figures



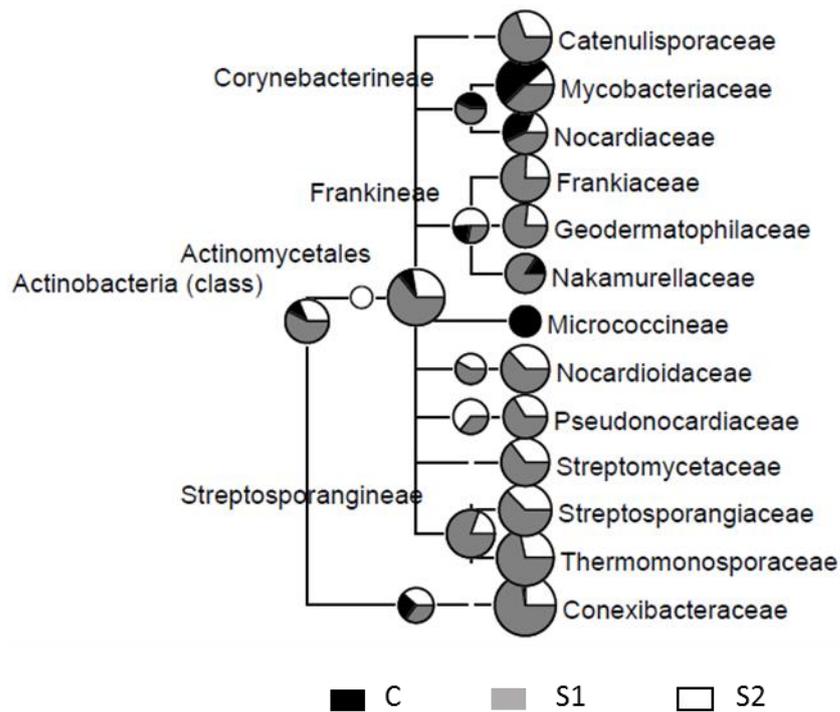
Analysis of aromatic hydrocarbon contamination through CG-MS method. The graph **A** indicates the presence of toluene and xylenes in soil from site C. The graph **B** shows the profile for site S1. No peak is observed indicating no contamination of this site with aromatic hydrocarbons.

Figure 2 - Phyla identification and relative abundance of sequences



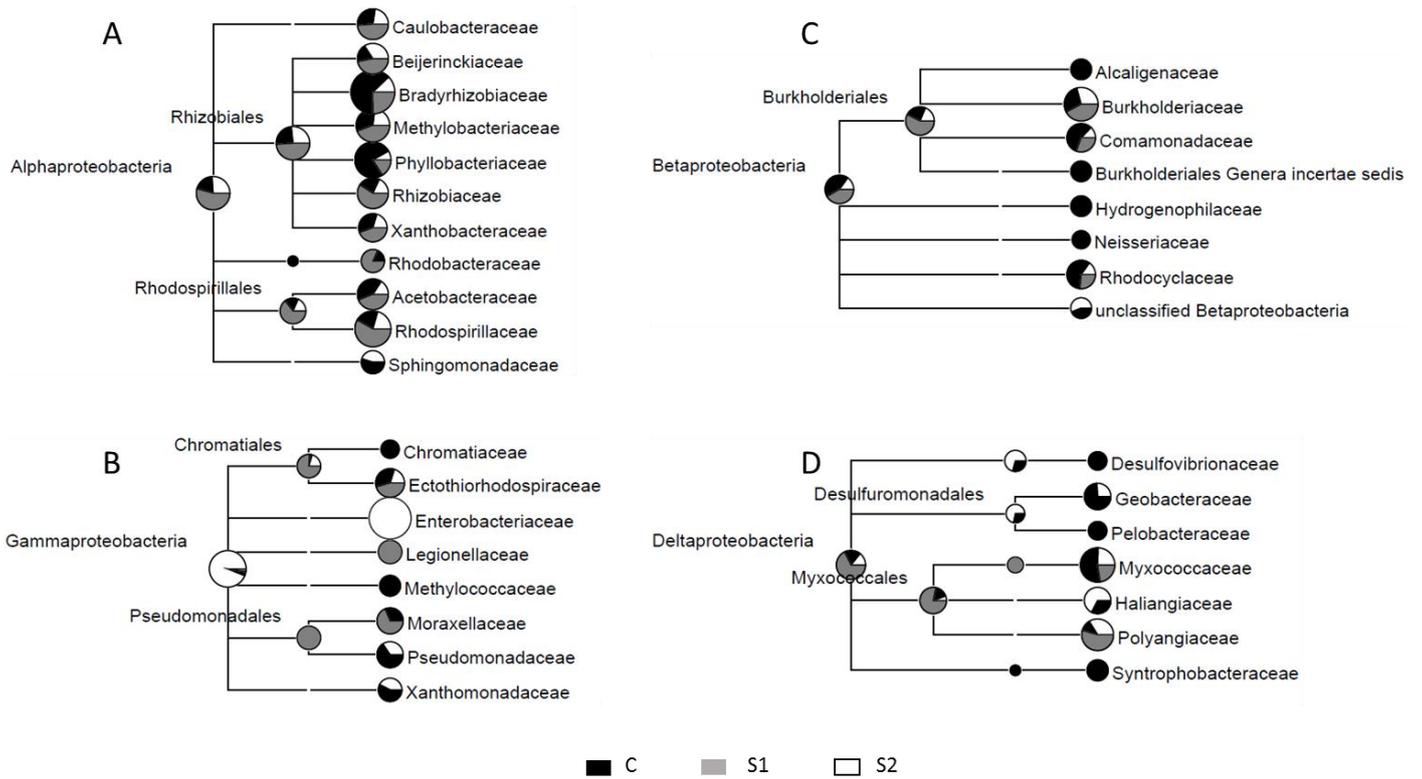
BlastN sequence identification distribution in phyla for Bacteria domain. Each circle represents one phylum found and its respective abundance for the three sites studied. The size of circles is proportional to the number of sequences grouped in each phyla. The distribution showed is normalized.

Figure 3 - Actinobacteria sub-groups



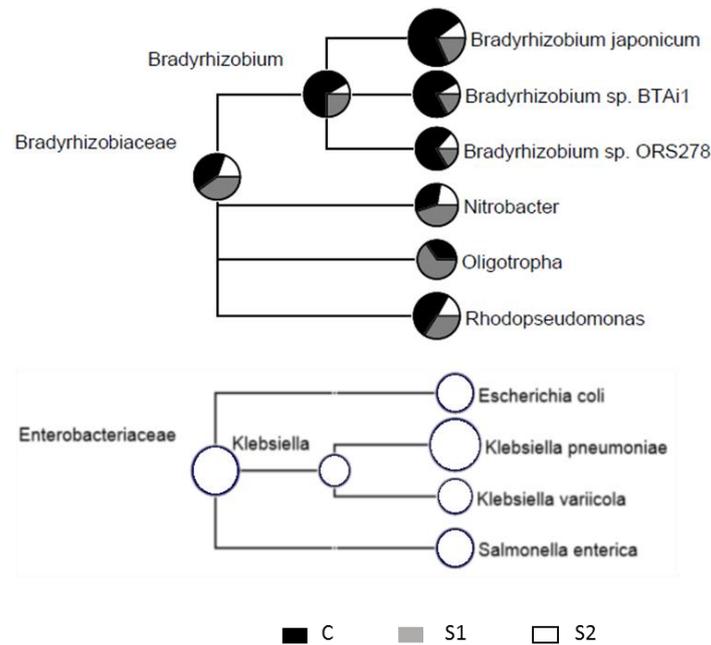
The figure represents the families of Actinobacteria class and the relative abundance found in sites studied for each family. Sequences were identified by BLASTN comparison with prokaryotes genomes of NCBI database and by MEGAN program. The size of circles represents the number of sequences attributed to each one. Sequences that could not be classified in families were included in previous levels.

Figure 4 - Proteobacteria sub-groups



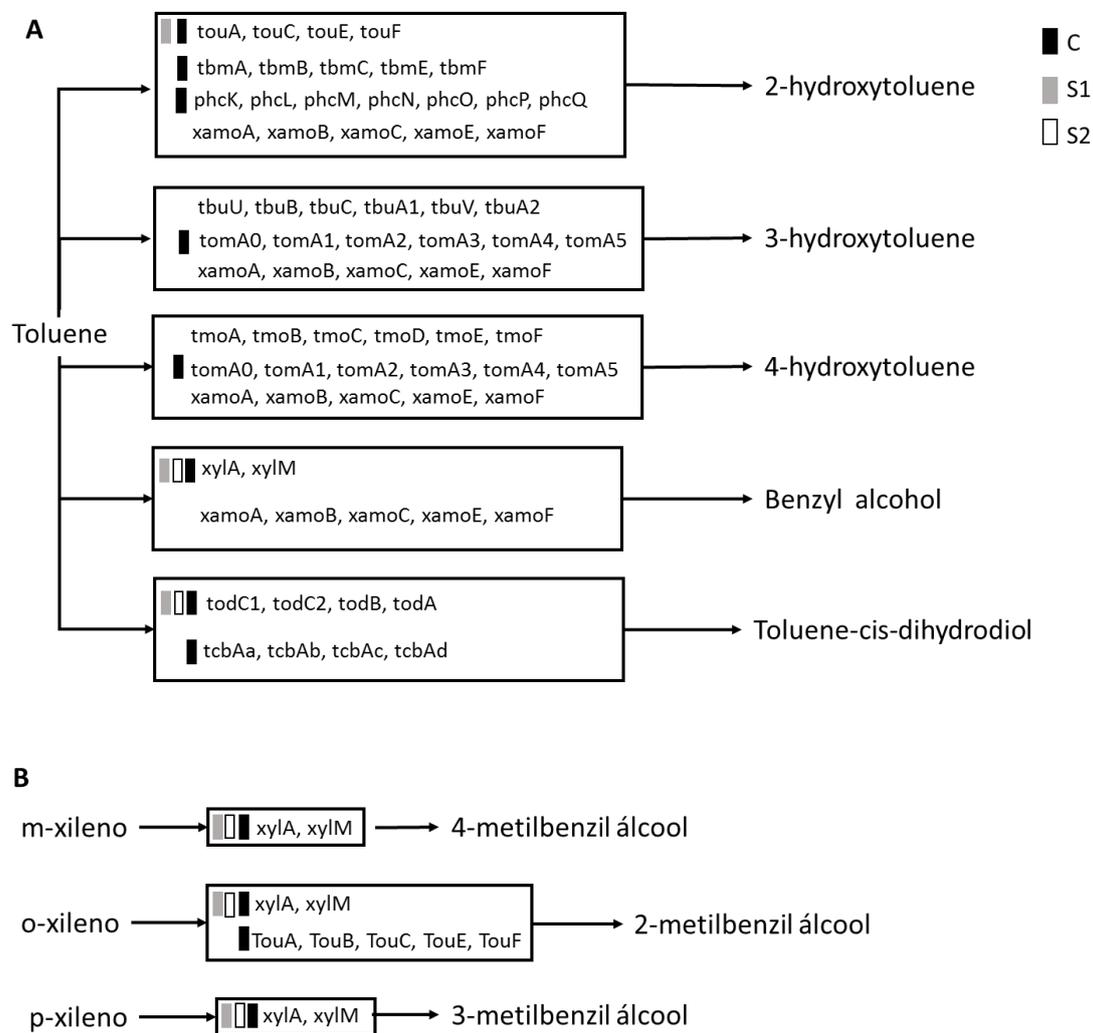
Distribution of identified sequences in families for each class of Proteobacteria phylum. The size of circles represents the normalized number of sequences attributed to each one. Sequences were identified by BLASTN with prokaryotic genomes from NCBI and organized by MEGAN.

Figure 5 - Families most represented



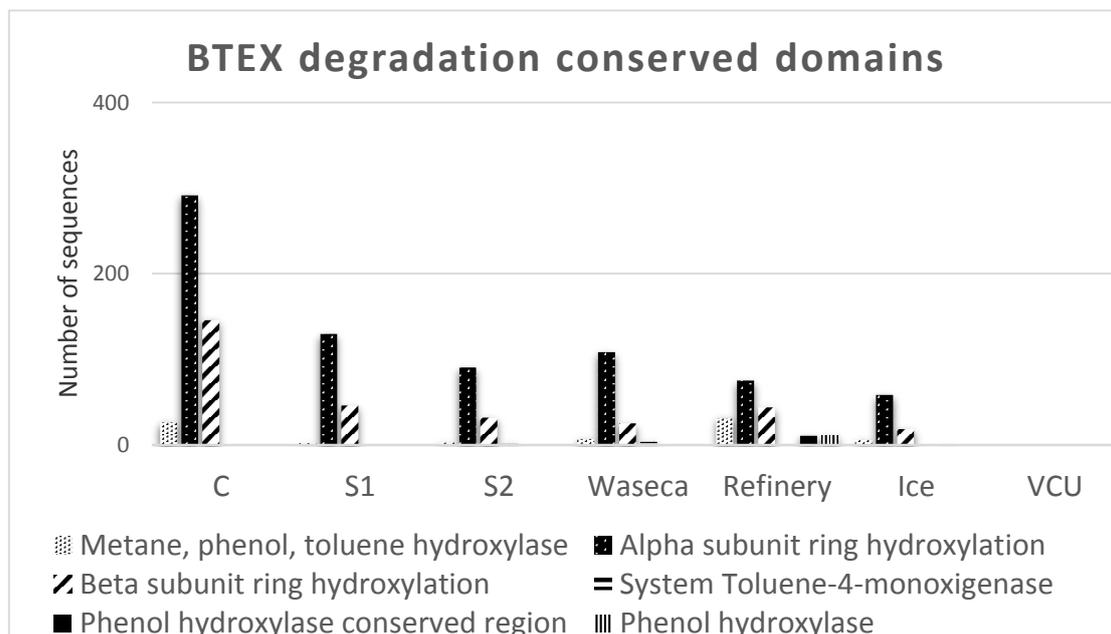
Distribution of sequences identified in Genus and species for family Bradyrhizobiaceae and Enterobacteriaceae, the most abundant families in our study. Sequences were compared in prokaryotic genomes of NCBI database by BLASTN and organized by MEGAN.

Figure 6 - Toluene and xylenes degradation operons



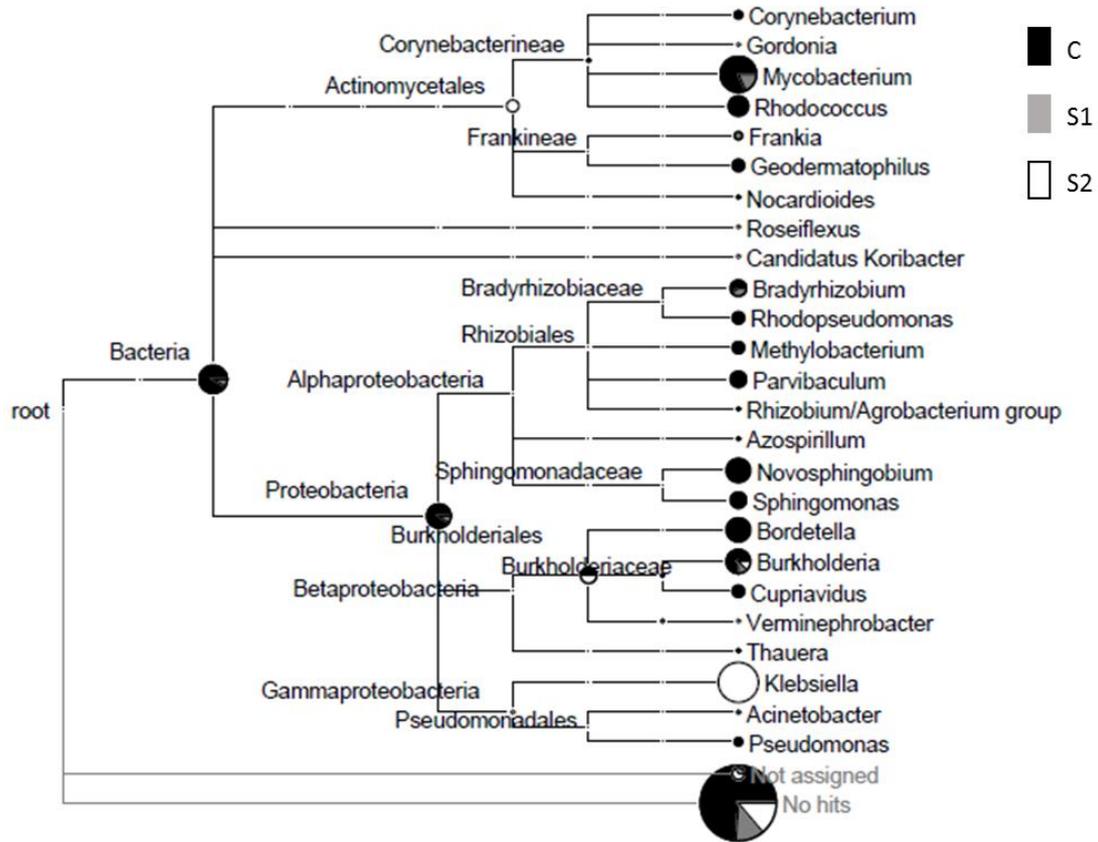
The figure represents the toluene (A) and xylenes (B) degradation operons and its genes. Each operon directs the degradation to one of the pathways showed. The boxes in front of operons' genes indicates the soil where the operon was found.

Figure 7 - Distribution of conserved domains related sequences



The graph shows the distribution of sequences with similarity to the six conserved degradation domains chosen for the metagenomes studied. All metagenomes created for this study (S1, S2 and C) and four metagenomes from NCBI database (Waseca, Refinery, Ice and VCU) are represented. The results were normalized for comparison.

Figure 8 - Taxonomic identification of sequences related to degradation conserved domains



Taxonomic structure proposed based on NCBI prokaryotic genomes classification of sequences with similarity to aromatic degradation domains. The size of circles represent the number of sequences found for each genera. The No hits node represents sequences with conserved degradation domains similarity but with no similarity with prokaryotic genomes through BLASTN analysis. The Not assigned node grouped metagenomic sequences with similarity with in database but with groups not assigned in the taxonomic structure. The sequences from the three sites studied are presented.

Tables

Table 1 - Accounting of metagenomic sequences

	C	S1	S2
<i>Number of sequences</i>	598,041	578,478	533,697
<i>Average size</i>	521	527	487
<i>De-replicated sequences</i>	432,288	238,766	485,022
<i>Identified sequences</i>	7,214	1,365	3,760

Overall information about sequences from pirosequencing method.

Table 2 - Activity confirmation of identified conserved domains holders

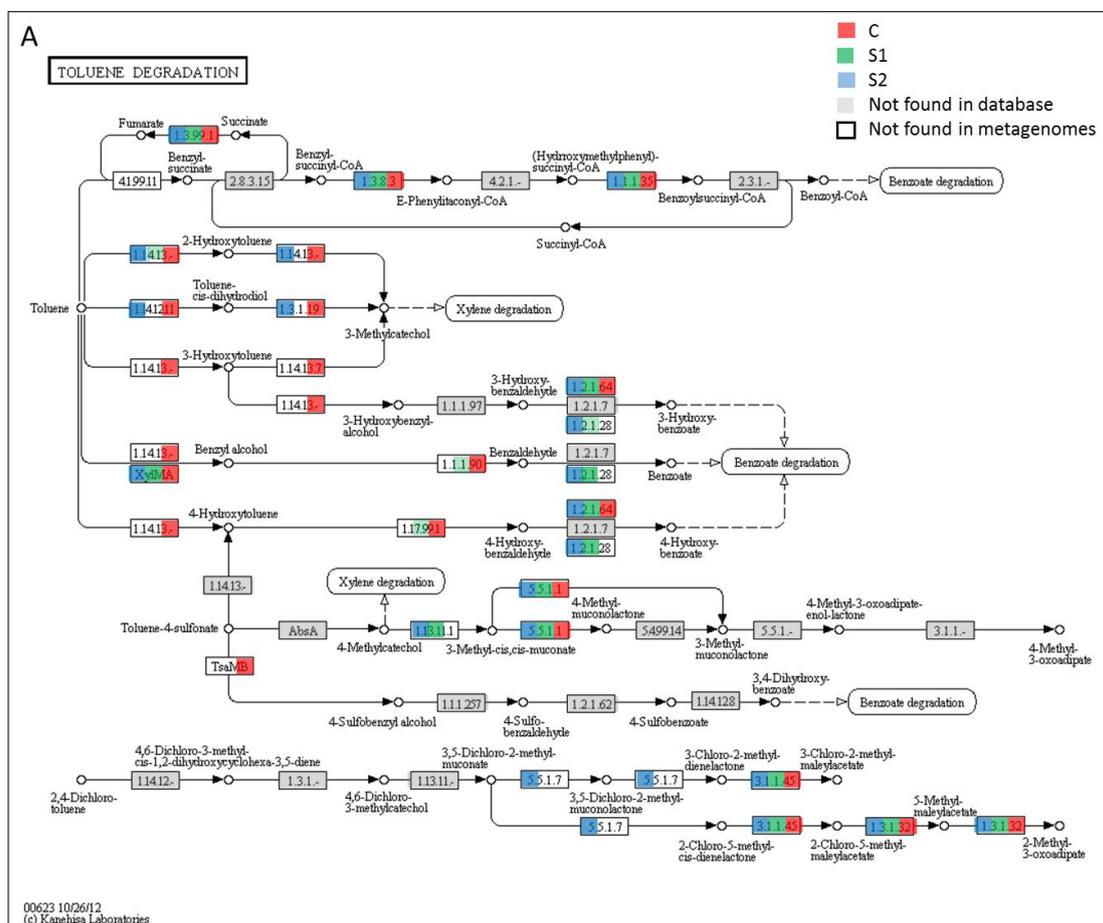
Genera	Degradation Activity	Reference
<i>Gordonia</i>	Proved	[72]
<i>Mycobacterium</i>	Proved	[27]
<i>Rhodococcus</i>	Proved	[45]
<i>Frankia</i>	Potential	[73]
<i>Geodermatophilus</i>	Potential	[74]
<i>Roseiflexus</i>	_____	_____
<i>Koribacter</i>	_____	_____
<i>Bradyrhizobium</i>	Potential	[33]
<i>Rhodopseudomonas</i>	Proved	[75]
<i>Methylobacterium</i>	Potential	[76]
<i>Parvibaculum</i>	Potential	[31]
<i>Sphingomonas</i>	Proved	[77]
<i>Bordetella</i>	Potential	[78]
<i>Burkholderia</i>	Proved	[42]
<i>Cupriavidus</i>	Proved	[79]
<i>Verminephrobacter</i>	_____	_____
<i>Thauera</i>	Proved	[80]
<i>Klebsiella</i>	Proved	[81]
<i>Acinetobacter</i>	Potential	[6]
<i>Pseudomonas</i>	Proved	[30]

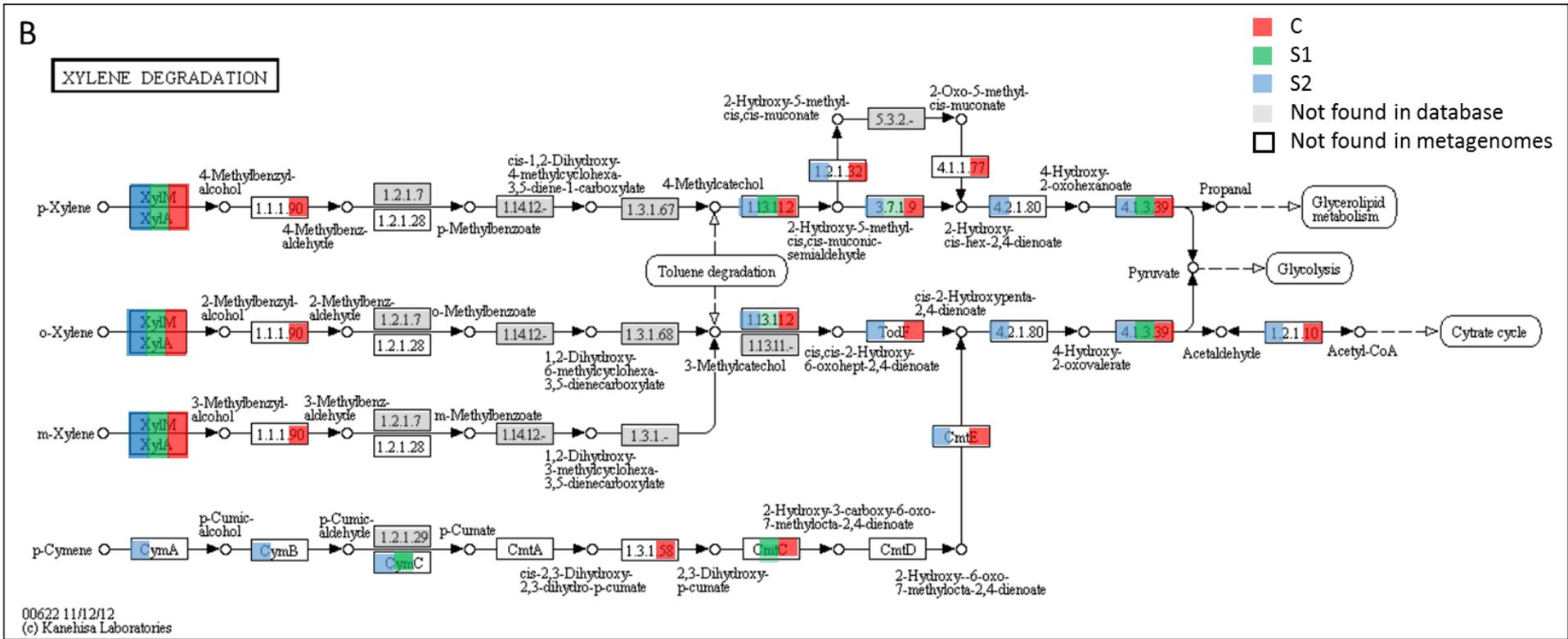
The table shows the genera identified for sequences with conserved degradation domains similarity. The sequences were identified in NCBI prokaryotic genome database. Genera with proved degradation activity are identified with the word “Proved” in the table followed by the study that found the degradation activity. Genera found in sites contaminated with aromatic hydrocarbon are identified with the word “Potential” followed by the study which observed the genera.

...

Additional files

Identification of degradation genes of toluene and xylene in KEGG database metabolic pathways





The figure shows the genes found for sites S1, S2 and C related to toluene and xylenes degradation domains. The genes found in three sites studied are represented