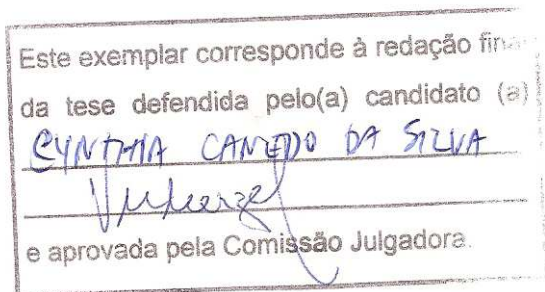




UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

CYNTHIA CANÊDO DA SILVA

**“PROSPECÇÃO DA ATIVIDADE DE DEGRADAÇÃO DE
FENOL EM METAGENOMA MICROBIANO ORIGINADO DE
EFLUENTE DE REFINARIA DE PETRÓLEO”**



Tese apresentada ao Instituto de Biologia,
para a obtenção do Título de Doutor em
Genética e Biologia Molecular na área de
Genética de Microorganismos.

Orientadora: Dra. Valéria Maia Merzel

Campinas, 2010

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

Si38p

Silva, Cynthia Canedo da
Prospecção da atividade de degradação de fenol em metagenoma microbiano originado de efluente de refinaria de petróleo / Cynthia Canedo da Silva. – Campinas, SP: [s.n.], 2010.

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

1. Lodo - Microbiologia. 2. Metagenoma. 3. Diversidade bacteriana. 4. Biblioteca de fosmídeos. 5. Fenol. 6. RNA ribossômico. I. Oliveira, Valéria Maia de. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Prospection of phenol-degrading activity in microbial metagenome from petroleum refinery wastewater.

Palavras-chave em inglês: Sludge - Microbiology; Metagenome; Bacterial diversity; Fosmid library; Phenol; Ribosomal RNA.

Área de concentração: Genética de Microorganismos.

Titulação: Doutora em Genética e Biologia Molecular.

Banca examinadora: Valéria Maia de Oliveira, Anete Pereira de Souza, Márcio Rodrigues Lambais, Ricardo Henrique Kruger, Wellington Luiz Araújo.

Data da defesa: 28/06/2010.

Programa de Pós-Graduação: Genética e Biologia Molecular.

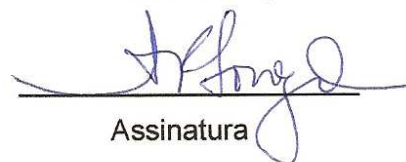
Campinas, 28 de Junho de 2010

BANCA EXAMINADORA

Dr^a. Valéria Maia Merzel (Orientadora)


Assinatura

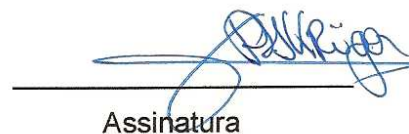
Prof^a. Dra. Anete Pereira de Souza


Assinatura

Prof. Dr. Márcio Rodrigues Lambais

Assinatura

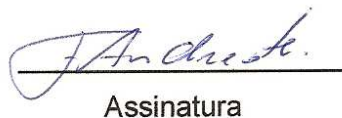
Prof. Dr. Ricardo Henrique Kruger


Assinatura

Prof. Dr. Wellington Luiz Araújo

Assinatura

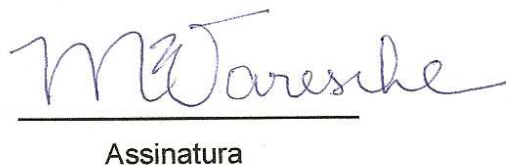
Prof. Dr. Fernando Dini Andreote


Assinatura

Prof^a. Dr^a. Siu Mui Tsai

Assinatura

Prof^a. Dr^a. Maria Bernadete Amâncio Varesche Silva


Assinatura

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha Tese de Doutorado intitulada **Prospecção da atividade de degradação de fenol em metagenoma microbiano originado de efluente de refinaria de petróleo:**

() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

(X) tem autorização da(s) seguinte(s) Comissão(ões) de Bioética ou Biossegurança*:
CIBio/ CPQBA, sob Protocolo(s) nº 10/06.

** Caso a Comissão seja externa à UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*

Cynthia Canêdo da Silva

Aluno:

Isabela Inês de Oliveira

Orientador:

Para uso da Comissão ou Comitê pertinente:

(X) Deferido () Indeferido

Helena Coutinho F. de Oliveira

Nome:

Função:

Profª. Dra. HELENA COUTINHO F. DE OLIVEIRA

Presidente

Comissão Interna de Biossegurança

CIBio/IB - UNICAMP

Dedico esta Tese com muito carinho:

*À minha família que amo
incondicionalmente, meus pais **Rita** e **Erli** e
meus irmãos **Marcos Vinícius** e
Ívina,*

*E ao **Sérgio**, quem aprendi a
amar e respeitar ao longo desta caminhada.*

AGRADECIMENTOS

À Deus, pelo dom da vida, por sempre me proteger e guiar meus caminhos, tornando-os mais fáceis e agradáveis;

À Universidade Estadual de Campinas (UNICAMP) em especial ao Instituto de Biologia (IB) e ao Departamento de Genética e Biologia Molecular, pela oportunidade e execução deste trabalho;

À minha orientadora, Prof^a Dr^a Valéria Maia de Oliveira, pelo apoio, amizade e confiança no transcorrer deste trabalho, contribuindo para meu crescimento pessoal e profissional;

Aos membros da Banca Prof^a Dr^a Anete Pereira, Prof Dr Márcio Lambias, Prof. Dr. Ricardo Kruger e Prof. Dr. Wellinton Araújo, pelas precisas correções e sugestões realizadas nesta Tese;

Aos professores da Pré-banca, Dra Lara D. Sette e Prof. Dr. Márcio R. Lambais, pelas valiosas sugestões para finalização deste trabalho;

Aos professores da Qualificação, Prof. Dr. Marcelo Brocchi (UNICAMP – SP), Prof. Dr. Wellington L. Araújo (UNIRARAS) e Profa. Dra. Profa.Dra. Irma Nelly Gutierrez Rivera (USP), pelas sugestões para desenvolvimento do meu projeto;

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de doutorado;

À PETROBRAS pelo apoio financeiro do meu projeto de pesquisa e formação profissional, e em especial à Ana Paula Rodrigues Torres, Maíra P. Sousa e Vânia M. J. Santiago;

Às equipes de trabalho das refinarias REGAP e REVAP pela boa vontade na coleta de material para este estudo;

Aos integrantes do Laboratório de Gerência de Biotecnologia e Tratamentos Ambientais / CENPES pela colaboração nas análises, utilização do laboratório e equipamentos;

Aos funcionários do Instituto de Biologia (IB), pelo apoio e colaboração;

À secretaria de pós-graduação de Genética e Biologia Molecular, pela paciência e disposição na parte burocrática;

Aos funcionários do Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas, pelo apoio e colaboração;

As pesquisadoras da Divisão de Recursos Microbianos, Lara Sette Durães e Fabiana Fantinatti pelos conselhos, discussões e sugestões;

Aos todos os colegas da Divisão de Recursos Microbianos, que farei um esforço para citar todos: Milena, João, Gisele, Mariana, Patrícia, Samantha, Rafaella, Cristina, Barbara, Paula, Bruna, Natália, Cláudia, Michel, Leandro, Tiagão e Fernando pela amizade, companheirismo e agradáveis momentos de convivência e incentivo;

Ao Department of Primary Industries/Victoria AgroBioscience (Melbourne, Australia) e seus integrantes, em especial a Helen, Tim, Damian e Pauline, pela

hospitalidade, paciência, orientação, disponibilização do laboratório e equipamento na etapa do pirosequenciamento;

À Comunidade Dagmar, em especial Dagma e Juarez, pelo apóio e agradáveis momentos de convivência e confraternização;

Às amigas: Eloisa, Luciana, Sabrina, Suzan, pela amizade, apoio, conselhos e palavras de incentivo;

Às mais novas amigas: Andrea Giordano, Maia, Yui Osanai e Monica Romero pela amizade, companhia e apoio durante minha estadia na Austrália;

Enfim, meu muito obrigada a todos que colaboraram direta ou indiretamente para execução deste trabalho,

Cynthia

*“Apesar do tamanho diminuto, os micro-organismos representam
mais da metade da biomassa da Terra.”*

Brock Biology of Microorganisms

RESUMO

Os efluentes de refinarias contêm vários compostos tóxicos, que podem provocar sérios danos aos seres vivos se não tratados previamente. Os fenóis são poluentes encontrados em efluentes de diversas indústrias, incluindo as refinarias de petróleo, e são conhecidos como um dos compostos mais tóxicos encontrados no meio ambiente. Portanto, há necessidade constante de buscar formas de remover ou reduzir a presença destas substâncias nos efluentes. Este trabalho teve como objetivos analisar e explorar a diversidade microbiana presente em lodo de sistemas de tratamento de efluentes de refinarias petróleo, através da construção de bibliotecas de genes RNAr 16S e bibliotecas metagenômicas em vetores do tipo fosmídeo. Análises filogenéticas das bibliotecas de genes RNAr 16S e metagenômicas mostraram uma comunidade bacteriana bastante diversa, com predominância do Filo Proteobacteria, seguido de Actinobacteria, Planctomycetes, Verrucomicrobia, Chloroflexi e Bacteroidetes. Dentre os gêneros mais abundantes, se destacaram *Thauera*, *Comamonas*, *Diaphorobacter* e *Thiobacillus*. Espécies pertencentes a estes gêneros realizam o processo de desnitrificação e espécies de *Thauera* e *Comamonas* estão relacionadas com a degradação de compostos aromáticos. A grande diversidade bacteriana refletiu em uma ampla diversidade metabólica, sendo obtido um grande número de resultados para genes relacionados com processos biológicos importantes para o tratamento de efluentes, como o metabolismo de nitrogênio, enxofre, fósforo, além de genes relacionados com o catabolismo de compostos aromáticos, tais como benzoato, bifenil e fenol. Adicionalmente, foram observadas possíveis novas sequências para os genes que codificam enzimas-chaves responsáveis pela *meta*-clivagem de fenol e outros aromáticos, como fenol hidroxilase e catecol 2,3-dioxigenase. No presente estudo foram também realizadas triagens funcionais dos 13.200 clones para enzimas hidrolíticas, com detecção de dois hits para esterase e um para lipase. Este trabalho mostrou que a bioprospecção genômica microbiana é uma estratégia inovadora na área de tratamento de efluentes, que pode contribuir para o conhecimento da microbiota responsável pela degradação dos poluentes.

ABSTRACT

Refinery wastewater has several toxic compounds that may cause serious damage to life. Phenolic compounds are found in several industrial effluents, including petroleum refinery, and represent a significant environmental toxicity hazard. Therefore, there is a constant need to search for new technologies able to remove and reduce the presence of these substances in wastewater. This study aimed to analyze and explore the microbial diversity present in sludge from refinery wastewater treatment systems through the construction of 16S rRNA genes libraries and metagenomic libraries in fosmid vectors. Sequencing and phylogenetic analyses of libraries showed a highly diverse bacterial community, with predominance of the Proteobacteria phylum, followed by Actinobacteria, Planctomycetes, Verrucomicrobia, Chloroflexi and Bacteroidetes. The most abundant genera were *Thauera*, *Comamonas*, *Diaphorobacter* and *Thiobacillus*. Species belonging to these genera are facultative anaerobic and may be involved with the denitrification process, and *Thauera* and *Comamonas* species are related with degradation of aromatic compounds. This bacterial diversity reflected in a broad metabolic diversity. Analysis of the metagenomic data derived from pyrosequencing revealed a lot of hits related with biological processes of great relevance for wastewater treatment, such as genes for nitrogen, sulfur and phosphorus metabolism, in addition to genes related to the catabolism of aromatic compounds, such as benzoate, biphenyl and phenol. Additionally, sequences representing possible new genes encoding for key-enzymes responsible for the *meta*-cleavage of phenol and other aromatic compounds, such as phenol hydroxylase and catechol 2,3-dioxygenase, were found in the fosmid libraries. Finally, high-throughput functional *screening* of 13,200 clones for hydrolytic enzymes yielded two positive hits for lipase and one for esterase. The data gathered together in this work showed that microbial genomic prospecting is an innovative strategy for the wastewater treatment system that may undoubtedly contribute to the knowledge of microorganisms responsible by pollutant degradation.

SUMÁRIO

SUMÁRIO.....	xii
1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	5
2.1. Efluentes hídricos de refinarias	5
2.2. Fenóis	7
2.3. Biodegradação dos compostos fenólicos.....	9
2.4. Bibliotecas metagenômicas e sua exploração biotecnológica.....	12
2.5. Estratégias de enriquecimento de bibliotecas metagenômicas	17
2.6. Hidrolases.....	19
2.7. REFERÊNCIAS BIBLIOGRÁFICAS	23
3. OBJETIVOS	28
3.1. Objetivos gerais	28
3.2. Objetivos específicos.....	28
4. APRESENTAÇÃO DA TESE	30
CAPÍTULO 1 - Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load	31
CAPÍTULO 2 – Investigation of bacterial diversity in membrane bioreactor and convectional activated sludge processes from petroleum refineries using phylogenetics and statistical approaches.	41
CAPÍTULO 3 – Diversidade molecular em bibliotecas metagenômicas de lodo bacteriano provenientes de estação de tratamentos de efluentes analisados pela tecnologia de pirosequenciamento 454.	55
3.1. INTRODUÇÃO	56
3.2. MATERIAL E MÉTODOS	58
3.2.1. Amostras e extração de DNA.....	58

3.2.2. Aclimação do lodo.....	59
3.2.3. Taxa de degradação do fenol após aclimação.....	60
3.2.4. Construção das bibliotecas metagenômicas	61
3.2.5. Extração do DNA fosmidial e pirosequenciamento	65
3.2.6. Análise das sequências	67
3.3. RESULTADOS E DISCUSSÃO	68
3.3.1. Aclimação do lodo REGAP a altas concentrações de fenol	68
3.3.2. Construção das bibliotecas metagenômicas	70
3.3.3. Dados do pirosequenciamento e montagem de contigs.....	73
3.3.4. Perfil taxonômico e fisiológico das sequências metagenômicas	73
3.3.5. Análise dos contigs >500 pb.....	77
3.3.6. Análise das sequências de genes RNAr 16S dos dados metagenômicos	80
3.3.7. Análise do conteúdo GC das sequências metagenômicas.....	83
3.3.8. Perfil funcional das sequências metagenômicas.....	84
3.4. CONCLUSÕES.....	87
3.5. REFERÊNCIAS BIBLIOGRÁFICAS	88
CAPÍTULO 4. Prospecção de novos genes relacionados à degradação de fenol e derivados em bibliotecas metagenômicas de lodo de refinaria de petróleo.	91
4.1. INTRODUÇÃO	92
4.2. MATERIAL E MÉTODOS	94
4.2.1. Construção e validação das bibliotecas metagenômicas	94
4.2.2. Screening do gene que codifica a enzima fenol hidroxilase	95
4.2.3. Screening para atividade de degradação de fenol	96
4.2.4. Ensaios de degradação do fenol por cromatografia gasosa.....	97
4.2.5. Extração do DNA fosmidial e pirosequenciamento	98

4.2.6. <i>Análise das sequências</i>	99
4.3. RESULTADOS E DISCUSSÃO	101
4.3.1. <i>Triagens funcionais</i>	101
4.3.2. <i>Cromatografia gasosa para degradação de fenol</i>	102
4.3.4. <i>Análise dos dados da corrida de pirosequenciamento</i>	103
4.3.5. <i>Análise funcional dos dados de pirosequenciamento</i>	104
4.3.6. <i>Análise de genes e rotas relacionadas com degradação fenol</i>	106
4.4. CONCLUSÕES	112
4.5. REFERÊNCIAS BIBLIOGRÁFICAS	113
CAPÍTULO 5. <i>Screening</i> de enzimas hidrolíticas em bibliotecas metagenômicas de lodo bacteriano proveniente de estações de tratamento de efluentes de refinarias de petróleo	115
5.1. INTRODUÇÃO	116
5.2. MATERIAL E MÉTODOS	118
5.2.1. <i>Amostras, construção e validação das bibliotecas metagenômicas</i>	118
5.2.2. <i>Screening para atividade de lipase e esterases</i>	119
5.2.3. <i>Screening para atividade de epóxido hidrolase</i>	120
5.2.4. <i>Screening para atividade de protease</i>	120
5.2.5. <i>Ensaio HTS e de biocatálise convencional</i>	121
5.3. RESULTADOS E DISCUSSÃO	124
5.4. CONCLUSÕES	128
5.5. REFERÊNCIAS BIBLIOGRÁFICAS	130
5. DISCUSSÃO GERAL	133
5.1. REFERÊNCIAS BIBLIOGRÁFICAS	140
6. CONCLUSÕES GERAIS	142
7. PERSPECTIVAS FUTURAS	144

1. INTRODUÇÃO

Entre todos os problemas provocados pelo fenômeno de poluição ambiental, a contaminação das águas, e a consequente escassez de recursos hídricos de boa qualidade, se encontram entre os mais importantes, não só no Brasil, mas em todo planeta. Frente a esta situação, as regulamentações relacionadas com a qualidade das águas vêm se tornando mais restritivas e a fiscalização cada vez mais presente (Chamarro *et al.*, 2001). Neste contexto, a contaminação por petróleo e derivados tem causado bastante preocupação, seja pela frequência dos eventos de contaminação ou pelo elevado potencial poluente dos mesmos.

Além dos derramamentos, uma importante parcela do processo de contaminação pode ser atribuída às atividades das refinarias de petróleo e seus derivados (Freire *et al.*, 2000). No campo da poluição hídrica, as refinarias de petróleo têm feito um esforço considerável no sentido de tratar os seus efluentes em nível primário e secundário e, em alguns casos, em nível terciário para a remoção de nitrogênio amoniacal. Apesar desse esforço, os efluentes tratados ainda apresentam compostos recalcitrantes e alguma toxicidade crônica, que se constituem em fatores limitantes para o reuso dessa água. No caso das refinarias, a geração de água com qualidade adequada para reutilização industrial constitui-se em uma difícil tarefa, em função da complexidade e da variedade da composição química desses efluentes.

Compostos como fenol e derivados, habitualmente presentes nos efluentes das refinarias, são muito tóxicos para a saúde humana e à vida aquática (Mota, 1997). Estes compostos, mesmo quando presentes em baixas concentrações, 5-25 mg L⁻¹, podem provocar danos à fauna e à flora devido à toxicidade advinda dos efeitos sinérgicos das interações entre os mesmos (Braile, 1979, Contreas *et al.*, 2008).

O tratamento biológico dos efluentes é intensamente aplicado nas refinarias, pois a utilização de micro-organismos para remoção de poluentes, como compostos fenólicos e hidrocarbonetos (principais contaminantes presentes nos efluentes de refinarias), é uma prática internacional (Mariano, 2001). Assim, a biorremediação, que emprega micro-organismos para mineralizar ou complexar compostos poluentes orgânicos e inorgânicos, transformando-os em compostos inertes ou não tóxicos, tem se tornado uma metodologia avançada para o tratamento e limpeza dos efluentes de refinarias, mostrando também ser uma próspera tecnologia para o estabelecimento da reutilização de água nestas indústrias.

Entretanto, tem sido demonstrado que alguns poluentes apresentam uma recalcitrância incomum, isto é, não são metabolizados por micro-organismos ou são transformados em metabólitos capazes de se acumular no sistema. Nestes casos, a descoberta de novas vias catabólicas levando à completa mineralização do poluente seria extremamente valiosa. Embora as técnicas de cultivo tenham sido aprimoradas, sabe-se que apenas uma pequena fração da diversidade microbiana na natureza, em torno de 1 a 10%, pode ser hoje cultivada em laboratório (Ward *et al.*, 1990; Amann *et al.*, 1995). Este fato é demonstrado pela alta taxa de re-isolamento das mesmas espécies quando empregamos os métodos tradicionais de

cultivo para recuperar micro-organismos que degradam poluentes, resultando em uma visão limitada das vias de degradação microbiana.

Metodologias moleculares desenvolvidas nas últimas décadas, como extração de ácidos nucleicos diretamente a partir de amostras ambientais, amplificação por PCR, clonagem e sequenciamento de DNA, têm sido otimizadas e adaptadas para superar as limitações impostas pela abordagem clássica de estudo de populações microbianas, evitando o isolamento e cultivo dos micro-organismos. A utilização destas metodologias vem permitindo uma avaliação mais precisa da diversidade microbiana no ambiente e acelerando a descoberta de novos grupos de organismos e funções, nunca antes explorados (Amann *et al.* 1995; Hugenholtz *et al.*, 1998; Steele *et al.*, 2009). O emprego destas metodologias independentes-de-cultivo, associado à estratégia de clonagem de grandes fragmentos de DNA, nos permite ter acesso ao potencial metabólico destes novos organismos. Esta abordagem denominada de metagenômica, consiste na utilização de vetores, tipo BACs, cosmídeos e fosmídeos, utilizados para clonagem de fragmentos de DNA (40 a 300 kb), obtidos diretamente da comunidade microbiana ambientais, seguido de análises baseadas na sequência e função da biblioteca resultante (Handelsman *et al.*, 1998; Rondon *et al.*, 2000; Uchiyama & Miyazaki, 2009). Uma das vantagens da clonagem de genes ou operons em vetores do tipo fosmídeo é que, além de permitir maior estabilidade do inserto, dispensa tediosas metodologias de caracterização gênica, como por exemplo, *primer walking* (Uchiyama & Miyazaki, 2009). Um refinamento da estratégia de construção de bibliotecas metagenômicas consiste em enriquecer a biblioteca para os membros de interesse na comunidade. Desta forma, através da adição de determinado composto como única fonte de carbono, pode-se favorecer a seleção daquelas bactérias que utilizam este nutriente como fonte de

energia, aumentando as possibilidades de obtenção de clones com a atividade de degradação desejada (Schloss & Handelsman, 2003). Embora o enriquecimento de culturas no laboratório produza uma biodiversidade limitada, esta estratégia tem gerado uma alta eficiência de isolamento de fragmentos de DNA e clonagem de operons e genes de elevado valor biotecnológico (Streit & Schmitz, 2004).

No caso do estudo de comunidades microbianas para degradação de compostos poluentes, as bibliotecas metagenômicas constituem-se em uma ferramenta que abre perspectivas para a descoberta de novas vias catabólicas a partir de membros da microbiota, principalmente porque inclui os micro-organismos fastidiosos ou ainda não-cultivados, ou mesmo daqueles que vivem em consórcios e que não podem ser recuperados isoladamente. Ainda, a partir de uma biblioteca de metagenoma, uma variedade maior de genes catabólicos pode ser obtida simultaneamente, em comparação ao método tradicional baseado em isolamento, cultivo e triagem de linhagens puras de micro-organismos.

Neste sentido, o foco deste trabalho foi a busca de genes relacionados a novas atividades de degradação de fenol, bem como de outros compostos poluentes encontrados em efluentes de refinarias de petróleo, como hidrocarbonetos, visando a aplicação futura destes clones ou da informação de sequência derivada dos mesmos no aprimoramento dos sistemas de tratamento biológico. Isto representa uma estratégia inovadora no país, pois envolve o uso de metodologias independentes de cultivo, capazes de explorar de maneira mais abrangente a diversidade microbiana presente no lodo proveniente de reatores de sistemas de tratamento de efluentes de refinarias de petróleo.

2. REVISÃO BIBLIOGRÁFICA

2.1. Efluentes hídricos de refinarias

Devido ao elevado potencial de agressão ao meio ambiente, as indústrias têm despertado um especial interesse para o controle de seus processos e para regulamentação da emissão de seus poluentes. A regulamentação sobre o gerenciamento de recursos hídricos (Lei 9433/97) tem levado os setores produtivos à racionalização do consumo de água (Monteiro *et al.*, 2004). Como outras indústrias, as refinarias são grandes consumidoras de água ou de soluções aquosas, uma vez que o processo de refino consiste de uma série de etapas de beneficiamento, classificadas como processos de separação ou processos de conversão, as quais empregam grande volume de água (Pereira, 2005). A racionalização do uso deste recurso é prioridade no planejamento estratégico das refinarias de petróleo, o que torna premente o investimento em processos que permitam o gerenciamento das águas e efluentes, como uma forma de possibilitar a reutilização deste recurso.

A composição e a quantidade de efluente hídrico gerado pelo processo de refino variam em função do tipo de óleo processado, das unidades de processamento e da forma de operação dessas unidades (Mariano, 2001). De um modo geral, as refinarias geram uma quantidade de efluentes líquidos que é

relativamente proporcional às quantidades de óleo refinado. No caso do Brasil, as onze refinarias do sistema Petrobrás geram entre 0,40 - 0,60 m³ efluente / m³ de óleo refinado na planta (Piras, 1993). Devido ao contato direto da água com o óleo em algumas etapas de refino e da adição de insumos, os efluentes hídricos internos resultantes podem ser muito contaminados, podendo acarretar, quando liberados sem o devido tratamento, em impactos ecológicos e até sócio-econômicos (Pereira, 2005).

Nas refinarias, os efluentes líquidos consistem em águas de resfriamento, águas de processo, água dos esgotos sanitários e águas de chuva, e são direcionados para as estações de tratamento situadas na própria refinaria. Após o tratamento, os efluentes são despejados em estações de tratamento públicas ou em corpos receptores, desde que atendam à legislação ambiental concernente. Nestes efluentes estão presentes diversos tipos de contaminantes, como fenóis, sulfetos, cloretos, ácido sulfídrico, hidróxido de sódio, cálcio e amônia, hidrocarbonetos, carbonatos e metais pesados (Cd, Pb, Cu, Ni, Zn e Cr). O sabor e o odor dos efluentes se devem, principalmente, à presença de compostos fenólicos, naftênicos, nitrogenados e sulfurosos (Mariano, 2001).

Devido à diversidade de efluentes, as refinarias apresentam sistemas separadores para segregar as águas oleosas, de processo, de chuva drenada, de refrigeração, e as de esgoto sanitário. Este processo se faz necessário, uma vez que tais efluentes não receberão os mesmos tratamentos. Estes tratamentos são classificados como pré-tratamento, primário, secundário e terciário. O pré-tratamento baseia-se no emprego de separadores gravitacionais; o primário consiste no processo de neutralização, coagulação química seguida de sedimentação e processos de filtração e flotação. O tratamento secundário inclui o tratamento

biológico, onde a matéria orgânica (poluente) é consumida por micro-organismos. Nesta categoria inclui-se o uso de lodos ativados, lagoas aeradas, filtros biológicos, lagoas de estabilização, torres de oxidação, filtração, adsorção em carvão ativado e osmose reversa. O tratamento terciário consiste na remoção de determinados nutrientes, como nitrogênio e fósforo, que podem potencializar isoladamente ou em conjunto a eutrofização das águas do corpo receptor (Pereira, 2005).

O tratamento biológico é uma prática internacional, intensamente aplicado nas refinarias de petróleo para remoção de poluentes, como fenóis e hidrocarbonetos (Mariano, 2001). Dentre os vários tipos de tratamento biológico se destaca a biorremediação, técnica que emprega micro-organismos para complexar ou mineralizar compostos poluentes orgânicos e inorgânicos a gás carbônico e água ou, transformando-os em compostos inertes ou não tóxicos ao meio ambiente. A biorremediação tem se tornado uma metodologia avançada para tratamento e limpeza de efluentes de refinarias, mostrando ser uma próspera ferramenta para o estabelecimento da prática de reutilização de água nas indústrias.

2.2. Fenóis

O fenol é uma substância incolor, de odor específico e solúvel em vários compostos orgânicos como clorofórmio, éster e óleos. Consiste em uma hidroxila ligada ao anel benzênico (aromático), sendo classificado na função orgânica Fenol. A origem dos fenóis no ambiente ocorre de forma natural, pela liberação de compostos fenólicos pelas plantas, ou antropogênica, neste caso podendo ser xenobiótica. É conhecido como um dos poluentes mais comuns no meio ambiente,

que causa sérios danos aos seres vivos e a diversas reações bioquímicas (Arutchelvan *et al.*, 2005). Tem sido associado a doenças cardiovasculares quando inalado por tempo prolongado, se ingerido causa sérios danos ao sistema gastrointestinal e pode causar sérias queimaduras quando em contato com a pele, além de ser conhecido como uma substância mutagênica e teratogênica (Agency for Toxic Substances and Disease Registry, 2010).

O fenol é amplamente aplicado na indústria, sendo utilizado como matéria-prima ou obtido como resíduo em diversos tipos de indústrias: petroquímicas, refinarias de petróleo, papel celulose, plástica, têxtil, pesticida, farmacêutica, dentre outras. Nas refinarias de petróleo e indústrias petroquímicas este composto é utilizado como matéria-prima para lavagem e condicionamento de produtos ácidos e alcalinos. Isto torna o fenol e seus derivados, poluentes largamente encontrados em diversos efluentes industriais (Breining *et al.*, 2000; Barrios-Martinez *et al.*, 2006).

Neste contexto, diversos processos para a remoção de compostos fenólicos dos efluentes industriais têm sido bastante estudados. O tratamento biológico tem sido uma alternativa preferencialmente escolhida por ser eficiente e permitir uma remoção em larga escala, com baixos custos operacionais quando comparados aos métodos físico-químicos convencionais, além de ser uma metodologia não agressiva ao meio ambiente (Barrios-Martinez *et al.*, 2006). Este tratamento se baseia no emprego de micro-organismos com capacidade de assimilar o poluente como fonte de carbono e/ou energia, podendo promover a completa mineralização destes compostos (Wei *et al.*, 2008).

2.3. Biodegradação dos compostos fenólicos

Várias espécies bacterianas têm sido reportadas como degradadoras de fenóis, dentre elas: *Pseudomonas*, *Acidovorax*, *Rhizobium*, *Bacillus*, *Ochrobactrum*, *Rhodococcus*, *Acinetobacter*, *Thauera*, *Azoarcus*, *Comamonas*, etc (Buswell, 1975; Arai *et al.*, 2000; Breining *et al.*, 2000; El-Sayed *et al.*, 2003; Manefield *et al.*, 2002, Wei *et al.*, 2008, Cordova-Rosa *et al.* 2009, Sueoka *et al.*, 2009). Embora existam micro-organismos aeróbios e anaeróbios capazes de degradar o fenol, o processo aeróbio é realizado pela maioria das bactérias estudadas (Contreas *et al.*, 2008).

Pelas rotas aeróbias de degradação, as bactérias podem utilizar o fenol tipicamente de duas maneiras, via *ortho*- e *meta*-oxidação. Ambas apresentam o catecol como composto intermediário (Figura 1). O anel aromático é primeiro monohidroxilado pela fenol hidroxilase (PH, fenol 2-monooxigenase) na posição *ortho* para formação do grupo hidroxil. O próximo passo é catalisado pela catecol 1,2-dioxigenase (C12O), que inicia a via *ortho*, levando à formação de succinil-CoA e acetil-CoA, ou pela catecol 2,3-dioxigenase (C23O), que inicia a via *meta* levando à formação de piruvato e acetaldeído (Nozaki *et al.*, 1970; Manson & Cammack 1992).

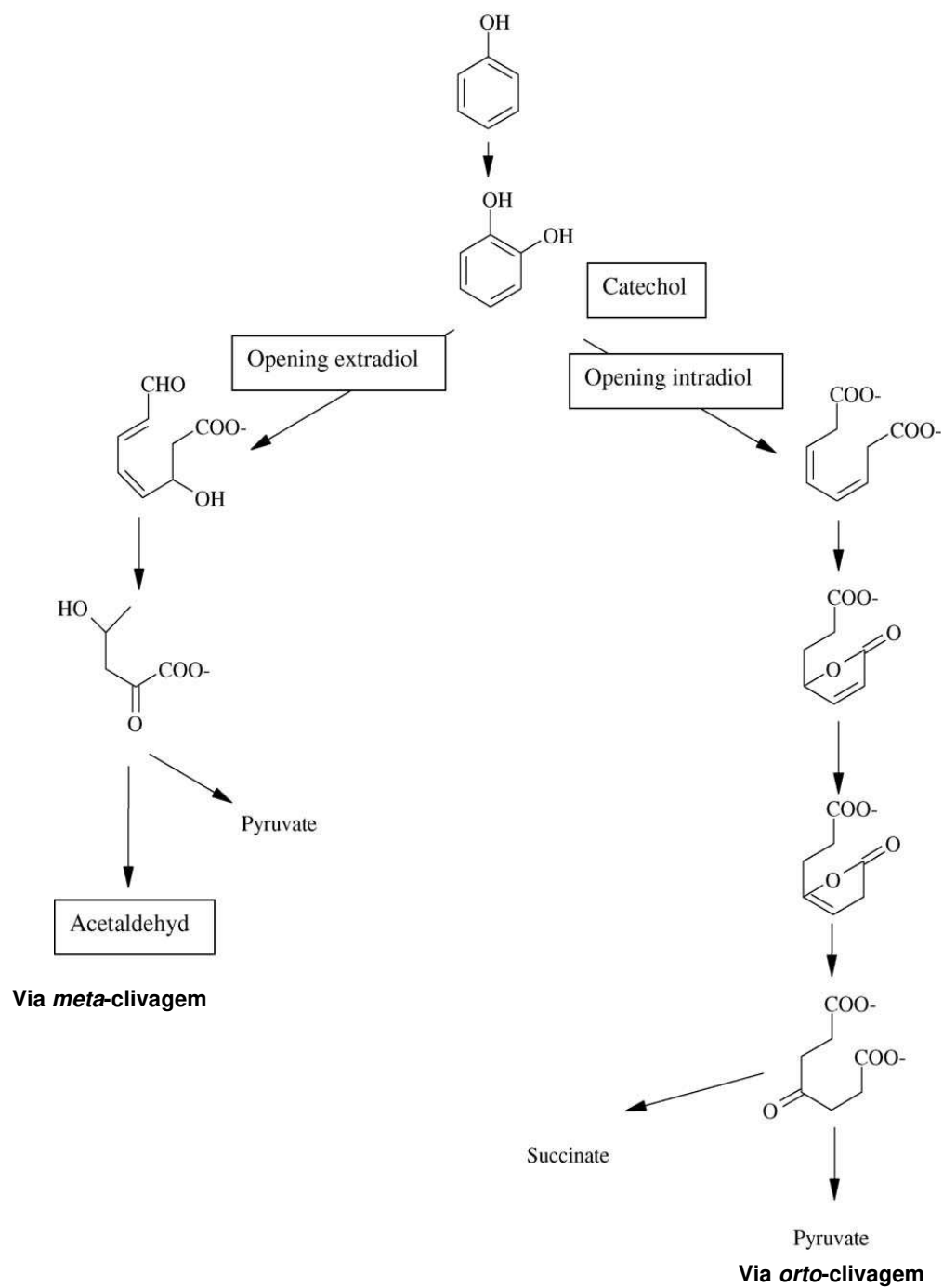


Figura 1. Vias *meta*- e *ortho*-clivagem de degradação aeróbia do fenol. (Fonte: Barrios-Martinez et al., 2006)

A *meta*-oxidação é a mais utilizada entre as bactérias. Espécies de pseudomonaceas mineralizam o fenol pela rota de *meta*-oxidação, produzindo 2 - ácido hidroxi mucônico semialdeído (2-HMAS) como o primeiro derivado do catecol (Barrios-Martinez *et al.*, 2006). *Comamonas testosteroni* AT441 também utiliza esta rota para degradação do fenol após um período de aclimação, o que permite a ocorrência de uma mutação espontânea no operon *aphKLMNOPQB*, que codifica para as enzimas que iniciam os passos da degradação do fenol, fenol hidroxilase e catecol-2,3-dioxygenase (Arai *et al.*, 2000).

A degradação do fenol é limitada pela aclimação da biomassa ao fenol e seus derivados. A ação de toxicidade do fenol está associada com a integridade da membrana citoplasmática (Yap *et al.*, 1999), resultando na disrupção da transdução de energia, perturbação à função de barreira da membrana, inibição da função das proteínas de membrana, causando a subsequente morte celular (Yap *et al.*, 1999). A aclimação consiste no aumento gradual da concentração do substrato ao longo do tempo. Este tempo parece ser necessário para que ocorra a multiplicação das populações capazes de degradar fenol e derivados, além de trocas gênicas entre estas e aquelas não degradadoras destes compostos. Barrios-Martinez e colaboradores (2006) observaram a importância da fase de aclimação para o aumento da taxa de degradação bacteriana e da tolerância de certas populações às altas concentrações de substrato.

Como detalhado acima, os passos iniciais para conversão de fenol em catecol é realizado por diferentes enzimas. Dois tipos de fenol hidroxilase têm sido identificadas, uma flavoproteína de cadeia simples (PHs) e outra multicomponente (PHm). As bactérias *Pseudomonas pickettii* PKO1 e *Pseudomonas* sp. EST1001 apresentam a fenol hidroxilase simples, codificada pelos genes *tbuD* e *pheA*,

respectivamente (Marima *et al.*, 2006). O gene *tbuD* é co-transcrito com o gene *tbuE*, que codifica a enzima catecol 2,3-dioxigenase, e o gene *pheA* divide o operon com o gene *pheB*, que codifica a catecol 1,2-dioxigenase. A enzima fenol hidroxilase multicomponente é composta por seis subunidades de estruturas protéicas semelhantes, com o sítio catabólico localizado na maior subunidade (Watanabe *et al.*, 1998). A enzima catecol 2,3-dioxigenase é uma enzima chave para rota de degradação de muitos compostos aromáticos, principalmente os monocíclicos, e a maioria destas enzimas está relacionada à subfamília 1.2. A da família das proteínas 1.2 extradiol dioxigenase (Eltis & Bolin, 1996).

Até o momento, os genes relacionados à degradação de compostos fenólicos foram descritos com base em bactérias isoladas em meios de cultivo. Considerando que somente 1-10% das bactérias existentes foram cultivadas (Amman *et al.*, 1995), acredita-se que exista uma fonte metabólica inexplorada relacionada com a degradação de fenol daquelas ainda não cultivadas. Esta lacuna vem sendo preenchida pela aplicação de métodos independentes de cultivo como as bibliotecas metagenômicas. Estas têm permitido o conhecimento e acesso às comunidades microbianas oriundas dos mais diversos ambientes para a exploração de novos genes com potencial biotecnológico.

2.4. Bibliotecas metagenômicas e sua exploração biotecnológica

Na última década, estudos de diversidade microbiana empregando microscopia e contagem de células têm revelado que podemos recuperar apenas cerca de 1-10% dos micro-organismos de uma amostra ambiental através do

isolamento e cultivo (Amann *et al.*, 1995; Hugenholtz *et al.*, 1998). Isto se deve basicamente às limitações inerentes às técnicas de plaqueamento e cultivo, uma vez que todos os meios de cultura são seletivos em maior ou menor extensão para os diversos grupos de micro-organismos, e, na maioria das vezes, incapazes de reproduzir as condições encontradas no ambiente. Este é o caso de micro-organismos que vivem sob a forma de consórcios, onde uma espécie depende de produtos do metabolismo de outras espécies para sua sobrevivência e multiplicação; de micro-organismos simbiotes, os quais vivem no interior de outros organismos, e de micro-organismos extremófilos, como aqueles adaptados a condições extremas de temperatura ou pressão. Existem ainda espécies que não podem ser recuperadas porque entraram num estado não-cultivável, e/ou espécies desconhecidas que nunca foram cultivadas antes por falta de métodos adequados.

Nas últimas décadas, metodologias moleculares independentes do cultivo microbiano, como extração de ácidos nucléicos, amplificação por PCR, clonagem e sequenciamento, têm sido otimizadas e adaptadas para superar as limitações impostas pela abordagem clássica, evitando o isolamento e cultivo dos micro-organismos. Dentre os métodos utilizados para acessar a diversidade genética de organismos ainda não cultivados, a metagenômica vem se destacando por ser uma ferramenta robusta, que vem permitindo uma avaliação mais precisa das comunidades microbianas no ambiente e a descoberta de novos grupos de organismos e funções de potencial biotecnológico (Giovannoni *et al.*, 1990; Ward *et al.*, 1990; Handelsman, 2004). A análise metagenômica consiste da extração direta do ácido nucléico de amostras ambientais, seguida de sua clonagem em vetores adequados, transformação em células hospedeiras e *screening* para os genes e/ou funções de interesse (Handelsman, 2004), sem passar por etapas de amplificação

do DNA (Figura 2). Estas bibliotecas metagenômicas podem ser construídas a partir de grandes fragmentos de DNA (25 a 200 kb) extraídos diretamente de amostras ambientais e clonados em vetores específicos, como BACs (*Bacterial Artificial Chromosome*), cosmídeos e fosmídeos, permitindo a captura e investigação de operons completos ou genes pertencentes a rotas metabólicas relacionadas com síntese ou catabolismo de moléculas complexas, tais como antibióticos e compostos aromáticos (Rondon *et al.*, 2000, Schloss & Handelsman, 2003). A escolha dos vetores irá depender do tamanho do inserto a ser clonado, sendo que BACs suportam fragmentos de DNA entre 100-200 Kb, cosmídeos entre 25-35 Kb e fosmídeos entre 25-40 Kb.

A triagem das bibliotecas metagenômicas para seleção de genes ou função de interesse pode ser realizada com base em similaridade de sequência ou com base na expressão da atividade biológica (Lorenz *et al.*, 2002). Os métodos baseados em sequências utilizam oligonucleotídeos para identificação dos genes-alvo. Os *primers* e sondas são desenhados com base em regiões conservadas das sequências de aminoácidos conhecidas, de maneira a garantir uma alta probabilidade de pareamento com sequências gênicas desconhecidas presentes no ambiente. É considerada uma estratégia atraente porque qualquer gene que apresente identidade com a sequência de DNA utilizada será detectado, independente da sua expressão na célula hospedeira (Lorenz *et al.*, 2002).

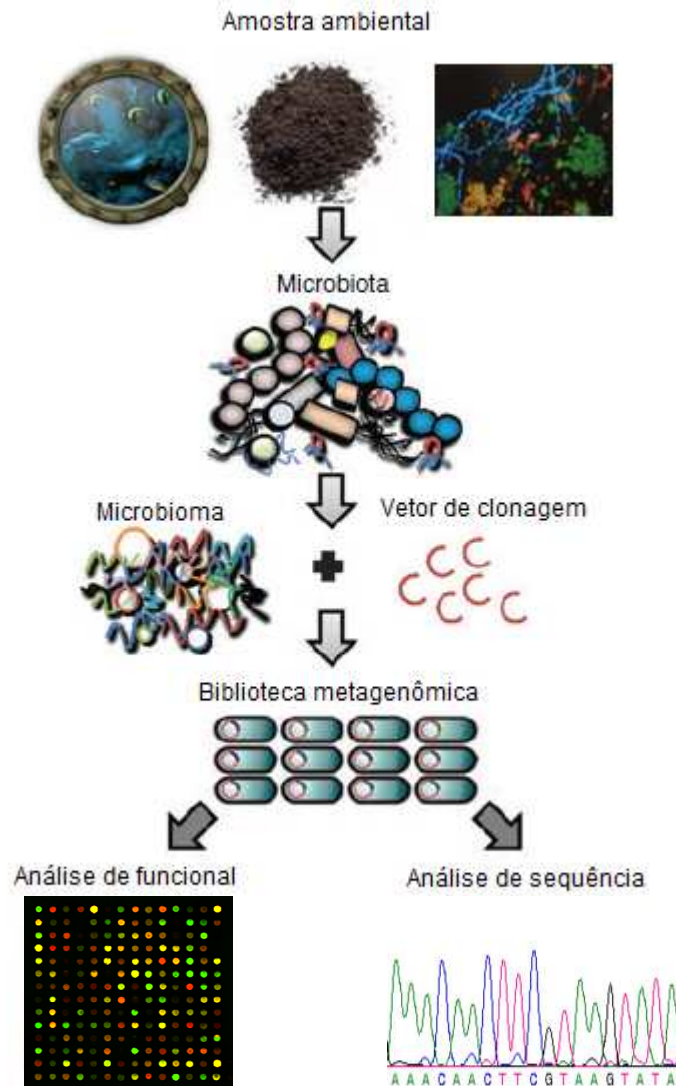


Figura 2. Esquema da construção de bibliotecas metagenômicas a partir de amostras ambientais, seguido das triagens de sequência e funcionais. (Figura modificada de *Letters in Applied Microbiology* 47 (2008) 361–366)

A estratégia funcional baseia a identificação de novos processos e proteínas na atividade biológica obtida pela expressão heteróloga, oferecendo a possibilidade de acessar sequências gênicas ainda desconhecidas e de detectar apenas enzimas que estão ativas (Sleator *et al.*, 2008). Entretanto, esta estratégia

pode ser comprometida pelos problemas decorrentes da expressão gênica heteróloga, ou seja, expressão dirigida por fatores de transcrição e tradução da célula hospedeira, que na maior parte das vezes é *Escherichia coli*, sendo que os genes e suas sequências reguladoras são originários de outro micro-organismo (Uchiyama & Miyazaki, 2009). A expressão heteróloga pode implicar em baixos níveis ou ausência de expressão da molécula-alvo, ou ainda, limitações no transporte da molécula sintetizada para fora da célula, comprometendo sua detecção nos ensaios. Warren *et al.* (2008) compararam os níveis de transcrição de DNA exógeno expresso em células de *E. coli* para verificarem a influência da expressão heteróloga. Utilizando RT-PCR e *microarray* os pesquisadores mediram a transcrição de sequências de DNA provenientes de *Haemophilus influenzae*, de *Pseudomonas aeruginosa* e do homem, clonadas em vetores BACs e propagadas em *E. coli*. Observaram que mais da metade dos genes de *Haemophilus influenzae* foram transcritos em *E. coli*. Isto se deve provavelmente ao fato de que *H. influenzae* apresenta regiões promotoras similares a sítios de ligação-RpoD s70 de *E. coli*. Comparado ao resultado de *H. influenzae*, um pequeno número de genes de *P. aeruginosa* foi transcrito e quase nenhuma transcrição foi observada para o DNA humano.

Apesar deste problema, o *screening* funcional de bibliotecas metagenômicas vem tendo maior sucesso na descoberta de genes catabólicos (Rondon *et al.*, 2000). Suenaga *et al.* (2007) construíram uma biblioteca metagenômica de fosmídeos para procurar novos genes que codificassem a enzima extradiol dioxigenase (EDOs), enzima relacionada com catabolismo de compostos aromáticos. Eles encontraram 91 clones positivos para esta enzima, dos quais 38 foram selecionados para sequenciamento, resultando na descrição de 43 genes,

dos quais 25 foram identificados como pertencentes a novas subfamílias EDO. Outros estudos de triagem de bibliotecas metagenômicas revelaram a detecção de 25 clones positivos para *naftaleno dioxigenase* (Ono, 2007), uma nova lipase alcali-termostável e genes de lipase (Meilleur *et al.*, 2009).

Portanto, para as comunidades microbianas que degradam poluentes, as bibliotecas metagenômicas constituem-se em uma ferramenta que abre a perspectiva de descoberta de novas vias catabólicas a partir de membros da microbiota, principalmente de micro-organismos fastidiosos ou ainda não-cultivados, ou mesmo daqueles que vivem em consórcios e não podem ser recuperados isoladamente.

2.5. Estratégias de enriquecimento de bibliotecas metagenômicas

Uma das frustrações associadas às análises de bibliotecas metagenômicas reside na baixa frequência de clones positivos. A fim de aumentar a proporção de clones que carregam o gene de interesse ou a característica desejada na biblioteca, diferentes estratégias foram desenhadas de modo a enriquecer as sequências de interesse antes da clonagem (Scholloss & Handelsman, 2003).

Uma estratégia de enriquecimento simples é selecionar para determinado conteúdo de GC dos genomas, como, por exemplo, para bactérias que têm alto conteúdo de GC no seu DNA, como os actinomicetos, que são de grande importância para aplicações biotecnológicas, incluindo biorremediação. Neste caso, o DNA ambiental é sujeito à ultra-centrifugação para separação da porção da

comunidade com alto GC. Apesar desta não ser uma estratégia muito refinada, pode aumentar a representação de certos genomas na biblioteca.

Outra estratégia seria o simples enriquecimento de populações específicas através da adição ao meio de cultivo de determinados compostos de interesse como únicas fontes de carbono e energia, a fim de se obter uma maior proporção de populações que utilizem este composto (Eyers *et al.*, 2004). Entcheva *et al.* (2001) combinaram as técnicas de enriquecimento microbiano e clonagem direta para construção de bibliotecas de cosmídeos com insertos de consórcios microbianos maiores de 30 Kb. As culturas de enriquecimento foram inoculadas com amostras do ambiente e grandes quantidades de avidina foram adicionadas para favorecer o crescimento de micro-organismos produtores de biotina, resultando em bibliotecas com número total de clones entre 6.000 a 35.000.

Os enriquecimentos também favorecem o crescimento de clones com grandes potenciais para identificação de genes ou enzimas de grande interesse biotecnológico, como foi observado por Knietzsch *et al.* (2003). Estes utilizaram meio enriquecido com glicerol e 1,2- propanodiol para o isolamento de genes que conferissem a oxidação de cadeias pequenas de poliol ou redução de carbonil. Foram identificados neste estudo nove plasmídios contendo genes para novas enzimas de álcool dioxigenase, oxiredutases, glicerol quinase, além de um gene que codificava uma proteína que não mostrou similaridade com nenhum produto gênico conhecido.

Uma estratégia de enriquecimento mais elaborada é baseada na utilização de substratos marcados com isótopos estáveis (SIP - *Stable Isotope Probing*). Esta metodologia consiste na utilização de átomos que são isotopicamente enriquecidos, como por exemplo, C^{13} ou N^{15} , para marcação de substratos que serão assimilados

pelas populações microbianas metabolicamente ativas (Murrell & Radajewski, 2000). O conceito de SIP-DNA está na diferença de densidade entre o DNA pesado (marcado isotopicamente) e o DNA leve (não marcado), que são separados pelo gradiente de densidade na ultra-centrifugação em cloreto de cério ou trifluoroacetato de cério. O DNA marcado e não marcado formam duas bandas distinguíveis, visíveis sob luz ultravioleta em presença de brometo de etídio (Friedrich, 2006). A detecção do isótopo incorporado pelos micro-organismos normalmente é realizada no DNA, mas também pode ser observado no RNA e nos fosfolípidios derivados de ácidos graxos (PLFAs).

2.6. Hidrolases

Os micro-organismos têm sido uma grande fonte de compostos naturais com atividades biológicas importantes em aplicações industriais e ambientais. Muitos compostos químicos encontrados em diferentes produtos comerciais têm sido lentamente substituídos por enzimas biológicas, que dentre outras vantagens são biodegradáveis, portanto minimizando os impactos causados ao meio ambiente (Meilleur *et al.*, 2009). Atualmente, devido à alta demanda do mercado, a biotecnologia tem aumentado a busca de biocatalistas de aplicação industrial e ambiental. Ferramentas de metagenômica, como extração direta do DNA de amostras ambientais, clonagem e subsequente triagem das bibliotecas obtidas, têm acelerado o processo de descobrimento de novos biocatalistas e metabólitos secundários (Elend *et al.*, 2006; Steele *et al.*, 2009). A metagenômica vem, portanto, revelando o grande potencial biocatalítico, antes inacessível, pertencente aos micro-organismos ainda não cultivados (Amann *et al.*, 1995).

Muitos dos biocatalistas metagenômicos encontrados têm mostrado novos espectros de utilização de substrato e formação de produto, sendo muitas vezes reações altamente estáveis, ocorrendo sob condições extremas, como altas ou baixas temperaturas ou pH, resultando em grande potencial de aplicação nos processos industriais (Steele *et al.*, 2009).

Dentre as enzimas de interesse biocatalítico estudadas, o grupo das enzimas hidrolíticas ou hidrolases é particularmente importante devido à ampla especificidade de substratos, alta estereo- e regioseletividade, independência de co-fatores, estabilidade e atividade em solventes orgânicos (Lammle *et al.*, 2007). Estas são aplicadas em vários setores industriais, como a área têxtil (amilase, celulase, pectinase), de detergentes (celulases, lipase, protease), de alimentos (celulases, lactases, pectinases, proteases) e de papel (lipases, xilanases). E enzimas como lipase, esterase, epóxido hidrolase e proteases também têm apresentado um grande potencial de aplicação nas estações de tratamento de efluentes industriais.

As proteases são enzimas que catalisam a hidrólise de ligações peptídicas, e estão dentre o grupo de hidrolases mais bem estudadas, pois têm uma grande importância na regulação de diversos processos do metabolismo celular e fisiológicos, como digestão, fertilização, apoptose, dentre outros (Chattopadhyaya & Yao, 2006). Apresentam também um grande valor econômico, devido à sua ampla aplicação industrial, como nas indústrias de detergentes, alimentos, farmacêuticas, diagnósticos e tratamento de resíduos. As proteases representam 40% das enzimas vendidas no mercado (Gupta *et al.*, 2002). Dentre as diversas aplicações destas enzimas nos setores industriais e ambientais, em sistemas de tratamento de efluentes especificamente, elas têm auxiliado muito na degradação do lodo

excedente gerado nas estações de lodos ativados, diminuindo o volume de material destinado ao processo de incineração. Estas enzimas agem no lodo, quebrando as macromoléculas de proteínas e facilitando a digestão do lodo por outros micro-organismos (Chenel *et al.*, 2008; Li *et al.*, 2009).

As lipases e esterases representam um grande grupo dentro das hidrolases que vem sendo extensamente estudado. São muito utilizadas em síntese orgânica devido a vários fatores, como versatilidade catalítica, disponibilidade comercial, baixo custo, não requerem co-fatores, atuam em uma faixa de pH relativamente grande e são muito estáveis (Henne *et al.*, 2000). Diversas bibliotecas metagenômicas têm sido construídas com DNA proveniente dos mais diversos tipos de amostras ambientais para triagem de esterases e lipases de aplicação industrial. Muitas novas enzimas com características termofílicas têm sido encontradas em inúmeras triagens funcionais (Demirjian *et al.*, 1999; Meilleur *et al.*, 2009). Bertram e colaboradores (2008) analisaram 350 novas lipases e esterases provenientes de amostras ambientais e selecionaram várias enzimas termoestáveis para serem utilizadas na síntese de triglicerídeos. Nos tratamentos de efluentes, a aplicação destas enzimas tem sido testada no processo de pré-tratamento do efluente para hidrolisar e dissolver os ácidos graxos. Estas são adicionadas ao sistema para auxiliar a degradação biológica, resultando também na redução do tempo gasto no processo de tratamento de efluentes (Lin *et al.*, 1997; Commarota & Freire, 2006).

As epóxido hidrolases são consideradas importantes devido ao seu papel na detoxificação celular e no metabolismo de inúmeros compostos biológicos (Pan *et al.*, 2010). São enzimas amplamente encontradas na natureza por serem produzidas por uma diversidade de organismos, tais como bactérias, fungos, plantas e mamíferos (Kotik *et al.*, 2010). Apresentam grande relevância industrial,

principalmente para o setor farmacêutico pela constante busca de epóxidos intermediários na síntese de substâncias biologicamente ativas, como por exemplo, óxido de *p*-nitroestireno, óxido de *trans* e *cis*-1-fenilpropano, óxido de *meta*- e *para*-coroestireno (Kotik *et al.*, 2010). Devido à importância destas enzimas, elas têm sido intensivamente procuradas em metagenomas microbianos. Henne *et al.* (2000) encontraram uma estireno-monoxigenase, enzima catalisa a epoxidação do estireno, proveniente de biblioteca metagenômica de amostra de solo. Kotik *et al.* (2010) identificaram dois genes codificando epóxido hidrolases funcionais em biomassa bacteriana recuperada de biofiltro para purificação de estireno, as quais mostraram promissoras características para hidrólise de vários epóxidos aromáticos.

2.7. REFERÊNCIAS BIBLIOGRÁFICAS

- Agency for Toxic Substances and Disease Registry. Disponível em: ><http://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=27><. Acesso em: 10 abril de 2010.
- Amann, R.I., Ludwig, W. & Schleifer, K. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiology Reviews*. 59: 143-169.
- Arai, H., Akahira, S., Ohishi, T., Maeda, M. & Kudo, T. 1998. Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: organization and regulation of the genes involved in phenol degradation. *Microbiology*. 144: 2895-2903.
- Barrios-Martinez, A., E. Barbot, B. Marrot, P. Moulin & N. Roche. 2006. Degradation of synthetic phenol-containing wastewaters by MBR. *Journal Membrane Science*. 281: 288-296.
- Bertram, M., Hildebrandt, P., Weiner, D.P., Patel, J.S., Bartnek, F., Hitchman, T.S. & Bornscheuer, U. T. 2008. Characterization of Lipases and Esterases from Metagenomes for Lipid Modification. *Journal American Oil Chemistry Society*. 85:47-53. Doi: 10.1007/s11746-007-1161-2).
- Breinig, S., Schiltz, E. & Fuchs, G. 2000. Genes Involved in Anaerobic Metabolism of Phenol in the Bacterium *Thauera aromatica*. *Journal of Bacteriology*. 182: 5849-5863.
- Buswell S. A. 1975 Metabolism of phenol and cresols by *Bacillus stearothermophilus*. *Journal of Bacteriology*. 72:248- 54.
- Chattopadhyaya, S & Yao, S. Q. 2006. In **Enzyme assays: High-throughput Screening, Genetic Selection and Fingerprinting**. Editores: Jean-Louis Reymond. Wiley-VCH, Weinheim, Alemanha. ISBN: 3-527-31095-9.
- Chenel, J. P., Tyagi, R. D. & Surampalli, R. Y. 2008. Production of thermostable protease enzyme in wastewater sludge using thermophilic bacterial strains isolated from sludge. *Water Science & Technology*. 57: 639-645.
- Cammarota, M. C. & Freire, D. M. G. 2006. A review on hydrolytic enzymes in the treatment of wastewater with high oil and grease content. *Bioresource Technology*. 97: 2195-2210.
- Cordova-Rosa, S. M., Dams, R. I., Cordova-Rosa, E. V., Radetski, M. R., Corrêa, A.X.R. & Radetski, C. M. 2009. Remediation of phenol-contaminated soil by a

- bacterial consortium and *Acinetobacter calcoaceticus* isolated from an industrial wastewater treatment plant. *Journal of Hazardous Materials*. 164: 61-66.
- Demirjian, D. C., Shah, P. C. & Morís-Varas, F. 1999. *Screening for novel enzymes. Topics in Current Chemistry*. 200: 1-29.
- Elend, C., Leggewie, C., Babiak, P., Carballeira, J. D., Steele, H. L., Reymond, J-L, Jaeger, K-E. & Streit, W. R. 2006. Isolation and biochemical characterization of two novel metagenome derived esterases. *Applied and Environment Microbiology*. 72: 3637-3645.
- El-Sayed, W. S., Ibrahim, M. K., Abu-Shady, M., El-Bih, F., Ohmura, M., Saik, H. & Ando, A. 2003. Isolation and characterization of phenol-catabolizing bacteria from a coke plant. *Bioscience, Biotechnology and Biochemistry*. 67:2026-2029.
- Eltis, L. D. & Bolin, J. T. 1996. Evolutionary relationships among extradiol dioxygenases. *Journal of Bacteriology*. 178: 5930-5937.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W. R. 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environment Microbiology*. 67: 89-99.
- Eyers, L., George, I., Schuler, L., Stenuit, B., Agathos, S. N. & Said El Fantroussi. 2004. Environmental genomics: exploring the unmined richness of microbes to degrade xenobiotics. *Applied Microbiology and Biotechnology*. 66: 123-130.
- Friedrich, M. W. 2006. Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. *Current Opinion in Biotechnology*. 17: 1-8.
- Giovannoni, S. J.; Britschgi, T. B.; Moyer, C. L. & Field, K. G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature*. 345: 60-63.
- Gupta, R.; Beg, Q. K. & Lorenz, P. 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied and Environment Microbiology*. 59: 15-32.
- Handelsman, J. 2004. Metagenomics: Application of Genomics to Uncultured microorganisms. *Microbiology and Molecular Biology Reviews*. 68: 669-685.
- Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G. & Daniel, R. 2000. *Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. Applied and Environment Microbiology*. 66: 3113-3116.

- Hugenholtz, P., Goebel, B. M. & Pace, N. R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*. 180: 4765-4774.
- Knietsch, A., Waschowitz, T., Bowien, S., Henne, A. & Daniel, R. 2003. Construction and *screening* of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Applied and Environment Microbiology*. 69: 1408-1416.
- Kotik, M., Stepanek, V., Grulich, M., Kyslik, P., Archelas, A. 2010. Access to enantiopure aromatic epoxides and diols using epoxide hydrolases derived from total biofilter DNA. *Journal of Molecular Catalysis B: Enzymatic*. Doi:10.1016/j.molcatb.2010.01.016. (*In press*).
- Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. & Rupp, S. 2007. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology*. 127: 575-592.
- Li, X., Ma, H., Wang, Q., Matsumoto, S., Maeda, T. & Ogawa, H. I. 2009. Isolation, identification of sludge-lysing strain and its utilization in thermophilic aerobic digestion for waste activated sludge. *Bioresource Technology* 100: 2475-2481.
- Lin, J. G., Chang, C. N. & Chang, S. C. 1997. Enhancement of anaerobic digestion of waste activated sludge by alkaline solubilization. *Bioresource Technology*. 62: 85-90.
- Lorenz, P., Liebeton, K., Niehaus, F. & Eck, J. 2002. *Screening* for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Current Opinion in Biotechnology*. 13: 572-577.
- Manefield, M., A. S. Whiteley, G. I. Griffiths & M. J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Applied Environment and Microbiology*. 68: 5367-5373.
- Mason, J. R. & Cammack, R. 1992. The electron-transport proteins of hydroxylating bacterial dioxygenases. *Annual Review of Microbiology*. 46: 277-305.
- Mariano, J. B. (2001) Impactos ambientais do refino de petróleo. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- Meilleur, C., Hupe, J-F., Juteau, P. & Shareck, F. 2009. Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *Journal of Industrial Microbiology and Biotechnology*. 36: 853-861.

- Merimaa, M. Heinaru, E., Liivak, M. Vedler, E. & Heinaru, A. 2006. Grouping of phenol hydroxylase and catechol 2,3-dioxygenase genes among phenol- and *p*-cresol-degrading *Pseudomonas* species and biotypes. *Archives Microbiology*. 186: 287-296.
- Mota, S. (1997) Introdução à Engenharia Ambiental, 1ª. ed. Rio de Janeiro, ABES.
- Murrell, J.C. & Radajewski, S. 2000. Cultivation-independent techniques for studying methanotroph ecology. *Research Microbiology*. 151: 807-814.
- Nozaki, M., Kotani, S., Ono, K. & Seno, S. 1970. Metapyrocatechase. III. Substrate specificity and mode of ring fission. *Biochimica Biophysica Acta*. 220, 213-223.
- Ono, A., Miyazaki R., Sota, M., Ohtsubo, Y., Nagata, Y. & Tsuda, M. 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Applied Microbiology and Biotechnology*. 74:501-10.
- Pan, H. F., Wen, N. B., Zhi, P. X., Jian, G. Z. & Yongquan, L. 2010. Molecular cloning and characterization of a cis-epoxysuccinate hydrolase from *Bordetella* sp. BK-52. *Journal of Microbiology and Biotechnology*. 20: 659-665.
- Pereira, R. C. 2005. Desenvolvimento e aplicação de uma nova metodologia para análise de especiação de selênio em efluentes hídricos de refinarias de petróleo. Tese de M.Sc., PUC- Rio de Janeiro, RJ, Brasil.
- Piras, P. R. F. 1993. Tratamento Biológico de Efluentes de Refinaria de Petróleo em Lagoas Aeradas em Série, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- Rondon, M. R., August, P. R., Bettermann, A. D., *et al.* 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied Environment and Microbiology*. 66: 2541-2547.
- Schloss, P. D. & Handelsman, J. 2003. Biotechnological prospects from metagenomics. *Current Opinion*. 14: 303-310.
- Sleator, R. D., Shortall, C. & Hill, C. 2008. Metagenomics. *Letters in Applied Microbiology*. 47: 361-366.
- Steele, H.L., Jaeger, K-E., Daniel, R. & Streit, W. R. 2009. Advances in Recovery of Novel Biocatalysts from Metagenomes. *Journal of Molecular Microbiology Biotechnology*. 16: 25-37.

- Suenaga, H., Ohnuki, T. & Miyazaki, K. 2007. Functional *screening* of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology*. 9: 2289–2297.
- Sueoka, K., H. Satoh, M. Onuki, & T. Mino. 2009. Microorganisms involved in anaerobic phenol degradation in the treatment of synthetic coke-oven wastewater detected by RNA stable-isotope probing. *FEMS Microbiology Letter*. 291: 169-174.
- Uchiyama, T. & Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient *screening*. *Current Opinion in Biotechnology*. 20: 616-622.
- Ward, D.M., Weller, R. & Bateson, M. M. 1990. 16S RNAr sequences reveal numerous uncultured microorganisms in a natural community. *Nature*. 345:63-65.
- Warren, R. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S. & Holt, R. A. 2008. Transcription of foreign DNA in *Escherichia coli*. *Genome Research*. 18: 1798-1805.
- Watanabe, K., Teramoto, M. Futamata, H. and Harayama, S. 1998. Molecular Detection, Isolation, and Physiological haracterization of Functionally Dominant Phenol-Degrading Bacteria in Activated Sludge. *Applied Environment and Microbiology*. 64: 4396-4402.
- Wei, G., Yu, J., Zhu, Y., Chen, W., Wang, L. 2008. Characterization of phenol degradation by *Rhizobium* sp. CCNWTB 701 isolated from *Astragalus chrysopteru* in mining tailing region. *Journal of Hazardous Materials*. 151:111-117.
- Yap, L. F, Lee, Y. K. & Poh, C. L. 1999. Mechanism for phenol tolerance in phenol-degrading *Comamonas testosteroni* strain. *Applied Microbiology and Biotechnology*. 51: 833-840.

3. OBJETIVOS

3.1. Objetivos gerais

- ✓ Exploração da microbiota presente em lodo de efluente de refinarias através da construção de uma biblioteca metagenômica de cultivo enriquecido com fenol e subsequente triagem visando à obtenção de clones com capacidade de degradação deste composto;
- ✓ Avaliação das sequências gênicas responsáveis pela degradação de compostos aromáticos recalcitrantes visando contribuir para o conhecimento da diversidade destes genes catabólicos em efluentes de refinarias e compreender as vias metabólicas envolvidas nos processos de degradação de fenol e derivados.

3.2. Objetivos específicos

- ✓ Caracterização da diversidade de lodos proveniente de diferentes refinarias de petróleo pela construção de bibliotecas de genes de RNAr 16S;
- ✓ Aclimação e enriquecimento do lodo com fenol como única fonte de energia;
- ✓ Construção das bibliotecas metagenômicas de DNA de alto peso molecular enriquecidas com fenol;

- ✓ Triagem baseada sequência, por PCR, dos clones de fosmídeos para presença do gene que codifica a enzima fenol hidroxilase;
- ✓ Triagem funcional para os clones capazes de crescer em meio mineral acrescido de fenol como única fonte de carbono;
- ✓ Ensaios por cromatografia gasosa com clones capazes de crescer em meio mineral acrescido de fenol para degradação de fenol;
- ✓ Triagens funcionais para enzimas hidrolíticas: lipase, esterase, protease e epóxido hidrolase;
- ✓ Pirosequenciamento dos clones derivados das bibliotecas metagenômicas para a caracterização da diversidade molecular presente no lodo bacteriano;
- ✓ Pirosequenciamento de clones que apresentaram resultados positivos nas triagens para fenol, para identificação de novos genes relacionados à degradação de compostos fenólicos.

4. APRESENTAÇÃO DA TESE

Este trabalho se baseou na construção de bibliotecas metagenômicas utilizando DNA proveniente de lodo de biorreatores pilotos pertencentes a estações de tratamento de efluentes de refinarias de petróleo. As bibliotecas foram construídas após período de aclimação do lodo ao fenol e o DNA de alto peso molecular foi extraído e clonado em fosmídeos. Após obtenção das bibliotecas metagenômicas, foram realizadas triagens baseadas em similaridade de sequência, por reações de PCR, e funcionais, através de ensaios em meio líquido e em placas. Nos **Capítulos 1 e 2** serão apresentados os resultados da caracterização da diversidade bacteriana presente em lodos provenientes dos sistemas de tratamento de efluentes de refinaria de petróleo, os quais já foram publicados em revistas internacionais indexadas. No **Capítulo 3** será apresentado o perfil funcional da biblioteca metagenômica obtido por pirosequenciamento de todos os 13.200 clones obtidos. No **Capítulo 4** serão apresentados os resultados referentes às triagens para degradação de fenol dos 13.200 clones fosmidiais e ao pirosequenciamento de 100 clones positivos selecionados nos ensaios de triagem. No **Capítulo 5**, serão apresentados os resultados de triagens das bibliotecas para as enzimas hidrolíticas lipase, esterase, epóxido hidrolase e protease. Os resultados do trabalho como um todo serão discutidos de maneira integrada no item **Discussão Geral**, e, finalmente, na **Conclusão** serão apresentadas as considerações finais deste trabalho de Tese.

**CAPÍTULO 1 - MONITORING THE BACTERIAL COMMUNITY DYNAMICS IN A PETROLEUM
REFINERY WASTEWATER MEMBRANE BIOREACTOR FED WITH A HIGH
PHENOLIC LOAD**

Cynthia C. Silva, Aline F. Viero, Ana Carolina F. Dias, Fernando D. Andreote, Ederson C. Jesus, Sergio O. De Paula, Ana Paula R. Torres, Vania M. J. Santiago & Valeria M. Oliveira. **Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load.** *Publicado no Journal of Microbiology and Biotechnology.*

J. Microbiol. Biotechnol. (2010), 20(1), 17–25
doi: 10.4014/jmb.0906.06001
First published online 11 November 2009



Monitoring the Bacterial Community Dynamics in a Petroleum Refinery Wastewater Membrane Bioreactor Fed with a High Phenolic Load

Silva, Cynthia C.^{1*}, Aline F. Viero², Ana Carolina F. Dias³, Fernando D. Andreote⁴, Ederson C. Jesus⁵, Sergio O. De Paula³, Ana Paula R. Torres⁶, Vania M. J. Santiago⁶, and Valeria M. Oliveira¹

¹Division of Microbial Resources, Research Center for Chemistry, Biology and Agriculture (CPQBA), Campinas University - UNICAMP, CP 6171, CEP 13081-970, Campinas, SP, Brazil

²Alberto Luiz Coimbra Institute - Graduate School and Research in Engineering (COPPE), Federal University of Rio de Janeiro, Rio de Janeiro, RJ, CEP 21945-970, Brazil

³Department of General Biology, Federal University of Viçosa - UFV, Viçosa, MG, CEP 36571-000, Brazil

⁴Brazilian Agricultural Research - Embrapa, CP 69, Jaguariúna, SP, CEP 13820-000, Brazil

⁵Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824, U.S.A.

⁶PETROBRAS R&D Center, Cidade Universitária, Quadra 7, Ilha do Fundão, Rio de Janeiro, RJ, CEP 21949-900, Brazil

Received: June 1, 2009 / Revised: August 23, 2009 / Accepted: September 3, 2009

The phenolic compounds are a major contaminant class often found in industrial wastewaters and the biological treatment is an alternative tool commonly employed for their removal. In this sense, monitoring microbial community dynamics is crucial for a successful wastewater treatment. This work aimed to monitor the structure and activity of the bacterial community during the operation of a laboratory-scale continuous submerged membrane bioreactor (SMBR), using PCR and RT-PCR followed by denaturing gradient gel electrophoresis (DGGE) and 16S rRNA libraries. Multivariate analyses carried out using DGGE profiles showed significant changes in the total and metabolically active dominant community members during the 4-week treatment period, explained mainly by phenol and ammonium input. Gene libraries were assembled using 16S rDNA and 16S rRNA PCR products from the fourth week of treatment. Sequencing and phylogenetic analyses of clones from the 16S rDNA library revealed a high diversity of taxa for the total bacterial community, with predominance of *Thauera* genus (ca. 50%). On the other hand, a lower diversity was found for metabolically active bacteria, which were mostly represented by members of Betaproteobacteria (*Thauera* and *Comamonas*), suggesting that these groups have a relevant role in the phenol degradation during the final phase of the SMBR operation.

Keywords: Sludge, bacterial community, DGGE, 16S rRNA gene library, multivariate analysis

Phenol and its derivatives are known to be the most common toxic environmental pollutants found in many industrial effluents, such as cellulose source plants, industrial resin manufacturing, petrochemical, pharmaceutical, plastic, and varnish industries [8, 39]. These compounds are toxic to animals at relatively low concentrations (5–25 mg/l) [6, 41]. Recent environmental regulations have become more and more restrictive, making the efficient removal of such compounds from industrial aqueous effluents of great concern [41]. Biological treatments have been the most employed strategy for phenol removal in effluents, mainly because of their relatively low costs when compared with physicochemical methods and also of their potential to promote the complete mineralization of such toxic compounds [39].

It is well known that the degradation efficacy of a biological treatment depends on the stability of the microbial community involved in the decontamination process [24]. Recently, bacterial phenol degradation in bioreactor conditions has been successfully demonstrated by several authors [1, 4, 7, 33]. However, despite the great relevance of biological treatments to industrial wastewater plants, most of the related studies have focused mainly on treatment performance and associated parameters, without giving proper attention to the microbial community involved. The understanding of the ecology, structure, and dynamics of those populations are major points for the success of the wastewater treatment and may provide a basis for future application of methods for *in situ* analyses of the function of specific microorganisms, helping to establish the links between microbial biodiversity and process stability in wastewater treatment plants [22].

*Corresponding author
Phone: +55-19-2139-2874; Fax: +55-19-2139-285;
E-mail: canedasilva@yahoo.com.br

Methods based on 16S rDNA analysis provide extensive information about the composition, spatial distribution, and relative abundance of taxonomical groups present in a specific environment [32]. Fingerprint techniques, such as denaturing gradient gel electrophoresis (DGGE), have proved to be fast and straightforward tools for monitoring shifts in microbial communities submitted to pollutant impact, and can be used to evaluate the genetic diversity and the temporal or spatial dynamics of the community [25].

This study aimed to monitor the dynamics of the bacterial community present in a submerged membrane bioreactor (SMBR) fed with high phenolic load, by using PCR and RT-PCR followed by DGGE, in an attempt to establish correlations between shifts in bacterial community and the phenolic stream and evaluate a possible impact on specific populations. The assembly and analysis of 16S clone libraries were subsequently employed for a detailed characterization of the bacterial community composition.

MATERIAL AND METHODS

Membrane Bioreactor and Operation

Experiments were carried out in a laboratory-scale continuous submerged membrane bioreactor (SMBR), consisting of an acrylic tank with volume of 4.4 l with a polyetherimide hollow fiber membrane skein ($2.78 \times 10^{-2} \text{ m}^2$ area) inserted. Filtration was alternated with air backwashing in order to keep the permeate flux constant. Air was used as the oxygen source, supplied through a porous stone diffuser placed in the bottom of the tank. Further details on the membranes, equipment, and operating conditions were reported elsewhere [38]. Before seeding the bioreactor, sludge was sampled from a Petrobras oil refinery in Brazil, and settled to achieve the total suspended solids concentration of 10 g/l. Acclimatization of the sludge took place for 30 days, when the SMBR processed the oily stream (OS) only. After the acclimatization step, phenolic wastewater (PW) was added to the oily stream (1:6, PW:OS), to increase the concentration of recalcitrant compounds, resulting in an average phenol feed of $86 \pm 2 \text{ mg/l}$, as previously described in Viero *et al.* [38]. The oily wastewater was pre-treated in a sand filter to reduce the free oil and grease content and, further, sodium tripolyphosphate was added as a phosphorus source to improve the microbial activity. Both wastewater and sludge were sampled in the same oil refinery.

Sampling and Nucleic Acid Extraction

Sludge samples were collected weekly during the 30-day period of the SMBR operation. The four sludge samples were collected in triplicate using sterilized microtubes and submitted to co-extraction of DNA and RNA by the TRIzol method (Invitrogen, Grand Island, NY, U.S.A.), according to the manufacturer's instructions. Aliquots of 150 μl of each nucleic acid sample were treated with DNase (Fermentas, U.S.A.) for 1 h at 37°C, in order to remove contaminant DNA for RT-PCR assays. The remaining nucleic acid solution, 50 μl , was used in PCR reactions for subsequent analysis of total bacterial community.

PCR and RT-PCR Amplification

PCR and RT-PCR targeting 16S rDNA and rRNA, respectively, were performed using the universal primers 968f (attached to a 40-nucleotide GC-rich sequence) and 1401r [13], which are homologous to conserved bacterial 16S rDNA regions. Reverse transcription of 16S rRNA and subsequent PCR amplification were performed after checking the absence of DNA in the total RNA extract. The first cDNA strand was synthesized in a 50- μl reaction containing 0.1 mM dNTP mix, 7.5 U of SuperScript (Invitrogen, Grand Island, NY, U.S.A.), 5 μl of 10 $\times$ buffer (GIBCO-BRL, Grand Island, NY, U.S.A.), 25 pmol of p(dN)₆ random hexamer primers (GE Healthcare, U.S.A.), and 2 μg of extracted RNA. The reaction was incubated at 42°C for 1 h, followed by 1 min at 70°C, to promote the enzyme denaturation. The RT-PCR and PCR amplifications were performed in 50- μl reactions containing 1 μl of cDNA and 30 ng of DNA, respectively, 5 μl of 10 $\times$ Tris-HCl reaction buffer, 1.5 mM MgCl₂, 0.4 μM primers 968f and 1401r, 0.2 mM dNTP mix, and 2 U *Taq* DNA polymerase (Invitrogen, Grand Island, NY, U.S.A.). The PCR amplifications were done using an initial denaturation step of 5 min at 94°C, 10 cycles of 1 min at 94°C, 30 s at 58°C, decreasing 1°C each cycle, and 2 min at 72°C, followed by another 25 cycles of 1 min at 94°C, 30 s at 53°C and 2 min at 72°C. DNA-free RNA was used as the control in the PCR reactions to check the presence of contaminating DNA in the RNA preparations. The amplicons were first checked on 1.2% agarose gels prior to DGGE analyses.

DGGE and Statistical Analyses

DGGE analysis of RT-PCR and PCR products was carried out in the D-Code Universal Mutation Detection System (Bio-Rad, U.S.A.) using a linear denaturing gradient of urea and formamide ranging from 45% to 65% [100% denaturant corresponding to 7 M urea and 40% (v/v) deionized formamide]. Gels (6% polyacrylamide) containing about 250 ng of the PCR or RT-PCR products for each sample, in triplicate, were run at 50 V and 60°C for 16 h in 1 $\times$ TAE buffer. Gels were stained with SYBR Green 1 $\times$ solution and documented under UV light.

DGGE patterns were analyzed using GelCompar v. 4.1 (Applied Maths, Kortrijk, Belgium). Gels were normalized and UPGMA-based dendrograms constructed from Pearson (product-moment) correlation coefficient matrices [26]. In addition, band relative surfaces were estimated and used to allow comparison between different samples.

The multivariate analyses were conducted using Canoco (Canoco v. 4.5; Biometrics, Wageningen, The Netherlands). Input data consisted of combining the table containing band relative surfaces with the table containing values for chemical compound removal. Bands were considered as bacterial groups and the relative areas as the frequency of group occurrence. The chemical removal percentages for phenol, chemical oxygen demand (COD), and ammonium (NH₄⁺-N) were determined for the same SMBR samples and used to construct the table of environmental factors.

Prior to multivariate analyses, bacterial group tables were generated from DGGE patterns based on DNA and RNA. They were then subjected to a detrended correspondent analysis (DCA) to estimate the data distribution and choose the appropriate model for further correlations between bacterial groups and chemical data. Values of data distribution around 3.0 led to the choice of the redundancy analysis (RDA) and values above 4.0 led to the choice of canonical correlation analysis [2, 35].

16S rDNA and 16S rRNA Libraries and Sequencing

Two libraries, 16S rDNA and 16S rRNA, were constructed for the fourth-week bioreactor samples. The 16S rDNA library was constructed following the amplification of partial 16S rRNA genes from total community DNA, using 27f and 1100r primers [20]. The PCR products were gel-purified using GFX PCR DNA and the Gel Band Purification Kit (GE Healthcare, U.S.A.), according to the manufacturer's recommendations, and used for cloning. The 16S rRNA library was constructed using the region excised from the RT-PCR–DGGE gel that showed the greatest number of dominant bands. Briefly, one block of acrylamide gel containing fragments of interest were excised and DNA was extracted by using the “crush and soak” method [28]. Pellets were suspended in 15 µl of Milli-Q water for subsequent PCR and cloning, using 968f and 1400r primers, without CG-clamp.

After purification, 16S rDNA- and 16S rRNA-derived PCR products (50 ng) were ligated into pGEM T-Easy Vector (Promega, Madison, WI, U.S.A.) according to the manufacturer's instructions and transformed into *Escherichia coli* JM109 competent cells. The 16S rRNA gene inserts were amplified from selected clones using the universal M13 forward and reverse primers. The PCR products were purified as previously described for subsequent automated sequencing in the MegaBACE DNA Analysis System 1000 (GE Healthcare). The sequencing was carried out using M13 primers and the DYEnamic ET Dye Terminator Cycle Sequencing Kit for automated MegaBACE 500 system (GE Healthcare), in accordance with the recommendations of the manufacturer.

Phylogenetic Analysis

Partial 16S rRNA gene sequences obtained from clones were assembled in a contig using the phred/Phrap/CONSED program [11, 12]. The identification was achieved by comparing the contiguous 16S rRNA gene sequences obtained with data from reference and type strains, as well as environmental clones, available in the public databases GenBank (<http://www.ncbi.nlm.nih.gov>) and RDP (Ribosomal Database Project - Release 10; <http://rdp.cme.msu.edu/>) using BLASTn and Classifier routines, respectively. The sequences were aligned using the CLUSTAL X program [36] and analyzed with MEGA software Version 4.0 [34]. Evolutionary distances were derived from the sequence-pair dissimilarities, calculated as implemented in MEGA using the DNA substitution model reported by Kimura [19]. The phylogenetic reconstruction was done using the neighbor-joining (NJ) algorithm [27], with bootstrap values calculated from 1,000 replicate runs, using the routines included in the MEGA software.

GenBank Accession Numbers

Sequences obtained from the clone libraries were deposited in the GenBank database under the accession numbers FJ438995 to FJ439118.

Analytical Methods

Standard Methods [31] procedures were used to measure COD (chemical oxygen demand) and ammonium concentration ($\text{NH}_4^+\text{-N}$). TOC (total organic carbon) determination was performed in a TOC analyzer (Model 5000, Shimadzu, Japan). Phenol concentrations were

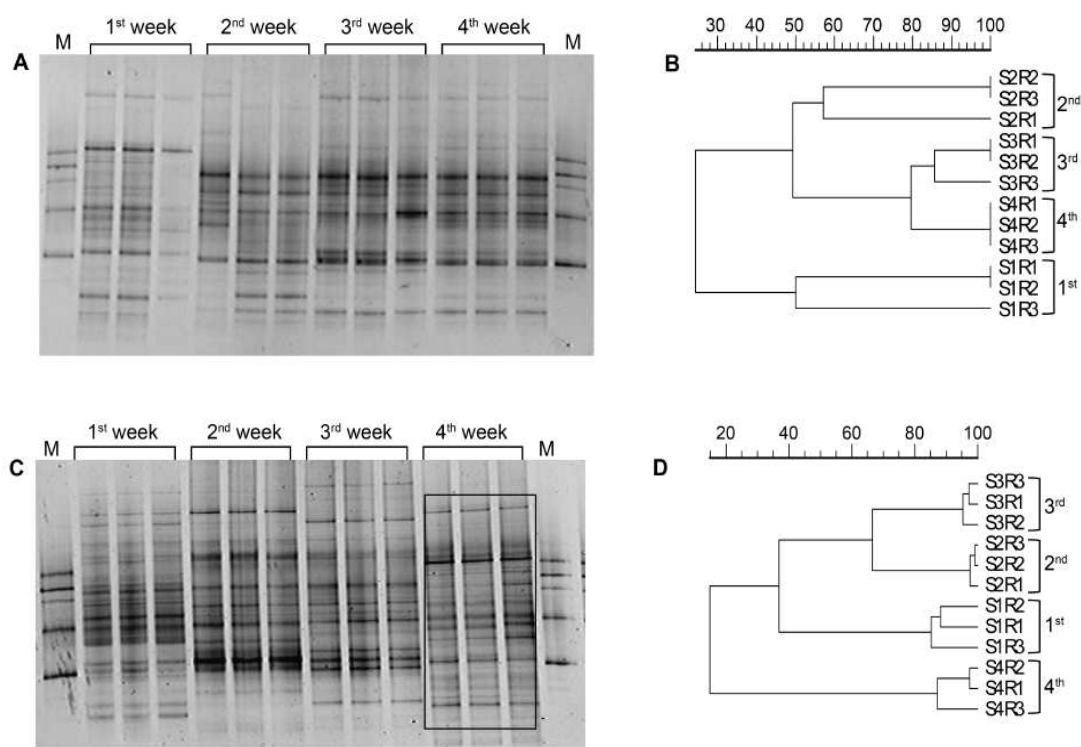


Fig. 1. UPGMA clustering by pattern similarity of DGGE bands obtained from MBR sludge samples during four weeks with phenolic feeding. A. 16S rDNA; B. 16S rRNA. Lanes: M ladder. Square: DGGE prominent bands excised for library construction of the metabolically active bacterial community.

analyzed with R-8012 kits supplied by Chemetrics, Inc. The analyses were performed weekly.

RESULTS

The structure and activity of dominant populations of the SMBR sludge were monitored, respectively, by PCR–DGGE and RT–PCR–DGGE during 30 days of bioreactor operation with high phenolic load feed. The extraction of nucleic acids and PCR amplifications, including RT–PCR, were successful for all samples. RNA yielded no PCR products when subjected to amplification reaction, indicating the absence of residual DNA in the RNA preparations (data not shown).

The DGGE profiles revealed pronounced changes in the diversity of total and metabolically active bacterial communities during the evaluated period (Fig. 1A and 1C). Cluster analyses of total community fingerprints yielded two main clusters presenting <30% similarity with each other (Fig. 1B). The first cluster corresponded to the 1st-week replicate samples, and the second group corresponded to the 2nd-, 3rd-, and 4th-week samples. The 2nd week samples showed around 50% similarity with the 3rd- and 4th-week samples, which in turn showed 80% similarity between each other, revealing a progressive differentiation of the bacterial populations along the bioreactor operation period (Fig. 1A and 1B).

The DGGE fingerprints of the metabolically active community revealed three main clusters (Fig. 1D). The first cluster was formed by the samples of the 2nd and 3rd

Table 1. Percentage of chemical removal in the SMBR bioreactor evaluated during the four weeks of the experiment.

Removals	1 st week	2 nd week	3 rd week	4 th week
Ammonium	75.0%	45.5%	40.0%	72.0%
Phenol	99.6%	99.5%	99.6%	98.9%
COD	61.0%	61.2%	65.6%	68.2%
TOC	78.5%	70.2%	70.0%	75.3%

weeks, which clustered with 66% similarity. The 1st-week replicate samples formed the second cluster, which showed 35–40% similarity with the first cluster. The third cluster was formed by the 4th-week samples, which showed <20% similarity with the other two clusters. The band profiles observed in the RT–PCR–DGGE gel indicated a different dynamics for the metabolically active populations and more profound changes along the 4 weeks of monitoring, based on similarity values, when compared with the total bacterial populations of the sludge samples.

According to the results of COD, TOC, ammonium, and phenol analyses, for feed and permeate samples, the SMBR biomass showed good efficacy in the removal of compounds, especially phenol, present in the wastewater during a 30-day operation (Table 1). Canonical ordination was applied in order to evaluate our hypothesis that the high phenol load might be determining the variations observed for both total and metabolically active bacterial populations in the SMBR. Values of DCA, 2.87 and 3.17, for DNA- and RNA-based DGGE, respectively, revealed the linear distribution of data, leading to the application of RDA as the best model for multivariate analysis [35].

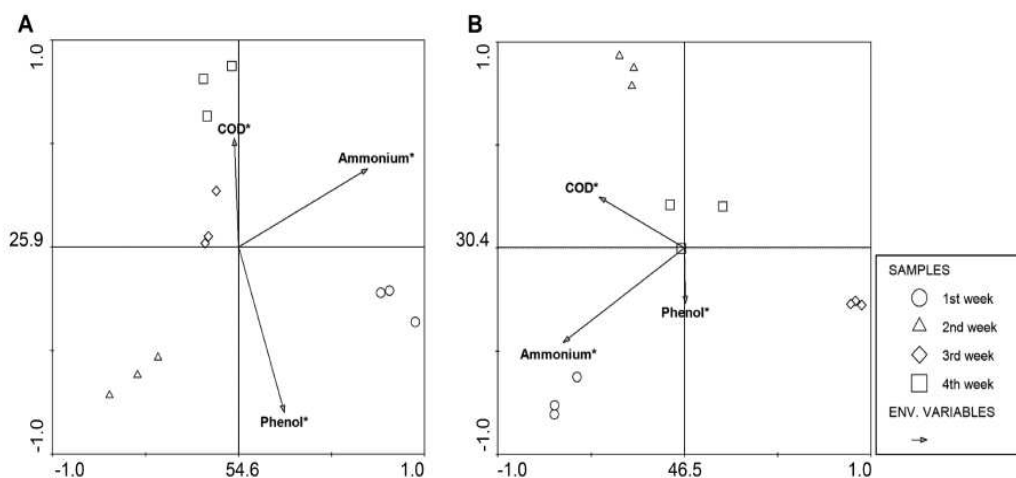


Fig. 2. Redundancy analyses (RDA) based on DNA (A) and RNA (B) DGGE patterns correlations with the removing of chemical compounds.

Values in axes indicate the percentage of species-variable correlation. Chemical quantifications are shown (environmental variables), and the indication with * mark these significant factors according to the Monte Carlo permutation test with 499 unrestricted permutations under full model selection. COD means chemical oxygen dissolved.

RDA showed that $\text{NH}_4^+\text{-N}$, phenol, and COD were the major factors explaining data variation. Higher $\text{NH}_4^+\text{-N}$ and phenol concentrations were mainly related with total and metabolically active communities of the 1st week. Higher COD factor was related with the total community of the 3rd and 4th weeks and also with the metabolically active community of the 1st and 2nd weeks (Fig. 2).

The Monte Carlo permutation test, based on the RDA model, showed that *p* values for the removal of phenol, ammonium, and COD were significant in both DNA- and RNA-based DGGE analyses (Table 2). Lambda1 values indicated that, in DNA-based analysis, environmental factors explained data variation as follows: ammonium>phenol>COD. In the case of RNA-based analysis, the variation was explained as ammonium>COD>phenol (Table 2).

The 16S rDNA and 16S rRNA libraries, assembled from DNA and RNA templates, respectively, from the 4th-week samples, allowed the identification of the bacterial populations selected in the last phase of the SMBR operation. A total of 117 clone sequences of the 16S rDNA library presented good quality for subsequent sequence alignment and phylogenetic inference. BLASTn analyses revealed that clones were affiliated to the phyla Bacteroidetes (2.5%), Nitrospirae (2.5%), and Proteobacteria (95%), the latter being represented mainly by the class Betaproteobacteria (ca. 83% of all clones). The betaproteobacteria were Hydrogenophilales, and Burkholderiales. The first one was the most abundant (56% of all clones) and exclusively represented by *Thauera* spp.-related clones (Fig. 3A).

For the 16S rRNA library, 80 clones were subjected to phylogenetic analyses, which showed relationships with the phyla Planctomycetes, Bacteroidetes, and Proteobacteria. Similarly to 16S rDNA library results, the clones were mainly affiliated to the class Betaproteobacteria (61.3%), which was represented by the orders Rhodocyclales and Burkholderiales. Clones belonging to the order Rhodocyclales were related to the genus *Thauera* (20% of all clones) and the ones belonging to the order Burkholderiales were related to *Comamonas* spp. (29%). The phylum Bacteroidetes was also well represented, accounting for 25% of all clones in the 16S rRNA library (Fig. 3B). The comparison of the

bacterial composition between the two libraries revealed that, except for a few clones affiliated to Planctomycetes, the taxonomic groups observed in the 16S rRNA library consist of a subset of the 16S rDNA library (Fig. 4).

DISCUSSION

Biological treatment has been usually preferred for large-scale phenol removal in wastewaters, since it allows for phenol degradation to innocuous low molecular weight compounds. In the petrochemical industry and oil refineries, the presence of high concentrations of phenol reduces significantly the biological degradation of the other components, in addition to the significant environmental toxicity hazard [6].

In this study, it was verified that a SMBR biomass was able to degrade almost all phenol added in the bioreactor. Studies have revealed that the use of the membrane bioreactor process (MBR) has been effective in the biological treatment of wastewater with high concentration of phenol [1, 6]. The main advantage of MBR is that it can achieve complete sludge retention (including the floating populations), resulting in high sludge concentrations and long sludge retention times (SRTs) [42].

Molecular approaches based on PCR have helped in the of mechanisms to improve the control of the biological process [16, 21, 23]. In this study, the monitoring of the SMBR biomass was done by DGGE fingerprints, which clearly distinguished between the profiles of total and metabolically active communities, assessed by DNA and RNA, respectively. The results suggested a greater change for the active populations during phenol degradation, probably because the responses at the RNA level are more rapid and have greater amplitude than those at the DNA level [10]. Similar results were found by Ibekwe and Lyon [15], who observed distinct clustering of DNA- and RNA-derived DGGE profiles when evaluating alterations in the bacterioplankton community composition of a drinking water aquifer.

Redundancy analysis was applied in order to test the correlation between community structure and physicochemical factors of the bioreactor. COD, phenol, and $\text{NH}_4^+\text{-N}$ removal results were shown to be significantly related with the changes in the bacterial communities. The phenol and $\text{NH}_4^+\text{-N}$ were related mainly with the total and active bacterial communities of the 1st week. Possibly, during this week, the bacterial populations had to cope with the initial selective pressure owing to high load of phenol and the presence of $\text{NH}_4^+\text{-N}$ in the influent. During the other three weeks, the populations were progressively differentiated, what might reflect a gradual acclimation to the new conditions of the bioreactor.

Table 2. Significance of chemical compounds variation on the diversity of the bacterial community estimated by multivariate analysis.

	DNA-based		RNA-based	
	Lambda1	<i>p</i> value	Lambda1	<i>p</i> value
Phenol	0.21	0.002*	0.16	0.030*
Ammonium	0.30	0.006*	0.33	0.018*
COD	0.16	0.002*	0.24	0.002*
TOC	0.30	0.422	0.26	0.098

*Significance is considered for values of *p*<0.05.

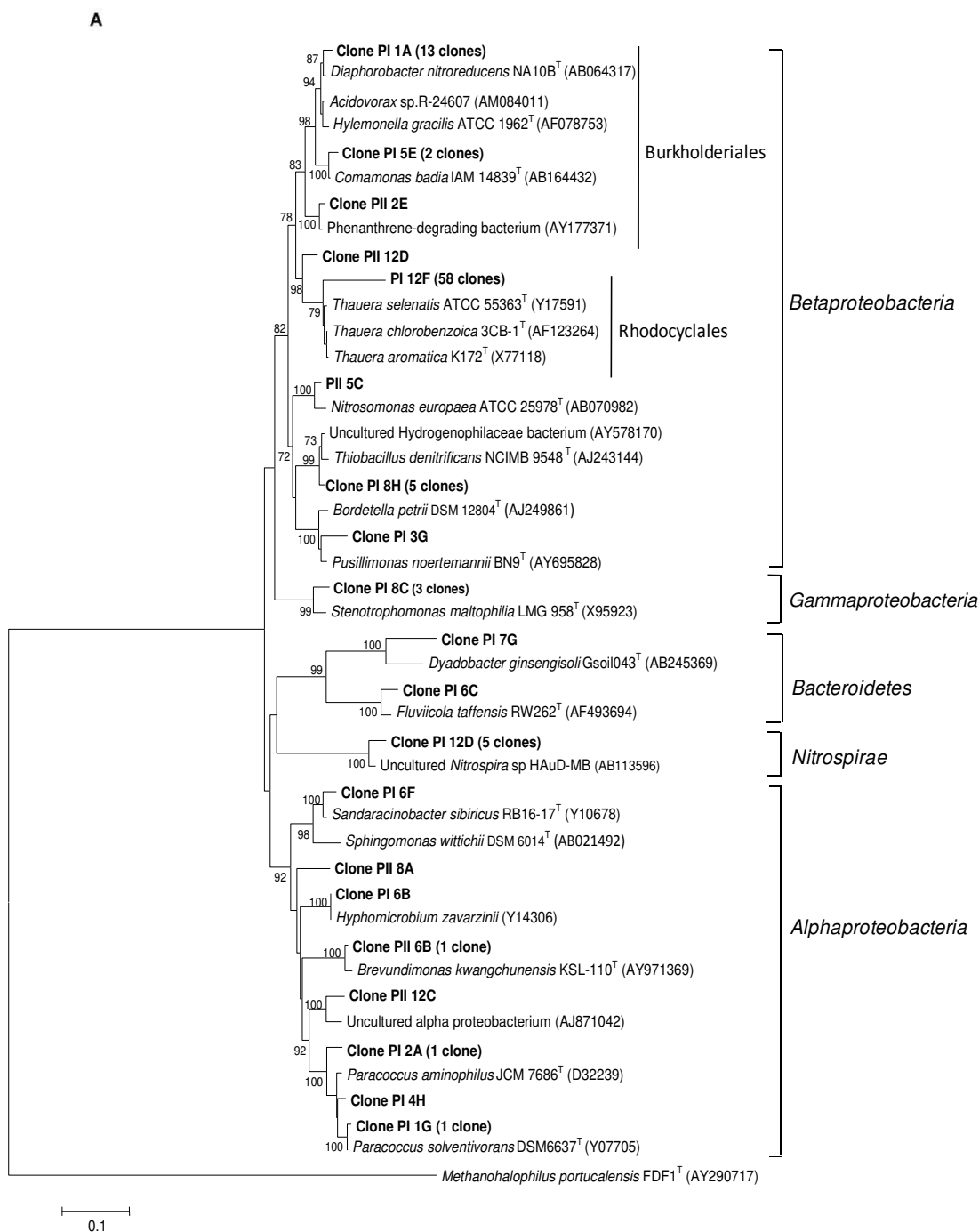


Fig. 3. Phylogenetic analysis of partial 16S rRNA gene sequences from fourth-week MBR bioreactor obtained with (A) PCR systems and (B) RT-PCR-DGGE systems.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. The number of clones in parenthesis showed $\geq 97\%$ similarity with the clone represented in the branch. GenBank accession numbers are listed after species names. *Methanohalophilus portucalensis* was used as the outgroup.

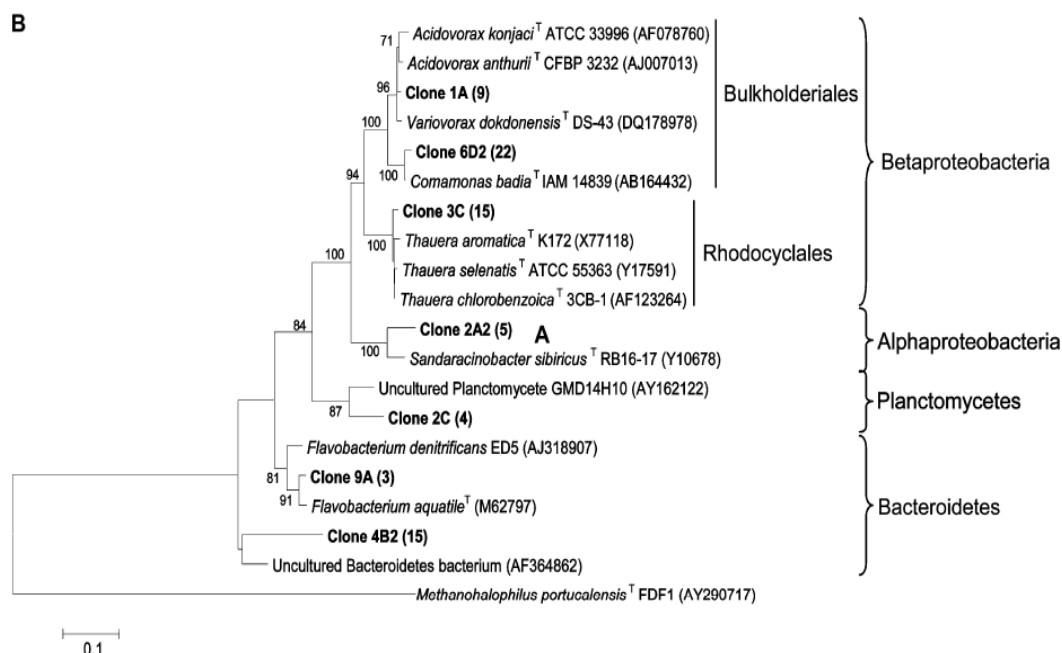
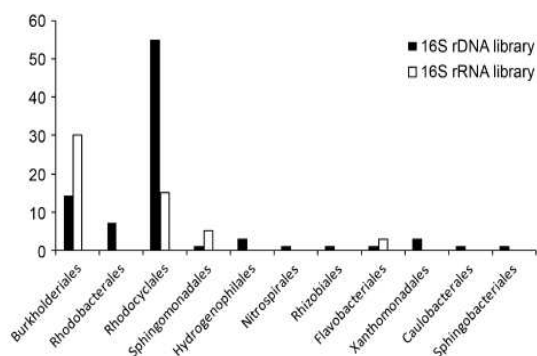


Fig. 3. Continued.

The COD is used as an index of the total organic carbon in a sample, and may be used to quantify all organic substances with potential to be used as nutrients by microorganisms [4]. The COD factor showed a higher correlation with the variation in the RNA-based patterns of the 1st and 2nd weeks, suggesting that metabolic responses such as the degradation of organic matter are first noticed at the rRNA level than at the rDNA level.

Amongst the factors analyzed, the $\text{NH}_4^+\text{-N}$ results explained more than those of phenol did the data variation of total and metabolically active populations. This is

Fig. 4. Abundance of clones belonging to each bacterial order in libraries based on DNA or RNA extracted from a SMBR bioreactor at the 4th sampling.

probably due to the relevant role of ammonium in biological processes conducted by bacterial communities inside the bioreactor. The ammonium is important for all microorganisms, since it is used in a variety of metabolic pathways related to energy metabolism and biosynthesis [40]. Nonetheless, the higher correlation of the $\text{NH}_4^+\text{-N}$ factor with the biological data does not exclude the important role of phenol in modulating the composition of bacterial communities present in the bioreactor.

Phylogenetic analyses of 16S rDNA and 16S rRNA clone libraries revealed that bacterial communities of the fourth week were mostly represented by members of the class Betaproteobacteria, which were related to *Thauera* spp. and *Diaphorobacter* spp. in the total community and *Thauera* spp. and *Comamonas* spp. in the active community. The relevance of the last two genera in phenol metabolism is broadly supported by literature data. *Thauera* spp. are commonly found in biological treatment systems and are able to utilize different kinds of aromatic compounds, such as phenol [28], polyphenol [9], toluene [29], and halobenzoate [30] as carbon source. Functional studies based on SIP-RNA have revealed that members of *Thauera* genus dominated the phenol degradation process in bioreactor sludges [23, 37], corroborating our findings. Additionally, studies based on culturing and/or DGGE fingerprint analyses have showed that *Comamonas* populations can be involved with the metabolism of aromatic compounds, including phenol [3, 5, 16], and can even supplant *Thauera*

populations as the dominant functional members of an industrial wastewater treatment reactor during phenol degradation in response to the presence of *N*-acyl-L-homoserine lactones, suggesting that both organisms are responsible for phenol metabolism. [37].

Members of the *Diaphorobacter* genus have already been related with ammonium metabolism in wastewater treatment plants [14, 17], occasionally displaying simultaneous nitrification and denitrification by aerobic conversion of ammonium to N_2 gas [18].

Sequencing of clones from both libraries indicated that dominant taxonomic groups (Rhodocyclales and Burkholderiales) observed in the RNA-derived library were essentially the same verified in the DNA-derived library, suggesting that these bacterial populations play essential roles in the metabolic processes that take place in the final phase of the bioreactor operation. These results are in agreement with those reported by Duineveld *et al.* [10], who observed that numerically dominant populations were responsible for most of the metabolic activity in the community. Some bacterial groups were observed only in one of the 16S libraries, such as the orders Hydrogenophilaes, Nitrospirales, and Rhizobiales, found in the rDNA library, as well as the phylum Planctomycetes, found exclusively in the rRNA library. In the first case, those bacteria could represent cells of high abundance and/or with multiple operons, although with low metabolic activity [15]. In the second case, the Planctomycetes-related bacteria could represent active members of the complex community with more ribosomes present in their cells and/or low cellular abundance [15]. This could also explain why the most abundant clone in the 16S rRNA library (clone 6D2) is poorly represented in the 16S rDNA library.

Although populations related with ammonium metabolism were found in the 16S rDNA library, such as *Diaphorobacter* spp., *Nitrospira* spp., and *Nitrosomonas* spp., these populations do not appear in the metabolically active library, corroborating the low value of ammonium removal observed in the bioreactor. In contrast, populations related with phenol degradation were found in high abundance in both libraries and seemed not to be affected by the NH_4^+ -N concentration in the SBR wastewater.

Finally, investigating bacterial populations responsible for the degradation of toxic compounds may contribute to our understanding of the community dynamics and to increasing the efficiency of biological treatment wastewater plants. The combined use of DGGE, multivariate analysis, and 16S rRNA gene libraries was shown to be a relatively rapid and efficient approach in monitoring the changes of bacterial communities in response to exogenous perturbation. This work certainly helps to provide a future direction for the functional analyses of specific microorganisms of the community.

Acknowledgments

The authors are grateful to Petrobras for technical and financial support. C.C.S. was supported by grants from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

REFERENCES

1. Ahn, S., S. Congeevaram, Y. K. Choung, and J. Park. 2008. Enhanced phenol removal by floating fungal populations in a high concentration phenol-fed membrane bioreactor. *Desalination* **221**: 494–501.
2. Andreote, F. D., J. L. Azevedo, and W. L. Araújo. 2009. Assessing the diversity of bacterial communities associated with plants. *Braz. J. Microbiol.* (In Press).
3. Arai, H., S. Akahira, T. Ohishi, M. Maeda, and T. Kudo. 1998. Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: Organization and regulation of the genes involved in phenol degradation. *Microbiology* **144**: 2895–2903.
4. Atlas, R. M. and R. Bartha. 1998. *Microbial Ecology: Fundamentals and Applications*, pp. 329–477, 4th Ed. Bejaming/Cumming Science Publishing Company, Inc., Menlo Park, CA.
5. Bae, H. S., J. M. Lee, Y. B. Kim, and S. T. Lee. 1996. Biodegradation of the mixtures of 4-chlorophenol and phenol by *Comamonas testosteroni* CPW301. *Biodegradation* **7**: 463–469.
6. Barrios-Martinez, A., E. Barbot, B. Marrot, P. Moulin, and N. Roche. 2006. Degradation of synthetic phenol-containing wastewaters by MBR. *J. Membrane Sci.* **281**: 288–296.
7. Basile, L. A. and L. Erijman. 2008. Quantitative assessment of phenol hydroxylase diversity in bioreactors using a functional gene analysis. *Appl. Microbiol. Biotechnol.* **78**: 863–872.
8. Contreras, E. M., M. E. Albertario, N. C. Bertola, and N. E. Zaritzky. 2008. Modelling phenol biodegradation by sludges evaluated through respirometric techniques. *J. Hazard. Mater.* **150**: 366–374.
9. Dibenedetto, A., R. M. Lo Noce, M. Narracci, and M. Aresta. 2006. Structure–biodegradation correlation of polyphenols for *Thauera aromatica* in anaerobic conditions. *Chem. Ecol.* **22**: 133–143.
10. Duineveld, B. M., G. A. Kowalchuk, A. Keijzer, J. D. van Elsas, and J. A. van Veen. 2001. Analysis of bacterial communities in the rhizosphere of chrysanthemum via denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA as well as DNA fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **67**: 172–178.
11. Ewing, B., L. Hillier, M. Wendt, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
12. Godon, J. J., E. Zumstein, P. Dabert, F. Habouzit, and R. Moletta. 1997. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63**: 2802–2813.
13. Heuer, H., M. Krsek, P. Baker, K. Smalla, and E. M. Wellington. 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic

- separation in denaturing gradients. *Appl. Environ. Microbiol.* **63**: 3233–3241.
14. Heylen, K., B. Vanparys, L. Wittebolle, N. Verstraete, N. Boon, and P. de Vos. 2006. Cultivation of denitrifying bacterium: Optimization of isolation conditions and diversity study. *Appl. Environ. Microbiol.* **72**: 2637–2643.
 15. Ibekwe, A. M. and S. R. Lyon. 2007. Microbial characteristics through drinking water aquifer sand material. *Eng. Life Sci.* **1**: 81–89.
 16. Jiang, H. J., J. H. Tay, A. M. Maszenan, and S. T. L. Tay. 2004. Bacterial diversity and function of aerobic granules engineered in a sequencing batch reactor for phenol degradation. *Appl. Environ. Microbiol.* **70**: 6767–6775.
 17. Khan, S. T. and A. Hiraishi. 2002. *Diaphorobacter nitroreducens* gen. nov., sp. nov., a poly(3-hydroxybutyrate) degrading denitrifying bacterium isolated from sludge. *J. Gen. Appl. Microbiol.* **48**: 299–308.
 18. Khardenavis, A. A., A. Kapley, and H. J. Purohit. 2007. Simultaneous nitrification and denitrification by diverse *Diaphorobacter* sp. *Appl. Microbiol. Biotechnol.* **77**: 403–409.
 19. Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
 20. Lane, D. L., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Prog. Nat. Sci.* **82**: 6955–6959.
 21. Lin, C.-W., C.-Y. Lai, L.-H. Chen, and W.-F. Chiang. 2007. Microbial community structure during oxygen-stimulated bioremediation in phenol-contaminated groundwater. *J. Hazard. Mater.* **140**: 221–229.
 22. Löffler, F. E. and E. A. Edwards. 2006. Harnessing microbial activities for environmental cleanup. *Curr. Opin. Microbiol.* **17**: 274–284.
 23. Manefield, M., A. S. Whiteley, G. I. Griffiths, and M. J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.* **68**: 5367–5373.
 24. Mao, Y., X. Zhang, X. Yan, B. Liu, and L. Zhao. 2008. Development of group-specific PCR–DGGE fingerprinting for monitoring structural changes of *Thauera* spp. in an industrial wastewater treatment plant responding to operational perturbations. *J. Microbiol. Meth.* **75**: 231–236.
 25. Park, J. Y. and B. Sang. 2007. Change of sludge consortium in response to sequential adaptation to benzene, toluene, and o-xylene. *J. Microbiol. Biotechnol.* **17**: 1772–1781.
 26. Pearson, K. 1926. On the coefficient of racial likeness. *Biometrika.* **18**: 105–117.
 27. Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
 28. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, U.S.A.
 29. Shinoda, Y., Y. Sakai, H. Uenishi, Y. Uchihashi, A. Hiraishi, H. Yukawa, H. Yurimoto, and N. Kato. 2004. Aerobic and anaerobic toluene degradation by a newly isolated denitrifying bacterium, *Thauera* sp. strain DNT-1. *Appl. Environ. Microbiol.* **70**: 1385–1392.
 30. Song, B., N. J. Palleroni, and M. M. Haggblom. 2000. Isolation and characterization of diverse halobenzoate-degrading denitrifying bacteria from soils and sediments. *Appl. Environ. Microbiol.* **66**: 3446–3453.
 31. *Standard Methods for the Examination of Water and Wastewater*. 1998. American Public Health Association, American Water Works association and Water Pollution Control Federation, 20th edition, Washington, DC.
 32. Streit, W. R. and R. A. Schmitz. 2004. Metagenomics – The key to the uncultured microbes. *Curr. Opin. Microbiol.* **7**: 492–498.
 33. Sueoka, K., H. Satoh, M. Onuki, and T. Mino. 2009. Microorganisms involved in anaerobic phenol degradation in the treatment of synthetic coke-oven wastewater detected by RNA stable-isotope probing. *FEMS Microbiol. Lett.* **291**: 169–174.
 34. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.
 35. ter Braak, C. J. F. and P. Smilauer. 2002. *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination* (version 4.5).
 36. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.
 37. Valle, A., M. J. Bailey, A. S. Whiteley, and M. Manefield. 2004. N-Acyl-L-homoserine lactones (AHLs) affect microbial community composition and function in sludge. *Environ. Microbiol.* **6**: 424–433.
 38. Viero, A. F., T. M. Melo, A. P. R. Torres, N. R. Ferreira, G. L. Sant'Anna Jr., C. P. Borges, and V. M. J. Santiago. 2008. The effects of long-term feeding of high organic loading in a submerged membrane bioreactor treating oil refinery wastewater. *J. Membrane Sci.* **319**: 223–230.
 39. Wei, G., J. Yu, Y. Zhu, W. Chen, and L. Wang. 2008. Characterization of phenol degradation by *Rhizobium* sp. CCNWTB 701 isolated from *Astragalus chrysopterus* in mining tailing region. *J. Hazard. Mater.* **151**: 111–117.
 40. White, D. 2000. *The Physiology and Biochemistry of Prokaryotes*, 2nd Ed. Oxford University Press, Oxford, NY.
 41. Yan, J., W. Jianping, B. Jing, W. Daoquan, and H. Zongding. 2006. Phenol biodegradation by the yeast *Candida tropicalis* in the presence of m-cresol. *Biochem. Eng. J.* **29**: 227–234.
 42. Zhao, W.-T., X. Huang, D.-J. Lee, X.-H. Wang, and Y.-X. Shen. 2009. Use of submerged anaerobic–anoxic–oxic membrane bioreactor to treat highly toxic coke wastewater with complete sludge retention. *J. Membrane Sci.* **330**: 57–64.

**CAPÍTULO 2 – INVESTIGATION OF BACTERIAL DIVERSITY IN MEMBRANE BIOREACTOR
AND CONVECTIONAL ACTIVATED SLUDGE PROCESSES FROM
PETROLEUM REFINERIES USING PHYLOGENETICS AND STATISTICAL
APPROACHES.**

Cynthia C. Silva, Aline F. Viero, Ana Carolina F. Dias, Fernando D. Andreote, Ederson C. Jesus, Sergio O. De Paula, Ana Paula R. Torres, Vania M. J. Santiago & Valeria M. Oliveira. **Investigation of bacterial diversity in membrane bioreactor and conventional activated sludge processes from petroleum refineries using phylogenetic and statistical approaches.** *Publicado no Journal of Microbiology and Biotechnology.*



Investigation of Bacterial Diversity in Membrane Bioreactor and Conventional Activated Sludge Processes from Petroleum Refineries Using Phylogenetic and Statistical Approaches

Silva, Cynthia C.¹, Ederson C. Jesus², Ana P. R. Torres³, Maira P. Sousa³, Vânia M. J. Santiago³, and Valéria M. Oliveira^{1*}

¹Division of Microbial Resources, Research Center for Chemistry, Biology and Agriculture (CPQBA), Campinas University - UNICAMP, CP 6171, CEP 13081-970, Campinas, SP, Brazil

²Center for Microbial Ecology, Michigan State University, East Lansing, MI, 488242, U.S.A.

³PETROBRAS R&D Center, Cidade Universitária, Quadra 7, Ilha do Fundão, Rio de Janeiro, RJ, CEP 21949-900, Brazil

Received: June 24, 2009 / Revised: November 9, 2009 / Accepted: December 3, 2009

Bacterial diversity of two distinct wastewater treatment systems, conventional activated sludge (CAS) and membrane bioreactor (MBR), of petroleum refineries were investigated through 16S rRNA gene libraries. Sequencing and phylogenetic analysis showed that the bacterial community composition of sludge samples was distinct between the two wastewater treatment systems. MBR clones belonged predominantly to Class *Betaproteobacteria*, represented mainly by genera *Thiobacillus* and *Thauera*, whereas CAS clones were mostly related to Class *Alphaproteobacteria*, represented by uncultured bacteria related to Order *Parvularculales*. Richness estimators ACE and Chao revealed that the diversity observed in both libraries at the species level is an underestimate of the total bacterial diversity present in the environment and further sampling would yield an increased observed diversity. Shannon and Simpson diversity indices were different between the libraries and revealed greater bacterial diversity for the MBR library, considering an evolutionary distance of 0.03. LIBSHUFF analyses revealed that MBR and CAS communities were significantly different at the 95% confidence level ($P \leq 0.05$) for distances $0 \leq D \leq 0.20$. This work described, qualitatively and quantitatively, the structure of bacterial communities in industrial-scale MBR and CAS processes of the wastewater treatment system from petroleum refineries and demonstrated clearly differentiated communities responsible for the stable performance of wastewater treatment plants.

Keywords: Wastewater treatment, sludge, bacterial community, 16S rRNA genes library

*Corresponding author

Phone: +55 (19) 2139-2874; Fax: +55 (19) 2139-2852;
E-mail: vmaia@cpqba.unicamp.br

Petroleum refineries are great water consumers, since the refining process consists of a series of steps of petroleum amelioration classified as separation or conversion processes that use great volumes of water or aqueous solutions [43]. This situation contributes to worsening the picture of environmental pollution, with potential hazards to the water quality. As a consequence, strict regulations were created in Brazil to rule the use of water resources, leading the productive sectors to rationalize water consumption [40].

In that sense, petroleum refineries have made a considerable effort to treat their effluents, which may pose a difficult task because of the complexity of their chemical composition [37]. Compounds such as phenols and mercaptans are the main constituents of refineries of petrochemical industries, and may be harmful to human health and to the environment, even in low concentrations, as a consequence of the toxicity derived from the synergic effects of their interaction [6, 41]. In this context, treated effluents may still present recalcitrant compounds and some chronic toxicity, which make water reuse limited.

Biological treatment, or bioremediation, of wastewaters uses microorganisms to mineralize complex pollutant organic and inorganic compounds, transforming them into inert or nontoxic compounds. This technology has been broadly employed in petroleum refineries as an international practice, and has shown promising results for the efficient wastewater treatment and the establishment of the water reuse in such industries [37]. Additionally, the inefficiency, high costs, and operational complexity of the existing physicochemical treatments seem to stimulate the intense application of bioremediation. Conventional activated sludge (CAS) and membrane bioreactor (MBR) processes have

been widely used in the biological treatment of petrochemical wastewater [49, 57]. The CAS process usually presents three stages: sedimentation of gross solids in the influent followed by aerobic degradation of the organic matter, and then a secondary clarifier to remove the biomass. The MBR is one of the next-generation wastewater treatment processes and consists of a suspended growth-activated sludge system that utilizes microporous membranes for solid/liquid separation instead of secondary clarifiers. This system presents a high sludge age that gives the bacteria time to adapt to pollutant compounds and a low food/microorganism ratio that is expected to force bacteria to also mineralize poorly degradable organic compounds [27, 59], resulting in an effluent water quality that is significantly higher than the one generated by conventional treatment.

However, despite the great relevance of biological treatments to industries, most of the related studies have focused mainly on treatment performance and associated parameters, without giving proper attention to the microbial communities involved. Communities of bacteria present in activated sludge or biofilm reactors are responsible for most of the carbon and nutrient removal from the effluent and represent the major component of every biological wastewater treatment plant. Thus, a better understanding of the microbial ecology in these systems is required to reveal factors influencing the efficiency and stability of biological treatments and to develop strategies for improved process performance.

Microbial communities in environmental samples have been traditionally characterized by means of cultivation-based techniques or optical microscopy. These approaches allowed the selection of model organisms for the majority of microbiological processes in wastewater treatments [22]. Although very useful for taxonomical, physiological, and genetic studies, these techniques are inadequate for a proper description of the composition and dynamics of the microbial communities, since it is now well known that only a small fraction (0.1 to 10%) of the microbial diversity in nature can be recovered in the laboratory [1, 24].

Molecular tools developed in the last decades to overcome the limitations imposed by traditional cultivation techniques have been used to study bacterial diversity in wastewater biological treatments [33]. These studies demonstrated that most of the suggested model organisms are functionally not relevant in the bioreactors, and that other microorganisms, some of them yet uncultured, may be related to key ecological processes that take place *in situ*.

Based on previous knowledge that MBR and CAS processes produce effluent of different quality, this work aimed at a cultivation-independent characterization and comparison of the bacterial communities present in sludge samples originated from wastewater treatment plants of petroleum refineries operating in conventional activated sludge and membrane bioreactor systems.

MATERIALS AND METHODS

Conventional Activated Sludge and Membrane Bioreactor Processes

We investigated two wastewater treatment plants of petroleum refinery with operation in continuous flow. The influent of both plants had similar composition (data not shown). The first one has been in continuous operation for more than 10 years and is part of the industrial Wastewater Treatment Plant (WWTP) of Petrobras Refinery REVAP (São Paulo State, Brazil). It comprises a conventional activated sludge process as a secondary treatment, with an anoxic/oxic zone tank and a secondary clarifier. The treatment capacity of the WWTP is about 16,800 m³/d (700 m³/h) and the HRT (hydraulic retention time) is around 6.5 h. The second one is a submerged membrane bioreactor pilot that has been in continuous operation for 18 months and is part of the industrial WWTP of Petrobras Refinery REGAP (Minas Gerais State, Brazil). The treatment capacity of the MBR is about 24 m³/d (1 m³/h) and the HRT is around 8 h. This bioreactor operates with a microfiltration flat-sheet membrane module (Kubota, Osaka, Japan) with nominal porosity of 0.4 µm. This bioreactor effective volume is 8 m³ and aeration is continuously carried out. Filtration was intermittently carried out in the constant flow rate operation mode.

We considered the common analytical factors measured in the routine of the plants. The physicochemical analyses were performed during the month of sampling. COD (chemical oxygen demand), BOD (biochemical oxygen demand), VSS (volatile suspended solid), ammonium (NH₄⁺-N) concentration, and alkalinity were measured according to the procedures described in Standard Methods [12]. Measures of pH were recorded by using a pH meter (Thermo Orion, Model 868; U.S.A.).

Sampling and DNA Extraction

Sludge samples were collected after 18 months of continuous operation at REGAP and after 10 years of continuous operation at REVAP, with logistic support of CENPES/PETROBRAS. Samples were collected in triplicate using 1-l sterilized bottles, which were kept on ice during transportation to the laboratory and stored at -20°C for subsequent community DNA extraction.

DNA extraction from sludge samples was carried out using a protocol based on Großkopf *et al.* [21] and Neria-González *et al.* [42], with modifications, as follows: microbial cells were retrieved from sludge samples by centrifuging 4-ml aliquots at 13,400 ×g for 5 min. The pellets were suspended in 600 µl of PBS buffer and homogenized by vortex. The solution of lysozyme was added at a final concentration of 17 mg/ml and incubated at 37°C for 30 min, with manual shaking every 10 min. Proteinase K and SDS were subsequently added (final concentration of 0.7 mg/ml and 2%, respectively) and the solution was incubated at 60°C for 30 min, with manual shaking every 10 min. The tubes were subjected to 3 freeze-thaw cycles, being 2 min in liquid nitrogen followed by 2 min at 65°C. An equal volume of saturated phenol (pH 8.0) was added to the solution, followed by homogenization and centrifugation at 13,400 ×g for 5 min. The supernatant was extracted with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged at 13,400 ×g for 5 min. For DNA precipitation, 5 M NaCl (10% of total volume) and 2 volumes of cold ethanol were added to the solution. The pellet was washed once with 70% ethanol, dried, and suspended in 25 µl of Milli-Q H₂O.

The yield and integrity of DNA obtained from sludge samples were confirmed through electrophoresis in 1% agarose gel.

16S rRNA Gene Libraries

For 16S rRNA gene library construction, the amplification was performed using the bacterial primer set 27f and 1100r [31]. Fifty μ l reaction mixtures contained 50–100 ng of total DNA, 2 U of *Taq* DNA polymerase (Invitrogen), 1 $\times$ *Taq* buffer, 1.5 mM $MgCl_2$, 0.2 mM of dNTP mix (GE Healthcare), and 0.4 mM each primer. The PCR amplifications were done using an initial denaturation step of 2 min at 95°C, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 3 min at 72°C, followed by a final extension step at 72°C for 3 min, in an Eppendorf thermal cycler.

Cloning and Sequencing of 16S rRNA PCR Products

Two 16S rRNA gene libraries were constructed, one for each wastewater treatment plant, REGAP and REVAP. Three PCR replicates of each sample were pooled, purified using GFX PCR DNA and Gel Band Purification kit (GE Healthcare), and concentrated in a speed vacuum concentrator 5301 Eppendorf, A-2-VC rotor. The purified PCR product (200 ng) was ligated into a pGEM-T Easy Vector (Promega), according to the manufacturer's instructions, and transformed into *E. coli* JM109 competent cells. Approximately 200 positive clones were selected from each library for subsequent sequencing. The 16S rRNA inserts were amplified from plasmid DNA of selected clones using the universal M13 forward (5'-CGC CAG GGT TTT CCC AGT CAC GAC-3') and M13 reverse (5'-TTT CAC ACA GGA AAC AGC TAT GAC-3') primers. PCR was performed in a 50- μ l reaction volume, containing 1–2 μ l of an overnight clone culture, 0.4 μ M each primer, 0.2 mM dNTP mix, 2 U *Taq* DNA polymerase (Invitrogen), 1 $\times$ *Taq* buffer, and 1.5 mM $MgCl_2$. The amplification program consisted of an initial denaturation step at 94°C for 3 min, followed by 30 cycles of 94°C/20 s, 60°C/20 s, and 72°C/90 s. PCR products were purified as previously described for automated sequencing in the MegaBace DNA Analysis System 500 (GE Healthcare). The sequencing was carried out using 10f and 1100r primers [30] and the DYEnamic ET Dye Terminator Cycle Sequencing Kit (GE Healthcare), according to the manufacturer's recommendations.

Sequence Analysis

Partial 16S rRNA sequences obtained from clones were assembled in a contig using the phred/Phrap/CONSED program [15, 19]. Identification was achieved by comparing the contiguous 16S rRNA sequences obtained with 16S rRNA sequence data from reference and type strains, as well as environmental clones available in the public databases of GenBank (<http://www.ncbi.nlm.nih.gov>) and RDP (Ribosomal Database Project - Release 9; <http://rdp.cme.msu.edu/>) using the BLASTn and Classifier routines, respectively. The sequences were aligned using the CLUSTAL X program [56] and analyzed with MEGA software v.4 [53]. Evolutionary distances were derived from sequence-pair dissimilarities calculated as implemented in MEGA, using Kimura's DNA substitution model [28]. The phylogenetic reconstruction was done using the neighbor-joining (NJ) algorithm, with bootstrap values calculated from 1,000 replicate runs, using the routines included in the MEGA software.

Statistical Analyses

Physicochemical data were submitted to statistical analysis by using the program SISVAR [17]. The normal distribution of data was

tested with the Shapiro–Wilk Test. Analysis of variance was carried out on the data and the difference between means was assessed with the Student's t-test with Bonferroni correction. The VSS variable was transformed to its neperian logarithm before being subjected to the analysis of variance.

Diversity index calculations (α -diversity measures) were performed individually for both 16S rRNA gene libraries with the programs MOTHUR version 1.1.0 and DOTUR v. 1.3 [46]. DOTUR was used to assign sequences to OTUs and to calculate collector's curves using 2 distance levels among sequences (3 and 20%), and MOTHUR was used to calculate, for each sludge sample, the Shannon (H') diversity index, Simpson index [34], and the nonparametric richness estimators ACE (abundance-based coverage estimator) [8] and Chao1 [7]. Chao1 richness estimates were based on singletons and doubletons as described by Chao [7], and ACE was based on the distribution of abundant (>10) and rare (≤ 10) species.

Statistical comparisons of coverage of bacterial 16S rDNA libraries from the two sludge samples were performed using the LIBSHUFF computer program [51], which used the formula of Good [18] to generate homologous and heterologous coverage curves from the clone libraries. Sequences were randomly shuffled 999 times between samples prior to the distance between the curves being calculated using the Cramér–von Mises statistical test [44]. The DNADIST program of PHYLIP [16], using Kimura's model [28] for nucleotide substitution, was employed to generate the distance matrix analyzed by LIBSHUFF.

Nucleotide Sequence Accession Numbers

The 16S rDNA partial sequences determined in this study for the environmental clones were deposited at the GenBank database under the accession numbers FJ933260 to FJ933490.

RESULTS AND DISCUSSION

Physicochemical Characteristics of Effluent Water in CAS and MBR Processes

Physicochemical parameters of processes were analyzed during 30 days in the month of sampling (September 2005 at REGAP and March 2006 at REVAP) and the average values are shown in Table 1. Evolution of chemical oxygen demand, biochemical oxygen demand, alkalinity, ammonia nitrogen (NH_3 -N), pH, and volatile suspended solid were used to monitor the performance of the CAS and MBR processes.

Statistical tests were applied over the average values of parameters to assess if the CAS and MBR processes were significantly different. Results demonstrated stability and efficiency in the performance during the month of sampling for both processes. Only VSS and alkalinity values were significantly different (Table 1). The VSS parameter in the MBR was approximately 2.5 times more than in the CAS. This parameter reflects the quantity of biomass; therefore, this difference could be explained by the complete biomass retention that occurs in the MBR process. Ammonia, BOD, and COD were efficiently removed in both processes (Table 1). In spite of the significant difference between alkalinity

Table 1. Average values of operational parameter characteristics of CAS and MBR.

Parameters	CAS			MBR		
	Influent	Effluent	Removal %	Influent	Effluent	Removal %
SSV (mg/l) ^a	3,500±500			8,506±504		
pH	7.81±0.60	7.60±0.20		8.15±0.57	6.60±0.30	
NH ₄ ⁺ -N (mg/l)	41.8±14.3	1.1±1.1	97.4	46.9±17.9	2.4±1.1	94.9
COD (mg/l)	412±155	63±29	84.7	458±53	68±27	85.2
BOD (mg/l)	165±44	5±5	97.0	148±51	3±1	98.0
Alkalinity (mgCaCO ₃ /l) ^a	247±64	10±5	96.0	395±79	28±14	92.9

Values are given as mean concentration±standard deviation.

^aStatistical significance at the 5% level.

values, the high values found for both processes allowed pH to be maintained around 7–8, necessary for efficient ammonia removal by nitrification.

Composition of the Sludge Bacterial Communities

The bacterial community composition of sludges from the two different petrochemical wastewater systems was determined by analysis of 16S rRNA gene clone libraries. Conventional activated sludge and membrane bioreactor systems were studied operating in industrial scale. A total of 231 partial sequences were obtained from both libraries (110 sequences from MBR and 121 from CAS), each consisting of 850 nucleotides on average, which presented good quality for subsequent analyses. These 16S rRNA sequences were compared with sequences from reference and type strains, as well as environmental clones, available at the GenBank and RDP II databases.

In the MBR sample, 11 major bacterial taxa were identified. Seventy-eight sequences (71%) were related to the phylum *Proteobacteria*, being 38 (48.7%) affiliated to Class *Betaproteobacteria*, 28 to Class *Alphaproteobacteria*, 10 to class *Deltaproteobacteria*, and 1 to class *Gammaproteobacteria* (Fig. 1A and 1B). The remaining sequences were significantly less numerous, being related to the phyla *Acidobacteria* (11), *Actinobacteria* (3), *Bacteroidetes* (2), *Chloroflexi* (3), *Firmicutes* (1), *Nitrospirae* (8), *Planctomycetes* (1), *Chlorobi* (1), Candidate Division TM7 (1), and *Verrucomicrobia* (2) (Fig. 1A). In the CAS sample, 7 major bacterial taxa were identified. The phylum *Proteobacteria* was the most abundant one, with 77 sequences (64%) distributed among the classes *Alphaproteobacteria* (44 sequences), *Betaproteobacteria* (15 sequences), *Gammaproteobacteria* (7 sequences), and unaffiliated *Proteobacteria* (11 sequences) (Fig. 1A and 1B). The remaining sequences were related to the phyla *Acidobacteria* (2), *Actinobacteria* (8), *Chlorobi* (3), *Chloroflexi* (15), *Firmicutes* (3), and *Nitrospirae* (10), and three clones were considered unaffiliated (Fig. 1A). All taxonomic groups found in both libraries have already been described in the literature as commonly occurring organisms in wastewater treatment systems [27, 29, 49].

The composition of phylotypes at the phylum level was similar between the libraries from distinct wastewater treatments. All phyla found in the CAS library were present in the MBR library (Fig. 1A). The phylum *Proteobacteria* was predominant in both libraries, corroborating literature data that report the predominance of this group in sludge samples from sewage and industrial wastewater treatment systems [27, 29]. However, the abundance of clones in *Alpha*- and *Betaproteobacteria* classes was discrepant between the libraries (Fig. 1B).

Differences were more pronounced when comparing the orders present in each library. In the MBR library, the most

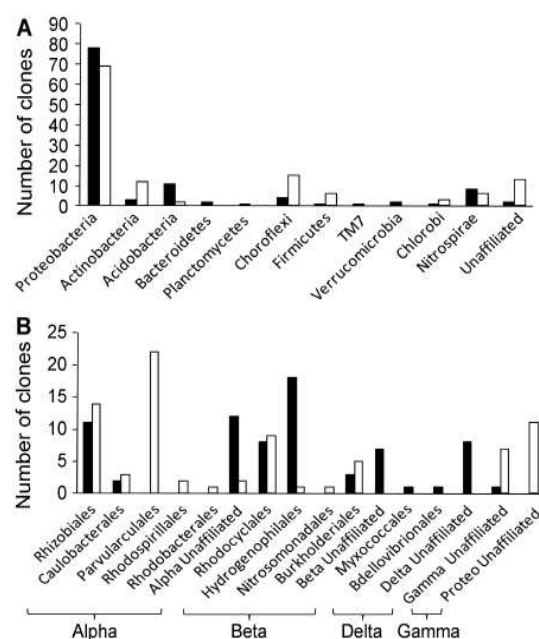


Fig. 1. Proportion of 16S rRNA gene clones from MBR (filled bar) and CAS (opened bar) processes among bacterial phyla (A) and orders within phylum *Proteobacteria* (B). Sequence classification was based on the RDP classifier and BLASTn results.

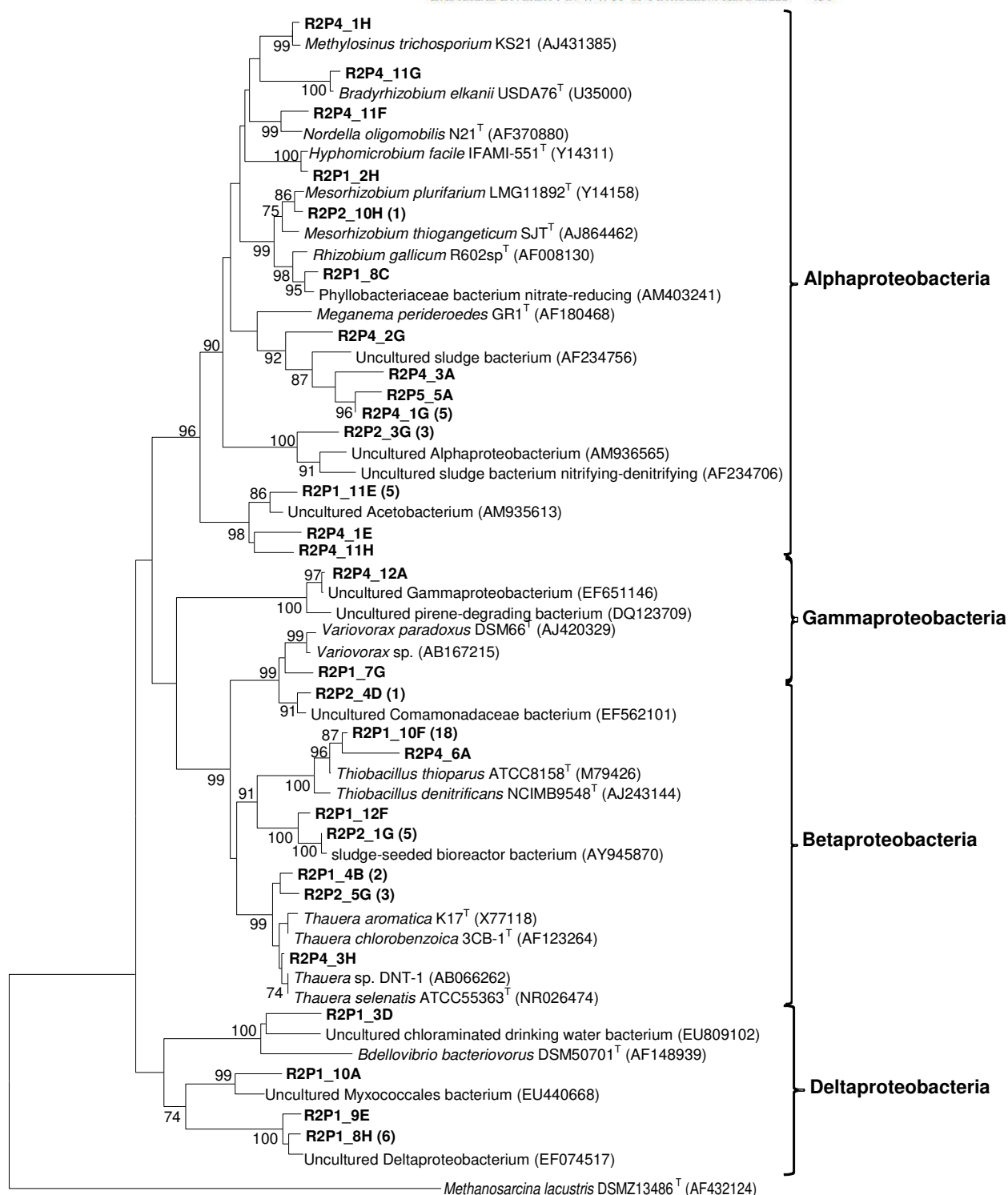


Fig. 2. Phylogenetic analysis based on partial 16S rRNA sequences from the MBR clone library (R2) representing members of phylum *Proteobacteria* and related species. Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanosarcina lacustris* was used as outgroup.

abundant orders in the class *Betaproteobacteria* were *Hydrogenophilales* and *Rhodocyclales* (Fig. 1B), mainly represented by the genera *Thiobacillus* and *Thauera*, respectively (Fig. 2). These genera have been described in other studies of sludge diversity of wastewater treatment [5, 35, 48, 50, 52]. *Thauera* spp. have been reported as representing denitrifying bacteria in a wastewater treatment plant [55]. In addition, a recent work suggested that these bacteria may be explored as an important biological indicator for population optimization and performance monitoring in wastewater treatment processes [36]. Several researchers have demonstrated that *Thauera* species are related with aerobic and anaerobic degradation of aromatic compounds, such as phenol and toluene, by means of distinct catabolic pathways [35, 48]. The genus *Thiobacillus* is able to oxidize reduced inorganic or organic sulfur compounds commonly found in petroleum refining industry wastewaters (such as hydrogen sulfide, thiosulfate, methanethiol, ethanethiol, *etc.*) to sulfates [10, 52]. *Thiobacillus* strains have also been related to the degradation of polycyclic aromatic hydrocarbons (PAHs) [5]. Clones found in the MBR library were more closely related to *T. thioparus* (Fig. 2). These bacteria are used for the removal of hydrogen sulfide (H₂S) and other reduced sulfur compounds from petroleum refinery and sewage waste streams, in order to treat odorous emissions [2, 14]. Tanji *et al.* [54] described that immobilized cells of *Thiobacillus thioparus* were able to remove great quantities of dimethyl sulfide, methyl mercaptan, and hydrogen sulfide from sulfur-containing malodorous gas in laboratory-scale tower-type reactors. Clones affiliated to the order *Rhizobiales* in MBR library were as abundant as the ones affiliated to *Rhodocyclales* or *Hydrogenophyllales*, although more diverse, being related to bacteria capable of oxidizing methane (AJ431385), fixing nitrogen (AF041442), and reducing sulfate (AJ864462) (Fig. 2).

The results of the CAS library revealed that the dominant sequence types belonged to *Alphaproteobacteria*, which included the orders *Rhodobacterales*, *Caulobacterales*, *Rhodospirillales*, *Parvularculales*, and *Rhizobiales* (Fig. 1B). In this case, the majority of the clones were related to the order *Parvularculales*, but clone identification at higher taxonomic ranks was not possible by comparison of 16S rRNA gene sequences with databases and subsequent phylogenetic analyses, suggesting that the corresponding organisms may represent new taxa (Fig. 3). Literature data about *Parvularculales* are scarce, and the only valid species described to date is *Parvularcula bermudensis*, a bacterium isolated from a seawater sample [9]. In the class *Betaproteobacteria*, the order *Rhodocyclales* was the most abundant, mainly represented by *Thauera* spp. (Fig. 3). The phylum *Chloroflexi* was the second more abundant in the CAS library (Fig. 1A), accounting for 14% of total clones. Filamentous representatives of this phylum have

been commonly found in activated sludge wastewater treatment plants, and they have occasionally been associated with bulking incidences [3, 4]. Members of *Chloroflexi* are ecologically significant in the MBR process, since they are responsible for degradation of carbohydrates and cellular materials, which consequently reduces membrane fouling potential [38].

The remaining phyla, such as *Acidobacteria* and *Chlorobi* for the CAS library, and *Bacteroidetes*, *Firmicutes*, TM7, and *Planctomycetes* for the MBR library, showed low abundance in the bacterial communities under study (Fig. 4 and 5). Schloss and Handelsman [47] suggested that these poorly sampled members represent rare members in microbial communities, but may play significant functional roles in the microbial community. Other speculative explanation for the low abundance of such taxa would be PCR bias introduced by the use of bacterial primers designed based on rRNA sequences from cultured organisms, consequently amplifying environmental sequences similar to those already existing in the databases [45].

Richness Analyses

Collector's curves, also called species accumulation curves, plot the cumulative number of species recorded as a function of sampling effort [13]. Collector's curves illustrate the rate at which new species are found. The complete coverage of data set is expected to result in a plateau-shaped curve. In this study, collector's curves at the estimated phylum level (distance=0.20) reached clear saturation for the CAS library, suggesting that the sampling effort was sufficient to reveal all phyla present in this sample. For the MBR library, the curve tends to an asymptote, but additional sampling is necessary for saturation. At the species level (distance=0.03), the weakly curvilinear plots showed that phylotype richness in both wastewater libraries was underestimated, suggesting that further sampling from both libraries may have revealed more diversity at the species level (Fig. 6). These results were not expected, since the sampled refinery wastewater treatment plants consist of closed systems that receive large amounts of pollutant compounds and therefore tend to select specific populations capable of utilizing and/or degrading such compounds, thus reducing diversity. Other microbial population studies corroborate our results in the sense that the number of sampled clones was insufficient to reach saturation, when the OTU definition was set at 3% [20, 23]. Hughes *et al.* [25] argued that microbial diversity cannot be fully estimated because of the high bacterial diversity in environmental samples, small sample sizes, or both, resulting in linear or close to linear accumulation curves. On the other hand, exhaustive inventories of microbial communities still remain impractical for most laboratories of microbial ecology.

Nonparametric richness estimators, such as ACE (abundance-based coverage estimators) and Chao1, were

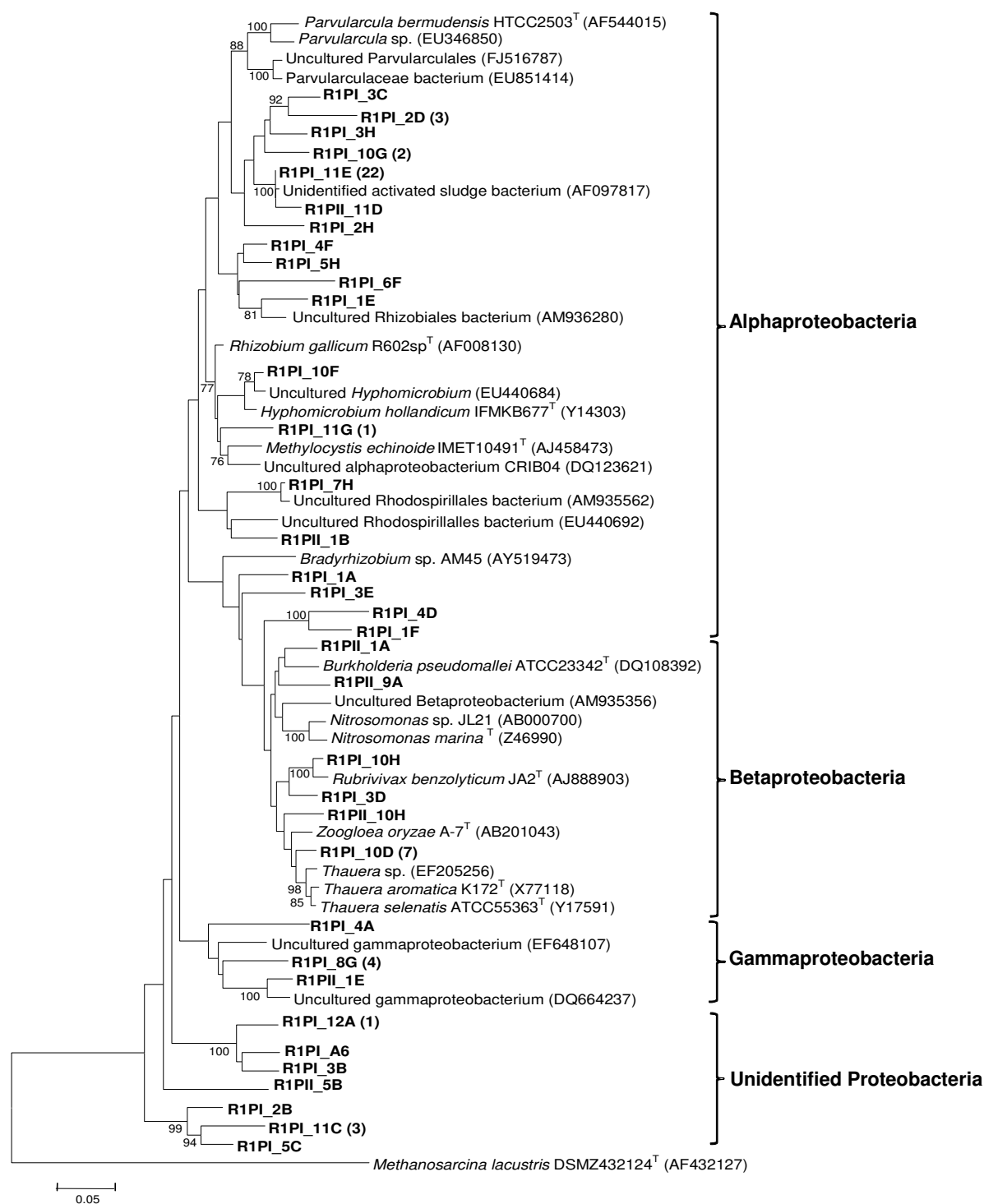


Fig. 3. Phylogenetic analysis based on partial 16S rRNA sequences from the CAS clone library (R1) representing members of phylum *Proteobacteria* and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanosarcina lacustris* was used as the outgroup.

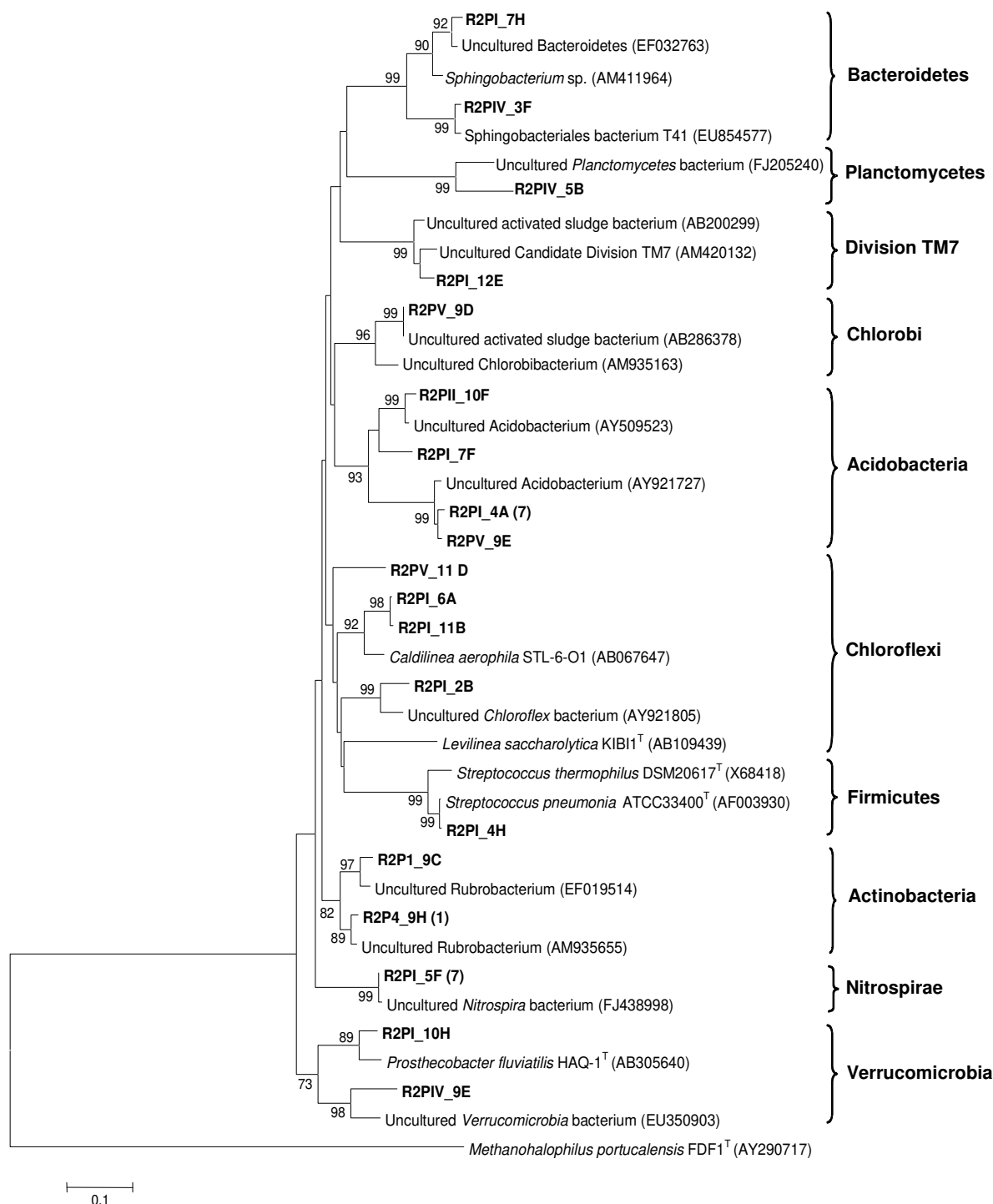


Fig. 4. Phylogenetic analysis based on partial 16S rRNA sequences from MBR clones (R2) representing members of major phyla in Bacteria Domain, except *Proteobacteria*, and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanohalophilus portucalensis* was used as the outgroup.

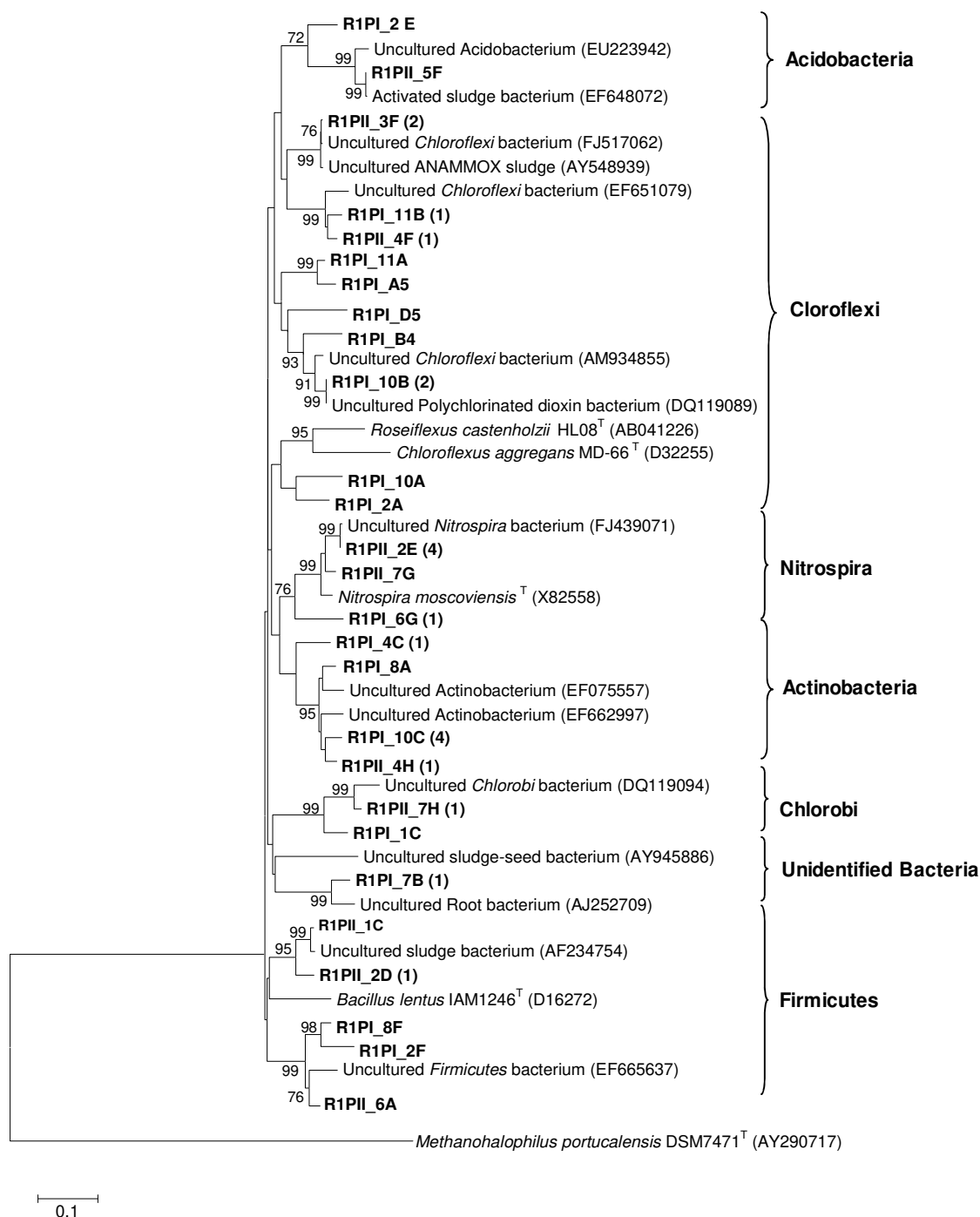


Fig. 5. Phylogenetic analysis based on partial 16S rRNA sequences from CAS clones (R1) representing members of major phyla in Bacteria Domain, except *Proteobacteria*, and related species. Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanohalophilus portucalensis* was used as the outgroup.

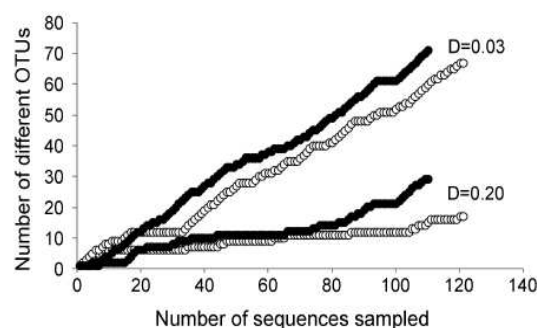


Fig. 6. Collector's curve of observed OTU richness in MBR (filled circle) and CAS (opened circle) sludge samples at evolutionary distances (D) of 0.03 and 0.20.

described to estimate total phylotype richness from the frequency of phylotypes in a clone library [25]. These indices were used in this study to estimate the total number of OTUs in bacterial communities of the sludge samples. In both libraries, the estimated number of phylotypes at the species level was higher than the observed numbers (Table 2). On the other hand, the estimated phylotype richness at the phylum level was similar to the observed one for both MBR and CAS libraries, suggesting that all phyla present in the sludges were sampled in this study. The combined data of richness analyses led us to conclude that the diversity that was observed in both libraries at the species level is an underestimate of the total bacterial diversity present in the environment, and further sampling would yield an increased observed diversity.

Diversity Analyses

Diversity indices serve as a valuable tool to quantify diversity in a community and describe its numerical structure, combining richness and evenness components. Two diversity indices, the Shannon index and the reciprocal of Simpson's index ($1/D$), were used to assess the bacterial community diversity. Simpson's index is heavily weighted towards the most

abundant species in the sample, while being less sensible to species richness. The Shannon index is positively correlated with species richness and evenness and gives more weight per individual to rare than common species, being sensitive to sample size [34]. Higher numbers indicate greater diversity. Both diversity indices were different between the libraries and revealed greater bacterial diversity for the MBR library, considering an evolutionary distance of 0.03 (Table 2). Magurran [34] described that the value of the Shannon index obtained from empirical data usually falls between 1.5 and 3.5 and rarely surpasses 4. Based on diversity index values found in the literature for sludge [32], soil [20], and marine sponge samples [39], the Shannon diversity indices of the CAS and MBR libraries were considered high (distance=0.03). According to previous data [60], the value of the reciprocal Simpson's index ($1/D$) below 50 found for CAS (Table 2) might indicate a dominance profile in this community. These results could be explained by the high abundance of the sequence types related to the order *Parvularculales* in the CAS library (Fig. 1B). Several researchers agree that the microbial population dynamics of MBR systems is different from that of the conventional activated sludge process because of the long sludge retention (>SRT), low food-to-microorganism ratio, complete biomass retention, and lack of sludge recirculation [11, 59]. These conditions allow for adaptation of microorganisms in general and of potentially slow growing specialist bacteria in particular, which establish a more diverse microbial community with broader physical capabilities in the system [58].

The computer program LIBSHUFF was used to compare the genetic coverage obtained in the MBR and CAS 16S rRNA libraries. Two gene libraries are assumed to be different if their homologous and heterologous coverage curves differ significantly [51]. These analyses showed statistically significant differences between the genetic coverage obtained in libraries MBR and CAS, with low P values for both analyses ($P=0.001$) (Fig. 7A and 7B). More information on these differences was obtained by examination

Table 2. Richness and diversity estimates for bacterial 16S rRNA gene clone libraries from MBR and CAS processes (sequence classification based on the cutoff value determined by DOTUR).

Source (n ^a)	Distance ^b	Richness ^c	ACE ^d	Chao1 ^e	Shannon ^f	1/Simpson ^g
MBR (110)	0.20	29	34	30	2.7	14
	0.03	71	145	167	4.0	97
CAS (121)	0.20	17	18	17	2.3	9
	0.03	67	179	150	3.6	30

^an, Number of gene sequences analyzed.

^b80% Identity was estimated as the phylum-level distance (D=0.20), and 97% identity was estimated as the species-level distance (D=0.03).

^cRichness is based on observed unique OTUs.

^dNonparametric statistical prediction of total richness of different OTUs based on distribution of abundant (10) and rare (10) OTUs.

^eNonparametric statistical predictions of total richness of OTUs based on distribution of singletons and doubletons.

^fShannon diversity index. A higher number represents more diversity.

^gReciprocal of Simpson's diversity index. A higher number represents more diversity.

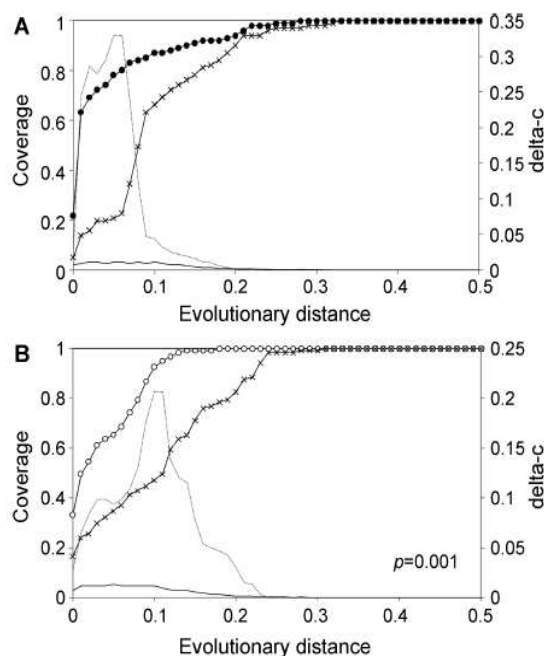


Fig. 7. Genetic coverage comparison between libraries MBR and CAS.

A. MBR homologous coverage (filled circle). B. CAS homologous coverage (opened circle). Heterologous coverages are represented by crosses. Broken lines indicate the value of $(C_x - C_{xy})^2$ for the original samples and solid lines show $P=0.05$ value of $(C_x - C_{xy})^2$ for the randomized samples.

of the distribution of $(C_x - C_{xy})^2$ with D , which revealed that the values calculated for the real data exceeded the values calculated by using randomized data (at a P value of 0.05) at taxonomic distances lower than 0.23. These results showed that the CAS and MBR libraries differed greatly at the level of closely related sequences, but shared most deep taxa. Homologous coverage curves indicated that the MBR library had high coverage estimates at taxonomic distances correspondent to deeply branched taxa, such as phylum ($D \geq 0.20$). The CAS library, on the other hand, had high coverage estimates at $D \geq 0.10$, suggesting that higher taxa such as orders and classes have been fully sampled in the library. These results are consistent with the nonparametric statistical calculations of total richness (ACE and Chao), which yielded greater values for the MBR library at $D=0.20$.

Analysis of bacterial diversity in MBR and CAS samples originated from wastewater treatments of petroleum refineries revealed that the composition of bacterial communities was significantly different between the two processes. To date, a lot of literature have been published regarding the organic removal performance, sludge characteristics, operational conditions, and so on, and illustrated the differences between MBRs and the CAS system. These characteristics obviously contribute to yield

differences in microbial community structure between the processes, which were clearly demonstrated in this work. Diversity calculations revealed high Shannon and Simpson indices for both libraries. Communities highly diverse offer the advantage of flexibility to adapt to changes in environments, what could explain the stable performance of both CAS and MBR processes. This work was one of a few that described, qualitatively and quantitatively, the structure of bacterial communities in MBR and CAS processes of wastewater treatment systems from petroleum refineries. A thorough understanding of the microbial community composition provides a basis for future application of methods for *in situ* analyses of the function of specific microorganisms, and will thus help to establish the links between microbial biodiversity and process stability in wastewater treatment plants.

Acknowledgments

We thank PETROBRAS for the financial support and authorization to publish this work. C.C.S. was supported by grants from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

REFERENCES

1. Amann, R. I., W. Ludwig, and K. Schleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**: 143–169.
2. Barbosa, V. L., S. D. Atkins, V. P. Barbosa, J. E. Burgess, and R. M. Stuetz. 2006. Characterization of *Thiobacillus thioautotrophicus* isolated from an activated sludge bioreactor used for hydrogen sulde treatment. *J. Appl. Microbiol.* **101**: 1269–1281.
3. Beer, M., E. M. Seviour, Y. Kong, M. Cunningham, L. L. Blackall, and R. J. Seviour. 2002. Phylogeny of the filamentous bacterium Eikelboom Type 1851, and design and application of a 16S rRNA targeted oligonucleotide probe for its fluorescence *in situ* identification in activated sludge. *FEMS Microbiol. Lett.* **207**: 179–183.
4. Björnsson, L., P. Hugenholtz, G. W. Tyson, and L. L. Blackall. 2002. Filamentous *Chloroflexi* (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology* **148**: 2309–2318.
5. Bodour, A. A., J.-M. Wang, M. L. Brusseau, and R. M. Maier. 2003. Temporal change in culturable phenanthrene degraders in response to long-term exposure to phenanthrene in a soil column system. *Environ. Microbiol.* **5**: 888–895.
6. Braille, P. M. 1979. *Manual of Industrial Wastewater Treatment*, 1st Ed. São Paulo, CETESB.
7. Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**: 783–791.
8. Chao, A., M.-C. Ma, and M. C. K. Yang. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80**: 193–201.

9. Cho, J. C. and S. J. Giovannoni. 2003. *Parvularcula bermudensis* gen. nov., sp. nov., a marine bacterium that forms a deep branch in the α -Proteobacteria. *Intl. J. Syst. Evol. Microbiol.* **53**: 1031–1036.
10. Cho, K.-S., M. Hirai, and M. Shoda. 1991. Degradation characteristics of hydrogen sulfide, ethanethiol, dimethyl sulfide and dimethyl disulfide by *Thiobacillus thioparus* DW 44 isolated from peat biofilter. *J. Ferment. Bioeng.* **71**: 384–389.
11. Choi, J.-H., S. H. Lee, K. Fukushi, and K. Yamamoto. 2007. Comparison of sludge characteristics and PCR–DGGE based microbial diversity of nanofiltration and microfiltration membrane bioreactor. *Chemosphere* **67**: 1543–1550.
12. Clescerl, L. G., A. E. Greenberg, and A. D. Eaton. 1998. *Standard Methods for the Examination of Water and Wastewater*. APHA, AWWA, WEF, Washington, U.S.A.
13. Colwell, R. K. and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* **345**: 101–118.
14. Duncan, K. E., K. L. Sublette, P. A. Rider, A. Stepp, R. R. Beitle, J. A. Conner, and R. Kolhatkar. 2001. Analysis of a microbial community oxidizing inorganic sulfide and mercaptans. *Biotechnol. Prog.* **17**: 768–774.
15. Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
16. Felsenstein, J. 1989. PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5**: 164–166.
17. Ferreira, D. F. 2000. Análises estatísticas por meio do Sisvar para Windows versão 4.0. In: 45ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, pp. 255–258. UFSCar, São Carlos, SP.
18. Good, I. J. 1953. The population frequency of species and the estimation of population parameters. *Biometrika* **40**: 237–264.
19. Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**: 195–202.
20. Gremion, F., A. Chatzinotas, and H. Harms. 2003. Comparative 16S rDNA and 16S rRNA sequence analysis indicates that *Actinobacteria* might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. *Environ. Microbiol.* **5**: 896–907.
21. Grobkopf, R., S. Stubner, and W. Liesack. 1998. Novel Euryarchaeota lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. *Appl. Environ. Microbiol.* **64**: 960–969.
22. Henze, M., P. Harremoës, J. la Cour Jansen, and E. Arvin. 1997. Wastewater treatment. In U. Forstner, R. J. Murphy, and W. H. Rulkens (eds.). *Environmental Engineering*. 2nd Ed. Springer, Berlin.
23. Horn, M. A., H. L. Drake, and A. Schramm. 2006. Nitrous oxide reductase genes (*nosZ*) of denitrifying microbial populations in soil and the earthworm gut are phylogenetically similar. *Appl. Environ. Microbiol.* **72**: 1019–1026.
24. Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**: 4765–4774.
25. Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan. 2001. Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**: 4399–4406.
26. Judd, S. The status of membrane bioreactor technology. 2008. *Trends Biotechnol.* **26**: 109–116.
27. Juretschko, S., L. Loy, A. Lehner, and M. Wagner. 2002. The microbial community composition of a nitrifying–denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *System. Appl. Microbiol.* **25**: 84–99.
28. Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Mol. Biol. Evol.* **16**: 111–120.
29. Kraigher, B., T. Tina Kosjek, E. Heath, B. Kompare, and I. Mandic-Mulec. 2008. Influence of pharmaceutical residues on the structure of activated sludge bacterial communities in wastewater treatment bioreactors. *Water Res.* **42**: 4578–4588.
30. Lane, D. L., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 6955–6959.
31. Lane, D. 1991. 16S/23S rRNA sequencing, pp. 115–175. In E. Stackebrandt and M. Goodfellow (eds.). *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons, New York, N.Y.
32. Lin, S., S. Jin, L. Fu, C. Quan, and Y. S. Yang. 2009. Microbial community variation and functions to excess sludge reduction in a novel gravel contact oxidation reactor. *J. Hazard. Mater.* **165**: 1083–1090.
33. Loy, A., H. Daims, and M. Wagner. 2002. Activated sludge: Molecular techniques for determining community composition, pp. 26–43. In G. Bitton (ed.). *The Encyclopedia of Environmental Microbiology*. Wiley, New York.
34. Magurran, A. E. 2004. *Measuring Biological Diversity*. Blackwell Publishing, Oxford.
35. Manefield, M., A. S. Whiteley, R. I. Griffiths, and M. J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.* **68**: 5367–5373.
36. Mao, Y., X. Zhang, X. Yan, B. Liu, and L. Zhao. 2008. Development of group-specific PCR–DGGE fingerprinting for monitoring structural changes of *Thauera* spp. in an industrial wastewater treatment plant responding to operational perturbations. *J. Microbiol. Meth.* **75**: 231–236.
37. Mariano, J. B. 2001. Environmental impacts of petroleum refining. MSc. Dissertation, COPPE/UF RJ, Rio de Janeiro, RJ, Brasil.
38. Miura, Y., Y. Watanabe, and S. Okabe. 2007. Significance of *Chloroflexi* in performance of submerged membrane bioreactor (MBR) treating municipal wastewater. *Environ. Sci. Technol.* **41**: 7787–7794.
39. Mohamed, N. M., J. J. Enticknap, J. E. Lohr, S. M. McIntosh, and R. T. Hill. 2008. Changes in bacterial communities of the marine sponge *Mycale laxissima* on transfer into aquaculture. *Appl. Environ. Microbiol.* **74**: 1209–1222.
40. Monteiro, J. P., A. Magrini, A. S. Szklo, and R. T. S. Nunes. 2004. Management of water in the petroleum refinery and perspective of reuse. In: 1st Academic Congress of Environment and Development of Rio de Janeiro, RJ, Brazil. CADMA. FGV, Rio de Janeiro.
41. Mota, S. 1997. *Introduction to Environmental Engineering*. 1st Ed. Rio de Janeiro, Brazil, ABES.
42. Neria-González, I., E. T. Wang, F. Ramírez, J. M. Romero, and C. Hernández-Rodríguez. 2006. Characterization of bacterial

- community associated to biofilms of corroded oil pipelines from southeast of Mexico. *Anaerobe* **12**: 122–133.
43. Pereira, R. C. 2005. Development and application of a new methodology for the speciation analysis of selenium in aqueous effluents from petroleum refineries. MSc. Dissertation, PUC-Rio de Janeiro, RJ, Brazil.
 44. Pettitt, A. N. 1982. Cramer–von Mises statistic, pp. 220–221. In S. Kotz and N. L. Johnson (eds.), *Encyclopedia of Statistical Sciences*. Wiley-Interscience, New York, N.Y.
 45. Schloss, P. D. and J. Handelsman. 2004. Status of the Microbial Census. *Microbiol. Mol. Biol. Rev.* **68**: 686–691.
 46. Schloss, P. D. and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**: 1501–1506.
 47. Schloss, P. D. and J. Handelsman. 2006. Toward a census of bacteria in soil. *PLoS Comp. Biol.* **2**: e92.
 48. Shinoda, Y., J. Akagi, Y. Uchihashi, A. Hirais, H. Yukawa, H. Yurimoto, Y. Sakai, and N. Kato. 2005. Anaerobic degradation of aromatic compounds by *Magnetospirillum* strains: Isolation and degradation genes. *Biosci. Biotechnol. Biochem.* **69**: 1483–1491.
 49. Shokrollahzadeh, S., F. Azizmohseni, F. Golmohammad, H. Shokouhi, and F. Khademhaghighat. 2008. Biodegradation potential and bacterial diversity of a petrochemical wastewater treatment plant in Iran. *Bioresour. Technol.* **99**: 6127–6133.
 50. Silva, C. C., A. F. Viero, A. C. F. Dias, F. D. Andreote, E. J. Jesus, S. O. De Paula, A. P. R. Torres, V. M. J. Santiago, and V. M. Oliveira. Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load. *J. Microbiol. Biotechnol.* Doi: 10.4014/jmb.0906.06001. [In press.]
 51. Singleton, D. R., M. A. Furlong, S. L. Rathbun, and W. B. Whitman. 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* **67**: 4374–4376.
 52. Sipma, J., A. Svitelskaya, B. van der Marka, L. W. H. Pol, G. Lettinga, C. J. N. Buismana, and A. J. H. Janssen. 2004. Potentials of biological oxidation processes for the treatment of spent sulfidic caustics containing thiols. *Water Res.* **38**: 4331–4340.
 53. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.
 54. Tanji, Y., T. Kanagawa, and E. Mikami. 1989. Removal of dimethyl sulfide, methyl mercaptan, and hydrogen sulfide by immobilized *Thiobacillus thioautotrophicus* TK-m. *J. Ferment. Bioeng.* **68**: 280–285.
 55. Tarlera, S. and E. B. Denner. 2003. *Sterolibacterium denitrificans* gen. nov., sp. nov., a novel cholesterol-oxidizing, denitrifying member of the β -Proteobacteria. *Int. J. Syst. Evol. Microbiol.* **53**: 1085–1091.
 56. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.
 57. Viero, A. F., T. M. Melo, A. P. R. Torres, N. R. Ferreira, G. L. Sant'Anna Jr., C. P. Borges, and V. M. J. Santiago. 2008. The effects of long-term feeding of high organic loading in a submerged membrane bioreactor treating oil refinery wastewater. *J. Membrane Sci.* **319**: 223–230.
 58. Wang, Z. and Z. Wu. 2009. Distribution and transformation of molecular weight of organic matters in membrane bioreactor and conventional activated sludge process. *Chem. Eng. J.* **150**: 396–402.
 59. Wever, H., S. Weiss, T. Reemtsma, V. Vereecken, J. Muller, T. Knepper, et al. 2007. Comparison of sulfonated and other micropollutants removal in membrane bioreactor and conventional wastewater treatment. *Water Res.* **41**: 935–945.
 60. Zhou, J., B. Xia, D. S. Treves, L.-Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje. 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* **61**: 326–334.

**CAPÍTULO 3 – DIVERSIDADE MOLECULAR EM BIBLIOTECAS METAGENÔMICAS DE
LODO BACTERIANO PROVENIENTES DE ESTAÇÃO DE TRATAMENTOS DE
EFLUENTES ANALISADOS PELA TECNOLOGIA DE PIROSEQUENCIAMENTO
454.**

Cynthia C. Silva, Tim Sawbridge, Helen Hayden, Ana Paula R. Torres, Vânia M. J. Santiago & Valéria M. Oliveira. *Parte deste trabalho foi realizada em colaboração com Department of Primary Industries / Victoria AgroBioscience (Melbourne, Austrália).*

3.1. INTRODUÇÃO

Sistemas de tratamentos de efluentes são habitats que apresentam continuas mudanças de composição química (Szczepanowski *et al.*, 2008). Em refinarias de óleo, os efluentes que chegam às plantas são contaminados com uma fração leve de hidrocarbonetos alifáticos e aromáticos do petróleo e de uma fração organoclorada proveniente de líquidos utilizados no resfriamento do processo (Stepnowski *et al.*, 2002), além de outros compostos usualmente encontrados, como fenóis, mercaptanos, sulfetos, cloretos, ácido sulfídrico, hidróxido de sódio, cálcio e amônia, carbonatos e metais pesados (Braile, 1979, Mariano, 2001). A presença desta variedade de substratos na composição dos efluentes resulta em densa variedade de micro-organismos capazes de se adaptar rapidamente às alterações ambientais (Silva *et al.*, 2010a). Isto torna estes micro-organismos ferramentas fundamentais na biodegradação de compostos poluentes e consequentemente na proteção ambiental (Sanapareddy *et al.*, 2009)

Tradicionalmente, comunidades microbianas de amostras ambientais têm sido caracterizadas com técnicas baseadas em cultivo ou microscopia óptica. Embora muito utilizadas para taxonomia, fisiologia e estudos genéticos, estas metodologias são insuficientes para o conhecimento da composição filogenética e funcional das comunidades microbianas, considerando que apenas 1-10% dos

micro-organismos de uma amostra ambiental são cultiváveis (Amann *et al.*, 1995; Hugenholtz *et al.*, 1998).

Nas últimas décadas o desenvolvimento de ferramentas moleculares independentes de cultivo tem permitido o acesso a micro-organismos antes inacessíveis em diversos nichos ambientais (Schluter *et al.*, 2008a, Sleator *et al.*, 2008b). Entretanto, a maior parte da avaliação tem sido limitada pelo número de sequências que pode ser obtido pelos métodos de sequenciamento de Sanger (Zhang *et al.*, 2009). Os estudos de diversidade que usam construção de bibliotecas de genes RNA ribossômico 16S seguida de sequenciamento Sanger, geralmente têm mostrado uma subestimada diversidade bacteriana, o que pode ser observado pelas curvas de rarefação (Zhang *et al.*, 2009; Silva *et al.*, 2010b).

Recente tecnologia de sequenciamento, chamada de sequenciamento de 2ª geração, como o pirosequenciamento, tem alcançado um alto número de sequências, suficientes para revelar a diversidade de complexas comunidades microbianas. Esta tecnologia vem sendo considerada de baixo custo quando comparada ao sequenciamento de Sanger, devido à quantidade de informação gerada na corrida de apenas uma placa (Sogin *et al.*, 2006). A corrida de uma placa do equipamento 454-Roche Genomic Sequencer FLX pode produzir 400.000 sequências com o tamanho de 100 – 500 pares de bases, enquanto o sequenciamento de Sanger gera 96 sequências de aproximadamente 700 pares de bases (Zhang *et al.*, 2009).

Até o momento, os trabalhos de análises genômicas e/ou funcionais de lodo proveniente de estações de tratamento de efluentes usando pirosequenciamento são escassos. Szczepanowski *et al.* (2008) e Schluter *et al.* (2008a) empregaram esta ferramenta para procurar plasmídeos de resistência a

antibióticos em lodo de estações de tratamento de esgoto doméstico. Ambos os estudos verificaram, através da comparação com sequências de genes de resistência a antibióticos, que estes plasmídeos apresentam genes de resistência pertencentes à maioria das classes de drogas antimicrobianas conhecidas. Outro estudo, realizado por Zhang e colaboradores (2009) em amostras de sistema de digestor anaeróbico, empregou o pirosequenciamento com primers para a região V6 do gene RNAr 16S para avaliação da diversidade bacteriana, representada por 36.979 sequências.

Devido à importância e potencial dos micro-organismos presentes em sistemas de tratamentos de efluentes como fonte de novos genes e rotas catabólicas relacionadas com a degradação de poluentes, o presente estudo teve o objetivo de avaliar a diversidade taxonômica e funcional dos clones derivados das bibliotecas metagenômicas de DNA de alto peso molecular provenientes de lodo de efluente de refinaria de petróleo usando a técnica de pirosequenciamento.

3.2. MATERIAL E MÉTODOS

3.2.1. Amostras e extração de DNA

Duas amostras de lodo foram coletadas. A primeira amostra de lodo foi proveniente de um biorreator a membrana (MBR) operando em escala piloto no laboratório da COPPE/UFRJ, sob supervisão da Dra. Aline F. Viero (Viero *et al.*, 2008). A partida deste biorreator foi realizada com lodo proveniente da refinaria REVAP – São José dos Campos - SP. Após alimentação contínua do lodo por 30

dias com altas concentrações fenólicas ($0,086 \pm 2 \text{ mg}\cdot\text{L}^{-1}$), a amostra foi coletada para extração do DNA. A segunda amostra foi proveniente de um sistema de enriquecimento em batelada, após aclimação do lodo com $1,0 \text{ g/L}$ de fenol. O lodo para este experimento foi proveniente do biorreator a membrana piloto do Sistema de Tratamento de Efluentes da Refinaria Gabriel Passos (REGAP – Betim, MG).

3.2.2. Aclimação do lodo

Para a construção da biblioteca metagenômica enriquecida com fenol como única fonte de carbono, foi primeiramente realizada a aclimação do lodo ao fenol. O lodo proveniente do biorreator a membrana pertencente à Estação de Tratamento de Efluentes da REGAP foi inoculado em meio mínimo MP (Watanabe *et al.*, 1998), suplementado com concentrações crescentes de fenol e decrescentes de extrato de levedura.

Dois experimentos de aclimação foram realizados. No primeiro, a concentração máxima de fenol foi $1,5 \text{ g}\cdot\text{L}^{-1}$ e no segundo foi $1,0 \text{ g}\cdot\text{L}^{-1}$. O experimento foi realizado por batelada em frasco *Erlenmeyer* de 1 L contendo 300 mL de meio de cultivo (MP: $2,75 \text{ g}\cdot\text{L}^{-1} \text{ K}_2\text{HPO}_4$, $2,25 \text{ g}\cdot\text{L}^{-1} \text{ KH}_2\text{PO}_4$, $0,1 \text{ g}\cdot\text{L}^{-1} \text{ NaCl}$, $1,0 \text{ g}\cdot\text{L}^{-1} (\text{NH}_4)_2\text{SO}_4$, $0,2 \text{ g}\cdot\text{L}^{-1} \text{ MgCl}_2\cdot 6\text{H}_2\text{O}$, $0,01 \text{ g}\cdot\text{L}^{-1} \text{ CaCl}_2$) suplementado com concentrações decrescentes de extrato de levedura e crescentes de fenol. A concentração de extrato de levedura inicial foi de $1,0 \text{ g}\cdot\text{L}^{-1}$, sendo que esta foi diminuída gradativamente até a depleção, ao passo que a do fenol aumentou gradativamente de $0,2 \text{ g}\cdot\text{L}^{-1}$ até $1,5 \text{ g}\cdot\text{L}^{-1}$ (1º experimento) e até $1,0 \text{ g}\cdot\text{L}^{-1}$, (2º experimento), sendo que a depleção ou o incremento nas concentrações se deu a cada 3 dias (tempo proposto com base em revisão de literatura). Estes ensaios tiveram a duração de 45

dias. O fenol adicionado foi filtrado à vácuo em membrana hidrofílica com poro de 0,22 μM (Millipore - USA). A concentração inicial de inóculo (lodo) foi de 2,0 gL^{-1} e nos repiques subsequentes foram utilizados como inóculo 10% do volume total de meio. O tratamento meio + inóculo + fenol foi realizado em triplicata e os controles C_1 : meio + fenol, C_2 : meio+ inóculo e C_3 : meio, foram realizados em duplicata. A depleção do fenol foi acompanhada por cromatografia gasosa acoplada a detector de massa (GC/MS).

3.2.3. Taxa de degradação do fenol após aclimação

A taxa de degradação do fenol foi determinada para verificar se o lodo estava aclimatado após 45 dias de enriquecimento. A concentração do fenol no meio de cultivo foi monitorada e quantificada por cromatografia gasosa acoplada a detector de massa (GC/MS) pelo método de extração descrito no Phenol-604 da EPA. O monitoramento foi realizado a cada 3 dias de cultivo, durante os 30 dias de aclimação. A injeção dos extratos no cromatógrafo foi realizada pela Dra. Marili Villa Nova Rodrigues, da Divisão de Química Orgânica e Farmacêutica (CPQBA/UNICAMP). A quantificação de fenol por GC-MS foi realizada em cromatógrafo a gás Agilent 6890N (Agilent – USA) com detector seletivo de massas 5975 operando no modo de quantificação (modo SIM) e injetor automático 7683B. Os componentes das amostras foram separados em uma coluna capilar HP-5 MS (30 m x 0,25 mm x 0,25 μm) na seguinte programação de temperatura: 60°C (3%/min) 240°C/10 min. A temperatura do injetor foi de 220°C e a temperatura da interface foi de 250°C. Razão de split 1/100 e os fragmentos monitorados foram 94 (íon de quantificação), 66 e 65 (íons de identificação).

Após o período de 45 dias de aclimação, foi determinada a taxa de degradação do fenol pelo lodo aclimatado em meio de cultivo MP, por cromatografia gasosa. O teor de fenol foi monitorado por 72 horas, para isso foram realizadas oito coletas nos pontos: 0 h, 6 h, 12 h, 24 h, 36 h, 48 h, 60 h e 72 h. O experimento foi realizado em frascos *Erlenmeyers* de 1 L contendo 300 mL de meio MP adicionado de $1,2 \text{ gL}^{-1}$ de fenol. Frascos contendo meio + $1,2 \text{ gL}^{-1}$ fenol, sem inóculo, foram utilizados como controle. O experimento foi realizado em duplicata.

3.2.4. Construção das bibliotecas metagenômicas

Foram construídas duas bibliotecas metagenômicas de DNA de alto peso molecular clonadas em foscídios: i) DNA total extraído a partir da amostra de lodo do biorreator MBR alimentado continuamente com alta concentração fenólica por 30 dias (Biblioteca MBR) e ii) DNA total extraído do lodo proveniente da Refinaria Gabriel Passos (REGAP – Betim, MG) aclimatado com $1,0 \text{ gL}^{-1}$ de fenol (Biblioteca REGAP).

A extração do DNA de alto peso molecular de ambas as amostras foi realizada como descrito por Silva *et al.*, (2010a). O peso molecular do DNA foi verificado em gel de agarose submetido à eletroforese em campo pulsado (*Pulsed-field CHEF DRIII System* - BioRad - USA) nas condições de corrida: TEB 0,5X, 5 h, 9 Vcm^{-1} , 120° , 0,5 s - 0,5 s.

Para construção das bibliotecas metagenômicas foi utilizado o Kit *CopyControl™ HTP Fosmid Library Production* (Epicentre - USA), com algumas modificações propostas pelo Prof. Ricardo Krüger da Universidade Católica de

Brasília (comunicação pessoal). As modificações se resumiram no aumento da concentração de DNA em todas as etapas da construção da biblioteca. Portanto, antes de todas as reações o DNA foi precipitado por cerca de 16 h (*overnight*) e suspenso em água para o volume máximo de DNA que se poderia utilizar na próxima reação. Por exemplo, após o tratamento do DNA com Gelase (Epicentre - USA), este foi precipitado com acetato de sódio, como descrito no protocolo do kit, e suspenso em água para o volume de 6 μ L, exatamente o que seria empregado na próxima reação.

Primeiramente, foi selecionado o DNA de tamanho adequado para o empacotamento no fago, aproximadamente 36 Kb. Um gel de agarose *low melting point* 1% foi preparado, sendo que nele foram aplicados o DNA de interesse e o *DNA control* (36 Kb) - o qual serviu de parâmetro para excisão do DNA no tamanho correto. O gel foi submetido à eletroforese em campo pulsado (*Pulsed-field CHEF DRIII System* - BioRad- USA) nas condições de corrida: TEB 0,5X, 5 h, 9 Vcm^{-1} , 120°, 0,5 s - 0,5 s.

Após excisão do DNA com o tamanho de interesse do gel, a recuperação do DNA foi realizada pelo tratamento com a enzima Gelase. As amostras excisadas foram colocadas em microtubos e deixadas em banho-maria a 45°C para fusão da agarose. Após este passo, foi adicionada a enzima gelase (1 U/ 100 μ L de gel) e um volume apropriado de tampão Gelase 50X (previamente aquecido a 45°C) para concentração final da reação de 1X. Esta reação foi incubada a 45°C por pelo menos 1 hora, seguido de um choque a 70°C por 10 minutos para inativação da enzima. Logo após, foram transferidas alíquotas de 500 μ L para novos microtubos, incubadas no gelo por 5 minutos e centrifugadas por 20 minutos a 12000 rpm. Cerca de 90-95% do sobrenadante foi recuperado e precipitado com 1/10 volume de

acetato de sódio 3,0 M, pH 7,0 e 2,5 volumes de etanol absoluto. Após precipitação *overnight*, as amostras foram centrifugadas por 20 minutos a 12000 rpm, e o sobrenadante descartado. O precipitado (*pellet*) foi lavado com etanol 70% gelado e cuidadosamente dissolvido em água ultrapura. O tamanho e a quantidade do DNA de interesse foram confirmados em gel de agarose *pulsed field* 1% (BioRad - USA) em eletroforese de campo pulsado (*Pulsed-field CHEF DRIII System* - BioRad®), nas mesmas condições de corrida descritas anteriormente.

Após a confirmação do tamanho do DNA de interesse, foi realizada a reação de desfosforilação das extremidades do DNA, utilizando aproximadamente 100 ng de DNA para tratamento com *End-Repair*. Para isto foi preparada uma reação de 40 µL com os seguintes reagentes: 4 µL de tampão *End repair* 10X, 4 µL de dNTPs 2,5 mM, 4 µL de ATP 10 mM, 22 µL de DNA de interesse e 2 µL de enzima *End-repair Mix*. Esta solução foi incubada à temperatura ambiente por 45 minutos, seguido de 10 minutos a 37°C para desnaturação da enzima.

Após a reação de ligação do DNA no vetor, foi realizada a reação de empacotamento dos clones de foscídeos. Um tubo contendo 50 µL de *MaxPlax Lambda Packaging Extracts* (utiliza-se 1 tubo por ligação) foi descongelado à temperatura de 0 °C (gelo), sendo metade do volume adicionada à reação de ligação e incubada por 90 minutos a 30 °C. O restante do extrato foi temporariamente armazenado no freezer a -70 °C para ser, após os primeiros 90 minutos, adicionado à reação (25 µL de extrato + 10 µL de foscídeos) e incubados por mais 90 minutos a 30 °C. Passado este tempo, o volume da reação foi completado para 1 mL com tampão de diluição de fagos (10 mM Tris-HCl pH 8,3, 100 mM NaCl, 10 mM MgCl₂) e misturado cuidadosamente. Após este passo, 25 µL

de clorofórmio foram adicionados e a solução foi homogeneizada manualmente. Nesta etapa os fagos podem ser transfectados ou armazenados a 4 °C.

Para a transfecção, as células foram preparadas pela inoculação de 5 mL de cultura de *Escherichia coli* EPI300 T1^R (crescidas *overnight* a 37 °C, sob agitação de 150 rpm) em 45 mL de meio LB líquido adicionado de 10 mM de MgCl₂. Estas foram incubadas a 37 °C, sob agitação de 200 rpm até DO₆₀₀= 0,8-1,0. Após este tempo, foram aliquotados 100 µL de células para cada 10 µL de partículas de fago diluídas. Após cuidadosa homogeneização, estas soluções foram incubadas a 37 °C por 20 minutos. Para seleção dos clones de fosmídios transfectados, estes foram plaqueados em meio sólido de LB + cloranfenicol 12,5 µg/mL e incubados a 37 °C por 24 h.

Para validação da biblioteca, 5 clones de cada biblioteca foram selecionados aleatoriamente, inoculados em meio LB + cloranfenicol a 12,5 µg/mL e incubados a 37 °C por 16 horas, sob agitação de 200 rpm. O DNA fosmidial foi extraído exatamente como descrito no protocolo do Kit *FosmidMax DNA Purification* (Epicentre, USA). O DNA fosmidial foi clivado utilizando 10 U da enzima de restrição *NotI* (Promega, USA). A concentração ótima da enzima foi alcançada após sucessivos testes de reação de restrição variando a sua concentração. A reação foi incubada a 37 °C *overnight* e a 65 °C por 10 min para inativação da enzima. Após digestão, o DNA fosmidial foi submetido à eletroforese em campo pulsado nas seguintes condições: 10,5 h, 120°, 6 Vcm⁻¹, 1 s - 12 s.

3.2.5. Extração do DNA fosmidial e pirosequenciamento

Os clones das duas bibliotecas foram crescidos em placas de 96 poços contendo 1,5 mL de meio LB + cloranfenicol a 12,5 µg/mL e incubados a 37°C por 17 horas, sob agitação de 150 rpm. Uma alíquota 0,5 mL foi coletada de 960 diferentes clones formando um *pool* de 500 mL que foi utilizado como amostra para extração do DNA fosmidial. Um total de 14 *pools*, contendo 960 clones cada, foi obtido a partir dos 13.200 clones. O DNA fosmidial dos *pools* foi extraído utilizando o Kit *Large-Construction* (Qiagen - USA), de acordo com as instruções do fabricante. Este kit tem um passo para remoção do DNA da célula hospedeira *E. coli*.

Após extração do DNA fosmidial, volumes contendo concentrações equivalentes de DNA dos *pools* provenientes das duas bibliotecas foram reunidos formando uma única amostra. Assim, 5 µg de DNA fosmidial foram submetidos ao pirosequenciamento. O DNA foi nebulizado por 2 minutos, produzindo fragmentos de tamanho médio de 500 pb. O tamanho dos fragmentos foi verificado em gel de agarose. Após o tratamento das extremidades, a amostra foi purificada para remoção de fragmentos < 500 pb, de acordo com as instruções do protocolo 454/Roche GS-FLX (Roche, USA). Em seguida, adaptadores foram ligados aos fragmentos de DNA usando DNA T4 ligase para facilitar o sequenciamento no 454. Dois adaptadores foram utilizados: o primeiro adaptador dupla-fita parcial contendo o primer B de sequenciamento (5'-TCCCTGCGTGTCTCCGACTCAGCCGCGCAGGT-3') e seu complemento parcial anti-B (5'-ACCTGCGCGG-3'), e o segundo adaptador dupla-fita parcial contendo o primer A de sequenciamento (5'-TGTGTGCCTTGGCAGTCTCAGACGAGCGGCCA-3') e seu complemento parcial anti-A (5'-TGGCCGCTCGT-3'). A estrutura dupla-

fita parcial nestes adaptadores garante que eles sejam ligados na orientação correta, além da DNA ligase T4 requerer DNA dupla-fita como molde. O mix de ligação foi preparado no volume final de 25 μ L contendo: 1X de tampão NEB 2 (Invitrogen), 1X BSA (Promega, USA), 1 mM ATP, 6,5 U de DNA ligase T4 (Fermentas, USA), 1 μ M de cada adaptador e 50 ng de DNA nebulizado. Esta reação foi incubada durante 16 h a 12°C. O produto ligado foi purificado em colunas Qiagen Minielute (Qiagen, USA) para um volume final de 30 μ L.

A reação de ligação foi testada por meio de reações de PCR para verificar se as corretas construções foram produzidas. A amplificação foi realizada usando 0,1 μ M de cada um dos primers A e B (acima descritos), acrescido de 0,2 mM dNTP, 1X tampão Immolase, 0,6 U Immolase (New England Biolabs), 1 μ L do DNA ligado e o volume completado para 30 μ L. As reações de PCR foram realizadas a 94°C por 60 s, seguida de 16 ciclos de 94°C/40 s, 58°C/60s, 72°C/90 s e um ciclo a 72°C por 10 min para elongação final.

Após seleção do tamanho dos produtos de PCR de acordo com o protocolo *454/Roche GS-FLX*, a distribuição do tamanho dos fragmentos e sua quantificação foi realizada no equipamento *Agilent 2100 BioAnalyzer DNA 7500 LabChip* (Agilent - USA). Os passos seguintes foram a reação de PCR em emulsão e o sequenciamento seguindo as instruções do protocolo *454 Life Sciences GS-FLX* (Roche, USA).

3.2.6. Análise das sequências

A análise do conjunto de sequências metagenômicas, bem como a montagem de *contigs*, foi realizada utilizando o Software *DeNovo Assembler* versão 2.0.01.14 do *Genome Sequencer FLX* (454 GS Flexi, Roche, EUA). Este software triou as sequências de boa qualidade e de tamanho adequado para as análises subseqüentes e montou os *contigs* utilizando as sequências selecionadas.

Nas etapas subseqüentes, foram realizadas as análises de bioinformáticas. Foram utilizados diversos programas computacionais para anotação e agrupamento manual e automático das sequências e dos *contigs*, como BLASTn, BLASTp, MEGAN (<http://www-ab.informatik.uni-tuebingen.de/software/megan>), MG-RAST (<http://metagenomics.nmpdr.org/metagenomics.cgi?page=Logout>) e RDP (<http://rdp.cme.msu.edu/classifier/classifier.jsp>).

O BLASTn e BLASTp foram utilizados para anotação manual de aproximadamente 2.200 *contigs* (≥ 500 bases). Estes foram agrupados de acordo com a função anotada. O MEGAN foi utilizado para agrupar os dados metagenômicos (aproximadamente 123.000 sequências) em uma árvore filogenética, e também foi utilizado para o agrupamento das sequências com base nos atributos fisiológicos da comunidade. Entretanto, antes da análise no MEGAN, foi necessário anotar automaticamente todos os *contigs* no BLASTp para geração de um arquivo FASTA, que foi utilizado para entrada dos dados no programa.

A plataforma MG-RAST também foi utilizada para uma anotação *on line* de todas as sequências. Estas foram agrupadas em categorias funcionais e pelo conteúdo GC. Para a análise do conjunto de dados metagenômicos por este

programa, foi necessário salvar os dados metagenômicos em um arquivo de texto (.txt), contendo todas as sequências dos *contigs* para entrada no programa.

A análise das sequências afiliadas ao gene RNAr 16S foi realizada pelo RDP *Classifier*. Portanto, primeiramente, utilizando o BLASTn, todas as 322.752 sequências provenientes das bibliotecas metagenômicas foram comparadas às sequências de genes RNAr 16S depositadas no banco *European Ribosomal RNA Database* (<http://bioinformatics.psb.ugent.be/webtools/rRNA/>), neste passo foram selecionadas todas as sequências que apresentaram ao menos um resultado no banco ribossomal, com valor de *E-value* $\geq 1e-2$. Um valor de corte (*cutoff*) baixo para garantir a presença de todas as sequências dos genes RNAr 16S. E posteriormente, as sequências selecionadas foram submetidas à classificação automática pelo RDP *Classifier*.

3.3. RESULTADOS E DISCUSSÃO

3.3.1. Aclimação do lodo REGAP a altas concentrações de fenol

O enriquecimento do lodo com altas concentrações de fenol já havia sido realizado em experimentos anteriores, mas sem sucesso, pois o lodo não estava aclimatado, portanto foi incapaz de degradar o fenol em um período de 21 dias. Diante dos resultados anteriores, percebeu-se a necessidade da etapa de aclimação do lodo às altas concentrações de fenol para ocorrer o sucesso do enriquecimento de micro-organismos capazes de degradar ou tolerar altas

concentrações deste composto. O fenol foi adicionado gradativamente até a concentração final de $1,5 \text{ gL}^{-1}$, durante 45 dias.

Durante o experimento foi observada a formação de agregados bacterianos (flocos) e o desenvolvimento no meio de cultivo de uma coloração amarelada, que segundo Futamata *et al.* (2001) seria indicativo da degradação do fenol via *meta*-clivagem (Figura 1a). Entretanto, estas características visuais desapareceram quando foram adicionados $1,5 \text{ gL}^{-1}$ de fenol ao meio. Neste caso, após 24 horas observou-se a ruptura dos flocos e a mudança de coloração do meio, que se tornou leitoso e esbranquiçado (Figura 1b). Existem trabalhos na literatura que evidenciam a quebra do lodo ativado com o aumento da concentração fenólica. Watanabe *et al.* (1996), trabalharam com biorreatores de fluxo contínuo, observaram que cinco dias após o aumento da concentração fenólica de $0,3$ para $2,0 \text{ gL}^{-1}$ ocorreu a ruptura do lodo no tanque de aeração. Amor *et al.* (2005) observaram que concentrações fenólicas acima de $1,0 \text{ gL}^{-1}$ podem inibir fortemente a ação de micro-organismos.

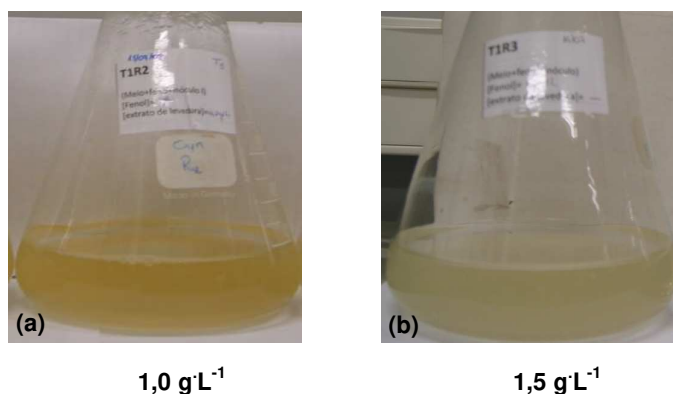


Figura 1. Enriquecimentos contendo 300 mL de meio mineral MP + fenol + 10% de lodo inóculo.

Como houve ruptura dos flocos com $1,5 \text{ gL}^{-1}$ de fenol, foi necessário repetir a fase de aclimação do lodo antes da construção da biblioteca. Portanto, o experimento de aclimação foi realizado até a concentração máxima de $1,0 \text{ gL}^{-1}$. A adaptação da biomassa foi alcançada após 45 dias, onde verificamos a degradação de $1,0 \text{ gL}^{-1}$ de fenol em menos de 24 horas (Figura 2).

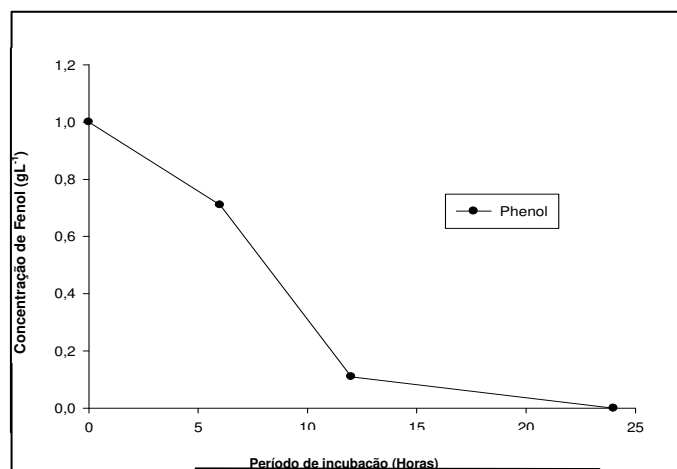


Figura 2. Degradação de $1,0 \text{ gL}^{-1}$ de fenol em meio mínimo após aclimação do lodo.

3.3.2. Construção das bibliotecas metagenômicas

Cerca de 13.200 clones foram obtidos das bibliotecas MBR e REGAP. O DNA total de alto peso molecular das comunidades bacterianas dos biorreatores MBR e REGAP foi purificado com sucesso. O tamanho do DNA a ser clonado foi então selecionado utilizando-se como parâmetro o marcador *High Molecular DNA* (Invitrogen) (Figura 3a). Toda a região do gel contendo DNA com peso molecular acima de 25 Kb foi excisada e tratada com enzima Gelase (Epicentre – USA) para recuperação do DNA da agarose. A Figura 3b evidencia o tamanho do DNA recuperado para a clonagem.

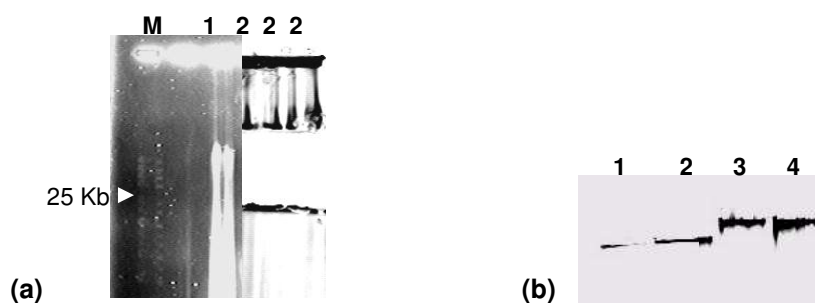


Figura 3. (a) Eletroforese de campo pulsado realizada para excisão do DNA de alto peso molecular proveniente do lodo MBR; **Canaletas:** **M** - Marcador de peso molecular *High molecular DNA* (Invitrogen), *seta*: banda de 25 kb. **1**- DNA total extraído do lodo do biorreator MBR; **2**- Região de excisão no gel para recuperação do DNA com tamanho de interesse; (b) Eletroforese em gel de agarose 1%; **Canaletas:** **1** - DNA do fago lambda (50 ng), **2** - DNA controle do Kit (control DNA - 36 kb) **3** - DNA de alto peso molecular da amostra MBR que será ligado no vetor fosmidial (1 µL), **4**- DNA de alto peso molecular da amostra MBR, que será ligado no vetor fosmidial (2 µL).

A construção das bibliotecas de fosmídios foi realizada com sucesso, sendo que cerca de 10.000 clones de *Escherichia coli* recombinantes foram obtidos do lodo MBR e 3.200 clones do lodo do biorreator da REGAP.

Para a etapa de validação, isto é, verificar se houve distintos eventos de inserção do DNA metagenômico no fosmídeo, 6 clones de cada biblioteca foram selecionados aleatoriamente e o DNA fosmidial extraído. A eletroforese em gel de agarose evidenciou a presença dos fosmídios nos clones e uma pequena contaminação de DNA cromossomal de *E. coli* EPI300T1R, o que não acarretou problemas para as etapas subsequentes (Figura 4).

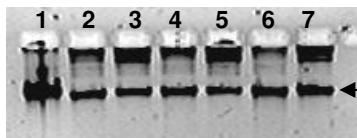


Figura 4. Extração de DNA fosmidial dos clones REGAP em gel de agarose 1%. **Canaletas:** 1 - DNA do fago Lambda (50 ng); 2 a 7 - DNA fosmidial de diferentes clones. *Seta:* contaminação de DNA cromossomal de *E. coli*.

Em seguida, o DNA dos fosmídios foi submetido a uma reação de clivagem com *NotI* (Figura 5), a fim de confirmar se os eventos de inserção eram realmente distintos. Observamos um padrão variado de restrição dos diferentes fosmídios, com presença da banda do vetor em todas as amostras, evidenciando que houve ligação do DNA metagenômico de interesse no fosmídeo.

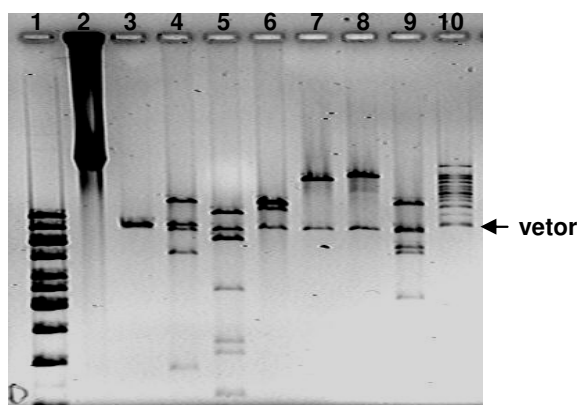


Figura 5. DNA fosmidial dos clones MBR, clivado com *NotI*, em gel de agarose 1% submetido à eletroforese de campo pulsado. **Canaletas:** 1- marcador de 1 Kb (Invitrogen); 2- DNA fosmidial não-clivado; 3- vetor pCC2Fos; 4-9: DNA fosmidial de clones distintos; **M-** Marcador *High Molecular DNA* (Invitrogen).

3.3.3. Dados do pirosequenciamento e montagem de contigs

O pirosequenciamento dos 13.200 fosmídeos resultou em 322.742 sequências (86 Mpb), cujo tamanho médio foi de 286 pb, sendo que a maior delas apresentou 735 pb. A montagem dos contigs provenientes do conjunto de dados foi realizada pelo software *GS De Novo Assembler*, versão 2.0.01.14 do *Genome Sequencer FLX* (Roche, EUA). Os contigs consistem na combinação de sequências individuais, formando longas sequências. Cerca de 108.000 sequências (33,40%) foram recrutadas para formação de 22.267 contigs, e, destes, somente 2.170 foram considerados longos, > 500 pb, pelo 454 *Sequence Assembler*.

Os resultados obtidos a partir do pirosequenciamento foram considerados satisfatórios, considerando que foram provenientes de bibliotecas metagenômicas geradas a partir de um complexo ecossistema e que foi utilizada meia placa (capacidade máxima de 180 - 280 Mpb) para o sequenciamento do DNA fosmidial contido nos 13.200 clones (aproximadamente 580,8 Mpb). Sanapareddy e colaboradores (2009) realizaram o pirosequenciamento de amostra de DNA, não clonado, proveniente de lodo de efluentes de estação de tratamento de esgoto e observaram, após a montagem dos contigs, a obtenção de somente 117 sequência maiores que 500 pb.

3.3.4. Perfil taxonômico e fisiológico das sequências metagenômicas

A análise taxonômica das 122.053 sequências metagenômicas foi realizada pelo MEGAN (*Metagenome analysis software*), onde observou-se que a maioria das sequências foi afiliada ao Domínio Bacteria (64%) e o restante se dividiu nos

Domínios Archaea e Eukarya, e 0,1% das sequências (368/122.053) foram afiliadas a DNA viral (Figura 6). Esta diversidade biológica refletiu o ambiente dos sistemas de tratamento biológico de efluentes estudados, os quais não são constituídos somente por bactérias, mas também por arqueias, fungos e protozoários. Uma parte considerável das sequências metagenômicas, 34%, não foi anotada, por serem sequências muito pequenas ou mesmo por não estarem representadas nos bancos de dados. Isto reflete o grande número de genes presentes em sistemas biológicos que ainda não foram descritos nos bancos de dados, representando possíveis novas espécies e funções importantes para o entendimento da ecologia e funcionamento deste ecossistema microbiano.

Dentre os organismos celulares o Domínio Bacteria foi representado por 94% (74.200/78.932) das sequências analisadas, que foram agrupadas em 18 diferentes filos (Figura 6). Destes, se destacaram como os mais abundantes os filos Proteobacteria (66%), seguido de Bacteroidetes (7%), Actinobacteria (5,5%), Planctomycetes (4,9%), Verrucomicrobia (1,3%) e Firmicutes (1%). Estes resultados são corroborados por dados de literatura baseados em análise de diversidade de bibliotecas de genes RNAr 16S (Miura *et al.*, 2007, Ahmed *et al.*, 2007, Li *et al.*, 2009), que reportam estes filos como normalmente presentes nos tratamentos biológicos de efluentes, sendo também as proteobacterias as mais abundantes. Os dados de diversidade filogenética derivados do pirosequenciamento são também coerentes com aqueles obtidos através da análise de bibliotecas de genes RNAr 16S para as mesmas amostras de lodo, extensivamente discutidos nos Capítulos 1 e 2 desta Tese.

Utilizando o programa MEGAN também foi gerado um perfil dos atributos fisiológicos da comunidade microbiana representada pelos dados metagenômicos

(Figura 7). Neste perfil verificou-se que a maioria das espécies é gram-negativa, apresenta respiração aeróbia, tem motilidade, não é patogênica, vive em habitat aquático, não forma endósporo, é mesofílica e de salinidade desconhecida. É importante notar, entretanto, que estes atributos representam a comunidade microbiana como um todo, englobando outros micro-organismos presentes no lodo além de bactérias. Sabendo que as bactérias formaram o maior grupo representado pelos dados metagenômicos, com 64% das sequências, acredita-se que os atributos microbianos de maiores hits reflitam o perfil da comunidade bacteriana. Consequentemente, o resultado dos atributos fisiológicos foi coerente com a composição da comunidade bacteriana descrita nos Capítulos 1 e 2 deste trabalho, que cita a predominância de proteobactérias, filo que inclui as bactérias gram-negativas, vários grupos com respiração aeróbia, que se movem utilizando flagelo, e, apesar de incluir importantes membros patogênicos, como *Escherichia coli* e *Helicobacter*, muitos membros não apresentam patogenicidade (Madigan *et al.*, 2008).

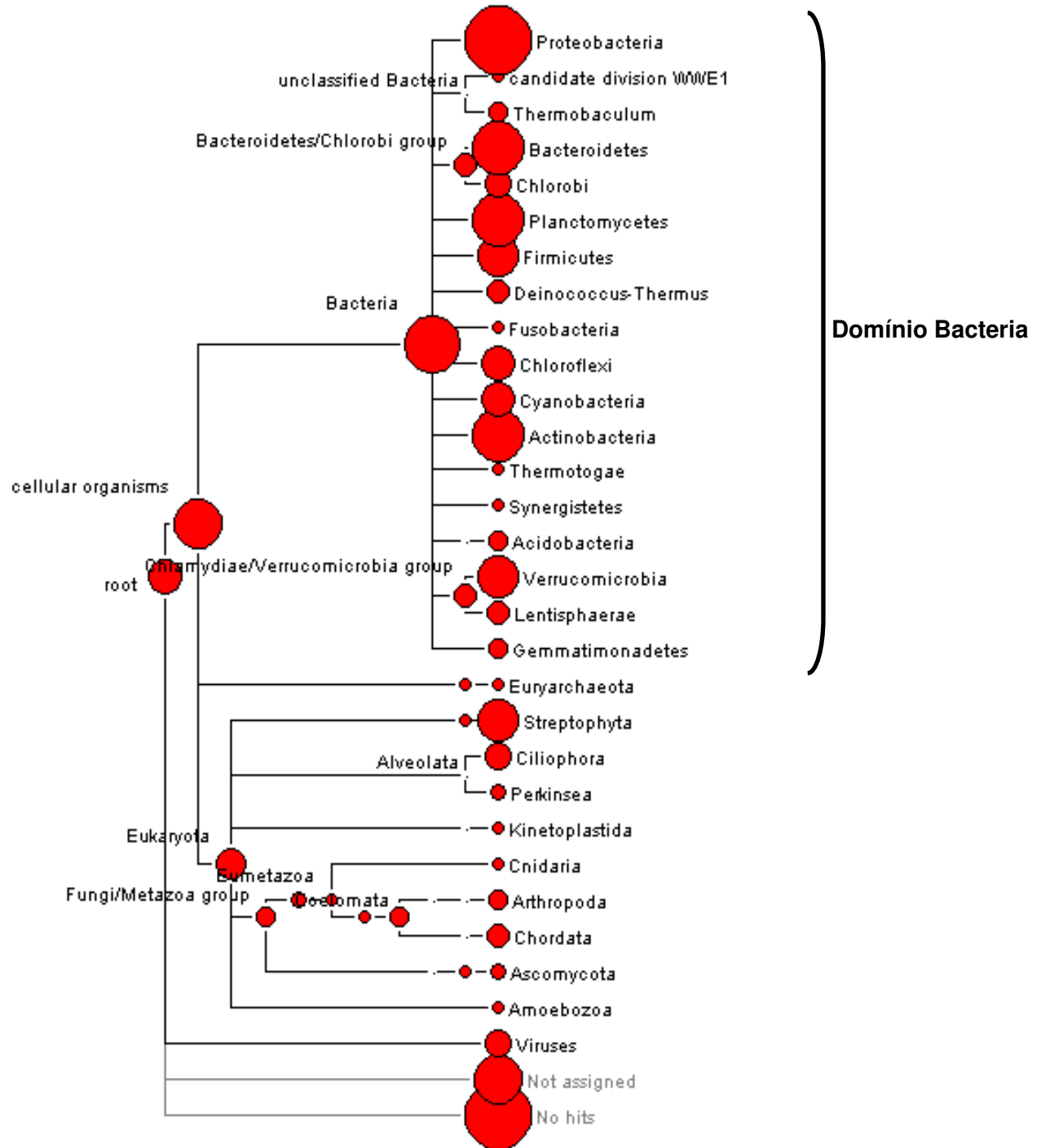


Figura 6. Árvore filogenética gerada pelo MEGAN com base no conjunto total de dados metagenômicos.

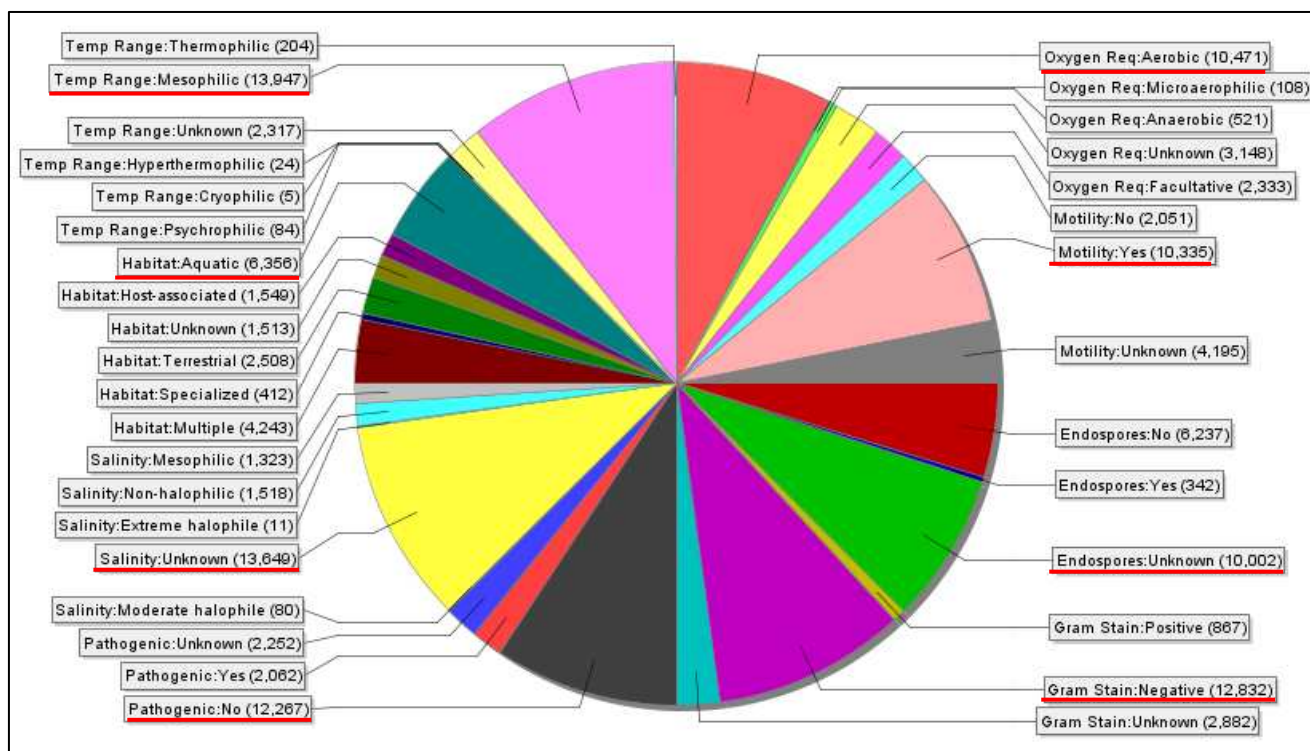


Figura 7. Perfil dos atributos fisiológicos da comunidade bacteriana gerado pelo MEGAN. O destaque em vermelho evidencia os maiores hits de cada atributo analisado.

3.3.5. Análise dos contigs >500 pb

A análise dos 2.170 contigs >500 pb para definir as possíveis funções foi realizada com auxílio do banco de dados do GenBank. O BLASTn foi utilizado para prever os genes e o BLASTp para definir as possíveis proteínas dentro do banco de dados das famílias protéicas Swissprot (banco de dados de proteínas).

Dentre os 2.170 contigs, 1.990 foram anotados, todos pertencentes ao Domínio Bacteria. Destes, 3,3% apresentaram genes que codificam proteínas de função desconhecida. Dentre as sequências de genes que codificam proteínas com função conhecida, 90% (Figura 8, “outras funções”) foram anotadas como genes constitutivos, aqueles que são expressos continuamente por todas as células, e

dentre os observados neste estudo, se destacaram aqueles que codificam para transportadores, reguladores, metabolismo de aminoácidos e ácidos nucleicos, divisão celular, sinalizadores, dentre outros (Figura 8).

Adicionalmente, foram observados genes relacionados a importantes processos biológicos que ocorrem nos sistemas de tratamento de efluentes para remoção de concentração orgânica. Exemplos destes foram os genes putativos que codificam para proteínas relacionadas com metabolismo de compostos aromáticos, como a fenol hidroxilase, que catalisa a monohidroxilação do fenol no carbono adjacente ao grupo hidroxil (Marima *et al.*, 2006), e a *p*-cresol metil-hidroxilase. Esta última enzima, em bactérias desnitrificantes, catalisa a reação de hidroxilação independente de oxigênio do *para*-cresol (Bossert & Young, 1986; Rabus *et al.*, 2005). Também foram identificados genes putativos que codificam proteínas relacionadas com a desnitrificação, processo responsável pela remoção da de nitrogênio efluente, como nitrato e nitrito redutase e 2-nitropropano dioxigenase. As enzimas nitrato e nitrito redutases catalisam a redução do nitrato e nitrito, respectivamente, a nitrogênio gasoso (Madigan *et al.*, 2008). A 2-nitropropano dioxigenase em fungos *Neurospora crassa* é responsável pela catalisação da reação de desnitrificação oxidativa de nitroalcanos (Francis & Gadda, 2009). Apesar de não termos encontrado trabalhos publicados sobre a ocorrência de 2-nitropropano dioxigenase em bactérias, a sequência metagenômica deste estudo apresentou similaridade com uma putativa enzima de *Hyphomicrobium denitrificans* ATCC 51888.

Um grupo que mereceu destaque, pelo número de hits apresentado, foram os genes que codificam para as transposases. Tal resultado nos permitiu inferir que esta comunidade provavelmente apresenta uma significativa taxa de transferência

genética horizontal entre os diferentes genomas, o que é reflexo de um ambiente que apresenta contínua oscilação em termos de tipos e concentrações de substratos. Semelhante resultado foi obtido por Sanapareddy *et al.* (2009), ao analisar por pirosequenciamento o metagenoma de lodo de estação de tratamento de esgoto.

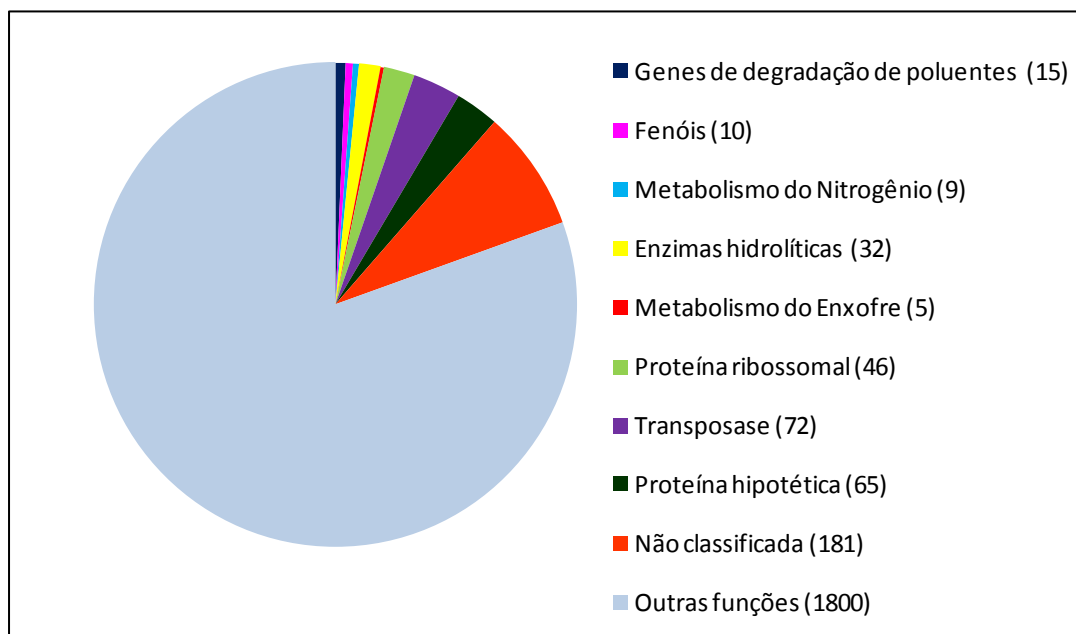


Figura 8. Análise dos contigs ≥ 500 pb usando as ferramentas de anotação BLASTn e BLASTp.

Algumas das sequências apresentaram similaridade com plasmídeos, como, por exemplo, sequência similar ao plasmídeo pTDN1 de *Pseudomonas putida* contendo genes que codificam para prováveis proteínas de conversão de anilina para catecol, reguladora e transposase. Adicionalmente, foram detectadas sequências similares a elemento de transposição composto (transposon composto) carregando genes para degradação de determinado substrato, como, por exemplo, para degradação de nonilfenol em *Sphingomonas* sp. NP5. Este foi constituído

pelos genes que codificam as putativas transposases IS20 e IS6100, os quais delimitam o gene que codifica para a enzima nonilfenol monooxigenase.

3.3.6. Análise das sequências de genes RNAr 16S dos dados metagenômicos

Das 322.752 sequências provenientes das bibliotecas metagenômicas, 6.099 sequências foram selecionadas pelo programa BLASTn como sendo relacionada ao gene 16S RNAr. Estas foram submetidas à classificação automática pelo RDP *Classifier*. Esta plataforma utiliza estatística Bayesiana para classificar as sequências gênicas de RNAr 16S dentre os taxa conhecidos. O *output* deste algoritmo inclui um *score* de confiança de 0-100, que indica um grau de confiança da anotação, semelhante ao teste de *bootstrap* 1-100 (Wang *et al.*, 2007). Segundo Sanapareddy *et al.* (2009), sequências com pelo menos 90 pb têm apresentado tamanhos suficientes para acurada caracterização taxonômica, mostrando não ser substancialmente diferente de uma análise de sequências maiores que 250 pb.

Um conjunto de 247 sequências foi classificado pelo menos em nível de Filo com base na sequência do gene RNAr 16S (Figura 9). Observou-se a presença de 6 diferentes filos, Proteobacteria, Actinobacteria, Chloroflexi, Planctomycetes, Verrucomicrobia e Bacteroidetes. O filo Proteobacteria foi o mais frequente. Este resultado já era esperado, uma vez que estudos anteriores mostraram que as proteobacterias são normalmente o filo mais abundante presente em lodo de sistemas de tratamento de efluentes de refinaria (Silva *et al.*, 2010a, 2010b, Miura *et al.*, 2007).

Foram encontradas 11 diferentes ordens, sendo Burkholderiales, que pertence à classe Betaproteobacteria, a ordem mais abundante, seguida de Rhizobiales (Alfaproteobacteria) e Pseudomonadales (Gamaproteobacteria). Somente 33% das sequências foram classificadas em gêneros, dos quais se destacam *Diaphorobacter*, seguido por *Pseudomonas*, *Hyphomicrobium*, *Comamonas*, *Planctomyces* e *Thauera*. Com exceção de *Planctomyces*, todos os outros gêneros pertencem ao filo Proteobacteria, se dividindo nas classes Beta- (*Diaphorobacter*, *Thauera*, *Comamonas*), Gama- (*Pseudomonas*) e Alfaproteobacteria (*Hyphomicrobium*). Alguns destes gêneros, como *Thauera*, *Pseudomonas* e *Comamonas*, apresentam espécies já descritas como relacionadas à degradação de compostos aromáticos, como *Pseudomonas putida* (Gonzales *et al.*, 2001), *Thauera aromatica* (Breining *et al.*, 2000) e *Comamonas testosteroni* (Arai *et al.*, 2000). Membros de *Diaphorobacter* e *Hyphomicrobium* já foram descritos como relacionados ao metabolismo do nitrogênio. Espécies de *Diaphorobacter* são capazes de realizar simultaneamente os processos de nitrificação e desnitrificação (Khardenavis *et al.*, 2007), e *Hyphomicrobium* spp. já foram isolados de sedimento de subsuperfície contaminado com nitrato (Green *et al.*, 2010).

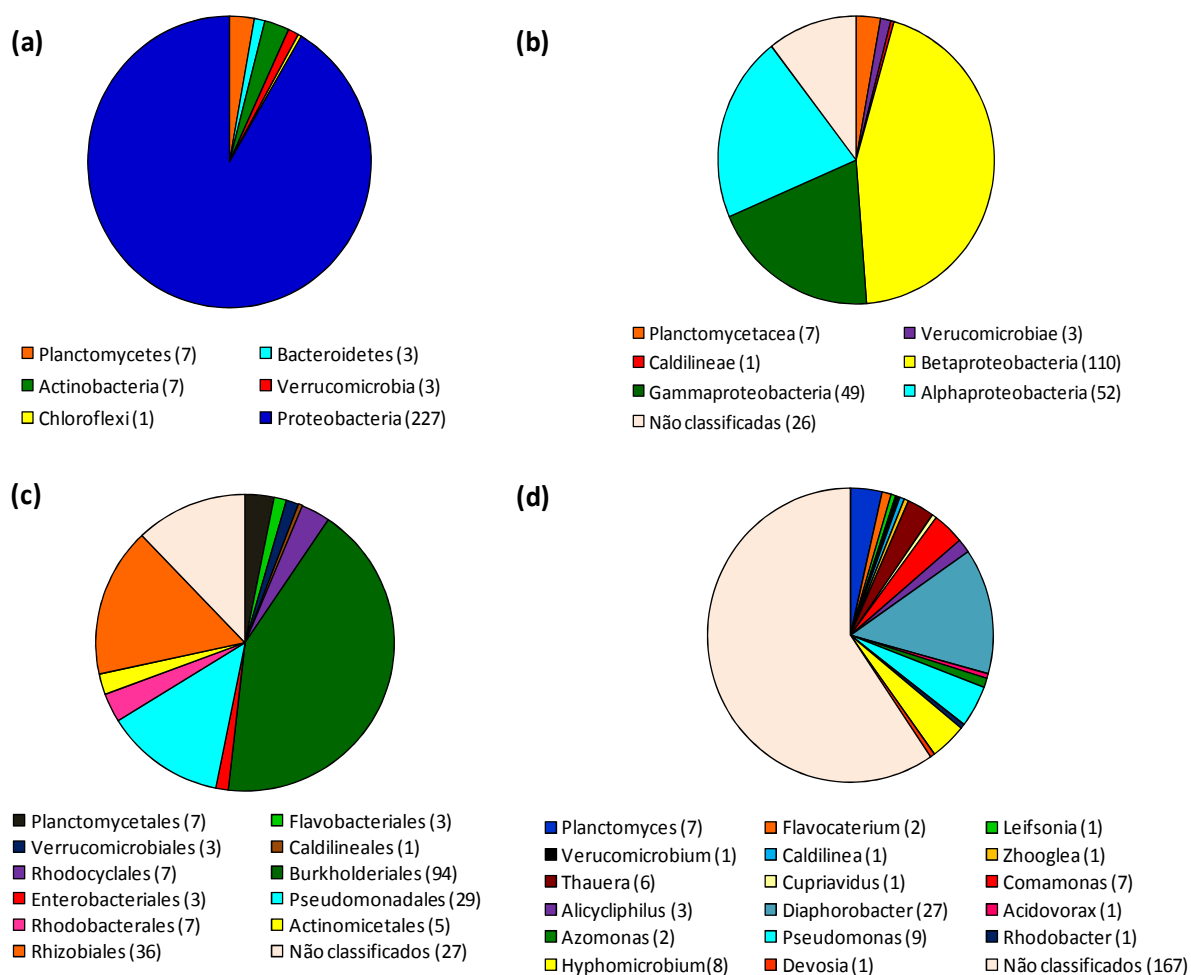


Figura 9. Classificação das sequências de genes RNAr 16S baseadas no RDP Classifier. **(a)** Filo, **(b)** Classe, **(c)** Ordem **(d)** Gênero.

Uma considerável parte das sequências encontradas não pôde ser classificada em nível de gênero, o que pode ser indicativo da presença de novos grupos taxonômicos, além de evidenciar a enorme diversidade microbiana que está presente nos lodos de tratamento de efluentes. Resultado semelhante foi encontrado por Sanapareddy *et al.* (2009) ao analisar dados de pirosequenciamento de lodo proveniente esgoto, onde mais de 60% das sequências não foram afiliadas.

3.3.7. Análise do conteúdo GC das sequências metagenômicas

A análise do conteúdo GC (%[G+C]), percentagem de ácido nucléicos que consiste de guanina e citosina, das sequências foi realizada pelo SEED/ MG-RAST (<http://metagenomics.nmpdr.org>), uma plataforma para análise de dados metagenômicos que consiste em um serviço automatizado para anotações de genomas de bactérias e arqueias. A Figura 10 mostra a percentagem da distribuição do conteúdo GC das sequências metagenômicas determinadas neste estudo.

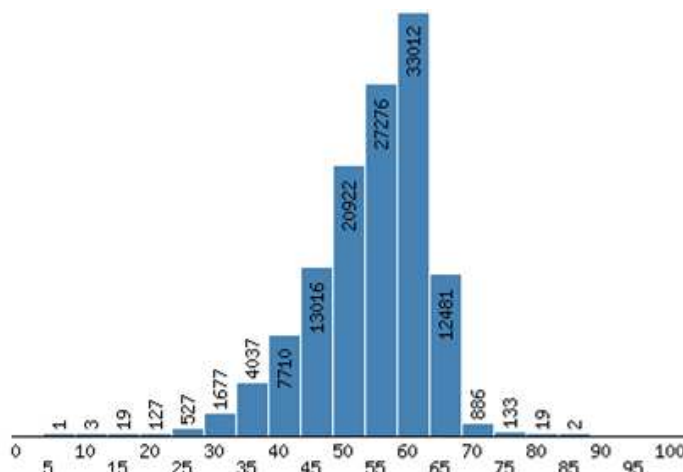


Figura 10. Perfil da distribuição das percentagens do conteúdo G+C das sequências metagenômicas. O Eixo X representa percentagem, e o eixo Y representa o número de sequências na faixa de conteúdo G+C.

Foi observado um perfil de distribuição de conteúdo G+C das sequências metagenômicas na faixa de 15 a 80%. No entanto, foi na faixa de 45-69% onde se verificou o maior número de sequências, sendo significativamente dominante o conteúdo G+C entre 50% a 64%. De acordo com Madigan *et al.* (2008), valores acima de 50% podem ser considerados como alto conteúdo de GC. Entretanto, não existe uma faixa definida de %[G+C] onde se enquadram as proteobacterias,

maioria encontrada neste estudo, como ocorre no filo Actinobacteria (alto GC). Portanto, fizemos um estudo do conteúdo G+C dos gêneros mais encontrados pelo RDP. Usando os sites <http://insilico.ehu.es/oligoweb/index2.php?m=all>, <http://www.kazusa.or.jp/codon/> e Manual do Bergey's, observamos que a %[G+C] do gênero *Diaphorobacter* foi de 66,8%, *Thauera* de 64-69%, *Hiphomicrobium* de 59-65%, *Pseudomonas* na faixa de 53,2 a 66,6%, *Comamonas* de 60-69% e *Planctomyces* de 53,97%.

Verificamos que a faixa de %[G+C] gerado pelas sequências metagenômicas foi coerente às percentagens encontradas para os gêneros mais abundantes presente no metagenoma estudado. Embora este agrupamento seja menos utilizado, atualmente, o conteúdo GC pode auxiliar na classificação dos micro-organismos, principalmente daqueles que não apresentam o gene RNAr 16S sequenciado e depositado no banco de dados (Madigan *et al.*, 2008).

3.3.8. Perfil funcional das sequências metagenômicas

Para uma análise do perfil funcional dos dados metagenômicos, foi utilizada a plataforma SEED/ MG-RAST. Este classifica as rotas metabólicas em estruturas hierárquicas, nas quais todos os genes requeridos para específica tarefa são organizados em um subsistema. Nos altos níveis de organização, os subsistemas incluem as funções de catabolismo e anabolismo, como por exemplo, metabolismo de DNA, proteínas, compostos aromáticos, e nos baixos níveis de subsistemas estão as vias específicas, como síntese de determinado aminoácido, metabolismo de determinado composto aromático.

Assim, utilizando o BLASTx e o *cutoff* do *E-value* de 0,01, a plataforma SEED/MG-RAST (Meyer *et al.*, 2008) foi capaz de anotar 42,4% das sequências, agrupando-as em sistemas funcionais estabelecidos pela plataforma SEED. Observamos que a distribuição das sequências metagenômicas mostrou-se bastante diversa (Figura 11). O perfil metabólico observado foi semelhante ao descrito por Sanapareddy *et al.* (2009), ao analisarem por pirosequenciamento amostras de lodo proveniente de sistemas de tratamento de esgoto.

Apesar de alguns sistemas terem apresentado um baixo número de hits positivos, todas as categorias foram amostradas pelos dados metagenômicos. A baixa frequência de hits encontrados em algumas categorias funcionais é perfeitamente explicável, como, por exemplo, no caso das proteínas envolvidas com processos fotossintéticos, uma vez que estas bactérias vivem em biorreatores, sistemas que recebem baixa intensidade luminosa. Este resultado nos confirma que estes micro-organismos retiram sua fonte de energia primária dos próprios compostos orgânicos presentes na planta de tratamento. Observamos também que esta é uma comunidade que apresenta grande potencial metabólico, o que pôde ser evidenciado, principalmente, pelo alto número de genes que codificam para proteínas envolvidas nos processos de divisão celular (2,72%), respiração (4,17%), metabolismo de DNA (4,17%), RNA (2,85%) e proteínas (7,36%).

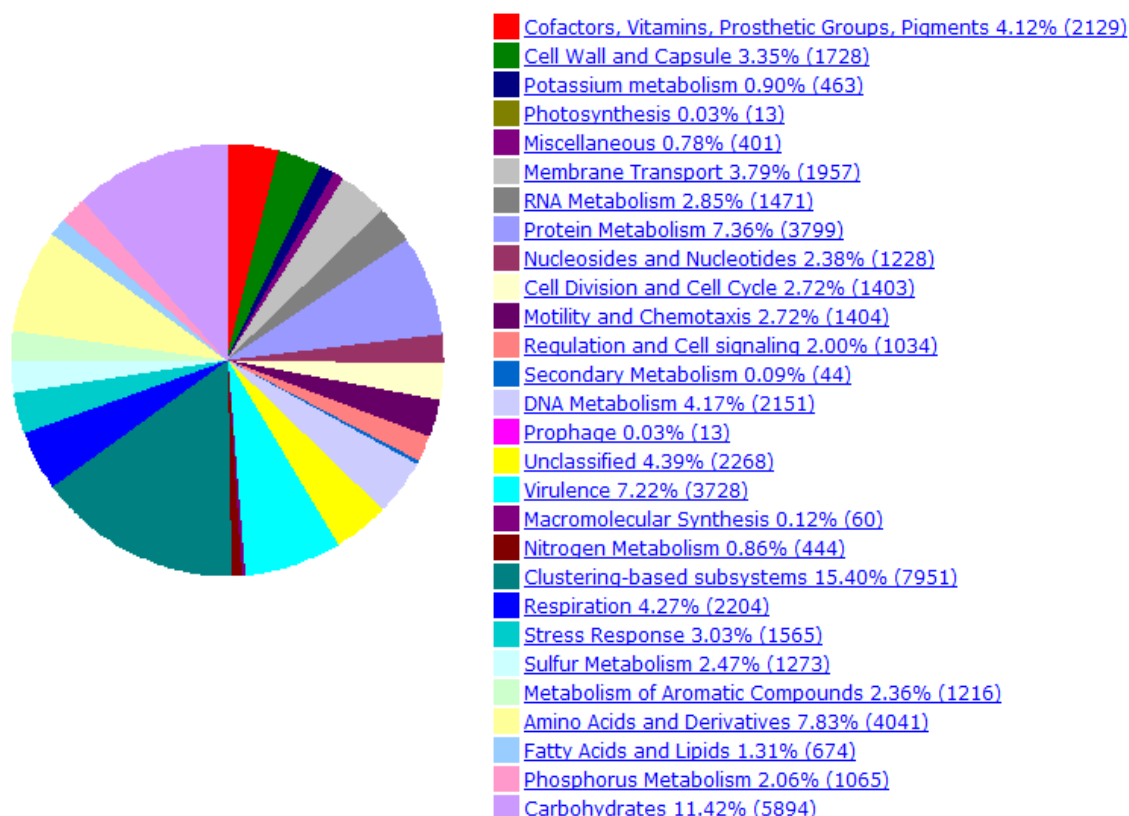


Figura 11. Distribuição das sequências metagenômicas nas categorias funcionais definidas pela plataforma SEED.

Adicionalmente, é importante chamar a atenção para os genes relacionados com o metabolismo de compostos aromáticos, como bifenil, benzoato, cresol, além de genes relacionados com a via meta-clivagem de degradação de compostos aromáticos, como fenol hidroxilase e catecol dioxigenase. Além destes, foram encontrados genes relacionados à assimilação orgânica e inorgânica de enxofre, ao metabolismo de fosfatos e fosfonatos, e ao metabolismo do nitrogênio, representado por genes para amonificação, fixação de nitrogênio, assimilação de amônia, desnitrificação, dentre outros. Estes resultados evidenciam o alto potencial desta comunidade bacteriana para realizar processos biológicos importantes para o tratamento de efluentes industriais.

Outro resultado que mereceu destaque foi o alto índice de genes relacionados com resposta ao estresse (3,03%), resultado condizente com o ambiente dinâmico de um biorreator de refinaria, que está a todo instante recebendo concentrações contendo os mais variados compostos tóxicos e poluentes. Enfim, considerando a quantidade de sequências deste conjunto metagenômico anotado e a diversidade de categorias ocupadas, acreditamos que esta comunidade apresenta um amplo perfil metabólico.

3.4. CONCLUSÕES

O pirosequenciamento dos 13.200 clones pertencentes às bibliotecas metagenômicas REGAP e MBR gerou aproximadamente 322.742 sequências (~ 86 Mpb), cuja análise revelou um perfil de comunidade microbiana bastante diverso, com a presença de 18 diferentes filos do Domínio Bacteria. Isto resultou num amplo perfil metabólico, representado por genes para as mais diversas funções responsáveis pela remoção da concentração orgânica presente nos efluentes de refinarias de petróleo. Estes genes mostraram-se também relacionados com a remoção de compostos recalcitrantes e poluentes, como amônia, fosfatos e hidrocarbonetos, incluindo fenol, tolueno, cresol, dentre outros. Outro aspecto que mereceu destaque nestas análises foi a grande quantidade de sequências com baixa ou nenhuma similaridade com sequências já conhecidas e depositadas nos bancos de dados do NCBI, sugerindo a presença nesta amostra de lodo de possíveis novas espécies, genes e funções ainda não descritas.

3.5. REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmed, Z., Cho, J., Lim, B-R., Song, K-G., Ahn, K-H. 2007. Effects of sludge retention time on membrane fouling and microbial community structure in a membrane bioreactor. *Journal of Membrane Science*. 287: 211-218.
- Amor, L., Eiroa, M., Kennes, C. & Veiga, M. C. 2005. Phenol biodegradation and its effect on the nitrification process. *Water Research*. 39: 2915-2920.
- Arai, H., Akahira, S., Ohishi, T., Maeda, M. & Kudo, T. 1998. Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: organization and regulation of the genes involved in phenol degradation. *Microbiology*. 144: 2895-2903.
- Bossert, I. D. & Young, L. Y. 1986. Anaerobic Oxidation of p-Cresol by a Denitrifying bacterium. *Applied and Environmental Microbiology*. 52: 1117-1122.
- Braile, P. M. 1979. Manual of Industrial Wastewater Treatment, 1ed. São Paulo, CETESB.
- Breener, D. J., Krieg, N. R. & Staley, J. T. 2005. **Bergey's manual of systematic bacteriology. Volume 2.** 2nd Edition. Springer Ed., New York, EUA. ISBN: 0-387-24143-4.
- Breinig, S., Schiltz, e. & Fuchs, G. 2000. Genes Involved in Anaerobic Metabolism of Phenol in the Bacterium *Thauera aromatica*. *Journal of Bacteriology*. 5849-5863.
- Futamata, H., Harayama, S. & Watanabe, K. 2001. Diversity in kinetics of trichloroethylene-degrading activities exhibited by phenol-degrading bacteria. *Applied Microbiology and Biotechnology*. 55: 248-253.
- González, G. Herrera, G., García, M. T. & Peña, M. 2001. Biodegradation of phenolic industrial wastewater in a fluidized bed bioreactor with immobilized cells of *Pseudomonas putida*. *Bioresource Technology*. 80: 17-142.
- Green, S. J., Prakash, O., Gihring, T. M., Akob, D. M., Jasrotia, P., Jardine, P. M., Watson, D. B., Brown, S. D., Palumbo, A. V., Kostka, J. E. 2010. Denitrifying bacteria from the terrestrial subsurface exposed to mixed waste contamination. *Applied and Environmental Microbiology*. (In print).
- Kevin, F. & Gadda G. 2009. Inflated Kinetic Isotope Effects in the Branched Francis Mechanism of *Neurospora crassa* 2-Nitropropane Dioxygenase. *Biochemistry*. 48: 2403-2410.

- Khardenavis, A. A., Kapley, Hemant, A. & Purohit, J. 2007. Simultaneous nitrification and denitrification by diverse *Diaphorobacter* sp. *Applied Microbiology and Biotechnology*. 77: 403-409.
- Lin, S., Jin, Y., Fu, L., Quan, C. & Yang, Y.S. 2009. Microbial community variation and functions to excess sludge reduction in a novel gravel contact oxidation reactor. *Journal of Hazardous Materials*. 165: 1083-1090.
- Madigan M., Martinko, J., Dunlap, P. & Clark, D. 2008. **Brock Biology of Microorganisms**. 12 Ed. ISBN: 0-13-232460-1.
- Mariano, J. B. (2001) Impactos ambientais do refino de petróleo. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- Merimaa, M. Heinaru, E., Liivak, M. Vedler, E. & Heinaru, A. 2006. Grouping of phenol hydroxylase and catechol 2,3-dioxygenase genes among phenol- and *p*-cresol-degrading *Pseudomonas* species and biotypes. *Archives Microbiology*. 186:287-296.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., *et al.* 2008. The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9:386.
- Miura, Y., Hiraiwa, M. N., Ito, T., Itonaga, T., Watanabe, Y. & Okabe, S. 2007. Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance. *Water Research*. 41: 627 - 637.
- Rabus, R., Kube, K., Heider, J., Beck, A., Heitmann, K., Widdel, F. Reinhardt, R. 2005. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Archives of Microbiology*. 183: 27-36.
- Sanapareddy, N., Hamp, T. J., Gonzalez, L. C., Hilger, H. A. Fodor, A. A. & Clinton, S. M. 2009. Molecular Diversity of a North Carolina Wastewater Treatment Plant as Revealed by Pyrosequencing. *Applied and Environmental Microbiology*. 175: 1688-1696.
- Silva, C. C., Jesus, E. C., De Paula, S. O., Torres, A. P. R., Sousa, M. P., Santiago, V. M. J. & Oliveira, V. M. 2010b. Investigation of Bacterial Diversity in Membrane Bioreactor and Conventional Activated Sludge Processes from Petroleum Refineries Using Phylogenetic and Statistical Approaches. *Journal of Microbiology and Biotechnology*. 20: 447-459.
- Silva, C. C., Viero, A. F., Dias, A. C. F., Andreote, F. D., Jesus, E. C., De Paula, S. O., Torres, A. P. R., Santiago, V. M. J. & Oliveira, V. M. 2010a. Monitoring the Bacterial Community Dynamics in a Petroleum Refinery Wastewater Membrane

- Bioreactor Fed with a High Phenolic Load. *Journal of Microbiology and Biotechnology*. 20:17-25.
- Sleator, R. D., Shortall, C. & Hill, C. 2008. Under the microscope: Metagenomics. *Letters in Applied Microbiology*. 47: 361-366.
- Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta & G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *PNAS – USA*. 103:12115-12120.
- Stepnowski, P., Siedlecka, E.M., Behrend, P. & Jastorff, B. 2002. Enhanced photo-degradation of contaminants in petroleum refinery wastewater. *Water Research*. 36: 2167-2172.
- Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., Eichler, W., Pühler, A., Schlüter, A. 2008. Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *Journal of Biotechnology*. 136: 54-64.
- Viero, A. F., Melo, T. M., Torres, A. P. R., Ferreira, N. R., Sant’Anna Jr, G. L., Borges, C. P. & Santiago, V. M. J. 2008. The effects of long-term feeding of high organic loading in a submerged membrane bioreactor treating oil refinery wastewater. *Journal of Membrane Science*. 319: 223-230.
- Watanabe, K., Hino, S. A. & Takahashi, N. 1996. Responses of activated sludge to an increase in phenol loading. *Journal of Fermentation and Bioengineering*. 82: 522-524.
- Zhang, H., Banaszak, J. E., Parameswaran, P., Alder, J., Krajmalnik-Brown, R. & Rittmann, B. E. 2009. Focused-Pulsed sludge pre-treatment increases the bacterial diversity and relative abundance of acetoclastic methanogens in a full-scale anaerobic digester. *Water Research*. 43: 4517- 4526.

**CAPÍTULO 4. PROSPECÇÃO DE NOVOS GENES RELACIONADOS À DEGRADAÇÃO DE
FENOL E DERIVADOS EM BIBLIOTECAS METAGENÔMICAS DE LODO DE
REFINARIA DE PETRÓLEO.**

Cynthia C. Silva, Tim Sawbridge, Helen Hayden, Ana Paula R. Torres, Vânia M. J. Santiago & Valéria M. de Oliveira. *Parte realizada em colaboração com Department of Primary Industries / Victoria AgroBioscience (Melbourne, Australia).*

4.1. INTRODUÇÃO

Fenol e compostos fenólicos são os principais poluentes orgânicos encontrados nos efluentes de indústrias químicas, têxteis, farmacêuticas, de refino de petróleo, dentre outras (Basile & Erijman, 2008). Além de causar sérios danos ao meio ambiente e seres vivos, a presença destes compostos nos efluentes pode inibir a taxa de crescimento microbiano, e consequentemente levar a uma redução significativa da degradação biológica de outros componentes presentes no sistema (Barrios-Martinez *et al.*, 2006; Cordova-Rosa *et al.*, 2009). Neste sentido, existe uma preocupação constante das refinarias de petróleo para remoção destes contaminantes dos efluentes (Barrios-Martinez *et al.*, 2006). Vários processos são utilizados para eliminação dos compostos fenólicos do ambiente, entretanto os tratamentos biológicos vêm se sobressaindo em relação a tratamentos físico-químicos existentes, pois além de ser um processo natural, que utiliza micro-organismos encontrados na natureza para mineralizar compostos orgânicos poluentes em água e gás carbônico, apresentam baixo custo operacional (Dong *et al.*, 2008).

Em decorrência deste processo, muitos micro-organismos que degradam fenol têm sido estudados, tais como bactérias, *Acinetobacter*, *Alcaliniger*, *Thauera*, *Azoarcus*, *Comamonas*, *Pseudomonas*, *Bacillus*, etc, e leveduras, como *Candida tropicalis* (Valle *et al.*, 2004, Nair *et al.*, 2007; Yan *et al.*, 2006; Sueoka *et al.*, 2009,

Cordova-Rosa *et al.*, 2009). A maioria destes micro-organismos utiliza uma típica rota de degradação aeróbia do fenol, composta por dois passos críticos: no primeiro passo, o anel aromático do fenol é inicialmente monohidroxilado no carbono adjacente ao grupo hidroxil existente pela fenol hidroxilase (ou fenol 2-monooxigenase) formando catecol; e, no segundo, ocorre a clivagem do anel do catecol pelas vias *orto* ou *meta*-clivagem. Na via *orto* o anel é clivado pela enzima catecol 1,2-dioxigenase (C12O), e, caso a via seja *meta*, a clivagem do catecol é realizada pela catecol 2,3-dioxigenase (C23O) (Heinaru *et al.*, 2006). Existem dois tipos de fenol hidroxilase presentes em bactérias, a simples componente e a multicomponente (mPHs), esta última, segundo Peters *et al.* (1997), é considerada a principal fenol hidroxilase encontrada no ambiente.

Numerosos estudos de catabolismo microbiano do fenol, em termos de enzimologia, diversidade microbiana e biorremediação, têm sido realizados com as enzimas fenol hidroxilases e catecol dioxigenases para o melhor entendimento da degradação de compostos aromáticos (Merimaa *et al.*, 2006). Entretanto, até o momento, nenhum estudo utilizou bibliotecas metagenômicas de DNA de alto peso molecular para explorarem a presença destas e novas enzimas relacionadas com a degradação de fenol e outros compostos aromáticos. As bibliotecas metagenômicas são construídas com DNA de alto peso molecular extraído diretamente de amostras ambientais e inseridas em vetores que suportam grandes fragmentos de DNA, variando de, aproximadamente, 25 a 150 kb. Estes fragmentos são ligados em vetores como cosmídeos, fosmídeos e BACs e são transformados em células hospedeiras, normalmente *E. coli*. As bibliotecas resultantes podem ser triadas para várias sequências ou funções, na busca de novos produtos, genes e rotas metabólicas de interesse presentes, principalmente, em organismos ainda não

cultivados. Portanto, este trabalho teve o objetivo de explorar bibliotecas metagenômicas de alto peso molecular provenientes de lodo de refinaria de petróleo, na busca de novos genes ou rotas metabólicas relacionadas com a degradação de fenol e derivados, utilizando estratégias de triagens baseadas em similaridade de sequência e atividade biológica.

4.2. MATERIAL E MÉTODOS

4.2.1. Construção e validação das bibliotecas metagenômicas

Foram construídas duas bibliotecas metagenômicas com fragmentos de DNA de 25-36 Kb. Uma amostra foi proveniente do biorreator MBR operado em escala piloto na COPPE (UFRJ), alimentado continuamente com alta concentração fenólica por 30 dias (Biblioteca MBR), e outra com lodo proveniente do biorreator MBR operado em escala piloto na Refinaria Gabriel Passos (REGAP – Betim, MG), após aclimação com $1,0 \text{ g} \cdot \text{L}^{-1}$ de fenol (Biblioteca REGAP).

A extração de DNA de alto peso molecular de ambas as amostras foi realizada como descrito por Silva *et al.* (2010a). Após seleção do fragmento de DNA de tamanho de interesse, a construção das bibliotecas metagenômicas foi realizada utilizando o Kit *CopyControl™ HTP Fosmid Library Production* (Epicentre, USA), como descrito anteriormente no Capítulo 3.

A validação das bibliotecas foram realizadas pela digestão de 5 clones de cada biblioteca, MBR e REGAP, selecionados ao acaso, e clivados com a enzima *NotI* (Promega, USA). Para isto, os clones foram inoculados em meio LB +

cloranfenicol a 12,5 µg/mL e incubados a 37°C por 16 horas, sob agitação de 200 rpm. O DNA fosmidial foi extraído como descrito no protocolo do Kit *FosmidMax DNA Purification* (Epicentre, USA) e clivado com a utilização de 10 U da enzima de restrição. A reação foi incubada a 37°C *overnight* e a 65°C por 10 min para inativação enzimática. Após digestão, o DNA fosmidial foi submetido à eletroforese em campo pulsado nas seguintes condições: 10,5 h, 120°, 6 Vcm⁻¹, 1 s - 12 s.

4.2.2. Screening do gene que codifica a enzima fenol hidroxilase

Todos os clones das bibliotecas metagenômicas, MBR e REGAP, foram triados para o gene que codifica a subunidade maior da enzima fenol hidroxilase. Estes foram cultivados em placas de 96 poços contendo 150 µL LB + 12,5 µg/mL de cloranfenicol, e incubadas a 37°C por 16 horas. O *screening* para o gene que codifica para a fenol hidroxilase foi feito através de reações de PCR e foi realizado em duas etapas, a primeira com alíquotas de *pools* contendo cultivos provenientes de 10 clones, e uma segunda etapa em que o PCR foi realizado somente com os clones provenientes dos *pools* que apresentaram sinal positivo de amplificação. Este segundo PCR foi necessário para identificação individual do clone positivo para o gene que codifica a enzima fenol hidroxilase. Os *pools* foram preparados pela adição de 2 µL de cultura de cada clone em 10 µL de água ultrapura, e após homogeneização da mistura 5 µL foram utilizados nas reações de PCR. As reações continham 0,5 pmol/µL dos primers pheUf/pheUr (Futamata *et al.*, 2001), genéricos para o gene da fenol hidroxilase, 0,2 mM de dNTPs, 1X de tampão da enzima, 0,1 mM de BSA e 2,5 U de *Taq* DNA Polimerase (Invitrogen, USA) para um volume final

de 50 μL . As condições de amplificação utilizadas foram 10 min a 94°C, 5 ciclos de 1 min a 94°C, 1 min a 58°C e 1 min a 72°C, seguidos de 25 ciclos de 1 min a 94°C, 1 min a 56°C e 1 min a 72°C, e um ciclo de extensão final a 72°C por 10 min.

4.2.3. Screening para atividade de degradação de fenol

Os clones das bibliotecas metagenômicas, MBR e REGAP, foram triados para a capacidade de crescer em fenol como única fonte de carbono. Estes foram reativados em placas de 96 poços contendo 150 μL LB + 12,5 $\mu\text{g/mL}$ cloranfenicol e 10% destas culturas foram utilizadas como o inóculo para o meio seletivo. Este foi constituído de 150 μL de meio mineral (5,0 g L^{-1} NaCl, 1,0 g L^{-1} K_2HPO_4 , 1,0 g L^{-1} NaHPO_4 , 1,0 g L^{-1} NH_3SO_4 , 0,2 g L^{-1} $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 3,0 g L^{-1} K_2NO_2), acrescido de 10% de solução de vitaminas (Vasconcellos *et al.*, 2009) e 0,02% de solução de fenol estéril. As culturas foram incubadas a 37°C por 24 h. A detecção dos clones que foram capazes de crescer em fenol como única fonte de carbono foi realizada pelo método colorimétrico do MTT (3-(4,5-dimetiltiazol-2-yl)- 2,5-difeniltetrazolium brometo), um corante sintético amarelo que em contato com células vivas é reduzido à Formazano, um composto de coloração violeta. Assim, em cada cultura de clone foram adicionados 30 μL de MTT 10%, seguido pela incubação das placas de 96 poços a 37°C por 1 hora. Os clones positivos foram detectados pela formação da cor violeta.

4.2.4. Ensaios de degradação do fenol por cromatografia gasosa

Antes de realizar os ensaios de degradação do fenol, testes preliminares foram feitos com intuito de otimizar o volume de meio de cultivo. Foram realizados dois tratamentos em meio mineral acrescido de 10% de solução de vitaminas (descritos acima) + 0,02% de fenol + 10% de inóculo (cultura de clone positivo nas triagens de sequência): tratamento 1 (T1), com volume final de 300 mL, e tratamento 2 (T2), com volume final de 50 mL. As amostras foram coletadas nos pontos: 0, 7, 14, 21 e 40 dias. Foram realizados os seguintes controles experimentais: C₁ = meio + 0,02% de fenol; C₂ = meio + 10% de inóculo; C₃ = meio.

Todos os pontos coletados foram submetidos à cromatografia gasosa acoplada a detector de massa (GC/MS) pelo método de extração descrito no Phenol-604 da EPA para determinação da concentração do fenol no meio. A quantificação de fenol por GC-MS foi realizada em cromatógrafo a gás Agilent 6890N (Agilent – USA) com detector seletivo de massas 5975 operando no modo de quantificação (modo SIM) e injetor automático 7683B. Os componentes das amostras foram separados em uma coluna capilar HP-5 MS (30 m x 0,25 mm x 0,25 µm), na seguinte programação de temperatura: 60°C (3%/min) 240°C/10 min. A temperatura do injetor foi de 220°C e a temperatura da interface foi de 250°C. Razão de split 1/100 e os fragmentos monitorados foram 94 (íon de quantificação), 66 e 65 (íons de identificação).

4.2.5. Extração do DNA fosmidial e pirosequenciamento

Foram selecionados, ao acaso, 100 clones que apresentaram resultados positivos nos *screenings* anteriores para a realização do pirosequenciamento. Os clones selecionados foram cultivados em 10 mL de LB + 12,5 µg/mL de cloranfenicol e incubados a 37°C por 17 horas, sob agitação de 150 rpm. O DNA fosmidial foi extraído a partir de um *pool* de 500 mL de culturas, formado por 5 mL de cada um dos 100 clones selecionados, utilizando o *Large-Construction Kit* (Qiagen, USA), de acordo com as instruções descritas pelo fabricante. Após extração do DNA fosmidial, 5 µg de DNA fosmidial foram submetidos ao pirosequenciamento. O DNA foi nebulizado por 90 segundos, produzindo fragmentos de tamanho médio de 500 pb. O tamanho dos fragmentos foi verificado através de eletroforese em gel de agarose. Após o tratamento das extremidades, a amostra foi purificada para remoção de fragmentos < 500 pb, de acordo com as instruções do protocolo 454/Roche GS-FLX (Roche, USA). Após esta etapa, adaptadores foram ligados aos fragmentos de DNA para facilitar o sequenciamento no 454. O primeiro adaptador dupla-fita parcial continha o primer B de sequenciamento (5'-TCCCTGCGTGTCTCCGACTCAGCCGCGCAGGT-3') e seu complemento parcial anti-B (5'-ACCTGCGCGG-3'), o segundo adaptador dupla-fita parcial continha o primer A de sequenciamento (5'-TGTGTGCCTTGGCAGTCTCAGACGAGCGGCCA-3') e seu complemento parcial anti-A (5'-TGGCCGCTCGT-3'). A estrutura dupla-fita parcial nestes adaptadores garante que eles sejam ligados na orientação correta, além da DNA ligase T4 requerer DNA dupla-fita como molde. O mix de ligação foi preparado no volume final de 25 µL contendo: 1X de tampão NEB 2 (Invitrogen, USA), BSA 1X (Promega,

USA), 1 mM de ATP, 6,5 U de DNA ligase T4 (Fermentas, USA), 1 μ M de cada adaptador e 50 ng de DNA nebulizado. Esta reação foi incubada durante 16 h a 12°C. O produto ligado foi purificado em colunas *Qiagen Minielute* para um volume final de 30 μ L.

A solução de ligação foi testada por meio de reações de PCR para verificar se as corretas construções haviam sido produzidas. A amplificação foi realizada usando 0,1 μ M de cada um dos primers A e B (acima descritos), acrescido de 0,2 mM de dNTP, tampão Immolase 1X, 0,6 U de Immolase (New England Biolabs), 1 μ L do DNA ligado e água ultrapura para o volume final de 30 μ L. O programa de amplificação consistiu de uma desnaturação inicial a 94 °C por 60 s, seguida de 16 ciclos de 94 °C/40 s, 58 °C/60 s, 72 °C/90 s, e um ciclo de elongação final de 72°C/10 min.

Após a purificação do produto de PCR, de acordo com o protocolo 454/Roche GS-FLX, a distribuição do tamanho dos fragmentos e sua quantificação foram realizadas pelo *Agilent 2100 BioAnalyzer DNA 7500 LabChip* (Agilent, USA). Os passos seguintes foram o PCR em emulsão e o sequenciamento, seguindo as instruções do protocolo 454 Life Sciences GS-FLX (Roche, USA).

4.2.6. Análise das sequências

A análise do conjunto de sequências metagenômicas, bem como a montagem de *contigs*, foi realizada utilizando o Software *DeNovo Assembler* versão 2.0.01.14 do *Genome Sequencer FLX* (454 GS Flexi, Roche, EUA). Este software

triu as sequências de boa qualidade e de tamanho adequado para montagem dos *contigs* e análises subsequentes.

Para as análises de bioinformática, foram utilizados vários programas computacionais para anotação funcional e agrupamento dos *contigs* metagenômicos, como BLASTp e MG-RAST (<http://metagenomics.nmpdr.org/metagenomics.cgi?page=Logout>).

No BLASTp foram anotados os *contigs* ≥ 1000 bases, seguido de sua análise filogenética utilizando os programas CLUSTAL X (Thompson *et al.*, 1997) e MEGA (versão 4.0). Assim as sequências foram alinhadas no programa CLUSTAL X e a distância evolutiva foi calculada pelo MEGA usando o modelo de substituição de DNA reportado por Kimura (Kimura, 1980). A construção da árvore filogenética foi realizada pelo MEGA, utilizando o algoritmo *Neighbor-joining* (NJ), com valores de bootstrap calculados a partir de 1000 replicas. Entretanto, antes da anotação pelo BLASTp, os *contigs* foram submetidos ao PRODIGAL (<http://prodigal.ornl.gov/>) para predição de todos os possíveis genes microbianos presentes no *contig*.

Para a análise do conjunto de dados metagenômicos pelo MG-RAST, um arquivo de texto (.txt) foi gerado, contendo todas as sequências dos *contigs*. Este arquivo foi utilizado para entrada dos dados no programa, onde foram realizados *on line* toda a anotação das sequências, que foram agrupadas em categorias funcionais e pelo conteúdo GC.

4.3. RESULTADOS E DISCUSSÃO

4.3.1. Triagens funcionais

Os 10.000 clones provenientes da biblioteca MBR e 3.200 clones da REGAP foram amplificados para o gene que codifica a subunidade maior da enzima fenol hidroxilase (LmPHs), destes 21 clones apresentaram amplificação. Esta enzima, também chamada de fenol 2-monooxigenase, atua no início da via de degradação aeróbia do fenol, pela adição inicial de um grupo hidroxila no anel aromático.

Os 13.200 clones também foram triados para atividade de degradação de fenol em meio mineral. O ensaio utilizando MTT em microplacas de 96 poços foi realizado para detecção de clones que conseguiram crescer em fenol como única fonte de carbono. Foi observado o crescimento de 211 clones provenientes da biblioteca MBR e 202 clones da biblioteca REGAP (exemplos na Figura 1).

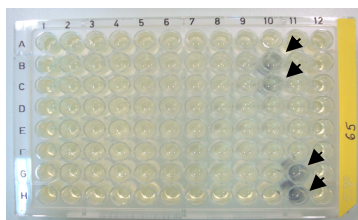


Figura 1. Ensaio MTT em microplacas de 96 poços. Os hits positivos, indicados pelas setas, apresentaram a coloração violeta.

Dois clones, que apresentaram resultados positivos em ambas as triagens, foram selecionados para os primeiros ensaios de cromatografia gasosa de degradação do fenol.

4.3.2. Cromatografia gasosa para degradação de fenol

Foi testada a capacidade de degradação do fenol dos clones por cromatografia gasosa. Após a análise dos resultados, observamos que foi possível a redução do volume da amostra para extração cromatográfica, de 300 mL (T1) para 50 mL (T2), pois o fenol foi detectado com sucesso no menor volume.

Outra questão que pretendíamos responder com estes experimentos era em relação ao tempo de incubação para a total degradação do fenol. Entretanto, os resultados encontrados em relação à degradação de fenol não foram conclusivos (dados não mostrados). Após 21 dias de incubação, as amostras dos tratamentos ainda apresentaram valores altos de concentração de fenol e valores semelhantes aos encontrados nos controles do meio acrescido de fenol. Esperando que houvesse a depleção do fenol, decidiu-se aumentar o tempo de coleta da última amostra para 40 dias. Entretanto, mesmo após este longo tempo de incubação, foram detectadas altas concentrações de fenol nos dois tratamentos.

Os resultados dos ensaios cromatográficos permitiram concluir, que o volume utilizado para extração cromatográfica pode ser reduzido para 50 mL. Porém, os resultados referentes ao tempo de incubação sugerem que a não degradação do fenol observada, provavelmente é decorrente da expressão heteróloga, isto é, expressão dirigida por enzimas transcricionais e traducionais da

célula hospedeira (*E. coli*) em um contexto onde os genes e as sequências regulatórias são originados de outros micro-organismos. A expressão heteróloga é uma das principais limitações dos ensaios funcionais das bibliotecas metagenômicas, pois pode produzir baixos níveis ou ausência de expressão da molécula alvo, ou ainda, seu transporte através da membrana celular pode ser afetada, comprometendo a detecção da molécula sintetizada (Uchiyama & Miyazaki, 2009).

Assim, baseado nestas conclusões, foi determinado para experimentos cromatográficos futuros, que o tempo de incubação será fixado em 28 dias, com cinco pontos de coleta: 0, 7, 14, 21 e 28 dias, onde a seleção dos melhores clones será realizada pela comparação das taxas de degradação obtidas para cada um. Ainda, o método cromatográfico será otimizado para se tornar mais sensível a pequenas variações na concentração do fenol. Ao final dos futuros experimentos cromatográficos, o melhor clone para degradação de fenol será utilizado para experimentos posteriores de evolução dirigida, que terá o objetivo de maximizar a degradação de fenol por meio de modificações genéticas no inserto.

4.3.4. Análise dos dados da corrida de pirosequenciamento

O pirosequenciamento dos 100 clones selecionados ao acaso dentre os que se mostraram positivos nas triagens anteriores, resultou em 129.635 sequências, aproximadamente 50 Mpb de DNA. A maioria destas sequências, 95%, formaram contigs, destes 57 de tamanho superior a 20 kb, sendo o maior deles de 41 kb, o que aumentou a probabilidade de encontrarmos operons completos. O número total de contigs foi de 609, destes 516 foram maiores que 500 pb, e destes

314 foram maiores que 1000 pb. Os dados metagenômicos apresentaram um resultado satisfatório, sendo obtida uma cobertura de 14X dos dados, considerando que os fragmentos de DNA metagenômico dos 100 fosmídeos somaram aproximadamente 3,6 Mpb. Os resultados desta corrida foram melhores que a obtida no primeiro pirosequenciamento (Capítulo 3).

4.3.5. Análise funcional dos dados de pirosequenciamento

Antes das análises funcionais, os 314 contigs maiores que 1000 pb foram analisados pela plataforma PRODIGAL (<http://compbio.ornl.gov/prodigal/>) na busca de todos os possíveis genes com base nos sítios de iniciação e terminação de um gene. Milhares de possíveis genes foram preditos, cada contig gerou entre 5 a 27 diferentes genes.

Para gerar o perfil metabólico desta comunidade microbiana, estas sequências foram submetidas à plataforma SEED/MG-RAST. O SEED classifica as rotas metabólicas em estruturas hierárquicas nas quais todos os genes requeridos para uma tarefa específica são organizados em um subsistema. Nos altos níveis de organização, os subsistemas incluem as funções de catabolismo e anabolismo, como por exemplo, metabolismo de DNA, proteínas, compostos aromáticos, e nos baixos níveis de subsistemas estão as vias específicas, como síntese de um aminoácido, metabolismo de determinado composto aromático.

Usando o BLASTx e o *cutoff* do E-value de 0,01, a plataforma SEED/MG-RAST (Meyer *et al.*, 2008) foi capaz de classificar 68,2% das sequências em categorias funcionais (Figura 2). Analisando os resultados dos dados

metagenômicos, observamos que os clones selecionados apresentam um perfil metabólico bastante diverso, relacionado com metabolismo de fósforo, enxofre, nitrogênio e de compostos aromáticos, sendo este último representado por 108 sequências. Estas se subdividiram em genes relacionados a rotas para o catabolismo de aromáticos (24,1%), como degradação dos compostos fenilpropanóide, bifenil e benzoato; degradação anaeróbia de benzoato (4,6%); metabolismo de intermediários centrais (64,8%), como intermediários na rota catabólica do ácido 4-fenilacético e na rota de *meta*-clivagem. Este perfil nos mostrou que nesta comunidade existem bactérias que realizam a degradação aeróbia dos fenóis por *meta*-clivagem, e que são capazes de degradar compostos como bifenil e benzoato.

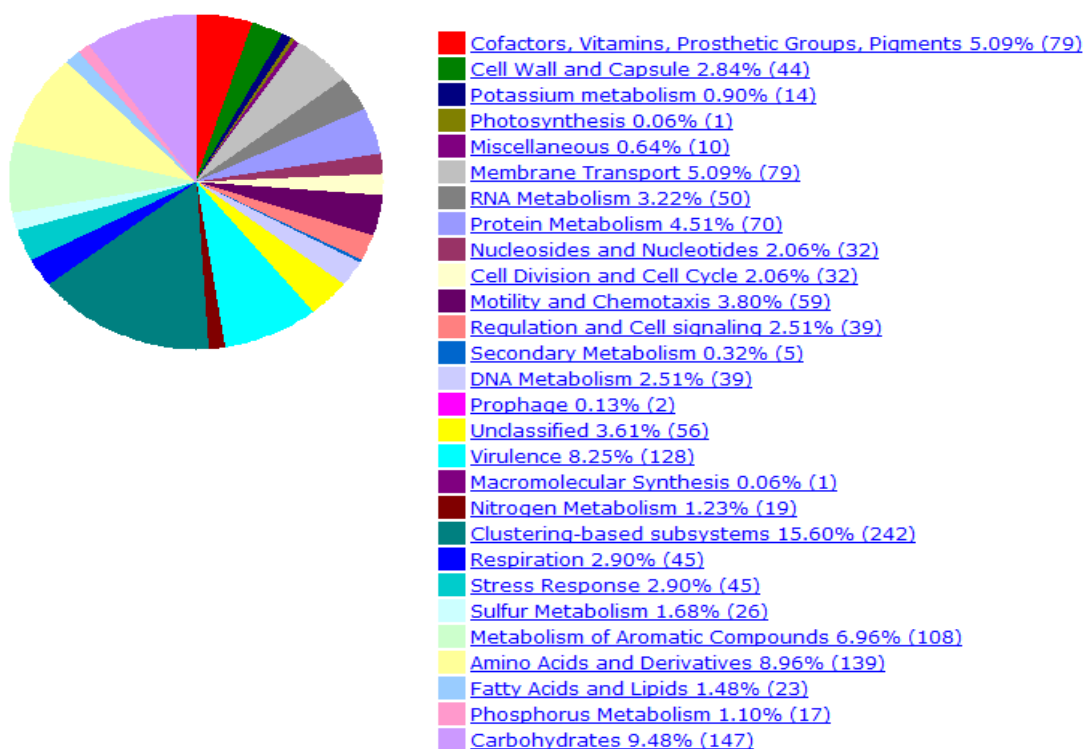


Figura 2. Categorias funcionais obtidas pela plataforma SEED. O número de sequências afiliadas à categoria está entre parênteses na legenda.

4.3.6. Análise de genes e rotas relacionadas com degradação fenol

Na tentativa de verificar a existência de novas sequências e rotas protéicas relacionados com a degradação de fenol, todos os genes preditos pelo PRODIGAL foram traduzidos em sequências de aminoácidos e analisados pelo BLASTp. Dos milhares de genes obtidos, 2.276 foram afiliados a proteínas conhecidas, com funções relacionadas com as categorias já descritas no item anterior.

Foram encontrados 19 proteínas diretamente relacionadas com a degradação de compostos fenólicos, subunidades da enzima multimérica fenol hidroxilase, subunidade 1 (4 hits), subunidade 3 (2 hits), catecol 2,3-dioxigenase (C23O) (6 hits), catecol 1,2-dioxigenase (C12O) (1 hit), metano/fenol/tolueno

hidroxilase (3 hits) e uma proteína MetA envolvida na via *meta*-clivagem de degradação aeróbia do fenol. A cobertura e a possível bactéria a qual estas enzimas são afiliadas estão listadas na Tabela 1. Observamos que muitas das bactérias as quais as enzimas foram relacionadas, estão envolvidas com degradação de compostos poluentes, como degradação de compostos aromáticos, como fenol, como *Acidovorax* (Manefield *et al.*, 2002) *Azoarcus* sp. (Breining *et al.*, 2000) e *Pseudomonas putida* (Gonzales *et al.*, 2001)

Para verificar a relação evolutiva das enzimas encontradas nos dados metagenômicos com proteínas depositadas em bancos de dados protéicos acurados, como *Swiss-Prot protein* e proteínas de referência (Refseq_prot), árvores filogenéticas foram construídas para cada uma das diferentes enzimas detectadas (Figura 3).

Tabela 1. Enzimas relacionadas com degradação de fenol presentes nos dados metagenômicos.

Contigs (Gene)	Enzimas	(Acesso GenBank) / Bactéria relacionada	Cobertura
06 20 (G17) 74 (G1)	Metano/fenol/ tolueno hidroxilase	(YP001669536) <i>Pseudomonas putida</i> GB-1 (YP583480) <i>Ralstonia metallidurans</i> CH34 (YP001669536) <i>Pseudomonas putida</i> GB-1	98% 96% 75%
579 (G2) 581 (G2)	Metano/fenol/ tolueno hidroxilase ou Fenol hidroxilase (sub. 3)	<i>Thauera</i> sp. MZ1T ou <i>Azoarcus</i> sp. BH72 <i>Thauera</i> sp. MZ1T ou <i>Azoarcus</i> sp. BH72	100% / 100% 100% / 100%
36 20 (G18) 74 (G2) 530 579 (G1) 581 (G1)	Fenol hidroxilase (sub. 1)	(P19730) <i>Pseudomonas</i> sp. CF600 (P19730) <i>Pseudomonas</i> sp. CF600 (YP001669538) <i>Pseudomonas putida</i> GB-1 (P19730) <i>Pseudomonas</i> sp. CF600 (YP933947) <i>Azoarcus</i> sp. BH72 (YP933947) <i>Azoarcus</i> sp. BH72	65% 99% 46% 72% 62% 56%
581 (G4) 579 (G4) 20 (G15) 79 80	Catecol 2,3- dioxigenase (C23O)	(Q04285) <i>Pseudomonas putida</i> (Q04285) <i>Pseudomonas putida</i> (P06622) <i>Pseudomonas putida</i> (Q04285) <i>Pseudomonas putida</i> (Q04285) <i>Pseudomonas putida</i>	98% 98% 99% 98% 98%
458	Catecol 1,2- dioxigenase (C12O)	(YP001669541) <i>Pseudomonas putida</i> GB-1	78%
87	MetA- pathway phenol degradation- like protein	(YP001669533) <i>Pseudomonas putida</i> GB-1	85%

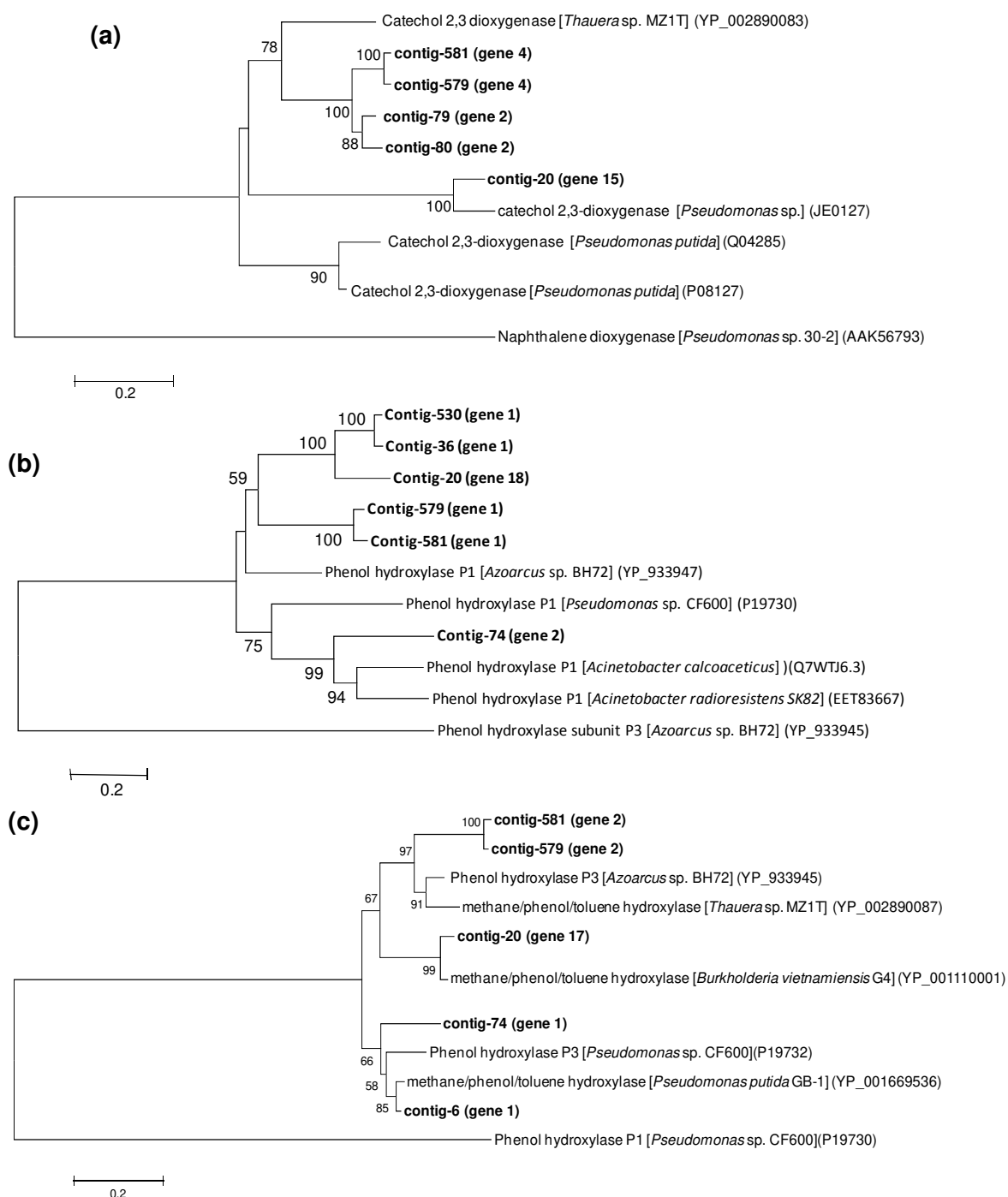


Figura 3. Árvores filogenéticas *Neighbour-joining* deduzidas de seqüências de aminoácidos das enzimas **(a)** catecol 2,3-dioxigenase, **(b)** fenol hidroxilase (sub. 1) e **(c)** fenol hidroxilase (sub. 3) e metano/fenol/tolueno hidroxilase das seqüências metagenômicas derivadas do pirosequenciamento. Todas as seqüências protéicas utilizadas foram obtidas de bancos de dados protéicos acurados, *SwissProt protein* e *Refseq_protein*. Os valores dos *bootstraps* maiores que 50% estão indicados nos nós.

Os resultados das árvores filogenéticas mostraram que algumas das sequências metagenômicas formaram agrupamentos ou *clusters* claramente distintos das sequências já conhecidas. Este resultado pode evidenciar a ocorrência de novas enzimas responsáveis pela degradação de fenol no metagênoma do lodo. Para a enzima C23O foi observada a formação de um cluster coeso de sequências semelhantes entre si e filogeneticamente distintas daquelas conhecidas, incluindo os contigs 79, 80, 579 e 581 (Figura 3a). Para a fenol hidroxilase sub. 1, também houve a formação de um cluster distinto, incluindo os contigs 530, 36, 20, 579 e 581, cuja sequência conhecida mais proximamente relacionada é a fenol hidroxilase sub. 1 de *Azoarcus* sp. BH72. Por outro lado, as sequências relacionadas às enzimas fenol hidroxilase sub. 3 e metano/fenol/tolueno hidroxilase não formaram um cluster distinto, e mostraram-se relacionadas a proteínas já conhecidas.

As sequências afiliadas às enzimas fenol hidroxilase sub. 3 e metano/fenol/tolueno hidroxilase foram analisadas no mesmo conjunto, pois a percentagem de identidade destas sequências à estas enzimas mostraram o mesmo valor. Um exemplo deste caso foi observado para os contigs 579 e 581 (gene 2), que apresentaram 100% de similaridade (*E-value* = 0,0) com a fenol hidroxilase sub. 3 de *Azoarcus* sp. BH72 e com metano/fenol/tolueno hidroxilase de *Thauera* sp. MZ1T. Portanto, achamos mais seguro trabalhar com os dados em conjunto para identificação das sequências na árvore filogenética. Esta ambiguidade ocorreu provavelmente porque estas proteínas pertencem à mesma família Phenol_Hydrox (PF02332), portanto, são proteínas relacionadas evolutivamente, que normalmente apresentam similaridade de sequência, e estrutura tridimensional e função similares. Esta família contém também as subunidades P1 e P2 da enzima multimérica fenol hidroxilase (mPHs), as cadeias alfa e beta da metano hidroxilase

(inicial oxigenação de metano para etanol e catalisa a monohidroxilação de vários alquenos e compostos, aromáticos, alicíclicos e heterocíclicos) e o sistema protéico tolueno-4-monooxigenase (hidroxila o tolueno para p-cresol). Apesar da fenol hidroxilase sub. 1 também pertencer à mesma família, ficou claro que existe uma maior similaridade das sequências com as enzimas fenol hidroxilase sub. 3 e metano/fenol/tolueno hidroxilase (Figura 3c).

O operon composto dos genes que codificam as enzimas mPHs foi detectado nos contigs 20, 579 e 581. No contig 20 foram encontrados 3 genes, que codificam para a C23O, e fenol hidroxilase sub. 1 e 3. Considerando que este contig está representado pela fita -1, os genes que codificam a enzima fenol hidroxilase sub. 1 (Gene 18) e sub. 3 (Gene 17) estão *upstream* ao gene que codifica para a enzima C23O (Gene 15). Esta organização gênica foi observada também para os contigs 579 e 581, com a presença do Gene 2, que codifica para a fenol hidroxilase sub. 3, seguido do Gene 4, que codifica para C23O. Entretanto, neste caso não foi observado o gene que codifica para a fenol hidroxilase sub. 1. Estes resultados são corroborados por dados encontrados na literatura para espécies de *Pseudomonas* e *Acinetobacter*, que apresentam seu operon da fenol hidroxilase com a mesma organização, ou seja, os genes que codificam para mPHs estão *upstream* aos genes que codificam para C23O e C12O (Shingler *et al.*, 1992; Ehrt *et al.*, 1995).

4.4. CONCLUSÕES

Os *screenings* por atividade biológica e por similaridade de sequência revelaram vários resultados positivos para a degradação do fenol, comprovando o sucesso da etapa de aclimação e enriquecimento com este composto anteriormente à construção da biblioteca.

O resultado do pirosequenciamento dos 100 clones positivos revelou a presença de novos genes relacionados com a degradação de fenol, como as sequências relacionadas às subunidades 1 e 3 de fenol hidroxilase, catecol 2,3-dioxigenase, além de genes relacionados com a degradação de outros compostos aromáticos poluentes, como bifenil, bifenilpropanóide e benzoato. Algumas rotas metabólicas relacionadas com a degradação de fenol e outros compostos aromáticos também foram observadas como a via *meta*-clivagem de degradação do fenol. A presença destes genes na comunidade indica que esta apresenta um grande potencial para degradação de compostos aromáticos presentes no sistema de tratamento de efluentes de refinaria de petróleo.

4.5. REFERÊNCIAS BIBLIOGRÁFICAS

- Barrios-Martinez, A., E. Barbot, B. Marrot, P. Moulin & N. Roche. 2006. Degradation of synthetic phenol-containing wastewaters by MBR. *Journal Membrane Science*. 281: 288-296.
- Basile, L. A. & L. Erijman. 2008. Quantitative assessment of phenol hydroxylase diversity in bioreactors using a functional gene analysis. *Applied Microbiology and Biotechnology*. 78: 863-872.
- Breinig, S., Schiltz, E. & Fuchs, G. 2000. Genes Involved in Anaerobic Metabolism of Phenol in the Bacterium *Thauera aromatica*. *Journal of Bacteriology*. 182: 5849-5863.
- Cordova-Rosa, S. M., Dams, R. I., Cordova-Rosa, E. V., Radetski, M. R., Corrêa, A.X.R. & Radetski, C. M. 2009. Remediation of phenol-contaminated soil by a bacterial consortium and *Acinetobacter calcoaceticus* isolated from an industrial wastewater treatment plant. *Journal of Hazardous Materials*. 164: 61-66.
- Dong, X., Hong, Q., He, L., Jiang, X. & Li, S. 2008. Characterization of phenol-degrading bacterial strains isolated from natural soil. *International Biodeterioration & Biodegradation*. 62: 257-262.
- Ehrt, S., F. Schirmer, & W. Hillen. 1995. Genetic organization, nucleotide sequence and regulation of expression of genes encoding phenol hydroxylase and catechol 1:2-dioxygenase in *Acinetobacter calcoaceticus* NCIB8250. *Molecular Microbiology*. 18: 13-20.
- Futamata, H., Harayama, S. & Watanabe, K. 2001. Diversity in kinetics of trichloroethylene-degrading activities exhibited by phenol-degrading bacteria. *Applied Microbiology and Biotechnology*. 55: 248-253.
- González, G. Herrera, G., García, M. T. & Peña, M. 2001. Biodegradation of phenolic industrial wastewater in a fluidized bed bioreactor with immobilized cells of *Pseudomonas putida*. *Bioresource Technology*. 80: 17-142.
- Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 16: 111-120.
- Manefield, M., Whiteley, A. S., Griffiths, G. I. & Bailey, M. J. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Applied Environment and Microbiology*. 68: 5367-5373.

- Merimaa, M. Heinaru, E., Liivak, M. Vedler, E. & Heinaru, A. 2006. Grouping of phenol hydroxylase and catechol 2,3-dioxygenase genes among phenol- and *p*-cresol-degrading *Pseudomonas* species and biotypes. *Archives Microbiology*. 186: 287-296.
- Nair, I. C., Jayachandran, K. & Shashidhar, S. 2007. Treatment of paper factory effluent using a phenol degrading *Alcaligenes* sp. under free and immobilized conditions. *Bioresource Technology*. 98: 714-716.
- Peters, M., Heinaru, E., Talpsep, E., Wand, H., Stottmeister, U., Heinaru A. Nurk A. 1997. Acquisition of a deliberately introduced phenol degradation operon, *pheBA*, by diVerent indigenous *Pseudomonas* species. *Applied and Environmental Microbiology*. 63: 4899-4906.
- Shingler, V., Powlowski, J. & Marklund, U. 1992. Nucleotide sequence and functional analysis of the complete phenol/3,4- dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *Journal of Bacteriology*. 174: 711-724.
- Sueoka, K., Satoh, H., Onuki, M. & T, Mino. 2009. Microorganisms involved in anaerobic phenol degradation in the treatment of synthetic coke-oven wastewater detected by RNA stable-isotope probing. *FEMS Microbiology Letter*. 291: 169-174.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, 423 D. G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 24: 4876-4882.
- Uchiyama, T. & Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*. 20: 616-622.
- Valle, A., Bailey, M. J., Whiteley, A. S. & Manefield, M. 2004. *N*-acyl- L -homoserine lactones (AHLs) affect microbial community composition and function in activated sludge. *Environmental Microbiology*. 6: 424-433.
- Vasconcellos, S. P, Crespim, E., Cruz, G. F., Senatore, D. B., Simioni, K. C. M., Santos Neto, E. V., Marsaioli, A. J. & Oliveira, V. M. 2009. Isolation, biodegradation ability and molecular detection of hydrocarbon degrading bacteria in petroleum samples from a Brazilian offshore basin. *Organic Geochemistry*. 40: 574-588.
- Yan, J., Jianping, W., Jing, B., Daoquan, W. & Zongding, H. 2006. Phenol biodegradation by the yeast *Candida tropicalis* in the presence of *m*-cresol. *Biochemical Engineering Journal*. 29: 227-234.

**CAPÍTULO 5. SCREENING DE ENZIMAS HIDROLÍTICAS EM BIBLIOTECAS
METAGENÔMICAS DE LODO BACTERIANO PROVENIENTE DE ESTAÇÕES
DE TRATAMENTO DE EFLUENTES DE REFINARIAS DE PETRÓLEO.**

Trabalho parcialmente apresentado: 3rd Congress of European Microbiologists - Gotemburgo, Suécia. ***Screening of lipases and esterases in a metagenomic library from petroleum refinery wastewater.*** 2009. Silva, Cynthia Canedo; Milagre, C. D. F.; Marsaioli, Anita J.; Torres, A. P. R.; Santiago, V. M. J.; Oliveira, Valéria M.

5.1. INTRODUÇÃO

Muitos compostos químicos encontrados em diferentes produtos comerciais têm sido lentamente substituídos por enzimas biológicas, que, por serem biodegradáveis, minimizam os impactos causados ao meio ambiente (Meilleur *et al.*, 2009). Atualmente, devido à alta demanda de mercado, a biotecnologia tem aumentado a busca por novos biocatalistas. Ferramentas de metagenômica, como extração direta do DNA de amostras ambientais, construção de bibliotecas e subsequente triagem para o composto de interesse, têm acelerado o processo de descobrimento de novos biocatalistas e metabólitos secundários com potencial biotecnológico (Elend *et al.*, 2006; Steele *et al.*, 2009). Estas técnicas têm auxiliado na revelação do grande potencial biocatalítico dos micro-organismos ainda não cultivados, até então inacessível por técnicas tradicionais de microbiologia (Amann *et al.*, 1995). Aliado a isto, há o fato de que alguns micro-organismos capazes de degradar a matéria orgânica pela produção de enzimas hidrolíticas não são capazes de produzir tais enzimas quando cultivados pelas técnicas convencionais (Burgess & Pletshke, 2008).

Assim, cada vez mais, bibliotecas metagenômicas têm sido triadas na busca de novos biocatalistas. Muitos dos compostos metagenômicos encontrados têm mostrado novos espectros de utilização de substrato e formação de produto, sendo muitas vezes reações altamente estáveis, ocorrendo sob condições extremas, resultando em grande potencial de aplicação nos processos industriais (Steele *et al.*,

2009). Dentre as enzimas de interesse biocatalítico estudadas, o grupo das enzimas hidrolíticas ou hidrolases, como por exemplo, lipases, esterases, proteases e epóxido hidrolase, é particularmente importante devido à ampla especificidade de substratos, alta estereo- e regioseletividade, independência de cofatores, estabilidade e atividade em solventes orgânicos (Bornscheuer & Kazlauskas, 1999).

As proteases estão dentre as enzimas hidrolíticas mais bem estudadas, devido à sua ampla aplicação na indústria de detergentes, de alimentos, farmacêutica, de diagnósticos e tratamento de resíduos, representando 40 % das enzimas vendidas no mercado (Gupta *et al.*, 2002). Em sistemas de tratamento de efluentes, elas têm sido testadas para tratamento do lodo excedente gerado, principalmente, pelas plantas de lodos ativados. Estas enzimas quebram macromoléculas de proteínas, facilitando, assim, a digestão do lodo (Chenel *et al.*, 2008; Li *et al.*, 2009). As lipases e esterases também representam um grupo extenso dentro das hidrolases, e grandes bibliotecas de esterases e lipases termofílicas têm sido originadas pelos inúmeros métodos de triagens enzimáticas (Demirjian *et al.*, 1999). Em tratamento de efluentes, a aplicação destas enzimas, pela técnica de bioaugmentação, tem sido testada no processo de pré-tratamento do efluente para hidrolisar e dissolver os ácidos graxos, resultando na melhora da degradação biológica e, conseqüentemente, numa redução de tempo para realização do processo de tratamento de efluentes (Lin *et al.*, 1997; Commarota & Freire, 2006). As epóxido hidrolases são enzimas amplamente encontradas na natureza e são de grande interesse industrial, principalmente para o setor farmacêutico pela constante busca de epóxidos intermediários na síntese de substâncias biologicamente ativas (Kotik *et al.*, 2007). Henne *et al.* (2000), através da triagem de uma biblioteca metagenômica de amostra de solo, encontraram uma

estireno-monoxigenase, enzima responsável pela epoxidação do estireno. Até o momento, não existem estudos de aplicação destas enzimas em sistemas de tratamentos de efluentes, mas como elas apresentam também um papel na degradação de compostos orgânicos (van Loo *et al.*, 2006), acredita-se que possam vir a auxiliar na degradação de compostos poluentes presentes em efluentes industriais.

O objetivo deste trabalho foi, através de ensaios de atividades enzimáticas em meio ágar, realizar o *screening* de enzimas hidrolíticas dos 13.200 clones das bibliotecas metagenômicas de DNA de alto peso molecular provenientes de lodo de biorreatores à membrana (MBR) utilizados em sistemas de tratamento de efluentes de refinarias.

5.2. MATERIAL E MÉTODOS

5.2.1. Amostras, construção e validação das bibliotecas metagenômicas

Foram construídas duas bibliotecas metagenômicas com fragmentos de DNA de 25-36 Kb. Uma amostra foi proveniente do biorreator MBR operado em escala piloto na COPPE (UFRJ), alimentado continuamente com alta concentração fenólica (média=75 mgL⁻¹), por 30 dias (Biblioteca MBR), e outra com lodo proveniente do biorreator MBR operado em escala piloto na Refinaria Gabriel Passos (REGAP – Betim, MG), após aclimação com 1,0 gL⁻¹ de fenol (Biblioteca REGAP).

A extração de DNA de alto peso molecular de ambas as amostras foi realizada como descrito por Silva *et al.* (2010a). Após seleção do fragmento de DNA de tamanho de interesse, a construção das bibliotecas metagenômicas foi realizada utilizando o Kit *CopyControl™ HTP Fosmid Library Production* (Epicentre, USA), como descrito anteriormente no Capítulo 3.

A validação das bibliotecas foram realizadas pela digestão de 5 clones de cada biblioteca, MBR e REGAP, selecionados ao acaso, e clivados com a enzima *NotI* (Promega, USA). Para isto, os clones foram inoculados em meio LB + cloranfenicol a 12,5 µg/mL e incubados a 37 °C por 16 horas, sob agitação de 200 rpm. O DNA fosmidial foi extraído como descrito no protocolo do Kit *FosmidMax DNA Purification* (Epicentre, USA) e clivado com a utilização de 10 U da enzima de restrição. A reação foi incubada a 37°C *overnight* e a 65°C por 10 min para inativação enzimática. Após digestão, o DNA fosmidial foi submetido à eletroforese em campo pulsado nas seguintes condições: 10,5 h, 120°, 6 Vcm⁻¹, 1 s - 12 s.

5.2.2. Screening para atividade de lipase e esterases

A produção de lipases e esterases foi testada através de ensaios em placas contendo meio LA + 12,5 µg/mL de cloranfenicol, acrescido de 3,1 % peso/volume de azeite de oliva purificado e 0,001 % peso/volume de rodamina B (Sigma-Aldrich - USA), também conhecido por rodamina 610, *basic violet* 10 ou C.I. 45170. Esta mistura foi homogeneizada, ainda quente, em liquidificador industrial, previamente esterilizado por luz UV durante 2 minutos (Kouker & Jaeger, 1987). Depois de vertido o meio, as placas foram deixadas em repouso por 24 horas em temperatura

ambiente. Após este período, os clones recém reativados em placas de 96 poços contendo 1 mL de LB + 12,5 µg/mL de cloranfenicol foram replicados nas mesmas condições e incubados por 48 h a 37°C. Os clones produtores de lipase formaram colônias de coloração rosa e emitiram fluorescência amarelo-alaranjado, e os não produtores de lipase, que acumularam rodamina B, formaram colônias rosa, mas não apresentaram fluorescência sob irradiação UV (366 nm).

5.2.3. Screening para atividade de epóxido hidrolase

Para identificar clones produtores de epóxido hidrolases, realizou-se um ensaio em placas de Petri baseado na seleção através de toxicidade de fosfomicina aos clones. A produção de epóxido hidrolase foi detectada em placas contendo meio LA acrescido de 12,5 µg/mL de cloranfenicol e 50 µg·mL⁻¹ do antibiótico fosfomicina (ácido L-cis-1,2-epoxipropilfosfônico). No Brasil, este composto é comercializado sob a forma de fosfomicina-trometamol (Monuril) pelo laboratório Zambom. Uma cepa resistente de *Pseudomonas aeruginosa* foi utilizada como controle positivo. Este ensaio foi recentemente implementado pelo grupo da Profa. Dra. Anita J. Marsaioli (IQ/UNICAMP) e está em fase de validação.

5.2.4. Screening para atividade de protease

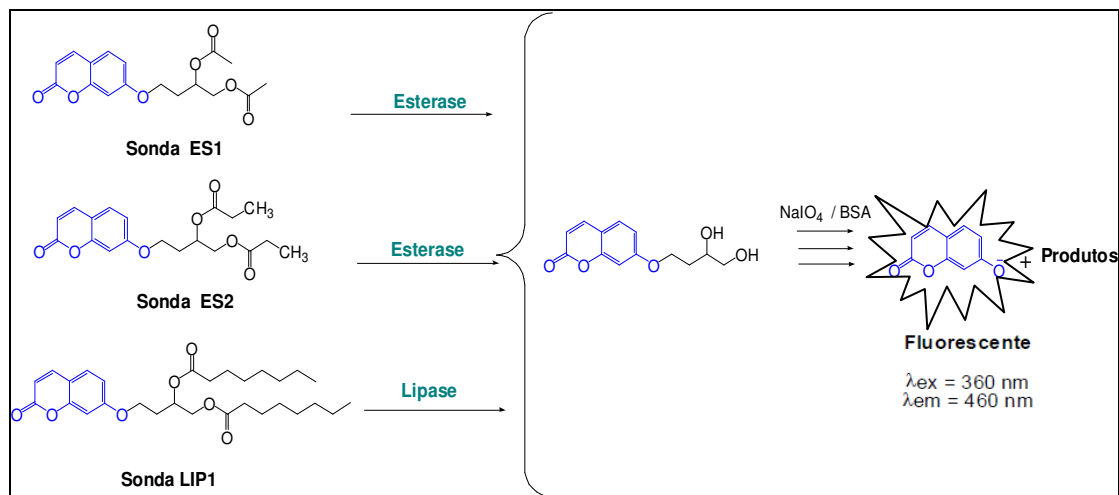
Os clones foram reativados em placas de 96 poços contendo 150 µL LB + 12,5 µg/mL de cloranfenicol, e incubados a 37°C por 16 horas. Com o auxílio de um replicador, os clones foram carimbados em placas contendo meio LA + 1% de Skim

Milk (Rossi *et al.*, 2007) + 12,5 µg/mL de cloranfenicol, e incubadas a 37°C por 4 dias. A detecção dos clones positivos foi realizada pela detecção visual da formação de um halo ao redor da colônia.

5.2.5. Ensaios HTS e de biocatálise convencional

Os clones considerados como hits positivos nos ensaios enzimáticos em placas foram submetidos às triagens de alta eficiência (HTS - *High Throughput Screening*) e, após confirmação da atividade enzimática, foram encaminhados aos ensaios de biocatálise convencional para legitimar os resultados obtidos nos ensaios de fluorescência.

O HTS em formato miniaturizado (microplacas de 96 poços) foi realizado utilizando sondas fluorogênicas distintas para a detecção das enzimas lipase e esterase. Nestes ensaios, a atividade enzimática é mensurada a partir da intensidade de fluorescência emitida a partir da liberação do ânion umbeliferona das moléculas dos substratos avaliados (sondas fluorescentes para determinação da atividade de lipases e esterases). Um modelo de ensaio enzimático utilizando substratos fluorogênicos foi desenvolvido por Klein e colaboradores (1998) e adaptada por Bicalho *et al.* (2003), para a avaliação de células microbianas íntegras. As sondas fluorogênicas investigadas foram sintetizadas e gentilmente cedidas pelo grupo de pesquisas em biocatálise liderado pela Profa. Dra. Anita J. Marsaioli (Figura 1).



Fonte: Suzan Pantarotto Vasconcellos

Figura 1. Sondas fluorogênicas utilizadas e reações para a detecção da atividade enzimática microbiana de esterases (ES1 e ES2) e lipases (LIP1).

As reações foram realizadas em microplacas e alíquotas de 10 μL das soluções dos substratos (2 mM) (H_2O – MeCN, 1:1) foram adicionadas a 80 μL de tampão borato (20 mM) pH 8,8, solução de BSA 2 mg/mL e 20 μL das suspensões de clones, previamente cultivados em LB + cloranfenicol 12,5 $\mu\text{g/mL}$. Por fim, foram adicionados 10 μL de uma solução aquosa de NaIO_4 a 20 mM.

O acompanhamento das reações de biotransformação foi realizado por meio de espectrofotômetro (*Flashscan 530 Analytic Jena*), utilizando filtro de λ_{em} 460 $\pm$ 20 nm para a detecção da emissão de fluorescência. Os experimentos foram monitorados por períodos de 48 horas, de acordo com procedimentos já consolidados no referido grupo de pesquisa.

Para os experimentos de biocatálise convencional, os clones foram cultivados em tubos Falcon (15 mL), contendo 5,0 mL de LB + cloranfenicol 12,5 $\mu\text{g/mL}$, durante 24 horas a 37°C, em agitador rotacional (150 rpm). A biomassa

gerada foi centrifugada a 8000 rpm e 4°C durante 15 minutos, sendo que o precipitado celular foi suspenso em 4 mL de tampão fosfato pH 7,0 (2,0 mM). Frascos *Erlenmeyers* (100 mL) contendo 20 mL da mesma solução tampão foram inoculados com 4,0 mL da suspensão celular e 2 µL de substrato para lipase ou esterase, seguindo a incubação em agitador rotacional a 37 °C, 150 rpm, durante 72 horas. O substrato 1, utilizado para detecção da atividade de lipase, foi um ester de cadeia longa, derivado do álcool isoprenílico com 18 carbonos, o estearato de 3-metilbuten-2-enila, e o substrato 2, utilizado para confirmação da atividade de esterase, foi um derivado de um composto esterificado de 8 carbonos do álcool metilbenzílico, denominado octanoato de feniletila (Figura 2).

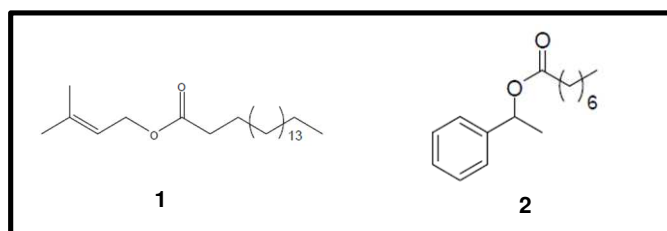


Figura 2. Substratos utilizados na biocatálise convencional: 1) estearato de 3-metilbuten-2-enila e 2) octanoato de feniletila.

A bioconversão dos substratos foi monitorada a cada 24 horas, através da extração de alíquotas (2,0 mL) das biorreações utilizando hexano como solvente orgânico. Os extratos orgânicos foram evaporados sob fluxo de N₂, e suspensos em solução de benzofenona (0,05 mg/mL). Os valores de conversão foram calculados com base na quantificação do substrato remanescente no meio reacional, quanto comparado ao padrão interno (Benzofenona) nas análises de CG-EM. Para as análises espectrométricas utilizou-se cromatógrafo a gás HP5890II acoplado a detector de massas HP 5970-MSD, injetor tipo split (10:1), He como gás de arraste,

fluxo de 5 mL / min, coluna capilar de sílica fundida do tipo HP5 (30 m x 0,25 mm x 0,25 μ m), cuja fase estacionária consiste de 5% de fenil metil silicone. O forno foi operado na seguinte programação: 60 °C por 3 min, seguindo o aquecimento de 60°C / min a 120°C, por 3 min. Após isso, a temperatura do forno foi operada em rampa de 80 °C / min a 200 °C, durante 7 min. O injetor foi utilizado à temperatura de 220 °C. Os dados foram adquiridos utilizando-se o modo SCAN e os valores de massa variaram entre 50 e 700 u.m.a (unidades de massa atômica).

O excesso enantiomérico dos produtos gerados foi obtido por CG-FID usando coluna capilar Chrompack® com fase quiral Chirasil- β -ciclodextrina.

5.3. RESULTADOS E DISCUSSÃO

Do total de 13.200 clones das bibliotecas metagenômicas avaliadas, três clones emitiram fluorescência sob luz UV no meio seletivo para lipase e esterase (Figura 3). O mecanismo molecular para formação de produtos fluorescentes gerados do ácido graxo e lipase ainda não foram completamente elucidados. Mackenzie *et al.* (1967) sugeriram que estes produtos fluorescentes resultam da formação de um complexo entre a rodamina B e o íon do ácido graxo livre proveniente da hidrólise pela lipase.

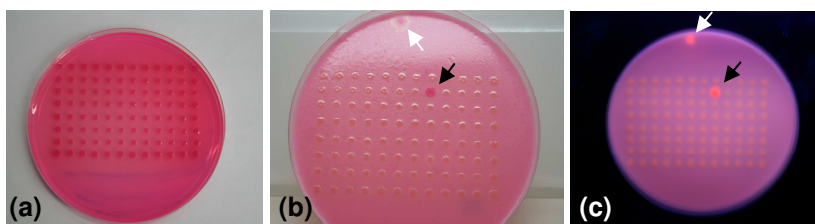


Figura 3. Ensaio para as enzimas lipase/esterase em meio sólido. **(a)** placa correspondente ao experimento controle; **(b)** placa mostrando um hit positivo para lipase/esterase (seta preta) e o controle, lipase comercial (seta branca), visualizados sob a luz branca; **(c)** placa

visualizando o hit positivo (seta preta) e o controle, lipase comercial (seta branca) sob luz UV (366 nm).

Ensaio HTS foram realizados para confirmação da atividade destes clones, bem como para definir se a atividade era de uma lipase ou esterase. Os resultados de intensidade de fluorescência mostraram que a atividade dos clones era resultado de uma lipase e duas esterases (Tabela 1). Foram verificados baixos valores de fluorescência para os três clones analisados. Os valores de conversão foram calculados com base na intensidade do controle positivo, uma lipase comercial (Novozin – Chiminova - Brasil), que foi estabelecido como 100% de fluorescência. Apesar dos baixos valores encontrados nos HTS, testes de biocatálise convencional foram aplicados para confirmação das atividades enzimáticas detectadas (Tabela 2). Estes testes mostraram que a atividade enzimática para estes clones realmente era baixa, mas confirmaram que se tratava de uma lipase e duas esterases. Todas as esterases apresentaram excesso enantiomérico, isto é, produziram os dois enantiômeros, levógiro e dextrógiro, e a lipase produziu um enantiômero. O desejável, dentre outras características, é que haja uma enantiosseletividade, ou seja, a produção de somente um dos enantiômeros, e de preferência aquele que apresente uma atividade de aplicação biotecnológica (Park *et al.*, 2006).

Tabela 1. Taxa de conversão do substrato nos ensaios de HTS

Identificação	% de conversão do substrato
MBR-97C4	5
MBR-40D6	3
MBR-27G5	5

Tabela 2. Taxa de conversão do substrato pela biocatálise convencional

Identificação	Substrato ^a	Convenção (ee) ^b	Atividade
MBR-97C4	2	32(32)	Esterase
MBR-27G5	2	46(4)	Esterase
MBR-40D6	1	20	Lipase

^aSubstrato 1 - detecta atividade para lipase

^aSubstrato 2 - detecta atividade para esterase

^bee- excesso enantiomérico

Quanto à detecção da produção de epóxido hidrolases, foram observados 30 hits positivos na triagem em meio contendo fosfomicina (Figura 4). Entretanto, quando estes foram submetidos aos ensaios de HTS, para comprovação da atividade da enzima, não houve emissão de fluorescência pelas sondas. Portanto, consideramos que não foi detectada atividade para esta enzima.

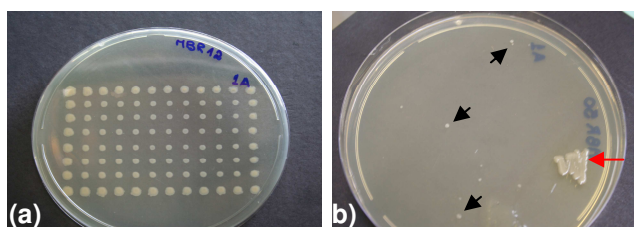


Figura 4. Ensaio para epóxido hidrolase em meio sólido. **(a)** Placa controle meio LA + Cloranfenicol; **(b)** placa LA + Cloranfenicol + Fosfomicina, as setas pretas evidenciam os hits positivos. Controle positivo - *Pseudomonas aeruginosa* (seta vermelha).

O ensaio para detecção da produção de enzimas proteolíticas não evidenciou a formação de halo hidrolítico ao redor de nenhuma colônia, entretanto observamos que 4 clones apresentaram um crescimento superior aos demais (Figura 5). Como não foi evidenciado a formação visual de um halo hidrolítico, a

inoculação dos mesmos em meio de triagem LA + Skim Milk foi repetida. Infelizmente, não houve repetibilidade do padrão de crescimento, e acreditá-se que o crescimento diferenciado observado na primeira triagem possa ter ocorrido devido a uma maior quantidade inicial de inóculo em relação aos demais clones. Portanto, concluímos também que não houve hits positivos para atividade proteolítica nestas bibliotecas.

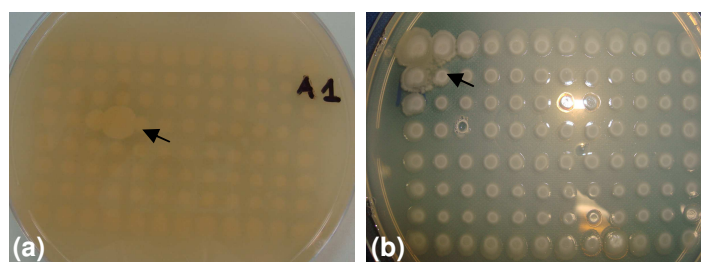


Figura 5. Ensaio em placa de LB + Skim Milk para detecção de protease. (a) e (b) setas indicando os clones que apresentaram crescimento superior aos demais no primeiro *screening*.

Uma das prováveis explicações para os baixos hits positivos obtidos nas triagens enzimáticas dos clones metagenômicos é a expressão heteróloga, condição em que a maquinaria do hospedeiro não reconhece as sinalizações de transcrição e tradução presentes no inserto exógeno e, conseqüentemente, falham na codificação da informação genética (Uchiyama & Miyazaki, 2009). Villegas & Kropinski (2008) reportaram que existe um códon preferencial de tradução em *E. coli*, AUG, mas que em bactérias de diversos genomas microbianos não existe preferência em um códon específico. Adicionalmente, Warren *et al.* (2008) demonstraram que o grau de expressão dos genes de organismos relacionados filogeneticamente com *E. coli* é maior que aquele de organismos filogeneticamente mais distantes. Isto evidencia a importância dos estudos de engenharia genética em

E. coli ou em outros potenciais hospedeiros para ampliação da capacidade de expressão dos diversos genes de interesse. Uma segunda explicação, é que apesar de tentarmos utilizar meios seletivos já descritos na literatura como bem sucedidos no *screening* de clones metagenômicos, a escolha dos mesmos para determinadas enzimas é bem complicada, pois muitas vezes só existem meios padronizados para seleção de isolados, podendo assim não apresentar constituintes necessários e adequados, como a presença de cofatores para produção de determinadas enzimas.

Apesar dos resultados das triagens não terem sido muito satisfatórios, existe a possibilidade de exploração futura destes clones positivos, por meio da evolução dirigida, na tentativa de aumentarmos a produção enzimática dos clones, principalmente sob condições extremas de temperatura, pH e salinidade.

5.4. CONCLUSÕES

De modo geral, as triagens para as enzimas hidrolíticas não geraram resultados satisfatórios. Alguns dos fatores que poderiam explicar estes resultados seriam o enriquecimento do lodo para degradadores de fenol antes da construção da biblioteca e a expressão heteróloga dos insertos exógenos no hospedeiro de *E. coli*. Entretanto, apesar da baixa atividade enzimática para lipase e esterase, estes clones terão a sequência gênica analisada para verificação da possibilidade de se tratar de um novo gene. Além disso, estes clones podem ser tomados como um ponto de partida para futuros experimentos de evolução dirigida, com objetivo de selecionar características de interesse para aplicação em processos industriais,

como aumento de atividade enzimática em condições extremas, como alta salinidade, pH, etc. E a principal conclusão obtida com este trabalho, é a necessidade de maiores pesquisas na área de engenharia genética para melhoramento dos sistemas hospedeiros, principalmente *E. coli*, para melhor reconhecimento dos sinais de transcrição e tradução presentes no DNA exógeno provenientes das mais diversas amostras ambientais.

5.5. REFERÊNCIAS BIBLIOGRÁFICAS

- Amann, R.I., Ludwig, W. & Schleifer, K. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiology Review*. 59:143-169.
- Bicalho, B., Chen, L. S., Grougnux, J., Reymond, J. L. & Marsaioli, A. J. 2004. Studies on whole cell fluorescence-based *screening* for epoxide hydrolases and Baeyer-Villiger monooxygenases. *Journal Brazilian Chemistry Society*. 15: 911-916.
- Bornscheuer, U. T. & Kazlauskas, R. J. 1999. **Hydrolases in Organic Synthesis: Regio- and Stereoselective Biotransformations**. Wiley-VCH Verlag, Weinheim.
- Burgess, J. E. & Pletshke, B. I. 2008. Hydrolitic enzymes in sewage sludge treatment: A mini review. *Water SA*. 34: 343-349.
- Cammarota, M. C. & Freire, D. M. G. 2006. A review on hydrolytic enzymes in the treatment of wastewater with high oil and grease content. *Bioresource Technology*. 97: 2195-2210.
- Chenel, J. P., Tyagi, R. D. & Surampalli, R. Y. 2008. Production of thermostable protease enzyme in wastewater sludge using thermophilic bacterial strains isolated from sludge. *Water Science & Technology*. 57: 639-645.
- Demirjian, D. C., Shah, P. & Morís-Varas, F. 1999. *Screening* for novel enzymes. *Topics in Current Chemistry*. 200:1-29.
- Elend, C., Leggewie, C., Babiak, P., Carballeira, J. D., Steele, H. L., Reymond, J-L, Jaeger, K-E. & Streit, W. R. 2006. Isolation and biochemical characterization of two novel metagenome derived esterases. *Applied and Environmental Microbiology*. 72: 3637-3645.
- Gupta, R., Beg, Q. K. & Lorenz, P. 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*. 59, 15-32.
- Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G. & Daniel, R. 2000. *Screening* of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Applied and Environmental Microbiology*. 66: 3113-3116.

- Klein, G. & Reymond, J. L. 1998. An enantioselective fluorimetric assay for alcohol dehydrogenases using albumin-catalyzed beta-elimination of umbelliferone. *Bioorganic & Medicinal Chemistry Letters*. 8:1113-1116.
- Kotik, M., Stepanek, V., Kyslik, P. & Maresova H. 2007. Cloning of an epoxide hydrolase-encoding gene from *Aspergillus niger* M200, overexpression in *E. coli*, and modification of activity and enantioselectivity of the enzyme by protein engineering. *Journal of Biotechnology*. 132: 8-15.
- Kouker, G. & Jaeger, K-E. 1987. Specific and sensitive plate assay for bacterial lipases. *Applied and Environmental Microbiology*. 53:211-213.
- Li, X., Mai, H. Wang, Q., Matsumoto, S., Maeda, T. & Ogawa, H.I. 2009. Isolation, identification of sludge-lysing strain and its utilization in thermophilic aerobic digestion for waste activated sludge. *Bioresource Technology* 100: 2475-2481.
- Lin, J. G., Chang, C. N. & Chang, S. C. 1997. Enhancement of anaerobic digestion of waste activated sludge by alkaline solubilization. *Bioresource Technology*. 62: 85-90.
- MacKenzie, R. D., Blohm, T. R., Auxier, E. M. & Luther, A. C. 1967. Rapid colorimetric micromethod for free fatty acids. *Journal of Lipid Research*. 8: 589-597.
- Meilleur, C., Hupe, J-F., Juteau, P. & Shareck, F. 2009. Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *Journal Industrial Microbiology and Biotechnology*. 36:853-861.
- Nabarlatz, D., Vondrysova, J., Jenicek, P., Stuber, F., Font, J., Fortuny, A., Fabregat, A. & Bengoa, C. 2008. Extraction of enzymes from activated sludge. *WIT Transactions on Ecology and the Environment* 109:249-257.
- Park, H. S., Nam, S-H., Lee, J.K., Yoon, C. N., Mannervik, B., Benkovic, S.J. & Kim, H-S. 2006. Design and evolution of new catalytic activity with an existing protein scaffold. *Science*. 311: 535-538.
- Rossi, B. P., Calenda, M., Vay, C. & Franco, M. 2007. Biofilm formation by *Stenotrophomonas maltophilia* isolates from device-associated nosocomial infections. *Revista Argentina de Microbiologia*. 39: 204-212.
- Steele, H. L., Jaeger, K-E., Daniel, R., Streit, W. R. 2009. Advances in Recovery of Novel Biocatalysts from Metagenomes. *Journal of Molecular Microbiology and Biotechnology*. 16: 25-37.
- Uchiyama, T. & Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*. 20: 616-622.

- van Loo, B., Kingma, J. Arand, M., Wubbolts, M.G., Janssen, D. B. 2006. *Applied and Environmental Microbiology*. 72: 2905-2917.
- Villegas, A. & Kropinski, A. M. 2008. An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. *Microbiology*. 154: 2559-2561.
- Warren, R. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S. & Holt, R. A. 2008. Transcription of foreign DNA in *Escherichia coli*. *Genome Research*. 18:1798-1805.

5. DISCUSSÃO GERAL

Dentre os problemas de poluição ambiental, a poluição da água tem sido uma das mais preocupantes, pois este é um recurso de grande importância em todos os setores, urbano, rural e industrial. Como as indústrias são grandes consumidoras de água, medidas legais foram implementadas para regulamentação do uso da água pelas indústrias (Lei 9433/97). A partir de então, as refinarias de petróleo brasileiras (Petrobras) por serem grandes consumidoras de água, começaram a adotar e criar procedimentos para reduzirem o consumo de água, o que resultou no aumento de seus investimentos em pesquisa de tratamentos de efluentes.

O tratamento biológico tem sido uma das estratégias mais empregadas para a remoção, principalmente, de concentração orgânica, como os compostos fenólicos, presentes em vários dos efluentes industriais (Barrios-Martinez *et al.*, 2006). Os fenóis são os principais poluentes encontrados no meio ambiente, e mesmo em baixas concentrações ele pode ser tóxico aos seres vivos. Além disso, sua presença pode inibir a atividade de biodegradação nos sistemas de tratamentos, uma vez que inibe o crescimento microbiano (Cordova-Rosa *et al.*, 2009).

Neste estudo, caracterizamos a diversidade microbiana presente em amostras de lodo provenientes de três sistemas de tratamento de efluentes: i) biorreator à membrana (MBR) piloto operado na COPPE (UFRJ-RJ), ii) MBR piloto

operando na refinaria REGAP (MG) e iii) sistema de lodo ativado convencional (CAS) na REVAP (SP). Ao se analisar os dados da comunidade microbiana destes 3 sistemas, percebemos que as proteobacterias constituíram o grupo predominante em todos eles. Segundo descrito no livro Brock (Madigan *et al.*, 2008), Proteobacteria é o filo mais amplo e metabolicamente mais diverso dentre todos do Domínio Bacteria. A diversidade metabólica é altamente necessária em ambientes de tratamento de efluentes industriais, um ambiente dinâmico que sofre variações quanto à constituição e concentração da carga orgânica a todo instante (Szczepanowski *et al.*, 2008). Resultados similares quanto à diversidade bacteriana também foram observados por vários estudos em sistemas de tratamento de efluentes (Ahmed *et al.*, 2007, Miura *et al.*, 2007, Lin *et al.*, 2008).

Apesar das proteobacterias terem sido as mais abundantes nas comunidades dos três sistemas, diferenças foram observadas em níveis hierárquicos inferiores. A classe Betaproteobacteria foi a mais abundante nos sistemas MBR da COPEE e da REGAP, entretanto no sistema de lodos ativados CAS da REVAP, as alfafroteobacterias foram predominantes. Provavelmente, os resultados de composição bacteriana das amostras de lodo refletiram as diferenças químicas e físicas encontradas nos diferentes sistemas de tratamento biológico utilizados, MBR e CAS. No sistema MBR há retenção total do lodo através de sua membrana, resultando na idade infinita do lodo, pois não ocorre purga do mesmo. Consequentemente, este sistema trabalha em uma alta concentração de micro-organismos, levando à uma baixa razão alimento/micro-organismos, ao passo que no sistema de lodos ativados, não há retenção total do lodo e periodicamente ocorre purga do sistema, o que torna a razão alimento/micro-organismos mais alta que no sistema anterior. Adicionalmente, a composição química dos efluentes, sob uma

ótica macro, é basicamente a mesma em termos de pH, concentração e constituição das concentrações orgânicas e inorgânicas. Entretanto sob uma ótica micro, estas diferenças existem, e mesmo estas pequenas variações na concentração da concentração de algum nutriente podem refletir na composição e função de comunidades microbianas, que estão em um constante equilíbrio dinâmico.

Os reflexos de tais diferenças físico-químicas entre os sistemas foram percebidas pronunciadamente nos níveis hierárquicos mais baixos, ou seja, a presença e abundância de determinados gêneros e espécies que caracterizam cada comunidade dos sistemas de tratamento. No sistema MBR da COPPE, espécies de *Thauera* e *Comamonas* foram as bactérias mais abundantes, resultado que foi relacionado à aplicação de alta concentração fenólica no sistema. É sabido que estas bactéria realizam a degradação de compostos aromáticos (Breining *et al.*, 2000, Arai *et al.*, 2000), e, portanto, os resultados obtidos sugerem que o aumento da carga orgânica de fenol promoveu o favorecimento destas populações em relação às demais. No sistema MBR da REGAP, houve a predominância de *Thiobacillus*, bactérias de metabolismo versátil que desnitrificam, oxidam compostos de enxofre reduzidos (Cho *et al.*, 1991) e utilizam compostos aromáticos policíclicos como fonte de carbono e energia (Bodour *et al.*, 2003). Já no sistema de tratamento de lodo ativado (CAS), houve a predominância de bactérias afiliadas à ordem *Parvularculales*. Estes clones formaram um agrupamento distinto na árvore filogenética, cuja espécie descrita mais proximamente relacionada foi *Parvularcula bermudensis*, única espécie descrita na ordem *Parvularculales*, sugerindo fortemente que tais clones sejam representantes de novas espécies ou até mesmo de novos gêneros de alfa-proteobactérias. Não se sabe nada sobre o metabolismo de bactérias da ordem *Parvularculales* em relação à degradação de compostos

poluentes, as únicas informações existentes são provenientes da bactéria marinha *Parvurlacula bermudensis*.

Um resultado que observamos para os sistemas MBR é que as bactérias mais abundantes, *Thauera*, *Comamonas* e *Thiobacillus*, apresentam uma característica em comum, são todas bactérias anaeróbias facultativas que podem desnitrificar quando necessário, utilizando o nitrato como aceptor de elétrons na respiração anóxica e produzindo compostos nitrogenados, como os óxidos nítrico, nitroso e nitrogênio gasoso, que são liberados na atmosfera. Tal característica é interessante para um sistema de tratamento de efluentes, onde se busca dentre outras funções, a remoção da concentração de compostos nitrogenados. Dentre as enzimas envolvidas neste processo estão a nitrato e a nitrito redutase, hits abundantes nos dados metagenômicos derivados do pirosequenciamento. Assim, estas bactérias mostram um metabolismo mais versátil, adequado ao sucesso de biorreatores MBR, onde há uma baixa relação alimento/micro-organismo, favorecendo as interações competitivas por todos os tipos de nutrientes e nichos (Wever *et al.*, 2007).

Uma característica comum às comunidades estudadas, além da alta diversidade microbiana, foi a presença de um considerável número de sequências de RNAr 16S que não puderam ser afiliadas a nenhuma espécie descrita, o que nos permite inferir que estamos trabalhando com comunidades complexas, e que ainda há muito para se descobrir sobre a diversidade bacteriana de sistemas de tratamento de efluentes.

A grande diversidade taxonômica e metabólica inferida pelos bancos de RNAr 16S foi confirmada pela análise das bibliotecas metagenômicas através do pirosequenciamento. A análise filogenética dos dados metagenômicos revelou uma

enorme diversidade de proteobacterias, compondo mais de 80% das sequências de genes RNAr 16S encontradas. A Classe Betaproteobacteria foi a mais abundante, corroborando os dados gerados pelo sequenciamento de Sanger dos lodos do sistema MBR, COPPE e REGAP. As espécies bacterianas que se destacaram na análise metagenômica, *Thauera*, *Comamonas*, *Diaphorobacter*, *Rhizobium*, *Pseudomonas* e *Planctomyces*, também foram as mesmas ou filogeneticamente relacionadas às que se destacaram nos estudos de biblioteca de genes RNAr 16S. Uma exceção, foi o gênero *Pseudomonas*, organismos de pronunciada abundâncias nos dados metagenômicos e que não foram representados nos bancos de genes RNAr 16S. Esta diferença observada entre as estratégias de análise de diversidade utilizadas poderia ser explicada pela amplificação preferencial decorrente do uso de primers considerados genéricos para a construção de bibliotecas de genes RNAr 16S, mas que invariavelmente apresentam uma afinidade em maior ou menor escala por determinados grupos taxonômicos em detrimento de outros. Neste caso, o par de primers utilizado não está sendo capaz de amplificar com sucesso a região dos genes RNAr 16S de espécies de *Pseudomonas* quando misturados no DNA total de uma comunidade bacteriana. Por outro lado, experimentos realizados em nosso laboratório utilizando DNA genômico obtido de cultura pura de espécies de *Pseudomonas* revelaram que este par de primers amplifica normalmente o gene RNAr 16S destas espécies (dados não mostrados).

A riqueza de espécies bacterianas observada refletiu em uma alta diversidade metabólica, caracterizada também por um perfil de população metabolicamente ativa, demonstrado pelos altos hits de genes relacionados a atividades como replicação, transcrição, tradução e síntese protéica. Adicionalmente, esta população se mostrou bastante versátil para degradação de

compostos poluentes aromáticos, uma vez que foram detectados genes e rotas catabólicas relacionadas a atividades de degradação de bifenil, benzoato, fenol, catecol, cresol, dentre outros. Alguns destes genes se mostraram distantes filogeneticamente das sequência de genes já descritos, resultado evidenciado nas análises dos dados metagenômicos derivados do segundo pirosequenciamento. Estes dados incluíram somente as sequências pertencentes a clones que cresceram em meio mineral contendo fenol como única fonte de carbono. As análises filogenéticas destas sequências revelaram a formação de clusters coesos, incluindo genes que codificam para as subunidades 1 e 3 da enzima fenol hidroxilase, e para catecol 2,3-dioxigenase, e que se mostraram claramente distintos das sequências já descritas.

O grande número de transposases observado sugere a existência de um grande número de transposons na comunidade presente nos lodos. As transposases são enzimas relacionadas com a translocação dos transposons. Estes elementos móveis são moléculas de DNA que se encontram inseridas no genoma dos micro-organismos e têm um importante papel na variação genética, pois podem se movimentar dentro do genoma e entre os genomas, estando relacionados com a transferência horizontal de genes entre as espécies (Madigan *et al.*, 2008). Portanto, eles representam um dos fatores responsáveis pela rápida adaptação das bactérias a determinadas alterações no ambiente, pois além de genes relacionados à translocação, carregam genes que podem estar relacionados com as mais diversas características, como resistência a determinado antibiótico.

As triagens funcionais das bibliotecas metagenômicas obtidas neste estudo se mostraram promissoras quando relacionadas às atividades de degradação do fenol, resultado que não se repetiu ao realizarmos triagens para enzimas

hidrolíticas. O alto número de hits encontrados nas triagens para o fenol evidencia que a estratégia de aclimatação e enriquecimento para degradadores de fenol antes da construção da biblioteca metagenômica foi bem sucedida. O enriquecimento de amostras ambientais utilizando um composto específico como substrato, favorece o aumento da representação de genomas na biblioteca relacionados à utilização de tal composto (Entcheva *et al.*, 2001). Portanto, ao se enriquecer para degradadores de fenol, pode ter havido uma contra-seleção de populações relacionadas com a produção das enzimas hidrolíticas investigadas. Outro problema, inerente às bibliotecas metagenômicas, é a expressão heteróloga, onde o hospedeiro não reconhece os sinais de transcrição ou tradução presentes no DNA exógeno e falham na codificação do material genético (Uchiyama & Miyazaki *et al.*, 2009).

Apesar das triagens funcionais não terem evidenciado um metagenoma metabolicamente diversificado, as análises de sequência mostraram o contrário, pelas inúmeras funções gênicas detectadas ou preditas, sugerindo que estas bibliotecas metagenômicas possam ser futuramente exploradas para diversas outras funções de interesse. Além disso, os resultados abrem a perspectiva de exploração de amostras de lodo provenientes de tratamento de efluentes de refinarias de petróleo para descoberta de novas espécies de bactérias, como demonstrado pelas milhares de sequências de genes RNAr 16S não afiliadas a nenhuma das espécies já descritas nas bases de dados pertinentes.

5.1. REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmed, Z., Cho, J., Lim, B-R., Song, K-G., Ahn, K-H. 2007. Effects of sludge retention time on membrane fouling and microbial community structure in a membrane bioreactor. *Journal of Membrane Science*. 287: 211-218.
- Arai, H., Akahira, S., Ohishi, T., Maeda, M. & Kudo, T. 1998. Adaptation of *Comamonas testosteroni* TA441 to utilize phenol: organization and regulation of the genes involved in phenol degradation. *Microbiology*. 144: 2895-2903.
- Barrios-Martinez, A., E. Barbot, B. Marrot, P. Moulin & N. Roche. 2006. Degradation of synthetic phenol-containing wastewaters by MBR. *Journal Membrane Science*. 281: 288-296.
- Bodour, A. A.; Wang, J-M.; Brusseau, M. L. & Maier, R. M. 2003. Temporal change in culturable phenanthrene degraders in response to long-term exposure to phenanthrene in a soil column system. *Environment Microbiology*. 5:888-895.
- Breinig, S., Schiltz, E. & Fuchs, G. 2000. Genes Involved in Anaerobic Metabolism of Phenol in the Bacterium *Thauera aromatica*. *Journal of Bacteriology*. 5849-5863.
- Cho, K.-S., Hirai, M. & Shoda, M. 1991. Degradation characteristics of hydrogen sulfide, ethanethiol, dimethyl sulfide and dimethyl disulfide by *Thiobacillus thioparus* DW 44 isolated from peat biofilter. *Journal Fermentation and Bioengineering*. 71: 384-389.
- Cordova-Rosa, S. M., Dams, R. I., Cordova-Rosa, E. V., Radetski, M. R., Corrêa, A.X.R., Radetski, C. M. 2009. Remediation of phenol-contaminated soil by a bacterial consortium and *Acinetobacter calcoaceticus* isolated from an industrial wastewater treatment plant. *Journal of Hazardous Materials*. 164:61-66.
- de Wever, H., S. Weiss, T. Reemtsm, V. Vereecken, J. Muller, T. Knepper, *et al.* 2007. Comparison of sulfonated and other micropollutants removal in membrane bioreactor and conventional wastewater treatment. *Water Research*. 41: 935-945.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W. R. 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology*. 67: 89-99.
- Lin, S., Jin, Y., Fu, L., Quan, C., Yang & Y. S. 2009. Microbial community variation and functions to excess sludge reduction in a novel gravel contact oxidation reactor. *Journal of Hazardous Materials*. 165: 1083-1090.

- Madigan M., Martinko, J., Dunlap, P. & Clark, D. 2008. **Brock Biology of Microorganisms**. 12 Ed. ISBN: 0-13-232460-1.
- Miura, Y., Hiraiwa, M. N., Ito, T., Itonaga, T., Watanabe, Y. & Okabe, S. 2007. Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance. *Water Research*. 41: 627 - 637.
- Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., Eichler, W., Pühler, A. & Schlüter, A. 2008. Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *Journal of Biotechnology*. 136: 54-64.
- Uchiyama, T. & Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*. 20:616-622.

6. CONCLUSÕES GERAIS

I) As análises de diversidade revelaram que as amostras de lodo bacteriano de sistemas de tratamento de efluentes de refinaria apresentam uma complexa comunidade bacteriana, sugerindo que estas comunidades apresentam uma ampla versatilidade metabólica e, conseqüentemente, capacidade de responder rapidamente a mudanças nos sistemas;

II) O pirosequenciamento dos clones derivados das bibliotecas metagenômicas revelou que estes lodos possuem uma alta diversidade metabólica, representada pelos inúmeros subsistemas mostrados pela plataforma MG-RAST, evidenciando o grande potencial da comunidade para degradação de compostos poluentes presentes em sistemas de tratamento biológico de efluentes de refinarias de petróleo;

III) As triagens funcionais mostraram a possibilidade de encontrarmos clones com alta taxa de degradação de fenol, visto que mais de 400 clones conseguiram sobreviver em fenol como única fonte de carbono;

IV) Os resultados combinados nos sugerem a possibilidade de existência de novas espécies de bactérias e funções metabólicas nas amostras de lodo analisadas, demonstrada pela quantidade de sequências com baixa ou nenhuma similaridade com sequência conhecidas depositadas nos bancos de dados do NCBI;

V) Os dados de sequencia gerados pelo pirosequenciamento nos permitirão explorar e melhor conhecer a ecologia das comunidades microbianas de vários ambientes, além daquelas provenientes de sistemas de tratamento de efluentes, pelo emprego de metodologias dependentes de sequências, como FISH, PCR, Microarranjos, etc;

VI) A bioprospecção genômica microbiana é uma estratégia inovadora e de grande potencial para área de tratamento de efluentes industriais, podendo contribuir para o estabelecimento de novas tecnologias que permitam o reuso da água.

7. PERSPECTIVAS FUTURAS

- As sequências gênicas de interesse, encontradas no pirosequenciamento, serão utilizadas para o desenho de *primers* e sondas (PCR e FISH) a serem empregados na caracterização de populações e funções microbianas presente nos lodos de tratamentos de efluentes da indústria do petróleo;
- Ajustes na metodologia de cromatografia gasosa para detecção de baixas concentrações de fenol no meio de cultivo;
- Os clones que cresceram em fenol como única fonte de carbono serão avaliados, quanto a degradação de fenol, por cromatografia gasosa;
- Os clones selecionados nas triagens de lipase/esterase e degradação de fenol, terão os genes caracterizados e serão utilizados em experimentos de evolução dirigida, com objetivo de selecionar características de interesse para aplicação em processos industriais, como aumento de atividade enzimática em condições extremas, como alta salinidade, pH e aumento da taxa de degradação de fenol.