

Geizecler Tomazetto

**“Diversidade e prospecção de metagenoma
microbiano em fermentadores de biogás
produzindo H₂”**

**CAMPINAS
2013**



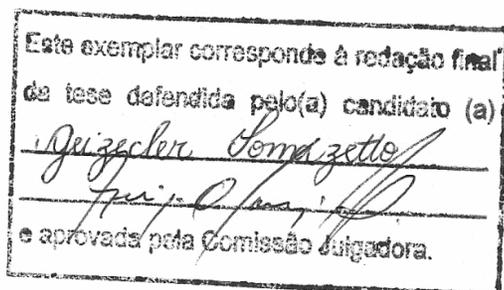
UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

GEIZECLER TOMAZETTO

“DIVERSIDADE E PROSPECÇÃO DE METAGENOMA MICROBIANO EM FERMENTADORES DE BIOGÁS PRODUZINDO H₂”

Tese apresentada ao Instituto de Biologia da UNICAMP para obtenção do Título de Doutora em Genética e Biologia Molecular, na área de Genética de Microorganismos.



Orientadora: Dra. Valeria Maia Merzel

CAMPINAS,
2013

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

T591d Tomazetto, Geizecler, 1979-
Diversidade e prospecção de metagenoma microbiano em fermentadores de biogás produzindo hidrogênio / Geizecler Tomazetto. – Campinas, SP : [s.n.], 2013.

Orientador: Valeria Maia Merzel.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Metagenoma. 2. Hidrogênio. 3. Lodo de esgoto. 4. Análise filogenética. 5. Bioinformática. I. Oliveira, Valeria Maia de. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em inglês: Diversity analysis and bioprospection of microbial metagenome in a H₂-producing biogas fermenter

Palavras-chave em inglês:

Metagenome

Hidrogen

Sewage sludge

Phylogenetic analysis

Bioinformatic

Área de concentração: Genética de Microorganismos

Titulação: Doutora em Genética e Biologia Molecular

Banca examinadora:

Valeria Maia Merzel [Orientador]

Suzete Aparecida Lanza Destefano

Fernando Dini Andreote

Suzan Pantaroto de Vasconcellos

Anete Pereira de Souza

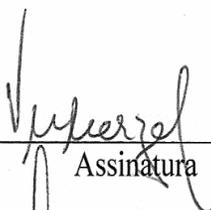
Data de defesa: 30-04-2013

Programa de Pós-Graduação: Genética e Biologia Molecular

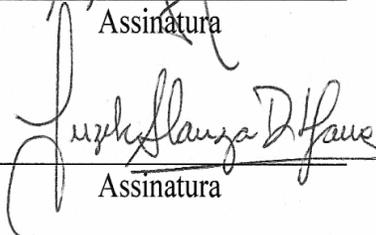
Campinas, 30 de abril de 2013

BANCA EXAMINADORA

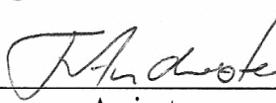
Dra. Valeria Maia Merzel (orientadora)


Assinatura

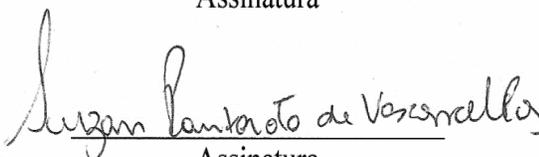
Dra. Suzete Aparecida Lanza Destefano


Assinatura

Prof. Dr. Fernando Dini Andreote


Assinatura

Profa. Dra. Suzan Pantaroto De Vasconcellos


Assinatura

Profa. Dra. Anete Pereira De Souza


Assinatura

Dra. Laura Maria Mariscal Ottoboni

Assinatura

Prof. Dr. Gonçalo Amarante Guimarães Pereira

Assinatura

Profa. Dra. Maria Bernadete Amâncio Varesche Silva

Assinatura

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha tese de Doutorado intitulada "Diversidade e prospecção de metagenoma microbiano em fermentadores de biogás produzindo hidrogênio":

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

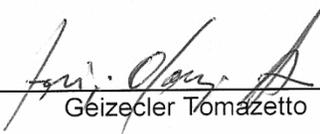
Tem autorização da(s) seguinte(s) Comissão(ões):

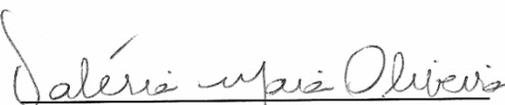
(x) CIBio – Comissão Interna de Biossegurança , projeto No. 22/2009, Instituição: Universidade Estadual de Campinas.

() CEUA – Comissão de Ética no Uso de Animais , projeto No. _____, Instituição: _____

() CEP - Comissão de Ética em Pesquisa, protocolo No. _____, Instituição: _____

** Caso a Comissão seja externa ao IB/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.*


Géizecler Tomázetto


Valéria M de Oliveira

Para uso da Comissão ou Comitê pertinente:

Deferido () Indeferido



Carimbo e assinatura

Prof. Dr. MARCELO LANCELLOTTI
Presidente da Comissão Interna de Biossegurança
Instituto de Biologia - UNICAMP

Para uso da Comissão ou Comitê pertinente:

() Deferido () Indeferido

Carimbo e assinatura



Título: FCIBio 01 – Formulário de proposta de projeto envolvendo OGM	Emissão: CIBio	Edição/Revisão: 2/0	nº p. 4
--	----------------	---------------------	---------

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

(a ser preenchido pela CIBio sob o número CPQBA/CQB 0189/03)

Número do projeto: <u>22 / 2009</u>
Data de entrada: <u>04 / 03 / 09</u>
Analisado por <u>MARTA C.I. DUARTE</u> em <u>16 / 10 / 09</u>
“Check list” no laboratório feito em: <u>16 / 10 / 09</u>
Campinas, Situação: <input checked="" type="checkbox"/> Aprovado <input type="checkbox"/> Reprovado Assinatura do Presidente da CIBio
OBS: FOI ENVIADO NOVO PEDIDO PARA ACESSAR CTNBIO 1006/09

Projeto de Pesquisa

Título: Diversidade e prospecção de metagenoma microbiano em fermentadores de biogás produzindo H₂	
Data de início do projeto (mês/ano): 03/2009	Data prevista de conclusão (mês/ano): 02/2012

Pesquisador Principal

Nome: Valéria Maia Merzel		
Endereço: DRM, CPQBA/UNICAMP, R: Alexandre Cazellato, 999 Vila Betel, Paulínia SP		
e-mail: vmaia@cpqba.unicamp.br	Telefone: (19) 2139-2874	Fax: (19) 2139-2852

Objetivo do Projeto

Este trabalho tem como objetivos a construção de uma biblioteca metagenômica a partir de um único modelo de fermentador de biogás, triagem de clones na busca de novos genes de hidrogenase e a caracterização da diversidade da comunidade microbiana produtora de hidrogênio.

AGRADECIMENTOS

À minha orientadora, Dra. Valéria Maia Merzel, pelo apoio, paciência, ensinamentos e confiança durante todo esse período.

À Universidade Estadual de Campinas (UNICAMP), ao Departamento de Genética e Biologia Molecular (IB) e ao Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas (CPQBA), pela oportunidade e execução deste trabalho.

Aos professores da pré-banca, Suzan Pantaroto de Vasconcellos e Rodrigo Taketani, pela disponibilidade e sugestões para a finalização do trabalho.

Aos membros da banca Suzete Aparecida Lanza Destéfano, Fernando Dini Andreote, Suzan Pantaroto de Vasconcellos, Anete Pereira de Souza, pelas correções e sugestões realizadas nesta Tese.

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

À Sociedade de Abastecimento de Água e Saneamento S/A (SANASA) pelo fornecimento e auxílio na coleta das amostras.

Ao *Institute for Genome Research and Systems Biology, CeBiTec, Bielefeld University*, em especial ao Dr. Andreas Schlüter pela disponibilização do laboratório, paciência e colaboração para o enriquecimento do trabalho.

Aos meus colegas da DRM (CPQBA) pelos favores e agradáveis momentos de convivência.

A toda a minha família pelo amor e compreensão.

"Se non é vero, é bene trovato."

RESUMO

O hidrogênio é apontado como o candidato mais promissor para substituição do combustível fóssil devido a sua maior eficiência na conversão de energia útil e ausência de emissão de substâncias tóxicas. A produção de hidrogênio a partir de resíduos orgânicos é realizada por meio de digestão anaeróbica, tornando-se uma alternativa ecologicamente correta para atender à futura demanda por hidrogênio. No entanto, os micro-organismos e os processos metabólicos envolvidos estão longe de serem exaustivamente caracterizados. Nesse trabalho, amostras de uma planta de tratamento de esgoto doméstico foram analisadas em dois estudos complementares visando à caracterização de sua diversidade filogenética e a descrição de novas hidrogenases. O primeiro trabalho combinou a análise dos genes de RNAr 16S e FeFe-hidrogenase (*hydA*) com ferramentas estatísticas para estimar a riqueza e diversidade da comunidade procariótica em nível filogenético e funcional. As análises filogenéticas e de diversidade das bibliotecas gênicas demonstraram que todas as sequências de arqueias foram afiliadas a Euryarchaeota não cultivadas e, com relação ao Domínio Bacteria, Proteobacteria foi grupo filogenético predominante apresentando os maiores índices de diversidade e riqueza. As sequências putativas de *hydA* foram identificadas como sequências de genes de FeFe-hidrogenases ainda não descritas. Na segunda abordagem, a biblioteca metagenômica de fosmideo construída nesse estudo foi analisada empregando a tecnologia de pirosequenciamento 454 e resultou em aproximadamente 218 Mb de dados. Os três diferentes classificadores aplicados permitiram uma visão geral dos grupos taxonômicos mais abundantes devido ao enorme número de sequências metagenômicas não classificadas. Contudo, análises taxonômicas revelaram Gammaproteobacteria e Deltaproteobacteria, respectivamente, como as classes taxonômicas predominantes, enquanto que as espécies do gênero *Methanospirillum* foram dominantes entre as arqueias metanogênicas. A análise do metabolismo da comunidade microbiana através das bases de dados COG e Carma revelou que a degradação da biomassa depende de diferentes grupos filogenéticos, como por exemplo, Bacteroidia e Gammaproteobacteria, os quais foram indicados como envolvidos na degradação de carboidratos e proteínas, respectivamente. Além disso, as análises sugerem Clostridia e Methanomicrobiales e Methanosarcinales como principais micro-organismos produtores de hidrogênio e metano, respectivamente. As análises das seis sequências codificantes de FeFe-hidrogenase identificadas no conjunto de dados metagenômicos revelaram que essas representam novas sequências do gene alvo. Contudo, quatro dessas sequências foram

identificadas na biblioteca de fosmídeo pela triagem gênica baseada no uso de PCR. O conjunto de resultados obtido nesse estudo permitiu elucidar a composição e o potencial metabólico dos micro-organismos residentes na planta de tratamento de esgoto analisada e sugere esse ambiente como um reservatório potencial de novos genes de hidrogenases para a exploração biotecnológica.

ABSTRACT

Hydrogen appears to be the most promising candidate for the replacement of fossil fuel due to its potentially higher efficiency of conversion to usable power and no toxic emission production. The production of hydrogen from organic wastes is performed through the anaerobic digestion, making it an environmentally friendly alternative for satisfying future hydrogen demands. Nonetheless, the microorganisms and metabolic processes involved are far from being exhaustively characterized. In this work, samples of a domestic sewage treatment plant were analyzed in two complementary studies aiming at the characterization of its phylogenetic diversity and the description of new hydrogenases. The first one, combined the analysis of 16S rRNA and [FeFe]-hydrogenase (*hydA*) genes with statistical tools to estimate richness and diversity of the prokaryotic community at the phylogenetic and functional levels. Phylogenetic analysis showed that all archaeal sequences were affiliated with yet uncultured Euryarchaeota and that Proteobacteria were the most predominant and diversified phylogenetic group within the bacterial library. The putative *hydA* sequences were identified as hitherto undetected [Fe-Fe]-hydrogenase gene sequences. Diversity statistical analysis confirmed a great richness and diversity of bacterial and *hydA* sequences retrieved from the sewage sludge sample. In the second approach, a fosmid metagenomic library was constructed and analyzed employing 454-pyrosequencing technology, resulting in approximately 218 Mb of data. Three different classifiers applied allowed a broad overview of the most abundant taxonomic groups due to a huge number of metagenome reads remained unidentified. However, taxonomic analysis revealed Gammaproteobacteria and Deltaproteobacteria, respectively, as the most abundant classes, whereas species of the genus *Methanospirillum* were dominant among methanogenic Archaea. The analysis of the microbial community metabolism by means of COG and Carma databases revealed that the degradation of biomass depends on different phylogenetic groups, for instance, Bacteroidia and Gammaproteobacteria were indicated as involved into the degradation of carbohydrate and proteins. Furthermore, the analysis suggested Clostridia and Methanomicrobiales and Methanosarcinales as the main microorganisms producing hydrogen and methane, respectively. Analysis of the six coding sequences of FeFe-hydrogenases identified into the dataset revealed that they represented novel target gene sequences. However, only four of these coding sequences could be detected into the fosmid library by PCR screening. The combined results obtained in this study allowed us to have an insight of the composition and potential metabolism of the microbes residing in the analyzed domestic sewage

treatment plant and suggested such environment as a potential reservoir for new hydrogenase genes to biotechnological exploration.

LISTA DE TABELAS

Tabela 1. O Calor de combustão dos combustíveis mais comuns.....	03
Tabela 2. Sequências dos <i>primers</i> utilizados para amplificação do gene RNAr 16S de bactérias e arqueias.....	25
Tabela 3. Sequências dos <i>primers</i> degenerados para amplificação dos genes hidrogenases.....	26
Tabela 4. Sequências dos <i>primers</i> degenerados para amplificação dos genes de NiFe-hidrogenases.....	27
Tabela 5. Sequências dos <i>primers</i> para amplificação das CDS de FeFe-hidrogenase.....	36
Tabela 6. Distribuição das OTUs bacterianas afiliadas aos grupos taxonômicos identificados nas amostras de RAFA.....	46
Tabela 7. Estimativas de riqueza e índices de diversidade das bibliotecas do gene de RNAr 16S do domínio Bacteria e Archaea.....	57
Tabela 8. Dados estatísticos do pirosequenciamento da biblioteca de fosmídeos.....	61
Tabela 9. Classificação taxonômica das sequências metagenômicas da biblioteca de fosmídeo baseado no menor ancestral comum.....	64
Tabela 10. Classificação taxonômica das sequências metagenômicas da biblioteca de fosmídeo gerados pelo Carma3.....	66
Tabela 11. Lista de EGTs associadas aos processos de hidrólise identificados na biblioteca de fosmídeo.....	69
Tabela 12. Lista de EGTs associadas à produção de hidrogênio e metano identificados na biblioteca de fosmídeo.....	72
Tabela 13. Distribuição e lista de EGTs atribuídas às subunidades maior e menor de FeFe-hidrogenases identificadas na biblioteca de fosmídeo.....	73

Tabela 14. Principais categorias de grupos ortólogos de proteínas identificadas nas sequencias codificantes do conjunto de <i>contigs</i> metagenômicos da biblioteca de fosmídeo.....	77
Tabela 15. Classificação taxonômica das sequencias codificantes de FeFe-hidrogenase identificadas na biblioteca de fosmídeos.....	80

LISTA DE FIGURAS

- Figura 1.** Rotas metabólicas de produção de hidrogênio em Clostridia. PFOR: piruvato-ferredoxina oxirredutase e NFOR: NADH ferredoxina oxidorredutase..... 07
- Figura 2.** Rota metabólica de produção de hidrogênio em micro-organismos facultativos. FHL: formato – hidrogênio - liase é uma via de produção de hidrogênio descrita em *Escherichia coli*..... 08
- Figura 3.** Classificação das hidrogenases (adaptado de KIM; KIM, 2011)..... 09
- Figura 4.** Esquema do funcionamento da Estação de Tratamento de Esgoto da Bacia do Piçarrão pertencente à SANASA (SANASA, 2013)..... 20
- Figura 5.** Eletroforese de DNA metagenômico em campo pulsado. **Painel A)** canaletas: 1, marcador de 36 Kb; 2, método de Fiantdt; 3, método de Fiantdt purificado com cloreto de céσιο; 4, método de Fiantdt purificado pelo Wizard® DNA Clean-Up System. **Painel B)** 1, método de Zhou; 2, método de Zhou purificado pelo Wizard® DNA Clean-Up System; 3, PowerSoil® DNA Isolation. M, marcado de massa molecular..... 39
- Figura 6.** Amplificação do gene RNAr 16S do DNA obtido pelo método de Fiantdt purificados com kit Wizard® DNA Clean-Up System. M, Marcador de peso molecular 1 kb DNA ladder; 1, controle negativo (sem DNA); 2, controle positivo utilizando DNA de uma linhagem de *Streptomyces* sp.; 3, amostra de DNA sem diluição; 4, amostra de DNA diluída 10X; 5 amostra de DNA diluída 100X..... 40
- Figura 7.** Eletroforese de DNA metagenômico extraído pelo método indireto. Canaletas: 1, marcador de 36 kb; 2, extração por lise indireta em *plugs*; M, marcador de peso molecular..... 41
- Figura 8.** Árvore filogenética das sequências parciais dos genes RNAr 16S representativas das OTUS do Domínio Archaea na amostra de lodo e organismos relacionados recuperados da base de dados, utilizando o método de Neighbour-

Joining e distância evolutiva Kimura 2p. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão listados. Os números entre parêntese correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. *Desulfurococcus kamchatkensis* 1221n^T utilizado como grupo externo (*outgroup*) 42

Figura 9. (página anterior). Árvore filogenética baseada nas sequências parciais do gene RNAr 16S dos clones representativos das OTUS do Domínio Bacteria e organismos relacionados recuperados das bases de dados. Os números entre parêntese correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão indicados. *Methanosarcina lacustris* foi utilizada como grupo externo (*outgroup*)..... 46

Figura 10 (página anterior). Árvore filogenética baseada nas sequências parciais do gene RNAr 16S dos clones representativos das OTUS dos membros do filo Proteobacteria e organismos relacionados recuperados das bases de dados. Os números entre parênteses correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão indicados. *Methanosarcina lacustris* foi utilizada como grupo externo (*outgroup*)..... 48

Figura 11. Painel A): Árvore filogenética baseada no alinhamento das sequências putativas de aminoácidos de FeFe-hidrogenase e suas sequências relacionadas recuperadas das bases de dados. As distâncias evolutivas foram calculadas pelo modelo de Poisson e a construção da árvore pelo método “Neighbor-joining”. Os números entre parêntese correspondem a clones adicionais com similaridade $\geq 82\%$ com o clone representado na árvore. Os valores de *bootstrap* (1000 re-amostragens) acima de 70% estão indicados. A sequência da proteína Nar1p pertencente à *Saccharomyces cerevisiae*, homóloga a *HydA* foi utilizada como grupo externo (*outgroup*) . **Painel B)** Sequências de aminoácidos deduzidos a partir dos clones deste estudo e sequências de referência de *HydA* ilustrado a coerência filogenética das novas substituições dentro do motif L2(MPCxxKxxE). *Clusters* designados de C1 a C4..... 52

Figura 12. Curvas de rarefação das OTUs identificadas com base nas sequências parciais de DNAr 16S. □ e Δ, distância evolutiva para o domínio Bacteria, 0.03 e 0.2, respectivamente. ○ e × distância evolutiva para o domínio Archaea , 0.03 e 0.2, respectivamente.....	55
Figura 13. Eletroforese em campo pulsado dos produtos de restrição do DNA fosmidial dos clones metagenômicos. Canaletas: 1, DNA fosmidial não clivado; 2-6 DNA fosmidial de clones distintos; 7, vetor clivado; 8, vetor intacto; 9, marcador de peso molecular (1 kb).....	57
Figura 14. Número de pares de bases por sequência e número de sequências no conjunto de dados metagenômicos.....	59
Figura 15. Composição dos domínios das sequências codificantes (CDS) de FeFe-hidrogenase. Domínios conservados e motifs foram identificados em todas as CDS usando InterProScan e Pfam. Os seguintes domínios foram identificados: fer2_4 – pfam13510, 2Fe-2S <i>cluster</i> [FeS]; NADH-G_4Fe-4S_3 – pfam10588, [FeS] oxidorreductase; Fe_hyd_lg_C – pfam02906, FeFe-hidrogenase subunidade maior, domínio C-terminal; Fe_hyd_SS – pfam02256, FeFe-hidrogenase subunidade menor; Fer4_10 – pfam13237, 4Fe-4S domínio <i>dicluster</i> ; FeS – pfam04060, putativo <i>cluster</i> [FeS]; Fer4 – pfam00037, 4Fe-4S domínio; Fer4_9 – pfam13187, 4Fe-4S domínio <i>dicluster</i>	77
Figura 16. Amplificação das sequências codificantes de FeFe-hidrogenase identificadas na biblioteca de fosmideos. Amplificação das seguintes CDS: CDS_959; CDS_14436; CDS_7499 e CDS_12638. M, marcador 1 kb; 1, clone_2F; 2, clone_1D; 3, clone_7D; 4 clone_1C; 5,clone_2F; 6, clone_1D;7, clone_1C; 8, clone_7D.....	80

SUMÁRIO

1.	Introdução.....	01
2.	Revisão Bibliográfica.....	03
2.1	Biomassa para a produção de biohidrogênio e metano.....	03
2.2	Digestão anaeróbica: metabolismo e micro-organismos envolvidos.....	04
2.3	Hidrogenases: produção fermentativa de biohidrogênio.....	06
2.3.1	Diversidade de hidrogenases.....	08
2.4	Metanogêneses: vias de produção de metano.....	11
2.5	Metagenômica: acessando micro-organismos não-cultivados.....	12
2.5.1	Estratégias de triagem de bibliotecas metagenômicas.....	14
2.6	Bioinformática aplicada à análise do conjunto de dados metagenômicos.....	15
3	Objetivos.....	19
3.1	Objetivos Gerais.....	19
3.2	Objetivos Específicos.....	19
4.	Material e Métodos.....	20
4.1	Amostragem.....	20
4.2	Extração de DNA.....	21
4.2.1	Extração de DNA por método direto.....	22
4.2.2	Extração de DNA por método indireto.....	22
4.3	Purificação da solução de DNA.....	23
4.4	Construção das bibliotecas de genes RNAr 16S.....	24
4.5	Construção das bibliotecas gênicas de hidrogenases.....	26
4.6	Desenho de <i>primers</i> de NiFe-hidrogenase.....	27
4.7	Sequenciamento das bibliotecas gênicas.....	28
4.8	Análises estatísticas.....	28
4.9	Análise filogenética.....	29
4.10	Construção da biblioteca metagenômica de fosmídeo.....	30
4.11	Triagem das sequências de FeFe-hidrogenase na biblioteca de metagenômica...	31
4.12	Extração de DNA fosmidial e triagem por PCR.....	31
4.13	Pirosequenciamento.....	32
4.14	Análise do conjunto de dados metagenômicos.....	33
4.14.1	Análises das sequências metagenômicas na plataforma MetaSAMS.....	34
4.14.2	Identificação de sequências codificantes e anotação automática.....	35
4.15	Desenhos de <i>primers</i> para as CDS de FeFe-hidrogenase	35
4.16	Triagem por PCR da biblioteca metagenômica para a identificação dos clones contendo as CDS de FeFe-hidrogenase	36
4.17	Sequenciamento dos fosmídeos contendo as CDS de FeFe-hidrogenase.....	37
5	Resultados e Discussão.....	38
5.1	Avaliação dos métodos de extração de DNA.....	38
5.2	Diversidade da comunidade microbiana.....	42

5.3	Diversidade de hidrogenases	51
5.4	Índices de riqueza e diversidade das bibliotecas gênicas.....	54
5.5	Construção da biblioteca metagenômica e triagem por PCR.....	56
5.6	Análises estatísticas dos dados metagenômicos.....	58
5.7	Composição taxonomia da biblioteca metagenômica.....	60
5.8	Potencial metabólico da comunidade microbiana.....	66
5.9	Análise funcional da biblioteca metagenômica de acordo com COG.....	73
5.10	Triagem por PCR dos clones contendo as CDS de FeFe-hidrogenase.....	79
6	Conclusões.....	81
7	Perspectivas Futuras.....	82
	REFERÊNCIAS.....	83

1. INTRODUÇÃO

Estimativas indicam que 80 % da energia consumida anualmente no mundo são provenientes da queima de combustíveis fósseis e que as projeções sobre as reservas desses combustíveis são discordantes devido aos tipos ou categorias de petróleo que são estimados (McGLADE, 2012; SORRELL et al., 2010). Esses dados sugerem a possibilidade de ocorrer um colapso no abastecimento energético mundial devido a nossa total dependência dessa fonte de recurso para a geração de energia. Somam-se ainda o aumento de preço do petróleo, tendo sempre como justificativa as suas reservas finitas, as instabilidades políticas nos principais países produtores e a degradação do meio ambiente (ELREFAEI et al., 2013). Trata-se de um enorme desafio a geração de energia e a sua diversificação nos próximos anos, uma vez que a demanda por energia é crescente, principalmente em países em desenvolvimento, os quais têm níveis de energia *per capita* muito menores que os países desenvolvidos (World Energy Council, 2011).

Nesse contexto, a biomassa é considerada como a principal fonte de energia renovável e menos poluente em substituição ao petróleo. É uma fonte de recurso de baixo custo e geograficamente distribuída que pode contribuir para a geração de energia descentralizada, evitando a existência de monopólio, o qual atualmente se observa com as reservas dos combustíveis fósseis (DEMIRBAS, 2007). Toda biomassa, a qual pode ser composta de plantas ou somente de resíduos orgânicos, tais como resíduos industriais, agrícolas, municipais ou mesmo dejetos de animais, pode ser convertida em energia via processo termoquímico ou biológico (CANTRELL et al., 2008; GUO et al., 2010).

O processo biológico mais utilizado para a obtenção de energia a partir desses resíduos orgânicos é a digestão anaeróbica. A digestão anaeróbica compreende a degradação destes resíduos por uma comunidade complexa de micro-organismos anaeróbios (ZIEMINSKI; FRAC, 2012). Esse processo biológico ocorre basicamente em quatro fases: hidrólise, acidogênese, acetogênese e metanogênese. Durante este processo biológico é produzido o hidrogênio por micro-organismos, denominado biohidrogênio, e o seu consumo é essencial para as interações sintróficas necessárias para o processo de digestão anaeróbica (BOYD; SPEAR; PETERS, 2009). Portanto, muito deste hidrogênio produzido é indetectável, pois é imediatamente consumido, principalmente por arqueias metanogênicas para a produção de metano (ZIEMINSKI; FRAC, 2012).

A partir da digestão anaeróbica de diferentes resíduos orgânicos pode se produzir hidrogênio e metano como fonte de energia alternativa e ecologicamente correta (LIN et al., 2012; HERRMANN; RATH, 2012). No entanto, o processo de digestão anaeróbica requer uma série complexa de reações bioquímicas, a qual é realizada por diversos grupos de microorganismos e necessita de mais estudos para completa elucidação das suas limitações (SCHLÜTER et al., 2008; JAENICKE et al., 2011; RADEMACHER et al., 2011). Porém, metodologias clássicas da microbiologia, tais como o isolamento e cultivo de microorganismos, entre outros, não são suficientes para a total recuperação da diversidade e análise do metabolismo microbiano.

Abordagens moleculares desenvolvidas nas últimas décadas (extração de ácidos nucleicos, amplificação por PCR, clonagem e sequenciamento de DNA) têm sido otimizadas e adaptadas para superar as limitações impostas pela metodologia clássica de estudo de populações microbianas. A utilização dessas abordagens moleculares vem permitindo uma avaliação mais precisa da composição da comunidade microbiana no ambiente, a descoberta de novos grupos de organismos e ainda a identificação de novas sequências de genes (HANDERMAN; SJÖLING, 2007; HJORT et al., 2010; JIMÉNEZ et al., 2011; JAENICKE et al., 2011; RADEMACHER et al., 2011).

As estratégias consistem na construção de bibliotecas de pequenos ou grandes insertos de DNA obtidos diretamente da comunidade microbiana (metagenoma) de ambientes como água, solo, sedimento ou lodo, e subsequentemente sequenciamento em larga escala e/ou análise dessas bibliotecas resultantes na busca da expressão de atividade biológica de interesse na linhagem hospedeira (HANDERMAN; SJÖLING, 2007; HJORT et al., 2010; CASTRO et al., 2011). Recentemente, o emprego do sequenciamento de segunda geração dessas bibliotecas e, mais frequentemente, do DNA extraído diretamente de amostras tem gerado um extenso conjunto de dados sem precedentes em comparação com as técnicas tradicionais de microbiologia (SIMON; DANIEL, 2011). A metagenômica tem sido uma poderosa ferramenta para revelar sequências genômicas inéditas de vários organismos em ambiente natural sem a necessidade de isolamento e cultivo da espécie individual (EKKERS et al., 2012).

2. REVISÃO BIBLIOGRÁFICA

2.1 Biomassa para a produção de biohidrogênio e metano

A perspectiva iminente de uma crise do petróleo e a crescente consciência dos danos ambientais em consequência da exploração e utilização maciça dessa fonte de recurso ao longo das últimas décadas estão conduzindo o mercado de combustíveis para a busca de combustíveis alternativos. Entre os combustíveis alternativos, estão os biocombustíveis que têm potencial para substituir parcialmente os combustíveis fósseis, especialmente no setor de transporte devido a sua disponibilidade ilimitada e seu reduzido impacto ambiental (LIN et al., 2012).

Neste contexto, a conversão de resíduos orgânicos em hidrogênio biológico, biohidrogênio, e metano por digestão anaeróbica é uma promissora abordagem para atender a crescente necessidade de energia. E ainda, o hidrogênio e o metano têm maior calor de combustão por quilograma quando comparados aos outros combustíveis (Tabela 1).

Tabela 1. Calor de combustão dos combustíveis mais comuns.

Combustível	Calor de Combustão (MJ/Kg)	Densidade (Kg/m ³)	Calor de combustão (MJ/m ³)
Hidrogênio	120,1	0,09	10,8
Metano	50,0	0,72	36
Propano	45,6	1,83	83,4
Gás natural (m ³)	29,3-35,7	0,7-0,9	20,5-32,1
Gasolina	41,7-44,1	740	(31 – 32) x 10 ³
Diesel	42 – 44	890	(37 – 39) x 10 ³
Etanol	27,7	720	20 x 10 ³

Adaptado de Da Silva (2010).

Resíduos orgânicos provenientes das atividades humanas, tais como resíduos agrícolas, municipais e alimentícios são uma fonte de recurso que não depende das condições climáticas, sazonalidade e pode ser armazenado para a sua utilização de acordo com a

demanda. Além disso, o uso deste tipo de biomassa representa também uma oportunidade para reduzir a emissão de gases do efeito estufa, gerar novas oportunidades de emprego e diminuir a dependência dos combustíveis fósseis (CARNEIRO; FERREIRA, 2012).

Águas residuais de fábrica de azeite, polpa de azeitona e soro de queijo são exemplos da utilização de resíduos orgânicos na produção de biohidrogênio em escala de laboratório (NTAIKOU; LYBERATOS, 2010; LIN et al., 2012). Por outro lado, a produção e o emprego do metano já são realizados em muitos países da Europa, por exemplo, na Alemanha mais de 5.800 biodigestores espalhados pelo país estão gerando energia a partir da digestão anaeróbica de resíduos agrícolas (RADEMACHER et al., 2012).

Adicionalmente, a geração de biohidrogênio ou metano a partir de resíduos orgânicos também é uma alternativa a outros biocombustíveis, tais como bioetanol e biodiesel, porque estes requerem a utilização de recursos como terra e água. Preocupações éticas têm sido levantadas quanto ao uso desses recursos naturais para a produção de combustíveis ao invés de serem utilizados na produção de alimentos uma vez que muitas pessoas estão morrendo de fome (RITTMANN, 2008; LIN et al., 2012).

2.2 Digestão anaeróbica: metabolismo e micro-organismos envolvidos

Digestão anaeróbica é um processo bioquímico de decomposição da matéria orgânica em metano e dióxido de carbono realizado por diferentes grupos de micro-organismos (ZIEMINSKI; FRAC, 2012). Dependendo da temperatura na qual o processo é realizado, a digestão anaeróbica pode ser classificada em: i) “termofílica”, realizada entre 45° a 60° C; ii) “mesofílica”, realizada entre 20° a 45° C; e iii) psicofílica, realizada abaixo de 20° C (KASHYAP; DADHICH; SHARMA, 2003 ; SCHLÜTER et al., 2008; RADEMACHER et al., 2011). Independente da temperatura do processo, a conversão da biomassa a metano sob condições anaeróbicas ocorre em quatro estágios: hidrólise, acidogênese, acetogênese e metanogênese.

A primeira etapa mediada por diferentes hidrolases extracelulares envolve a transformação de material orgânico insolúvel e compostos de alta massa molecular, tais como lipídeos, polissacarídeos e proteínas, em material orgânico solúvel, compostos adequados para a sua utilização como fontes de energia e carbono, tais como monossacarídeos, aminoácidos e outros compostos orgânicos simples (GUELLIL et al., 2001; ZIEMINSKI; FRAC,

2012). Essa etapa denominada hidrólise é considerada o estágio limitante da velocidade da digestão anaeróbica, pois durante a digestão de resíduos sólidos apenas 50 % dos compostos orgânicos são degradados (MYINT; NIRMALAKHANDAN; SPEECE, 2007). O restante desses compostos permanece no seu estado primário devido à falta de enzimas que participam do processo de degradação (PARAWIRA et al., 2008). A hidrólise é basicamente realizada pelas classes Clostridia, Bacteroidia e Gammaproteobacteria, todos pertencentes ao domínio Bacteria (SCHLÜTER et al., 2008; RADAMACHER et al., 2011).

Na acidogênese, segunda etapa do processo, as bactérias acidogênicas convertem os produtos da hidrólise em ácidos orgânicos de cadeia curta ou voláteis (propionato, butirato, succinato), álcoois e aldeídos. Nessa fase, as bactérias facultativas consomem o oxigênio acidentalmente introduzido no processo, proporcionando condições favoráveis para o desenvolvimento de anaeróbicos obrigatórios, tais como *Clostridium*, *Flavobacterium*, etc (ZIEMINSKI; FRAC, 2012). Subsequentemente, bactérias acetogênicas consomem os produtos gerados na fase anterior produzindo principalmente acetato, dióxido de carbono (CO₂) e hidrogênio (H₂) (MYINT; NIRMALAKHANDAN; SPEECE, 2007). No entanto, o H₂ e CO₂ são imediatamente consumidos pelas próprias bactérias acetogênicas, principalmente aquelas pertencentes aos gêneros *Syntrophomonas* e *Syntrophobacter* para a produção de acetato ou por arqueais metanogênicas para a produção de metano (DRAKE; KÜSEL; MATTHIES, 2002; ZIEMINSKI; FRAC, 2012). Esta associação sintrófica entre micro-organismos produtores e consumidores de hidrogênio é essencial para a metanogênese, última etapa da digestão anaeróbica (BOYD; SPEAR; PETERS, 2009). Nessa fase também, o excesso de elétrons gerados do metabolismo é transferido aos prótons pela atividade das hidrogenases, principalmente presentes em micro-organismos estritamente anaeróbicos (e.g. Clostridiaceae e Streptococcaceae) (CHEN et al., 2005; O-THONG et al., 2008).

A última etapa, a metanogênese, é realizada pelas arqueias metanogênicas hidrogenotróficas, tais como: *Methanobacterium* spp., *Methanomicrobiales* spp. e *Methanococcus* spp. que convertem o H₂ e CO₂ em metano (ROTHER; METCLAF, 2005; TALBOT et al., 2008), ou pelas metanogênicas acetoclásticas, tais como: *Methanosaeta* spp. e *Methanosarcina* spp. (TALBOT et al., 2008), que convertem o acetato em metano, todas metanogênicas pertencentes ao filo Euryarchaeota.

Essa conversão de compostos orgânicos complexos a metano e dióxido de carbono requer uma mistura de espécies bacterianas que dependem umas das outras (CANTRELL et al., 2008). Mas, apesar do conhecimento de alguns membros de bactérias e arqueias

metanogênicas na digestão anaeróbica, estudos metagenômicos e, recentemente, uma análise metatranscriptômica de reatores de biogás, os quais forneceram informações mais abrangentes sobre a diversidade e abundância dos membros das comunidades microbianas e daqueles metabolicamente ativos, concluíram que ainda pouco se sabe sobre a comunidade de micro-organismos envolvidos na digestão anaeróbica e suas interações (SCHLÜTER et al., 2008; JAENICKE et al., 2011; RADEMACHER et al., 2011; ZAKRZEWSKI et al., 2012).

2.3 Hidrogenases: produção fermentativa de biohidrogênio

A produção fermentativa de hidrogênio nos últimos anos tem sido ativamente estudada devido a sua flexibilidade em utilizar águas residuais ou resíduos sólidos como fonte de recurso renovável, caracterizando o processo como ecologicamente correto (MUDHOO; SÁNCHEZ, 2011; LIN et al., 2012; KIM; CHA, 2013).

Numerosos micro-organismos podem produzir hidrogênio por reações ligadas ao seu metabolismo. Dependendo da espécie, o hidrogênio pode ser produzido pela via piruvato-ferredoxina oxirredutase (PFOR) e NADH ferredoxina oxidorredutase (NFOR) (Figura 1) ou formato – hidrogênio - liase (FHL) (Figura 2). No entanto, qualquer via de produção de hidrogênio é fundamentalmente dependente da presença de enzimas produtoras de hidrogênio. Por exemplo, durante a digestão anaeróbica, o excesso de elétrons, produzidos durante o metabolismo microbiano, é transferido aos prótons reduzindo-os a hidrogênio. A reação ($2H + 2e^- \leftrightarrow H_2$) é reversível, e o seu sentido depende do potencial redox dos componentes que são capazes de interagir com as hidrogenases (VIGNAIS; BILLOUD, 2007), enzima chave dessa reação.

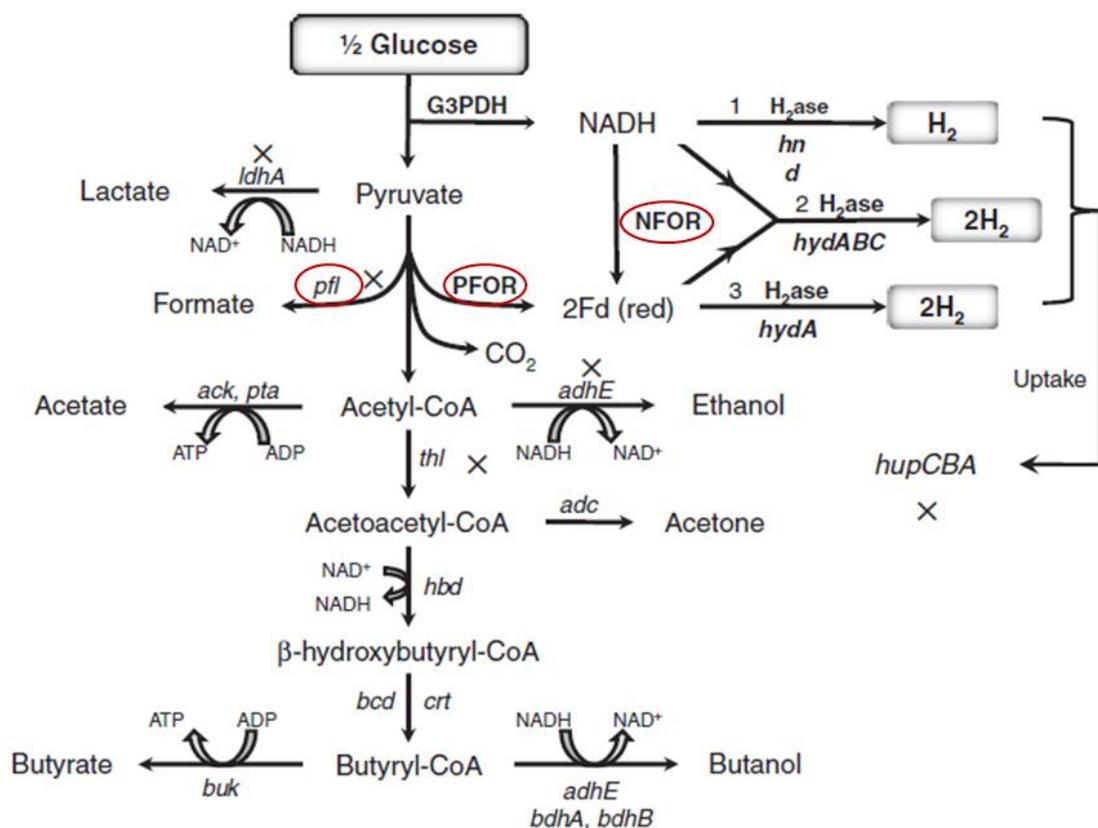


Figura 1. Rotas metabólicas de produção de hidrogênio em Clostridia. PFOR: piruvato-ferredoxina oxirredutase e NFOR: NADH ferredoxina oxidorredutase (ABO-HASHESH; WANG; HALLENBECK, 2011).

A elucidação da informação genética das hidrogenases tem permitido numerosas abordagens para a expressão dessas enzimas em hospedeiros heterólogos na tentativa de otimizar a produção de hidrogênio (ZHAO et al., 2010). E a exploração de sequências obtidas a partir da genômica e metagenômica auxiliará a identificação de novas hidrogenases potencialmente úteis. Além disso, essas enzimas possuem uma variedade de aplicações biotecnológicas, tais como em células biocombustíveis, biosensores e prevenção contra a corrosão induzida por micro-organismos (MERTENS; LIESE, 2004).

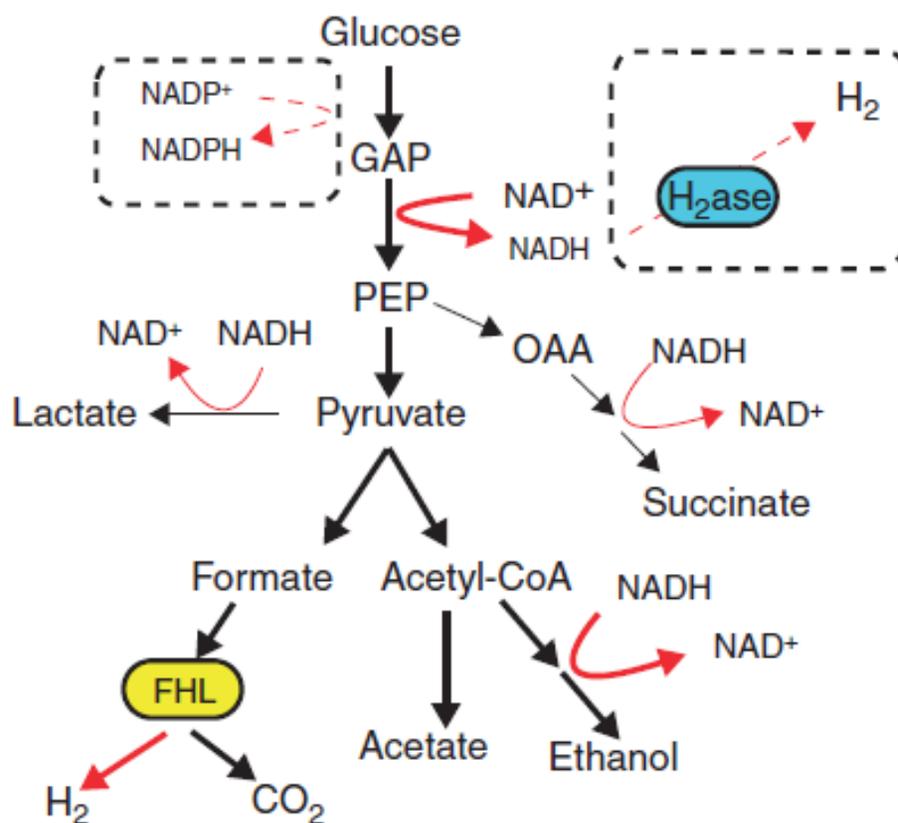


Figura 2. Rota metabólica de produção de hidrogênio em micro-organismos facultativos. FHL: formato – hidrogênio - liase é uma via de produção de hidrogênio descrita em *Escherichia coli* (WAYNE; NOGUERA; DONOHUE, 2012).

2.3.1 Diversidade de hidrogenases

As hidrogenases são classificadas em três grupos filogeneticamente distintos (Figura 3), de acordo com os íons metálicos presentes em seu sítio ativo: 1) Fe-hidrogenases, contêm um átomo de ferro em seu centro ativo; 2) FeFe-hidrogenases, contêm dois átomos de ferro em seu centro ativo; e 3) NiFe-hidrogenases, contêm níquel e

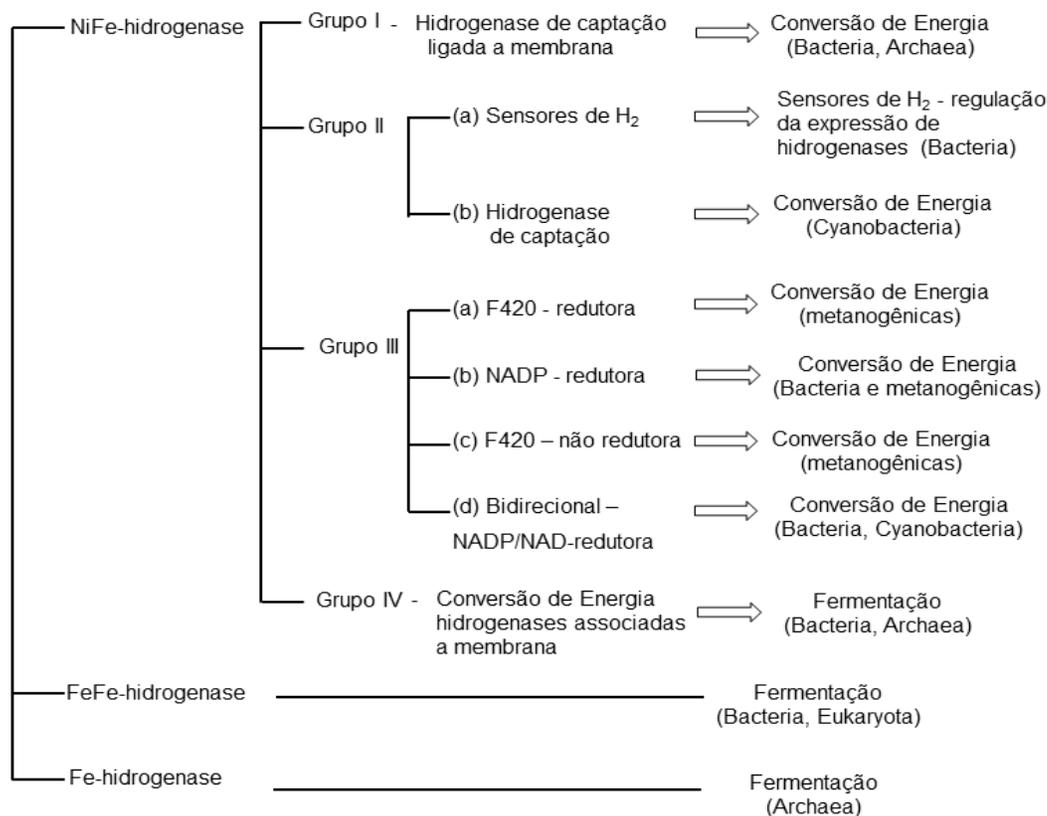


Figura 3. Classificação das hidrogenases (adaptado de KIM; KIM, 2011).

ferro em seu centro ativo (VIGNAIS; BILLOUD, 2007). NiFe-hidrogenases estão mais envolvidas no consumo de hidrogênio, enquanto que FeFe-hidrogenases estão preferencialmente direcionadas para a produção de hidrogênio (VIGNAIS; COLBEAU, 2004). Apesar de classificadas como hidrogenases, Fe-hidrogenases são descritas somente em algumas arqueias metanogênicas e catalisam a reação de produção de metano a partir de H₂ e CO₂.

NiFe-hidrogenases são as mais abundantes e mais estudadas das classes de hidrogenases. A análise da sequência completa dessa hidrogenase, subunidade maior e menor, permite a sua classificação em quatro grupos, o que é consistente com as funções das enzimas (KIM; KIM, 2011). Essas hidrogenases são codificadas por genes organizados em *clusters*, sendo que alguns desses *clusters* também possuem os genes regulatórios que

controlam a expressão dos genes estruturais. Dentro de cada uma das classes, a filogenia da hidrogenase é essencialmente consistente com a taxonomia das espécies das quais são derivadas.

O Grupo I consiste de NiFe-hidrogenases, as quais são geralmente associadas a *Wolinella succiogenes*, *Aquifex aeolicus* e alguns membros do gênero *Desulfovibrio*, são ligadas à membrana e estão envolvidas na oxidação do hidrogênio. O Grupo II consiste de NiFe-hidrogenases encontradas no citoplasma e que possuem a função de detectar a presença de hidrogênio no ambiente para desencadear uma cascata de reações celulares que controlam a síntese de hidrogenases (VIGNAIS; BILLOUD, 2007). Esse grupo de hidrogenase está associado à *Bradyrhizobium japonicum*, *Rhodobacter eutropha* e a cianobactérias (KIM; KIM, 2011).

O terceiro grupo é composto de NiFe-hidrogenases associadas a cofatores que exercem uma função bidirecional, pois dependendo das condições podem produzir hidrogênio. Membros dessas hidrogenases são encontradas em *R. eutropha*, *Methanosarcina mazei*, *Methanothermobacter marburgensis* (VIGNAIS; BILLOUD, 2007).

O grupo IV inclui as NiFe-hidrogenases associadas a membranas que estão envolvidas na produção de hidrogênio. Essas hidrogenases estão acopladas ao complexo de hidrogênio- formato-liase encontrado em *Escherichia coli* que degrada o ácido fórmico em hidrogênio e dióxido de carbono (VIGNAIS; COLBEAU, 2004).

Já as FeFe-hidrogenases são em sua maioria monoméricas e compostas de um domínio denominado “domínio H” onde está acomodado o seu “cluster H”. Este *cluster* é composto por um sítio ativo de [FeFe] ligado a um *cluster* [2Fe-4S] por uma ponte de cisteína (MEYER, 2007). As regiões mais conservadas do domínio do “cluster H” são aquelas em torno da ligação de cisteína do sítio ativo bimetálico, as quais podem ser identificadas em sua sequência primária por três diferentes motifs de ligação denominados L1 (TSCCPxW), L2 (MPCxxKxxE) e L3 (ExMACxxGCxxGGGxP). A presença desses motifs permite uma simples identificação das verdadeiras FeFe-hidrogenases e as distingue das sequências proximamente relacionadas (MEYER, 2007; MULDER et al., 2011).

As maiores diferenças estruturais entre as FeFe-hidrogenases estão relacionadas com seu respectivo complemento de *clusters* FeS (*F-cluster*), que putativamente funcionam de modo a regular a transferência de elétrons para o/ou do local do *cluster* H ativo (MULDER et al., 2011). A análise filogenética proposta por Meyer (2007) classifica as FeFe-hidrogenases

em oito grupos distintos, as quais são únicas de acordo com a arquitetura de domínio e/ou a composição de suas subunidades.

As FeFe-hidrogenases são encontradas em anaeróbicos procariotos, tais como clostridia e bactérias sulfato-redutoras, e em alguns eucariotos anaeróbicos. Porém, observou-se que a maioria das sequências de FeFe-hidrogenases está relacionada à classe *Clostridia* (CALUSINSKA et al., 2010).

Métodos para determinar a diversidade de FeFe-hidrogenases em uma população incluem a utilização de *primers* degenerados para este gene (*hyd*). Um estudo recente de amostras de pântano descreveu sequências parciais de *hydA* associadas aos gêneros *Clostridium* e *Thermosinus*. A análise de um consórcio microbiano produtor de hidrogênio e etanol, cultivado em águas residuais de melaço, também analisou a presença do gene *hydA* nessa população, constatando a grande variedade dessas enzimas. Todas, porém, foram relacionadas à classe Clostridia (DEFENG; REN; RITTMANN, 2008).

2.4 Metanogêneses: vias de produção de metano

Os organismos pertencentes ao domínio Archaea são considerados únicos, com propriedades metabólicas extraordinárias e filogenia particular. Certas arqueias apresentam vias metabólicas semelhantes às aquelas encontradas em organismos do domínio Bacteria. Porém, outras apresentam vias metabólicas raras, como as metanogênicas, que produzem gás metano através de sistemas enzimáticos únicos (ZIEMINSKI; FRAC, 2012). Existem dois grupos de metanogênicas: as hidrogenofílicas, as quais consomem hidrogênio para a produção de metano, e as acetoclásticas, que convertem o acetato em metano (BAUER et al., 2008; CONRAD et al., 2006).

A via hidrogenotrófica é composta por diferentes enzimas tais como: formil metano furano desidrogenase, formil metano furano tetrahydro-metanopterin (H4MP), formil transferase, metenil-H4MP-ciclohidrolase, coenzima redutase F420-dependente de metileno-H4MP e H4MP S-metiltransferase. Essas são as principais enzimas que catalisam o processo de conversão de CO₂ e hidrogênio a metano (ZIEMINSKI; FRAC, 2012). Na produção de metano pela via acetoclástica, o acetato é fosforilado e ativado a acetil-CoA para então ser processado pelo complexo acetil-CoA descarboxilase. Essa enzima cliva o acetil-CoA para a

transferência do grupo metil a coenzima M, produzindo metil-CoM. Esse último produto é convertido a metano pela atividade enzimática da metil coenzima M redutase (FERRY, 1997).

A metil-coenzima M redutase é a enzima chave de toda a metanogênese, pois, é necessária na última etapa de produção de metano das duas vias de produção desse gás.

A caracterização das arqueias metanogênicas é importante também como um auxílio na identificação das espécies da comunidade metanogênica, pois a reconstrução filogenética baseada no gene RNAr 16S parece não ser suficiente para a avaliação detalhada da composição dessa comunidade (LUTON et al., 2002; BAUER et al., 2008). Um valioso marcador para a análise da comunidade metanogênica é o gene *mcrA*, o qual codifica a subunidade alfa da metil-coenzima M redutase. Muitos estudos têm destacado o uso do gene *mcrA* como marcador para a detecção de metanogênicas em uma variedade de ambientes (JUOTTONEN; GALAND; YRJALA, 2006; RASTOGI et al., 2007; STEINBERG; REGAN, 2008; KRÖBER et al., 2009).

No entanto, a análise da diversidade desse gene não é suficiente para compreender o processo de produção de metano. Desta forma, o estudo da composição da comunidade metanogênica juntamente com a caracterização de genes e operons presentes em reatores anaeróbicos pode auxiliar na otimização da produção de metano.

2.5 Metagenômica: acessando micro-organismos não-cultivados

Os micro-organismos formam comunidades complexas no meio ambiente e estão envolvidos em todos os ciclos biogeoquímicos. Portanto, a identificação dos membros da comunidade e seus genes expressos têm sido vital para compreender as interações entre esses organismos e seus ambientes (ZWOLINSKI, 2007). Contudo, somente uma pequena fração da diversidade de micro-organismos no ecossistema, a qual está estimada entre 10^6 a 10^8 espécies, pode ser recuperada em cultura pura por métodos de cultivo disponíveis atualmente (AMANN; LUDWING; SCHLEIFER, 1995).

Neste sentido, para contornar as dificuldades e limitações associadas com as técnicas de microbiologia clássica, vários métodos moleculares baseados em ácidos nucleicos (DNA e RNA) foram desenvolvidos (SCHMEISSER et al., 2003). O desenvolvimento dessas técnicas possibilitou o início de uma nova era no campo da ecologia microbiana, cujo marco foi o uso da informação filogenética do gene RNA ribossomal (RNAr) 16S para identificar bactérias não cultiváveis do ambiente (PACE et al., 1985).

O gene RNAr 16S é uma sequência suficientemente bem conservada em várias espécies que pode ser amplificada por reação em cadeia da polimerase (PCR) utilizando *primers*, ou oligonucleotídeos iniciadores, homólogos a regiões altamente conservadas do gene. Esse gene pode ainda apresentar regiões suficientemente variadas para diferenciar espécies de bactérias ou arqueias (ZWOLINSKI, 2007). Embora a sequência de RNAr 16S seja uma ferramenta de alta resolução para identificação de micro-organismos, a desvantagem do seqüenciamento desse gene é a ausência de informações sobre a genômica funcional (WENG; RUBIN; BRSITOW, 2007).

Nos últimos anos, novas abordagens moleculares baseadas no isolamento de ácidos nucléicos diretamente das amostras ambientais, provaram ser ferramenta poderosa para comparar e explorar o perfil ecológico e metabólico de complexas comunidades microbianas, assim como identificar novas biomoléculas através do uso de bibliotecas construídas a partir do ácido nucléico isolado (RIESENFELD; SCHLOSS; HANDELSMAN, 2004; SIMON; DANIEL, 2011). Esta técnica, denominada metagenômica, tem como principal objetivo acessar sequências genômicas de vários micro-organismos em ambiente natural sem isolamento e cultivo da espécie individual, permitindo assim uma ampla investigação da diversidade taxonômica, de genes filogeneticamente relevantes, genes catabólicos e operon inteiros (HARDEMAN; SJÖLING, 2007; HESS et al., 2011; UHLIK et al., 2013).

O passo básico para construção de uma biblioteca metagenômica é a geração de fragmentos de DNA (isolados de amostras do ambiente) para a construção de bibliotecas de pequenos ou grandes fragmentos. Dependendo dos fragmentos de DNA, bibliotecas metagenômicas são construídas usando plasmídeos (até 8 Kb), fosmídeos, cosmídeos (cerca de 40 Kb) ou cromossomos artificiais bacterianos -BAC (acima de 40 Kb). A escolha do sistema de clonagem depende da qualidade do DNA, genes alvos, e estratégias de triagem da biblioteca. Em geral, bibliotecas de pequenos insertos podem ser exploradas para identificar novas biomoléculas codificadas por um simples gene ou pequenos operons. Enquanto que bibliotecas de grandes fragmentos são requeridas para a recuperação de *clusters* de genes, os quais codificam vias complexas do metabolismo (SIMON; DANIEL, 2011). Em princípio, as técnicas para a identificação de novas biomoléculas a partir das bibliotecas metagenômicas construídas por diferentes sistemas de clonagem, podem ser divididas em duas abordagens: baseada na expressão da atividade biológica e baseada em similaridade de sequências de nucleotídeos (RIESENFELD; SCHLOSS; HANDELSMAN, 2004).

2.5.1 Estratégias de triagem de bibliotecas metagenômicas

A triagem baseada em similaridade de sequências utiliza sequências conservadas de DNA para detectar genes-alvo que apresentam homologia às sequências já depositadas em bancos de dados públicos, tais como NCBI, Pfam, Greengenes. As sequências conservadas são usadas para desenhar sondas ou *primers* para garantir uma alta probabilidade de pareamento com sequências gênicas desconhecidas presentes no ambiente (SCHLOSS; HANDELSMAN, 2003). Nesse tipo de triagem, a técnica de amplificação por PCR é um método rápido para triar bibliotecas metagenômicas, especialmente quando *pools* de clones recombinantes são preparados. Por exemplo, Jogler e colaboradores (2010) identificaram *clusters* contendo vários genes putativos de magnetossomos em uma biblioteca metagenômica através da triagem por PCR. Uma vantagem adicional dessa triagem é a possibilidade da identificação de marcadores filogenéticos universais, tais como o gene RNAr 16S, que permitem realizar a correlação entre um determinado fragmento genômico a sua taxonomia (BÉJÀ et al., 2000).

Análise de polimorfismo do tamanho de fragmentos de restrição terminal (T-RFLP), hibridização, e recentemente, abordagens de alto desempenho, tais como, sequenciamento de segunda geração e microarranjos, são técnicas que podem ser utilizadas na triagem de bibliotecas metagenômicas baseadas nas sequências de nucleotídeos (HJORT et al., 2010; PARSLEY et al., 2011; DANHORN; YOUNG; DELONG, 2012; UHLIK et al., 2013).

A triagem de alto desempenho de microarranjos também tem sido empregada em análises de comunidade microbiana e, principalmente, no monitoramento de processos biogeoquímicos (UHLIK et al., 2013). Por exemplo, o microarranjo GeoChip permite uma análise de mais de 150 mil genes codificantes de enzimas responsáveis por processos biogeoquímicos, processos metabólicos, resistência a metais pesados, resistência a antibióticos, e degradação de poluentes (HE et al., 2007).

No entanto, a desvantagem dessa abordagem é a sua incapacidade de ligar as funções específicas às suas respectivas populações. Na tentativa de contornar esse obstáculo, a técnica SIP (do inglês *Stable Isotope Probing*) tem sido associada à abordagem metagenômica para direcionar a análise do gene alvo para a sua população ativa, e assim, suprimindo da análise a inatividade da maioria que não é o foco do estudo (UHLIK et al., 2013).

Já a triagem funcional é baseada na detecção da expressão de uma atividade biológica dos clones da biblioteca metagenômica. Como a informação das sequências não é necessária, essa é a única estratégia que tem o potencial de identificar genes inteiros, *clusters* de genes ou genes completamente novos que codificam funções conhecidas ou novas (SJÖLING; COWAN, 2008; SIMON; DANIEL, 2011). Nessa triagem os genes que codificam enzimas com baixa similaridade às sequências de proteínas conhecidas podem ser isolados, pois as sequências podem ser tão divergentes que a identificação baseada na homologia de sequências não seria bem sucedida (SJÖLING; COWAN, 2008).

Três diferentes abordagens têm sido utilizadas para a recuperação de novas biomoléculas: detecção fenotípica da atividade de interesse, complementação heteróloga em mutantes ou indução da expressão do gene (SIMON; DANIEL, 2011). Entretanto, essas abordagens exigem necessariamente a expressão heteróloga em uma célula hospedeira, presença de todos os genes necessários para sua expressão e disponibilidade de um ensaio para a função de interesse. Embora essas exigências têm sido alcançadas por algumas triagens, ainda é necessário não somente melhorar como aumentar a diversidade de sistemas de expressão vetor/hospedeiro (SCHLOSS; HANDELSMAN, 2003). Hospedeiros alternativos como, por exemplo, *Bacillus*, *Rhizobium* e *Pseudomonas* (HANDELSMAN et al., 1998), ou hospedeiros que possuem uma maquinaria de transcrição híbrida capaz de reconhecer promotores heterólogos, têm sido a base para a descoberta de novas moléculas (PAPOUTSAKIS, 2012).

Cada abordagem tem sua própria eficácia e limitações, no entanto, uma combinação da metagenômica baseada na sequência e metagenômica funcional é a abordagem mais indicada para explorar o mundo microbiano, pois nenhuma das estratégias isoladas pode definir toda a diversidade das funções dos genes, monitorar as interações dinâmicas com o ambiente e as mudanças na comunidade de micro-organismos (RIESENFELD; SCHLOSS; HANDELSMAN, 2004; ZWOLINSKI, 2007).

2.6 Bioinformática aplicada à análise do conjunto de dados metagenômicos

Recentemente, a metagenômica tem sido transformada pela aplicação do seqüenciamento de alto desempenho, pois, fornece uma profunda análise da amostra sem precedentes em comparação com abordagens tradicionais da biologia molecular, tais como a

eletroforese desnaturante em gel gradiente (DGGE), análise de polimorfismos do tamanho de fragmentos de restrição terminal (T-RFLP), ou mesmo o seqüenciamento de genes de RNAr 16S pelo método de Sanger (PETROSINO et al., 2009; CHISTOSERDOVA, 2010; SIMON; DANIEL, 2011).

Atualmente, existem cinco plataformas para o seqüenciamento metagenômico: seqüenciamento pelo método de Sanger, piroseqüenciamento (454 – Roche), SOLiD™ (Applied Biosystems), Illumina® (Solexa) e Ion Torrent™ (Life Tehnologies). Cada uma destas tecnologias possui suas vantagens e desvantagens no que diz respeito ao desempenho do seqüenciamento, incluindo o custo, tamanho das sequências geradas pela plataforma de seqüenciamento, e taxas de erros (WILLIAMSON; YOOSEPH, 2012). Entre essas, a tecnologia de piroseqüenciamento tem gerado bons resultados nos projetos de seqüenciamento de metagenoma, basicamente em função das sequências mais longas obtidas (SCHLÜTER et al., 2008; JAENICKE et al., 2011; RADEMACHER et al., 2011; HESS et al., 2011; ZAKRZEWSKI et al., 2012).

Após a obtenção dos dados brutos do seqüenciamento, vários passos são necessários para a interpretação do conjunto de dados metagenômicos, incluindo a predição da origem taxonômica, montagem, predição de genes, e anotação. No entanto, a metagenômica não pode ser vista como uma simples extensão de análises de genoma, ao contrário, essa análise requer ferramentas de bioinformática desenhadas especificamente para essa abordagem (DESAI et al., 2012).

A análise do conjunto de dados metagenômicos se inicia na tentativa de determinar a origem taxonômica de cada sequência metagenômica. Nessa análise, cada sequência metagenômica é considerada como um identificador exclusivo de um membro da comunidade microbiana permitindo uma profunda avaliação da composição taxonômica.

Atualmente, duas abordagens são utilizadas na identificação da origem taxonômica das sequências alvo: uma baseada na similaridade e outra baseada na composição. A abordagem baseada na similaridade explora algoritmos de alinhamento, tais como BLAST (ALTSCHUL; GISH; MILLER, 1990), BLAT (KENT, 2002), BOWTIE (LANGMEAD et al., 2009), na busca de sequências similares em bancos de dados referenciais e direciona a classificação das sequências metagenômicas (MANDE; MOHAMMED; GHOSH, 2012).

As principais plataformas de análises de conjunto de dados metagenômicos que geram um perfil taxonômico e funcional baseado na similaridade são o MEGAN (HUSON et al., 2007), Carma3 (GERLACH; STOYE, 2011; KRAUSE et al., 2008) e MG-RAST (MEYER et

al., 2008). No entanto, essas plataformas diferem quanto aos algoritmos utilizados para a classificação das sequências, assim como sua base de dados de referência. O Carma3 identifica as sequências ambientais baseado em semelhanças com as famílias de proteínas e domínios depositados no bando de dados de família de proteínas (Pfam), enquanto que o MEGAN classifica as sequências a partir de comparações contra as sequências não redundantes da base de dados de nucleotídeos e de proteínas do NCBI. Já no MG-RAST, a reconstrução filogenética é baseada na detecção e classificação de sequências de genes RNAr 16S e na base de dados SEED.

No entanto, existem várias limitações para essas abordagens: a base de dados de referência é atualmente incompleta e altamente tendenciosa para apenas três filos de bactérias (Proteobacteria, Firmicutes e Actinobacteria), dentre pelo menos 50 filos; são analisados fragmentos de DNA em torno de 600 pb, são produzidos alinhamentos incompletos; os genes conservados filogeneticamente são apenas uma fração do total dos dados metagenômicos (JAENICKE et al., 2011; RADEMACHER et al., 2011; HESS et al., 2011; ZAKRZEWSKI et al., 2012).

Métodos de classificação baseados em composição analisam as características intrínsecas, tais como o conteúdo GC, uso de códons, frequências de oligonucleotídeos, e comparam essas características com as sequências de genoma de origem conhecidas (SIMON; DANIEL, 2011). Esse método é mais rápido, pois, requer um esforço computacional menor comparado ao método baseado em alinhamentos. Por outro lado, esse método necessita de sequências metagenômicas de tamanho suficiente para essas análises (MANDE; MOHAMMED; GHOSH, 2012). Ferramentas como PhyloPythia (McHARDY et al., 2007) e TACOA (DIAZ et al., 2009) utilizam essa abordagem.

Para a determinação do perfil funcional do conjunto de dados metagenômicos, a maioria das plataformas integra programas para a predição funcional das sequências metagenômicas usando de múltiplas ferramentas (múltiplas variâncias de BLAST) e bases de dados, as quais são comumente empregadas em análises de genomas (PRAKASH; TAYLOR, 2012). Entre as bases de dados mais consultadas estão Gene Ontology (GO), *Clusters of Orthologous Group* (COG), Pfam, NCBI, SwissProt, SEED e Kyoto Encyclopedia of Genes and Genomes (KEGG).

Paralelamente às análises das sequências metagenômicas, a montagem das sequências com regiões sobrepostas em sequências contíguas, denominadas *contigs*, pode resultar em um bom desenho ou até mesmo genomas completos da espécie alvo, mas isso

requer uma quantidade substancial de sequenciamento. Por exemplo, um sequenciamento massivo permitiu a Pelletier e colaboradores (2008) obterem um bom desenho do genoma de "*Candidatus Cloacamonas acidaminovorans*" a partir de uma biblioteca metagenômica construída de amostra de um digestor anaeróbico. Por outro lado, a montagem de *contigs* também tem o risco de gerar quimeras, principalmente em ambientes onde há várias espécies filogeneticamente relacionadas entre si ou regiões altamente conservadas entre as diferentes espécies (MANDE; MOHAMMED; GHOSH, 2012).

Há uma variedade de programas desenhados para a montagem de *contigs* a partir de conjunto de dados metagenômicos: MAP (LAI et al., 2012), Genovo (LASERSON; JOJIC; KOLLER, 2011), GS De Novo Assembler (Roche Science), MetaVelvet (NAMIKI et al., 2012), dentre outros. Da mesma forma, programas de predição de genes aplicados a metagenoma também foram desenhados, pois a maioria dos *contigs* obtidos é curta. Entre os programas disponíveis estão MetaGene (NOGUCHI; PARK; TAKAGI, 2006), Orphelia (HOFF et al., 2009), e FragGeneScan (RHO; TANG; YE, 2010). No entanto, cada um aplica uma abordagem diferente na predição de genes.

3. OBJETIVOS

3.1 Objetivos Gerais

- ✓ Caracterizar a diversidade filogenética e funcional de genes de FeFe-hidrogenase e de NiFe-hidrogenase da comunidade microbiana residente em um reator anaeróbico utilizado no tratamento de resíduos municipais visando melhorar a compreensão da microbiota envolvida na digestão anaeróbica e avaliar o ambiente como potencial reservatório de novos genes de FeFe-hidrogenase;
- ✓ Construir uma biblioteca metagenômica de fosmídeos a partir de amostra de um reator anaeróbico utilizado no tratamento de resíduos municipais para a triagem de clones na busca de novos genes de FeFe-hidrogenase e enzimas de aplicação biotecnológica.

3.2 Objetivos Específicos

- ✓ Caracterizar a diversidade filogenética de bactérias e arqueias em um reator de fluxo ascendente (RAFA) no tratamento de resíduos municipais;
- ✓ Caracterizar a diversidade dos genes funcionais FeFe-hidrogenase e NiFe-hidrogenase presentes na comunidade microbiana do RAFA;
- ✓ Construir uma biblioteca metagenômica de fosmídeos a partir de DNA obtido de amostras do RAFA;
- ✓ Selecionar clones da biblioteca metagenômica contendo genes de FeFe-hidrogenase utilizando ensaios baseados em PCR;
- ✓ Sequenciar os clones positivos para FeFe-hidrogenase obtidos da triagem por PCR para a descrição da sequência completa do gene alvo.

4. MATERIAL E MÉTODOS

4.1 Amostragem

Amostras de lodo foram coletadas de um reator anaeróbico de fluxo ascendente (RAFA) instalado na Estação de Tratamento de Esgoto da Bacia do Piçarrão, pertencente à companhia SANASA (Sociedade de Abastecimento de Água e Saneamento S.A.), localizada no município de Campinas. Frascos esterilizados foram utilizados no armazenamento de 200 gramas de amostras retiradas de três diferentes alturas de um RAFA. As amostras foram transportadas em gelo, imediatamente após a coleta no intuito de preservar os micro-organismos e diminuir a produção de biogás, para o laboratório da Divisão de Recursos Microbianos do Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas (CPQBA/UNICAMP) para serem processadas.

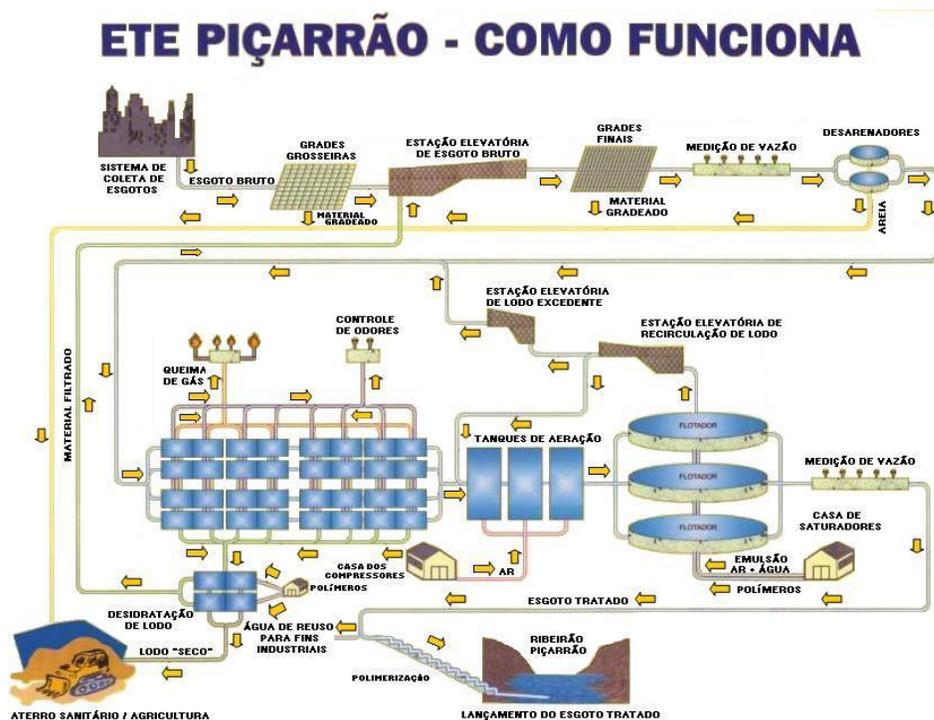


Figura 4. Esquema do funcionamento da Estação de Tratamento de Esgoto da Bacia do Piçarrão pertencente à SANASA (SANASA, 2013).

4.2 Extração de DNA

4.2.1 Extração de DNA por método direto

A partir das amostras coletadas do RAFA, cinco protocolos de extração de DNA por método direto foram avaliados. O primeiro método utilizado foi resultante da combinação de dois protocolos baseados em lise enzimática e mecânica (NERIA-GONZÁLES et al., 2006). Uma alíquota de 1,5 mL de amostra foi centrifugada em microtubo a 10.000 rpm durante 5 minutos. O sedimento foi suspenso em 1 mL de tampão TE (Tris-HCl 10 mM pH 8.0, EDTA 1 mM) e submetido a agitação vigorosa. Em seguida, foram adicionados 100 µL de lisozima (100 mg/mL) à solução e o tubo foi mantido em banho-maria a 37° C, durante 30 minutos. Após esse período, foram adicionados 8 µL de proteinase K (10 mg/mL) e 200 µL de SDS (10%), e essa nova solução foi mantida em banho-maria a 60° C por 30 minutos. A lise mecânica foi realizada por 3 ciclos de “freeze-thaw” seguida de uma agitação vigorosa durante 1 minuto. Igual volume de fenol saturado (Tris-HCl, pH 8.0) foi adicionado à solução, seguido da centrifugação a 10.000 rpm por 5 minutos. A fase aquosa obtida foi transferida para um novo microtubo, e o adicionou-se igual volume de uma solução clorofórmio:álcool isoamílico (24:1). Essa solução foi centrifugada a 10.000 rpm por 10 minutos. A fase superior foi transferida para um novo microtubo para precipitação do DNA, a qual foi realizada pela adição de 10% v/v de NaCl (5 M) e 2 volumes de etanol absoluto gelado. O sedimento obtido da centrifugação desta última solução foi lavado com 800 µL de etanol 70% gelado e centrifugado novamente. O sobrenadante foi descartado e o sedimento foi suspenso em 30 µL TE e estocado em freezer a – 20° C.

O segundo método foi realizado de acordo com o protocolo descrito por Fiantt (2000), no qual o sedimento obtido de 5 mL da amostra previamente centrifugada a 10.000 rpm por 5 minutos foi suspenso em 10 mL de SDS (1%) e adicionado de 20 µL de proteinase K (10 mg/mL). Essa solução foi somente homogeneizada e incubada a 65° C em banho-maria durante 1 hora. Após esse período, a solução foi resfriada em temperatura ambiente, em seguida, foram adicionados 2 µg de RNase por mL de solução. Após a incubação dessa solução em temperatura ambiente durante 30 minutos, a solução foi centrifugada e o sobrenadante obtido transferido para um novo microtubo. A precipitação e a suspensão do DNA foram realizadas como descrito anteriormente.

O método descrito por Stralis-Pavese e colaboradores (2004) consiste em homogeneizar o sedimento da amostra em 780 µL de tampão de lise (100 mM NaPO₄, 1% CTAB, 0,1M NaCl, 50mM, EDTA pH 8) e 40 µL de lisozima (100 mg/mL). Essa solução foi incubada em banho-maria a 37° C durante 30 minutos. Após esse período, foram adicionados 120 µL de tampão TE e pérolas de vidros à solução. A nova solução foi agitada vigorosamente durante 1 minuto. O sobrenadante obtido da centrifugação dessa última solução foi transferido para um novo microtubo e adicionou-se 50 µL de proteinase K (10 mg/mL). Essa solução foi incubada a 65° C durante 30 minutos. Em seguida, foram adicionados volumes iguais de clorofórmio:álcool isoamílico e esta nova a solução foi centrifugada. A fase aquosa desta última solução foi transferida para um novo microtubo. A precipitação e recuperação do DNA foram realizadas como previamente descrito.

O quarto método empregado foi baseado no protocolo descrito por Zhou e colaboradores (1996), com modificações. Volumes de 2 a 5 mL da amostra foram centrifugados e o sedimento foi suspenso em 3 mL de tampão de extração (100 mM NaPO₄, 1% CTAB, 0,1 M NaCl, 50 mM EDTA pH 8) juntamente com 45 µL de proteinase K. Essa solução foi incubada a 37° C durante 30 minutos sob uma agitação de 225 rpm. Após esse período, foram adicionados 600 µL de TE e 900 µL de SDS (10%) à solução, a qual foi mantida em banho-maria a 65° C durante 1 hora. Em seguida, a solução foi centrifugada e o sobrenadante foi transferido para um novo microtubo e volumes iguais de clorofórmio-álcool isoamílico foram adicionados. A solução aquosa obtida da centrifugação desta solução foi transferida para um novo microtubo. A precipitação do DNA e a sua recuperação nessa extração foram realizadas conforme descrito anteriormente.

Como método de extração comercial foi utilizado o PowerSoil® DNA Isolation Kit (MO BIO Laboratories), e a extração do DNA foi realizada de acordo com as instruções do fabricante.

4.2.2 Extração de DNA por método indireto

O primeiro método avaliado foi baseado no protocolo descrito por Hardeman e Sjöling (2007). Nessa metodologia, em um Erlenmeyer 100 gramas da amostra foram homogeneizados com 100 mL de tampão CTAB (50 mM Tris, 20 mM EDTA, 150 mM NaCl, 27 mM CTAB) e a solução foi mantida sob agitação branda durante 2 horas. Em seguida, essa mistura foi centrifugada em baixa rotação durante 10 minutos. O sobrenadante obtido foi

transferido para um novo tubo, o qual foi centrifugado para a recuperação das células. O sedimento celular foi suspenso em TE e em uma alíquota de 750 µL dessa solução foram adicionados 30 µL de SDS (10%) e 20 µL de proteinase K. Essa solução foi incubada em banho-maria a 37° C durante 1 h. Após esse período, foram adicionados 100 µL de NaCl (5M) e 80 µL de CTAB/NaCl (10% CTAB, 0,7 M NaCl) a esta solução e subsequentemente a solução foi incubada a 65° C por 10 minutos. Em seguida, a solução foi centrifugada e o sobrenadante transferido para um novo tubo para a recuperação do DNA como previamente descrito no item 4.2.1.

O segundo método de extração de lise indireta avaliado foi realizado conforme descrito por Liles e colaboradores (2008), com modificações. Nesse método, 100 g da amostra foi suspensa em 100 mL de TE (Tris-HCl 10 mM pH 8.0; EDTA 1 mM) e incubada a 4° C durante 16 h sob leve agitação. Em seguida, essa mistura foi filtrada com o auxílio de papel de filtro para a remoção de partículas grandes. A solução obtida foi centrifugada a 12.000 rpm durante 10 min e o sedimento obtido foi suspenso em TE. Para cada volume de sedimento suspenso em TE foi adicionado o mesmo volume de agarose de baixo ponto de fusão (*low-melting*) 0,5% para a construção dos *plugs* (DR II Sample Mold). Esses *plugs* foram adicionados à solução I (100 mM EDTA, 50 mM NaCl, 10 mM Tris, 1% lauril sarcosina e 1 mg/mL de lisozima) e mantidos durante 1 h a 37° C. Em seguida, os *plugs* foram transferidos para a solução II (2 mg/ml de proteinase K, 1% lauril sarcosina, 0,5% EDTA) e incubados a 50° C durante 24 h sob leve agitação. Durante esse período essa solução foi trocada por uma nova solução II. Após o período total de incubação, os *plugs* foram armazenados em geladeira em uma solução de TE.

4.3 Purificação da solução de DNA

Dois métodos de purificação de DNA foram avaliados para cada solução de DNA obtida das diferentes extrações por método direto. O método de purificação de cloreto de cério foi realizado pela adição de 100 mg desse sal a cada 100 µL de solução de DNA. A solução foi incubada durante 1 h e 30 minutos em temperatura ambiente. Em seguida, a solução foi centrifugada a 12.000 rpm durante 20 minutos e o sobrenadante transferido para um novo microtubo. A essa solução foram adicionados 400 µL de água Milli-Q e 300 µL de isopropanol, e esta foi mantida por 5 minutos em temperatura ambiente. Após esse período, a

solução foi centrifugada (sob as mesmas condições já mencionadas) e o precipitado obtido foi suspenso em 100 μ L TE. Essa solução foi transferida para um novo microtubo e foram adicionados 20 μ L de acetato de potássio (8 M). Após 15 minutos de incubação em temperatura ambiente, essa solução foi centrifugada e o sobrenadante transferido para um novo microtubo. A precipitação do DNA foi realizada com 2 volumes de etanol absoluto e o sedimento obtido suspenso TE.

O segundo método de purificação foi realizado com o Wizard[®] DNA Clean-Up System, de acordo com o manual de instruções fornecido pelo fabricante.

4.4 Construção das bibliotecas de genes RNAr 16S

Os genes RNAr 16S do domínio Bacteria foram amplificados por PCR em reações contendo tampão 1X, 0,2 μ M de uma mistura de dNTPs, 0,4 μ M de cada *primer* (Tabela 2), 1,5 mM de MgCl₂, 4 U de Taq DNA Polimerase (Invitrogen, Brasil) e cerca de 50 ng de DNA purificado a partir do RAFA, para um volume final de 50 μ L. A reação foi incubada no termociclador modelo Gene Amp PCR System 9700 (Applied Biosystems), para a amplificação do gene, de acordo com o seguinte programa: um ciclo de desnaturação inicial de 94^o C por 5 min; 30 ciclos de 94^o C por 1 min, 55^o C por 1 min e 72^o C por 3 min; um ciclo de extensão final a 72^o C por 5 min. Para amplificação do DNAr 16S de organismos do domínio Archaea, foram testados os pares de *primers* ARCH46F / ARCH915R (Tabela 2), segundo metodologia descrita por Yoshida e colaboradores (2005). As reações de amplificação foram realizadas nas mesmas proporções conforme descrito acima, no entanto, contendo *primers* para o gene RNAr 16S de arqueia. A amplificação no termociclador foi realizada da seguinte maneira: um ciclo de desnaturação inicial de 94^o C por 2 min., 10 ciclos de 94^o C por 30 segundos, 65^o C por 30 segundos, 72^o C por 2 min; 25 ciclos de 94^o C por 30 segundos, 60^o C por 30 segundos e 72^o C por 2 min.

Os produtos de PCR foram submetidos à eletroforese em gel de agarose (1%) contendo brometo de etídio (0,5 μ g/mL) e visualizados e fotografados utilizando o equipamento UVP GDS-8000. Em seguida, os fragmentos obtidos foram purificados utilizando *GFX 96 PCR and Band Purification Kit* (GE Healthcare), de acordo com as instruções do fabricante.

As soluções purificadas contendo os fragmentos amplificados dos genes RNAr 16S de Bacteria e Archaea foram quantificadas em gel de agarose (1%) e uma alíquota de cada

solução foi utilizada para a clonagem no sistema pGEM-T Easy Vector (Promega)/*Escherichia coli* DH10B por eletroporação.

Tabela 2. Sequências dos *primers* utilizados para amplificação do gene RNAr 16S de bactérias e arqueias.

Primers	Sequências 5'- 3'	Referências
27F	AGAGTTTGATCMTGGCTCAG	LANE, 1991
1100r	CGGTGTGTACAAGGCCCGGGAACG	LANE, 1991
ARCH46F	YTAAGCCATGCRACT	YOSHIDA et al., 2005
ARCH915R	GTGCTCCCCCGCCAATTCCT	STAHL; AMANN,1991

R é A ou G; **Y** é C ou T e **M** é A ou C.

A reação de clonagem dos fragmentos amplificados foi realizada da seguinte maneira: 5 µL de tampão de ligação, 1 µL de T4 DNA ligase, inserto e vetor pGEM-T na proporção de 6:1, e água ultra pura estéril para um volume final de 10 µL. Após a incubação dessas soluções a 16º C durante 12 h, uma alíquota de 1 µL de cada reação de ligação foi homogeneizada com 40 µL de células competentes, previamente preparadas e armazenadas a – 80º C, de acordo com o protocolo descrito por Kasap e colaboradores (2008). Cada solução, contendo o vetor ligado aos fragmentos amplificados e células competentes, foi incubada durante 5 minutos no gelo, e posteriormente transferida para cuvetas de 0,2 cm às quais foi aplicada uma voltagem de 2,5 kV. As células foram recuperadas da cuveta com 500 µL de LB e transferidas para microtubos. Após a incubação a 37º C durante uma 1 h, a solução foi plaqueada com o auxílio de uma alça de Drigalsky em placas de Luria Bertani ágar acrescidas de ampicilina (100 µg/mL), 100 µL de IPTG (0,1 M) e 20 µL de X-gal (50 mg/mL).

Após a incubação das placas a 37º C durante 24 h, a seleção dos clones recombinantes foi realizada pela resistência a ampicilina e identificação das colônias brancas, pois a inserção dos fragmentos no vetor de clonagem inativa o sistema β-galactosidase tornando as células brancas. Em seguida, as colônias brancas foram transferidas com auxílio de palitos estéreis para microplacas contendo LB, ampicilina (100 µg/mL) e glicerol (10%) para serem armazenadas a - 80º C até o início do sequenciamento.

4.5 Construção das bibliotecas gênicas de hidrogenases

Diferentes pares de *primers* degenerados (Tabela 3) foram utilizados nas reações de PCR para amplificação dos genes de FeFe-hidrogenase e NiFe-hidrogenase. Reações de amplificação para cada biblioteca gênica foram preparadas da seguinte maneira: tampão 1 X, 0,2 µM de uma mistura de dNTPs, 0,4 µM de cada *primer* degenerado, 1,5 mM de MgCl₂, 4 U de Taq DNA Polimerase (Invitrogen, Brasil) e 50 ng de DNA purificado, para um volume final de 50 µL. As reações foram incubadas no termociclador modelo Gene Amp PCR System 9700 (Applied Biosystems), para a amplificação dos fragmentos dos genes de FeFe-hidrogenase e NiFe-hidrogenase, de acordo com os programas descritos por Xing e colaboradores (2008) e Kim e colaboradores (2007), respectivamente.

As condições para amplificação da NiFe-hidrogenase utilizando os pares de *primers* desenhados neste estudo foram baseadas nas avaliações obtidas pelo software OLIGO ANALYZER. Um ciclo de desnaturação inicial de 94° C por 5 min; 10 ciclos de 94° C por 30 segundos, 52° C por 30 segundos, 72° C por 2 min; 20 ciclos de 94° C por 30 segundos, 55° C por 30 segundos e um ciclo de extensão a 72° C por 2 min.

O DNA genômico de *Desulfovibrio vulgaris* e *Citrobacter freundii* CBMAI 1187 foi utilizado como controle positivo nas reações de amplificação para FeFe-hidrogenase e NiFe-hidrogenase, respectivamente. As condições de amplificação foram realizadas de acordo com os autores destes *primers* (XING; REN; RITTMANN, 2008).

Os fragmentos dos diferentes genes amplificados foram clonados no sistema pGEM-T Easy Vector (Promega)/*Escherichia coli* DH10B por eletroporação, como descrito no item 4.4.

Tabela 3. Sequências dos *primers* degenerados para amplificação dos genes de hidrogenases.

<i>Primers</i>	Sequências 5'- 3'	Gene	Referências
hydF1	GCCGACCTKACMATMATGGA	FeFe-H ₂ ase	XING; REN; RITTMANN, 2008
hydR1	ATRCARCCRCCSGGRCAGGCCAT	FeFe- H ₂ ase	XING; REN; RITTMANN, 2008
NiFeF	ATGCGXCGXAGYTTYCXARTAYTGYAGY	NiFe- H ₂ ase	KIM;JUNG; CHA, 2007
NiFeR	RTGXGTXGARCAXAGRCAXGGRTC	NiFe- H ₂ ase	KIM;JUNG; CHA, 2007

K é G ou T, **R** é A ou G, **Y** é C ou T, **S** é C ou G e **M** é A ou C.

4.6 Desenho de *primers* de NiFe-hidrogenase

Para o desenho de *primers* de NiFe-hidrogenase, as sequências de aminoácidos relacionadas ao grupo I e II desse gene foram recuperadas do banco de dados GenBank, utilizando o repositório *Protein*. Em seguida, alinhamentos múltiplos das sequências de aminoácidos foram realizados usando o software Clustal W v. 2.0 (LARKIN et al., 2007) e as análises filogenéticas empregando o programa MEGA v. 5.04 (TAMURA et al., 2011). As distâncias evolutivas foram calculadas usando o modelo de Poisson, como implementado no programa MEGA v 5.04. A reconstrução filogenética foi realizada usando o algoritmo de *neighbour-joining* (NJ) (SAITOU; NEI, 1987), com valores de *bootstrap* calculados a partir de 1000 replicatas. Esta etapa teve a finalidade de evitar o alinhamento de sequências evolutivamente distantes, o que dificultaria a identificação de regiões relativamente conservadas.

Posteriormente, a sequências de nucleotídeos foram deduzidas a partir das respectivas sequências de aminoácidos utilizando o programa BioEdit (Ibis Biosciences). Essas sequências de nucleotídeos foram alinhadas com o auxílio do programa Clustal X v. 2.0 (LARKIN et al., 2007) para a identificação das regiões conservadas. Desta forma, os *primers* foram desenhados manualmente a partir dessas regiões conservadas.

Os pares de *primers* desenhados (Tabela 4) a partir das regiões conservadas foram submetidos à análise pelo software OLIGO ANALYZER para a predição de suas propriedades termodinâmicas. Essas propriedades foram: temperatura de anelamento, possíveis interações entre os *primers* e formação de grampo (*hairpin*). Após essas avaliações, os pares de *primers* foram encaminhados para síntese pela empresa Invitrogen (SP, Brasil).

Tabela 4. Sequências dos *primers* degenerados para amplificação dos genes de NiFe-hidrogenases.

<i>Primers</i>	Sequências 5'- 3'	Pares de bases (bp)
NiFeF1	CGTATTTGCCGGCRTKTGCM	1120
NiFeR1	GGTVGCCCCGGGTRRAS	
NiFeF2	KGARRTGGGCCCGVTG	587
NiFeR2	TVCGTGCCTAGTTTCGRTRC	

K é G ou T, **R** é A ou G, **S** é C ou G, **V** é A, C ou G e **M** é A ou C.

4.7 Sequenciamento das bibliotecas gênicas

Para o sequenciamento dos clones das bibliotecas gênicas, os clones foram cultivados individualmente em LB contendo ampicilina e uma alíquota dessa cultura foi utilizada para a amplificação do fragmento do gene de interesse na seguinte reação: tampão 1 X, 0,2 μ M de uma mistura de dNTPs, 0,4 μ M de *primers* universais M13F (5'-CGCCAGGGTTTTCCCAGTCACGAC-3') e M13R (5'-TTTCACACAGGAAAC AGCTATGAC-3'), 1,5 mM de MgCl₂, 4 U de Taq DNA Polimerase (Invitrogen, Brasil) e 4 μ L da cultura, para um volume final de 50 μ L.

Os fragmentos amplificados foram purificados com *GFX 96 PCR and Gel Band Purification Kit* (GE HealthCare) e aproximadamente 100 ng da solução purificada foram adicionados de 10 pmol de *primers*, 4 μ L de DYEnamic ET Terminator Cycle Sequencing Kit for MegaBace e água ultra pura estéril para a reação de sequenciamento. A reação foi incubada no termociclador modelo Gene Amp PCR System 9700 (Applied Biosystems) para 30 ciclos consecutivos de 15 s a 96^o C, 15 s a 50^o C e 4 min a 60^o C.

4.8 Análises estatísticas

As sequências parciais obtidas de cada fragmento sequenciado foram alinhadas para a obtenção de uma sequência consenso única, denominada *contig*, com o auxílio do programa phred/Phrap/CONSED (EWING et al., 1998). Posteriormente, todas as sequências de cada biblioteca gênica de RNAr 16S foram alinhadas usando o programa CLUSTAL X (LARKIN et al., 2007) e o arquivo contendo o alinhamento foi salvo em formato .phy. A partir desse arquivo o programa DNADIST, o qual compõe o pacote de ferramentas Phylip para análise filogenética (FELSENSTEIN, 1989), gerou as matrizes de distância.

Em seguida, o arquivo contendo as matrizes serviu de entrada no programa DOTUR (do inglês, *Distance based OTU Richness*) (SCHLOSS; HANDELSMAN, 2005). Baseado na similaridade entre as sequências, o programa agrupa as sequências em diferentes níveis de Unidade Taxonômica Operacional (OTU) (3, 5 e 20%). Adicionalmente, o programa calcula baseado no número de OTUs observado, os índices de diversidade de Shannon (H') e de Simpson, os índices não paramétricos de estimativa de riqueza CHAO e ACE, e curvas de rarefação, de acordo com os níveis filogenéticos desejados.

Para a determinação das OTUs dos genes funcionais, sequências de nucleotídeos foram traduzidas *in silico* para aminoácidos pelo programa Bioedit, alinhadas pelo programa CLUSTAL W (LARKIN et al., 2007) e os próximos passos foram executados de acordo como descrito acima.

Sequências representativas das distintas OTUs determinadas nesse estudo foram depositadas no banco de dados GenBank com os números de acesso JX843902 a JX844019 e JX875075 a JX875092 para as sequências afiliadas a Bacteria, JX548278 a JX548290 para as sequências atribuídas a Archaea e JX843877, JX843878, JX843880 a JX843886 e JX843888 a JX843901 para as sequências putativas de *hydA*.

4.9 Análise filogenética

Uma sequência representante de cada OTU definida nas bibliotecas de RNAr 16S pelo DOTUR foi selecionada para a reconstrução da árvore filogenética. Em seguida, a ferramenta BLAST buscou nas bases de dados RDP (Ribosomal Database Project) e Genbank por sequências similares às sequências de OTUs. As sequências que apresentavam alinhamento com *E-value* acima de 10^{-10} foram recuperadas para direcionar a classificação filogenética.

Um arquivo contendo as sequências de OTUs e as sequências similares recuperadas das bases de dados foi submetido ao alinhamento com o auxílio do programa CLUSTAL X (LARKIN et al., 2007) e analisadas com o software MEGA v 5.04. As distâncias evolutivas foram calculadas usando o modelo de substituição de DNA de Kimura (1980), como implementado no programa MEGA. A reconstrução filogenética foi realizada usando o algoritmo de *neighbour-joining* (NJ) (SAITOU; NEI, 1987), com valores de *bootstrap* calculados a partir de 1000 replicatas.

Para a análise de genes funcionais, as sequências de OTUs foram submetidas à busca por sequências similares no GenBank (banco de dados de aminoácidos) utilizando a ferramenta BLASTx. O alinhamento das múltiplas sequências de aminoácidos das OTUs e das sequências recuperadas do GenBank foi realizado pelo programa CLUSTAL W (LARKIN et al., 2007) e analisadas pelo programa MEGA v 5.04. A reconstrução da árvore filogenética utilizou o algoritmo de *neighbour-joining* (NJ) e modelo de substituição de resíduos de aminoácidos Poisson com valor de *bootstrap* de 1.000 repetições.

4.10 Construção da biblioteca metagenômica de fosmídeo

Para a construção da biblioteca metagenômica em fosmídeo, foram avaliados diferentes protocolos de extração de DNA, conforme descrito no item 4.2, para a obtenção de fragmentos de DNA de tamanho médio (30-40 Kb), assim como métodos de purificação de DNA, descritos no item 4.3.

A biblioteca metagenômica foi construída utilizando o “CopyControl™ HTP Fosmid Library Production Kit” de acordo com as instruções do fabricante. Fragmentos de DNA entre 25 Kb a 50 Kb foram excisados de um gel de agarose de baixo ponto de fusão 1% e transferidos para microtubos, os quais foram mantidos em banho-maria a 45° C para a fusão da agarose. Em seguida, foi adicionada a enzima GELase® (1 U/ 100 µL de gel) e o tampão dessa enzima (previamente aquecido a 45°C) para concentração final da reação de 1X. Esta reação foi incubada a 45°C durante 1 h, seguido de um choque a 70°C por 10 min para inativação da enzima. Alíquotas de 500 µL dessa última solução foram transferidas para novos microtubos, incubadas no gelo por 5 min e centrifugadas por 20 min a 12.000 rpm. O sobrenadante recuperado foi homogeneizado com 1/10 volume de acetato de sódio 3 M pH 7,0 e 2,5 volumes de etanol absoluto e mantido a 4° C durante 24 h. Após esse período, a solução foi centrifugada por 20 min a 12.000 rpm e o precipitado foi lavado com etanol 70% gelado e cuidadosamente dissolvido em água Milli-Q.

Uma alíquota da solução de DNA foi utilizada para a estimativa da concentração de DNA em gel de agarose 0,8% e cerca de 300 ng do DNA foram utilizados para a reação de ligação ao vetor pCC2Fos. Em seguida, essa solução foi incubada com 50 µL de “*MaxPlax Lambda Packaging Extracts*” para a reação de empacotamento dos fosmídeos. Após esse passo, 100 µL de uma cultura de *Escherichia coli* EPI300 T1^R, previamente preparada para a transfecção, foram incubados com 10 µL de partículas de fago diluídas a 37° C por 20 min. A seleção dos clones recombinantes foi realizada através do plaqueamento da solução em meio LA contendo cloranfenicol (12,5 µg/mL) e incubação a 37° C por 24 h. Os clones recombinantes foram individualmente cultivados em placas de 96 poços contendo LB suplementado com cloranfenicol 12,5 µg/mL a 37° C e 150 rpm durante 12 h. Após esse período, foram adicionados a cada poço glicerol para uma concentração final de 10% e as placas estocadas em freezer a -20° C.

A validação da biblioteca, ou seja, a confirmação da clonagem de diferentes insertos, foi realizada através da seleção aleatória de clones e cultivo dos mesmos em LB (contendo cloranfenicol 12,5 µg/mL) durante 16 h a 37° C sob 200 rpm. O DNA fosmidial foi extraído pelo método de lise alcalina e, em seguida, reações de clivagem para cada clone recombinante com a enzima de restrição *Xba*I (New England Biolabs) foram preparadas. As reações de clivagem foram incubadas a 37° C durante 16 h, com subsequente incubação a 65° C por 10 min para a inativação enzimática. Ao final dessas etapas, as reações foram submetidas à eletroforese em campo pulsado (11 h, 1s 12 s, 120°, 6 Vcm⁻¹) a fim de confirmar os perfis polimórficos dos clones.

4.11 Triagem das sequências de FeFe-hidrogenase na biblioteca metagenômica

Para triagem dos clones contendo as sequências de FeFe-hidrogenase na biblioteca de fosmídeos, reações de PCR em *pools* foram realizadas diretamente da cultura de clones. Cada *pool* foi constituído de 10 clones e uma alíquota de 5 µL da cultura de cada um dos clones estocado foi retirada para a composição desse *pool*. As reações de PCR foram preparadas da seguinte maneira: tampão 1 X, 0,2 µM de uma mistura de dNTPs, 0,4 µM de cada *primer* degenerado para FeFe-hidrogenase, 1,5 mM de MgCl₂, 4 U de Taq DNA Polimerase (Invitrogen, Brasil) e 3 µL de um *pool* de clones, para um volume final de 50 µL. As condições para amplificação dos fragmentos desse gene foi realizada conforme descrito no item 4.5.

Os produtos de amplificação das reações foram submetidos à eletroforese em um gel de agarose 1% e visualizados com o auxílio do Luminescent Image Analyzer Modelo Image Quant LAS 4000 (GE).

4.12 Extração de DNA fosmidial e triagem por PCR

Para a extração do DNA dos fosmídeos, cada clone foi cultivado em 1 mL de meio LB, suplementado com 12,5 µg/mL de cloranfenicol e 0,01% de arabinose (a fim de induzir o maior número de cópias do vetor por células), e incubado a 37° C durante 16 h sob agitação de 180 rpm. Uma alíquota de 0,5 mL de cada clone cultivado foi coletada para a obtenção do *pool*

com volume final de 480 mL, ou seja, cada *pool* foi constituído de 960 clones. Foram preparados 10 *pools*, totalizando a extração de 9.600 clones.

O DNA fosmidial dos *pools* foi extraído utilizando o QIAGEN *Large-Construct kit* (QIAGEN, Hilden), segundo especificações do fabricante. O DNA fosmidial de cada *pool* foi utilizado como molde em reações de PCR independentes para amplificação do gene de FeFe-hidrogenase. As reações de PCR e as condições para a amplificação foram realizadas conforme descrito no item 4.5.

4.13 Pirosequenciamento

O DNA fosmidial extraído de cada *pool* foi quantificado com o auxílio de PicoGreen®, reagente fluorescente ultra sensível com alta afinidade por DNA de dupla fita, para em seguida adicionar concentrações equivalentes dos diferentes *pools* para a composição de uma única amostra contendo 1 µg de DNA fosmidial.

O pirosequenciamento foi realizado a partir de 9.600 clones da biblioteca metagenômica, a qual contém um total de 17.280 clones. A amostra de DNA fosmidial foi preparada segundo o protocolo do GS *Titanium General library Prep Kit* (Roche Applied Science) para o sequenciamento utilizando a GS *FLX Titanium platform*. A solução contendo 1 µg de DNA fosmidial foi submetida à nebulização para a geração de fragmentos de 500 a 800 pares de bases (pb), os quais foram purificados com o *Qiagen MinElute PCR Purification kit* (QIAGEN, Germany), de acordo com as instruções do fabricante. Os fragmentos foram tratados com *End Repair* para gerar extremidades abruptas necessárias para a ligação de adaptadores nas extremidades 3' e 5', a qual foi realizada pela RL Ligase®. Os adaptadores foram utilizados para a ligação dos fragmentos de DNA às microesferas. Cada fragmento de DNA único da amostra foi imobilizado em uma microesfera. As microesferas contendo DNA foram adicionadas à emulsão de óleo para a reação em cadeia da polimerase. Na emulsão, cada micela formada continha uma microesfera ligada a um único fragmento de DNA juntamente com reagentes de PCR, os quais formam um micro-reator.

Desta forma, no final da PCR em emulsão, cada microesfera continha várias cópias de fragmentos idênticos gerados a partir de uma única cópia inicial. As primeiras PCRs em emulsão foram realizadas com diferentes concentrações de DNA para determinar a melhor proporção de DNA por microesfera. A titulação das reações de PCR em emulsão foi realizada

com *GS Titanium SL emPCR Kit* (Roche), de acordo com as instruções do fabricante. Após essa titulação, uma nova PCR em emulsão foi realizada de acordo com a melhor concentração de DNA por microesfera. Em seguida, essa emulsão foi quebrada, as microesferas contendo DNA amplificado foram recuperadas e contadas.

Em torno de dois milhões de microesferas foram aplicadas na *PicoTiterPlate*, placa suporte para o sequenciamento. Em seguida, as microesferas contendo sulfilase, luciferase e microesferas específicas para a estabilidade da reação foram adicionadas em camadas sucessivas. O suporte *PicoTiterPlate*, onde cada poço dessa placa recebeu uma microesfera contendo os fragmentos amplificados, foi então carregado no aparelho GS FLX, o qual realiza a leitura por sinais de quimioluminescência, juntamente com os reagentes necessários para a reação de sequenciamento. Após o término do sequenciamento, as imagens geradas foram processadas pelo programa *GS Run Processor* (Roche).

4.14 Análise do conjunto de dados metagenômicos

O programa *GS de Novo Assembler* foi utilizado para remoção das sequências de baixa qualidade, ou seja, duplicadas, sequências menores que 100 pares de bases (pb), homopolímeros (repetição de 20 nucleotídeos ou dinucleotídeos) e filtragem das sequências da espécie hospedeira, neste caso foi utilizado o genoma da *Escherichia coli* DH5 α como referência, e das sequências que compõem o fosmídeo sem o inserto, utilizando 90% de similaridade para o valor de corte. Após a realização dessa triagem, o arquivo contendo as sequências metagenômicas foi utilizado para a montagem das sequências em regiões contíguas, denominadas *contigs*, gerando um segundo conjunto de dados. Os dados estatísticos, tais como o número e a média do tamanho das sequências metagenômicas e *contigs*, foram gerados a partir desse conjunto de dados filtrados. Desta forma, dois conjuntos de dados metagenômicos foram analisados em duas abordagens distintas. As sequências metagenômicas foram analisadas na plataforma MetaSAMS (ZAKRZEWSKI et al., 2012), enquanto que o conjunto de *contigs* foi analisado no programa GenDB 2.0 (MEYER et al., 2003).

4.14.1 Análises das sequências metagenômicas na plataforma MetaSAMS

A plataforma MetaSAMS foi utilizada para a anotação automática, e classificação taxonômica e funcional das sequências metagenômicas. Essa plataforma inclui três diferentes abordagens de classificação taxonômica: o RDP Classifier (WANG et al., 2007), o CARMA (GERLACH; STOYE, 2011; KRAUSE et al., 2008) e uma abordagem baseada no algoritmo de atribuição do menor ancestral comum (LCA – *Lowest Common Ancestor*), assim como aplicado no MEGAN (HUSON et al., 2007). O CARMA também foi utilizado na determinação do perfil funcional das sequências metagenômicas.

Os fragmentos de genes RNAr 16S foram identificados utilizando a ferramenta BLAST, o qual buscou no banco de dados Ribosomal Database Project (DRP) por sequências similares às sequências-alvo. Alinhamentos que apresentavam um *E-value* acima de 10^{-10} e sequências contendo no mínimo 50 pb foram extraídas e classificadas utilizando a ferramenta RDP Classifier, a qual é baseada em inferência Bayesiana simples. Somente os fragmentos de RNAr 16S com no mínimo 80% de confiança foram consideradas na análise. A segunda classificação taxonômica que integra a plataforma MetaSAMS utiliza uma abordagem semelhante à do MEGAN. Resumidamente, as sequências metagenômicas foram comparadas contra o banco de dados NCBI (nucleotídeos e aminoácidos) utilizando a ferramenta BLAST e *E-value* acima 10^{-10} . A partir desses resultados, foi direcionada a classificação das sequências metagenômicas baseado no menor ancestral comum.

O CARMA3 é o terceiro programa integrado ao MetaSAMS que realiza a classificação taxonômica e a caracterização funcional de sequências metagenômicas baseado no modelo oculto de Markov. Nessa abordagem, foram identificados os domínios conservados (Pfam) e as famílias de proteínas nas sequências metagenômicas e, posteriormente, foi construída a árvore filogenética de cada família Pfam correspondente a cada sequência identificada. Desta forma, as sequências metagenômicas são classificadas com base na correlação entre as proteínas identificadas e sua relação filogenética com as sequências de banco de dados onde estão determinadas a sua origem taxonômica.

Essa plataforma gerou ainda gráficos de conteúdo GC, assim como o cálculo da abundância relativa de cada táxon identificado de acordo com a abordagem taxonômica selecionada.

4.14.2 Identificação de sequências codificantes e anotação automática

O arquivo contendo os *contigs* foi carregado no programa GenDB 2.0 (MEYER et al., 2003) para a predição das sequências codificantes (CDS) e anotação automática. A plataforma Reganor (LINKE et al., 2004), a qual está integrada ao GenDB, combina e compara os resultados do Glimmer3 (DELCHER et al., 2007) e Critica (BADGER; OLSEN, 1999) para a predição das sequências codificantes. Para cada sequência codificante identificada, a plataforma Metanor, também integrada ao GenDB, determina a sua função genética baseada em pesquisa BLAST contra diferentes bancos de dados, tais como NCBI (banco de dados de nucleotídeos e de proteínas), COG, KEGG, Swiss-Prot, Pfam, TIGRFAM e InterPro. Em seguida, as sequências codificantes foram então classificadas funcionalmente de acordo com a sua categoria e número correspondente no COG (TATUSOV et al., 2003) e os números no GO (Gene Ontology). Adicionalmente, as sequências codificantes identificadas como hidrogenase foram analisadas manualmente e a sua origem taxonômica foi determinada utilizando os bancos de dados citados acima como referência.

4.15 **Desenhos de *primers* para as CDS de FeFe-hidrogenase**

Os *primers* para as sequências codificantes de FeFe-hidrogenase foram desenhados no programa FastPCR (KALENDAR; LEE; SCHULMAN, 2011). As CDS foram inseridas no programa e os *primers* foram desenhados automaticamente, assim como a predição de suas propriedades físico-químicas e termodinâmica. Essas propriedades foram: conteúdo de GC, temperatura de anelamento, formação de dímeros (*self dimer* e *cross dimer*) e de grampo (*hairpin*). Adicionalmente, o genoma da *Escherichia coli* DH5 α (genoma de referência) foi recuperado do banco de dados GenBank para que uma PCR *in silico* fosse realizada a fim de testar os *primers* desenhados. Os *primers* que apresentaram temperatura de anelamento acima de 30° C e não apresentavam amplificação *in silico* de regiões do genoma da célula hospedeira foram selecionados para a síntese. Desta forma, evitou-se a amplificação de hidrogenases que compõem o genoma da *E. coli*.

Após essas avaliações, os pares de *primers* (Tabela 5) foram encaminhados para síntese na empresa Invitrogen (SP, Brasil).

Tabela 5. Sequências dos *primers* para amplificação das CDS de FeFe-hidrogenase.

CDS/Primer	Sequências 5'- 3'	Pares de bases (bp)
CDS_959		1872
F_959	ATAACAACACCACCCAGAGC	
R_959	TTATGCCATATATTTACCGGT	
CDS_13737		1864
F_13737	GAAGACTGACCCGTTGTTGC	
R_13737	TTTTGAGGATGGGTTCCGGT	
CDS_16334		1720
F_16334	GTTGAATATAGAAATAGATAA	
R_16334	TTATACCTTAATTTCTCG	
CDS_7499		1579
F_7499	GTGTTACCTCTGCTTCAGA	
R_7499	TTAATATACACCTCTCTCTAC	
CDS_14436		1114
F_14436	GGAACGCAAAAAGGCAGA	
R_14436	TTAGAAGCTTGACTGTTCCATA	
CDS_12638		1084
F_12638	GGAAACGAATATCACCATAAC	
R_12638	CTAGTTTTTTGCGGTGCAAGG	

4.16 Triagem por PCR da biblioteca metagenômica para a identificação dos clones contendo as CDS de FeFe-hidrogenase

A triagem por PCR foi realizada em *pools* compostos de 10 clones cada um. Uma alíquota de 3 µL de cada cultura estoque de clone da biblioteca metagenômica foi coletada para a composição de cada *pool*. Adicionalmente, os clones foram cultivados individualmente (como descrito no item 4.12), em seguida, alíquotas foram recuperadas de cada cultivo para a formação de um *pool* e, subsequentemente utilizado na tiragem por PCR.

Para a amplificação das sequências codificantes de FeFe-hidrogenase, cada reação foi composta de tampão 1 X, 0,2 µM de uma mistura de dNTPs, 0,4 µM de cada *primer* desenhado (descrito no item 4.15), 1,5 mM de MgCl₂, 4 U de Taq DNA Polimerase (Invitrogen,

Brasil) e 3 μL de um *pool*, para um volume final de 50 μL . DNA metagenômico extraído do lodo do reator anaeróbico e usado para a construção da biblioteca de fosmídeos foi utilizado como controle positivo desta reação.

As reações foram incubadas no termociclador modelo Gene Amp PCR System 9700 (Applied Biosystems) de acordo com o seguinte programa: um ciclo de desnaturação inicial de 94° C por 5 min; 30 ciclos de 94° C por 1 min, 35° C por 1 min e 72° C por 3 min; um ciclo de extensão final a 72° C por 5 min.

Os produtos de amplificação das reações foram submetidos à eletroforese em um gel de agarose 1% e visualizados com o auxílio do Luminescent Image Analyzer Modelo Image Quant LAS 4000 (GE).

4.17 Sequenciamento dos fosmídeos contendo as CDS de FeFe-hidrogenase

Os clones nos quais foi detectado um produto de PCR com tamanho próximo ao número de pares de bases determinado *in silico* para cada sequência codificante de FeFe-hidrogenase foram cultivados sob as mesmas condições mencionadas no item 4.12. Em seguida, os fosmídeos foram extraídos individualmente a partir de cada clone positivo utilizando o QIAGEN Large-Construct kit (QIAGEN, Hilden), segundo especificações do fabricante. As amostras de DNA fosmidial foram encaminhadas para o sequenciamento na plataforma 454 GS FLX Titanium (Roche Applied Sciences), o qual será realizado pela empresa Macrogen (Seul, Coréia do Sul).

Após o sequenciamento dos fosmídeos, o programa *GS de Novo Assembler* será utilizado para a remoção dos fragmentos contaminantes do genoma de *E. coli* e montagem dos fosmídeos. Cada fosmídeo será analisado manualmente para a correta ordenação de suas sequências, e posteriormente anotado com a ferramenta GenDB 2.0 (MEYER et al., 2003).

5 RESULTADOS E DISCUSSÃO

5.1 Avaliação dos métodos de extração de DNA

Os métodos de extração de DNA metagenômico podem ser classificados em procedimento direto, quando o DNA é extraído diretamente da amostra, e indireto, quando as células recuperadas da matriz ambiental são imobilizadas em *plugs* de agarose seguida da lise celular e recuperação do DNA.

A primeira metodologia avaliada na extração de DNA por método direto foi baseada em lise enzimática e mecânica. Por esse método não se observou DNA em gel de agarose. No segundo método avaliado, descrito por Stralis-Pavese e colaboradores (2004), foi obtido um rendimento razoável de DNA devido à presença do detergente CTAB (brometo de cetiltrimetilamônio) no tampão de extração, uma vez que o CTAB forma um complexo insolúvel com ácidos nucleicos, os quais se precipitam e podem ser coletados por centrifugação. No entanto, uma eletroforese em campo pulsado revelou que o tamanho médio dos fragmentos de DNA obtidos não ultrapassou 1 Kb. Estes resultados inviabilizaram o uso desta metodologia para a construção da biblioteca de fosmídeos, uma vez que são necessários fragmentos de DNA acima de 25 Kb.

Baseado nos protocolos descritos por Zhou e colaboradores (1996) e Fiantt (2000) foi possível a obtenção de fragmentos de DNA em torno de 30 Kb (Figura 5). Porém, independente de quaisquer métodos, a solução final de DNA era turva, ou seja, ainda continha grande quantidade de matéria orgânica. Este fato justificou a avaliação de várias metodologias para a extração de DNA no intuito de diminuir a quantidade de ácidos húmicos presentes na solução final de DNA, pois sabe-se que essas substâncias afetam quase todos os métodos de biologia molecular, tais como hibridização, digestão por enzima de restrição e PCR (RAJENDHRAN; GUNASEKARAN, 2008).

Reações de PCR para a amplificação do gene RNAr 16S a partir das duas últimas extrações de DNA foram preparadas a fim de se avaliar a qualidade do DNA obtido. Porém, não foi observada a amplificação desse gene em gel de agarose, mesmo utilizando uma diluição de DNA 1:1000, confirmando que a presença de ácidos húmicos inibe a atividade enzimática da polimerase e, principalmente, demonstrando a inviabilidade de construção das bibliotecas.

Esses resultados confirmaram que amostras de lodos são muito diferentes de amostras de outros ambientes para os quais a maioria dos métodos de extração foi desenvolvida, pois a biomassa está tipicamente organizada em agregados celulares ou em flocos (McILROY et al., 2009). Adicionalmente aos protocolos descritos na literatura, o kit comercial de extração de DNA *PowerSoil® DNA Isolation Kit* também foi utilizado como método de extração. No entanto, o emprego desse kit foi descartado devido à fragmentação das moléculas de DNA (Figura 5).

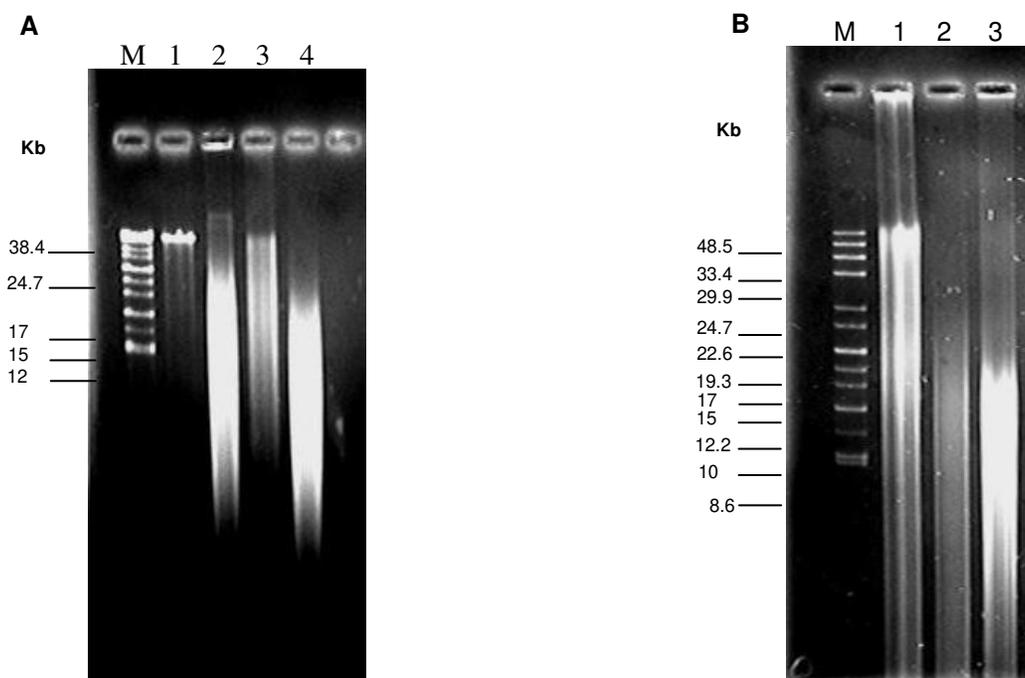


Figura 5. Eletroforese de DNA metagenômico em campo pulsado. **Painel A)** canaletas: 1, marcador de 36 Kb; 2, método de Fiantdt; 3, método de Fiantdt purificado com cloreto de céσιο; 4, método de Fiantdt purificado pelo Wizard® DNA Clean-Up System. **Painel B)** 1, método de Zhou; 2, método de Zhou purificado pelo Wizard® DNA Clean-Up System; 3, PowerSoil® DNA Isolation. M, marcador de massa molecular.

Tendo em vista somente as metodologias descritas por Fiantdt (2000) e Zhou e colaboradores (1996) como as mais adequadas inicialmente para a construção de todas as bibliotecas (gênicas e metagenômicas), foram avaliados os métodos de purificação com cloreto de céσιο e o kit comercial *Wizard® DNA Clean-Up System*.

Pelo método de cloreto de cério não ocorreu fragmentação das moléculas de DNA (Figura 5, canaleta 3). Porém, observou-se uma considerável diminuição da concentração de DNA, o que poderia restringir a detecção da diversidade de micro-organismos e genes que codificam as enzimas de interesse. Já a purificação com o kit *Wizard® DNA Clean-Up System* levou à fragmentação de DNA (Figura 5), no entanto, observou-se pouca diminuição na concentração de DNA. A determinação da concentração de DNA das diferentes soluções de extração e purificação foi quantificada em outra análise de eletroforese não demonstrada.

Devido às limitações expostas acima para a padronização de uma metodologia adequada para a construção de todas as bibliotecas, optou-se pelo método descrito por Fian dt (2000) seguida da purificação com kit *Wizard® DNA Clean-Up System* para a construção de bibliotecas de genes RNAr 16S e das enzimas-alvo. Isto porque a fragmentação do DNA não inviabilizaria a construção das bibliotecas gênicas, pois os fragmentos amplificados não estariam acima de 1,5 Kb.

Além disso, as duas metodologias juntas apresentaram uma amplificação do gene RNAr 16S de melhor qualidade quando comparada com aquela verificada para o DNA extraído utilizando o protocolo de Zhou e colaboradores ou o kit *Wizard® DNA Clean-Up System* (Figura 6).

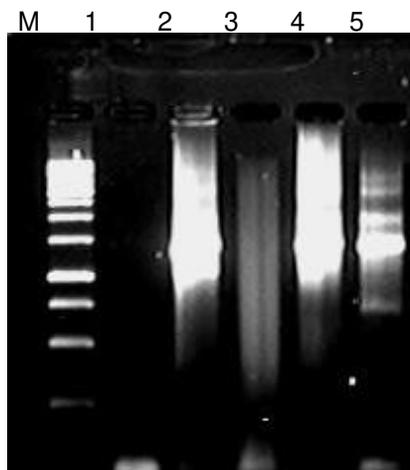


Figura 6. Amplificação do gene RNAr 16S do DNA obtido pelo método de Fian dt purificados com kit *Wizard® DNA Clean-Up System*. M, Marcador de peso molecular 1 kb DNA ladder; 1, controle negativo (sem DNA); 2, controle positivo utilizando DNA de uma linhagem de *Streptomyces* sp.; 3, amostra de DNA sem diluição; 4, amostra de DNA diluída 10X; 5 amostra de DNA diluída 100X.

Para a construção da biblioteca metagenômica em fósforo, optou-se pelo método de extração indireto, descrito por Liles e colaboradores (2008). O protocolo descrito por Hardeman e Sjöling (2007), o qual também é um método indireto, foi descartado devido à observação de rápida degradação de DNA e co-extração de impurezas.

Fragmentos de DNA acima de 40 kb foram selecionados através de eletroforese em campo pulsado (Figura 7), pois em tentativas prévias de construção da biblioteca foi constatada a degradação dos fragmentos nas sucessivas reações enzimáticas necessárias para a ligação do inserto nos fósforos. Cerca de 2 µg de DNA foram obtidos.

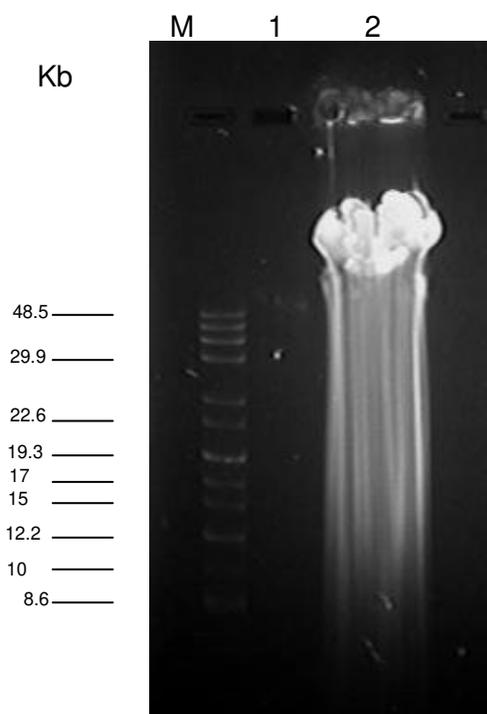


Figura 7. Eletroforese de DNA metagenômico extraído pelo método indireto. Canaletas: 1, marcador de 36 kb; 2, extração por lise indireta em *plugs*; M, marcador de peso molecular

5.2 Diversidade da comunidade microbiana

A análise filogenética da comunidade microbiana residente em reator anaeróbico de fluxo ascendentes (RAFA) foi baseada em 338 sequências parciais do gene RNAr 16S. As análises revelaram a identificação de 149 unidades taxonômicas operacionais pertencentes ao domínio Archaea e Bacteria.

Um total de 130 clones derivados da biblioteca de DNAr 16S do domínio Archaea foi sequenciado e 13 OTUs foram identificadas, entre essas 23% representando novas unidades taxonômicas. A análise filogenética demonstrou que todas as sequências foram afiliadas a Euryarchaeota não cultivadas, as quais foram distribuídas entre as ordens Methanosarcinales e Methanomicrobiales (Figura 8). Methanosarcinales foi o grupo predominante, sendo representado por 90 sequências distribuídas entre 5 OTUs.

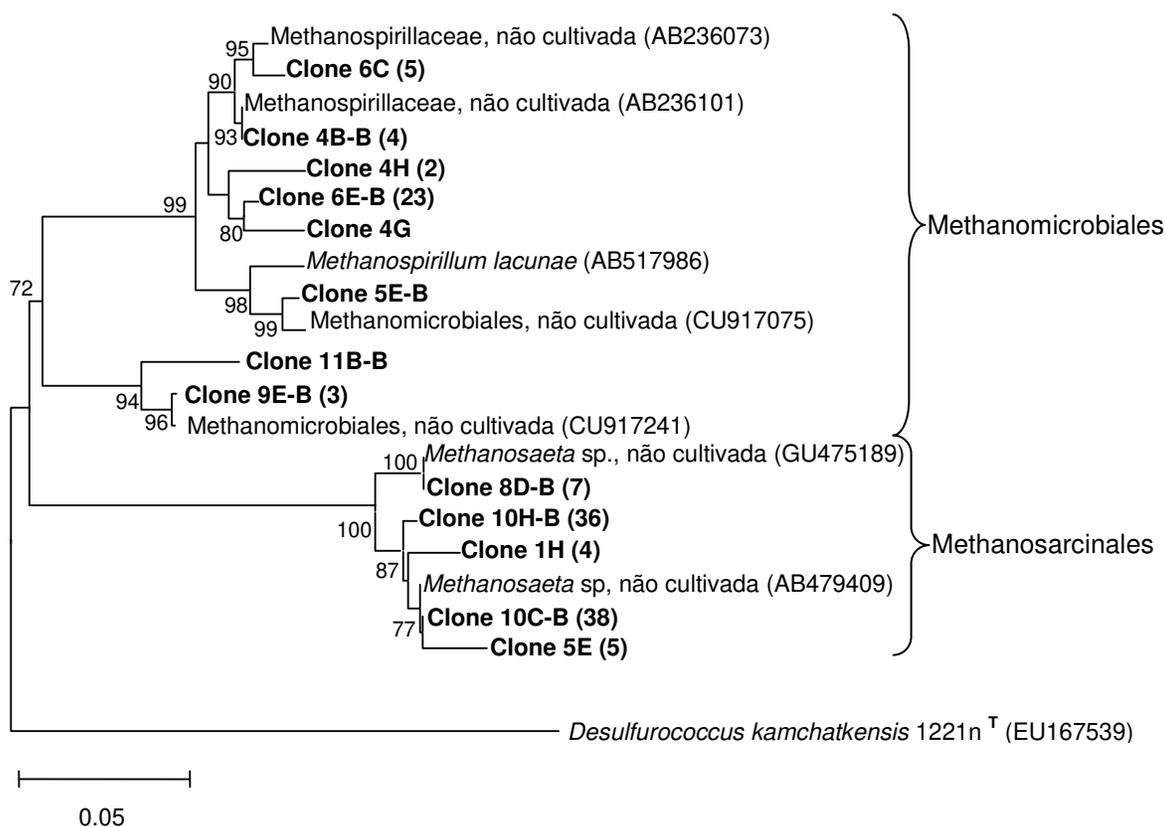


Figura 8. Árvore filogenética das sequências parciais dos genes RNAr 16S representativas das OTUs do Domínio Archaea na amostra de lodo e organismos relacionados recuperados da base de dados, utilizando o método de Neighbour-Joining e distância evolutiva Kimura 2p. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão listados. Os números entre parênteses correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. *Desulfurococcus kamchatkensis* 1221n^T utilizado como grupo externo (*outgroup*).

Já a ordem Methanomicrobiales foi representada por 40 sequências e 8 OTUs, sendo 3 novas unidades taxonômicas operacionais. As novas OTUs representaram 65% do grupo de Methanomicrobiales. Também se observou a predominância de Methanosarcinales (arqueias acetoclásticas) sobre as Methanomicrobiales (arqueias hidrogenotróficas). Por outro lado, as arqueias hidrogenotróficas foram mais diversas do que as acetoclásticas.

Trabalhos prévios mostram uma alternância da predominância dessas ordens independentemente do resíduo orgânico fermentado e das metodologias moleculares de análises, tornando a comparação inconclusiva. Por exemplo, Jaenicke e colaboradores (2011) identificaram a predominância do gênero *Methanoculleus*, pertencente à ordem Methanomicrobiales, em reatores anaeróbicos de resíduos agrícolas, através de uma análise de pirosequenciamento. Enquanto que Karakashev e colaboradores (2005) utilizando sondas identificaram a predominância de Methanosarcinales em reatores anaeróbicos de dejetos de animais e de tratamento de esgoto.

O melhor exemplo dessa alternância de grupos filogenéticos vem dos trabalhos de Rivière e colaboradores (2009) que identificaram a predominância de Methanosarcinales enquanto que Chouari e colaboradores (2010) descrevem a predominância de Methanobacteriales. Esses dois trabalhos utilizaram a mesma metodologia molecular, extração de DNA e *primers* para amplificação do gene RNAr 16S, e amostras de diferentes reatores anaeróbicos no tratamento de esgoto. Possivelmente, esta divergência entre os estudos quanto à predominância dos grupos de arqueias possa ser explicada pela complexidade da comunidade microbiana de ambientes anaeróbicos, a qual está sujeita a alterações recorrentes durante todo o processo biológico de digestão, e pela possibilidade da constante alimentação dos reatores com resíduos orgânicos, resultando na alteração da composição da microbiota.

As análises filogenéticas da biblioteca de DNAr 16S do domínio Bacteria revelaram 136 OTUs, das quais 50 OTUs foram identificadas como espécies cultivadas (similaridade \geq

97%) e 86 OTUs afiliadas a micro-organismos não cultivados (Tabela 6). Vinte e cinco OTUs não identificadas, representadas por 68 sequências, corresponderam a um novo taxa dentro do filo Synergistetes (Figura 9).

As OTUs mais abundantes afiliadas a grupos filogenéticos conhecidos foram distribuídas da seguinte maneira: Proteobacteria (35,9%), Synergistetes (12,5%), Firmicutes (11%) e Bacteroidetes (8,1%), correspondendo a 67,5 % do total das sequências. O restante das OTUs foram menos numerosas, sendo associadas aos filos Acidobacteria, Verrucomicrobia, Actinobacteria, Chloroflexi e Spirochaetes (Tabela 6).

O filo Proteobacteria foi o grupo filogenético mais abundante dessa biblioteca, sendo 36 sequências (17,3%) afiliadas à classe Deltaproteobacteria, 11 (5,3%) à classe Alphaproteobacteria, 8 (3,8%) à classe Betaproteobacteria e 3 (1,4%) à classe Gammaproteobacteria. Essas sequências mostraram-se relacionadas tanto a micro-organismos cultivados como a não cultivados, apresentando similaridades que variaram entre 74% a 100 % (Figura 10, Tabela 6). Esses resultados corroboraram dados de estudos prévios que descrevem Proteobacteria como o grupo filogenético predominante e mais diverso em análises filogenéticas de amostras de lodo de esgoto (NELSON; MORRISON; ZHONGTANG, 2011; CHOUARI et al., 2010).

Dentro da classe Deltaproteobacteria, 16 sequências foram distribuídas em 8 OTUs atribuídas à família Syntrophaceae e 1 OTU afiliada à Syntrophobacteraceae, representadas principalmente pelos gêneros *Syntrophus* e *Syntrophobacter*, respectivamente (Figura 10). Esses micro-organismos são sintróficos e consumidores de propionato, o qual é produto intermediário na fase de degradação da matéria orgânica e o seu consumo é importante para evitar o acúmulo de ácidos produzidos durante a segunda fase da fermentação (ARIESYADY et al., 2007; NELSON; MORRISON; ZHONGTANG, 2011).

Desta forma, não é surpresa a presença desses consumidores sintróficos. Além disso, a degradação do propionato ocorre somente quando a pressão parcial de H₂ é baixa. Essa pressão é mantida por micro-organismos tais como bactérias sulfato-redutoras consumidoras de hidrogênio pertencentes ao gênero *Desulfovibrio* (RODRÍGUEZ et al., 2012).

Tabela 6. Distribuição das OTUs bacterianas afiliadas aos grupos taxonômicos identificados nas amostras de RAFA.

	Proteobacteria				Candidate									Total
	Apha	Beta	Gamma	Delta	Synergistetes	Bacteroidetes	Firmicutes	Synergistetes	Acidobacteria	Actinobacteria	Verrucomicrobia	Chloroflexi	Spirochaetes	
No. de clones	11	8	3	36	23	21	16	68	7	5	4	4	2	208
% clones	5,3	3,8	1,4	17,3	11,1	10,1	7,7	32,7	3,4	2,4	1,9	1,9	0,9	100
No. de OTUs	11	8	3	27	17	11	15	25	6	4	4	3	2	136
% OTUs	8,1	5,8	2,2	19,8	12,5	8,1	11	18,4	4,4	2,9	2,9	2,2	1,4	100
OTUs \geq 97% cultivadas ^a	7	8	0	16	3	0	6	0	6	2	0	0	2	50
OTUs \geq 97% não-cultivadas ^b	4	0	3	11	14	11	9	25	0	2	4	3	0	86

^a similaridade \geq 97% com a sequência mais relacionada a um representante cultivado.

^b similaridade \geq 97% com a sequência mais relacionada a um representante não cultivado.

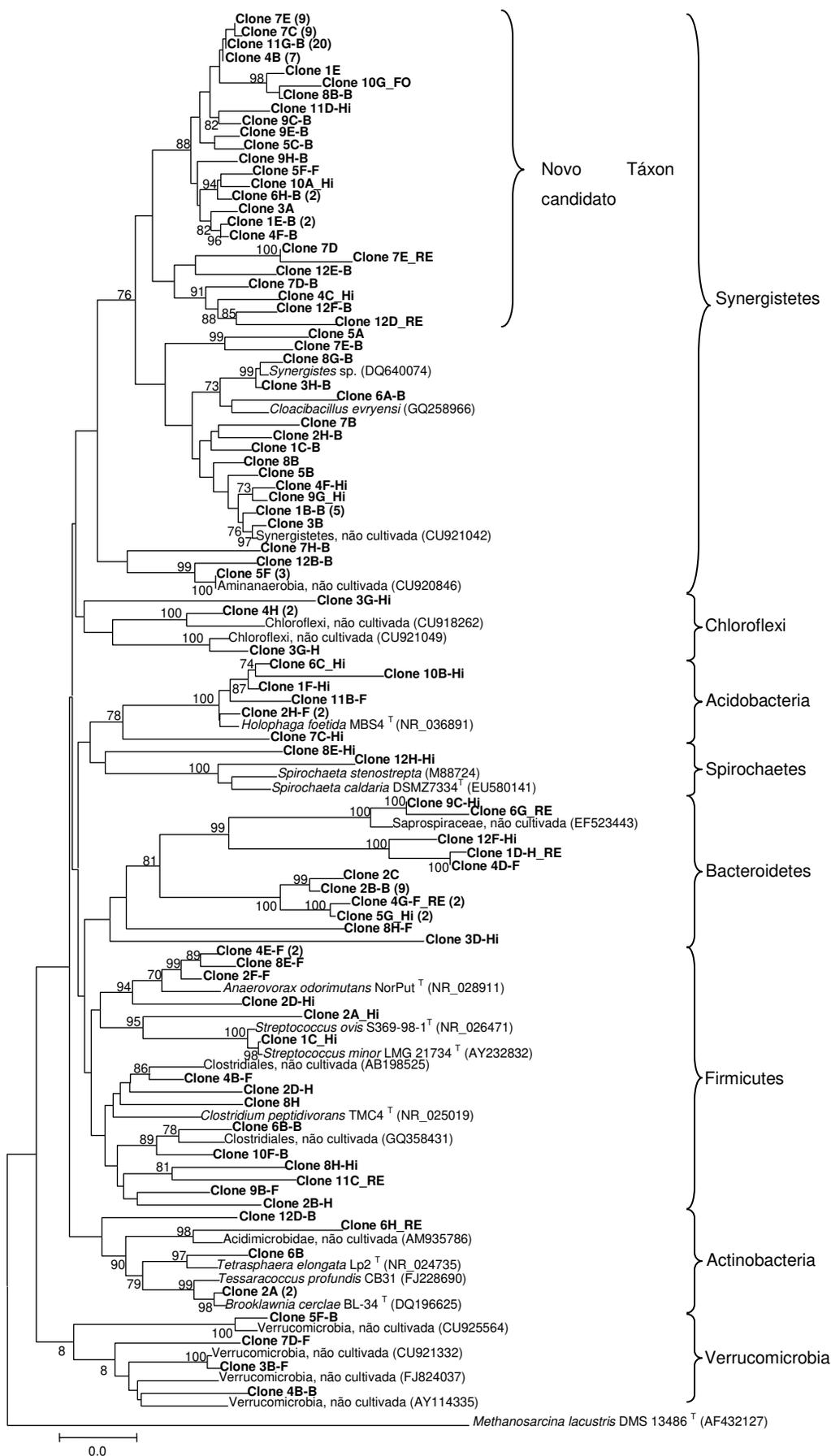


Figura 9. (página anterior). Árvore filogenética baseada nas sequências parciais do gene RNAr 16S dos clones representativos das OTUs do Domínio Bacteria e organismos relacionados recuperados das bases de dados. Os números entre parêntese correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão indicados. *Methanosarcina lacustris* foi utilizada como grupo externo (*outgroup*).

Nesse estudo, foram identificadas 7 sequências atribuídas a 7 OTUs afiliadas ao gênero *Desulfovibrio* e 5 sequências distribuídas em 4 OTUs associadas à família Desulfobacteraceae, representada principalmente pelo gênero *Desulfofaba* (Figura 10).

Já as sequências mais abundantes da classe Betaproteobacteria foram afiliadas aos gêneros *Zoogloea* e *Dechloromonas* (total de 4 OTUs), pertencentes à ordem Rhodocyclales, e ao gênero *Comamonas* (3 OTUs), pertencente à ordem Burkholderiales.

Muitos membros da classe Betaproteobacteria e ordem Rhodocyclales são comuns e abundantes em sistemas de tratamento de águas residuais e possuem diferentes capacidades fisiológicas que são benéficas para a degradação e transformação de poluentes, tais como nitrogênio, fósforo e compostos aromáticos (LIU et al., 2006; HESSELSON et al., 2009). Algumas cepas de *Dechloromonas*, as quais pertencem à ordem Rhodocyclales, degradam derivados de óleos tais como tolueno, e podem, assim, contribuir para reduzir o impacto ecológico de compostos aromáticos no ambiente (CHAKRABORTY et al., 2005; DINCER; MUEZZINOGLU, 2008; HESSELSON et al., 2009).

Alguns trabalhos descreveram sequências de *Comamonas* recuperadas de efluentes industriais, tais como petroquímica, águas residuais de plásticos e celulose, relacionadas com organismos envolvidos na degradação de fenol e seus derivados (JIANG et al., 2004; SILVA et al., 2010).

As duas classes do filo Proteobacteria menos abundantes foram Alphaproteobacteria, representada principalmente pela família Hyphomicrobiaceae (4 OTUs), e Gammaproteobacteria, representada por 3 OTUs afiliadas a micro-organismos não cultiváveis (Tabela 6). Membros da família Hyphomicrobiaceae foram descritos como desnitrificantes, desempenhando importante função em lodos de esgoto (OSAKA et al., 2008).

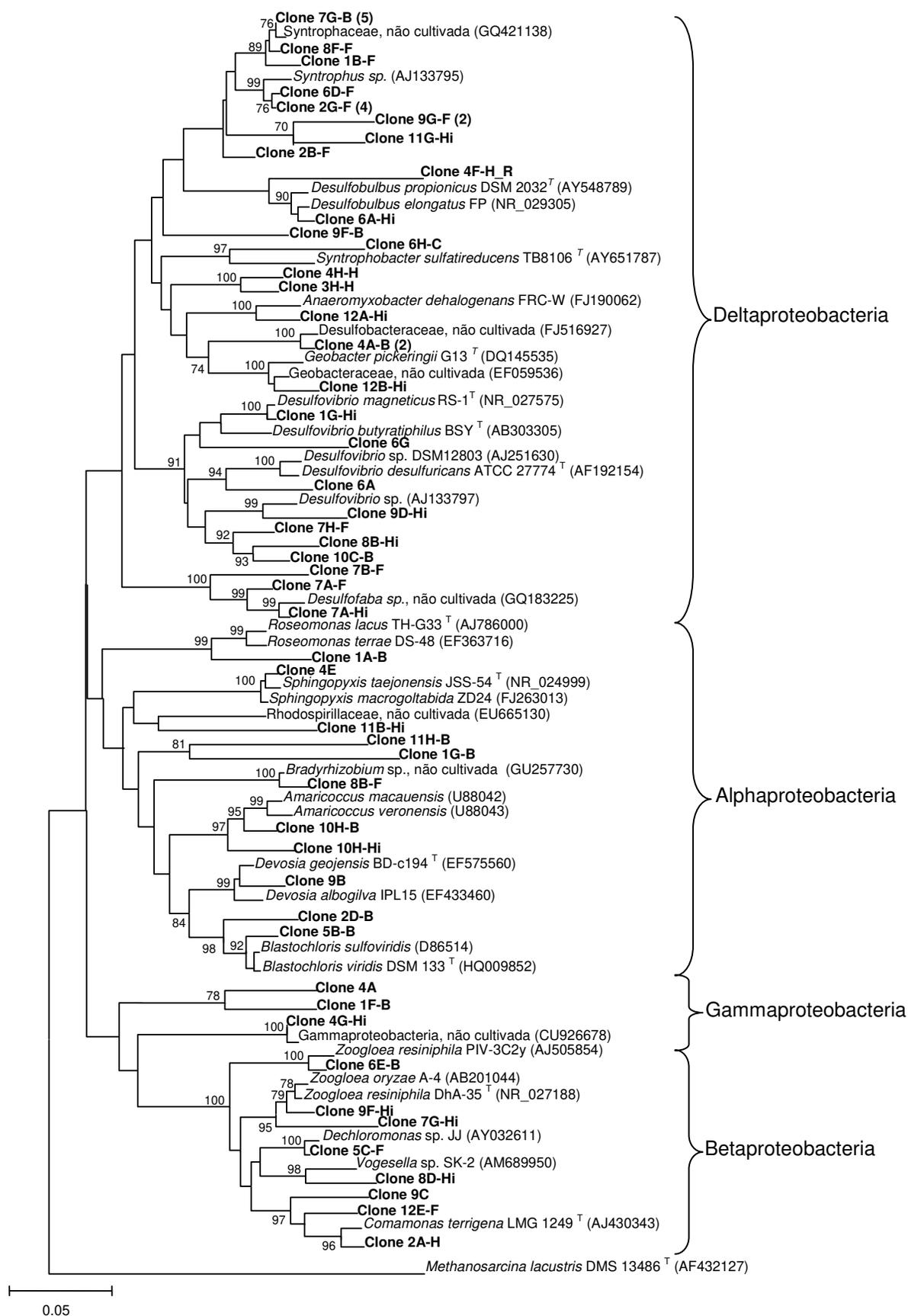


Figura 10 (página anterior). Árvore filogenética baseada nas sequências parciais do gene RNAr 16S dos clones representativos das OTUs dos membros do filo Proteobacteria e organismos relacionados recuperados das bases de dados. Os números entre parênteses correspondem a clones adicionais com similaridade $\geq 97\%$ com o clone representado na árvore. Os valores de bootstrap (1000 re-amostragens) acima de 70% estão indicados. *Methanosarcina lacustris* foi utilizada como grupo externo (*outgroup*).

As funções metabólicas de Alphaproteobacteria e Gammaproteobacteria não são muito discutidas em estudo de diversidade da comunidade microbiana de digestores anaeróbicos, talvez isso seja devido ao número inexpressivo de OTUs pertencentes a estas classes detectadas nesse tipo de amostra (CHOUARI et al., 2005; CHOUARI et al., 2010). Essa hipótese é amparada pelo estudo de Nelson e colaboradores (2011). Nesse estudo, os pesquisadores descreveram que de todas as sequências depositadas no GenBank que foram recuperadas de análises de digestores anaeróbicos, somente 378 sequências foram atribuídas a classe Gammaproteobacteria.

O segundo grupo filogenético bacteriano predominante da biblioteca do lodo de esgoto foi o filo Synergistetes, representado por 23 sequências (11,1%) e 17 OTUs (Tabela 6). Essas sequências foram afiliadas à família Synergistaceae e a micro-organismos não cultiváveis (Figura 9). Além dessas sequências relatadas acima, destacaram-se as 25 OTUs (68 sequências) que podem constituir um novo táxon dentro do filo Synergistetes. Esses resultados divergiram de estudos prévios que descreveram, independentemente da abordagem molecular empregada, a recuperação de poucas, ou nenhuma, OTUs afiliadas a Synergistetes de diferentes reatores anaeróbicos (RIVIÈRE et al., 2009; CHOJNACKA et al., 2011; CHOUARI et al., 2010; JAENICKE et al., 2011).

O terceiro grupo filogenético mais predominante na biblioteca foi Bacteroidetes, com 21 clones distribuídos em 11 OTUs (8,1%). Todas as sequências desse filo foram representadas por OTUs afiliadas a micro-organismos não cultiváveis (Tabela 6). O filo Firmicutes foi representado por 16 sequências distribuídas em 15 OTUs. Dentre essas, 13 OTUs (14 clones) foram associadas à ordem Clostridiales, principalmente representada pelo gênero *Clostridium*, e 2 OTUs foram atribuídas à ordem Lactobacillales, representada pelo gênero *Streptococcus* (Figura 9).

A detecção de Bacteroidetes e Firmicutes entre os grupos filogenéticos predominantes nessa análise corroborou estudos prévios de amostras de lodo de esgoto (RIVIÈRE et al., 2009, CHOUARI et al., 2010; WANG et al., 2010; JAENICKE et al., 2011). Embora todas as

sequências afiliadas a Bacteroidetes tenham sido relacionadas a micro-organismos não cultiváveis, sugere-se que Bacteroidetes têm a função de degradar ácidos graxos de cadeia longa, assim como fermentar carboidratos complexos (CHOAURI et al., 2005; CHOAURI et al., 2010; JAENICKE et al., 2011, LI et al., 2013). Já os micro-organismos pertencentes ao filo Firmicutes são comumente descritos como produtores de hidrogênio em processos anaeróbicos e envolvidos na degradação de polissacarídeos (JAENICKE et al., 2011; RADEMACHER et al., 2012).

Em relação aos filios menos abundantes, foi possível determinar a afiliação de algumas sequências, tais como, 6 OTUs afiliadas ao gênero *Holophaga* (filo Acidobacteria), 2 OTUs atribuídas à ordem Actinomycetales (filo Actinobacteria) e 2 OTUs afiliadas ao gênero *Spirochaeta* (filo Spirochaetes). Todas as sequências afiliadas ao filo Chloroflexi e Verrucomicrobia foram identificadas como OTUs potencialmente novas (Figura 9).

Acidobacteria, Actinobacteria, Verrucomicrobia, Chloroflexi e Spirochaetes representaram um total de 22 sequências (13,6%) dentro da biblioteca de clones do domínio Bacteria (Tabela 6). A maioria dessas sequências foi afiliada a micro-organismos não cultivados (Figura 9). Somente 3 das 19 OTUs foram atribuídas a espécies cultiváveis, sendo 1 OTU afiliada à Acidobacteria e 2 OTUs afiliadas à Actinobacteria. Com exceção do filo Chloroflexi, todos estes filios citados como pouco abundantes neste trabalho tem sido frequentemente associados a poucas sequências, ou nem sequer foram detectados, em outros estudos de comunidade microbiana de digestores anaeróbicos (RIVIÈRE et al., 2009; CHOUARI et al., 2010; JAENICKE et al., 2011). Rivière e colaboradores (2009) descreveram Chloroflexi como grupo filogenético predominante em bibliotecas de clones construídas de diferentes plantas de tratamento de esgoto municipal. Adicionalmente, uma meta-análise de bancos de dados públicos demonstrou que as sequências recuperadas de digestores anaeróbicos e afiliadas ao domínio Bacteria foram pertencentes ao filo Chloroflexi. Apesar disso, não é possível inferir a função metabólica desse filo (NELSON; MORRISON; ZHONGTANG, 2011).

Em resumo, vários trabalhos identificaram não somente novas OTUs a partir de amostras retiradas de lodo de esgoto, mas também uma nova divisão candidata no domínio Bacteria (GUERMAZI et al., 2008; RIVIÈRE et al., 2009; CHOUARI et al., 2010). Estes resultados demonstram que digestores anaeróbicos são ecossistemas específicos, com uma elevada diversidade e que contém muitas sequências que não são encontradas em outros ambientes, as quais representam grupos taxonômicos potencialmente novos (RIVIÈRE et al., 2009).

5.3 Diversidade de hidrogenases

Um total de 62 clones positivos contendo os fragmentos do gene de FeFe-hidrogenase (*hydA*) foi sequenciado. A confirmação de todas as sequências *hydA* como sequências verdadeiras de FeFe-hidrogenase foi realizada pela identificação do motif L2 (MPCxxKxxE) (Figura 11).

Análises do DOTUR permitiram identificar 22 OTUs entre as 62 sequências putativas de *hydA*. De acordo com a reconstrução filogenética derivada do alinhamento das sequências putativas de *hydA* e das sequências de aminoácidos mais relacionadas recuperadas do Genbank, as sequências foram distribuídas em 4 *clusters* filogenéticos (Figura 10).

Dentro do *cluster* I, as sequências (frequência de clones = 19,3 %) apresentaram entre 61 a 74 % de identidade com sequências de *Ignavibacterium album* JCM 16511, *Thiorhodococcus drewsii* AZ1 e *Dehalococcoides* sp. VS recuperadas do GenBank. O segundo *cluster* foi composto de 2 OTUs (frequência de clones = 3,2 %), as quais apresentaram 63% de identidade com a sequência de *Thermoanaerobacterium saccharolyticum* JW/SL-YS485 e 61% de identidade com *Caldithrix abyssi* DSM 13497.

A maioria das sequências putativas de *hydA* (frequência de clones = 67,7%) foi agrupada no *cluster* III, o qual não apresentou nenhuma relação filogenética com as sequências recuperadas do GenBank. Dentro do *cluster* IV, 5 OTUs (frequência de clones = 8,1 %) foram relacionadas a *Thermanaerovibrio acidaminovorans* DSM 6589 (71 % de identidade) e a bactérias não cultiváveis (90 % de identidade).

Independentemente do *cluster*, a maioria das sequências putativas de *hydA* se mostrou mais relacionada filogeneticamente entre si do que às sequências recuperadas do GenBank, principalmente aquelas agrupadas no *cluster* III amparadas pelo valor de 99% de *bootstrap* (Figura 10).

Figura 11. Painel A): Árvore filogenética baseada no alinhamento das sequências putativas de aminoácidos de FeFe-hidrogenase e suas sequências relacionadas recuperadas das bases de dados. As distâncias evolutivas foram calculadas pelo modelo de Poisson e a construção da árvore pelo método “Neighbor-joining”. Os números entre parêntese correspondem a clones adicionais com similaridade \geq 82% com o clone representado na árvore. Os valores de *bootstrap* (1000 re-amostragens) acima de 70% estão indicados. A sequência da proteína Nar1p pertencente à *Saccharomyces cerevisiae*, homóloga a *HydA* foi utilizada como grupo externo (*outgroup*). **Painel B)** Sequências de aminoácidos deduzidos a partir dos clones deste estudo e sequências de referência de *HydA* ilustrando a coerência filogenética das novas substituições dentro do motif L2(MPCxxKxxE). *Clusters* designados de C1 a C4.

A detecção do motif L2 em todas as sequências putativas de *hydA* permitiu uma simples e rápida identificação das verdadeiras sequências de FeFe-hidrogenases, distinguindo essas sequências das sequências homólogas de *hydA*.

Ao contrário dos resultados descritos por Xing e colaboradores (2008), a amplificação direta de genes de *hydA* forneceu uma visão incompleta da população bacteriana associada à produção de hidrogênio quando comparada aos resultados obtidos das análises do gene RNAr 16S. No entanto, nossos resultados corroboraram estudos anteriores (SCHMIDT; DRAKE; HORN, 2010; XU et al., 2013) que descrevem que ainda não é possível definir a espécie usando análises baseadas em genes de hidrogenase. Nesse estudo, a maioria das sequências putativas de *hydA* foram filogeneticamente distintas das sequências de FeFe-hidrogenases recuperadas das bases de dados, ou relacionadas a micro-organismos que não foram frequentemente detectados em reatores anaeróbicos.

Desta forma, a utilização do gene de FeFe-hidrogenase como marcador filogenético ainda não é eficiente na caracterização da comunidade bacteriana produtora de hidrogênio, provavelmente devido ao baixo número de sequências de FeFe-hidrogenase disponíveis em banco de dados. Além disso, quando as sequências putativas *hydA* foram comparadas às sequências depositadas no GenBank, a média de identidade variou entre 53 a 74%, sugerindo a identificação de novas sequências de hidrogenases. Esses resultados demonstram a necessidade de mais sequências de FeFe-hidrogenase disponíveis em banco de dados para o desenho de novos *primers* visando uma maior confiabilidade na identificação de tais genes em amostras ambientais.

Para a construção da biblioteca de NiFe-hidrogenase, três pares de *primers*, sendo dois pares de *primers* desenhados nesse trabalho e um descrito por Kim e colaboradores (2007), foram utilizados para a amplificação do gene de NiFe-hidrogenase a partir da amostra de lodo. No entanto, após a avaliação de diferentes condições de temperatura de anelamento e ciclos para a amplificação do gene alvo, um sinal positivo fraco foi detectado somente para *Citrobacter freundii* CBMAI 1187 (KCTC 2509), micro-organismo utilizado como controle positivo das reações, quando utilizados os *primers* descritos na literatura.

Apesar de ser a mais numerosa e mais bem estudada classe de hidrogenases do domínio Bacteria (VIGNAIS; BILLOUD, 2007), não há trabalhos descritos até o momento na literatura que avaliam a diversidade de NiFe-hidrogenase a partir de amostras ambientais. Diferentemente do gene de FeFe-hidrogenase, cuja literatura relata diferentes trabalhos avaliando não somente a diversidade desses genes em amostras ambientais como a também o desenho de diferentes *primers* degenerados para sua detecção (XING; REN; RITTMANN, 2008; BOYD; SPEAR; PETERS, 2009; SCHMIDT; DRAKE; HORN, 2010). Para NiFe-hidrogenase há somente um trabalho descrito que sugere um par de *primers* degenerados para a amplificação desse gene (KIM; JUNG; CHA, 2007), o qual foi um dos pares de *primers* utilizados nesse trabalho. Possivelmente haja a necessidade de um trabalho de bioinformática mais acurado para o desenho de *primers* degenerados mais eficazes, uma vez que a classe de NiFe-hidrogenase é dividida em 4 grupos, o que pode dificultar a detecção dessa diversidade em uma amostra ambiental.

5.4 Índices de riqueza e diversidade das bibliotecas gênicas

Índices de diversidade e estimativas de riqueza foram calculados para todas as bibliotecas gênicas (Tabela 7, Figura 12). Para a estimativa de riqueza, os índices não paramétricos ACE e Chao1 foram empregados para estimar o número de OTUs presente na amostra a partir da frequência de OTUs únicas e raras identificadas na biblioteca de clones em estudo. Adicionalmente, foram calculados os índices de Simpson e Shannon para avaliar a diversidade da amostra. O índice de Simpson é positivamente correlacionado com a frequência das espécies, ao passo que o índice de Shannon está mais correlacionado com a riqueza de espécies e frequência, além de atribuir mais peso às espécies raras do que às mais comuns.

Tabela 7. Estimativas de riqueza e índices de diversidade das bibliotecas do gene RNAr 16S dos domínios Bacteria e Archaea.

Domínio (n ^a)	Distância	ACE ^b	Chao ^c	Shannon ^d	1/Simpson's ^e
Archaea (130)	0.03	14	14	1.9	5.1
Bacteria (208)	0.03	984	630	4.5	58.8

^an, número de seqüências analisadas.

^bPredição estatística não-paramétrica da riqueza total de OTUs diferentes baseado na distribuição de OTUs abundantes e raras.

^cPredição estatística não-paramétrica da riqueza total de OTUs baseada na distribuição de um e dois representantes na amostra.

^dÍndice de diversidade de Shannon.

^eÍndice de diversidade de Simpson.

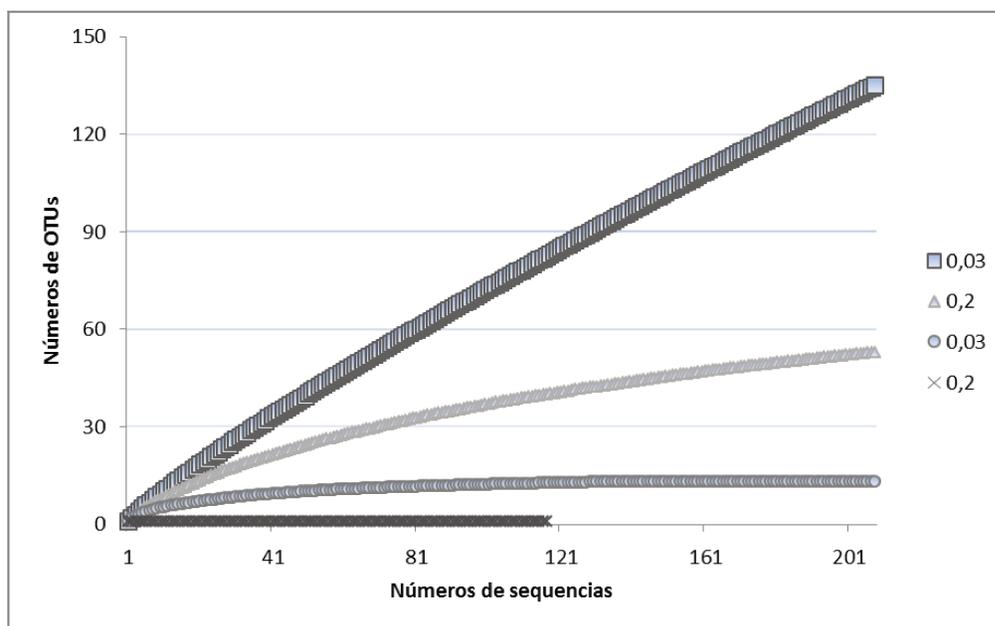


Figura 12. Curvas de rarefação das OTUs identificadas com base nas seqüências parciais de DNAr 16S. □ e Δ, distância evolutiva para o domínio Bacteria, 0.03 e 0.2, respectivamente. ○ e × distância evolutiva para o domínio Archaea, 0.03 e 0.2, respectivamente.

Ambos os índices não paramétricos indicaram um filo e 14 espécies para o domínio Archaea, sugerindo que a amostragem foi suficiente para revelar todos os filios e espécies presentes na amostra (Figura 7 e Tabela 7). O conjunto de dados estatísticos, sumarizados na Tabela 7, indicou uma baixa diversidade e um perfil de dominância na população. De acordo com Zhou e colaboradores (2002), valores de índice de Simpson abaixo de 50 indicam um perfil de dominância em uma comunidade.

Para a biblioteca de bactéria, os índices não paramétricos ACE e Chao1 estimaram a presença de 984 e 630 espécies na amostra, respectivamente (Tabela 7). Desta forma, as estimativas de riqueza e diversidade calculadas para o domínio Bacteria corroboraram estudos anteriores que descreveram a grande diversidade da comunidade bacteriana presente nesse tipo de amostra ambiental (RIVIÈRE et al., 2009; LI et al., 2013). Por outro lado, Rivière e colaboradores (2009) afirmaram que os estimadores de riqueza são sensíveis ao número de OTUs raras e devem ser utilizados com precaução, e que o intervalo de confiança é a chave para auxiliar na determinação precisa da estimativa de riqueza.

Já a estimativa de riqueza para as sequências putativas de *hydA* variou entre 63 a 73 OTUs. Portanto, menos de 50% dessa estimativa foi detectada pela amostragem realizada. Além disso, os índices de diversidade de Shannon e Simpson, 2,39 e 5,25, respectivamente, indicaram que um esforço amostral adicional permitiria a observação de novas OTUs. Trabalhos prévios descreveram resultados similares, sugerindo também que a diversidade de genes de FeFe-hidrogenase em amostras ambientais é alta (XING; REN; RITTMANN, 2008; BOYD; SPEAR; PETERS, 2009).

5.5 Construção da biblioteca metagenômica e triagem por PCR

Após a obtenção dos fragmentos acima de 25 Kb pelo método descrito por Liles e colaboradores (2008), a biblioteca metagenômica foi construída e 17.280 clones foram obtidos. Para a confirmação da clonagem de diferentes insertos nesta biblioteca, foram selecionados 5 clones aleatoriamente para a extração de fosmídeos pelo método de lise alcalina. Em seguida, os fosmídeos submetidos a reações de restrição com a enzima *XbaI* revelaram um perfil de restrição diferente para cada clone, permitindo a validação da biblioteca (Figura 13).

Para o isolamento e caracterização de micro-organismos ou enzimas envolvidas na produção de hidrogênio tem sido empregados métodos tradicionais, incluindo cultivo

quimioautótrofico e detecção da atividade de hidrogenases, análises amperométricas, captação de trítio, e cromatografia gasosa (KIM; CHA, 2013). Entretanto, essas abordagens demandam muito tempo e são laboriosas, não sendo adequadas para uma rápida triagem de alto desempenho.

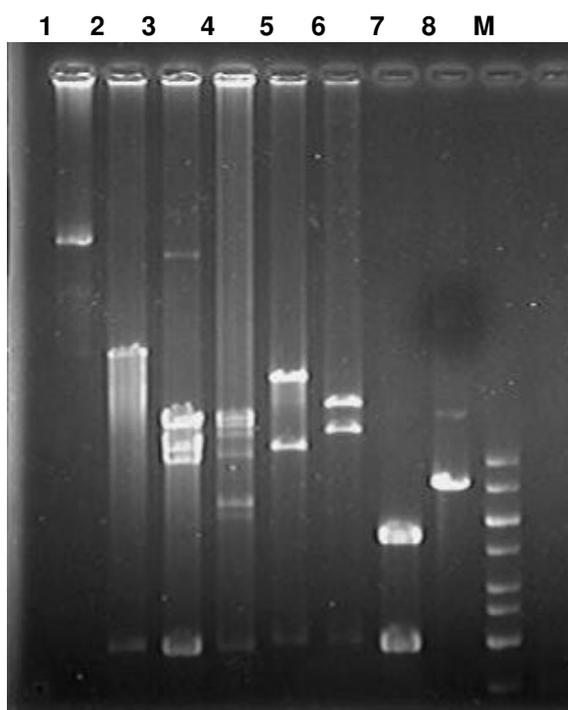


Figura 13. Eletroforese em campo pulsado dos produtos de restrição do DNA fosmidial dos clones metagenômicos. Canaletas: 1, DNA fosmidial não clivado; 2-6 DNA fosmidial de clones distintos; 7, vetor clivado; 8, vetor intacto; 9, marcador de peso molecular (1 kb).

Da mesma maneira, a purificação de hidrogenases também é árdua devido ao sistema de proteínas acessórias necessárias para o seu enovelamento e à sua sensibilidade ao oxigênio (VIGNAIS; BILLOUD, 2007). Devido a isso, uma triagem por PCR para verificar na biblioteca metagenômica a presença de genes de FeFe-hidrogenase foi realizada a partir das culturas dos clones.

No entanto, essa triagem foi descartada uma vez que foram detectadas inúmeras amplificações inespecíficas em todos os clones inicialmente avaliados. Essas amplificações

provavelmente ocorreram devido ao uso de *primers* degenerados de FeFe-hidrogenase na reação e, conseqüentemente, amplificação do gene de hidrogenase da célula hospedeira, *Escherichia coli* EPI300-T1R. Apesar da *Escherichia coli* não possuir o gene alvo, esse micro-organismo possui hidrogenases específicas encontradas somente em seu gênero (VIGNAIS; BILLOUD, 2007).

Na tentativa de eliminar essa amplificação cruzada e identificar os *pools* que contêm os clones de FeFe-hidrogenase, para subsequente pirosequenciamento, os clones fosmidiais foram cultivados individualmente e extraídos em *pools* com o auxílio do *QIAGEN Large-Construct kit* (*QIAGEN, Hilden*). Porém, o kit comercial não foi eficiente para a total remoção do DNA genômico da célula hospedeira, pois ainda foram detectadas amplificações inespecíficas, similarmente ao observado quando a reação de PCR foi realizada diretamente a partir da cultura de clones.

Desta forma, optou-se pelo sequenciamento de 10 *pools* de fosmídeos totalizando 9.600 clones no sistema 454 GS FLX (Roche, Life Sciences) por meio da química GS FLX Titanium.

5.6 Análises estatísticas dos dados metagenômicos

Os recursos financeiros disponíveis permitiu a utilização de meia placa na plataforma 454, então, optou-se pelo sequenciamento de 9.600 clones para aumentar a cobertura dos fragmentos clonados ao invés de sequenciar toda a biblioteca e diminuir a cobertura desses fragmentos.

O pirosequenciamento de 9.600 fosmídeos da biblioteca metagenômica resultou em 327.682 sequências com tamanho médio de 650 pb (Figura 14), representando cerca de 218 milhões de pb. A montagem das sequências metagenômicas em regiões contíguas (*contigs*) resultou em um total de 12.723 *contigs*, sendo a maioria desses com tamanho médio de 2 Kb. Os dados estatísticos do sequenciamento estão sumarizados na Tabela 8.

Outro dado importante do pirosequenciamento da biblioteca de fosmídeo foi o conteúdo de GC resultante. Observou-se que a maioria das sequências metagenômicas foi composta de 40 a 45% de GC, enquanto que o restante das sequências esteve acima de 50%. Um fator que contribui para esse conteúdo de GC resultante é que a fragmentação do DNA ambiental ocorre mais facilmente em regiões com menor conteúdo de GC, devido ao menor número de pontes de hidrogênio, tornando a cadeia de DNA mais susceptível a quebras.

Portanto, os fragmentos acima de 30 kb necessários para a construção de uma biblioteca de fosmídeos são ricos em GC, enquanto que as bibliotecas de pequeno inserto (bibliotecas de plasmídeos) não apresentam elevado conteúdo de GC (TEMPERTON et al., 2009 ; DANHORN; YOUNG; DELONG, 2012).

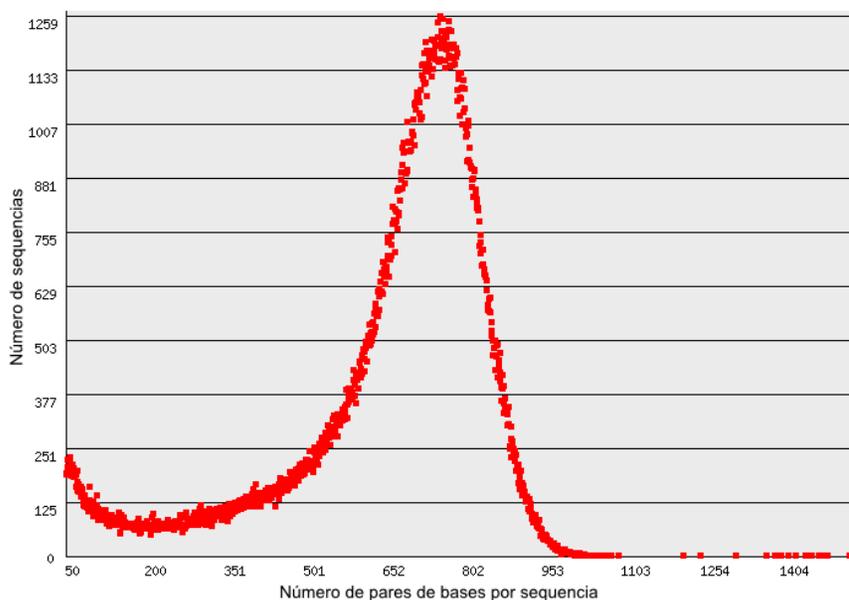


Figura 14. Número de pares de bases por sequência e número de sequências no conjunto de dados metagenômicos.

Tabela 8. Dados estatísticos do pirosequenciamento da biblioteca de fosmídeos.

Número de pares de bases	218 Mpb
Número de sequências	327.682
Média de pares de bases das sequências	665 pb
Média de pares de bases em um contig	2008 pb
Número total de <i>contigs</i>	12.723

5.7 Composição taxonomia da biblioteca metagenômica

A classificação taxonômica baseada no Classifier RDP identificou 39 e 28 fragmentos do gene RNAr 16S pertencentes aos domínios Bacteria e Archaea, respectivamente.

Euryarchaeota e Bacteroidetes foram os filos com maior número de sequências identificadas dentro dos seus respectivos domínios. Apesar do gene RNAr 16S ser um marcador filogenético de confiança, esse baixo percentual de sequências metagenômicas associadas a esse gene é comum em conjuntos de dados metagenômicos (SANAPAREDDY et al., 2009; JAENICKE et al., 2011), e nesse tipo de abordagem são insuficientes para a análise da comunidade microbiana. Para contornar essa limitação e obter uma análise mais profunda da composição taxonômica da biblioteca de fósdeos, foram utilizados dois algoritmos para a dedução da origem taxonômica das sequências metagenômicas.

A classificação taxonômica baseada no menor ancestral comum (LCA) permitiu classificar 47.300 sequências metagenômicas dentro de um taxa, enquanto que o Carma identificou 86.029 EGTs (do inglês *Environmental Gene Tag*), representando a classificação de 14,4% e 26,2% do total das sequências do conjunto de dados metagenômicos, respectivamente. A porcentagem de identificação da origem taxonômica das sequências metagenômicas nessa análise é semelhante a estudos prévios de pirosequenciamento de DNA extraído diretamente de amostras de biodigestores, indicando que ainda existe uma imensa quantidade de espécies de micro-organismos não identificados ou não cultiváveis em sistemas anaeróbicos de digestão (JAENICKE et al., 2011; LI et al., 2013).

As 38.729 sequências metagenômicas a mais classificadas pelo Carma3 foram, em sua maioria (90,5%), associadas ao domínio Bacteria. No entanto, um quadro semelhante da composição taxonômica da biblioteca de fósdeos foi obtido em nível de domínio e filo (Tabelas 9 e 10) utilizando as duas abordagens. Em ambos os métodos, o domínio Bacteria foi dominante e o número de filos detectados foi similar (23 e 22 filos identificados pelo Carma3 e LCA, respectivamente), sendo Proteobacteria, Bacteroidetes e Firmicutes os filos dominantes. Contudo, algumas diferenças foram identificadas em função do algoritmo utilizado. Por exemplo, o expressivo número de EGTs identificadas pertencentes aos filos Firmicutes e Bacteroidetes foi maior quando comparado ao número de sequências classificadas pelo LCA. Betaproteobacteria, Gammaproteobacteria e Methanomicria foram as classes mais abundantes, independentemente do algoritmo empregado. Entretanto, a classificação

taxonômica realizada pelo LCA identificou maior número de sequências metagenômicas associadas a Betaproteobacteria e Methanomicrobia (Tabela 9). Por outro lado, apesar do grande número de EGTs pertencentes às classes Bacteroidia, Flavobacterillia, Clostridia e Bacilli, essa discrepância foi esperada devido ao número de EGTs pertencentes a Bacteroidetes e Firmicutes (Tabela 10), como descrito anteriormente. Apesar das diferenças apresentadas acima, e do fato de somente metade da biblioteca de fosmídeo ter sido sequenciada, Jaenicke e colaboradores (2011) relataram que o perfil taxonômico em nível de filo e classe não sofre mudanças expressivas devido ao maior número de sequências identificadas. Portanto, a detecção de Proteobacteria e Bacteroidetes como filos dominantes poderia ser explicada pela habilidade desses grupos filogenéticos para eficiente degradação de moléculas orgânicas complexas presentes em amostras de lodo de esgoto e, principalmente, pela interação sintrófica entre as espécies de micro-organismos destas classes, a qual é essencial na degradação anaeróbica (CHOUARI et al., 2010; SIEBER et al., 2010).

Por exemplo, a degradação de ácidos graxos por *Syntrophus* e *Syntrophobacter* (principais membros identificados pertencentes à Deltaproteobacteria) ocorre somente quando o hidrogênio, formiato e acetato, são mantidos em baixas concentrações por micro-organismos tais como as arqueias metanogênicas e *Desulfovibrio*, os quais são capazes de acoplar a oxidação dos componentes orgânicos e hidrogênio para reduzir o sulfato e produzir H₂S (STAMS et al., 2005; SIEBER et al., 2010; RODRÍGUEZ et al., 2011).

Apesar do filo Bacteroidetes ser comumente identificado em processos anaeróbicos, sua função ainda é somente sugerida em um nível mais geral do metabolismo, tais como a degradação de ácidos graxos e a fermentação de carboidratos (CHOAURI et al., 2005; CHOAURI et al., 2010; JAENICKE et al., 2011, LI et al., 2013).

Em nível de ordem, Burkholderiales, Aeromonadales e Desulfovibrionales foram as mais abundantes identificadas pelo LCA, enquanto que a maioria das EGTs foi associada às ordens Enterobacteriales, Flavobacteriales e Burkholderiales (Tabelas 9 e 10). Burkholderiales e Enterobacteriales, pertencentes às classes Betaproteobacteria e Gammaproteobacteria, respectivamente, têm importante função no metabolismo de lodo de RAFA (LEE; VERMAAS; RITTMANN, 2010; LEFEBVRE et al., 2010). Estudos descrevem Burkholderiales como micro-organismos desnitrificantes dominantes e como a ordem mais abundante pertencente à Betaproteobacteria nesse ecossistema (CHON et al., 2010; LEFEBVRE et al., 2010).

Bacteria	Firmicutes	628	Clostridia	518	Clostridiales	444	<i>Pelotomaculum</i>	44	<i>thermopropionicum</i>	44
							<i>Clostridium</i>	114		
	Verrucomicrobia	616	Opitutae	463						
	Actinobacteria	363	Actinobacteria	363	Actinomycetales	154	<i>Propionibacterium</i>	39	<i>acnes</i>	39
					Bifidobacteriales	143	<i>Bifidobacterium</i>	143	<i>adolescentis</i>	79
	Chlorobi	462	Chlorobia	462	Chlorobiales	462	<i>Chlorobium</i>	248	<i>phaeobacteroides</i>	176
	Acidobacteria	229	Solibacteres	155	Solibacterales	155	<i>Candidatus</i>	155		
	Planctomycetes	283	Planctomycetia	283	Planctomycetales	283	<i>Pirellula</i>	186	<i>staleyii</i>	186
Chloroflexi	164	Chloroflexi	66	Chloroflexales	65	<i>Roseiflexus</i>	44	<i>castenholzii</i>	22	
Synergistetes	133	Synergistia	133	Synergistales	133	<i>Thermanaerovibrio</i>	122	<i>acidaminovorans</i>	122	
Eukaryota	7									
Viruses	53									

^a – Somente os Filos que contêm no mínimo 130 sequências metagenômicas identificadas foram inseridos na tabela.

Tabela 10. Classificação taxonômica das sequências metagenômicas da biblioteca de fosmídeo gerados pelo Carma3.

Domínio	Filo ^a	60077	Classe	47053	Ordem	37101	Gênero	18613	Espécie	8427		
Archaea	Euryarchaeota	3697	Methanomicrobia	2898	Methanomicrobiales	1778	<i>Methanospirillum</i>	617	<i>hungatei</i>	617		
					Methanosarcinales	808	<i>Methanosarcina</i>	298	<i>barkeri</i>	51		
	Crenarchaeota	254	Thermoprotei	254	Thermoproteales	88	<i>Pyrobaculum</i>	51				
					Sulfolobales	59	<i>Sulfolobus</i>	49				
Bacteria	Proteobacteria	33629	Gammaproteobacteria	10466	Enterobacteriales	4025	<i>Escherichia</i>	2478	<i>coli</i>	2432		
					Pseudomonadales	1024	<i>Pseudomonas</i>	629				
					Alteromonadales	908	<i>Shewanella</i>	311				
					Aeromonadales	459	<i>Aeromonas</i>	404	<i>hydrophila</i>	149		
					Vibrionales	439	<i>Vibrio</i>	292	<i>cholerae</i>	70		
					Deltaproteobacteria	8453	Desulfovibrionales	1785	<i>Desulfovibrio</i>	1663	<i>vulgaris</i>	517
						Desulfuromonadales	1485	<i>Geobacter</i>	873			
							<i>Pelobacter</i>	149				
						Syntrophobacterales	1498	<i>Syntrophus</i>	999	<i>aciditrophicus</i>	999	
						<i>Syntrophobacter</i>	377	<i>fumaroxidans</i>	377			
					Myxococcales	733	<i>Anaeromyxobacter</i>	156				
						<i>Stigmatella</i>	100					
						Desulfobacterales	377	<i>Desulfococcus</i>	298	<i>oleovorans</i>	298	
					Betaproteobacteria	4679	Burkholderiales	2938	<i>Acidovorax</i>	505	sp.	211
									<i>Burkholderia</i>	360		
Rhodocyclales	678	<i>Dechloromonas</i>	467	<i>aromatica</i>			467					
Neisseriales	219	<i>Neisseria</i>	86	<i>meningitidis</i>			38					
Alphaproteobacteria	2873	Rhizobiales	1063	<i>Rhodopseudomonas</i>	121							
		Rhodobacterales	584									

				Rhodospirillales	219				
		Epsilonproteobacteria	734	Campylobacterales	505	<i>Campylobacter</i>	170	<i>jejuni</i>	51
		Zetaproteobacteria	16						
Bacteroidetes	12921	Bacteroidia	5658	Bacteroidales	5650	<i>Bacteroides</i>	4117	<i>uniformis</i>	257
								<i>fragilis</i>	210
		Flavobacteriia	3044	Flavobacteriales	2967	<i>Flavobacterium</i>	188	<i>johnsoniae</i>	90
Firmicutes	4596	Clostridia	2440	Clostridiales	1878	<i>Clostridium</i>	701		
		Bacilli	1476	Bacillales	829	<i>Bacillus</i>	330		
				Lactobacillales	490	<i>Streptococcus</i>	158		
Actinobacteria	1185	Actinobacteria	1174	Actinomycetales	816	<i>Mycobacterium</i>	168		
Planctomycetes	990	Planctomycetacia	964	Planctomycetales	964				
Chlorobi	780	Chlorobia	763	Chlorobiales	749	<i>Chlorobium</i>	547	<i>phaeobacteroides</i>	196
Cyanobacteria	649								
Spirochaetes	568	Spirochaetia	568	Spirochaetales	563	<i>Leptospira</i>	288	<i>borgpetersenii</i>	100
Chloroflexi	446	Chloroflexi	292	Chloroflexales	222	<i>Roseiflexus</i>	119		
Acidobacteria	362	Solibacteres	198	Solibacterales	198	<i>Candidatus Solibacter</i>	198	<i>usitatus</i>	198

^a – Somente os Filos que contêm no mínimo 130 sequências metagenômicas identificadas foram inseridos na tabela.

Já a identificação de membros de Enterobacteriales sugere duas funções para esses micro-organismos nesse ambiente: podem estar envolvidos na produção de hidrogênio e serem consumidores de oxigênio, o que asseguraria um ambiente anaeróbico que é essencial para o crescimento de Clostridiales, principais micro-organismos identificados pertencentes ao filo Firmicutes, os quais são produtores de hidrogênio bem caracterizados (HALLENBECK; GHOSH, 2009; LEE; VERMAAS; RITTMANN, 2010). Além disso, Clostridiales são os principais micro-organismos envolvidos na digestão de polissacarídeos e oligossacarídeos em digestores de resíduos orgânicos (KRAUSE et al., 2008; WANG et al., 2010; JAENICKE et al., 2011).

As sequências metagenômicas associadas ao domínio Archaea foram atribuídas aos filios Euryarchaeota e Crenarchaeota, sendo o último filo identificado somente pelo Carma3 (Tabela 10). Dentro do filo Euryarchaeota, o grupo hidrogenotrófico, representado principalmente por *Methanospirillum hungatei*, foi predominante em relação ao grupo de acetoclásticos, esse representado por diferentes gêneros tais como *Methanosarcina* e *Methanosaeta*, identificados pelo Carma3 e LCA, respectivamente. Diferentes abordagens moleculares descrevem o grupo hidrogenotrófico como predominante sobre o grupo acetoclástico, sugerindo uma alta produção de hidrogênio e o consumo desse hidrogênio associado ao CO₂ pelo grupo hidrogenotrófico como a principal via de produção de metano nesse ambiente (STEINBERG; REGAN, 2008; CHOARI et al., 2010; JAENICKE et al., 2011; CARDINALI-REZENDE et al., 2012). Apesar do filo Crenarchaeota ter sido identificado na amostra, representado por arqueias não metanogênicas, a sua função metabólica em digestão anaeróbica ainda é desconhecida (CHOARI et al., 2005; RIVIÈRE et al., 2009; CHOARI et al., 2010; JAENICKE et al., 2011).

5.8 Potencial metabólico da comunidade microbiana

Adicionalmente à determinação da origem taxonômica das sequências metagenômicas, a função dessas sequências foi descrita de acordo com as famílias Pfam identificadas, gerando um perfil funcional da biblioteca de fosmídeos.

Desta forma, do total de 327.682 sequências metagenômicas, foi determinada a função gênica de 101.379 dessas sequências pela análise Pfam. A hidrólise de macromoléculas, tais como carboidratos, proteínas e lipídeos, é o primeiro passo dessa conversão. Assim, focando nas enzimas envolvidas no metabolismo de carboidratos, foram identificadas 21 famílias de

hidrolases (Tabela 11), tais como a família glicosil-hidrolase 5 (celulase, PF00150) e glicosil-hidrolase 10 (xilanase, PF00331), as quais estão entre as 115 famílias de hidrolases que compõem a superfamília glicosil-hidrolases.

A classificação taxonômica dessas sequências metagenômicas identificadas como enzimas indicou Bacteroidetes, Proteobacteria e Firmicutes como os filos predominantes. Esses resultados corroboraram estudos anteriores que descreveram Firmicutes, representado principalmente por Clostridia, e Bacteroidetes como grupos filogenéticos predominantes na degradação de carboidratos em digestão anaeróbica (SCHLÜTER et al., 2008; RADAMACHER et al., 2011; van der LELIE et al., 2012; LI et al., 2013).

Na degradação de proteínas as famílias identificadas foram a peptidase_M29 (aminopeptidase T, PF02073), e a peptidase_M13 (PF01431), entre outras (Tabela 11). A maioria das EGTs identificadas na degradação de proteínas foi associada às classes Bacteroidia, Gammaproteobacteria e Deltaproteobacteria. No entanto, a degradação de proteínas é pouco discutida na maioria dos estudos de comunidade microbiana anaeróbica, e tampouco associada a membros do filo Bacteroidetes.

Tabela 11. Lista de EGTs associadas aos processos de hidrólise identificados na biblioteca de fosmídeo.

Acesso FAM	Descrição PFAM	Número def EGTs
<i>Degradação de Carboidratos</i>		
PF01055	Família glicosil-hidrolase 31	248
PF10566	Hidrolase 97	135
PF00150	Família glicosil-hidrolase 5 (Celulose)	107
PF07971	Família glicosil-hidrolase 92	93
PF04616	Família glicosil-hidrolase 43	82
PF07748	Família glicosil-hidrolase 38 C-terminal	82
PF00251	Família glicosil-hidrolase 32 N-terminal	82
PF00722	Família glicosil-hidrolase 16	45
PF07745	Família glicosil-hidrolase 53	42
PF02156	Família glicosil-hidrolase 26	35
PF07470	Família glicosil-hidrolase 88	32
PF00295	Família glicosil-hidrolase 28	23
PF01183	Família glicosil-hidrolase 25	22
PF01501	Família glicosil-transferase 8	15
PF01229	Família glicosil-hidrolase 39	14
PF00331	Família glicosil-hidrolase 10	14
PF03664	Família glicosil-hidrolase 62	8

PF01270	Família glicosil-hidrolase 8	5
PF02057	Família glicosil-hidrolase 59	3
PF00840	Família glicosil-hidrolase 7	1
PF03639	Família glicosil-hidrolase 81	1

Degradação de Proteína

PF01343	Família Peptidase S49	264
PF07969	Família amido hidrolase	170
PF10459	Peptidase S46	169
PF04389	Família peptidase M28	158
PF01432	Família peptidase M3	149
PF01136	Família peptidase U32	143
PF01434	Família peptidase M41	111
PF02163	Família peptidase M50	106
PF04909	Família amido hidrolase_2	93
PF03577	Família peptidase C69	80
PF01433	Família peptidase M1	76
PF01640	Família peptidase C10	76
PF01676	Superfamília metaloenzima	67
PF03576	Família peptidase S58	53
PF01364	Família peptidase C25	49
PF03412	Família peptidase C39	48
PF07722	Peptidase C26	44
PF10502	Peptidase S26	41
PF03575	Família peptidase S51	33
PF05649	Família peptidase M13	31
PF02073	Família peptidase M29	19
PF01431	Família peptidase M13	16
PF03571	Família peptidase M49	14
PF05057	Putativa serine esterase	10
PF10462	Peptidase M66	9
PF07998	Família peptidase M54	7
PF08291	Peptidase M15	7
PF06397	Desulfoferrodoxina, N-terminal domain	6
PF01650	Família peptidase C13	6
PF10460	Peptidase M30	4
PF03566	Família peptidase A21	4
PF08496	Família peptidase S49 N-terminal	1

Degradação de lipídeos

PF00756	Putativa esterase	157
PF02230	Fosfolipase/Carboxilesterase	32
PF01674	Lipase (classe 2)	19
PF01764	Lipase (classe 3) (Lipase_3)	12
PF03583	Lipase (LIP)	2

Já no metabolismo de lipídeos, as sequências metagenômicas foram atribuídas às famílias de lipases (PF01674), esterases putativas (PF00756) e hidrolases alfa/beta_2 (carboxylesterases, PF02230) (Tabela 11).

A origem taxonômica dessas EGTs previstas foi associada às classes Actinobacteria, Flavobacteriia e Deltaproteobacteria, indicando esses micro-organismos como potenciais degradadores de lipídeos nesse ecossistema. Porém, estudos prévios descrevem os gêneros *Bacteroides* e *Flavobacterium* (pertencentes às classes Bacteroidetes e Flavobacteriia) como responsáveis pela degradação de carboidratos e não associados à degradação de lipídeos na digestão anaeróbica (McBRIDE et al., 2009; RADAMACHER et al., 2011). Além disso, o número relativamente pequeno de sequências metagenômicas atribuídas a lipases é particularmente interessante, dada a carga orgânica depositada diariamente na estação de tratamento de esgoto municipal.

A literatura descreve três rotas metabólicas para a produção microbiana de hidrogênio, a saber: piruvato-ferredoxina oxirredutase (PFOR), NADH ferredoxina oxidorredutase (NFOR) e formato – hidrogênio - liase (FHL). Nesse estudo, foram identificadas sequências metagenômicas atribuídas a todas as rotas citadas acima para a produção de hidrogênio. Entre essas se destacam: EGTs pertencentes às famílias FeS (ferredoxina, PF04060), Fe_hyd_lg_C (subunidade maior da FeFe-hidrogenase, PF02906), Fe_hyd_Ssu (subunidade menor da FeFe-hidrogenase, PF02256), e piruvato ferredoxina (PFOR, PF01558) (Tabela 12). A classificação taxonômica das EGTs de ferredoxina e FeFe-hidrogenase (subunidade maior e menor) revelou Clostridia como classe predominante (Tabela 13).

Esses resultados corroboraram estudos prévios que descreveram PFOR, ferredoxina e [FeFe]-hidrogenase como principais vias e enzimas envolvidas na produção de hidrogênio em Clostridia, e este último como o grupo filogenético responsável pela produção desse gás em digestão anaeróbica (ABO-HASHESH; WANG; HALLENBECK, 2011; HALLENBECK; GHOSH, 2010).

Para uma análise mais profunda dos diferentes micro-organismos potencialmente envolvidos na produção de hidrogênio em ambiente anaeróbico, os resultados da análise taxonômica das EGTs previstas para FeFe-hidrogenase desse estudo foram comparados aos resultados obtidos do pirosequenciamento de uma amostra de biodigestor de resíduo agrícola (JAENICKE et al., 2011).

Tabela 12. Lista de EGTs associadas à produção de hidrogênio e metano identificados na biblioteca de fosmídeo.

Acesso PFAM	Descrição PFAM	Número de EGTs
Hidrogênio		
PF04060	Putativo <i>cluster</i> [FeS]	47
PF02906	Subunidade maior [FeFe]-hidrogenase, C-terminal	177
PF02256	Subunidade menor [FeFe]-hidrogenase	24
PF01558	Piruvato ferredoxina (POR)	242
PF09163	Formato desidrogenase N	29
PF00374	Hidrogenase Niquel-dependente	49
PF01155	Síntese de hidrogenase	27
PF01924	Síntese de hidrogenase	20
PF04809	Síntese de hidrogenase (HupH)	1
Metanogênese		
PF02552	CO desidrogenase /acetyl-CoA sintetase subunidade-ε	1
PF03599	CO dehydrogenase/acetyl-CoA sintase subunidade-δ	19
PF02289	Ciclohidrolase (MCH) (MCH)	6
PF01993	Metileno-5,6,7,8-tetrahydro-methanopterin desidrogenase	6
PF02663	Formil metano furano desidrogenase	77
PF01515	Acetil fosfato	118
PF04208	S-metiltransferase-tetrahydro-metanopterin, subunidade A	2
PF04207	S-metiltransferase-tetrahydro- metanopterin, subunidade D	2
PF04206	S-metiltransferase-tetrahydro- metanopterin, subunidade E	3
PF02007	S-metiltransferase-tetrahydro- metanopterin, subunidade H	3
PF00871	Família acetoquinase	229
PF02241	Metil-coenzime M redutase subunidade-β, C-terminal	2
PF02249	Metil-coenzime M redutase subunidade-α, C-terminal	2
PF02783	Metil-coenzime M redutase subunidade-β, N-terminal	1

Apesar da expressiva diferença entre o número de sequências desses dois estudos, essa comparação também corroborou Clostridia como potencial classe produtora de hidrogênio, uma vez a maioria das EGTs preditas de FeFe-hidrogenase identificadas no biodigestor também foi atribuída a esse grupo filogenético.

Ainda pela análise Pfam foram identificadas sequências pertencentes às famílias formato desidrogenase (PF09163), NiFe-hidrogenase (PF00374) e suas proteínas acessórias (PF01155 e PF01924) e de maturação (PF01750) (Tabela 12).

Tabela 13. Distribuição e lista de EGTs atribuídas às subunidades maior e menor de FeFe-hidrogenases identificadas na biblioteca de fosmídeo.

Filo	Classe	Ordem	Gênero	Espécie
PF02906^a – subunidade maior				
Bacteroidetes (21)	Cytophagia (7)	Cytophagales (7)	<i>Microscilla</i> (7)	<i>marina</i> (7)
	Bacteroidia (14)	Bacteroidales (14)	<i>Parabacteroides</i> (8) <i>Bacteroides</i> (6)	
Firmicutes (119)	Clostridia (117)	Clostridiales (92)	<i>Syntrophomonas</i> (2) <i>Clostridium</i> (31)	<i>wolfei</i> (2)
		Halanaerobiales (10)	<i>Halothermothrix</i> (10)	<i>oreonii</i> (10)
Proteobacteria (6)	Deltaproteobacteria (5)	Syntrophobacterales (4)		
		Desulfovibrionales (1)	<i>Desulfovibrio</i> (1)	
	Alphaproteobacteria (1)	Rhizobiales (1)	<i>Rhodopseudomonas</i> (1)	<i>palustris</i> (1)
PF02256^a – subunidade menor				
Bacteroidetes (1)				
Firmicutes (9)	Clostridia (6)	Clostridiales (2)		
Proteobacteria (5)	Deltaproteobacteria (2)	Desulfovibrionales (1)	<i>Desulfovibrio</i> (1)	<i>vulgaris</i> (1)
		Desulfuromonadales (1)	<i>Pelobacter</i> (1)	<i>propionicus</i> (1)
	Alphaproteobacteria (1)	Rhizobiales (1)	<i>Rhodopseudomonas</i> (1)	<i>palustris</i> (1)
	Gammaproteobacteria (2)	Alteromonadales (2)	<i>Shewanella</i> (2)	

^a - Identificação para o acesso da família na base de dados Pfam. Entre parênteses estão listados os números de EGTs atribuídas a cada grupo taxonômico.

Interessantemente, todas as EGTs identificadas como formato desidrogenase, enzima que compõe o complexo FHL, foram pertencentes à *Syntrophus aciditrophicus*. Por outro lado, a maioria das EGTs de NiFe-hidrogenase associadas ao domínio bactéria foi atribuída a *Dechloromonas aromatica*.

O complexo FHL juntamente com NiFe-hidrogenase produzem hidrogênio a partir de formato para equilibrar o pH da fermentação e diminuir a concentração desse composto na célula. Esse processo é descrito em *Escherichia coli*, a qual é um micro-organismo facultativo, e não em *S. aciditrophicus*. McInerney e colaboradores (2007) descreveram que a análise genômica desse micro-organismo revelou um estilo de vida não convencional comparado com o conhecimento da microbiologia. Por outro lado, as hidrogenases descritas em *D. aromatica* foram associadas ao ciclo de fixação de nitrogênio (SALINERO et al., 2009).

Nesse estudo, as EGTs de NiFe-hidrogenase preditas foram atribuídas principalmente a Proteobacteria. Uma análise taxonômica comparativa entre as EGTs de NiFe-hidrogenase preditas nesse estudo e aquelas preditas pelos resultados do pirosequenciamento de uma amostra de biodigestor (JAENICKE et al., 2011) também foi realizada. A maioria das EGTs de NiFe-hidrogenase do biodigestor foi atribuída a *Halothermothrix orenii*, pertencente à classe Clostridia. Esses resultados sugerem a necessidade de estudos mais profundos para a análise de NiFe-hidrogenases, uma vez que essas enzimas estão envolvidas tanto na produção de hidrogênio quanto na sua captação para a produção de metano em ambientes anaeróbicos.

Já os dois tipos de hidrogenases associadas ao domínio Archaea são enzimas que agem preferencialmente na captação de hidrogênio. EGTs de NiFe-hidrogenase (PF00364) foram atribuídas a *Methanosarcina mazei*, e EGTs da coenzima F420-redutase (PF04422 e PF04432) foram associadas a *Methanospirillum hungatei* e *Methanococcus maripaludis*.

A oxidação do hidrogênio é a primeira etapa para a produção de metano pelas arqueias, utilizando a via hidrogenotrófica. EGTs atribuídas a várias enzimas envolvidas nessa via de produção de metano foram identificadas: S-metiltransferase-tetrahydro- metanopterin (PF04208, PF04207 PF04206 e PF02007), ciclohidrolase (PF02289) e Formil metano furano desidrogenase (PF02663) (Tabela 12). Em nível de ordem, Methanomicrobiales foi o maior representante das EGTs preditas para essa via de produção.

Por outro lado, as sequências metagenômicas associadas à via acetoclástica foram: acetoquinase (PF00871), e CO desidrogenase/acetil-CoA sintetase (PF02552 e PF03599), sendo a maioria das EGTs pertencente à Methanosarcinales. Em ambas as rotas, a última etapa da metanogênese é catalisada pela metil-coenzima M redutase (PF02241, PF02249 e

PF02783) (Tabela 12), sendo todas as EGTs preditas para esta enzima associadas a *Methanomicrobiales*.

A análise de todas as EGTs preditas relacionadas à metanogênese confirma que todos os genes das rotas hidrogenotrófica e acetoclástica para a produção de metano estão presentes na comunidade microbiana do lodo de esgoto. Desta forma, esses resultados corroboraram estudos prévios que descreveram *Methanomicrobiales* e *Methanosarcinales* como grupos filogenéticos predominantes nas vias hidrogenotrófica e acetoclástica, respectivamente, assim como principais representantes de arqueias em digestão anaeróbica (STEINBERG; REGAN, 2008; RIVIÈRE et al., 2009; JAENICKE et al., 2011).

5.9 Análise funcional da biblioteca metagenômica de acordo com COG

Além da caracterização taxonômica e funcional das sequências metagenômicas, os *contigs* obtidos também foram utilizados para a identificação de sequências completas de genes e operons. Como mencionado anteriormente, 12.723 *contigs* foram obtidos a partir do conjunto de sequências metagenômicas.

Para a predição e anotação das sequências codificantes, o programa GenDB 2.0, o qual é usualmente empregado na anotação de genomas, foi utilizado. Esse por sua vez, realizou a anotação automática de 21.083 sequências codificantes (CDS) identificadas nos 12.723 *contigs*. Essas sequências codificantes foram classificadas de acordo com as categorias de grupos ortólogos (COG). As classes metabolismo e transporte de carboidratos (G), metabolismo de lipídeos (I), modificação pós-tradução (O) e produção e conversão de energia (C) foram enfocadas nessa análise, uma vez que a matéria orgânica convertida em hidrogênio ou metano é composta por uma mistura de macromoléculas, tais como carboidratos, lipídeos e proteínas, que são convertidas em moléculas mais simples por diversas enzimas produzidas pela comunidade microbiana. Sendo assim, sequências codificantes de enzimas envolvidas na degradação de polissacarídeos, oligossacarídeos e dissacarídeos e relacionadas à classe de metabolismo de carboidratos tais como endoglucanase, glicosidase, xilose isomerase, beta-glucosidase foram identificadas, entre outras (Tabela 14).

Na degradação de proteínas e lipídeos foram identificadas as enzimas subtilisina (COG1404), e esterase (COG0657) e carboxilesterase (COG2272), respectivamente.

A categoria produção e conversão de energia foi analisada com maior ênfase, uma vez que descreve as proteínas envolvidas na produção de hidrogênio ou metano (Tabela 14). As CDS das principais proteínas envolvidas em todas as rotas de produção de hidrogênio descritas na literatura foram identificadas: piruvato-ferredoxina oxirredutase (PFOR), NADH ferredoxina oxidorrredutase (NFOR) e formato-hidrogênio-liase (FHL).

Na primeira via de produção de hidrogênio, o piruvato é clivado pela ação enzimática da PFOR (COG0674, COG1013 e COG1014), gerando acetil-CoA e reduzindo a ferredoxina (COG1146, COG3411 e COG1145). A ferredoxina reduzida é reoxidada pela transferência de seus elétrons a FeFe-hidrogenase (COG4624), direcionando a produção de hidrogênio. A função do NFOR (COG1034) é regenerar o NADH em NAD⁺ e promover a redução da ferredoxina. Já em condições ácidas, o formato pode ser convertido em H₂ e CO₂ pela ação enzimática do complexo FHL (COG1143 e COG0650) e NiFe-hidrogenase (COG0374, COG1740, COG3261 e COG3260) para equilibrar o pH do processo anaeróbico.

Assim como as EGTs de hidrogenases foram analisadas com maior detalhe, as sequências codificantes desses genes acima de 700 bp preditas pelo GenDB foram consideradas para análise, uma vez que a média das sequências metagenômicas esteve em torno de 665 pb. E para a identificação de genes e sequências candidatas a hidrogenases entre todas as sequências preditas, a classificação numérica para as enzimas (EC), COG, e termo GO foram empregados. Dez sequências codificantes foram identificadas como hidrogenases, sendo seis de FeFe-hidrogenase e quatro de NiFe-hidrogenase (Tabela 14). Adicionalmente, a determinação da origem taxonômica e os domínios que compõem essas sequências foram identificados manualmente empregando a pesquisa BLASTX contra os bancos de dados GenBank (NR) e Pfam.

Todas as sequências codificantes de FeFe-hidrogenase foram compostas por diferentes números de domínios e atribuídas a Clostridia (Figura 15, Tabela 15). No entanto, todas foram compostas de um *cluster* de [FeS] e sua subunidade maior, os quais são essenciais na composição desse gene. Essa diversidade no número de domínios que compõem o gene de FeFe-hidrogenase é comum em genes de *Clostridium* (MEYER, 2007), principalmente na subunidade catalítica que apresenta variações significativas de tamanho devido à presença adicional de domínios que acomodam diferentes *clusters* de [FeS].

Essa variação na composição do gene de FeFe-hidrogenase em Clostridia é fundamental, uma vez que permite a esses micro-organismos responderem rapidamente a uma mudança nas condições ambientais (CALUSINSKA et al., 2010).

Tabela 14. Principais categorias de grupos ortólogos de proteínas identificadas nas sequências codificantes do conjunto de *contigs* da biblioteca metagenômica de foscímico.

Acesso COG	Categoria	Número de CDS^a
Metabolismo e transporte de carboidratos (G)		
COG2730	Endoglucanase	3
COG3405	Endoglucanase Y	1
COG3459	Cellobiose fosforilase	11
COG0366	Glicosidases	7
COG2115	Xilose isomerase	2
COG3507	Beta-xilosidase	1
COG4213	ABC-sistema de transporte de xilose	1
COG2160	L-arabinose isomerase	2
COG2407	L-fucose isomerase	6
COG2132	Putativa oxidases	3
COG5520	O- hydrolase glicosil	2
COG345	Alfa-galactosidase	1
COG1874	Beta-galactosidase	2
COG0153	Galactoquinase	5
COG1472	Beta-glucosidase	21
COG5309	Exo-beta-1,3-glucanase	1
Modificação pós-tradução (O)		
COG1404	Proteases (subtilisina)	11
Metabolismo de lipídeos (I)		
COG0657	Esterase/lipase	7
COG2272	Carboxilesterase tipo B	6
Produção e conversão de energia (C)		
<i>Hidrogênio</i>		
COG1146	Ferredoxina	2
COG3411	Ferredoxina	1
COG1145	Ferredoxina	8
COG0674	Piruvato:ferredoxina oxidorreductase: subunidade α	16
COG1013	Piruvato:ferredoxina oxidorreductase: subunidade β	13
COG1014	Piruvato:ferredoxina oxidorreductase: subunidade γ	10
COG1034	NADH desidrogenase	4
COG0650	Formato-hidrogênio-liase, subunidade 4	1
COG1143	Formato-hidrogênio-liase subunidade 6	2
COG0374	NiFe -hidrogenase I subunidade maior	1

COG1740	NiFe -hidrogenase I subunidade menor	1
COG3261	NiFe -hidrogenase III subunidade maior	1
COG3260	NiFe -hidrogenase III subunidade menor	1
COG4624	FeFe- hidrogenase, subunidade maior, C-terminal	6
<i>Metanogêneses</i>		
COG0280	Fosfotransacetilase	3
COG0282	Acetoquinase	5
COG1456	CO desidrogenase/acetil-CoA sintetase subunidade γ	2
COG1614	CO desidrogenase/acetil-CoA sintetase subunidade β	1
COG2069	CO desidrogenase/acetil-CoA sintetase subunidade δ	1
COG1229	Formil metano furano desidrogenase subunidade A	2
COG1029	Formil metano furano desidrogenase subunidade B	2
COG2218	Formil metano furano desidrogenase subunidade C	2
COG1153	Formil metano furano desidrogenase subunidade D	1
COG2191	Formil metano furano desidrogenase subunidade E	4
COG2141	Coenzima F420-dependente N5,N10-metileno H ₄ MP	3
COG1035	Coenzima hidrogenase F420-redutora subunidade β	2
COG1908	Coenzima hidrogenase F420-redutora subunidade δ	4
COG1941	Coenzima hidrogenase F420-redutora subunidade γ	4
COG3259	Coenzima hidrogenase F420-redutora subunidade α	4

A identidade das sequências de FeFe-hidrogenase identificadas nesse estudo quando comparadas às sequências recuperadas do GenBank esteve em torno de 31 a 62% (Tabela 15). Esses resultados sugerem a identificação de novos genes de FeFe-hidrogenase.

Já as sequências codificantes de NiFe-hidrogenases foram compostas somente da subunidade maior, a qual acomoda o sítio ativo. Entretanto, para determinar ao qual dos quatro grupos de NiFe-hidrogenase essas sequências pertencem, é necessário o alinhamento de sequências completas do gene para uma maior discussão da sua função e sua origem taxonômica. Vignais e colaboradores (2001) explicam que essas enzimas consistem de duas subunidades estreitamente associadas a sua história evolutiva.

A produção de metano pode ser realizada a partir de acetato ou H₂/CO₂ por dois grupos únicos de arqueias estritamente anaeróbicas. O acetato, principal substrato da via acetoclástica, pode ser produzido pela ação da acetatoquinase (COG0282) ou/e fosfotransacetilase (COG0280). Também foram identificadas sequências codificantes de CO desidrogenase/acetil-CoA sintetase (COG1456, COG1614 e COG2069), a qual converte o acetil-CoA a metil-tetrahidrosarcinopterina (H₄SP), formando um intermediário na via acetoclástica.

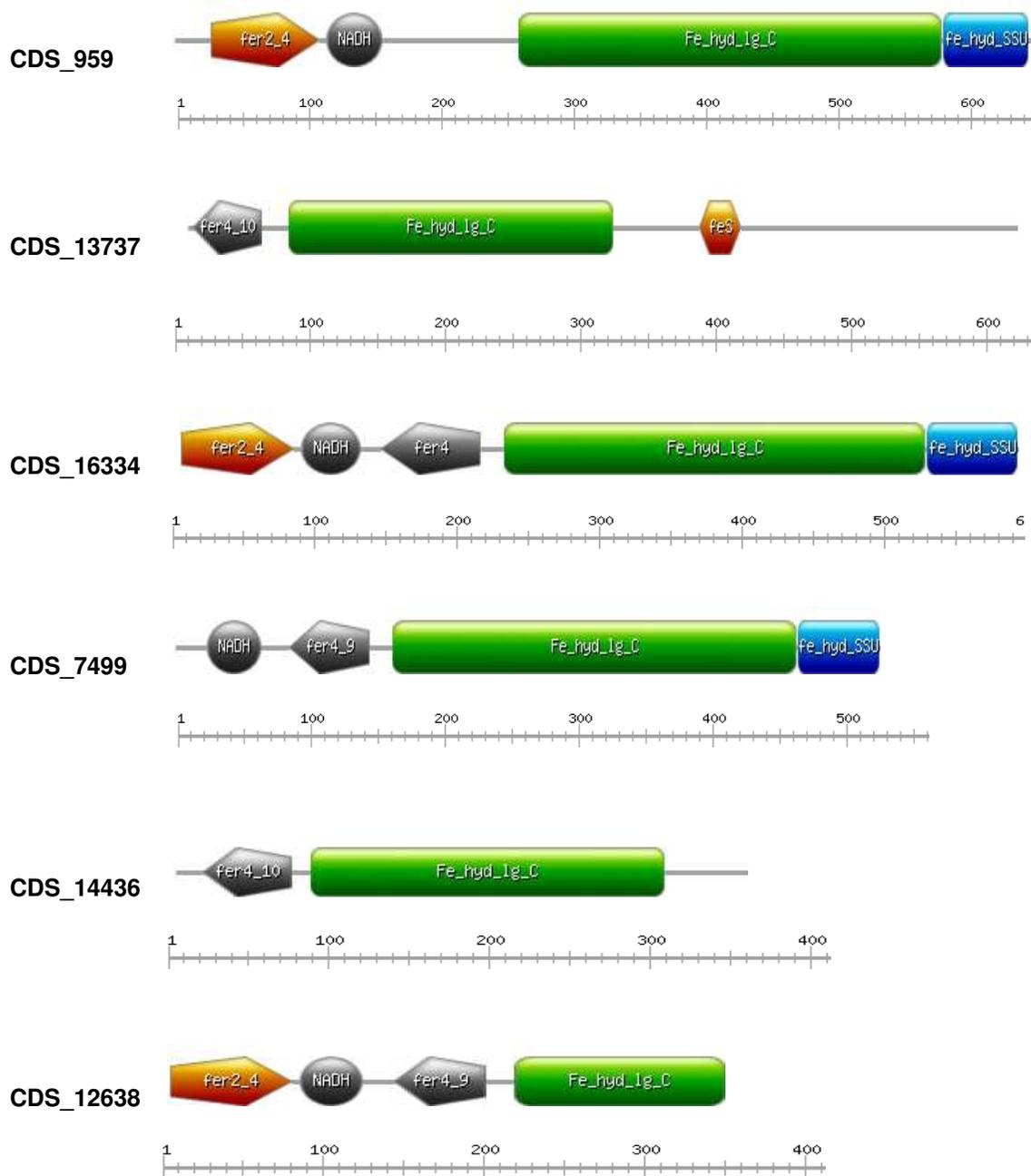


Figura 15. Composição dos domínios das sequências codificantes (CDS) de FeFe-hidrogenase. Domínios conservados e motifs foram identificados em todas as CDS usando InterProScan e Pfam. Os seguintes domínios foram identificados: fer2_4 – pfam13510, 2Fe-2S *cluster* [FeS]; NADH-G₄Fe-4S₃ – pfam10588, [FeS] oxidorreductase; Fe_hyd_Ig_C – pfam02906, FeFe-hidrogenase subunidade maior, domínio C-terminal; Fe_hyd_SS – pfam02256, FeFe-hidrogenase subunidade menor; Fer4_10 – pfam13237, 4Fe-4S domínio *dicluster*; FeS – pfam04060, putativo *cluster* [FeS]; Fer4 – pfam00037, 4Fe-4S domínio; Fer4_9 – pfam13187, 4Fe-4S domínio *dicluster*.

Tabela 15. Classificação taxonômica das sequências codificantes de FeFe-hidrogenase identificadas na biblioteca de fosmídeos.

CDS ^a	Pares de bases	Bactéria relacionada (Acesso GenBank)	E-value	Identidade (%) ^b
959	1878	<i>Caldicellulosiruptor bescii</i> DSM 672 (YP_002573172)	0	51%
		<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903 (YP_001180640)	0	50%
		<i>Clostridium thermocellum</i> DSM 2360 (ZP_05430688)	0	51%
13737	1869	<i>Spirochaeta caldaria</i> DSM 7334 (YP_004696894)	1E-105	31%
		<i>Odoribacter splanchnicus</i> DSM20712 (YP_004254419)	1E-90	31%
		<i>Holophaga foetida</i> DSM 6591 (ZP_09579251)	1E-86	33%
16334	1722	<i>Caldicellulosiruptor kristjanssonii</i> 177R1B (YP_004026404)	0	52%
		<i>Caldicellulosiruptor obsidiansis</i> OB47 (YP_003840536)	0	52%
		<i>Caldicellulosiruptor owensensis</i> OL (YP_004002453)	0	52%
7499	1587	<i>Thermosinus carboxydivorans</i> Nor1 (ZP_01665950)	0	62%
		<i>Desulfotomaculum acetoxidans</i> DSM 771 (YP_003189750)	0	60%
		<i>Acetonema longum</i> DSM 6540 (ZP_08624201)	0	59%
14436	1116	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223 (YP_001664229)	5E-55	31%
		<i>Thermoanaerobacter wiegelii</i> Rt8.B1 (YP_004821126)	1E-54	31%
12638	1086	<i>Pelotomaculum thermopropionicum</i> SI (YP_001212560)	3E-148	58%
		<i>Desulfotomaculum kuznetsovii</i> DSM 6115 (YP_004517644)	5E-139	55%

^a Número de identificação da sequência codificantes (CDS).

^b Porcentagem de identidade de região alinhada.

Na via hidrogenotrófica, a redução do CO₂ e a oxidação do H₂ ocorrem devido à ação de sete diferentes enzimas. Entre essas enzimas foram identificadas neste estudo as seguintes sequências codificantes: Formil metano furano desidrogenase (COG1229, COG1029, COG2218, COG1153 e COG2191), Coenzima F420-dependente N₅,N₁₀-metileno H₄MP (COG2141) e Coenzima hidrogenase F420- redutora (COG1035, COG1908, COG1941 e COG3259). A coenzima hidrogenase F420-redutora é um dos quatro subtipos de NiFe-hidrogenase encontradas no citoplasma de micro-organismos pertencentes ao domínio Archaea que realizam preferencialmente a captação de hidrogênio.

5.10 Triagem por PCR dos clones contendo as CDS de FeFe-hidrogenase

As análises das sequências codificantes (CDS) de FeFe-hidrogenase sugerem que essas sequências compreendem novos genes de FeFe-hidrogenase. Neste sentido, com o intuito de caracterizar esses possíveis novos genes, pares de *primers* foram desenhados para cada sequência codificante de FeFe-hidrogenase e usados para identificar por PCR o clone contendo as sequências descritas nesse estudo (Figura 15).

A triagem por PCR permitiu a identificação de quatro das seis sequências codificantes de FeFe-hidrogenase em cinco clones: dois clones positivos para a CDS_7499; um clone para cada CDS_12638, CDS_14436 e CDS_959 (Figura 16).

A triagem por PCR não foi eficiente na identificação de todas as sequências de interesse ou ocorreu um erro na montagem dos *contigs*. A obtenção de *contigs* a partir de conjunto de dados metagenômicos é difícil devido a vários fatores, tais como: diferentes representações de diferentes micro-organismos em uma comunidade, variação genômica entre micro-organismos relacionados, regiões genômicas conservadas compartilhadas com micro-organismos filogeneticamente distantes e sequências repetidas dentro de um genoma. Apesar das várias ferramentas de bioinformática disponíveis para a montagem de *contigs*, Treangen e colaboradores (2013) explicam que estas ainda estão sendo otimizadas para esse fim.

Os clones identificados foram cultivados para a extração do DNA fosmidial e encaminhados para o sequenciamento na plataforma 454 GS-FLX Titanium (Roche). A montagem de cada fragmento em torno de 40 kb, predição e anotação dos genes, assim como a determinação da origem taxonômica dos fragmentos clonados serão determinadas em breve, assim que os dados finais do pirosequenciamento forem recebidos.

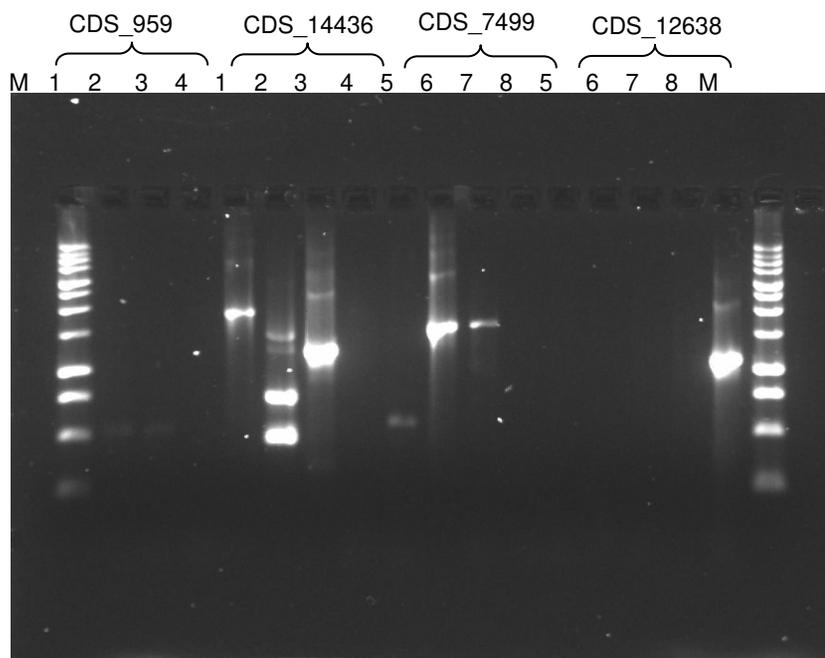


Figura 16. Amplificação das sequências codificantes de FeFe-hidrogenase identificadas na biblioteca de fosmídeos. Amplificação das seguintes CDS: CDS_959; CDS_14436; CDS_7499 e CDS_12638. M, marcador 1 kb; (1), clone_2F; (2), clone_1D; (3), clone_7D; (4) clone_1C; (5), clone_2F; (6), clone_1D; (7), clone_1C; (8), clone_7D.

6. CONCLUSÕES

- I) A análise filogenética das bibliotecas gênicas revelou a presença de novas unidades taxonômicas operacionais pertencentes aos domínios Archaea e Bacteria e novas sequências de genes de FeFe-hidrogenase nas amostras de lodo de esgoto doméstico;
- II) A combinação da análise filogenética e ferramentas de estatística confirmou a alta diversidade da comunidade procariótica, indicando que ainda é necessária a combinação de diferentes abordagens moleculares para o conhecimento detalhado da complexa comunidade microbiana presente em reatores de digestão anaeróbia;
- III) Os *primers* de NiFe-hidrogenase descritos na literatura e aqueles desenhados nesse estudo não foram eficientes para a amplificação desses genes, inviabilizando a análise de sua diversidade diretamente a partir do DNA da amostra;
- IV) O pirosequenciamento de 9.600 clones da biblioteca metagenômica de fosmídeos confirmou a grande diversidade de micro-organismos presente nesse ambiente e que ainda permanece inexplorada devido ao enorme número de sequências metagenômicas não identificadas, mesmo utilizando três abordagens distintas de classificação taxonômica;
- V) A análise do metabolismo da comunidade microbiana revelou que a degradação da biomassa depende de diferentes grupos filogenéticos, evidenciando a alta diversidade de enzimas envolvidas nesse processo biológico e as interações sintróficas entre os grupos microbianos;
- VI) O conjunto de dados metagenômicos indicou seis novas sequências codificantes de FeFe-hidrogenase, das quais quatro puderam ser identificadas em seus clones fosmidiais de origem;
- VII) O conjunto dos resultados permitiu a elucidação da comunidade procariótica na amostra e sugeriu o ambiente como potencial reservatório de novos genes de FeFe-hidrogenase.

7 PERSPECTIVAS FUTURAS

- ✓ Uma análise de bioinformática mais acurada para o desenho de *primers* degenerados de NiFe-hidrogenase eficientes no estudo da diversidade desse gene em amostra ambiental, tendo-se em vista a sua aplicação biotecnológica;
- ✓ Os dados do sequenciamento dos fósdeos contendo as sequências codificantes de FeFe-hidrogenase permitirão a caracterização completa desses operons, visando direcionar posteriormente a sua clonagem em um vetor de expressão para avaliar a sua aplicação biotecnológica na produção de hidrogênio em biorreatores.

8 REFERÊNCIAS

- ABO-HASHESH, M.; WANG, R.; HALLENBECK, P. C. Metabolic engineering in dark fermentative hydrogen production; theory and practice. *Bioresource technology*, v.102, n.18, p. 8414-22, 2011.
- ALTSCHUL, S.; GISH, W.; MILLER, W. Basic local alignment search tool. *Journal of Molecular Biology*, v.215, n.3, p. 403–10, 1990.
- AMANN, R. I.; LUDWING, W. ; SCHLEIFER, K .H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, v.59, p. 143-169, 1995.
- ANDERSON, J.A. The use of tributyrin agar in dairy bacteriology. - Ber. 3. Int. Mikrobiol. Kongress, v. 3, p. 726-728, 1939.
- ARIESYADY, H. D.; ITO, T.; YOSHIGUCHI, K.; OKABE, S. Phylogenetic and functional diversity of propionate-oxidizing bacteria in an anaerobic digester sludge. *Applied microbiology and biotechnology*, v. 75, p. 673–83, 2007.
- BADGER, J.H.; OLSEN, G.J. CRITICA: coding region identification tool invoking comparative analysis. *Molecular Biology and Evolution*, v. 16, n. 4, p. 512-24, 1999.
- BAUER, C.; KORTHALS, M.; GRONAUER, A ; LEBUHN, M. Methanogens in biogas production from renewable resources--a novel molecular population analysis approach. *Water science and Technology*, v.58, p. 1433–9. 2008.
- BÉJÀ, O.et al. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, v. 289, p.1902–1906, 2000.
- BOYD, E. S.; SPEAR, J. R.; PETERS, J. W. [FeFe]-hydrogenase genetic diversity provides insight into molecular adaptation in a saline microbial mat community. *Applied and environmental microbiology*, v. 75, p. 4620–3, 2009.
- CALUSINSKA, M.; HAPPE, T.; JORIS, B.; WILMOTTE, A. The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology*, v.156, p.1575-88, 2010.
- CANTRELL, K.C.; DUCEY, T.; RO, K.S.; HUNT, P. G. Livestock waste-to-bioenergy generation opportunities. *Bioresource Technology*, v. 99, p. 7941-7953, 2008.
- CARDINALI-REZENDE, J. et al. Bacterial and archaeal phylogenetic diversity associated with swine sludge from an anaerobic treatment lagoon. *World journal of microbiology & biotechnology*, v. 28, p. 3187–95, 2012.
- CARNEIRO, P.; FERREIRA, P. The economic, environmental and strategic value of biomass. *Renewable Energy*, v. 44, p. 17-22, 2012.
- CASTRO, A. P. et al. Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content. *Biotechnology letters*, v. 33, n.11, p. 2169-75, 2011.

CHAKRABORTY, R.; O'CONNOR, S. M.; CHAN, E.; COATES, J. D. Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene compounds by Dechloromonas strain RCB. *Applied and Environmental Microbiology*, v. 71, p. 8649, 2005.

CHEN, W.M.; TSENG, Z.J.; LEE, K.S.; CHANG, J.S. Fermentative hydrogen production with *Clostridium butyricum* CGS5 isolated from anaerobic sewage sludge. *International Journal of Hydrogen Energy*, v. 30, p.10,:1063e70, 2005.

CHISTOSERDOVA, L. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology letters*, v. 32, p. 1351–9, 2010.

CHOJNACKA, A. et al. Comparative analysis of hydrogen-producing bacterial biofilms and granular sludge formed in continuous cultures of fermentative bacteria. *Bioresource Technology*, v. 102, p. 10057–64, 2011.

CHON, K. ; KIM, Y.; CHANFG, N.K.; CHO, J. Evaluating wastewater stabilizing constructed wetland, through diversity and abundance of the nitrite reductase gene *nirS*, with regard to nitrogen control. *Desalination*, v. 264, n.3, p. 201-205, 2010.

CHOUARI, R. et al. Molecular analyses of the microbial community composition of an anoxic basin of a municipal wastewater treatment plant reveal a novel lineage of proteobacteria. *Microbial ecology*, v. 60, p. 272–81, 2010.

CHOUARI, R. et al. Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environmental microbiology*, v. 7, p.1104–15, 2005.

CONRAD, R.; ERKEL, C.; LIESACK, W. Rice Cluster I methanogens, an important group of Archaea producing greenhouse gas in soil. *Current opinion in biotechnology*, v. 17, p. 262–7, 2006.

DANHORN, T.; YOUNG, C. R.; DELONG, E. F. Comparison of large-insert, small-insert and pyrosequencing libraries for metagenomic analysis. *The ISME Journal*, p.1–11, 2012.

DEFENG, X. D.; REN, N. ; RITTMANN, B. E. Genetic diversity of hydrogen-producing bacteria in an acidophilic ethanol-H₂-coproducing System, analyzed using the [Fe]- hydrogenase gene. *Apply and Environmental Microbiology*, v. 74, p.1232–1239, 2008.

DELCHER, A.L.; BRATKE, K.A. ; POWERS, E.C.; SALZBERG, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, v. 23, p. 673–679, 2007.

DEMIRBAS, A. Progress and recent trends in biofuels. *Progress in Energy and Combustion Science*, v.33, p. 1-18, 2007.

DESAI, N.; ANTONOPOULOS, D.; GILBERT, J. A.; GLASS, E. M.; MEYER, F. From genomics to metagenomics. *Current opinion in biotechnology*, v. 23, p. 72–6, 2012.

DIAZ, N.; KRAUSE, L.; GOESMANN, A.; NIEHAUS, K.; NATTKEMPER, T. TACO - Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics*, v. 10, p. 56, 2009.

- DINCER, F.; MUEZZINOGLU, A. Odor-causing volatile organic compounds in wastewater treatment plant units and sludge management areas. *Journal Environmental Science Health*, v. 43, p; 1569–1574, 2008.
- DRAKE, H. L.; KÜSEL, K.; MATTHIES, C. Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie van Leeuwenhoek*, v.81, p. 203–13, 2002.
- DUAN, C. J. et al. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal Applied Microbiology*, v. 107, p. 245–256, 2009.
- EKKERS, D. M.; CRETOIU, M. S.; KIELAK, A. M.; ELSAS, J. D. V. The great screen anomaly--a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, v. 93, n. 3, p. 1005-20, 2012.
- EL SADEK, G.M. ;RICHARDS, T. Nile blue, aniline blue and neutral red as indicators of lipolysis. – *Journal Applied Microbiology*, v. 20, p. 137- 144,1959.
- ELREFAEI, H.; BIDA, A.; HALLOUDA, M.; ELSOBKY, M. Renewable Energy Market Competence Index part: The methodology. *Renewable Energy*, v. 53, p. 400-412, 2013.
- EWING, B.; HILLIER, L.; WENDL, M. C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, v. 8, p. 175-185, 1998.
- FELSENSTEIN, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164-166.
- FERRY, J.G. Enzymology of the fermentation of acetate to methane by *Methanosarcina thermophila*. *Biofactors*, v. 6, n. 1, p. 25-35, 1997.
- FIANDT, M. Construction of an Environmental Genomic DNA library from Soil using the EpiFOSTM Fosmid Library Production Kit. *EPICENTRE Forum*, v. 7, n.4, 2000.
- GERLACH, W.; STOYE, J. Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Research*, v. 39, e91, 2011.
- GUPELLIL, A. et al. Hydrolysis of wastewater colloidal organic matter by extracellular from activated sludge flocs. *Water Science Technology*, v. 43, n.6, p. 33-40, 2001.
- GUERMAZI, S. et al. Discovery and characterization of a new bacterial candidate division by an anaerobic sludge digester metagenomic approach. *Environmental microbiology*, v. 10, p.2111–23, 2008.
- GUO, X. M.; TRABLY, E.; LATRILLE, E.; CARRÈRE, H.; STEYER, J.-P. Hydrogen production from agricultural waste by dark fermentation: A review. *International Journal of Hydrogen Energy*, v. 35, n.19, p. 10660-10673, 2010.
- HALLENBECK, P. C. Fermentative hydrogen production: Principles, progress, and prognosis. *International Journal of Hydrogen Energy*, v.34, p. 17, 7379-7389, 2009.
- HALLENBECK, P. C. ; GHOSH, D. Advances in fermentative biohydrogen production: the way forward? *Trends in biotechnology*, v. 27, p. 287–97, 2009.

HANDELSMAN, J.; RONDON, M.R.; BRADY, S.F.; CLARDY, J.; GOODMAN, R.M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, v. 5, p.245– 49, 1998.

HÅRDEMAN, F.; SJÖLING, S. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiology Ecology*, v.59, n.2, p. 524-34, 2007.

HE, Z. et al. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME Journal*, p.67–7, 2007.

HEATH, C.; HU, X. P.; CARY, S. C.; COWAN, D. Identification of a novel alkaliphilic esterase active at low temperatures by screening a meta- genomic library from Antarctic desert soil. *Applied Environmental Microbiology*, v.75, p. 4657–4659, 2009.

HERRMANN, A. : RATH, J. Biogas Production from Maize: Current State, Challenges, and Prospects. 1. Methane Yield Potential. *BioEnergy Research*, v. 5, n. 4, p. 1027-1042, 2012.

HESS, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, v. 331, p.463–7, 2011.

HESSELSON, M. et al. Isotope array analysis of Rhodocyclales uncovers functional redundancy and versatility in an activated sludge. *ISME Journal*, v. 3, p. 1349–64, 2009.

HJORT, K. et al. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiology Ecology*, v.71, n.2, p. 197-207, 2010.

HOFF, K.J.; LINGNER, T.; MEINICKE, P.; TECH, M. Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Research*, v.37, 2009.

HUSON, D. H.; AUCH, A. F.; QI, J.; SCHUSTER, S. C. MEGAN analysis of metagenomic data. *Genome research*, v. 17, p. 377–86, 2007.

JAENICKE, S. et al. Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS one*, v. 6, n.1, e14519, 2011.

JIANG, H. J.; TAY, J. H.; MASZENAN, A. M.; TAY, S. T. L. Bacterial diversity and function of aerobic granules engineered in a sequencing batch reactor for phenol degradation. *Applied Environmental Microbiology*, v. 70, p.6767- 6775, 2004.

JIMÉNEZ, D. J.; MONTAÑA, J. S.; ÁLVAREZ, D.; BAENA, S. A novel cold active esterase derived from Colombian high Andean forest soil metagenome. *World Journal of Microbiology and Biotechnology*, v. 28, n. 1, p. 361-70,2012.

JOGLER, et al. Conservation of proteobacterial magnetosome genes and structures in an uncultivated member of the deep-branching Nitrospira phylum. *PNAS*, v. 108, p 3 1134-1139, 2010.

- JUOTTONEN, H.; GALAND, P. E. YRJALA, K. Detection of methanogenic Archaea in peat: comparison of PCR primers targeting the *mcrA* gene. *Research in Microbiology*, v.157, p. 914-921, 2006.
- KAKIRDE, K. S.; PARSLEY, L. C.; LILES, M. R. Size Does Matter: Application-driven Approaches for Soil Metagenomics. *Soil biology & biochemistry*, v. 42, p. 1911–1923, 2010.
- KALENDAR, R.; LEE, D.; SCHULMAN, A. H. Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Genomics*, v.98, n. 2, p. 137-144, 2011.
- KARAKASHEV, D.; BATSTONE, D.J.; ANGELIDAKI, I. Influence of environmental conditions on methanogenic compositions in anaerobic biogas reactors. *Applied Environmental Microbiology*, v. 71, p.331–338, 2005.
- KASAP, M.; TOROL, S. ; GACAR, G. Comparison of Electrocompetencies of Some Commonly Used Laboratory Strains of *Escherichia coli*. *Journal of Applied Biological Sciences*, v. 2, p. 12-22, 2008.
- KASHYAP, D. R.; DADHICH, K. S.; SHARMA, S. K. Biomethanation under psychrophilic conditions: a review. *Bioresource technology*, v. 87, p. 147–53, 2003.
- KENT, W.J. BLAT - the BLAST-like alignment tool. *Genome Research*, v. 12, n.4, p. 656–64, 2002.
- KIM, D.-H.; KIM, M.-S. Hydrogenases for biological hydrogen production. *Bioresource technology*, v. 108, n. 18, p. 8423-8431, 2011.
- KIM, J. Y. H. ; CHA, H. J. Recent progress in hydrogenase and its biotechnological application for viable hydrogen technology. *Korean Journal of Chemical Engineering*, v.30, p.1–10, 2013.
- KIM, J.; JUNG, H. ; CHA, H. Universal degenerate oligonucleotide-primed-polymerase chain reaction for detection and amplification of NiFe-hydrogenase genes. *Enzyme and Microbial Technology*, v. 42, p. 1–5, 2007.
- KRAUSE, L. et al. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*, v. 36, 2230–2239, 2008.
- KRÖBER, M. et al. Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. *Journal of biotechnology*, v.142, n. 1, p.38-49, 2009.
- LAI, B. et al. A de novo metagenomic assembly program for shotgun DNA reads. *Bioinformatics*, v.28, p. 1455–62, 2012.
- LANE, D.J. 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*. Stackebrandt, E. & Goodfellow, M. John Wiley & Sons, New York, pp. 115-175, 1991.
- LANGMEAD, B.; TRAPNELL, C.; POP, M.; SALZBERG, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, p.10:R25, 2009.
- LARKIN, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics*, v. 23, p.:2947-2948, 2007.

- LASERSON, J.; JOJIC, V.; KOLLER, D. Genovo: de novo assembly for metagenomes. *Journal Computational Biology*, v.18, n.3, p.429–43, 2011.
- LEE, H.-S.; VERMAAS, W.F.J.; RITTMANN, B.E. Biological hydrogen production: prospects and challenges. *Trends Biotechnology*, v. 28, p.262–271, 2010.
- LEFEBVRE, O. et al. T-RFLP reveals high β -Proteobacteria diversity in microbial fuel cells enriched with domestic wastewater. *Journal of applied microbiology*, v.109, n.3, p. 839-50, 2010.
- LI, A. et al. A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor. *Biotechnology for biofuels*, v. 6, 2013.
- LI, J.; SUN, K.; HE, J.; CHEN, Q. Using an Amylase Pretreatment of Pig Manure to Enhance Biogas Production. *Institute of Electrical and Electronics Engineers*, p. 1-5, 2011.
- LILES, M. R. et al. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Applied and Environmental Microbiology*, v. 74, p.3302–5, 2008.
- LIN, C.-Y. et al. Fermentative hydrogen production from wastewaters: A review and prognosis. *International Journal of Hydrogen Energy*, v. 37, n. 20, p. 15632-15642, 2012.
- LINKE, B.; McHARDY, A.C.; NEUWEGER, H.; KRAUSE, L.; MEYER, F. REGANOR: a gene prediction server for prokaryotic genomes and a database of high quality gene predictions for prokaryotes. *Applied Bioinformatics*, v. 5, p. 193–198, 2006.
- LIU, B. et al. Thauera and Azoarcus as functionally important genera in a denitrifying quinoline-removal bioreactor as revealed by microbial community structure comparison. *FEMS Microbiology Ecology*, v.55, p. 274–286, 2006.
- LUTON, P. E.; WAYNE, J. M.; SHARP, R. J.; RILEY, P. W. The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology*, v.148, p. 3521-3530, 2002.
- MAEDA, T.; SANCHEZ-TORRES, V.; WOOD, T.; K. Hydrogen production by recombinant Escherichia coli strains. *Microbial biotechnology*, v. 5, p. 214–25, 2012.
- MANDE, S. S.; MOHAMMED, M. H. ; GHOSH, T. S. Classification of metagenomic sequences: methods and challenges. *Briefings in bioinformatics*, v. 13, p. 669-681, 2012.
- McBRIDE, M.J. et al. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by Genome Sequence Analysis. *Applied Environmental Microbiology*, v. 75, p. 6864–6875, 2009.
- McGLADE, C. E. A review of the uncertainties in estimates of global oil resources. *Energy*, v. 47, p. 262-270, 2012.
- McHARDY, A. C.; MARTIN, H. G.; TSIRIGOS, A.; HUGENHOLTZ, P.; RIGOUTSOS, I. Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods*, v. 4, p. 63–72, 2007.

- McILROY, S. J.; PORTER, K.; SEVIOUR, R. J.; TILLET, D. Extracting nucleic acids from activated sludge which reflect community population diversity. *Antonie van Leeuwenhoek*, v. 96, p. 593–605, 2009.
- McINERNEY, M.J. et al. The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *PNAS*, v.104, n.18, p. 7600-05, 2007.
- MERTENS, R.; LIESE, A. Biotechnological applications of hydrogenases. *Current opinion in biotechnology*, v.15, p. 343–8, 2004.
- MEYER, F. et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, v. 9, p.386, 2008.
- MEYER, F. GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Research*, v. 31, p. 2187–2195, 2003.
- MEYER, J. [FeFe] hydrogenases and their evolution: a genomic perspective. *Cellular and molecular life sciences*, v. 64, n. 9, p. 1063-84, 2007.
- MUDHOO, A.; FORSTER-CARNEIRO, T.; SÁNCHEZ, A. Biohydrogen production and bioprocess enhancement: a review. *Critical reviews in biotechnology*, v.31, n. 3, p. 250-63, 2001.
- MULDER, D. W. et al. Insights into [FeFe]-Hydrogenase Structure, Mechanism, and Maturation. *Structure*, v. 19, p. 1038–52, 2011.
- MYINT, M.; NIRMALAKHANDAN, N.; SPEECE, R.E. Anaerobic fermentation of cattle manure: modeling of hydrolysis and acidogenesis. *Water Research*, v. 41, p. 323–332, 2007.
- NAMIKI, T.; HACHIYA, T.; TANAKA, H.; SAKAKIBARA, Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research*, v.40, n. 20, e.155, 2012.
- NELSON, C.M.; MORRISON, M.; ZHONGTANG, Y. A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresource Technology*, v.102, p. 3730-3739, 2011.
- NOGUCHI, H.; PARK, J.; TAKAGI, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, v.19, p. 5623-5630, 2006.
- NTAIKOU, I.; ANTONOPOULOU, G.; LYBERATOS, G. Biohydrogen Production from Biomass and Wastes via Dark Fermentation: A Review. *Waste and Biomass Valorization*, v.1, p. 21–39, 2010.
- OSAKA, T.; EBIE, Y.; TSUNEDA, S.; Y, INAMORI. Identification of the bacterial community involved in methane-dependent denitrification in activated sludge using DNA stable-isotope probing. *FEMS Microbiology Ecology*, v. 64, p.494–506, 2008.
- O-THONG, S. et al. Optimization of simultaneous thermophilic fermentative hydrogen production and COD reduction from palm oil mill effluent by thermoanaerobacterium-rich sludge. *International Journal of Hydrogen Energy*, v. 33, n. 4, p.1221e31, 2008.

- PACE, N. R.; STAHL, D. A.; LANE, D. J.; OLSEN, G. J. Analyzing natural microbial populations by rRNA sequences. *ASM News*, v. 51, p. 4-12, 1985.
- PAPOUTSAKIS, E. T. Exploring the Capabilities of the Geobiosphere's Microbial Genome. *AlchE Journal*, v.59, p. 688-698, 2013.
- PARAWIRA, W.; READ, J. S.; MATTIASSON, B.; BJÖRNSSON, L. Energy production from agricultural residues: High methane yields in pilot-scale two-stage anaerobic digestion. *Biomass and Bioenergy*, v.32, n. 1, p. 44-50, 2008.
- PARSLEY, L. C. et al. Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiology Ecology*, v. 78, p. 176–187, 2011.
- PELLETIER, E. et al. "Candidatus Cloacamonas acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. *Journal of Bacteriology*, v.190, n.7, p.2572-9, 2008.
- PETROSINO, J. F.; HIGHLANDER, S.; LUNA, R. A.; GIBBS, R. A.; Versalovic, J. Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*, v.55,p. 856–66, 2009.
- PRAKASH, T.;TAYLOR, T. D. Functional assignment of metagenomic data: challenges and applications. *Briefings in bioinformatics*, v.13, n.6, p711-727, 2012.
- RADEMACHER, A. et al. Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing. *FEMS Microbiology Ecology*, v. 79, p. 785-799, 2011.
- RAJENDHRAN, J.; GUNASEKARAN, P. Strategies for accessing soil metagenome for desired applications. *Biotechnology advances*, v. 26, p. 576–90, 2008.
- RASTOGI, G.; RANADE, D. R.; YEOLE, T. Y.; PATOLE, M. S.; SHOUCHE, Y. S. Investigation of methanogen population structure in biogas reactor by molecular characterization of methyl-coenzyme M reductase A (mrcA) genes. *Bioresource Technology*, v. 99, p. 5317 – 5326, 2008.
- RHO, M.; TANG, H.; YE, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research*, v.38, n.20, 2010.
- RIESENFELD, C. S.; SCHLOSS, P. D.; HANDELSMAN, J. Metagenomics: genomic analysis of microbial communities. *Annual review of genetics*, v. 38, p. 525–52, 2004.
- RITTMANN, B. E. Opportunities for renewable bioenergy using microorganisms. *Biotechnology and Bioengineering*, v. 100, n. 2, p. 203-12, 2008.
- RIVIÈRE, D. et al. Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *The ISME Journal*, v. 3, p.700–14, 2009.
- RODRÍGUEZ, E. et al. Molecular analysis of the biomass of a fluidized bed reactor treating synthetic vinasse at anaerobic and micro-aerobic conditions. *Applied microbiology and biotechnology*, v. 93, p. 2181–9, 2012.

- RONDON, R.M. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganism. *Applied Environmental Microbiology*, v. 66, p. 2541-2547, 2000.
- ROTHER, M.; METCALF, W. W. Genetic technologies for Achaea. *Current Opinion Microbiology*, v. 8, p. 745-751, 2005.
- SAITOU, N.; NEI, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology Evolution*, v. 4, p.406-425, 1987.
- SALINERO, K. K. et al. Metabolic analysis of the soil microbe *Dechloromonas aromatica* str. RCB: indications of a surprisingly complex life-style and cryptic anaerobic pathways for aromatic degradation. *BMC Genomics*, v.10, 2009.
- SANAPAREDDY, N. et al. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. *Applied and environmental microbiology*, v. 75, p.1688–96, 2009.
- SCHLOSS, P. D.; HANDELSMAN, J. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, v. 14,p. 303–310, 2003.
- SCHLOSS, P.; HANDELSMAN, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental*, v. 71, p. 1501–1506, 2005.
- SCHLÜTER, A. et al. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *Journal of Biotechnology*, v. 136, p. 77-90, 2008.
- SCHMEISSER, C.; STÖCKIGT, C.; RAASCHA, C et al. Metagenome survey of biofilms in drinking-water networks. *Apply and Environmental Microbiology*, v. 69, p. 7298-7309, 2003.
- SCHMIDT, O.; DRAKE, H. L. ; HORN, M. Hitherto unknown [Fe-Fe]-hydrogenase gene diversity in anaerobes and anoxic enrichments from a moderately acidic fen. *Applied and environmental microbiology*, v. 76, p. 2027–31, 2010.
- SIEBER, J. R. et al. The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environmental Microbiology*, v.12, p. 2289–2301, 2010.
- SILVA, C.C. et al. Monitoring the Bacterial Community Dynamics in a Petroleum Refinery Wastewater Membrane Bioreactor Fed with a High Phenolic Load. *Journal Microbiology Biotechnology*, v. 20, p.17-25, 2010.
- SIMON, C.; DANIEL, R. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology*, v. 77, p. 1153–61, 2011.
- SJÖLING, S.; COWAN, D. A. Metagenomics : Microbial Community Genomes Revealed. *Psychrophiles: from biodiversity to biotechnology*, p.313–332, 2008.
- SORRELL, S.; SPEIRS, J.; BENTLEY, R.; BRANDT, A.; MILLER, R. Global oil depletion: A review of the evidence. *Energy Policy*, v. 38, p. 5290-5295, 2010.

- STAHL, D.A.; AMANN, R. Development and application of nucleic acid probes in bacterial systematics. In E. Stackebrandt and M. Goodfellow (ed.), *Nucleic acid techniques in bacterial systematics*, p. 205–248, 1991.
- STAMS, A. J. M. et al. Metabolic interactions in methanogenic and sulfate-reducing bioreactors. *Water Science Technology*, v.52, p. 13–20, 2005.
- STEINBERG, L. M. ; REGAN, J. M. Phylogenetic comparison of the methanogenic communities from an acidic, oligotrophic fen and an anaerobic digester treating municipal wastewater sludge. *Applied and Environmental Microbiology*, v. 74, p. 6663–71, 2008.
- STRALIS-PAVESE, N. et al. Optimization of diagnostic microarray for application in analyzing landfill methanotroph communities under different plant covers. *Environmental Microbiology*, v. 6, n. 4, p. 347–363, 2004.
- TALBOT, G.; TOPP, E.; PALIN, M. F.; MASSÉ, D. I. Evaluation of molecular methods used for establishing the interactions and functions of microorganisms in anaerobic bioreactors. *Water Research*, v. 42, p. 513–537, 2008.
- TAMURA, K. et al. MEGA5: molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology Evolution*, v. 28, p.2731–2739, 2011.
- TATUSOV, R.L. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, v. 29, p.22–28, 2001.
- TEMPERTON, B. et al. Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal*, v. 3, p. 792–6, 2009.
- TREANGEN, T.J. et al. MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biology*, 2013 . doi:10.1186/gb-2013-14-1-r2
- UHLIK, O. et al. Stable isotope probing in the metagenomics era: A bridge towards improved bioremediation. *Biotechnology advances*, v. 31, p.154–65, 2013.
- van der LELIE, D. et al. The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS one*, v. 7, e36740, 2012.
- VIGNAIS, P. M.;BILLOUD, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chemical reviews*, v.107, p. 4206–72, 2007.
- VIGNAIS, P. M.; BILLOUD, B.; MEYER, J. Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews*, v. 25, n.4, p. 455–501, 2001.
- VIGNAIS, P. M.; COLBEAU, A. Molecular biology of microbial hydrogenases. *Current issues in molecular biology*, v. 6, n. 2, p. 159–88, 2004.

WANG, Q.; GARRITY, G.M.; TIEDJE, J.M.; COLE, J.R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environment Microbiology*, v. 73, p. 5261–5267, 2007.

WENG, L.; RUBIN, .M.; BRISITOW, J. Application of sequence-based methods in human microbial ecology. *Genome Research*, v.16, p. 316-322, 2007.

WILLIAMSON, S. J.; YOOSEPH, S. From Bacterial to Microbial Ecosystems. *Methods in Molecular Biology*, v.804, p.35-55, 2012.

XING, D.; REN, N. ; RITTMANN, B. E. Genetic diversity of hydrogen-producing bacteria in an acidophilic ethanol-H₂-coproducing system, analyzed using the [Fe]-hydrogenase gene. *Applied and Environmental Microbiology*, v. 74, p. 1232–9, 2008.

XU, S.-Y. et al. Hydrogen-Producing Microflora and Fe-Fe Hydrogenase Diversities in Seaweed Bed Associated with Marine Hot Springs of Kalianda, Indonesia. *Current microbiology*, (2013).doi:10.1007/s00284-013-0302-0

YOSHIDA N. et al. Bacterial communities in petroleum oil in stockpiles. *Journal Biosciences Bioengineering*, v. 99, p. 143–149, 2005.

ZAKRZEWSKI, M. et al. Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. *Journal of Biotechnology*, p. 1-11, 2012.

ZAKRZEWSKI, M. et al. MetaSAMS-A novel software platform for taxonomic classification, functional annotation and comparative analysis of metagenome datasets. *Journal of biotechnology*, (2012).doi:10.1016/j.jbiotec.2012.09.013

ZHAO, J.; SONG, W.; CHENG, J.; ZHANG, C. Heterologous expression of a hydrogenase gene in *Enterobacter aerogenes* to enhance hydrogen gas production. *World Journal of Microbiology and Bioetchnology*, v.26, p.177-181, 2010.

ZHOU, J.; BRUNS, M.A. ; TIEDJE, J.M. DNA Recovery from Soils of Diverse Composition. *Applied Environmental Microbiology*, v. 62, n.2, p. 316-22, 1996.

ZIEMIŃSKI, K.; FRĄC, M. Methane Fermentation Process as Anaerobic Digestion of Biomass: Transformations, Stages and Microorganisms. *African Journal of Biotechnology*, v. 11, n. 18, p.: 4127–4139, 2012.

ZWOLINSKI, M.D. DNA Sequencing: Strategies for soil microbiology. *Soil Science Society American Journal*, v.71, p. 592-600, 2007.

WANG, H.; VUORELA, M.; KERÄNEN, A-L.; LEHTINEN, T. M.; LENSU, A.; LEHTOMÄKI, A.; RINTALA, J. Development of microbial populations in the anaerobic hydrolysis of grass silage for methane production. *FEMS Microbiology Ecology*, v. 72, p. 496–506, 2010.

World Energy Council. Disponível em: <http://www.worldenergy.org/>. Acesso 03 de janeiro. 2011.

WAYNE, S.K.; NOGUERA, D.R.; DONOHUE, T. Maximizing reductant flow into microbial H₂ production. *Current Opinion Biotechnology*, v.23, p. 382-389, 2012.

NERIA-GONZÁLES, I.; WANG, E. T.; RAMÍREZ, F.; ROMERO, J. M. ; HERNÁNDEZ-RODRÍGUEZ, C. Characterization of bacterial community associated to biofilms of corroded oil pipelines from southeast of Mexico. *Anaerobe*, v. 12, p.:122-133, 2006.

SOCIEDADE DE ABASTECIMENTO DE ÁGUA E ESGOTO E SANEAMENTO S/A (SANASA). Disponível em: <http://www.sanasa.com.br/>. Acesso 15 de março, 2013.