



Fernanda Buchli

**ANÁLISE METAPROTEOGENÔMICA DE COMUNIDADES BACTERIANAS
ENRIQUECIDAS VISANDO A BIOPROSPECÇÃO DE ENZIMAS DE INTERESSE
BIOTECNOLÓGICO**

***PROSPECTION OF BIOTECHNOLOGICAL ENZYMES THROUGH
METAPROTEOGENOMIC ANALYSIS OF A MICROBIAL CONSORTIUM***

CAMPINAS

2014



UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Biologia



Fernanda Buchli

**Análise metaproteogenômica de comunidades bacterianas enriquecidas visando a
bioprospecção de enzimas de interesse biotecnológico.**

**Prospection of biotechnological enzymes through metaproteogenomic analysis of a microbial
consortium.**

*Dissertação apresentada ao Instituto de Biologia da
Universidade Estadual de Campinas como parte dos
requisitos exigidos para a obtenção do título de Mestra,
em Biologia Funcional e Molecular, na área de bioquímica.*

**Dissertation presented to the Institute of Biology of
the University of Campinas in partial fulfillment of the
requirements for the degree of Master, in
Functional and Molecular Biology, in the area of biochemistry.**

Supervisor/Orientador: DR. FABIO MARCIO SQUINA

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA DISSERTAÇÃO DEFENDIDA PELA ALUNA FERNANDA BUCHLI,
E ORIENTADA PELO DR. FABIO MARCIO SQUINA.**

CAMPINAS

2014

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

Buchli, Fernanda, 1989-
B853a Análise metaproteogenômica de comunidades bacterianas enriquecidas visando a bioprospecção de enzimas de interesse biotecnológico / Fernanda Buchli. – Campinas, SP : [s.n.], 2014.

Orientador: Fabio Marcio Squina.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Metagenômica. 2. Metaproteômica. 3. Bioetanol. I. Squina, Fabio Marcio. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Prospection of biotechnological enzymes through metaproteogenomic analysis of a microbial consortium

Palavras-chave em inglês:

Metagenomic
Metaproteomic
Bioethanol

Área de concentração: Bioquímica

Titulação: Mestra em Biologia Funcional e Molecular

Banca examinadora:

Fabio Marcio Squina [Orientador]
Fernando Segato
Claudio Chrysostomo Werneck

Data de defesa: 02-04-2014

Programa de Pós-Graduação: Biologia Funcional e Molecular

Campinas, 2 de abril de 2014

BANCA EXAMINADORA



Dr. Fabio Marcio Squina



Dr. Fernando Segato



Prof. Dr. Claudio Chrysostomo Werneck

Dr. Jörg Kobarg

Dr. André Ricardo de Lima Damásio

Resumo

A Biomassa vegetal tem sido reconhecida como uma potencial fonte de açúcares fermentescíveis para a produção de biocombustível, principalmente pelo crescente incentivo do uso de fontes renováveis de combustíveis e sustentabilidade. Atualmente, no Brasil, o etanol é quase exclusivamente produzido pela fermentação da sacarose, um açúcar que pode ser facilmente extraído da cana-de-açúcar, esse etanol produzido a partir da sacarose é chamado de etanol de primeira geração. O processo de extração dos açúcares da biomassa da cana, através da hidrólise enzimática para a produção do chamado etanol de segunda geração, ainda apresenta um baixo rendimento e elevado custo de produção. O objetivo deste trabalho foi à busca por enzimas capazes de promover uma degradação mais eficiente, contribuindo para a viabilidade da produção do bioetanol de segunda geração. Para este estudo foi utilizada uma abordagem de metagenômica e metaproteômica. A análise metagenômica baseou-se em uma amostra de solo de canavial a qual teve seu DNA extraído e sequenciado. Em paralelo utilizou-se este solo para o estabelecimento de dois consórcios microbianos utilizando o bagaço de cana-de-açúcar como única fonte de carbono, estes consórcios também foram sequenciados. As sequências foram anotadas e analisadas na plataforma MG-Rast. Para a abordagem metaproteômica foram utilizadas proteínas extraídas diretamente do solo e o secretoma de ambos os consórcios. A análise do sequenciamento revelou a predominância de bactérias, que representaram 94,86% do metagenoma de solo de canavial, sendo o filo Proteobacteria o grupo mais abundante em todos os metagenomas avaliados. Durante as análises foi possível observar mudanças populacionais entre os metagenomas, a exemplo, as classes Bacteroidia, Alphaproteobacteria e Gammaproteobacteria se mostraram mais abundantes nos metagenomas dos consórcios do que do solo. Analisando as proteínas identificadas nas análises de metaproteômica pertencentes à família das glicosil hidrolases nota-se uma predominância das hemicelulases seguida das celulases entre as enzimas mais abundantes identificadas para as três comunidades analisadas. Dentre as celulases identificadas as mais abundantes foram a GH1, GH3 e GH9, entres as hemicelulases as mais abundantes foram GH2, GH43 e GH51. As análises de metagenômica e metaproteômica sugerem que os consórcios apresentam um enriquecimento das enzimas de interesse e revelam o potencial destas comunidades para prospecção de novas enzimas envolvidas na degradação da biomassa lignocelulósica.

Abstract

Plant Biomass has been recognized as a potential source of fermentable sugars for biofuel production, mainly by the growing concern with renewable fuels and sustainability. Ethanol is currently almost exclusively produced by fermentation of sucrose, a sugar that can be easily extracted from sugar cane and thus, this process is called first generation ethanol. The process of extracting the sugars from sugarcane biomass, through enzymatic hydrolysis to produce the so-called second-generation ethanol, still has a low yield and high cost process. The objective of this study was to search for enzymes capable of promoting a more efficient degradation, making possible the production of second generation bioethanol. We used the metagenomic and metaproteomic approaches. The metagenomic analysis was based on a soil sample of sugar cane which had its DNA extracted and sequenced. In parallel the soil was used to establish two microbial consortia using sugarcane bagasse as a sole carbon source, these consortia were also sequenced. The sequences were annotated and analyzed in MG-Rast platform. Proteins extracted directly from soil and the secretome of both consortia were used for metaproteomic approach. The sequencing analysis revealed the predominance of bacteria, representing 94.86 % of the soil metagenome, phylum Proteobacteria is the most abundant group in all metagenomas reviews. During the analysis it was observed population changes between the metagenomes, we notice some groups that seem to be more abundant in consortia's metagenomes than in soil. Between these enriched classes of microorganisms we have the Bacteroidia, Alphaproteobacteria and Gammaproteobacteria classes. Among the proteins identified in the metaproteomic 61% of the soil's proteins represent glycosyl hydrolases and 25% glycosyl transferases, the consortia presented a similar profile. Analyzing the enzymes belonging to the family of glycoside hydrolases we can notice a predominance of hemicellulases then cellulases among the most abundant enzymes identified for the three communities. Among the most abundant cellulases identified were the GH1, GH3 and GH9, among hemicellulases the most abundant were GH2, GH43 and GH51. The metagenomic and metaproteomic analyzes suggest that consortia have an enrichment of the enzymes of interest and reveal the potential of these communities to search for new enzymes involved in the degradation of lignocellulosic biomass.

SUMÁRIO

| | |
|---|-------------|
| RESUMO | VII |
| ABSTRACT | IX |
| AGRADECIMENTOS | XIII |
| LISTA DE ILUSTRAÇÕES | XV |
| LISTA DE TABELAS | XVII |
| 1. INTRODUÇÃO | 1 |
| 1.1. BIOETANOL NO BRASIL | 1 |
| 1.2. BIOTECNOLOGIA E O BIOETANOL | 2 |
| 1.3. BIOMASSA LIGNOCELULÓSICA | 4 |
| 1.4. ENZIMAS QUE DEGRADAM A BIOMASSA | 9 |
| 1.5. ORGANISMOS E A CONVERSÃO NATURAL DA BIOMASSA | 14 |
| 1.6. METAGENÔMICA E METAPROTEÔMICA | 16 |
| 1.7. BIOLOGIA DE SISTEMAS E BIOLOGIA SINTÉTICA..... | 20 |
| 2. JUSTIFICATIVA E OBJETIVOS | 23 |
| 3. METODOLOGIA | 25 |
| 3.1. COLETA DE AMOSTRA DO SOLO | 25 |
| 3.2. CONSÓRCIOS DEGRADADORES DE BAGAÇO DE CANA-DE-AÇÚCAR | 25 |
| 3.3. EXTRAÇÃO DE PROTEÍNAS DO SOLO | 26 |
| 3.4. CONCENTRAÇÃO DAS PROTEÍNAS DO SOBRENADANTE..... | 26 |
| 3.5. EXTRAÇÃO DO DNAMET DO SOLO | 26 |
| 3.6. SEQUENCIAMENTO E ANÁLISE DE DADOS METAGENÔMICOS | 27 |
| 3.7. MG-RAST | 28 |
| 3.8. ANÁLISE PROTEÔMICA | 30 |
| 4. RESULTADOS E DISCUSSÃO | 33 |
| 4.1. EXTRAÇÃO DO DNAMET | 33 |
| 4.2. EXTRAÇÃO DE PROTEÍNAS DO SOLO E SECRETOMA..... | 33 |
| 4.3. SEQUENCIAMENTO | 35 |
| 4.4. ANÁLISE DO PERFIL TAXONÔMICO DO METAGENOMA DO SOLO DE CANA-DE-AÇÚCAR E COMUNIDADES MICROBIANAS DERIVADAS | 37 |
| 4.5. ANÁLISE DO PERFIL FUNCIONAL DO METAGENOMA DO SOLO DE CANA-DE-AÇÚCAR E COMUNIDADES MICROBIANAS DERIVADAS | 50 |
| 4.6. ANÁLISE DO PERFIL METAPROTEÔMICO DO SOLO DE CANA-DE-AÇÚCAR E COMUNIDADES MICROBIANAS DERIVADAS | 53 |
| 4.7. PROSPECÇÃO DE GENES DE INTERESSE E ESTUDOS FUTUROS | 61 |
| 5. CONCLUSÃO | 65 |
| 6. REFERÊNCIAS | 67 |
| 7. ANEXOS | 75 |

AGRADECIMENTOS

Ao Pesquisador Dr. Fabio Marcio Squina, pela oportunidade, por me aceitar como aluna de mestrado, pelas dicas sempre pertinentes, ensinamentos, experiência e pelos direcionamentos do projeto.

Ao quase Doutor João Paulo Lourenço Franco Cairo de Orleans e Bragança Cerqueira Leite, por ser praticamente meu co-orientador, advogado, por todos os ensinamentos, paciência, viagens e principalmente pela eterna amizade.

À UNICAMP, através do Instituto de Biologia e a pós graduação em Biologia Funcional e Molecular.

Ao Laboratório Nacional de Ciência e Tecnologia do Bioetanol pelo suporte técnico.

A Oklahoma State University em especial ao Dr. Rolf Prade pelo apoio e dedicação ao projeto. Aos bioinformatas Brian Couger e Diego Mauricio Riano Pachon que realizaram a anotação dos metagenomas e montagem dos bancos de dados.

Aos membros da banca de qualificação Dr. Fernando Segato, Dr. Claudio Werneck e Dra. Cristiane Sanchez Farinas obrigada por todas as sugestões.

Aos queridos amigos Érico, Ricardo e Pedro que me auxiliaram com a Bioinformática e estavam sempre prontos para criarem programinhas para salvarem a minha vida!!!

A toda equipe do Laboratório de Biologia Molecular do CTBE, pelas risadas, companheirismo e apoio a minha pesquisa. A Thabata por ter sido minha primeira “chefinha” e por tudo que aprendi com ela, pela amizade, conversas e desabafos. Ao Douglas pelas dicas e pela sua incansável dedicação a auxiliar a todos do laboratório, assim como os queridos técnicos que já se foram: Rodrigo, Ana e Dani. Obrigada!!!

À Zaíra coração gelado, que no fundo é uma menina maravilhosa, competente e muito metida que todos amamos. À Fernanda (Branca de Neve) pelo companheirismo, amizade e incentivo na corrida. À Carla (marida) pelas conversas, amizade e sua

felicidade contagiante. À Rosana pela amizade. À Rebeca pela ordem no laboratório, joguinhos no whatsapp e amizade. À Lívia pelas dicas de moda, amizade e compreensão. Ao Beto e Evandro pelas risadas no almoço. Ao André e Thiago pelas dicas e conversa. À toda equipe do CTBE. Obrigada a todos vocês !!!!

A equipe do Laboratório de Espectrometria de Massas do LNBio em especial a pesquisadora Adriana Paes Leme. Obrigada a todos!

À querida Aline que me acompanhou nessa conquista e estava sempre disposta a me ouvir e incentivar.

Ao CNPQ e FAPESP pelo suporte financeiro para o desenvolvimento da dissertação.

E principalmente à minha família. Mãe, Pai, Flavinha e Edward muito obrigada pela dedicação, amor e apoio incondicional de sempre! Amo vocês!

Lista de Ilustrações

| | |
|---|----|
| Figura 1: Esquematização da parede celular primária, secundária (camadas s1, s2 e s3) e lamela média (TAIZ E ZEIGER, 2009)..... | 6 |
| Figura 2: Esquema de parede celular primária, evidenciando as microfibras de celulose, a hemicelulose e pectina (TAIZ E ZEIGER, 2002)..... | 7 |
| Figura 3: Esquema de ação das celulases. Nota-se que as exo-glucanases clivam a cadeia pelas extremidades, enquanto as endo-glucanases promovem quebras pela região central da cadeia. Essa região central clivada forma então novas pontas acessíveis as exo-glucanas. Adaptado de A. Koivula et al., 2012. | 11 |
| Figura 4: Esquema de ação das hemicelulases. Adaptado de A. Koivula et al., 2012..... | 13 |
| Figura 5: Extração do DNA metagenômico do solo, MCS e MCF (Gel de agarose 1% corado com brometo de etídio). | 33 |
| Figura6: Eletroforese em gel de poliacrilamida do secretoma das comunidades MCS e MCF. As linhas representam como foi realizada a divisão da lane para a análise metaproteômica..... | 34 |
| Figura7: Eletroforese em gel de poliacrilamida da extração de proteínas do solo. As linhas representam como foi realizada a divisão da lane para a análise metaproteômica..... | 34 |
| Figura 8: Curva de rarefação do sequenciamento, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco M5NR usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. | 36 |
| Figura 9: Curva de rarefação do sequenciamento, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco SEED usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. | 37 |
| Figura 10: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco SEED usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1. | 38 |
| Figura 11: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco M5NR usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1. | 38 |
| Figura 12: Abundância dos filios dos metagenomas nos bancos de dados RDP, SEED e M5NR. | 39 |
| Figura 13: Gráficos das abundância dos filios dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR. | 40 |
| Figura 14: Heatmap dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco RDP usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Foram utilizados também dados de metagenomas disponíveis no servidor MG-Rast para comparação. Foram utilizados 2 metagenomas de gramado do Reino Unido, um metagenoma de floresta do Canadá, 1 metagenoma de floresta de Porto Rico e 1 metagenoma de uma fazenda dos Estados Unidos. | 41 |
| Figura 15: Abundância das Classes dos metagenomas nos bancos de dados RDP, SEED e M5NR..... | 42 |
| Figura 16: Gráficos das abundância das classes dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR. | 43 |
| Figura 17: Abundância das Ordens dos metagenomas nos bancos de dados RDP, SEED e M5NR..... | 44 |
| Figura 18:Gráficos das abundância das ordens dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR. | 45 |
| Figura 19: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco Subsystems usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1..... | 50 |

| | |
|--|----|
| Figura 20: Gráfico perfil funcional nível 1 dos metagenoms MSi, MSp, MCS e MCF de acordo com o banco de dados Subsystems. | 51 |
| Figura 21:Gráfico perfil funcional nível 3 (carboidratos) dos metagenoms MSi, MSp, MCS e MCF de acordo com o banco de dados Subsystems. | 52 |
| Figura 22: Famílias CAZy identificadas pela análise metaproteômica das amostras MS, MCS e MCF. No gráfico apresenta-se o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Trnasferase; PL – Pectato Liase. | 54 |
| Figura 23: Famílias CAZy identificadas pela análise metaproteômica da amostra MS (Metagenoma Solo). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase. | 55 |
| Figura 24: Famílias CAZy identificadas pela análise metaproteômica da amostra MCS (Metaproteômica da Comunidade Shaker) No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase..... | 56 |
| Figura 25: Famílias CAZy identificadas pela análise metaproteômica da amostra MCF (Metaproteômica da Comunidade Fermentador). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase. | 57 |
| Figura 26: Enzimas da Familia GH identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se o número de proteínas únicas identificadas, seguido pelo percentual correspondente. | 58 |
| Figura 27: Celulases identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. | 59 |
| Figura 28: Hemicelulases identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. | 59 |
| Figura 29: Esquema da estratégia de busca por genes de interesse. | 61 |
| Figura 30: Anotação do Contig referente a GH5 pelo programa Geneious. Setas em verde representama predição de genes pelo software glimmer, em azul a predição de ORFs pelo software ORFFinder e em vermelho estão anotados os dominios encontrados nas buscas das sequencias no banco de dados do NCBI. | 63 |
| Figura 31: Anotação do Contig referente a catalase pelo programa Geneious. Setas em verde representama predição de genes pelo software glimmer, em azul a predição de ORFs pelo software ORFFinder e em vermelho estão anotados os dominios encontrados nas buscas das sequencias no banco de dados do NCBI. | 63 |

Lista de Tabelas

| | |
|---|-----|
| Tabela 1 :Comparação dos clones positivos obtidos a partir do número total de clones submetidos ao screening provenientes de bibliotecas metagenômicas construídas com DNAMet resultante ou não da etapa de enriquecimento. Adaptado de NG, 2010..... | 18 |
| Tabela 2: Dados dos sequenciamentos dos metagenomas de solo e dos consórcios degradadores de bagaço de cana-de-açúcar. | 35 |
| Tabela 3: Perfil da abundância normalizada dos domínios dos metagenomas utilizando o banco de dados SEED. Legenda: metagenoma de solo de canavial sequenciado na plataforma 454 (MSp), metagenoma de solo de canavial sequenciado na plataforma ilumina (MSi), Metagenoma Comunidade Shaker (MCS) e Metagenoma Comunidade Fermentador (MCF).. | 38 |
| Tabela 4: Grupos que eram mais abundantes nos metagenomas de solo e que se tornaram menos abundantes nas comunidades enriquecidas, dados analisados com o banco de dados SEED..... | 48 |
| Tabela 5: Grupos que eram menos abundantes nos metagenomas de solo e que se tornaram mais abundantes nas comunidades enriquecidas, dados analisados com o banco de dados SEED..... | 49 |
| Tabela 6: Contigs selecionados para anotação. | 62 |
| Tabela 7: Lista de proteínas encontradas nas análises metaproteômicas da comunidade do solo, o banco de dados utilizado foi o banco do CAZy..... | 75 |
| Tabela 8: Lista de proteínas encontradas nas análises metaproteômicas da comunidade MCS , o banco de dados utilizado foi o banco do CAZy. | 89 |
| Tabela 9: Lista de proteínas encontradas nas análises metaproteômicas da comunidade MCF, o banco de dados utilizado foi o banco do CAZy. | 100 |

1. Introdução

O solo se apresenta como um ambiente promissor para a prospecção de novas enzimas com interesse industrial, principalmente enzimas envolvidas na degradação de materiais lignocelulósicos, devido à presença de micro-organismos capazes de processar este tipo de biomassa (RUBIN, 2008). Os métodos mais tradicionais para acessar a diversidade presente em amostras ambientais baseiam-se no isolamento e cultivo dos micro-organismos em laboratório. No entanto, cerca de 95 a 99.9% dos micro-organismos não são passíveis de cultivo (COWAN et al, 2005). Alternativamente, tem sido crescente a utilização da abordagem metagenômica, um novo campo de investigação que, fazendo uso de ferramentas de biologia molecular e bioquímica, acessa o genoma de micro-organismos incultiváveis.

No entanto, a metagenômica não é a única alternativa para o estudo dos micro-organismos não cultiváveis, pode-se combinar a construção e triagem de bibliotecas e o sequenciamento completo do metagenoma com a análise proteômica dessa comunidade. Dessa maneira passa-se a ter um olhar mais amplo sobre o micro-ambiente e um crescente entendimento dos processos investigados. A análise proteômica de comunidades microbianas complexas é uma nova área que visa o acesso direto ao potencial catalítico da comunidade.

Diante do exposto, este trabalho teve como objetivo relacionar a diversidade genética e o potencial catalítico de uma comunidade proveniente de solo de canal enriquecido (cultura contendo bagaço de cana-de-açúcar como única fonte de carbono) através dos dados de metaproteômica e sequenciamento do metagenoma da mesma.

1.1. Bioetanol no Brasil

A crise energética mundial vem se agravando desde a década de 1970, isso se deve principalmente pela crescente demanda por recursos não renováveis. Segundo a Agência Internacional de Energia, estima-se que os países em crescimento como a Índia e a China demandarão um aumento de 60% no consumo de energia entre 2010 e 2030. Em pleno desenvolvimento econômico a crescente demanda por combustíveis fósseis se torna insustentável, tanto pela progressiva escassez das reservas quanto por pressão da sociedade pelo

uso de energia limpa e sustentável. Arelado a este quadro também temos a problemática dos conflitos sócio-políticos nas regiões de maior concentração das reservas dos combustíveis fósseis, que geram incertezas sobre o abastecimento mundial.

Em meio a este panorama da crise energética é crescente o incentivo dos governos pela busca de alternativas energéticas limpas e renováveis, abrindo caminho para os chamados biocombustíveis. Os biocombustíveis são fontes de energia renovável, derivados de produtos agrícolas como cana-de-açúcar, sorgo e milho; plantas oleaginosas como a mamona; biomassa florestal e outras fontes de matéria orgânica (KIM et al, 2004). Além da possibilidade de serem produzidos de maneira sustentável, os biocombustíveis apresentam uma redução na emissão de gases tóxicos durante a sua queima.

Entretanto, a transformação da matriz energética requer um grande esforço por parte dos governantes, esforço que deve ser tanto de países ricos como dos países emergentes e que se não bem administrado pode levar para uma grave crise no mercado de grãos destas nações. O aumento excessivo sobre a procura poderá fazer o preço dos grãos dispararem nestes países, gerando graves crises às economias bem como uma crise social (SEXTON et al, 2009).

Por outro lado, a produção do bioetanol poderia fazer do Brasil uma referência mundial em tecnologias alternativas. Com mais de quarenta anos de experiência nesta área, possuímos importantes pesquisadores destas alternativas. Já durante a década de 1970, o Brasil desenvolveu tecnologia de ponta e lançou um dos maiores programas de utilização de "combustíveis limpos", o Pró-álcool. Assim, a economia brasileira dispõe do conhecido etanol da cana-de-açúcar e vem também desenvolvendo o bioetanol a partir de biomassa (bioetanol de segunda geração).

1.2. Biotecnologia e o Bioetanol

A biotecnologia é baseada na procura e utilização de sistemas biológicos, organismos e/ou seus derivados para a produção de produtos e serviços. A biocatálise industrial, em particular, tem se desenvolvido como um setor majoritário com aplicações em várias áreas, como no tratamento de resíduos tóxicos, na indústria de produtos de limpeza, alimentos e cosméticos, no processamento de materiais (algodão, couro, papel) e na produção de biocombustíveis entre outras transformações (BULL et al, 2000).

O interesse na utilização das enzimas na indústria, no lugar dos catalisadores químicos, tem relação à sua biodegradabilidade e a crescente importância da questão ambiental em todos os segmentos. Além disso, enzimas podem catalisar reações em condições moderadas de temperatura e pressão e possuem alta especificidade, levando a um menor acúmulo de produtos indesejados (BULL et al, 1999). Os principais entraves para a substituição dos catalisadores químicos pelos biológicos são o elevado custo na produção das enzimas e a dificuldade de se adequar uma determinada enzima a uma condição reacional específica (BULL et al, 2000).

Visando superar os entraves e viabilizar uma indústria sustentável, a biotecnologia busca biocatalizadores mais eficientes. Esta busca vem ganhando força na última década, em especial a busca por biocatalizadores capazes de promover a conversão de biomassa em biocombustíveis (LEE et al, 2008). A ideia de produzir etanol e outros biocombustíveis, a partir de biomassa vegetal, através de processos que permitam alcançar um balanço energético positivo e em uma escala suficientemente grande para ter um impacto significativo que promova a sustentabilidade nunca foi tão defendida como atualmente. Isso porque, a biomassa vegetal é a principal fonte de biopolímeros existente no planeta, o que a torna uma importante fonte de açúcares para a produção de biocombustíveis e outros compostos (LYND et al, 2008). Neste contexto, a biomassa vegetal é vantajosa por se tratar de uma fonte renovável, apresentar propriedades que não resultam em agressão ao meio ambiente, uma considerável redução na emissão de dióxido de carbono e compostos sulfurados quando comparada aos combustíveis fósseis e por apresentar significativo potencial econômico, tendo em vista as previsões de aumento no preço dos derivados de petróleo em um futuro próximo (DEMIRBAS, 2008).

A produção de bioetanol ocorre basicamente a partir dos seguintes tipos de biomassa: açucarada (cana, beterraba), amilácea (milho, trigo, mandioca) e lignocelulósica (resíduos agroindustriais como bagaço de cana, cascas de cereais, resíduos da indústria madeireira, entre outros) (KIM et al, 2004). Entretanto, a conversão de biomassa lignocelulósica em biocombustíveis não é simples e requer a desconstrução dos polímeros que compõem a parede celular em açúcares fermentescíveis por meio da hidrólise enzimática, a partir de processos que englobam o pré-tratamento da biomassa (SIMMONS et al, 2008). O pré-tratamento é fundamental neste processo, visto que é nesta etapa que a estrutura da parede celular vegetal é modificada e as enzimas empregadas na hidrólise enzimática passam a ter melhor acesso ao substrato (SADDLER 1993).

A produção do bioetanol de segunda geração envolve quatro etapas: (1) o pré-tratamento da biomassa, onde são utilizados processos químicos, como lavagens ácidas, e/ou processos físicos, como a explosão a vapor; (2) produção das enzimas (celulases e hemicelulases) a serem utilizadas na hidrólise da parede celular visando à liberação de açúcares fermentescíveis; (3) hidrólise da biomassa pré- tratada; (4) processo de fermentação dos açúcares liberados, hexoses e pentoses (LYND et al, 2005).

Devido ao fato que, historicamente a demanda por biocombustíveis sempre foi relativamente baixa, cada um dos processos necessários para a obtenção de tais produtos se encontram em estágios iniciais de otimização, inviabilizando economicamente sua aplicação em larga escala (RUBIN, 2008). Em relação ao planejamento inicial do governo dos EUA a tecnologia para produção em escala comercial dos biocombustíveis celulósicos está pelo menos três anos atrasada (produção comercial é definida aqui como plantas com capacidade igual ou superior a 20 milhões de litros por ano). As instalações para a produção de biocombustíveis celulósicos estão atualmente em construção nos EUA e tem previsão para começarem a operar em 2014 (REGALBUTO, 2011). Conseqüentemente, investimentos na área de pesquisa e inovação tecnológica são necessários para que os desafios possam ser superados tornando o processo economicamente viável.

Visando a otimização das etapas de hidrólise e fermentação pode-se utilizar ferramentas biotecnológicas tais como sequenciamento massivo e bioinformática, assim como a biologia de sistemas na busca por novos biocatalizadores capazes de superar os desafios de recalcitrância da biomassa e da produção sustentável em larga escala.

1.3.Biomassa Lignocelulósica

Biomassa é toda matéria biológica gerada através de organismos vivos ou recentemente mortos (Biomass energy center). No contexto da produção de energia, biomassa quase sempre se refere à matéria orgânica vegetal gerada a partir da transformação de energia solar em energia química através da fotossíntese e seus derivados. Mas o termo biomassa pode ser aplicado igualmente para matéria biológica de origem vegetal ou animal. Dentre os diversos tipos de biomassa podemos citar: resíduos florestais e agrícolas, resíduos animais (excrementos) e a matéria orgânica contida em resíduos industriais.

Resíduos agroindustriais são gerados em unidades de processamento de culturas agrícolas, alimentares ou agroenergéticas (BRASIL, 2007). Estes resíduos, que historicamente são subaproveitados, vem ganhando destaque à medida que a biotecnologia vem desenvolvendo ferramentas para obtenção de subprodutos de alto valor agregado destes resíduos, como exemplo a partir do bagaço da cana pode-se produzir os biocombustíveis de segunda geração e partir lignina, presente na parede das células deste tipo de biomassa, podemos extrair ácido paracumarico, ácido ferúlico, ácido vanilínico, vanilina e compostos fenólicos (GHATAK et al, 2010; PAIVA et al, 2013) .

Uma das biomassas mais promissoras para produção destes compostos de alto valor agregado é o bagaço de cana-de-açúcar, que é um dos resíduos da produção de açúcar e álcool. O potencial de aplicação do bagaço de cana se dá tanto pelo seu conteúdo energético quanto pela cana-de-açúcar representar a segunda maior produção brasileira, perdendo somente para soja (BRASIL, 2007). O bagaço é constituído de fibras celulósicas moídas e em média representa de 12 a 15% do peso seco da cana. Em termos de produção foram gerados $5,82 \times 10^7$ toneladas de bagaço pela indústria sucroalcooleira em 2005 (BRASIL, 2007). O bagaço pode ser utilizado como matéria prima para produção de rações bovinas e celulose, mas é, atualmente, majoritariamente utilizado na co-geração de energia térmica e elétrica dentro das usinas.

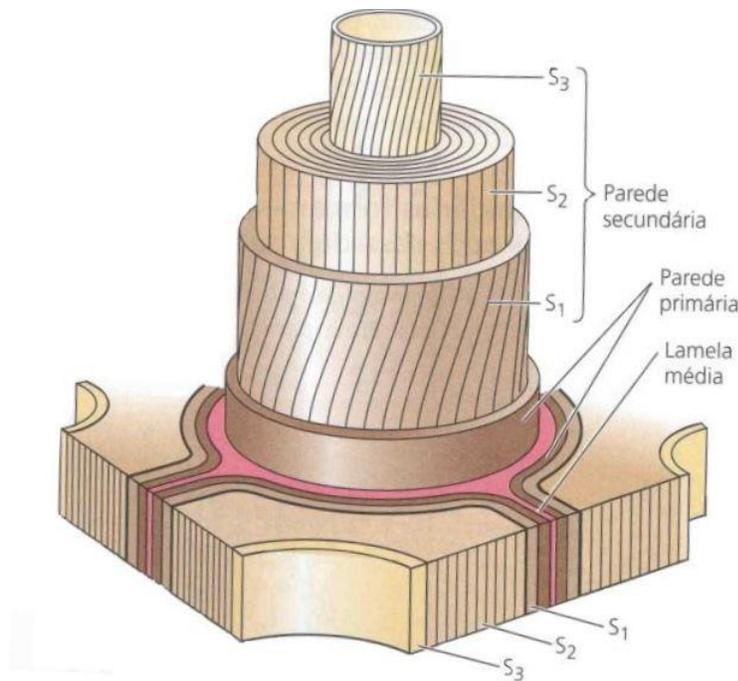
O bagaço da cana-de-açúcar é formado por células vegetais delimitadas por uma parede relativamente delgada, mas mecanicamente forte. Essa parede consiste de uma mistura de polímeros, proteínas estruturais, enzimas e complexos fenólicos (TAIZ E ZEIGER, 2009). A parede celular das plantas além de exercer a função de proteção do protoplasto, e determinação da forma e do volume, também oferece proteção mecânica e obstáculo ao ataque de patógenos (TAIZ E ZEIGER, 2009).

A parede celular pode variar muito em composição e aspecto microscópico dependendo do tipo celular, podendo se apresentar como uma parede mais delgada até possuir muitas camadas e maior grau de complexidade, como nas células do câmbio vascular e outras especializadas (TAIZ E ZEIGER, 2009). A parede celular pode ser primária ou secundária. A parede primária é formada por células em crescimento e apresenta estrutura extensível, permitindo o crescimento da planta e sua estrutura é considerada não especializada. A parede secundária é formada após cessar o crescimento da planta, apresenta um alto grau de diferenciação e está presente em tecidos que conferem a sustentação mecânica da planta. A parede secundária

organiza-se internamente à parede primária apresentando três camadas distintas (s1, s2 e s3) e normalmente ocorre impregnada de lignina. Na Figura 1 observa-se a parede primária e secundária (camadas s1, s2 e s3).

A lamela média (Figura 1) presente na junção entre as paredes primária e secundárias, possui uma composição química diferente das paredes, é rica em pectina, contém proteínas únicas e está relacionada ao transporte de substâncias e adesão da parede (TAIZ E ZEIGER, 2009).

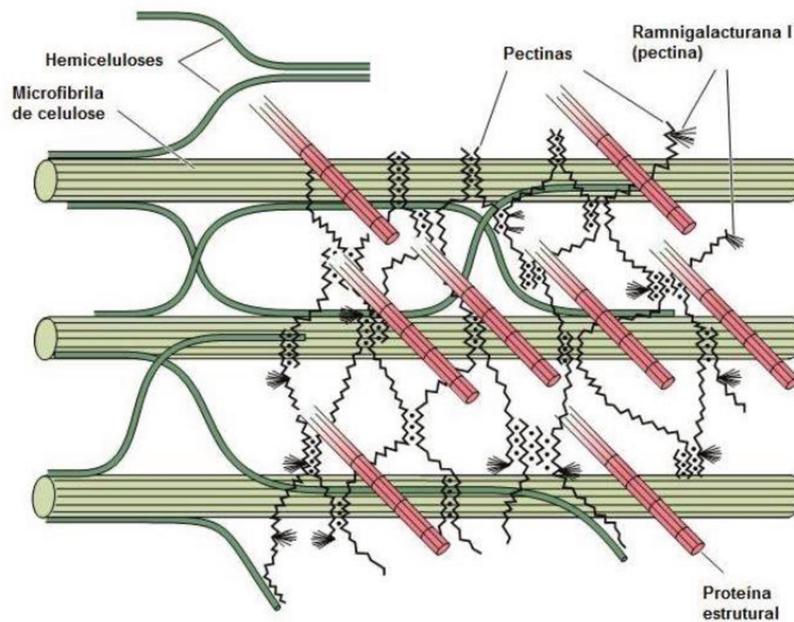
Figura 1: Esquematização da parede celular primária, secundária (camadas s1, s2 e s3) e lamela média (TAIZ E ZEIGER, 2009).



A parede celular primária é composta por microfibrilas de celulose, que estão imersas em uma matriz hidratada de polissacarídeos diversos. Essa matriz é composta principalmente por dois grupos de polissacarídeos, a hemicelulose e a pectina. A Figura 2 mostra a estrutura da parede celular, na qual nota-se as microfibrilas de celulose decoradas com hemicelulose e pectina. A celulose é uma estrutura relativamente rígida, o que confere estabilidade a parede celular. As moléculas de D-glicose estão unidas por ligações glicosídicas do tipo β -1,4 formando microfibrilas que podem ser compostas de 2000 até 25000 resíduos de glicose (BROWN et al, 1996). Essas microfibrilas constituem uma estrutura alinhada em forma de fita formando uma estrutura quimicamente estável e relativamente inacessível ao ataque enzimático (TAIS E ZEIGER, 2009). A

celulose também é formada por regiões com menor grau de organização (porção amorfa), que unem as regiões cristalinas por interações intermoleculares do tipo ponte de hidrogênio (VIETOR et al, 2002). Pelo fato de a configuração da ligação glicosídica entre os resíduos de glicose ser alternada, a unidade de repetição da celulose é a celobiose, um dissacarídeo, composto por duas moléculas de D-glicose com ligações do tipo β -1,4. A celulose pode ser encontrada em estruturas cristalinas variantes dependendo da forma do empacotamento (TAIZ E ZEIGER, 2009 e WEI et al, 2009).

Figura 2: Esquema de parede celular primária, evidenciando as microfibrilas de celulose, a hemicelulose e pectina (TAIZ E ZEIGER, 2002).



Compondo a matriz temos as hemiceluloses, polissacarídeos com certo grau de flexibilidade, que se ligam as moléculas de celulose formando uma rede coesa entre as microfibrilas e também impedindo o contato entre elas, funcionando como uma espécie de revestimento (TAIZ E ZEIGER, 2009). As hemiceluloses constituem um grupo heterogêneo de polissacarídeos, que variam de acordo com o tecido e a espécie em análise. As unidades desses polissacarídeos incluem principalmente aldopentoses, como D-xilose e L-arabinose, e aldo-hexoses como D-glicose, D-manose e D-galactose.

Além da heterogeneidade de monossacarídeos, a hemicelulose também contém substituintes como grupamentos acetila, L-arabinofuranosil, ácido D-glicurônico, ácido D-

galacturônico e ácido α -D-4-O-metilgalacturônico nas ramificações (BON et al, 2008). Os açúcares dos polissacarídeos em geral são unidos por ligações do tipo β -1,4 e ocasionalmente por ligações do tipo β -1,3 (PEREZ et al, 2002). A diversidade de combinações possíveis entre ligações, ramificações e unidades monoméricas contribui para a heterogeneidade da estrutura de hemicelulose que de acordo com o monossacarídeo mais abundante da cadeia pode ter diferentes denominações: xilana, arabinana, galactana ou glicomanana.

As pectinas constituem um gel hidratado que embebe o sistema celulose-hemicelulose, funcionando como um agente hidrofílico que atua impedindo o colapso dessa rede. São compostas por um número heterogêneo de polissacarídeos, contendo tanto açúcares neutros como ramnose, arabinose e galactose, quanto açúcares ácidos como o ácido galacturônico (TAIZ E ZEIGER, 2009).

Esses açúcares formam domínios de polissacarídeos distintos na parede celular, associados por ligações covalentes e intermoleculares. Dentro desses domínios podemos destacar os três principais:

- Homogalacturonana: Também chamada de ácido poligalacturônico, é formada por resíduos de D-ácido galacturônico unidos por ligações do tipo α -1,4, podendo representar até mais de 60% das pectinas na parede celular das plantas (CAFFAL E MOHNEN, 2009).
- Ramnogalacturonato I (RGI): a cadeia base desse domínio é composta por ramnose e resíduos de ácido galacturônico alternantes. Além dessa pode conter cadeias laterais longas de arabinanos e galactanos.
- Ramnogalacturonato II (RGII): contem uma estrutura básica de ácido poligalacturônico, além de cadeias laterais compostas por pelo menos 10 açúcares distintos. É um domínio mais complexo altamente ramificado. Outro aspecto exclusivo é a ligação de éster borato existente entre cadeias laterais de RGII (ISHII et al, 1999), importante para a resistência mecânica da parede celular.

As homogalacturonanas formam géis, redes frouxas formadas por polímeros altamente hidratados (TAIZ E ZEIGER 2009 e CAFFAL E MOHNEN 2009). Esses géis pécticos possuem grupos carbonila (coo-) não-metilados carregados e ligados às cadeias vizinhas via íons de cálcio, formando pontes iônicas.

Já na parede secundária vegetal, a pectina é geralmente substituída por uma macromolécula chamada lignina. A lignina é uma macromolécula fenólica tridimensional com alto grau de complexidade constituída por um polímero de grupos fenilpropanóides altamente ramificados (TAIZ E ZEIGER, 2009). Suas subunidades básicas são álcoois de fenilpropanóides com diferentes graus de metoxilação: cumaril, coniferil e sinapil (MARTINEZ et al, 2005; TAIZ E ZEIGER, 2009). A lignina é encontrada principalmente em tecidos de sustentação e vascularização, podendo ocorrer nas paredes primárias e secundárias (principalmente) e na lamela média, formando juntamente com a hemicelulose, uma matriz amorfa onde as fibras de celulose ficam inseridas (MARTINEZ et al, 2005). A lignina, além de conferir resistência mecânica à planta, permite a condução de água e sais minerais sob pressão negativa e protege os polissacarídeos da parede do processo de degradação, agindo como agente de defesa contra patógenos, insetos e outros herbívoros (HATFIELD E VERMIRIS, 2001). A estrutura precisa da lignina não é conhecida pelo fato de o polímero ser ligado covalentemente ao conteúdo de celulose e hemicelulose, mas sabe-se que as unidades de lignina não aparentam estar ligadas de um modo repetido (TAIZ E ZEIGER, 2009).

1.4.Enzimas que degradam a biomassa

Carboidratos raramente apresentam seu potencial biológico na forma de açúcares livres. Ao invés disso é quando eles estão ligados a outras moléculas ou carboidratos que os açúcares revelam seu potencial. Glicoconjugados, oligossacarídeos e polissacarídeos estão relacionados com diversas funções, desde estrutural ou de reserva até sinalização celular. Por tanto, as enzimas que seletivamente constroem (glicosiltransferases, GTs) e clivam (glicosil hidrolases, GHs, e polissacarídeo liases) ligações glicosídicas, detêm um papel importante em diversas funções celulares (HENRISSAT et al, 2008).

Dentre as proteínas ativas em polissacarídeos são de especial interesse, do ponto de vista da biotecnologia aplicada à produção de bioetanol, as enzimas capazes de hidrolisar esses polissacarídeos ou de ainda desestabilizar e/ou reorganizar esses tipos de cadeias. Os principais produtores destas enzimas são os fungos decompositores, bactérias e animais que se alimentam de compostos lignocelulósicos tais como cracas, baratas e principalmente os cupins (WATANABE E TOKUDA, 2010). Entre essas proteínas destacam-se as hidrolases glicolíticas. Essas enzimas quebram a cadeia polissacarídica gerando principalmente mono, di e/ou oligossacarídeos,

enquanto outras proteínas, chamadas de acessórias, tais como as transferases sacarídicas (glicosil transferases), expansinas (swollenina), carboxilesterases e módulos de ligação a carboidratos (CBMs- “*Carbohydrate Binding Modules*”) auxiliam nas modificações de polissacarídeos.

Visando a utilização eficaz da biomassa lignocelulósica, isto é, aproveita-la de forma integral, adotando métodos e processos que transformem seus principais constituintes em produtos úteis, tanto para a produção de substâncias químicas, como para uma variedade de combustíveis e produtos alternativos se faz necessária à atuação não de uma, mas de um conjunto de enzimas (LYND et al, 2005). Entre as enzimas necessárias para que ocorra uma completa hidrólise da biomassa são fundamentais celulases, hemicelulases, pectinases e proteínas acessórias, que irão auxiliar nas modificações dos polissacarídeos.

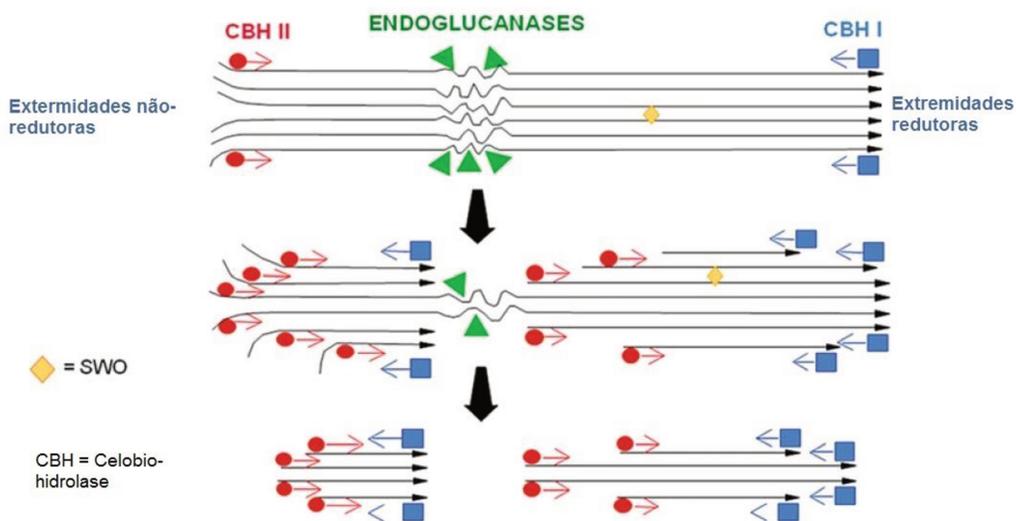
As celulases compõem um grupo de enzimas que, coletivamente, apresentam afinidade para ligações glicosídicas do tipo β -1,4 da fibra de celulose. São enzimas que agem em sinergia (LYND et al, 2002). A hidrólise completa da celulose requer, pelo menos, três grupos enzimáticos distintos:

- Endo- β -1,4 glucanase (E.C. 3.2.1.4) → Hidrolisa cadeias de celulose clivando randomicamente a porção central da cadeia. Atua sobre a região não-cristalina e região amorfa (TEERI, 1997). Endoglucanases são responsáveis pelo encurtamento e quebra da cadeia facilitando a ação das exoglucanases.
- Exo- β -1,4 glucanase como as celodextrinases (1,4- β -glicana-glicanohidrolases, E.C. 3.2.1.176) e as celobiohidrolases, ou CBHs (1,4- β -glicana-celobiohidrolases, E.C. 3.2.1.91) → Hidrolisa sequencialmente a cadeia de celulose removendo as unidades de celobiose das extremidades redutoras e/ou não redutoras gerando glicose e celobiose, respectivamente (TEERI et al, 1998; WOOD E BHAT 1988). As celobiohidrolases ocorrem em duas formas morfológicamente distintas: CBH tipo I, que catalisa a hidrólise de regiões redutoras, e CBH tipo II, que catalisa a hidrólise de regiões não-redutoras.
- β -glicosidases (E.C. 3.2.1.21) → Finaliza o processo de hidrólise clivando os oligossacarídeos e celobiose solúveis em glicose.

Na Figura 3 está representada a ação destas enzimas sobre a fibra de celulose. As enzimas exo-glucanases clivam as pontas terminais da cadeia, enquanto as endo-glucanases clivam a região

central da cadeia, formando novas pontas acessíveis as exo-glucanases (representadas pelas celobiohidrolases I e II). Para que a hidrólise seja finalizada as β -glicosidases clivam os oligossacarídeos e celobiose em glicose. Na Figura 3 é ainda representada a ação das expansinas (swollenina), que atuam ligando-se a microfibrila de celulose e enfraquecendo as ligações de hidrogênio entre os polímeros anti-paralelos de celulose (LEE et al, 2001) promovendo um aumento da eficiência da hidrólise (BAKER et al, 2000).

Figura 3: Esquema de ação das celulasas. Nota-se que as exo-glucanases clivam a cadeia pelas extremidades, enquanto as endo-glucanases promovem quebras pela região central da cadeia. Essa região central clivada forma então novas pontas acessíveis as exo-glucanases. Adaptado de A. Koivula et al., 2012.



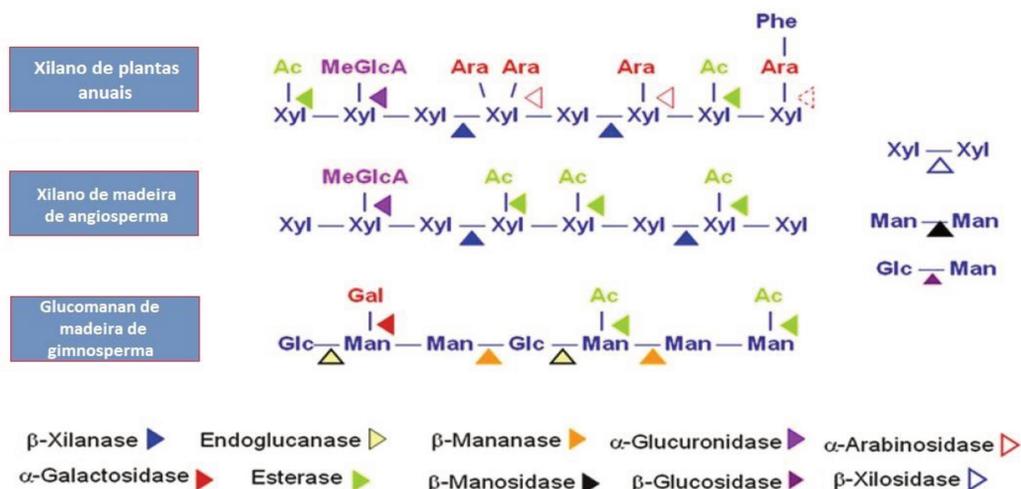
O coquetel enzimático necessário para a hidrólise completa da celulose cristalina requer uma mistura de diferentes tipos de celulasas atuando em sinergismo. Sinergismo se refere à habilidade de a mistura de enzimas obterem uma atividade maior do que a soma das atividades destas enzimas quando utilizadas separadamente.

A maioria das celulasas são enzimas modulares e apresentam ao menos um domínio catalítico e um ou mais módulos de ligação ao carboidrato (CBM – “*Carbohydrate Binding Modules*”) (DAVIES E HENRISSAT, 1995). Além do domínio de ligação ao carboidrato (CBM) as celulasas também podem apresentar domínios de ligações à superfície celular. O CBM é responsável pela concentração da enzima na superfície do substrato aumentando sua atividade em substratos cristalinos sem interferir na atividade sobre substratos solúveis ou amorfos.

Devido às diferenças estruturais cada cadeia polissacarídica necessita de um conjunto de enzimas específico para sua degradação. A exemplo, na hemicelulose a variedade de açúcares nas ramificações, impede a formação de regiões cristalinas, como ocorre na celulose. Essa característica heteropolissacarídica exige um arsenal enzimático mais complexo e específico para a degradação da hemicelulose, as hemicelulases (BON et al, 2008). As hemicelulases são produzidas por diversas espécies de bactérias e fungos e constituem um grupo formado por enzimas como as xilanases, mananases, xilosidases, manosidases e arabinofuranosidases entre outras (KOIVULA et al, 2012). As hemicelulases também trabalham em sinergia, o que aumenta a susceptibilidade do polímero de hemicelulose, ao ataque enzimático. Na Figura 4 observa-se os mecanismos de ação dos principais grupos de hemicelulases que degradam xilana, dentre eles:

- endo-(1,4)- β -xilanases (ou endoxilanases, E.C. 3.2.1.8) \rightarrow Hidrolisa regiões internas da xilana, despolimerizando sua estrutura e gerando polissacarídeos menores. Este grupo atua de forma análoga às endo β -(1,4)-glucanases;
- 1,4- β -xilosidase (E.C. 3.2.1.37) \rightarrow Atua em extremidades não-redutoras da xilana ou em moléculas de xilobiose, e tem como produto de hidrólise a D-xilose. Esse grupo atua de forma análoga às β -glicosidases;
- α -L-arabinofuranosidases (E.C. 3.2.1.55) \rightarrow Enzimas desramificadoras das cadeias laterais ligadas à cadeia principal, que atuam em ligações do tipo α -(1,3) ou α -(1,5) em resíduos de arabinose; outras enzimas desramificadores são as acetil-xilana esterases (E.C. 3.1.1.72), que catalisam a desacetilação de xilo-oligossacarídeos e de resíduos xilopiranosídicos de xilana; as feruloil esterases (E.C. 3.2.1.73), que catalisam a hidrólise entre a xilana e a lignina (ácido hidroxicinâmico); e as α -glucuronidases (E.C. 3.2.1.139), que catalisam a hidrólise de resíduos, como o ácido glucurônico das cadeias laterais (BEG et al, 2001).

Figura 4: Esquema de ação das hemicelulases. Adaptado de A. Koivula et al., 2012.



Embora seja um substrato com menor recalcitrância ao acesso enzimático, a pectina é constituída por um grupo heterogêneo de polissacarídeos. Sua hidrólise requer um grupo de enzimas com atividades distintas, as pectinases. Os principais grupos são (UENOJO E PASTORE, 2007):

- endo- α -(1,4)-poligalacturonases (E.C. 3.2.1.15), responsáveis por catalisar a hidrólise da região interna entre os resíduos de ácido D-homogalacturônico, resultando em polímeros menores;
- exo-poligalacturonase (E.C. 3.2.1.67), que catalisa a hidrólise de ligações α -(1,4) entre resíduos de ácido galacturônico, liberando seus monômeros como produto;
- exo-poligalacturonosidase (E.C. 3.2.1.82), que catalisa a hidrólise de extremidades não-redutoras da pectina, liberando digalacturonato;
- ramnogalacturonase (E.C. 3.2.1.-), que age em ramnogalacturonanas, catalisando a hidrólise entre resíduos de ácido-D- galacturônico e L-ramnose.

A fim de catalogar e unificar essas enzimas foi criado o CAZy (“*Carbohydrate Active Enzymes*”), um banco de dados especializado em enzimas que polimerizam, hidrolisam ou modificam carboidratos. Em setembro de 2013 o banco de dados possuía 159247 sequencias de glicosil hidrolases classificadas em 127 famílias, 119910 sequencias de glicosil-transferases classificadas em 93 famílias, 4043 sequencias de polissacarídeo liases classificadas em 22 famílias, 15856 sequencias de carboidrato esterases classificadas em 16 famílias, 32259 domínios de ligação ao carboidrato (CBM) classificadas em 67 famílias e 5801 sequencias de enzimas auxiliares classificadas em 11 famílias (LOMBARD et al, 2013). Essas famílias foram criadas com base na

caracterização experimental das proteínas e em sequências com similaridade significativa depositadas em bancos de dados públicos (ex: NCBI). As informações bioquímicas das proteínas são reavaliadas constantemente com base na literatura disponível e novas informações da estrutura das proteínas. Mais de 6400 proteínas possuem classificação bioquímica classificada em “EC number” e 700 tem sua estrutura depositada no PDB (“*Protein Data Bank*”, banco de dados de estruturas de proteínas). A classificação com base na estrutura gera melhores resultados do que quando apenas classificada pela afinidade do substrato, possibilitando um melhor entendimento do modo de atuação das enzimas.

Este recurso está disponível para toda comunidade científica a mais de 15 anos, e periodicamente sofre atualizações de acordo com a literatura incluindo novas enzimas e/ou mudando a classificação de outras. Em abril de 2013 o CAZy passou por uma reformulação visando à inclusão de uma nova classe de enzimas, as enzimas de “atividades auxiliares”. Essa inclusão se deve ao fato de a lignina ser, invariavelmente, encontrada juntamente com polissacarídeos na parede celular das plantas e por esses fragmentos de lignina serem propensos a degradação em conjunto com os polissacarídeos, portanto ocorreu a junção das famílias de enzimas envolvidas na degradação de lignina para a família Monooxigenases Líticas de Polissacarídeo (LPMO). Famílias de enzimas de degradação de lignina e as LPMOs foram usadas para definir uma nova classe do CAZy, chamada de 'Atividades Auxiliares', para acomodar uma ampla variedade de mecanismos enzimáticos e substratos relacionados com a conversão de lignocelulose (LOMBARD et al, 2013).

1.5.Organismos e a conversão natural da biomassa

O crescente número de projetos visando o desenvolvimento de processos industriais relacionados à conversão de biomassa em bicomcombustíveis se justifica pela necessidade de tornar essa substituição economicamente viável. Há necessidade de desenvolvimento de ferramentas capazes de degradar a biomassa lignocelulósica de maneira eficaz e a baixo custo.

Para tentar melhorar a eficiência do processo de conversão da biomassa, pesquisadores estão buscando alternativas para a conversão da biomassa na natureza. Sendo a biomassa um composto naturalmente presente em diversos ambientes, vários animais e microrganismos se especializaram ao longo da evolução na sua conversão. Eles utilizam processos bioquímicos

altamente específicos e eficazes para promover sua digestão, fermentação e compostagem (BIOMASS ENERGY CENTER).

Os organismos que utilizam a biomassa estão amplamente distribuídos entre archaea, bactérias, fungos, protistas, plantas e animais (incluindo os microrganismos simbiotes e gastrointestinais). Esses organismos possuem inúmeras enzimas lignocelulolíticas que agem sobre a estrutura da celulose, hemicelulose e lignina. Muitas destas enzimas são secretadas, quer isoladamente ou formando celulosomas, tornando-os assim potenciais biocatalisadores industriais para conversão de biomassa (SWEENEY E XU, 2012).

Dentre os animais podemos destacar os cupins, considerados os menores e mais eficazes biorreatores do planeta (BRUNE, 1998). Dentro de um microlitro do conteúdo intestinal de um cupim, é possível realizar a conversão de material lignocelulósico em monossacarídeos, com um rendimento de 90% (KATSUMATA et al, 2007). Nos últimos cinco anos, uma série de esforços em metagenômica e metatranscriptômica de cupins vem possibilitando a descoberta de novas enzimas envolvidas nessa conversão (FRANCO CAIRO et al, 2011).

Destacam-se na natureza microrganismos, principalmente bactérias e fungos, como produtores de enzimas capazes de degradar a biomassa. Esses microrganismos celulolíticos podem evoluir como degradadores individuais ou como parte de uma "reação em cadeia" em comunidades microbianas complexas de diversos ecossistemas. Enzimas secretadas por esses microrganismos pertencem às classes de glicosil hidrolases (GH), além de enzimas capazes de modificar a lignina. A combinação de enzimas e microrganismos variam de acordo com a fonte de biomassa inicial e fatores ambientais como temperatura, humidade e composição química do ambiente (YANG et al, 2011). Para a degradação eficiente desses polissacarídeos, fungos e bactérias produzem um extenso arsenal de enzimas envolvidas na degradação de polissacarídeo vegetal. São atribuídos aos fungos pelo menos 35 famílias de glicosil hidrolases, três famílias de esterases e seis famílias de liases (VAN DEN BRINK et al, 2011).

Com as ferramentas de biotecnologia emergentes, há um grande potencial para a descoberta de novas fontes de enzimas que oferecem recursos mais desejáveis incluindo atividades específicas, maior sinergismo, melhor estabilidade térmica, maior resistência aos inibidores ambientais e melhor combinação enzimática (por exemplo, celulase, hemicelulase,

atividades pectinase e ligninases) que poderiam maximizar a produção de biocombustíveis a baixo custo.

1.6. Metagenômica e Metaproteômica

Na busca por tornar a produção do etanol de segunda geração economicamente competitiva é crescente o número de trabalhos que buscam melhorias do processo de pré-tratamento do bagaço, maior rendimento da fermentação e a descoberta de novas enzimas que possam tornar a hidrólise mais eficiente. A busca por essas enzimas é realizada nos mais diversos organismos e ambientes como formigueiros (AYLWARD et al, 2012), cupins (FRANCO CAIRO et al, 2011; BURNUM et al, 2010), composteira (MARTINS et al, 2013), solo (GUAZZARONI et al, 2013) entre outros.

O solo vem ganhando destaque como um ambiente promissor para a prospecção de novas enzimas de interesse industrial, principalmente no que diz respeito à degradação do material lignocelulósico. Um dos fatores que contribuí para o destaque do solo na prospecção de enzimas é por ele ter como principal componente de sua biomassa os procariotos. Estima-se que um grama de solo contenha até 18.000 genomas de procariotos, o que representa uma grande quantidade de informação a ser decifrada (DANIEL, 2005).

Mas, o estudo das vias metabólicas desses procariotos não é simples. Métodos tradicionais demandam isolamento e cultivos dos organismos para posterior produção e purificação de enzimas, o que limita as possibilidades de estudo, não apenas pelo tempo e dinheiro envolvidos, mas principalmente pelo fato de que cerca de 95 a 99% dos micro-organismos presentes no ambiente não são cultiváveis (COWAN et al, 2005).

Para superar essa barreira podemos utilizar a metagenômica, que visa à aplicação de técnicas genômicas para estudo de comunidades microbianas de ambientes naturais, sem a necessidade de isolamento (CHEN et al, 2005). Para o estudo, o DNA metagenômico pode ser clonado em vetores apropriados e sequenciados ou massivamente sequenciado através de técnicas de sequenciamento de segunda geração (MEYER et al, 2008). Os fragmentos de DNA obtidos através da metagenômica podem conter genes, operons, ou até mesmo vias metabólicas completas, permitindo acesso a características de micro-organismos ainda não cultiváveis ou de difícil isolamento por métodos tradicionais. Por outro, lado a informação do sequenciamento se

limita aos genes que apresentem regiões conservadas, homólogas as de genes já descritos. Devido a isso, tem sido crescente a utilização conjunta de *screening* de bibliotecas metagenômicas na busca por novas enzimas, pois através do *screening* funcional podemos acessar genes com novos domínios, ainda não descritos.

A metagenômica envolve a extração direta de DNA genômico de amostras ambientais e posterior análise por três métodos principais: primeiramente o DNA metagenômico (DNAmet) é usado como alvo para amplificação das regiões 16S/18S do RNA ribossômico (rRNA) por técnicas de PCR (*“Polimerase Chain Reaction”*), com o intuito de se estudar a diversidade microbiana presente na amostra; posteriormente o DNAmet é utilizado na construção de bibliotecas em vetores e hospedeiros adequados para a busca de novos biocatalisadores, através de abordagens dirigidas por sequências nucleotídicas ou por triagem funcional. Finalmente, o sequenciamento do metagenoma através das tecnologias de sequenciamento em larga escala apresenta-se como a ferramenta mais poderosa para o estudo de microrganismos não cultiváveis. Tais abordagens atuam de forma complementar na prospecção por novos biocatalisadores (DUAN et al, 2010).

Estudos como o conduzido por Hess e colaboradores em 2011 utilizam a metagenômica aliada ao sequenciamento massivo para estudar a decomposição de biomassa no rumem bovino. Neste estudo, a microbiota do rumem foi submetida a sequenciamento Solexa/Illumina gerando 268 gigabases. Utilizando bioinformática foi feita uma remontagem deste metagenoma que previu 2.547.270 sequências de leitura aberta (ORFs – *“Open Reading Frames”*). A média de comprimento das ORFs era de 542 pb, e 55% representam genes completos. Todos os genes previstos foram rastreados na busca por candidatos com potencial atividade enzimática sobre polissacarídeos da parede celular vegetal. Visando minimizar a dependência de similaridade de sequência a enzimas conhecidas os pesquisadores buscaram genes que apresentavam domínios funcionais. Foram identificados 27.755 genes candidatos com pelo menos um domínio catalítico relevante ou um domínio de ligação a carboidratos. Os resultados revelam o potencial enzimático dessa comunidade que contém alguns dos microrganismos mesófilos mais eficientes na decomposição de biomassa de qualquer habitat.

Com o objetivo de aumentar as chances de que clones positivos de interesse biotecnológico sejam encontrados, por exemplo, enzimas como celulasas e hemicelulasas, pode-se enriquecer a microbiota utilizando como única fonte de carbono uma determinada biomassa (NG, 2010). A esse enriquecimento microbiano damos o nome de consórcio. Com base na literatura

pode-se afirmar que bibliotecas construídas a partir de amostras enriquecidas e submetidas ao “*screening*” apresentaram maior número de clones positivos quando comparadas a amostras que não foram submetidas ao enriquecimento (Tabela 1) (FENG et al, 2007; GRANT et al, 2004; KIM et al, 2008; PANG et al, 2009 e REES et al, 2003).

Tabela 1 :Comparação dos clones positivos obtidos a partir do número total de clones submetidos ao screening provenientes de bibliotecas metagenômicas construídas com DNAmet resultante ou não da etapa de enriquecimento. Adaptado de NG, 2010.

| Sem passo de enriquecimento vs Com passo de enriquecimento | Tipo de biblioteca metagenômica | #clones positivos/# clones testados | Gene de Interesse |
|--|--|-------------------------------------|-------------------|
| Sem enriquecimento | Solo | 1 / 70000 | Celulase |
| | Solo composto | 1 / 25000 | Celulase |
| | Água do Lago Nakuru | 1 / 60000 | Celulase |
| | Solo | 1 / 105000 | Oxidoreductase |
| Com passo de enriquecimento | Solo com glicerol e 1,2-propanodiol para micro-organismos consumidores de poliol | 1 / 60000 | Oxidoreductase |
| | Água do lago Nakuru e CMC | 1 / 15000 | Celulase |
| | Solo enriquecido por atividade agarolítica | 1 / 213 | Celulase |
| | Digestores anaeróbios termofílicos alimentados com capim elefante seco e grama-bermudas seca | 1 / 1250 | Celulase |
| | Ceco de coelho alimentado apenas com grama | 1 / 8125 | Celulase |
| | Sedimentos do lago Wadi el Natrun enriquecidos com celulose | 1 / 8750 | Celulase |
| | Solo básico do Wadi el Natrun enriquecido com celulose | 1 / 3083 | Celulase |

No entanto, a construção e triagem de bibliotecas metagenômicas não é a única alternativa para acessar toda essa informação a ser decifrada nos organismos não cultiváveis. Visando explorar de forma mais ampla o que o ambiente tem a oferecer podemos combinar o sequenciamento massivo do metagenoma com a proteômica dessa comunidade. Dessa maneira passa-se a ter um olhar mais amplo sobre o microambiente e um crescente entendimento dos processos investigados.

A análise proteômica de comunidades microbianas complexas é uma nova área que visa o acesso direto ao potencial catalítico da comunidade. Em 2004, Wilmes cunhou o termo metaproteômica como sinônimo de caracterização em larga escala de complexos enzimáticos inteiros, provenientes da microbiota ambiental em um determinado momento. Nesse estudo modelo, as proteínas produzidas por uma microbiota de lodo foram analisadas por eletroforese bi-dimensional e espectrometria de massa (WILMES et al, 2004 e SIMON et al, 2011).

Vários estudos foram conduzidos desde então utilizando a metaproteômica como ferramenta para elucidar diversas comunidades complexas como: intestino de cupins (WARNECKE et al, 2007 e FRANCO CAIRO et al, 2011), rumem de ovelha (TOYODA et al, 2009), fezes humanas (KLAASSENS et al, 2007 e VERBERKMOES et al, 2009) , saliva humana (RUDNEY et al, 2010), amostras marinhas (KAN et al, 2005 e SOWELL et al, 2009), matéria orgânica de lagos e solos de floresta (SCHULZE et al, 2005), solos e águas subterrâneas contaminadas (BENNDORF et al,2007) entre outros.

No primeiro semestre de 2012 Aylward e colaboradores apresentaram seu estudo sobre a caracterização da diversidade microbiana e do potencial biotecnológico de jardins de fungos simbióticos (*Leucoagaricus gongylophorus*) cultivados por formigas do gênero *Atta* utilizando técnicas de metagenômica e metaproteômica. O estudo revelou que a comunidade microbiana residente nos jardins de fungos, é representada principalmente pelos gêneros: *Enterobacter*, *klebsiella*, *Citrobacter*, *Escherichia* e *Panoea*. A presença constante de grupos de bactérias indica que estes possuem um papel importante nos jardins de fungo sugerindo que esses ambientes representam comunidades altamente estruturadas, em vez de uma coleção aleatória de micro-organismos oportunistas. Analisando o potencial funcional do metagenoma da comunidade residente nos jardins encontram-se diversas enzimas relacionadas na degradação da biomassa e biossíntese de vitaminas. Os dados de metaproteômica também evidenciam a presença destas enzimas, confirmando os dados das análises do metagenoma e revelando o potencial desse ambiente na busca por novas enzimas envolvidas tanto na degradação de biomassa quanto na biossíntese de vitaminas.

No entanto, é uma tarefa árdua detectar e identificar todas as proteínas produzidas por uma microbiota complexa no ambiente. Os desafios da metaproteômica incluem a distribuição desigual das espécies, os diferentes níveis de expressão das proteínas e a grande heterogeneidade genética dentro das comunidades microbianas (SCHNEIDER et al, 2010). Mas apesar desses obstáculos, a metaproteômica apresenta um enorme potencial para relacionar a diversidade genética e a atividade microbiana com seu impacto e função no ecossistema.

Para superar ou contornar esses obstáculos é necessário um investimento na criação do banco de dados correto. Isso porque em geral as proteínas que nos interessam não são as mais expressas e com um banco de dados muito abrangente elas podem ficar mascaradas ou até mesmo serem perdidas. A quantidade de informação gerada é muito grande e é necessário

sabermos o que queremos avaliar, somente com as perguntas definidas podemos montar o banco de dados mais correto para avaliação do metaproteoma.

O banco de dados ideal é formado a partir do sequenciamento e anotação dos genes do organismo ou microambiente estudado, sendo por vezes selecionados apenas os genes envolvidos nas rotas investigadas. Mas como na maioria das vezes há o tempo ou o dinheiro necessários, como alternativa ao sequenciamento utiliza-se bancos de dados públicos. Esses bancos são eficientes para um primeiro “*insight*” sobre o metaproteoma, auxiliando principalmente na escolha de qual ambiente/organismo é promissor. Dentre os bancos de dados de domínio público podemos citar o NCBI e o CAZy como ferramentas fundamentais para busca de glicosil-hidrolases.

1.7. Biologia de Sistemas e Biologia Sintética

A Biologia de Sistemas tem emergido na última década como uma das temáticas de maior potencial de crescimento em toda a área das Ciências Biológicas, sendo em muitos casos apontada como a ciência do século XXI. Na sua definição mais simples, biologia de sistemas é o estudo de um organismo ou um conjunto de organismos vistos como uma rede integrada e interativa de genes, proteínas e reações bioquímicas. Ao invés de analisar componentes individuais a biologia de sistemas foca no conjunto dos componentes e nas interações entre eles como parte de um sistema.

Inicialmente pensada nos anos 50, a Biologia de Sistemas só ganhou força recentemente devido ao grande aumento das bases de dados contendo “catálogos” cada vez mais completos de genes, proteínas, mapas metabólicos, e variabilidade biológica em ecossistemas que permitem sustentar a ideia que muitos sistemas e níveis de organização biológica são já suficientemente bem conhecidos para poderem ser abordados de forma sistêmica.

Por outro lado, a biologia sintética utiliza partes bem caracterizadas do sistema para construir um sistema artificial que desempenhe uma nova função. A biologia sintética visa através da re-engenharia de sistemas biológicos a criação de novos sistemas que desempenhem melhor uma dada função.

A biologia de sistemas e a biologia sintética são novos campos de investigação cujos caminhos se cruzam à medida que um gera novas informações para o outro. A biologia de

sistemas nos informa sobre as partes e vias do sistema orientando a sua re-engenharia pela biologia sintética (KOIDE et al, 2009).

2. Justificativa e Objetivos

O bagaço de cana-de-açúcar, assim como outras fontes de biomassa são, produzidas em grandes quantidades pelas usinas sucro-alcooleiras e biorefinarias (CAMPBELL et al, 2010) e atualmente apresentam baixo valor agregado, sendo utilizados na geração de energia para a refinaria. Isso deve-se ao baixo rendimento do uso da biomassa como fonte de açúcares fermentescíveis (obtido pela degradação da celulose e hemicelulose). Em vista disso é item prioritário nos programas de bioenergia de muitas nações, pesquisas que viabilizem a produção de biocombustíveis de segunda geração.

Na natureza, a biomassa é inteiramente reciclada por complexas interações de sistemas microbianos. Isso significa que os polímeros recalcitrantes são decompostos em suas unidades monoméricas (açúcares e fenóis) e utilizados pelos micro-organismos para promover o crescimento vegetativo, produzindo CO_2 e H_2O . Sabe-se que o processo natural de reciclagem integral da biomassa requer um sistema enzimático, em vez de enzimas individuais. Portanto, a nossa proposta sugere entender este sistema enzimático da natureza e determinar a sua base genética.

Neste estudo, visamos relacionar a diversidade genética e o potencial catalítico de uma comunidade microbiana proveniente de solo de canavial. Através dos dados de metaproteômica e análise do sequenciamento espera-se que o arsenal catalítico de micro-organismos especializados na degradação de biomassa lignocelulósica possa ser acessado, possibilitando o uso destas análises na bioprospecção de enzimas envolvidas na degradação de polímeros da parede celular vegetal.

3. Metodologia

3.1. Coleta de amostra do solo

Neste estudo foram utilizadas amostras de solo provenientes de canavial do município de São Carlos - Estado de São Paulo: (GPS: S220477982O 48635299, altitude 600m). O material foi coletado à aproximadamente 2 cm do solo, após a colheita da cana-de-açúcar, enquanto a palha ainda estava sobre o solo, onde espera-se uma microbiota especializada na degradação da biomassa vegetal. As amostras coletadas foram utilizadas nos protocolos descritos ao longo do trabalho, liofilizadas, pulverizadas e estocadas em ultrafreezer a -80°C.

3.2. Consórcios degradadores de bagaço de cana-de-açúcar

Neste estudo foram estabelecidas duas comunidades microbianas enriquecidas (derivada do solo de canavial coletado em São Carlos). Aproximadamente um grama da amostra de solo coletada foi utilizado como inoculo em meio líquido mineral Bushnell Haas Broth, ou meio BHB (BUSHENELL & HAAS, 1941) e bagaço de cana-de-açúcar como fonte de carbono (PAIXÃO et al, 2010). O bagaço utilizado nos experimentos foi o bagaço de cana-de-açúcar pré-tratado por explosão e deslignificado (composição em porcentagem de massa: celulose 89,55; polioses 3,36; lignina 5,47; cinzas 1,44) (RIBEIRO et al, 2012).

Consórcio MCS: Uma comunidade foi mantida em frasco erlenmeyer de 250 ml sob agitação de 150 rpm a 30°C em shaker. Essa comunidade foi mantida com 100 ml de meio BHB, com 0,5% de bagaço de cana-de-açúcar pré-tratado por explosão e deslignificado (BED) como única fonte de carbono, com acréscimo de 0,15% de extrato de levedura. O repique deste consórcio foi realizado semanalmente. Para o repique 2 ml do consórcio eram transferidos para um novo frasco com meio BHB suplementado com 0,5% de BED e 0,15% de extrato de levedura. Na 15ª semana de cultivo o sobrenadante e pellets obtidos foram utilizados nas etapas de concentração de proteínas e extração de DNA metagenômico (DNAm_{et}). A este consórcio foi dado o nome de Metagenoma Comunidade Shaker (MCS).

Consórcio MCF: A segunda comunidade foi estabelecida em fermentador e cultivada com 1 litro de meio BHB com 0,5% de bagaço de cana-de-açúcar (BED) como única fonte de carbono, agitação de 300 rpm da pá e regulação do pH para 5,5. O repique desta comunidade se deu pela técnica de pé-de-cuba onde o consórcio era esgotado até 100 ml com reposição de 900 ml de meio BHB suplementado com 0,5% de BED. Esta comunidade foi enriquecida por 6 semanas

quando o sobrenadante e pellets foram coletados para serem utilizados nas etapas de extração de proteínas e extração de DNAm. A este consórcio foi dado o nome de Metagenoma Comunidade Fermentador (MCF).

3.3.Extração de Proteínas do solo

A extração de proteínas do solo foi realizada com TRIzol® reagent de acordo com o protocolo recomendado pelo fabricante (Disponível em: http://php.med.unsw.edu.au/cellbiology/images/5/5e/TRIzol_Method.pdf). O protocolo baseia-se em lise química e mecânica da parede celular e membranas, seguidas de etapas de precipitação e lavagens com trizol isopropanol e guanidina. As proteínas obtidas foram então quantificadas pelo método de Bradford (BRADFORD, 1976).

3.4.Concentração das proteínas do sobrenadante

Após 15 semanas de crescimento o consórcio crescido em shaker (MCS) foi coletado, filtrado em malha de Nylon (para retenção do bagaço) e centrifugado a 12000 rpm por 1 hora a 4°C. O sobrenadante foi então concentrado em concentrador AMICON (Millipore) com membrana de corte de 10 KDa. O sobrenadante concentrado foi armazenado a 4°C até utilização na análise proteômica.

Posteriormente foi necessário repetir a coleta do sobrenadante para obtenção de mais material para análises complementares. Foram então concentrados 20 ml do sobrenadante da 28ª semana em concentrador Vivaspin (GE Healthcare) de 20 ml com membrana de corte de 10 KDa até 200 µl. O sobrenadante concentrado foi armazenado em a 4°C até utilização na análise proteômica.

Após 6 semanas, uma amostra de 20 ml do consórcio crescido em fermentador (MCF) foi concentrada em concentrador Vivaspin (GE Healthcare) de 20 ml com membrana de corte de 10 KDa até 200 µl. O sobrenadante concentrado foi armazenado a 4°C até utilização na análise proteômica.

3.5.Extração do DNAm do solo

O DNAm do solo (MS) foi extraído e purificado com a utilização do FastDNA® SPIN Kit for soil (MP-Biomedicals), de acordo com o protocolo recomendado pelo fabricante (Disponível em: <http://www.bio101.com/protocols/dna-kits/p-fastdna-soil.html>).

O DNAmet do MCS foi extraído do pellet obtido após a 15ª semana de crescimento do consórcio. O DNAmet foi extraído e purificado com a utilização do FastDNA® SPIN Kit for soil (MP-Biomedicals), de acordo com o protocolo recomendado pelo fabricante (Disponível em: <http://www.bio101.com/protocols/dna-kits/p-fastdna-soil.html>).

O DNAmet do MCF foi extraído do pellet obtido após a 6ª semana de crescimento do consórcio. O DNAmet foi extraído e purificado com a utilização do FastDNA® SPIN Kit for soil (MP-Biomedicals), de acordo com o protocolo recomendado pelo fabricante (Disponível em: <http://www.bio101.com/protocols/dna-kits/p-fastdna-soil.html>).

Todas as amostras foram quantificadas utilizando o espectrofotômetro Nanodrop modelo 2000c (Thermo Scientific), seguindo as orientações de uso (Disponível em: <http://nanodrop.com/Support.aspx?Type=User%20Guides&at=NanoDrop%201000>).

3.6. Sequenciamento e Análise de dados metagenômicos

O sequenciamento do MS foi realizado no Laboratório ENGECORE situado na Universidade da Carolina do Sul nos Estados Unidos, primeiramente através da plataforma 454 GS FLX da Roche. As sequências obtidas foram aparadas e reunidas através do Newbler Assembler Software (version 2.0) da 454 Life Sciences e nomeadas MSp.

Para análise das leituras de sequenciamento obtidas foi utilizado o servidor MG-RAST (disponível em: <http://metagenomics.anl.gov/>), tornando possível a análise taxonômica, funcional e a comparação da amostra sequenciada com outros metagenomas já depositados no servidor (MEYER et al, 2008). Através do MG-RAST foram feitas as anotações das sequências geradas e a classificação filogenética através do SEED (National Microbial Pathogen Database Resource), M5NR (M5 non-redundant protein database), e RDP (Ribosomal Database Project). A predição do perfil funcional foi feita através da comparação com o banco de dados do Subsystem.

Uma segunda amostra do MS, o MCS e o MCF foram sequenciados na plataforma Illumina HiSeq 2000, o sequenciamento foi realizado pela empresa Ambry Genetics, na Califórnia, Estados Unidos e os dados foram analisados em colaboração com o grupo do Dr. Rolf Prade, da Universidade Estadual de Oklahoma (Estados Unidos). As sequências foram reunidas através do software "Velvet", especializado em reunir sequências curtas gerando sequências contíguas ("*contigs*", em inglês) (ZERBINO E BIRNEY, 2008). Devido ao grande volume de dados gerados, todas as análises foram feitas no Pittsburgh Supercomputing Center, associado à Universidade de

Pittsburgh, nos Estados Unidos. As análises funcionais e comparativas foram feitas como descrito anteriormente, através do MG-RAST. As amostras foram nomeadas MSi, MCS e MCF representando metagenoma do solo, do consórcio proveniente do Shaker e do consórcio proveniente do Fermentador respectivamente.

Como foram utilizadas mais de uma plataforma de sequenciamento, gerando quantidades de dados desproporcionais, os dados foram normalizados para que pudessem ser comparados. Para a normalização foi utilizado o “método de contagem de espectros não balanceados”, comumente empregado na análise dos dados experimentais gerados a partir de espectrometria de massas (OLD et al, 2005). A normalização se baseou primeiramente na identificação do número total, denominados N_{MSi} , N_{MSP} , N_{MCS} e N_{MCF} (no caso do MSi, MSP, MCS e MSF respectivamente), de regiões dentro das sequências contínuas que por comparação apresentaram classificação em classes enzimáticas do KEGG. A abundância de cada uma das classes também foi determinada. Posteriormente, foi determinada a média entre os valores de N_{MSi} , N_{MSP} , N_{MCS} e N_{MCF} , denominada M. O valor de M foi então dividido pelo N calculado para cada uma das amostras, obtendo-se os valores de P_{MSi} , P_{MSP} , P_{MCS} e P_{MCF} . A última etapa consistiu na multiplicação dos valores de P pelo número abundância de em cada uma das classes identificadas para cada uma das amostras.

3.7.MG-RAST

MG- RAST (*“Metagenomics Rapid Annotation using Subsystem Technology”*, ou em português, anotação rápida de metagenoma usando a tecnologia Subsystem) é um sistema acessível web que fornece um conjunto de ferramentas para análise e visualização de dados metagenômicos. O sistema é uma adaptação do sistema de servidor RAST, que foi implementado originalmente para permitir a anotação de alta qualidade de genomas de micro-organismo completos utilizando o banco de dados SEED. Os dados de micro-organismos do SEED são ainda usados no algoritmo do MG - RAST, no entanto, numerosos outros recursos foram adicionados ao sistema a fim de aumentar a acurácia das classificações taxonômicas e funcionais. Mais recentemente foram adicionados o GreenGenes, o RDP- II e bancos de dados de RNA ribossômico europeus para permitir a classificação 16s rRNA dos metagenomas avaliados.

O servidor MG-RAST é disponibilizado utilizando a tecnologia estabelecida pelo Argonne National Laboratory e da Universidade de Chicago. É necessário registro no site. A submissão de metagenomas pelos usuários, assim como as análises permanecem confidenciais, no entanto, é

possível tornar os seus dados 'públicos' e compará-los com outros conjuntos de dados públicos. O servidor fornece vários métodos para atender aos diferentes tipos de dados, incluindo reconstruções filogenéticas e metabólicas, além da capacidade de comparar as anotações de um ou mais metagenomas e genomas.

Atualmente, o servidor consegue tratar dados de arquivos em formato fasta sff, e fastq. Arquivos maiores que 50 MB podem ser carregados em formato zip ou gzip. Ambos fasta e fastq precisam ser enviados em formato ASCII texto simples. Informação de qualidade pode ser fornecida para arquivos fasta submetendo-o como um arquivo com o mesmo prefixo. Arquivos de dados de sequência multiplexados pode ser analisado através da apresentação de um arquivo identificador multiplex descritivo (MID), em formato ASCII texto simples.

O servidor do MG-RAST disponibiliza uma série de ferramentas para análise dos metagenomas e comparação com outros metagenomas já depositados no bando, uma dessas ferramentas é a construção de curvas de rarefação da riqueza de espécies anotadas. Esta curva é um gráfico do número total de anotações espécies distintas como uma função do número de sequências de amostra. Na interpretação do gráfico, uma encosta íngreme, à esquerda, indica que uma grande fração da diversidade de espécies continua a ser descoberta. Se a curva se torna mais plana, à direita, um número razoável de indivíduos é amostrado, indicando que uma amostragem mais ampla apresentará apenas poucas espécies adicionais. Estas curvas de rarefação são calculadas a partir da tabela de abundância das espécies. As curvas representam o número médio de diferentes espécies de anotações para sub-amostras do conjunto de dados completo. Com base nas curvas de rarefação podemos, portanto inferir se o sequenciamento foi capaz de evidenciar uma parcela significativa ou não do metagenoma.

Outra análise é a análise de componentes principais (PCA), um método bem conhecido para a redução de dimensões de grandes conjuntos de dados. A Redução de dimensionalidade é um processo que permite que a variação encontrada em um grande conjunto de dados complexos (por exemplo, os valores de abundância de milhares táxons em dezenas de amostras metagenômicas) possa ser reduzida a um número muito menor de variáveis, próprios para a interpretação humana. No contexto do MGRAST, PCA permite que amostras que exibem perfis semelhantes (abundância taxonômica ou funcional) sejam agrupadas o que permite sabermos se os metagenomas avaliados são semelhantes ou diferentes entre si e entre outros metagenomas de domínio público. Na interpretação do gráfico as amostras são consideradas semelhantes

quando se agrupam no mesmo quadrante, quanto mais próximas as amostras maiores as semelhanças.

O heatmap é uma ferramenta que permite uma enorme quantidade de informação (por exemplo, os valores de abundância de milhares de táxons e/ou papéis funcionais em dezenas de amostras metagenômicas) para ser apresentado em uma forma visual. Heatmaps ou dendrogramas são árvores que indicam semelhanças e diferenças de um grupo de vetores (lista de valores, como as contagens a partir de uma única abundância do metagenoma) uns para os outros. Vetores no dendrograma são geralmente ordenados em relação ao seu nível de similaridade: vetores semelhantes são colocados ao lado uns dos outros, enquanto os vetores mais distantemente relacionados são colocados mais afastados. No servidor do MG-Rast temos dois dendrogramas, um indicando a semelhança/diferença entre as amostras metagenômicas (eixo X dendrograma) e outro para indicar a semelhança/diferença entre as categorias (por exemplo, papéis funcionais).

3.8. Análise Proteômica

As análises proteômica das amostras de proteínas extraídas do solo e do secretoma dos consórcios foram submetidas à espectrometria de massas de acordo com trabalho de Wilmes e colaboradores (2009) em espectrômetro de massas LC-MS/MS LTQ-Orbitrap. Os dados foram processados e analisados nas plataformas MASCOT e SCAFFOLD.

Para a análise das proteínas por espectrometria de massas primeiramente submetemos a amostra proteica à separação por SDS-PAGE. Foram aplicados aproximadamente 40 µg das proteínas em um gel poliacrilamida (12%) contendo SDS e β-mercaptoetanol como descrito por (LAEMMLI, 1970). Após a corrida do gel e visualização através da coloração com coomassie blue, o gel foi dividido em seis regiões de acordo com o peso molecular, a primeira banda compreende as proteínas na faixa de 85 a 110 KDa, a segunda banda a faixa de 50 a 85 KDa, a terceira a faixa de 40 a 50 KDa, a quarta a faixa de 26 a 40 KDa, a quinta a faixa de 19 a 26 KDa e a sexta banda incluí as proteínas com peso inferior a 19 KDa, na Figura 6 e na Figura 7 é apresentado um esquema dessa divisão. As bandas foram então descoradas, reduzidas, alquiladas e então submetidas à digestão *overnight* com tripsina (Promega, USA) como descrito por (SHEVCHENKO et al, 1996).

As amostras foram submetidas à espectrometria de massas em espectrômetro LTQ-Orbitrap Velos (Thermo Scientific), escolhido para as análises por apresentar uma maior

sensibilidade. O espectro foi adquirido usando o software Mass Lynx v. 4.1 e os arquivos com dados preliminares foram convertidos para uma lista picos sem somar as análises, pelo software Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.) e comparados no banco de dados do CAZy (com 67000 sequências de proteínas) usando o mecanismo MASCOT v.2.3.01 (Matrix Science Ltd.), com carbamidometilação como modificação fixa, e oxidação da metionina como a modificação variável, uma clivagem de tripsina perdida e tolerância da massa do fragmento do íon de 1 Da e tolerância para os fragmentos precursores dos íons de 10 PPM.

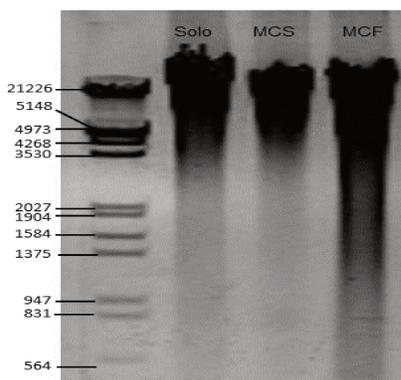
A plataforma Scaffold (versão Scaffold_3.6.1, Proteome Software Inc., Portland, OR) foi utilizada para validação MS/MS baseada na identificação dos peptídeos e proteínas. A identificação dos peptídeos é aceita se obter probabilidade maior de 60% como especificado no algoritmo "Peptide Prophet" (KELLER et al, 2002). A identificação das proteínas é aceita se obter probabilidade maior de 80% e apresentar um ou mais peptídeos identificados, as probabilidades são calculadas pelo algoritmo "Protein Prophet" (NESVIZHSKII, 2003).

4. Resultados e Discussão

4.1.Extração do DNAmet

Visando o sequenciamento foi realizada a extração de DNAmet da amostra do solo de canavial e dos “*pellets*” obtidos na 15ª semana de cultivo do MCS e 6ª semana de cultivo do MCF. A extração foi realizada como descrito no item 3.5, o perfil do DNAmet de cada amostra pode ser observado na Figura 5, destaca-se o alto peso molecular deste DNA, o que é esperado pelo fato de se tratar de DNA genômico. Estes DNAmet foram utilizados no sequenciamento descrito no item 3.6.

Figura 5: Extração do DNA metagenômico do solo, MCS e MCF (Gel de agarose 1% corado com brometo de etídio).



4.2.Extração de Proteínas do Solo e Secretoma

Na Figura 6 observa-se o perfil do secretoma das comunidades MCS e MCF obtidos após a concentração como descrito no item 3.4, o secretoma das comunidades apresentaram um perfil semelhante sendo possível a observação de diversas bandas aparentemente comuns aos dois secretomas. Visando sua utilização na análise metaproteômica como descrito no item 3.8 os geis foram seccionados dando origem a seis bandas, as linhas pontilhadas indicam a altura do corte.

Figura 6: Eletroforese em gel de poliacrilamida do secretoma das comunidades MCS e MCF. As linhas representam como foi realizada a divisão da lane para a análise metaproteômica.

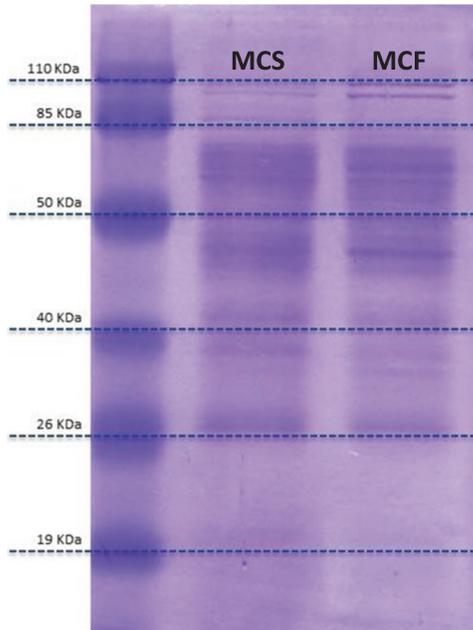
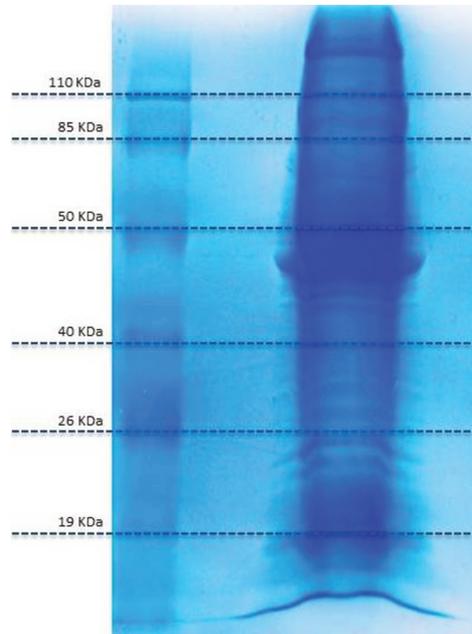


Figura 7: Eletroforese em gel de poliacrilamida da extração de proteínas do solo. As linhas representam como foi realizada a divisão da lane para a análise metaproteômica.



Além do secretoma também seguiu-se a análise metaproteômica do solo, para tanto, foi realizada a extração de proteínas a partir de 1 g de solo (Figura 7), como descrito no item 3.3. Foram obtidos 759 μg de proteínas quantificadas pelo método de Bradford (1976). Na Figura 7 observa-se o perfil proteico do solo, há presença de várias bandas indicando a complexidade deste extrato, as linhas da figura indicam como foi feita a secção da "lane" para o corte das bandas como descrito no item 3.8.

4.3. Sequenciamento

Diversos trabalhos relatam o uso do sequenciamento em larga escala como ferramenta de estudo de metagenomas extraídos de diversos ambientes, resultando em uma vasta quantidade de informação (GILL et al, 2006; DUAN et al, 2010; EDWARDS et al, 2010; HESS et al, 2011; MARX, 2013; JU et al, 2013; THUREBORN et al, 2013), auxiliando principalmente na identificação dos genes e vias metabólicas mais abundantes em um determinado ecossistema. Seguindo esta tendência, os metagenomas do solo e dos consórcios degradadores de bagaço de cana-de-açúcar foram sequenciados com o objetivo de identificar os organismos e os genes envolvidos no processo de degradação da biomassa. Para análise do sequenciamento, as sequências foram submetidas no servidor MG-RAST (disponível em: <http://metagenomics.anl.gov/>). Na Tabela 2 apresenta-se um resumo dos dados obtidos no sequenciamento das amostras.

Tabela 2: Dados dos sequenciamentos dos metagenomas de solo e dos consórcios degradadores de bagaço de cana-de-açúcar.



| | Metagenoma de Solo de Canavial | Metagenoma de Solo de Canavial | Metagenoma Consórcio Shaker | Metagenoma Consórcio Fermentador |
|-------------------------------------|--------------------------------|--------------------------------|-----------------------------|----------------------------------|
| Tamanho (bp) | 449,075,921 | 112,126,797 | 625,850,513 | 576,855,361 |
| Contagem de Sequências | 1,071,348 | 167,887 | 389,413 | 651,707 |
| Tamanho médio das Sequências | 419 ± 132 bp | 667 ± 1.689 bp | 1,607 ± 4,426 bp | 885 ± 3,581 bp |
| Porcentagem média de GC (%) | 63 ± 8% | 57 ± 10% | 55 ± 12% | 57 ± 11% |
| Predições de Proteínas | 1,121,574 | 234,989 | 860,693 | 1,059,676 |
| Predições de rRNA | 211,287 | 26,552 | 55,421 | 66,164 |
| Proteínas identificadas | 575,491 | 130,173 | 357,921 | 548,105 |
| rRNA identificados | 915 | 2,311 | 693 | 1,176 |
| Categorias de Funções identificadas | 536,392 | 119,835 | 317,533 | 501,303 |
| MG-RAST ID | 4460382.3 | 4486942.3 | 4475147.3 | 4487709.3 |
| Método de Sequenciamento | 454 | illumina | Hi-Seq 2000 | Hi-Seq 2000 |



O sequenciamento do MSp (metagenoma do solo de canavial na plataforma 454) gerou 1.071.348 “reads” de tamanho médio de 400 pares de base com um total de 449.075.921 pares de base sequenciadas. Dentre essas sequências 1.121.574 foram preditas como sequências de proteínas, mas apenas 575.941 dessas sequências foram identificadas como proteínas já descritas nos bancos de dados. É importante ressaltar que este dado de pirosequenciamento não foi submetido à montagem de “contigs”. Por outro lado os “reads” gerados pela plataforma illumina foram previamente reunidos em “contigs”, esses resultados são descritos a baixo.

O sequenciamento do MSi (metagenoma do solo de canavial na plataforma illumina) gerou 167.887 “contigs” de tamanho médio de 1.000 pares de base com um total de 112.126.797 pares de base sequenciadas. Dentre essas sequências 234.989 foram preditas como sequências de proteínas, mas apenas 130.173 sequências identificaram proteínas.

O sequenciamento do MCS (metagenoma comunidade shaker na plataforma illumina) gerou 389.413 “contigs” de tamanho médio de 2.000 pares de base com um total de 625.850.513 pares de base sequenciadas. Dentre essas sequências 860.693 foram preditas como sequências de proteínas, mas apenas 317.533 sequências identificaram proteínas.

O sequenciamento do MCF (metagenoma comunidade fermentador na plataforma illumina) gerou 651.707 “contigs” de tamanho médio de 2000 pares de base com um total de 576.855.361 pares de base sequenciadas. Dentre essas sequências 1.059.676 foram preditas como sequências de proteínas, mas apenas 548.105 sequências identificaram proteínas.

A análise do sequenciamento, no servidor MG-RAST, torna possível a análise taxonômica, funcional e a comparação da amostra sequenciada com outros metagenomas já depositados no servidor (MEYER et al, 2008). Através desta plataforma foram construídas as curvas de rarefação para cada um dos metagenomas. Nota-se que não há tendência à formação de platô dessas curvas, o que é um indicativo da grande diversidade desses metagenomas e de que apesar da grande quantidade de dados sequenciados, não foi possível fazer uma cobertura total desses metagenomas as curvas foram construídas utilizando o banco de dados do M5NR (Figura 8) e utilizando o banco de dados do SEED (Figura 9).

Figura 8: Curva de rarefação do sequenciamento, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco M5NR usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA.

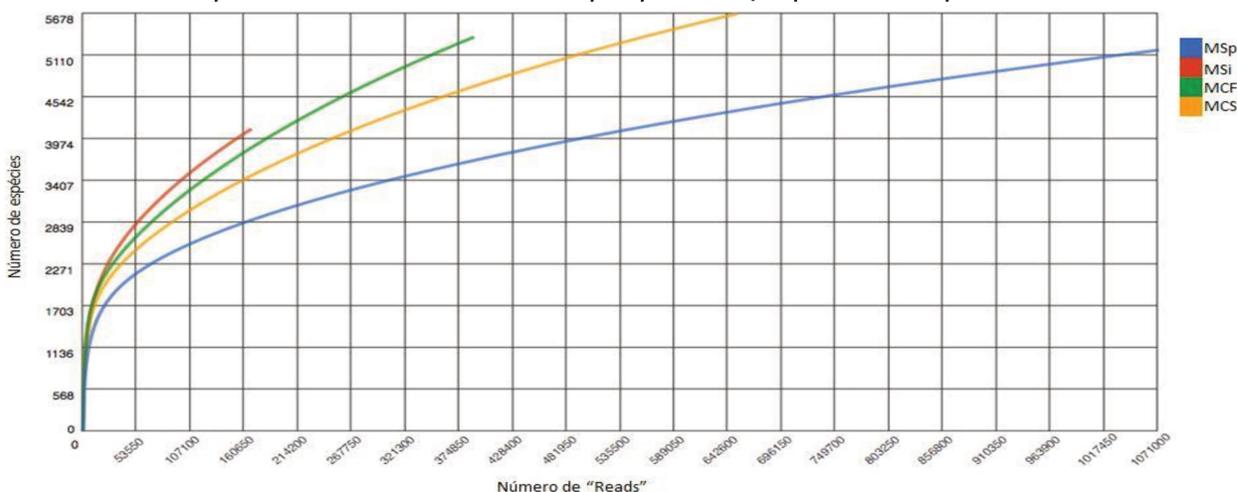
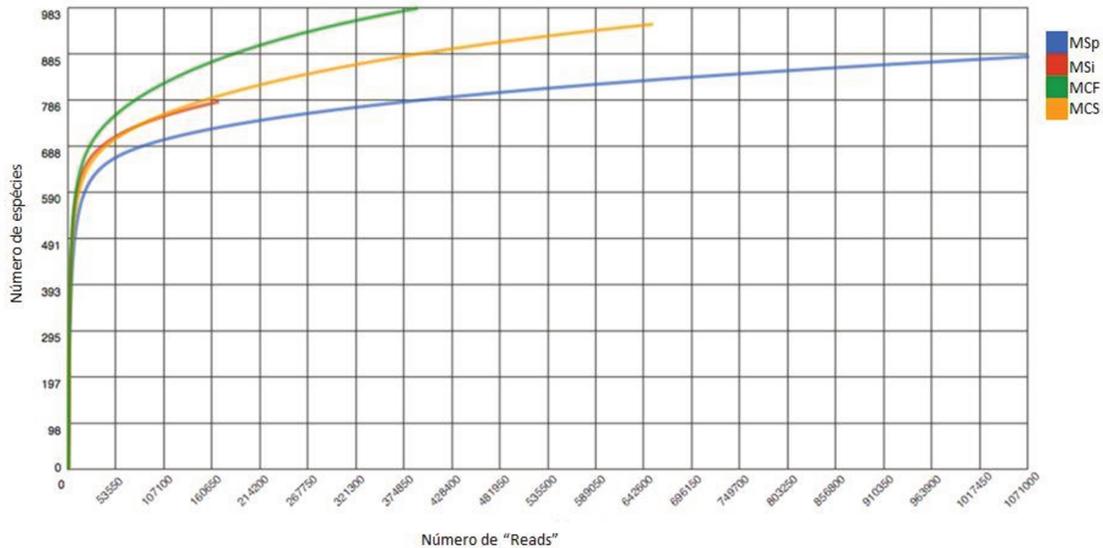


Figura 9: Curva de rarefação do sequenciamento, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco SEED usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA.



4.4. Análise do perfil taxonômico do metagenoma do solo de cana-de-açúcar e comunidades microbianas derivadas

A análise de PCA foi realizada na plataforma MG-RAST, a PCA é um método bem conhecido para a redução de dimensões de grandes conjuntos de dados. A Redução de dimensionalidade é um processo que permite que a variação encontrada em um grande conjunto de dados complexos (por exemplo, os valores de abundância de milhares táxons em dezenas de amostras metagenômicas) possa ser reduzida a um número muito menor de variáveis, próprios para a interpretação humana. No contexto do MG-RAST, PCA permite que amostras que exibem perfis semelhantes (abundância taxonômica ou funcional) sejam agrupadas.

Através da análise PCA verifica-se que os perfis taxonômicos dos metagenomas do MSp e MSi podem ser agrupados, se encontram no mesmo quadrante, ou seja apresentam perfis semelhantes. Isso já era esperado, por se tratarem de uma mesma amostra de solo, apenas sequenciada em plataformas diferentes. Já os metagenomas MCS e MCF se encontram afastados tanto entre si como do solo. Isso indica que ambas as comunidades sofreram adaptações e se especializaram às suas condições de cultivo, essas análises foram feitas com base no banco de dados M5NR (Figura 10) e do banco de dados do SEED (Figura 11).

Figura 10: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco M5NR usando um valor máximo de e-value igual a e-5, uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1.

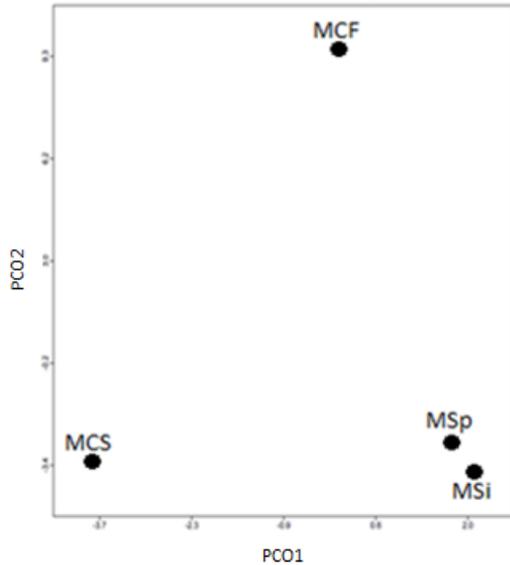
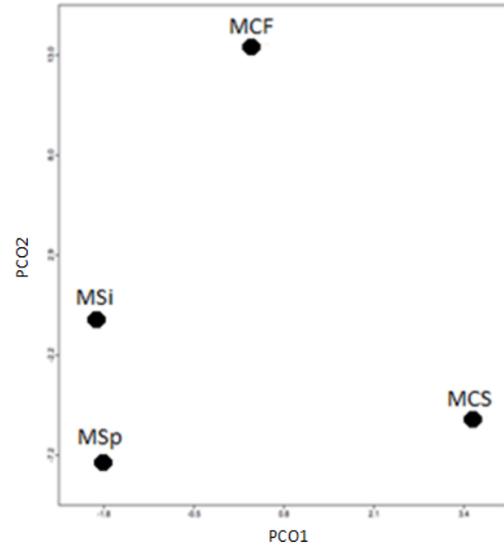


Figura 11: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco SEED usando um valor máximo de e-value igual a e-5, uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1.



A análise do sequenciamento revelou a predominância de bactérias, que representam 94,86% do metagenoma de solo de canavial sequenciado na plataforma 454 (MSp), 92,94% do metagenoma de solo de canavial sequenciado na plataforma Illumina (MSi), 94,87 do metagenoma do consórcio estabelecido em shaker (MCS) e 94,97% do metagenoma do consórcio estabelecido em fermentador (MCF), com base no número de hits observados no banco de dados SEED (Tabela 3).

Tabela 3: Perfil da abundância normalizada dos domínios dos metagenomas utilizando o banco de dados SEED. Legenda: metagenoma de solo de canavial sequenciado na plataforma 454 (MSp), metagenoma de solo de canavial sequenciado na plataforma illumina (MSi), Metagenoma Comunidade Shaker (MCS) e Metagenoma Comunidade Fermentador (MCF).

| Dominio | MSi | MSp | MCS | MCF |
|------------------|--------|--------|--------|--------|
| Archaea | 3,20% | 1,13% | 0,14% | 0,26% |
| Eubacteria | 92,94% | 94,86% | 94,87% | 94,97% |
| Eukarya | 0,25% | 0,68% | 0,17% | 0,09% |
| Não identificado | 3,59% | 3,33% | 4,80% | 4,67% |
| Vírus | 0,01% | 0,01% | 0,01% | 0,01% |

O filo Proteobacteria é o grupo mais abundante em todos os metagenomas avaliados, representando 45,67% do MSp, 45,12% do MSi, 57,71% do MCS e 58,82% do MCF, baseado no resultado do Blast da sequências com a taxonomia do SEED. Corroboram para estes dados o resultado da análise utilizando a plataforma do RDP (Ribosomal Data Project), onde o grupo proteobacteria representa 29,11% do MSp, 32,49% do MSi, 39,33% do MCS e 36,30% do MCF e o resultado da análise utilizando a plataforma do M5NR (M5 non-redundant protein database), onde o grupo proteobacteria representa 38,98% do MSp, 39,22% do MSi, 53,14% do MCS e 50,91% do MCF. Nota-se na Figura 12 e Figura 13 que os quatro filios mais abundantes nestes metagenomas incluem ainda Actinobacteria, Bacteroidetes e Firmicutes.

Figura 12: Abundância dos filios dos metagenomas nos bancos de dados RDP, SEED e M5NR.

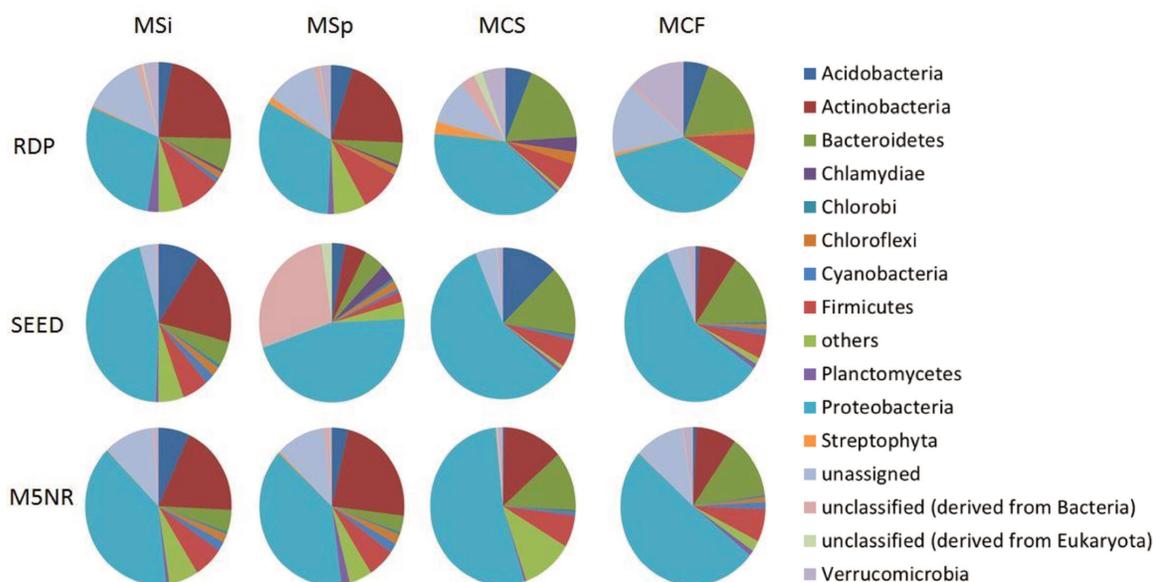
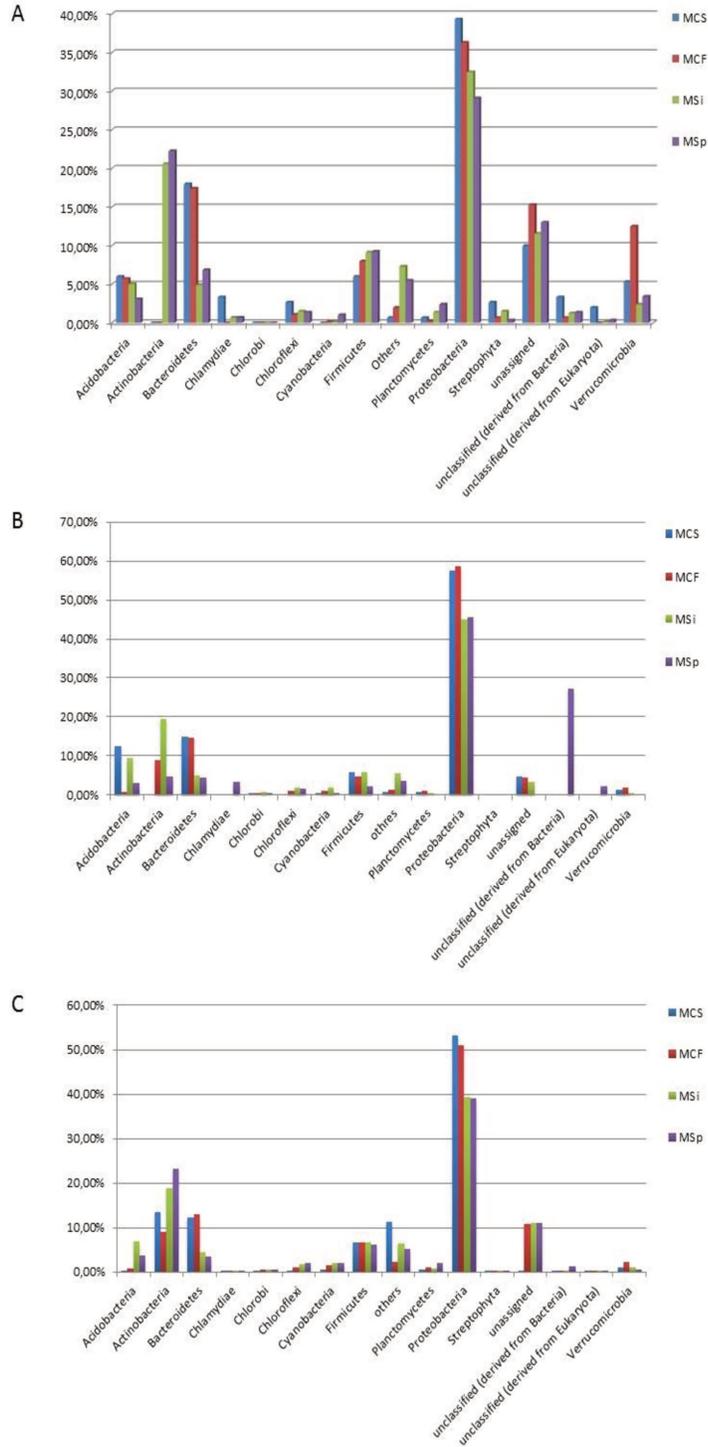
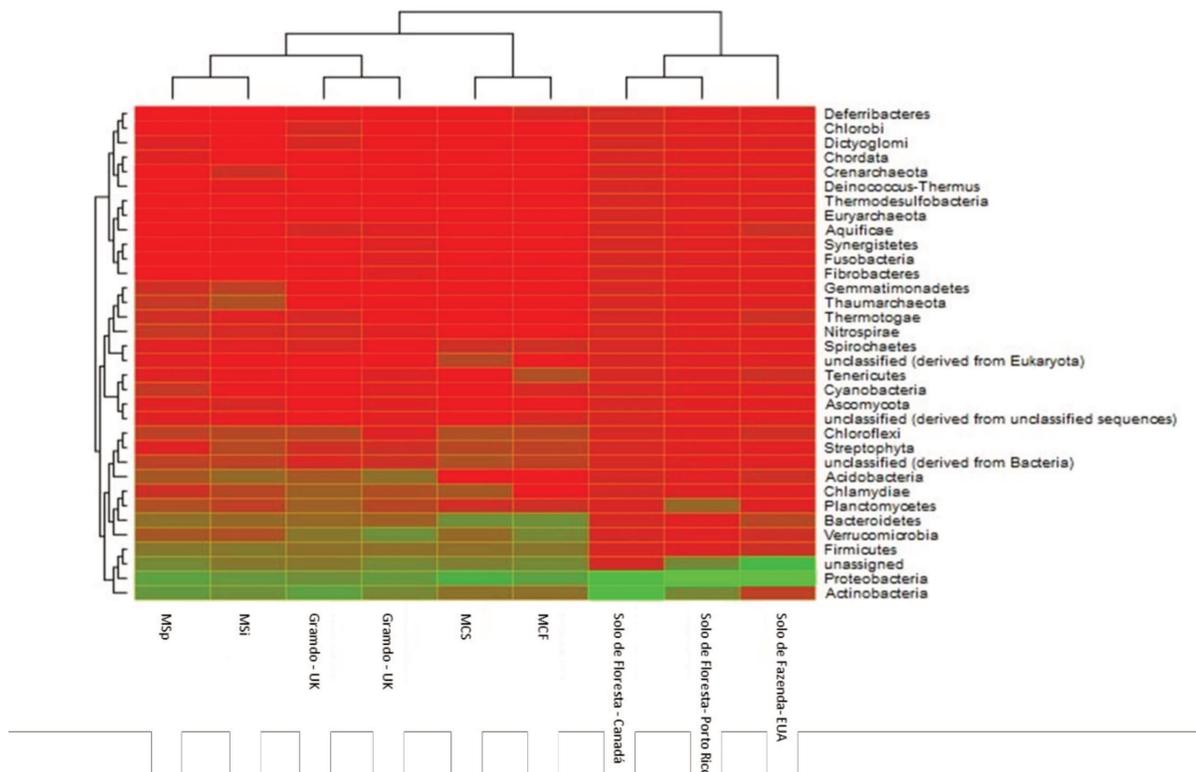


Figura 13: Gráficos das abundância dos filios dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR.



É necessário ressaltar que as características físico-químicas do solo assim como condições ambientais, tais como altitude, pluviosidade e temperatura são fatores que influenciam a composição da comunidade microbiana (FIERER et al, 2012), mas ainda assim o perfil taxonômico encontrado nos metagenomas está de acordo com o descrito na literatura, onde em outros trabalhos de metagenômica de solo os filos Proteobacteria, Actinobacteria, Bacteroidetes e Firmicutes já aparecem como principais componentes da comunidade microbiana em diversos solos como mangues (ANDREOTE et al, 2012), floresta temperada (SCHNEIDER et al, 2012) e desertos frios, desertos quentes, florestas tropical, pastagens e tundra (FIERER et al, 2012). Na Figura 14 observa-se um heatmap evidenciando as semelhanças dos metagenomas em estudo com outros metagenomas de solo depositados no servidor MG-Rast. Os metagenomas MSI e MSp apresentam mais semelhanças com os metagenomas de gramado do Reino Unido, e ainda o conjunto destes se apresenta semelhante aos metagenomas MCS e MCF, onde os filos mais abundantes também foram proteobacterias, firmicutes e actinobacterias. Diferenciam-se destes os solos de floresta que se agrupam entre si, e o solo de fazenda.

Figura 14: Heatmap dados calculados para os metagenomas MSp, MSI, MSC e MCF. Os dados foram comparados com o banco RDP usando um valor máximo de e-value igual a e-5, uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Foram utilizados também dados de metagenomas disponíveis no servidor MG-Rast para comparação. Foram utilizados 2 metagenomas de gramado do Reino Unido, um metagenoma de floresta do Canadá, 1 metagenoma de floresta de Porto Rico e 1 metagenoma de uma fazenda dos Estados Unidos.



Nota-se nas Figura 15 e Figura 16 dentre as classes, a mais abundante é Alphaproteobacteria, seguida por Gammaproteobacteria, Actinobacteria e Betaproteobacteria. Baseado no resultado do Blast da sequencias com a taxonomia do SEED a classe Alphaproteobacteria representa 23% do MSp, 19,68% do MSi, 25,69% do MCS e 34,50% do MCF. Corroboram para estes dados o resultado da análise utilizando a plataforma do RDP (Ribosomal Data Project), no qual o grupo Alphaproteobacteria representa 9,93% do MSp, 9,27% do MSi, 14% do MCS e 17,55% do MCF e o resultado da análise utilizando a plataforma do M5NR (M5 non-redundant protein database), no qual o grupo representa 18,08% do MSp, 17,07% do MSi, 22,98% do MCS e 27,83% do MCF.

Figura 15: Abundância das Classes dos metagenomas nos bancos de dados RDP, SEED e M5NR.

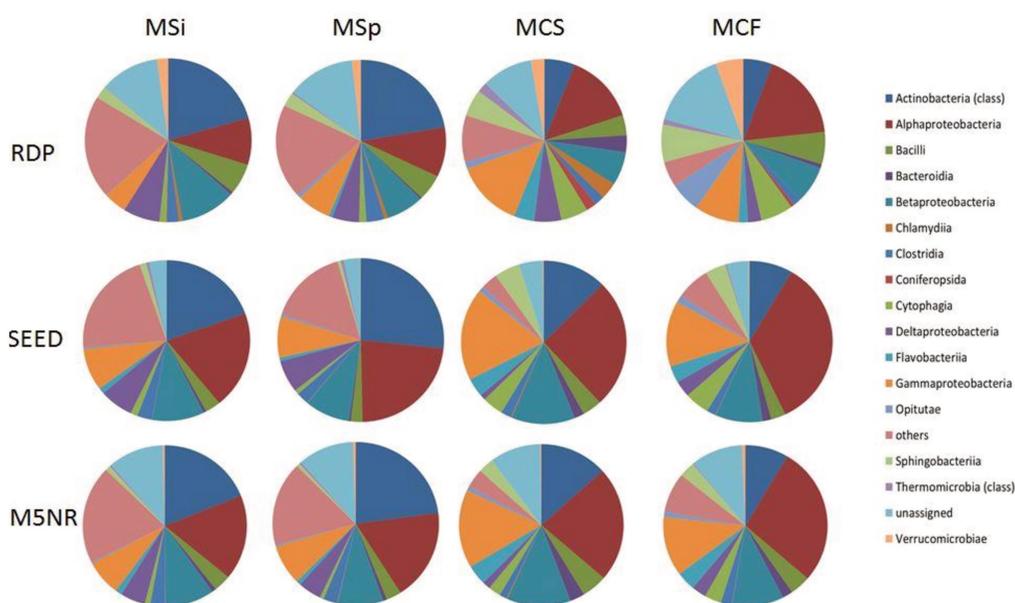
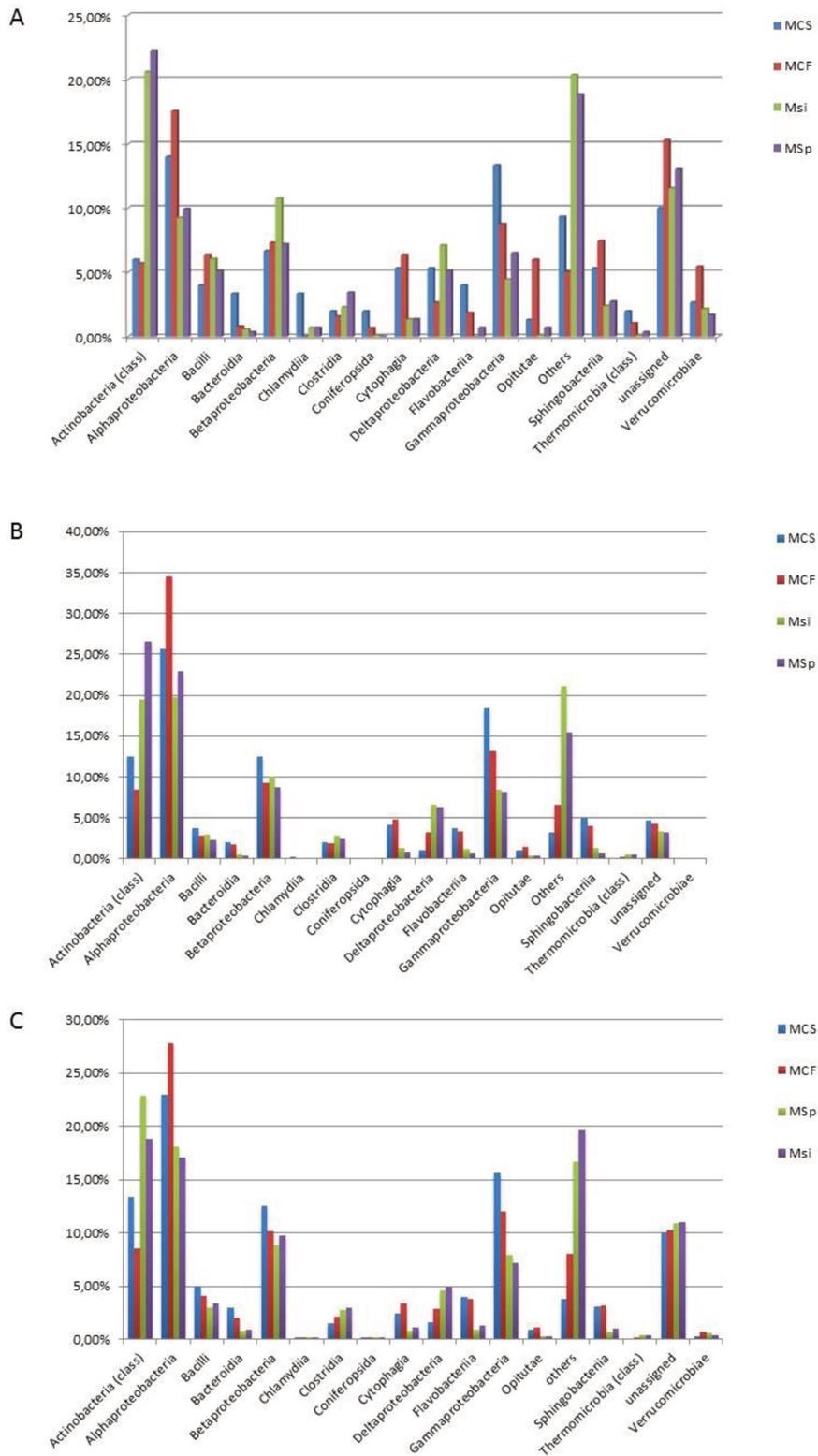
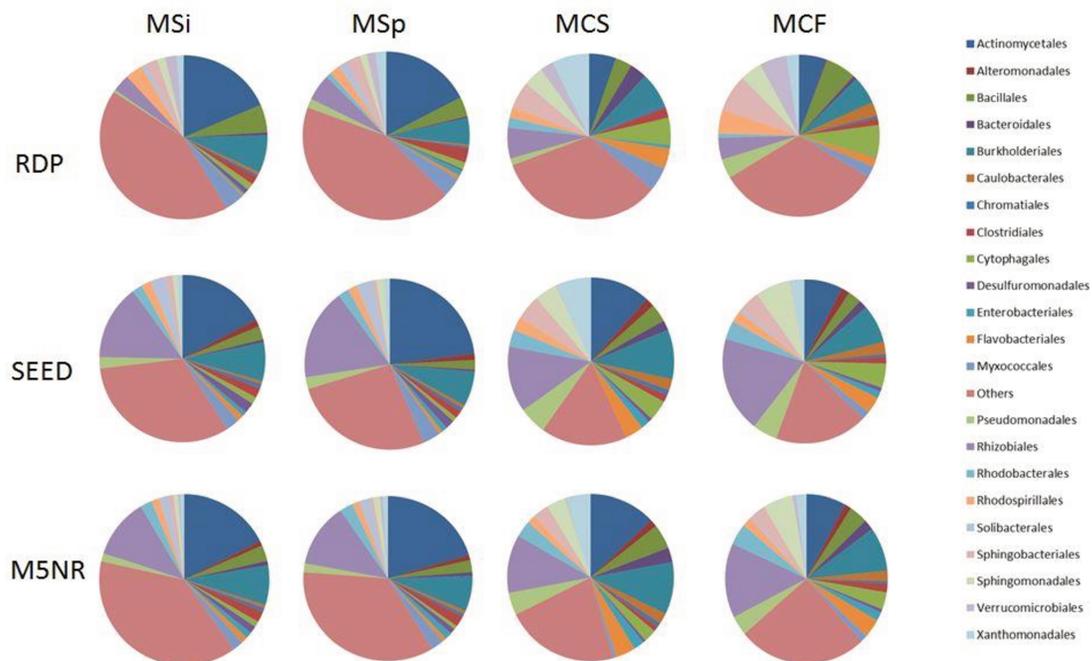


Figura 16: Gráficos das abundância das classes dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR.



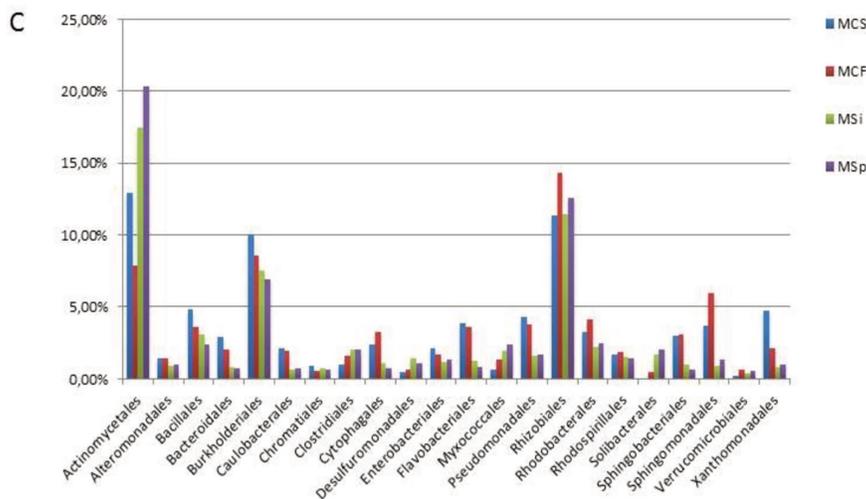
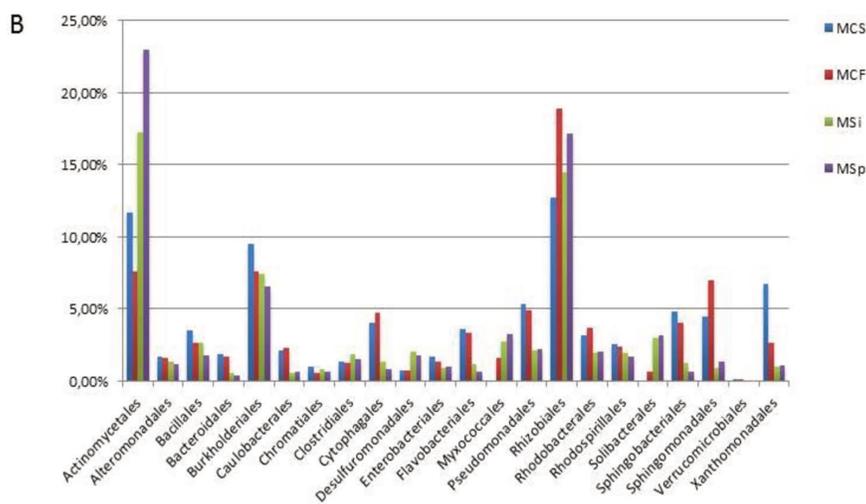
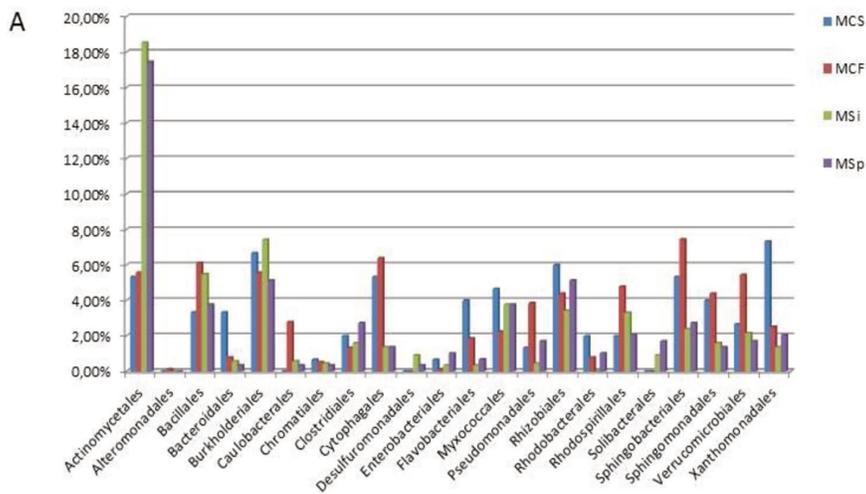
Nas Figura 17 e Figura 18 observa-se a distribuição das ordens, a mais abundante é Actinomycetales, baseado no resultado do Blast da sequencias com a taxonomia do SEED a ordem Actinomycetales representa 23% do MSp, 17% do MSi, 12% do MCS e 8% do MCF. Corroboram para estes dados o resultado da análise utilizando a plataforma do RDP (Ribosomal Data Project), no qual o grupo Actinomycetales representa 17% do MSp, 19% do MSi, 5% do MCS e 6% do MCF e o resultado da análise utilizando a plataforma do M5NR (M5 non-redundant protein database), no qual o grupo representa 20% do MSp, 17% do MSi, 13% do MCS e 8% do MCF. Destacam-se também entre as ordens a Rhizobiales, Burkholderiales e Bacilales.

Figura 17: Abundância das Ordens dos metagenomas nos bancos de dados RDP, SEED e M5NR.



O nome Proteobacteria, foi dado em referência ao deus grego Proteus, que podia assumir diversas formas, de modo a refletir a enorme diversidade de formas e fisiologia vistas dentro deste grupo (STACKEBRANDT et al, 1988 e MURRAY et al, 1990). Este grupo de organismos, anteriormente conhecido e ainda muitas vezes referido como bactérias púrpura (STACKEBRANDT et al, 1988, WOESE et al, 1987 e 1991), abrangem um conjunto muito complexo de fenótipos e características fisiológicas, incluindo organismos fototróficos (responsáveis pelas características roxas), heterotróficos e quimiolitotróficos (HOLT et al, 1994,

Figura 18: Gráficos das abundância das ordens dos metagenomas nos bancos de dados (A) RDP, (B) SEED e (C) M5NR.



COLLIER et al, 1998; ZAVARZIN et al, 1991 e TRÜPER et al, 1987). As proteobactérias pertencem ao maior e mais diverso filo de bactérias conhecidas, são bactérias gram negativas e a maioria dos seus representantes são anaeróbios, mas há exceções. Entre as proteobactérias encontram-se as Myxobacteria, um grupo único de bactérias que podem agrupar-se para formar corpos frutíferos (MADIGAN et al, 2005).

O filo Proteobactérias é subdividido em 5 classes: Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria e Gammaproteobacteria. Um dos principais nichos ecológicos do grupo é a fixação de nitrogênio no solo, no entanto o grupo proteobactéria é mais frequentemente encontrado em solos de cultivo do que solos de florestas, isso se deve ao fato do pH de solos de cultivos serem relativamente mais básico do que o de solos de floresta (NÜSSLEIN et al, 1999). Podemos relacionar a grande plasticidade metabólica deste filo ao fato da abundância deste grupo ter aumentado em aproximadamente 10% nas comunidades MCS e MCF (Figuras 12 e 13), as bactérias deste filo foram capazes de se adaptar as novas condições e continuaram a ser o filo mais abundante mesmo após o enriquecimento.

Os Bacteroidetes também são um grupo de bactérias gram negativas com grande plasticidade metabólica (MADIGAN et al, 2005). Os membros do filo Bacteroidetes colonizaram diferentes habitats, incluindo solo, oceano, água doce, e trato gastrointestinal de animais, onde eles exibem várias funções biológicas. Em particular, eles são conhecidos degradadores de matéria orgânica. Os Bacteroidetes tem sido encontrados em amostras ambientais de solo de vários locais, incluindo campos de cultivo (BORNEMAN et al, 1996; MARTINEZ-ALONSO et al, 2010), solos de efeito estufa (KIM et al, 2006) e áreas não exploradas (BUCKLEY E SCHMIDT, 2003; ZHOU et al, 2009).

Os metagenomas avaliados neste trabalho são provenientes de solo de canavial com pH 6,26, coerente com o sugerido pela embrapa (AGEITEC) que sugere que o cultivo da cana-de-açúcar deve ser feito em solos com pH entre 4,0 e 8,5, sendo que o ideal gira em torno de 6,5. A abundância do filo bacteroidetes nos metagenomas segundo o banco de dados RDP foi de 6,85% para o MSp e 4,92% para o MSi (Figura 10 A), abundância coerente com a literatura na qual um estudo publicado em 2009, incluindo solos de 88 locais diferentes revelou uma correlação positiva entre o pH do substrato e a abundância relativa de Bacteroidetes (LAUBER et al, 2009), que variam de 1,7% a um pH baixo (pH < 4) para 17% em solos alcalinos (pH > 8).

O filo Bacteroidetes é composto por quatro classes: Bacteroidia, Flavobactérias, Sphingobacteria e Cytophagia. O grupo tem como principal nicho ecológico a degradação da matéria orgânica complexa na biosfera, especialmente na forma de polissacáridos e proteínas (CHURCH et al, 2008 e KIRCHMAN et al, 2002). Como um grupo, eles são muito versáteis na gama de biopolímeros que podem ser usados como fonte de carbono e energia. Os resultados de triagem de amostras ambientais de bacteroidetes muitas vezes podem levar a descoberta de enzimas com potencial aplicação biotecnológica (BARBEYRON et al, 2001 e 2008; PANKRATOV et al, 2006 e LEE et al, 2010).

As actinobacterias são bactérias gram positivas e podem ser relacionadas à microbiota de solo, embora a maioria dos seus representantes esteja presente em ambientes de água doce. O filo tem também como nicho ecológico a decomposição de celulose e quitina sendo importante no ciclo do carbono (MADIGAN et al, 2005). As espécies do grupo também exibem uma enorme diversidade metabólica, tal como evidenciado pela produção de numerosas enzimas extracelulares e milhares de produtos metabólicos (CHATER et al, 2010 e SCHREMPF et al, 2001), muitos dos quais são os antibióticos (D’COSTA et al, 2006 e HOPWOOD, 2007).

Os representantes mais estudados deste grupo incluem *Streptomyces* spp. , que são as principais produtoras de antibióticos (BENTLEY et al, 2002, CHATER, 2006 e HOPWOOD, 2007) e importantes patógenos humanos do gênero *Mycobacterium* (*M. tuberculosis* e *M. leprae*), que são responsáveis pelo maior número de mortes por infecções bacterianas (BENTLEY et al, 2004; COLE, 1999 e SAVIOLA et al, 2006). No entanto, os gêneros *Streptomyces* e *Mycobacterium* constituem apenas dois gêneros dentro deste grande filo que contém mais de 300 gêneros (EUZEBY, 2011 e ZHI et al, 2009). Além disso, existem grandes populações de actinobactérias pouco estudadas, predominantes no solo, na água e em alto mar, ou em ambientes extremos, como o gelo do Ártico, locais contaminados quimicamente e ambientes radioativos (BARABOTE et al, 2009; BULL, et al, 2005; FERREIRA et al, 1999, MONGODIN et al, 2006; SMITH et al, 2006 e VENTURA et al, 2007) .

Nos últimos anos, com os rápidos avanços nas tecnologias de sequenciamento do genoma, foi possível o estudo da diversidade e biologia das actinobacterias. Os principais focos desses estudos foram sobre bactérias que produzem ou têm o potencial para a descoberta de novos produtos naturais úteis, por exemplo, *Streptomyces* (BALTZ et al, 2008; BEHAL et al, 2000; BERDY, 2005; BULL et al, 2007; FIEDLER et al, 2005 e OLANO et al, 2009). Além disso, a exploração

de outros gêneros industrialmente importantes, por exemplo, *Corynebacterium*, *Rhodococcus*, *Micrococcus*, *Cellulomonas*, *Acidothermus*, *Thermobifida* e *Nocardioides* e espécies benéficas para o ambiente, por exemplo, *Arthrobacter*, *Kocuria*, *Frankia*, *Kineococcus*, *Pseudonocardia*, e *Rubrobacter*, foi grandemente facilitada pelo desenvolvimento da tecnologia de sequenciamento (BAGWELL et al, 2008; BARABOTE et al, 2009; FERREIRA et al, 1999; INOUE et al, 2006; KALINOWSKI et al, 2003; LYKIDIS et al, 2007; MCLEOD et al, 2006; MONGODIN et al, 2006; NORMAND et al, 2007 e TAKARADA et al, 2008).

A presença destes grupos, proteobacteria, actinobacteria, bacteriodetes e firmicutes em ambientes diversos pode ser explicada pela sua grande versatilidade como já discutido, mas deve ser ressaltado que a presença constante destes grupos também se deve ao fato de que estes filós possuem milhares de espécies associadas representando um total de 90 a 95% de todas as espécies conhecidas de bactérias, o que faz com que seus bancos de dados sejam mais completos e, portanto os fragmentos de 16S que correspondem a esses filós conseguem mais eficientemente serem resgatados durante as buscas.

Durante as análises foi possível observar mudanças populacionais entre os metagenomas. A exemplo temos alguns grupos que parecem ser mais abundantes nos metagenomas do solo do que nos dos consórcios na Tabela 4. Essa diferença pode ter sido observada pela pressão seletiva aplicada aos consórcios pelas condições de cultivo (temperatura e agitação) e pela fonte de carbono (bagaço de cana-de-açúcar).

Tabela 4: Grupos que eram mais abundantes nos metagenomas de solo e que se tornaram menos abundantes nas comunidades enriquecidas, dados analisados com o banco de dados SEED.

| Classe | Ordem | MSp | Msi | MCS | MCF |
|----------------------------|----------------------------|------------|------------|------------|------------|
| Actinobacteria | Actinomycetales | 23% | 17% | 12% | 8% |
| Deltaproteobacteria | Myxococcales | 3% | 3% | 0% | 2% |
| Deltaproteobacteria | Desulfuromonadales | 2% | 2% | 1% | 1% |
| Solibacteres | Solibacterales | 3% | 3% | 0% | 1% |
| Actinobacteria | Solirubrobacterales | 2% | 1% | 0% | 0% |

Da mesma forma verifica-se que outros grupos apresentam maior abundância dentro das comunidades enriquecidas do que em relação ao metagenoma do solo, entre essas classes de microrganismos enriquecidas temos a Classe Bacteroidia, Alphaproteobacteria e Gammaproteobacteria (Tabela 5).

Tabela 5: Grupos que eram menos abundantes nos metagenomas de solo e que se tornaram mais abundantes nas comunidades enriquecidas, dados analisados com o banco de dados SEED.

| Classe | Ordem | MSp | MSi | MCS | MCF |
|----------------------------|---------------------------|-----|-----|-----|-----|
| Alphaproteobacteria | Sphingomonadales | 1% | 1% | 4% | 7% |
| Gammaproteobacteria | Pseudomonadales | 2% | 2% | 5% | 5% |
| Cytophagia | Cytophagales | 1% | 1% | 4% | 5% |
| Alphaproteobacteria | Rhodobacterales | 2% | 2% | 3% | 4% |
| Sphingobacteria | Sphingobacteriales | 1% | 1% | 5% | 4% |
| Flavobacteria | Flavobacteriales | 1% | 1% | 4% | 3% |
| Gammaproteobacteria | Xanthomonadales | 1% | 1% | 7% | 3% |
| Bacteroidia | Bacteroidales | 0% | 1% | 2% | 2% |

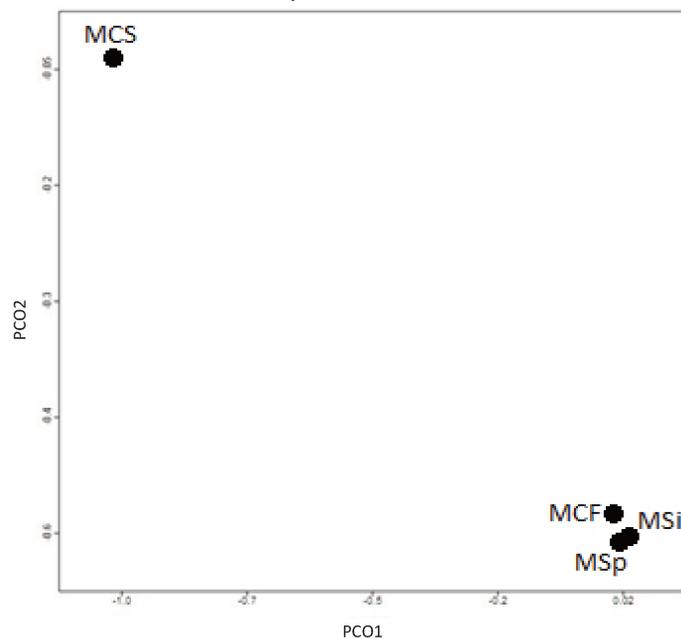
Dentre os grupos que foram enriquecidos nos consórcios convém destacarmos os Bacteroidetes (classes Cytophagia, Shingobacteria, Flavobacteria e Bacteroidia). Esse enriquecimento nas comunidades cuja a única fonte de carbono era a biomassa lignocelulósica pode ser explicado pelo fato de que os genomas de espécies de Bacteroidetes apresentam um número e diversidade de enzimas do CAZy (glicosil hidrolases) maior do que qualquer outro grupo, refletindo as estratégias moleculares desenvolvidas por esta comunidade microbiana de diferenciar, capturar e degradar carboidratos complexos. Conseqüentemente, como um resultado desta capacidade para degradar carboidratos são muitas vezes utilizados para isolar as enzimas específicas para a degradação de polissacárido (BERG et al, 1980; TIERNY et al, 1994; BERNARDET et al, 2006 e REICHENBACH , 2006).

Em cada uma das quatro classes do filo Bacteroidetes , há exemplos de enzimas do CAZY enriquecidas. Em Bacteroidia, os proteomas de *B. thetaiotaomicron*, *B. fragilis*, e *P. ruminicola* compreende 272, 137, e 130, respectivamente, GH e PL (XU et al, 2003 e KUWAHARA et al, 2004), muito mais do que os outros membros da microbiota intestinal, ou fora do filo Bacteroidetes, como o *Clostridium perfringens* (57 GH e PL) e *Bifidobacterium longum* (49 GH). O mesmo é verdadeiro para Flavobactérias, por exemplo, *G. forsetii* com 48 GH/PL (BAUER et al, 2006), *Z. profunda* com 120 GH (QIN et al, 2010), *F. johnsoniae* com 152 GH/PL (MCBRIDE et al, 2009), Cytophagia, *S. linguale* com 151 GH/PL (LAIL et al, 2010), *D. fermentans* com 98 GH/PL (LANG et al, 2009) e Sphingobacteria, por exemplo, *C. pinensis* com 184 e *P. heparinus* com 163 GH/PL, respectivamente (HAN et al, 2009 e DEL RIO et al, 2010).

4.5. Análise do perfil funcional do metagenoma do solo de cana-de-açúcar e comunidades microbianas derivadas

A análise de PCA realizada na plataforma MG-RAST para avaliação do perfil taxonômico dos metagenomas foi também realizada para avaliação do perfil funcional. Através da análise PCA podemos verificar que os perfis funcionais dos metagenomas do MSp e MSi podem novamente ser agrupados, ou seja apresentam perfis semelhantes. O metagenoma MCF também apresentou um perfil funcional semelhante ao apresentado pelos metagenomas MSp e MSi, já o metagenoma MCS se encontra afastado. Isso sugere que a comunidade do MCS sofreu adaptações e alterações no seu perfil funcional, esta adaptação da comunidade MCS pode ser explicada pelo maior tempo de cultivo desta comunidade, permitindo um enriquecimento de genes que codifiquem para as enzimas de interesse. As análises foram feitas com base no banco de dados Subsystems (Figura 19).

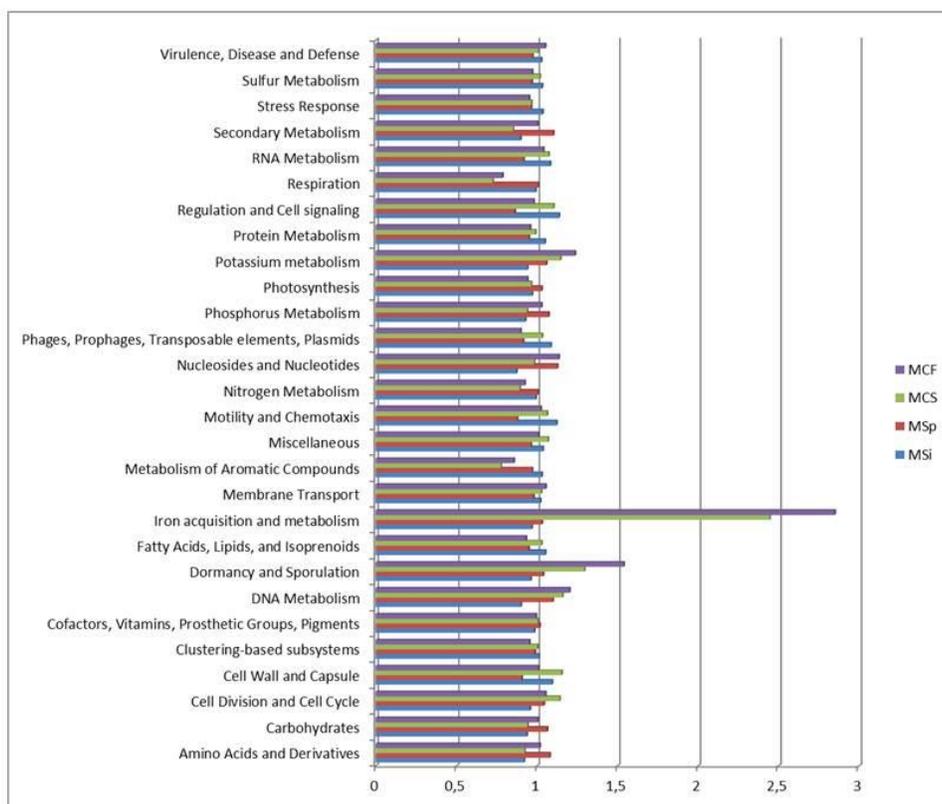
Figura 19: Análise de PCA, dados calculados para os metagenomas MSp, MSi, MSC e MCF. Os dados foram comparados com o banco Subsystems usando um valor máximo de e-value igual a e^{-5} , uma identidade mínima de 60%, e alinhamento de comprimento mínimo de 50 aminoácidos para proteínas e/ou pares de bases para RNA. Os dados foram normalizados para valores entre 0 e 1.



A análise do perfil funcional dos metagenomas foi realizada utilizando o banco Subsystems na plataforma do MG-RAST. Para o nível 1 (funções gerais) quando os dados são normalizados em relação a média da abundância dos solos ($M=1$), não é possível perceber variações significativas

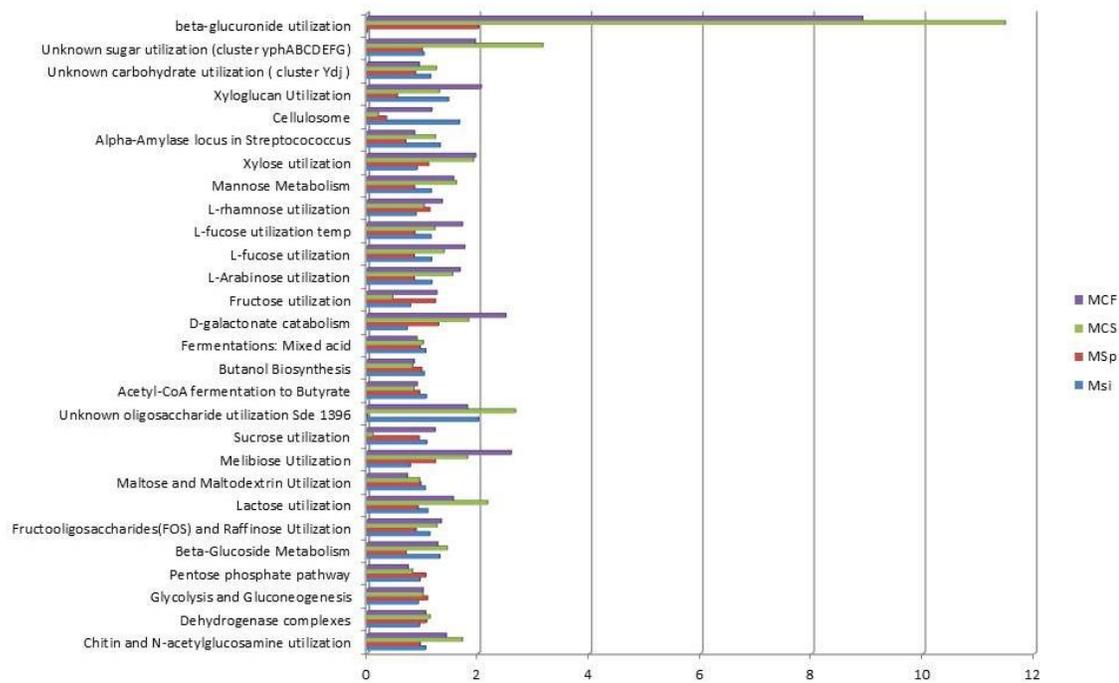
entre os metagenomas em relação ao metabolismo de carboidratos, o que também é esperado uma vez que o metagenoma do solo de canalial estudado já é um metagenoma no qual esperamos encontrar as enzimas de interesse enriquecidas (Figura 20). Mas podemos notar um aumento significativo no metabolismo de aquisição de ferro. Para uma avaliação mais completa passamos então a olhar algumas funções mais específicas.

Figura 20: Gráfico perfil funcional nível 1 dos metagenoms MSi, MSp, MCS e MCF de acordo com o banco de dados Subsystems.



O perfil funcional no nível 3 (Carboidratos), ainda utilizando o subsystems, revela pequenas diferenças entre os metagenomas. Quando os dados são normalizados em relação à média da abundância dos solos (M=1), podemos ver que há uma maior concentração de genes relativo à utilização de carboidratos complexos nas comunidades enriquecidas, MCS e MCF, um exemplo claro desse aumento pode ser notado observando os genes envolvidos na utilização de beta-glucoronideo que estão 8 a 10 vezes mais presentes nas comunidades MCF e MCS, respectivamente, do que nas comunidades MSi e MSp (Figura 21).

Figura 21: Gráfico perfil funcional nível 3 (carboidratos) dos metagenoms MSi, MSp, MCS e MCF de acordo com o banco de dados Subsystems.



Os glucuronídeos, também conhecidos como glucuronosídeos, são qualquer substância derivada da ligação com ácido glucurônico através de uma ligação glicosídica. Os glucuronídeos pertencem aos glicosídeos. As enzimas referidas como “utilização de glucuronídeo” incluem as beta-glucuronidases (EC 3.2.1.31) e transportadores de glucuronídeo (UidB). As β -glucuronidases são membros da família das hemicelulases, mais especificamente da família de glicosil hidrolases GH2. Essas enzimas promovem a clivagem de uma grande variedade de β -glucuronídeos (JEFFERSON et al, 1986). UidB é um transportador dependente de prótons específico para α e β -glucuronídeos. UidB é essencial para o transporte de glucuronídeo.

Nota-se ainda o enriquecimento de classes de genes envolvidos na degradação de biomassa como os genes das vias de utilização de xilose, que envolve enzimas como xilanases e xilosidases, genes de vias de metabolismos de manose, com endoglucanases e manosídeses, genes de utilização de arabinose, com arabinofuranosídeses e arabinanases além de um aumento dos genes das vias de utilização de lactose e quitina. Essas mudanças no perfil funcional das

comunidades enriquecidas sugere a especialização dessas comunidades na degradação de material lignocelulósico.

4.6. Análise do perfil metaproteômico do solo de cana-de-açúcar e comunidades microbianas derivadas

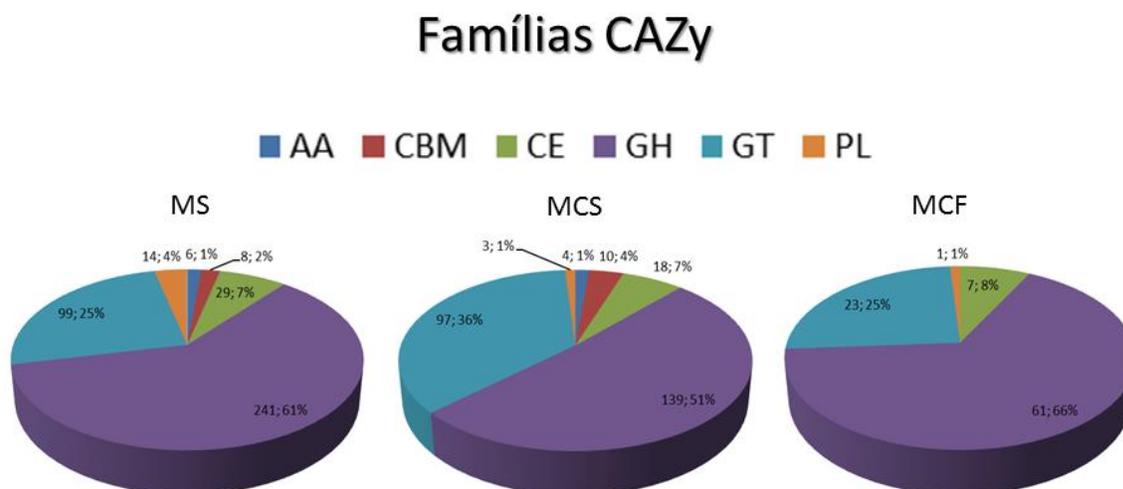
Foi realizada a análise do perfil metaproteômico das amostras de proteínas extraídas do solo e os secretomas dos consórcios, a fim de identificar as enzimas envolvidas na degradação da biomassa secretadas por estas comunidades. Os espectros de LC-MS/MS foram analisados com o software de busca MASCOT para a identificação de proteínas, utilizando o banco de dados do CAZy, como forma de validar essas identificações foi utilizado o software Scaffold, como descrito no item 3.8. O FDR (False Discovery Rate – Taxa de erro) foi calculado para cada amostra, sendo de 5,3% para o MS, 5,8% para o MCS e 4,9% para o MCF.

A metaproteômica do solo (MS) permitiu a identificação de 447 espectros, sendo 409 peptídeos únicos diferentes que representam 406 proteínas das quais três foram identificadas com 2 peptídeos, entre elas uma GH88 e uma GH1 β -glicosidase (Tabela 7, em anexo). As proteínas se agrupam em 151 famílias do CAZy. Nas análises metaproteômica da comunidade semi-aeróbica (MCS) foram identificados 317 espectros, sendo 290 peptídeos únicos diferentes que representam 282 proteínas, das quais uma foi identificada com 2 peptídeos, duas com 3 peptídeos e uma GH11, endo-1,4- β -xilanase, com 4 peptídeos e as proteínas se agrupam em 125 famílias do CAZy (Tabela 8, em anexo). E a metaproteômica da comunidade aeróbica (MCF) permitiu a identificação de 119 espectros, sendo 96 peptídeos únicos diferentes que representam 94 proteínas, das quais duas foram identificadas com 2 peptídeos, entre elas uma GH2 β -galactosidase. As proteínas se agrupam em 55 famílias do CAZy (Tabela 9, em anexo).

A degradação dos carboidratos da parede celular de plantas é mediada por diferentes enzimas classificadas no banco de dados do CAZy em famílias com base na sua sequência de aminoácidos e a atividade relacionada, são elas: as Carboidratos Esterases (CE), as Pectato Lases (PL), as Glicosil Hidrolases (GH), as Glicosil Transferases (GT) e as enzimas de Atividades Auxiliares (AA) e os Módulos de Ligação a Carboidratos (CBM). Dentre as proteínas identificadas em cada amostra 61% das proteínas do MS representam glicosil hidrolases e 25% glicosil transferases, os consórcios MCS e MCF possuem um perfil parecido e dentre as proteínas identificadas 51% e 66%, respectivamente, correspondem a glicosil hidrolases enquanto as glicosil transferases

representam 36% e 25%, respectivamente (Figura 22). Todas as famílias de enzimas do CAZy tiveram pelo menos uma enzima identificada pela proteômica com exceção da comunidade MCF onde não foi possível a identificação das famílias AA (“Auxilar Activity” – Atividade auxiliar) e CBM (“Carbohydrate binding module” - Módulo de ligação ao Carboidrato), a não identificação destas classes pode estar relacionada ao menor sucesso das corridas de espectrometria de massas do MCF, a comunidade teve um número significativamente menor de proteínas identificadas quando comparadas ao MCS e MS. Outra hipótese é que pelo uso de um bagaço com um menor conteúdo de lignina e hemicelulose, na comunidade MCF não ocorreu um enriquecimento das enzimas envolvidas na sua degradação.

Figura 22: Famílias CAZy identificadas pela análise metaproteômica das amostras MS, MCS e MCF. No gráfico apresenta-se o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Trnasferase; PL – Pectato Liase.

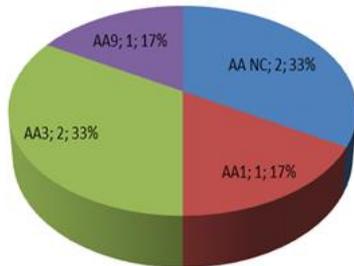


A análise da metaproteômica do MS revelou 6 proteínas que codificam AAs representando 4 famílias, 8 que codificam CBMs representando 6 famílias, 29 CEs, representando 9 famílias, 241 GHs representando 78 famílias, 99 GTs representando 45 famílias e 14 PLs representando 9 famílias (Figura 23). A metaproteômica da comunidade MCS identificou 4 AAs representando 3 famílias, 10 CBMs que representam 9 famílias, 18 CEs que representam 8 famílias, 139 GHs representando 64 famílias, 97 GTs representando 38 famílias e 3 PLs representando 3 famílias (Figura 24). A metaproteômica da comunidade MCF identificou 7 CEs que representam 4 famílias, 61 GHs representando 32 famílias, 23 GTs representando 18 famílias e 1 PL representando 1 família (Figura 25).

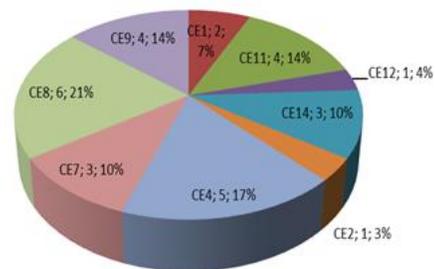
Figura 23: Famílias CAZy identificadas pela análise metaproteômica da amostra MS (Metagenoma Solo). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase.

MS

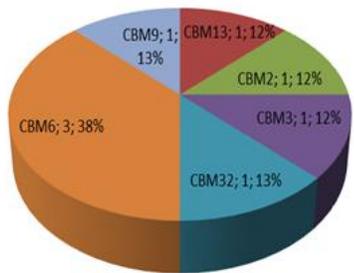
AA



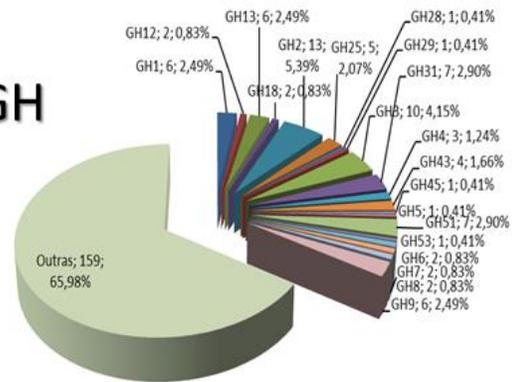
CE



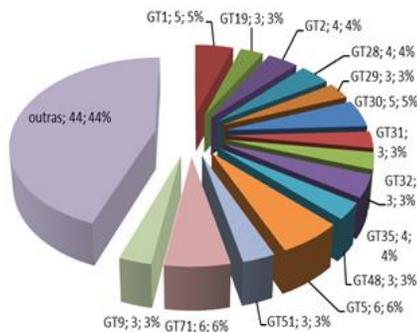
CBM



GH



GT



PL

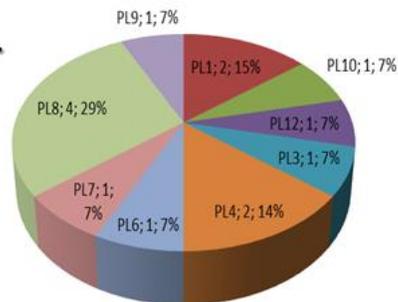
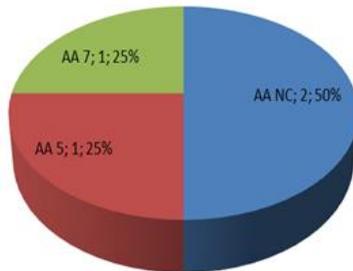


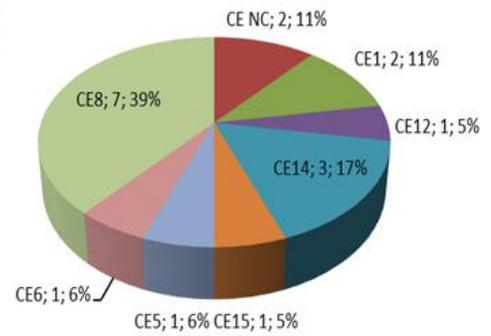
Figura 24: Famílias CAZY identificadas pela análise metaproteômica da amostra MCS (Metaproteômica da Comunidade Shaker) No gráfico apresenta-se a família CAZY e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. AA – Atividade Auxiliar; CBM - Molécula de ligação ao carboidrato; CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase.

MCS

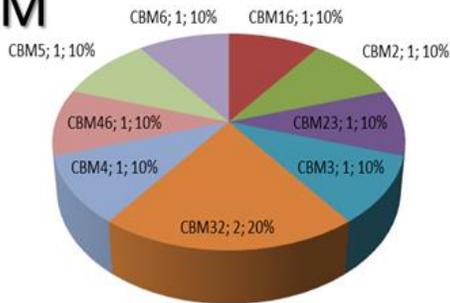
AA



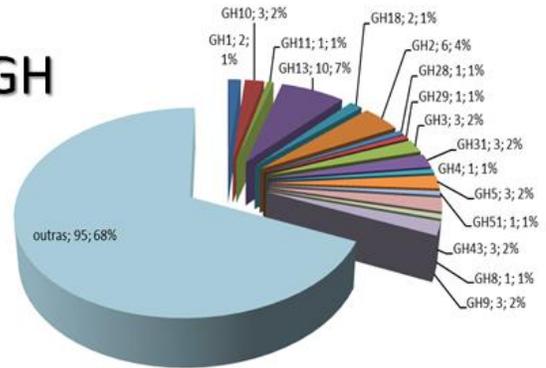
CE



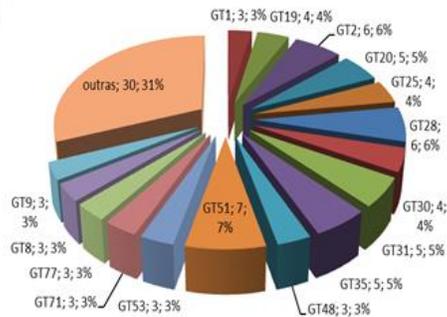
CBM



GH



GT



PL

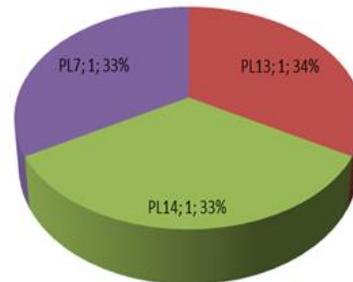
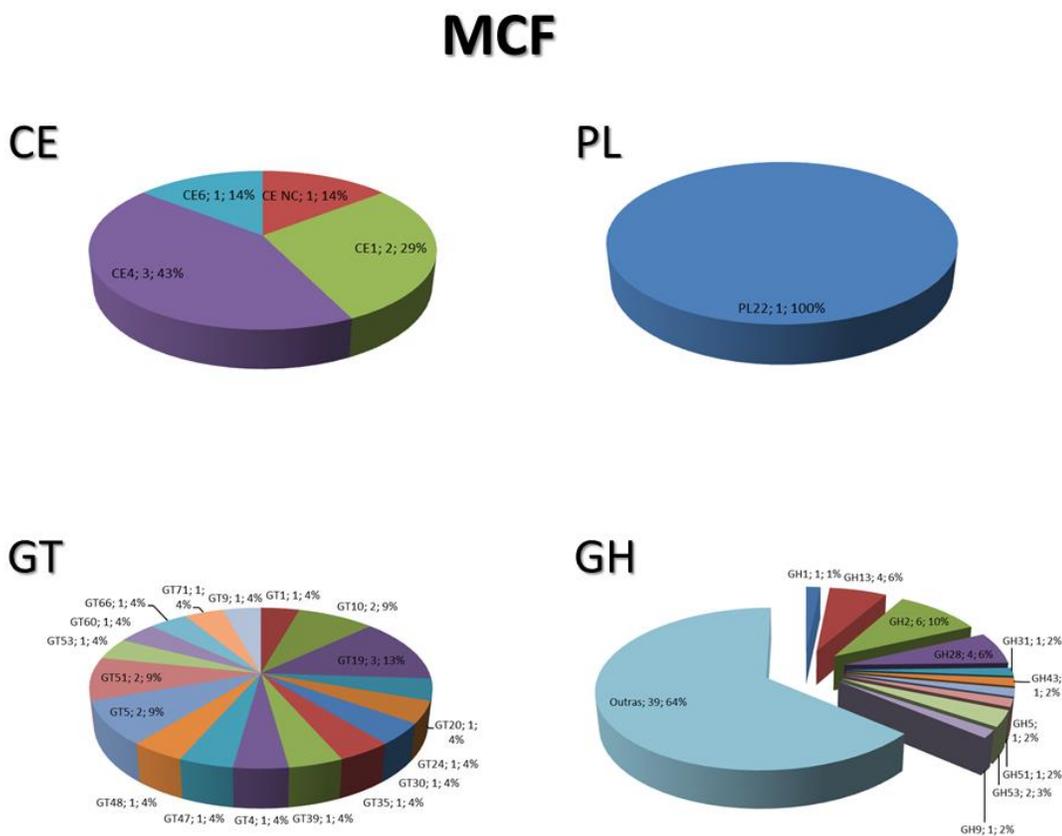


Figura 25: Famílias CAZy identificadas pela análise metaproteômica da amostra MCF (Metaproteômica da Comunidade Fermentador). No gráfico apresenta-se a família CAZy e o número de proteínas únicas identificadas, seguido pelo percentual correspondente. CE – Carboidrato Esterase; GH – Glicosil Hidrolase; GT – Glicosil Transferase; PL – Pectato Liase.



Visando a análise das famílias de glicosil hidrolases, estas foram subdivididas em celulases, hemicelulases, pectinases e outras enzimas que auxiliam na degradação da biomassa. Classificou-se como celulases enzimas β -glicosidase, celulase, endoglucanase e exoglucanase das famílias GH1, GH3, GH5, GH6, GH7, GH9, GH12 e GH45. Classificou-se como hemicelulases enzimas xilosidase, manosidase, galactosidase, β -manosidase, endomanosidase, xiloglucanase, xilanase, arabinofuranosidase e arabinanase das famílias GH2, GH3, GH8, GH10, GH11, GH12, GH43, GH51 e GH53. Classificou-se como pectinases enzimas endopoligalacturanase, fucosidase e β -galactosidase das famílias GH28, GH29 e GH43, para as análises deste estudo também foram consideradas as pectinases pertencentes a família pectato liase. Outras enzimas auxiliares na degradação da biomassa, que foram consideradas incluem enzimas amilase, quitinase e lisozima das famílias GH4, GH13, GH18, GH25 e GH31.

A análise da metaproteômica sugere uma predominância das hemicelulases seguida das celulases entre as enzimas mais abundantes identificadas para as três comunidades. No solo as hemicelulases representam 31% da amostra enquanto as celulases representam 28% do total de GH identificadas. Nas comunidades MCS e MCF as hemicelulases representam 28% e 43%, respectivamente, e as celulases representam 22% e 13% das enzimas pertencentes a família GH encontradas nessas comunidades (Figura 26). Dentre as Celulases identificadas as mais abundantes no solo foram a GH1, GH3 e GH9 com 22%, 26% e 22%, respectivamente, no solo ainda foi encontrada uma maior diversidade de celulases sendo identificadas também outras classes como GH5, GH6, GH7, GH12 e GH45. Na comunidade MCS as celulases mais abundantes foram as GH1, GH3, GH5 e GH9 com 18%, 28%, 27% e 27% respectivamente e para a comunidade MCF as mais abundantes são GH1, GH5 e GH9 com 34%, 33% e 33% respectivamente (Figura 27). Dentre as Hemicelulases as mais abundantes no solo foram GH2, GH43 e GH51 com 44%, 13% e 23%, respectivamente; no MCS GH2, GH10 e GH43 com 43%, 22% e 14%, respectivamente; e no MCF GH2, GH 43, GH 51 e GH 53 com 60%, 10%, 10% e 20% respectivamente (Figura 28).

Figura 26: Enzimas da Família GH identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se o número de proteínas únicas identificadas, seguido pelo percentual correspondente.

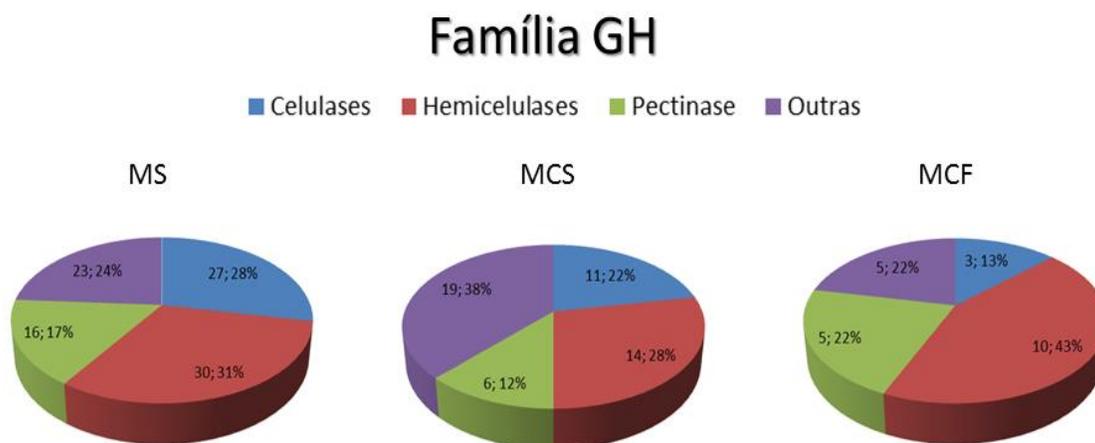


Figura 27: Celulases identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se a família CAZY e o número de proteínas únicas identificadas, seguido pelo percentual correspondente.

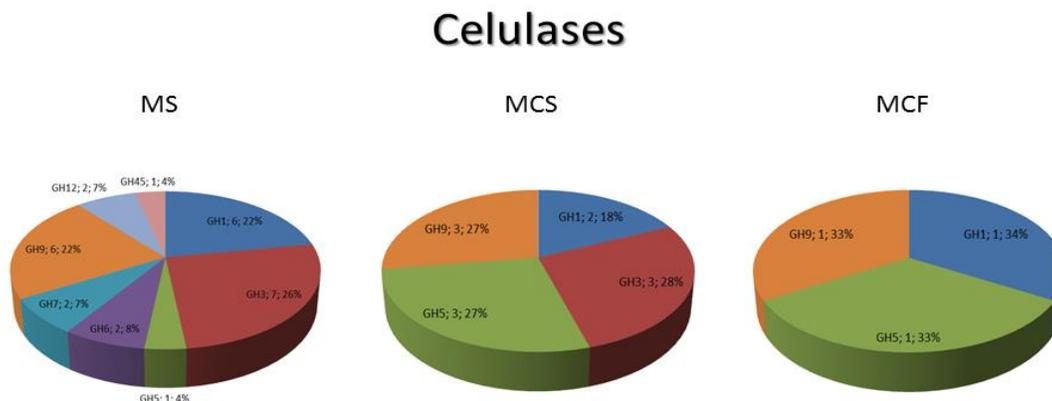
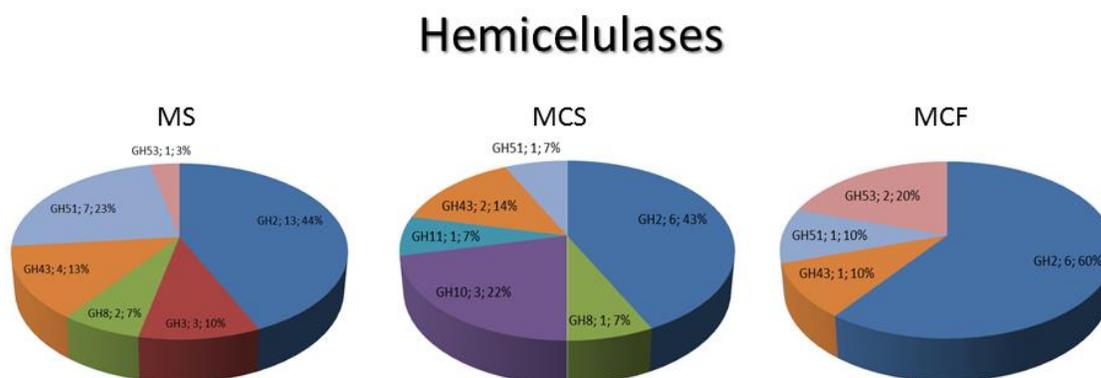


Figura 28: Hemicelulases identificadas pela análise metaproteômica das amostras MS (Metaproteômica do Solo), MCS (Metaproteômica Comunidade Shaker) e MCF (Metaproteômica de Comunidade Fermentador). No gráfico apresenta-se a família CAZY e o número de proteínas únicas identificadas, seguido pelo percentual correspondente.



O artigo publicado em 2013 por Ries e colaboradores compara as estratégias de decomposição da biomassa utilizada por *Trichoderma reesei* e *Aspergillus niger*. Na presença de palha de trigo, muitos dos genes que codificam para enzimas do CAZy foram supra-regulados nos dois fungos, que parecem codificar um coquetel de GHs e AAs similar e já previsto por análises do genoma. Apesar de algumas famílias serem espécie-específico, *T. reesei* e *A. niger* utilizam um mesmo conjunto de enzimas GH (3, 7, 11, 30 e 67) incluindo β -glicosidases, celobiohidrolase CEL7A e xilanases e AA (9) oxidoredutases, mas famílias diferentes de CE para a degradação da palha de trigo. As comunidades MCF, MCS e MS parecem se valer de estratégia semelhante sendo mais abundantes endoglucanases e mananases (GH5 e 9), β -glucosidases (GH1 e 3), manosidases (GH2), xilanases (GH10 e 43), arabinofuranosidades (GH51), arabinanases (GH53), enzimas da família AA 9, acetil xilano esterases (CE1 e 4) e pectina metil-esterase (CE8).

A análise desses perfis metaproteômicos podem auxiliar a elucidar a estratégia de degradação da biomassa utilizada nessas comunidades, no entanto é importante ressaltar que como os experimentos não foram realizados em triplicata, e portanto não é possível ser feita uma comparação quantitativa das proteínas encontradas. O foco dessas análises foi identificar enzimas presentes em cada uma das comunidades visando a prospecção de novas proteínas envolvidas na degradação da biomassa. Mas nota-se a presença de enzimas oxidantes e antioxidantes, no metaproteoma MS, no qual foram encontradas 5 enzimas e um aumento das mesmas no metaproteoma MCS, no qual foram encontradas 8 enzimas. A presença dessas enzimas sugere que radicais de oxigênio produzidos podem ser importantes na estratégia utilizada pela comunidade na desconstrução da lignina. Dentre as famílias AAs encontradas temos enzimas como lacases (AA1), celobiose dehidrogenases (AA3), mono-oxigenases (AA9) e oxidases (AA5 e 7). Além destas, na comunidade MCS, foram encontradas catalases e superóxido-dismutases, também corrobora com esse achado o enriquecimento de enzimas envolvidas no metabolismo de ferro notado nas análises do sequenciamento do metagenoma MCS.

D'haeseleer e colaboradores em maio de 2013 encontraram um padrão semelhante ao descrito acima analisando o proteoma de um consórcio termofílico adaptado para a desconstrução de gramíneas. Neste estudo foram encontradas xiloglucanases (GH74), arabinofuranosidases (GH51), xilanases (GH10), endoglucanases (GH5 e 9) e celobiohidrolases (GH48), proteínas esperadas dada a biomassa, no entanto foram identificadas também superóxido-dismutases e catalases cuja função na comunidade não foi esclarecida no artigo. No entanto os autores sugerem que estas estejam envolvidas na desconstrução da lignina dado a detecção de apenas uma enzima, lacase, que poderia estar envolvida neste processo.

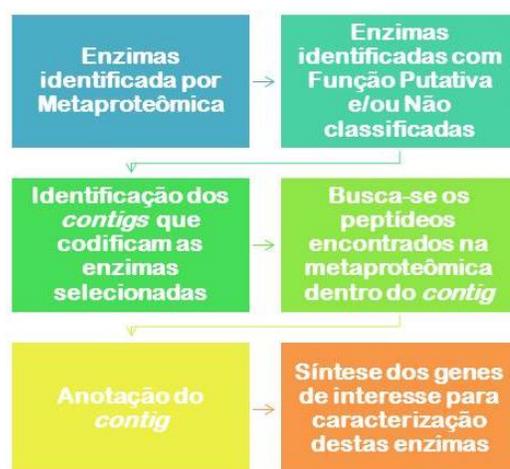
As glicosil hidrolases são as enzimas mais estudadas no processo de decomposição da biomassa, mas estudos vem mostrando que outras enzimas podem ter um papel importante neste processo. A exemplo alguns fungos dos gêneros *Postia* sp., *Phanerochaete* sp., *Wolfiporia* sp. e *Perenniporia* sp., são capazes de decompor biomassa lignocelulósica eficientemente apesar de seu arsenal enzimático ser incompleto, estes fungos não possuem enzimas importantes para a degradação da biomassa como celulasas e ligninasas. Alternativamente estes fungos utilizam a química de fenton para complementar o seu repertório enzimático, alcançando uma melhor eficiência no processo (ARANTES et al, 2011; MANSFIELD et al, 1997; MARTINEZ et al, 2009 e WYMELENBERG et al, 2010).

A química de Fenton mais conhecida por reação de Fenton foi descrita pela primeira vez por H.J.H Fenton na década de 1890, como uma potente reação de oxidação de compostos orgânicos na presença de peróxido de hidrogênio - H_2O_2 e ferro ferroso - Fe^{2+} (BARBUSIŃSKI, 2009). As espécies reativas de oxigênio geradas nessa reação são capazes de oxidar estruturas orgânicas como a biomassa lignocelulósica (KANEKO et al, 2005), auxiliando sua decomposição. Neste estudo não é possível afirmar que esses processos estão ocorrendo, para isso seriam necessários mais experimentos e a análise em triplicata das comunidades, mas a presença de enzimas relacionadas a processos oxidantes e antioxidantes, sugerem que as comunidades em estudo podem utilizar mecanismos semelhantes.

4.7. Prospecção de genes de interesse e estudos futuros

Visando utilizar os resultados descritos acima na busca por enzimas de interesse biotecnológico foi construído um banco de dados que reúne todos os *contigs* montados na etapa de sequenciamento. A estratégia empregada na busca por genes de interesse foi utilizar as proteínas identificadas pela metaproteômica como “iscas” para *contigs* com possíveis “clusters” enzimáticos, segundo o esquema apresentado na Figura 29.

Figura 29: Esquema da estratégia de busca por genes de interesse.



Segundo o esquema da figura 29, na busca por genes de interesse foram analisadas todas as proteínas identificadas por metaproteômica, o intuito dessa análise foi buscar possíveis novas enzimas, ou seja, enzimas anotadas com função putativa e/ou não classificadas dentro do banco do CAZy. Após essa análise foram encontradas 184 enzimas que representam possíveis candidatas. Estas 184 enzimas selecionadas foram então procuradas dentro do banco de dados gerado pelos

contigs, buscando assim, encontrar o *contig* que contém o gene que codifica cada uma destas enzimas. Dentre as enzimas cujos *contigs* foram encontrados, como forma de validar que estas enzimas realmente fazem parte da comunidades em estudo, foi realizada uma nova busca, na qual foi verificado se os peptideos resgatados por metaproteômica eram também encontrados no sequenciamento. Foram encontrados 10 *contigs* de interesse, destes foram selecionados 2 para anotação (Tabela 6).

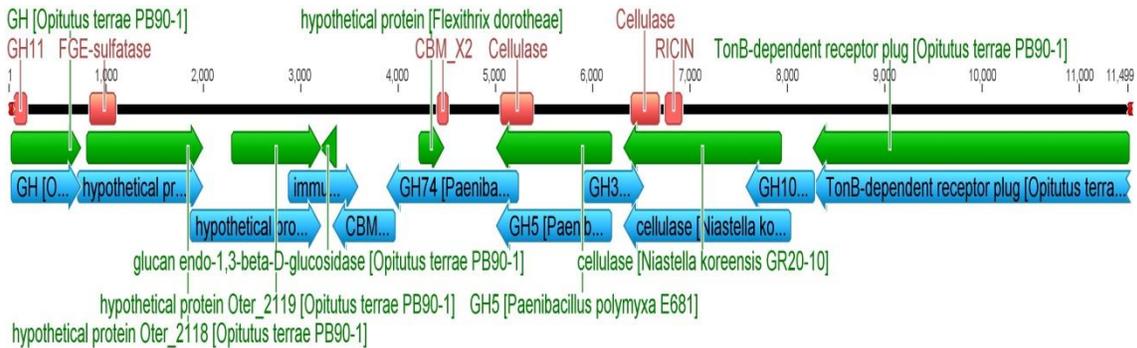
Tabela 6: Contigs selecionados para anotação.

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>Sequencia do Peptideo</u> | <u>Origem</u> | <u>Nome do Contig referente</u> | <u>e-Value</u> |
|---------------------|---------------------|--|------------------------------|---------------|---|----------------|
| A8EV24 | AA NC | Catalaseperoxidase [Arcobacter butzleri]. | DWWPNQLNL K | MS | NODE_47265 length_32667_co v_91.042274 | 0.0 |
| B9LP48 | AA NC | Catalase peroxidase [Halorubrum lacusprofundi] | TAWASASTFR | MCS | NODE_2058334 length_16777_co v_21.860703 | e-124 |
| BAL8621 7 | GH5 | putative glycosyl hydrolase [Actinoplanes missouriensis] | ELLAGIK | MS | | |

As sequências desses *contigs* foram analisadas pelo software Geneious Pro 5.4.6, através do qual é possível a identificação de ORFs. As ORFs podem ser identificadas através das ferramentas Glimmer e ORFFinder. Dentre as ORFs identificadas em cada um dos 6 frames, determinamos a correta a partir da comparação com o banco de dados do NCBI através da ferramenta blastp ou tblastx. As ORFs foram então anotadas e analisadas também a classificação em relação às famílias de glicosil-hidrolases (GHs) através da ferramenta pfam (disponível em: <http://pfam.sanger.ac.uk/>). Tal comparação foi possível, pois as diversas famílias de glicosil hidrolases existentes são caracterizadas pela existência de sequências de aminoácidos conservados em determinadas posições (GILBERT, 2010).

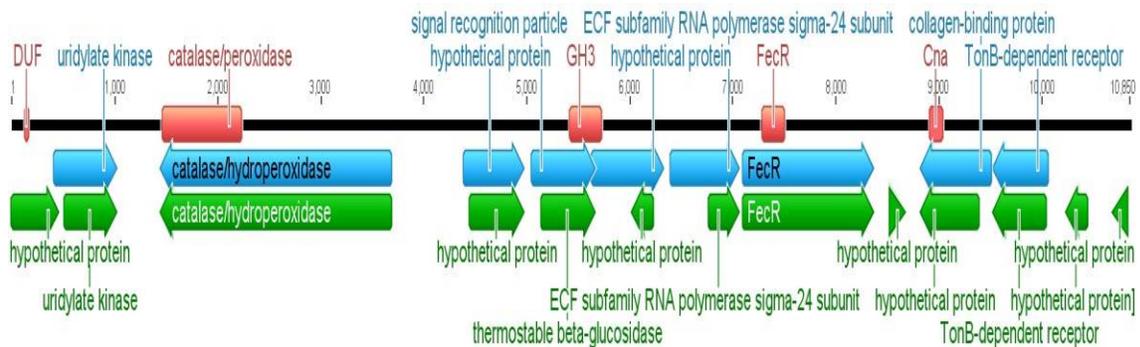
Na Figura 30 encontra-se o esquema da anotação do *contig* que contem a GH5. Esse *contig* apresenta um conjunto interessante de enzimas a serem estudadas como celulases, xyloglucanases e dominios de ligação ao carboidrato. Nota-se que a predição de genes tanto pelo Glimmer quanto pelo ORFFinder coincidiram e quando submetidas essas sequencias no banco NCBI são encontrados genes homologos. Nota-se ainda que os genes encontrados apresentam baixa homologia (e-Value alto e baixa identidade) com a sequencias sugerindo a possibilidade das enzimas codificadas pelo contig serem novas.

Figura 30: Anotação do Contig referente a GH5 pelo programa Geneious. Setas em verde representam predição de genes pelo software glimmer, em azul a predição de ORFs pelo software ORFFinder e em vermelho estão anotados os domínios encontrados nas buscas das sequencias no banco de dados do NCBI.



Na Figura 31 encontra-se o esquema da anotação do *contig* que contém a catalase. Esse *contig* apresenta um conjunto interessante de enzimas a serem estudadas como a enzima multifuncional catalase/peroxidase, uma GH3 termoestável e uma série de enzimas hipotéticas, que podem estar relacionadas ao metabolismo de carboidratos. Novamente a predição de genes tanto pelo Glimmer quanto pelo ORFFinder coincidiu e quando submetidas essas sequências ao banco NCBI são encontrados genes homólogos, porém de baixa homologia. Estudos futuros utilizarão as sequências aqui anotadas para síntese dos genes e caracterização funcional das enzimas apresentadas.

Figura 31: Anotação do Contig referente a catalase pelo programa Geneious. Setas em verde representam predição de genes pelo software glimmer, em azul a predição de ORFs pelo software ORFFinder e em vermelho estão anotados os domínios encontrados nas buscas das sequencias no banco de dados do NCBI.



5. Conclusão

- Neste estudo foram estabelecidas duas comunidades microbianas enriquecidas (MCS e MCF), visando a degradação de biomassa lignocelulósica.
- O enriquecimento é sustentado pelas análises de metagenômica na qual notou-se a reestruturação das comunidades e a predominância dos grupos bacterianos proteobacterias e bacteroidetes, filos conhecidos como decompositores de matéria orgânica. Dentre as ordens enriquecidas encontram-se Sphingomonadales, Pseudomonadales, Cytophagales Rhodobacterales, Shingobacterales, Flavobacterales, Xanthomonadales e Bacteroidales.
- As análises metaproteômicas também sugerem enriquecimento, sendo encontradas alterações no perfil funcional destas comunidades, possivelmente devido ao estabelecimento de micro-organismos degradadores de biomassa durante o cultivo do consórcio.
- O trabalho contribui para projetos futuros, os quais poderão utilizar os resultados obtidos neste trabalho para o estudo de enzimas de interesse, e através de ferramentas moleculares e bioquímicas poderão prospectar e avaliar o potencial das enzimas existentes nestas comunidades.
- As análises sugerem ainda que espécies reativas de oxigênio (ROS) podem ter papel importante na degradação da biomassa dessas comunidades, porém mais estudos são necessários para confirmação e entendimento do seu mecanismo de ação.

6. Referências

Ageitec (http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_20_3112006152934.html)

ANDREOTE, F.D. et al. The Microbiome of Brazilian Mangrove Sediments as Revealed by Metagenomics. **PLoS ONE**, v 7(6): e38600. 2012.

ARANTES, V. et al. Lignocellulosic polysaccharides and lignin degradation by wood decay fungi: the relevance of nonenzymatic Fenton-based reactions. **Journal of industrial microbiology & biotechnology**, v. 38, n. 4, p. 541-555. 2011.

AYLWARD, F.O. et al. Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. **ISME Journal**: 1–14. 2012.

BAGWELL, C.E. et al. Survival in nuclear waste, extreme resistance, and potential applications gleaned from the genome sequence of *Kineococcus radiotolerans* SRS30216. **PLoS One** 3:e3878. 2008.

BAKER, J.O. et al. Investigation of cell wall loosening protein expansin as a possible additive in the enzymatic saccharification of lignocellulosic biomass. **Applied Biochemistry and Biotechnology**. v. 84-86, p. 217 – 223. 2000.

BALTZ, R.H. Renaissance in antibacterial discovery from actinomycetes. **Curr. Opin. Pharmacol.** 8:557–563. 2008.

BARABOTE, R.D. et al. Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. **Genome Res.** 19:1033–1043. 2009.

BARBEYRON, T. et al. *Zobellia galactanovorans* gen. nov., sp. nov., a marine species of Flavobacteriaceae isolated from a red alga, and classification of [*Cytophaga*] *uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. **Int. J. Syst. Evol. Microbiol.** 51, 985–997. 2001.

BARBEYRON, T. et al. *Mariniflexile fucanivorans* sp. nov., a marine member of the Flavobacteriaceae that degrades sulphated fucans from brown algae. **Int. J. Syst. Evol. Microbiol.** 58, 2107–2113. 2008.

BARBUSIŃSKI, K. Fenton reaction-controversy concerning the chemistry. **Ecological chemistry and engineering**, v. 16, p. 347-358. 2009.

BAUER, M. et al. Whole genome analysis of the marine Bacteroidetes “*Gramella forsetii*” reveals adaptations to degradation of polymeric organic matter. **Environ. Microbiol.** 8, 2201–2213. 2006.

BEHAL V. Bioactive products from *Streptomyces*. **Adv. Appl. Microbiol.** 47:113–156. 2000.

BENNDORF, D. et al. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. **ISME J.** 1:224–234, 2007.

BENTLEY, S.D. et al. Genomics of Actinobacteria, the high GC Gram-positive bacteria, p 333–359. In FRASER, C.M.; READ, T.D.; Nelson, K.E. (ed), **Microbial genomes**. Humana Press, 2004, Totowa, NJ.

BENTLEY, S.D. et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). **Nature**, v417:141–147. 2002.

BERDY, J. Bioactive microbial metabolites. **J. Antibiot.**, Tokyo, v 58: 1–26. 2005.

BERG, J. O.; LINDQVIST, L.; NORD, C. E. Purification of glycoside hydrolases from *Bacteroides fragilis*. **Appl. Environ. Microbiol.** v40, 40–47. (1980).

BERNARDET, J.F.; NAKAGAWA, Y. An introduction to the family of Flavobacteriaceae. **Prokaryotes** v7, 455–480. 2006.

BIOMASS ENERGY CENTER. **Biomass** Disponível em <biomassenergycentre.org.uk>. Acesso em 23-ago-2013.

BON, E. P. S.; GIRIO F. PERREIRA JR., N. Enzimas na produção de etanol. In: BON, E. P. S.; FERRARA, M. A.; CORVO, M. A. **Enzimas em Biotecnologia: Produção, Aplicações e Mercado**. Rio de Janeiro: Editora Interciência, 2008. p. 241- 271.

BORNEMAN, J. et al. Molecular microbial diversity of an agricultural soil in Wisconsin. **Appl. Environ. Microbiol.** 62, 1935–1943. 1996.

BRADFORD, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Anal Biochem**, v72:248-254. 1976.

BRASIL, Ministério de Minas e Energia. **Geração termelétrica a partir da biomassa**. Plano Nacional de Energia 2030; colaboração Empresa de Pesquisa Energética. v.8. Brasília, 250p. 2007.

BROWN, R. M.; Jr., SAXENA, I. M.; KUDLICKA, K. Cellulose biosynthesis in higher plants. **Trends in Plant Science**. v. 1(5), p. 149-155. 1996.

BRUNE, A. Termite guts: the world's smallest bioreactors. **Trends Biotechnol**, v16:16-21. 1998.

BUCKLEY, D. H.; SCHMIDT, T. M. Diversity and dynamics of microbial communities in soils from agro-ecosystems. **Environ. Microbiol.**, v 5, 441–452. 2003.

BULL, A.T. et al. **Marine actinobacteria**: perspectives, challenges, future directions. *Antonie Van Leeuwenhoek* 87:65–79. 2005.

BULL, A.T.; STACH, J.E.M. Marine actinobacteria: new opportunities for natural product search and discovery. **Trends Microbiol.** 15:491– 499. 2007.

BULL, A. T.; BUNCH, A. W.; ROBINSON, G. K. Biocatalysts for clean industrial products and processes. **Curr Opin Microbiol**, v. 2, p. 246-51, 1999.

BULL, A. T.; WARD, A. C.; GOODFELLOW, M. Search and discovery strategies for biotechnology: the paradigm shift. **Microbiol Mol Biol Rev**, v. 64, p. 573-606, 2000.

BURNUM, K. E. et al. Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. **The ISME journal**. 2010.

BUSHNELL, L. D., & HAAS, H. F. Utilization of certain hydrocarbons by microorganisms. **J. Bacteriol.**, 41:653-673, 1941.

CAFFAL, K. H.; MOHNEN, D. The structure, function and biosynthesis of plant cell wall pectic polysaccharides. **Carbohydrate Research**. v. 344. p. 1879-1900. 2009.

CAMPBELL, J. E. and BLOCK, E. Land use and alternative Bioenergy Pathways for Waste Biomass. **Environ Sci Technol**, 2010.

CHATER, K.F. et al. The complex extracellular biology of *Streptomyces*. **FEMS Microbiol. Rev.** 34:171–198. 2010.

CHATER, K.F. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. **Philos. Trans. R. Soc. Lond. B Biol. Sci.** 361:761–768. 2006.

CHEN, K.; PACTER, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. **PLoS Comp Biol** 1(2): e24, 2005.

CHURCH, M. J. "Resource control of bacterial dynamics in the sea," in **Microbial Ecology of the Oceans**, 2nd Edn, ed. D. L. Kirchman (Hoboken, NJ: Wiley & Sons), (2008), 335–382.

COLE, S.T. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. **FEBS Lett.** 452:7–10. 1999.

COLLIER, L.; BALOWS, A.; SUSSMAN, M. Topley and Wilson's Microbiology and Microbial Infections. **Systematic Bacteriology** Vol. 2., Arnold, London. 1998.

COWAN, D. et al. Metagenomic gene discovery: past, present and future. **TRENDS in Biotechnology**, v. 23, 321-329, 2005.

D’COSTA, V.M. et al. Sampling the antibiotic resistome. **Science** v 311:374–377. 2006.

DAVIES, G.; HENRISSAT, B. Structures and mechanisms of glycosyl hydrolases. **Structure**, v 3:853-859. 1995.

DANIEL, R. The metagenomics of soil. **Nature Review Microbiology**, v.3, 470-478, 2005.

DEL RIO, T. G. et al. Complete genome sequence of Chitinophaga pinensis type strain (UQM 2034T). **Stand. Genomic Sci.** 2, 87–95. 2010.

DEMIRBAS, A. Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. **Energy Conversion and Management**, v.49, 2106-2116, 2008.

DUAN, C.J.; FENG, J.X. Mining Metagenomes for Novel Cellulases Genes. **Biotechnology Letters**, v.32, p.1765-1775, 2010.

EDWARDS, J. L. et al. Identification of Carbohydrate Metabolism Genes in the Metagenome of a Marine Biofilm Community Shown to Be Dominated by Gammaproteobacteria and Bacteroidetes. **Genes**, v1, 371-384. 2010.

EUZEBY, J.P. 2011. **List of prokaryotic names with standing in nomenclature.** Disponível em <<http://www.bacterio.cict.fr/a/actinobacteria>>. Acesso 3-fev-2012. 2011.

FENG, J. X. et al. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. **Applied Microbiology and Biotechnology**, v.75, p.319-328, 2007.

FERREIRA, A.C. et al. Characterization and radiation resistance of new isolates of Rubrobacter radiotolerans and Rubrobacter xylanophilus. **Extremophiles** 3:235–238. 1999.

FIEDLER, H.P. et al. Marine actinomycetes as a source of novel secondary metabolites. **Antonie Van Leeuwenhoek**, v87:37– 42. 2005.

FIERER, N. et al. Cross-biome metagenomics analyses of soil microbial communities and their functional attributes. **Proceedings of the National Academy of Sciences**, v 109: 21390–21395. 2012.

FRANCO CAIRO, J.P.L. et al. Functional characterization and target discovery of Glycoside Hydrolases from Lower Termite Coptotermes gestroi Digestome. **Biotechnol Biofuels**, v 4: 50. 2011.

Ghatak, S.B.; Panchal, S.J. Ferulic Acid – An Insight Into Its Current Research and Future Prospects, **Trends in Food Science & Technology**, doi: 10.1016/j.tifs.2010.11.004. 2010.

GILBERT, H. J. The biochemistry and structural biology of plant cell wall deconstruction. **Plant physiology**, v 153, 444-55. 2010.

GILL, S.R. et al. Metagenomic analysis of the human distal gut microbiome. **Science** 312:1355–1359. 2006.

GRANT, S. et al. A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. **Extremophiles**, v. 8, p.421-429, 2004.

GUZZARONI, M.E. et al. Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. **ISME J.** Jan;7(1):122-36. 2013

HAN, C. et al. Complete genome sequence of Pedobacter heparinus type strain (HIM 762-3T). **Stand. Genomic Sci.** v 1, 54–62. 2009.

HATFIELD, R., VERMERRIS, W. Lignin formation in plants. The dilemma of linkage specificity. **Plant Physiology**. v. 126. p. 1351-1357. 2001.

HESS, M. et al. fMetagenomic discovery of biomass-degrading genes and genomes from cow rumen. **Science** v. 331, pp. 463–467, 2011.

Holt, J.G. et al. **Bergey’s Manual of Determinative Bacteriology**. Williams and Wilkins (1994), Baltimore, MD.

HOPWOOD, D.A. **Streptomyces in nature and medicine**. Oxford University Press (2007), New York, NY.

INOUE, K. et al. Characterization of novel carbazole catabolism genes from gram-positive carbazole degrader *No-cardioides aromaticivorans* IC177. **Appl. Environ. Microbiol.** 72:3321– 3329. 2006.

ISHII, T. et al. The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalent cross-linked dimer. **The Journal of Biological Chemistry.** v. 274 (19). p. 13098-13104. 1999.

JEFFERSON, R.A.; BURGESS, S.M.; HIRSH, D. "beta-Glucuronidase from *Escherichia coli* as a gene-fusion marker." **Proc Natl Acad Sci U S A**;v 83(22);8447-51. PMID: 3534890 (1986).

JU, F. et al. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. **Environmental Microbiology Reports.** doi: 10.1111/1758-2229.12110. 2013.

KALINOWSKI, J. et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. **J. Biotechnol.**, v 104:5–25. 2003.

KAN, J. et al. Metaproteomic analysis of Chesapeake Bay microbial communities. **Saline Syst.** 1:7, 2005.

KANEKO, S. et al. Relationship between production of hydroxyl radicals and degradation of wood, crystalline cellulose, and a lignin-related compound or accumulation of oxalic acid in cultures of brown-rot fungi. **Journal of Wood Science**, v. 51, n. 3, p. 262-269. 2005.

KATSUMATA, K. et al. Structural changes in lignin of tropical woods during digestion by termite *Cryptotermes brevis*. **J Wood Sci**, v 53:419-426. 2007.

KELLER, A. et al. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. **Anal Chem.**, v 74(20):5383–5392. 2002.

KIM, S.; DALE, B. Global potential bioethanol production from wasted crops and crop residues. **Biomass Bioenergy**, v 26, 361. 2004.

KIM, B. Y. et al. *Flavobacterium daejeonense* sp. nov. and *Flavobacterium suncheonense* sp. nov., isolated from greenhouse soils in Korea. **Int. J. Syst. Evol. Microbiol.**, v 56, 1645–1649. 2006.

KIM, S. J. et al. Characterization of a gene encoding cellulase from uncultured soil bacteria. **FEMS Microbiology Letters**, v.282, p. 44-51, 2008.

KIRCHMAN, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. **FEMS Microbiol. Ecol.**, v 39, 91–100. 2002.

KLAASSENS, E. S.; DE VOS, W. M.; VAUGHAN, E. E. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. **Appl. Environ. Microbiol.**, v 73:1388–1392, 2007.

KOIDE, T.; PANG, W. L.; BALIGA, N. S. The role of predictive modelling in rationally re-engineering biological systems. **Nature Rev. Microbiol.**, v 7, 297–305. 2009.

Koivula, A. et al. Polysaccharide-Acting Enzymes and Their Applications. In **The European Polysaccharide Network of Excellence (EPNOE)** (pp. 375-392). Springer Vienna. (2013).

Kuwahara, T. et al. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. **Proc. Natl. Acad. Sci. U.S.A.** v 101, 14919–14924. 2004.

LAEMMLI, U.K. Cleavage of structural proteins during the assembly of head of bacteriophage T4. **Nature**, v 227:680-683. 1970.

LAIL, K. et al. Complete genome sequence of *Spirosoma linguale* type strain (1T). **Stand. Genomic Sci.**, v 2, 176–185. 2010.

LANG, E. et al. Complete genome sequence of *Dyadobacter fermentans* type strain (NS114T). **Stand. Genomic Sci.**, v 1, 133–140. 2009.

LAUBER, C. L. et al. Pyrosequencing- based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. **Appl. Environ. Microbiol.**, v 75, 5111–5120. 2009.

LEE, Y.; CHOI, D.; KENDE, H. Expansins: ever-expanding number and functions. **Current Opinion in Plant Biology**. v.4, p. 527 – 532. 2001.

LEE, M. et al. *Dyadobacter soli* sp. nov., a starch-degrading bacterium isolated from farm soil. **Int. J. Syst. Evol. Microbiol.**, v 60, 2577–2582. 2010.

LOMBARD, V. et al. The Carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Res.**, v 42:D490–D495. 2014.

LYKIDIS, A. et al. Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. **J. Bacteriol.**, v 189:2477–2486. 2007.

LYND, L.R. et al. Consolidated bioprocessing of cellulosic biomass: an update. **Current Opinion in Biotechnology**, v 16:577–583. 2005.

LYND, L.R. et al. How biotech can transform biofuels. **Nature Biotechnology**, v.26, 169-172, 2008.

MADIGAN, M.T.; MARTINKO, J.M. **Brock Biology of Microorganisms**. (11th ed.) Upper Saddle River, NJ, USA. Prentice Hall, (2005.)

MANSFIELD, S. D.; DE JONG, E.; SADDLER, J. N. Cellobiose dehydrogenase, an active agent in cellulose depolymerization. **Applied and environmental microbiology**, v. 63, n. 10, p. 3804-3809. 1997.

MARTINEZ-ALONSO, M. et al. Diversity of the bacterial community in the surface soil of a pear orchard based on 16S rRNA gene analysis. **Int. Microbiol.**, v 13, 123–134. 2010.

MARTÍNEZ, A.T. et al. Biodegradation of lignocelluloses: microbial, chemical and enzymatic aspects of the fungal attack of lignin. **International Microbiology**. v. 8, p. 195-204. 2005.

MARTINEZ, D. et al. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. **Proceedings of the National Academy of Sciences**, v. 106, n. 6, p. 1954-1959. 2009.

MARTINS, L.F. et al. Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. **PLoS ONE**, v 8(4): e61928. 2013.

MARX, C.J. Can You Sequence Ecology? Metagenomics of Adaptive Diversification. **PLoS Biol**, v 11(2): e1001487. 2013.

MCBRIDE, M. J. et al. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. **Appl. Environ. Microbiol.**,v 75, 6864–6875. 2009.

MCLEOD, M.P. et al. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. **Proc. Natl. Acad. Sci. U. S. A.**,v 103:15582–15587. 2006.

MEYER, F. et al. The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v 9:386. 2008.

MONGODIN, E.F. et al. Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. **PLoS Genet.**, v 2:e214. 2006.

MURRAY, R.G.E. et al. Report of the Ad Hoc Committee on approaches to taxonomy within the Proteobacteria. **Int. J. Syst. Bacteriol.**, v 40, 213-215. 1990.

NESVIZHSKII, A.I. et al. A statistical model for identifying proteins by tandem mass spectrometry. **Anal Chem.**, v 75(17):4646–4658. 2003.

NG, S. **Pure culture and metagenomic approaches to investigate cellulose and xylan degradation**. 150 págs. Dissertação de Mestrado. Faculty of Science Applied Bioscience. University of Ontario Institute of Technology, 2010.

NOGUEIRA, L. A. H. ; LORA, E.S. **Dendroenergia: Fundamentos e Aplicações**. 2. ed. Rio de Janeiro: Interciência, 2003. 199 p.

- NORMAND, P. et al. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. **Genome Res.**, v 17:7–15. 2007.
- NÜSSLEIN, K.; TIEDJE, J. M. Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. **Applied and Environmental Microbiology**, Baltimore, V.65 n8, p. 3622-3626. 1999.
- OLANO, C.; MENDEZ, C.; SALAS, J.A. Antitumor compounds from actinomycetes: from gene clusters to new derivatives by combinatorial biosynthesis. **Nat. Prod. Rep.**, v 26:628–660. 2009.
- OLD, W. M. et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. **Molecular & cellular proteomics**, v. 4, n. 10, p. 1487-1502, 2005.
- PAIVA, L. B. et al. Ferulic acid and derivatives: molecules with potential application in the pharmaceutical field. **Braz. J. Pharm. Sci.** [online], vol.49, n.3, pp. 395-411. ISSN 1984-8250. 2013.
- PAIXÃO, D.A.A. et al. Molecular analysis of the bacterial diversity in a specialized consortium for diesel oil degradation. **Revista Brasileira de Ciência do Solo**, v. 34, p.773-781, 2010.
- PANG, H. et al. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. **Current Microbiology**, v. 58, p.404-408, 2009.
- PANKRATOV, T. A. et al. Isolation of aerobic, gliding, xylanolytic and laminarinolytic bacteria from acidic Sphagnum peatlands and emended description of *Chitinophaga arvensicola* Kampfer et al., 2006. **Int. J. Syst. Evol. Microbiol.** 56, 2761–2764. 2006.
- PÉREZ, J. et al. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. **International Microbiology**, v. 5 (2), p. 53-63. 2002.
- QIN, Q. L. et al. The complete genome of *Zunongwangia profunda* SM-A87 reveals its adaptation to the deep-sea environment and ecological role in sedimentary organic nitrogen degradation. **BMC Genomics**, v 11, 247. 2010.
- RIBEIRO, D.A. et al. The *Penicillium echinulatum* secretome on sugar cane bagasse. **PloS one**, v. 7, n. 12, p. e50571. 2012.
- REES, H. C. et al. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. **Extremophiles**, v. 7, p.415-421, 2003.
- REGALBUTO, J.R. The sea change in US biofuels' funding: From cellulosic ethanol to green gasoline. **Biofuel Bioprod Bioref**, v 5(5):495–504. 2011.
- REICHENBACH, H. The order cytophagales. **Prokaryotes**, v 7, 549–590. 2006.
- RIES, L. et al. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. **BMC Genomics**, v 14:541. 2013.
- RUBIN, E.M. Genomics of cellulose biofuels. **Nature**, v 454 (14), 841-845. 2008.
- RUDNEY, J. D. et al. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. **Mol. Oral Microbiol.**, v 25:38–49. 2010.
- SADDLER, J.N. Bioconversion of Forest and Agricultural Plant Residues. **International Biotechnology Agricultural**, Ser, 9, C.A.B. Internacional, Wallingford, UK, 1993.
- SAVIOLA, B.; BISHAI, W. The genus *Mycobacterium*—medical, p 919– 933. In DWORKIN, M. et al. (ed), **The prokaryotes: an evolving electronic resource for the microbiological community**. Springer-Verlag, New York, NY. 2006.

SCHNEIDER, T.; RIEDEL, K. Environmental proteomics: analysis of structure and function of microbial communities. **Proteomics**, v 10:785–798, 2010.

Schneider, T. et al. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. **ISME J.**, v 6 1749–1762. 2012.

SCHREMPF, H. Recognition and degradation of chitin by streptomycetes. **Antonie Van Leeuwenhoek**, v 79:285–289. 2001.

SCHULZE, W. X. et al. A proteomic fingerprint of dissolved organic carbon and of soil particles. **Oecologia**, v 142:335–343. 2005.

SEXTON, S. et al. The role of biotechnology in a sustainable biofuel future. **AgBioForum**, v 12(1), 130-140. Available on the World Wide Web: <http://www.agbioforum.org>. 2009.

SHEVCHENKO, A. et al. In-gel digestion for mass spectrometric characterization of proteins and proteomes. **Analytical Chemistry**, v 68, 850-858. 1996.

SIMMONS, B.; LOQUE, D.; BLANCH, H.W. Next-generation biomass feedstocks for biofuel production. **Genome Biol**, v 9:242. 2008.

SIMON, C.; DANIEL, R. Metagenomic Analyses: Past and Future Trends. **Applied and Environmental Microbiology**, v.77, No. 4, p.1153-1161. 2011.

SMITH, J.J. et al. Bacterial diversity in three different Antarctic cold desert mineral soils. **Microb. Ecol.**, v 51:413–421. 2006.

SOWELL, S. M. et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. **ISME J.** 3:93–105. 2009.

STACKEBRANDT, E.; MURRAY, R.G.E.; TRUPPER, H.G. Proteobacteria classis nov., a name for the phylogenetic taxon that includes the 'purple bacteria and their relatives'. **Int. J. Syst. Bacteriol.**, v 38, 321-325. 1988.

SWEENEY, M. D.; XU, F. Biomass Converting Enzymes as Industrial Biocatalysts for Fuels and Chemicals: Recent Developments. **Catalysts**, v 2(2), 244-263. 2012.

TAIZ, L.; ZEIGER, E. Paredes celulares: Estrutura, Biogênese e Expansão. In: _____ **Fisiologia Vegetal**; tradução Eliane Romanato Santarém et al. 4 ed. Porto Alegre: Artmed, p. 407-434. 2009.

TAIZ, L., ZEIGER, E. Cell wall: Structure, Biogenesis and Expansion. In: _____ **Plant Physiology**; Sunderland, MA: Sinauer Associates, 2002. p. 313- 338.

TAKARADA, H. et al. Complete genome sequence of the soil actinomycete *Kocuria rhizophila*. **J. Bacteriol.**, v 190:4139–4146. 2008.

TERRI, T.T. et al. *Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose? **Biochem Soc Trans.**, v 26:173-178. 1998.

TEERI, T.T. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. **Tibtech.**, v 15:160-167. 1997.

Thureborn, P. et al. A Metagenomics Transect into the Deepest Point of the Baltic Sea Reveals Clear Stratification of Microbial Functional Capacities. **PLoS ONE**, v 8(9): e74983. doi:10.1371/journal.pone.0074983. 2013.

TIERNY, Y. et al. Molecular cloning and expression in *Escherichia coli* of genes encoding pectate lyase and pectin methylesterase activities from *Bacteroides thetaiotaomicron*. **J. Appl. Bacteriol.**, v 76, 592–602. 1994.

TILBURN, J. et al. Transformation by integration in *Aspergillus nidulans*. **Gene**, v 26(2-3), 205-21. 1983.

TOYODA, A. et al. Isolation and identification of cellulose-binding proteins from sheep rumen contents. **Appl. Environ. Microbiol.** 75:1667–1673, 2009.

TRÜPER, H.G. Phototrophic bacteria (an incoherent group of prokaryotes). A taxonomic versus phylogenetic survey. **Microbiologia**, v 3, 71-89. 1987.

UENOJO, M. & PASTORE, G. M. Pectinases: aplicações industriais e perspectivas. **Química Nova**. v. 30, n. 2, p. 388-394. 2007.

VAN DEN BRINK, J.; DE VRIES, R.P. Fungal enzyme sets for plant polysaccharide degradation. **Appl. Microbiol. Biotechnol.**, v 91, 1477–1492. 2011.

VENTURA, M. et al. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. **Microbiol. Mol. Biol. Rev.**, v 71:495– 548. 2007.

VERBERKMOES, N. C. et al. Shotgun metaproteomics of the human distal gut microbiota. **ISME J.**, v 3:179–189, 2009.

VIËTOR, R. J. et al. Conformation features of crystal-surface cellulose from higher plants. **The Plant Journal**. v. 30 (6), p. 721-731. 2002.

WARNECKE, F. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. **Nature**, v 450:560–565, 2007.

WATANABE, H.; TOKUDA, G. Cellulolytic Systems in Insects. **Annu. Rev. Entomol.**, v. 55, pp. 609–632, 2010.

WEI, H. et al. Natural paradigms of plant cell wall degradation. **Current Opinion in Biotechnology**., v. 20, p. 330-338. 2009.

WILMES, P.; BOND, P. L. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. **Environ. Microbiol.**, v 6:911–920. 2004.

WILMES, P.; BOND, P.L. Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. **Current Opinion in Microbiology**, v 12:310–317, 2009.

WOESE, C.R. The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria. In: **Evolution at Molecular Level** (Selander, R.K., Clark, A.G. and Whittmay, T.S., Eds.), pp. 1-24 Sinauer, Sunderland, MA. 1991.

WOESE, C.R. Bacterial evolution. **Microbiol. Rev.**, v 51, 221-271. 1987.

WOOD, T.M.; BHAT, K.M. Methods for measuring cellulase activities. **Methods Enzymol.**, v. 160:87-112. 1988.

WYMELENBERG, A. V. et al. Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. **Applied and environmental microbiology**, v. 76, n. 11, p. 3599-3610. 2010.

Xu, J. et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. **Science**, v 299, 2074–2076. 2003.

YANG, B. et al. Enzymatic hydrolysis of cellulosic biomass. **Biofuels**, v2:4, 421-450. 2011.

ZAVARZIN, G.A.; STACKEBRANDT, E.; MURRAY, R.G.E. A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. **Can. J. Microbiol.**, v. 37, 1-6. 1991.

ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome research**, v. 18, n. 5, p. 821-829, 2008.

ZHI, X.Y.; LI, W.J.; STACKEBRANDT, E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. **Int. J. Syst. Evol. Microbiol.**, v 59:589–608. 2009.

ZHOU, J. P.; HUANG, Y.; MO, M. Phylogenetic analysis on the soil bacteria distributed in karst forest. **Braz. J. Microbiol.**, v 40, 827–837. 2009.

7. Anexos

Tabela 7: Lista de proteínas encontradas nas análises metaproteômicas da comunidade do solo, o banco de dados utilizado foi o banco do CAZy.

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| A8EV24 | AA NC | Catalaseperoxidase [Arcobacter butzleri strain RM4018]. | 94,90% | 1 | 1 | 1 | 1,36% | DWWPNQLNLK |
| Q6ZXC2 | AA NC | Catalase [Claviceps purpurea]. | 93,80% | 1 | 1 | 1 | 2,54% | QIGADDAESATR |
| Q9X3V2 | AA1 | Multicopper oxidase CumA [Pseudomonas putida]. | 94,90% | 1 | 2 | 2 | 2,61% | TCADRPIATLQK |
| Q9UVU1 | AA3 | Alcohol oxidase 2 [Ogataea methanolica]. | 98,30% | 1 | 1 | 1 | 2,11% | DGPLTTNGIEAGVK |
| YP_001354501 | AA3 | Glucosemethanolcholine oxidoreductase [Janthinobacterium sp strain Marseille]. | 92,90% | 1 | 1 | 1 | 1,19% | MTIEMK |
| Q2YZG8 | AA9 | Putative uncharacterized protein [uncultured gamma proteobacterium]. | 93,90% | 1 | 1 | 1 | 5,52% | GWVAVGFSPK |
| CBG67596 | CBM13 | probable secreted alpha-galactosidase [Streptomyces scabiei 8722]. | 93,90% | 1 | 1 | 1 | 1,82% | WTPQADGSIR |
| BAL89256 | CBM2 | putative glycosyl hydrolase [Actinoplanes missouriensis 431]. | 93,30% | 1 | 1 | 1 | 2,77% | NVSWNGALAPR |
| ACL75605 | CBM3 | glycoside hydrolase family 9 [Clostridium cellulolyticum H10]. | 94,90% | 1 | 1 | 1 | 1,90% | DGAVNALDFANLK |
| ACU72301 | CBM32 | coagulation factor 58 type domain protein [Catenulispora acidiphila DSM 44928]. | 92,30% | 1 | 1 | 1 | 1,17% | YNAHMLVLTANELPIK |
| ACT03290 | CBM6 | alpha-1,2-mannosidase [Paenibacillus sp JDR2]. | 92,90% | 1 | 1 | 1 | 0,57% | LGSPTGLTVGK |
| AFH60159 | CBM6 | BglA2 [Paenibacillus mucilaginosus K02]. | 98,70% | 1 | 1 | 1 | 0,53% | LNVSQPGK |
| AGK76953 | CBM6 | Glycoprotein endo-alpha-1,2-mannosidase [Streptomyces fulvissimus DSM 40593]. | 93,90% | 1 | 1 | 1 | 1,80% | VSSAGAGGTIEVR |
| ABG58512 | CBM9 | CBM9 module, Glycoside Hydrolase Family 8 protein and Carbohydrate Esterase Family 4 protein [Cytophaga hutchinsonii ATCC 33406]. | 92,90% | 1 | 1 | 1 | 0,63% | GTTVSVLPAGR |
| ADF53787 | CE1 | putative esterase [Zunongwangia profunda SMA87]. | 93,60% | 1 | 1 | 1 | 1,07% | LYQELSGK |
| AER68962 | CE1 | Poly-3-hydroxybutyrate depolymerase [Corynebacterium pseudotuberculosis 106A]. | 93,90% | 1 | 1 | 1 | 4,77% | FAEGEAGNDALIVFPAGR |
| ABH76381 | CE11 | Sequence 6224 from patent US 7041814 [Unknown]. | 96,50% | 1 | 1 | 1 | 2,74% | YIEEIAGAR |
| ABS43647 | CE11 | UDP30 acyl-N-acetyl-glucosamine deacetylase [Campylobacter jejuni subsp doylei 26997]. | 91,30% | 1 | 1 | 1 | 2,04% | QLTLVK |
| AEI35086 | CE11 | UDP-3-O-acyl-N-acetylglucosamine deacetylase[Francisella sp TX077308]. | 93,90% | 1 | 1 | 1 | 3,86% | EAGIVEQESNR |
| AFJ91013 | CE11 | bifunctional UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase/(3R)-hydroxymyristoyl-ACP dehydratase [Blattabacterium sp Blaberus giganteus]. | 98,30% | 1 | 1 | 1 | 2,01% | NDIPEFDLK |
| CCA53866 | CE12 | hypothetical protein SVEN_0579 [Streptomyces venezuelae ATCC 10712]. | 93,10% | 1 | 1 | 1 | 1,68% | HLLTGMKPGER |
| ABY22908 | CE14 | conserved hypothetical protein [Renibacterium salmoninarum ATCC 33209]. | 93,10% | 1 | 1 | 1 | 3,67% | TAELNQAAAELGVSQR |
| BAJ62819 | CE14 | putative deacetylase [Anaerolinea thermophila UNI1]. | 93,20% | 1 | 1 | 1 | 3,19% | AAEITGVK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| CCG01889 | CE14 | 1D-myo-inositol-2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase [Blastococcus saxosidens DD2]. | 94,90% | 1 | 1 | 1 | 4,01% | AAMAALGVEDVR |
| ADU23224 | CE2 | lipolytic protein GDLS family [Ruminococcus albus 7]. | 94,90% | 1 | 1 | 1 | 2,27% | IAVLVDGER |
| AAP10088 | CE4 | Peptidoglycan N-acetyl-glucosamine deacetylase [Bacillus cereus ATCC 14579]. | 93,10% | 1 | 1 | 1 | 2,50% | WNSIPAK |
| ACL77751 | CE4 | Carbohydratebinding family 9 [Clostridium cellulolyticum H10]. | 93,00% | 1 | 1 | 1 | 1,16% | SASLEDLK |
| BAA23389 | CE4 | YfjS [Bacillus subtilis]. | 94,90% | 1 | 1 | 1 | 3,04% | QLNSLIEK |
| CAG67919 | CE4 | putative lipoprotein Haemin storage system HmsF [Acinetobacter sp ADP1]. | 94,20% | 1 | 1 | 3 | 1,74% | QMNLTDNNRPQK |
| Q87LH4 | CE4 | Deacetylase DA1 [Vibrio parahaemolyticus serotype O3:K6 strain RIMD 2210633]. | 94,90% | 1 | 1 | 1 | 1,41% | ANEFAR |
| AGA67828 | CE7 | acetyl esterase deacetylase [Desulfitobacterium dichloroeliminans LMG P21439]. | 99,00% | 1 | 1 | 1 | 2,26% | IELSLDER |
| AGA77907 | CE7 | acetyl esterase deacetylase [Echinicola vietnamensis DSM 17526]. | 93,40% | 1 | 1 | 1 | 2,77% | SEAAQTPMQIEK |
| CCI71608 | CE7 | cephalosporinC deacetylase [Paenibacillus polymyxa M1]. | 94,90% | 1 | 1 | 1 | 2,19% | ALQELER |
| AAL49830 | CE8 | putative pectin methylesterase [Arabidopsis thaliana]. | 93,20% | 1 | 1 | 1 | 2,18% | SIVDSSVGNLNR |
| ABD47735 | CE8 | pectin methylesterase [Eucalyptus globulus subsp globulus]. | 92,50% | 1 | 1 | 1 | 7,61% | MEELISR |
| ABE80620 | CE8 | Pectinesterase Pectinesterase inhibitor [Medicago truncatula]. | 90,10% | 1 | 1 | 1 | 1,03% | LLQDYK |
| ABN05750 | CE8 | Pectinesterase [Medicago truncatula]. | 93,10% | 1 | 1 | 1 | 3,86% | IMDAVMAAPNGSK |
| ADL52345 | CE8 | Pectinesterase [Clostridium cellulovorans 743B]. | 99,30% | 1 | 1 | 1 | 0,47% | LGGEGLAVK |
| CAW41288 | CE8 | unnamed protein product [Glycine max]. | 91,20% | 1 | 1 | 1 | 1,71% | NQNTGIIIQK |
| ACR30215 | CE9 | N-acetyl-glucosamine-6-phosphate deacetylase [Burkholderia glumae BGR1]. | 91,90% | 1 | 1 | 1 | 2,18% | VVAQLGER |
| ADO50532 | CE9 | Nacetylglucosamine6phosphate deacetylase [Enterobacter lignolyticus SCF1]. | 94,40% | 1 | 1 | 1 | 2,54% | TSSCAPAPR |
| AFT81264 | CE9 | N-acetyl-glucosamine-6-phosphate deacetylase [Leuconostoc carnosum JB16]. | 94,90% | 1 | 1 | 1 | 3,94% | AEGLNGISELGGQK |
| BAJ74392 | CE9 | N-acetyl-glucosamine-6-phosphate deacetylase [Microbacterium testaceum StLB037]. | 91,90% | 1 | 1 | 1 | 2,93% | LGGLDVEVR |
| ADY00258 | GH1 | glycoside hydrolase family 1 [Vulcanisaeta moutnovskia 76828]. | 94,90% | 1 | 1 | 1 | 1,43% | LDEAANK |
| AEK40976 | GH1 | beta-glucosidase [Amycolatopsis mediterranei S699]. | 99,90% | 2 | 2 | 2 | 4,65% | TAAAAGVDLR |
| | | | | | | | | TVADIGVDTSR |
| AGB78037 | GH1 | beta-galactosidase [Enterobacteriaceae bacterium strain FGI 57]. | 99,30% | 1 | 1 | 1 | 1,50% | FSISWPR |
| CBK85575 | GH1 | Beta-glucosidase [Enterobacter cloacae subsp cloacae NCTC 9394]. | 92,40% | 1 | 1 | 1 | 1,59% | VVEAFIR |
| EEV40795 | GH1 | hypothetical protein ECBG_03064 [Enterococcus casseliflavus EC20]. | 96,30% | 1 | 1 | 1 | 1,02% | EWQEK |
| P07206 | GH1 | Pullulanase [Klebsiella pneumoniae]. | 94,70% | 1 | 1 | 1 | 0,83% | LCEVNSAVK |
| ABD25030 | GH102 | MitA [Novosphingobium aromaticivorans DSM 12444]. | 94,50% | 1 | 1 | 2 | 1,72% | ALLLLPK |
| ABZ74159 | GH102 | MitA domain protein [Caulobacter sp K31]. | 94,90% | 1 | 1 | 1 | 3,15% | LAVALDVGGAIK |

| Nº de Acesso | Familia CAZy | Nome da Proteína [organismo] | ID % | Nº UP | Nº US | Nº TS | Seq. coverage (%) | Peptide sequence |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ACK65429 | GH102 | MitA domain protein [Cyanothecae sp PCC 8801]. | 91,90% | 1 | 1 | 1 | 1,54% | QELEGK |
| ACS17183 | GH102 | 3D domain protein [Variovorax paradoxus S110]. | 92,20% | 1 | 1 | 1 | 1,50% | LANGDIIR |
| ADN59026 | GH102 | MitA domain protein [Burkholderia sp CCGE1003]. | 95,60% | 1 | 1 | 2 | 2,67% | EMPSAEAMPK |
| AFC59093 | GH102 | murein transglycosylase A [Vibrio cholerae IEC224]. | 93,80% | 1 | 1 | 1 | 2,17% | YPVYGLPK |
| BAC08521 | GH102 | tlr0969 [Thermosynechococcus elongatus BP1]. | 91,90% | 1 | 1 | 1 | 3,08% | FVLDDQDTGSAIR |
| ADG60468 | GH103 | lytic murein transglycosylase [Moraxella catarrhalis BBH18]. | 94,40% | 1 | 1 | 1 | 1,69% | VGVMPPDGR |
| BAM87205 | GH103 | membranebound lytic murein transglycosylase signal peptide [Bradyrhizobium oligotrophicum S58]. | 94,90% | 1 | 1 | 1 | 2,78% | IMDLLDAQPEFTK |
| CBV43390 | GH103 | lytic murein transglycosylase [Halomonas elongata DSM 2581]. | 91,90% | 1 | 1 | 1 | 2,54% | GEIDEATLAR |
| ABI87408 | GH104 | glycoside hydrolase family 24 [Burkholderia ambifaria AMMD]. | 93,10% | 1 | 1 | 1 | 4,67% | YAIQQLR |
| ADM99981 | GH105 | predicted glycosyl hydrolase [Dickeya dadantii 3937]. | 93,80% | 1 | 1 | 1 | 2,12% | QIDALLDK |
| ADV83420 | GH106 | glycoside hydrolase family 2 sugar binding protein [Terriglobus saanensis SP1PR4]. | 98,20% | 1 | 1 | 1 | 0,71% | ITASDPK |
| AGK58082 | GH108 | hypothetical protein HYPDE_32038 [Hyphomicrobium denitrificans 1NES1]. | 93,80% | 1 | 1 | 1 | 1,83% | DDAPEILR |
| ACT93079 | GH110 | conserved hypothetical protein [Dyadobacter fermentans DSM 18053]. | 93,80% | 1 | 1 | 1 | 0,98% | LENMPK |
| AAO88279 | GH111 | endo-beta-N-acetyl-glucosaminidase [Bacillus circulans]. | 99,60% | 1 | 1 | 1 | 0,41% | GDLIVQIR |
| ABB75727 | GH114 | conserved hypothetical protein [Nitrosospora multiformis ATCC 25196]. | 99,00% | 1 | 1 | 1 | 1,87% | LDLIPIMR |
| AGB28982 | GH115 | hypothetical protein Prede_1683 [Prevotella dentalis DSM 3688]. | 93,00% | 1 | 1 | 1 | 1,09% | SWNAYVEIR |
| AGB29053 | GH115 | hypothetical protein Prede_1756 [Prevotella dentalis DSM 3688]. | 93,30% | 1 | 1 | 1 | 1,53% | HIGYTSWNDNFAK |
| AGL17375 | GH115 | hypothetical protein L083_3865 [Actinoplanes sp N902109]. | 92,20% | 1 | 1 | 1 | 0,61% | LAGWER |
| EDY82965 | GH115 | hypothetical protein VDG1235_2588 [Verrucomicrobiae bacterium DG1235]. | 94,40% | 1 | 1 | 1 | 1,44% | TASAVTVEPSNAAR |
| AAW74330 | GH12 | cellulase S [Xanthomonas oryzae pv oryzae KACC 10331]. | 93,00% | 1 | 1 | 1 | 3,17% | SFPQTGAR |
| ADP65783 | GH12 | cellulase B precursor [Streptomyces sp THW31]. | 94,50% | 1 | 1 | 1 | 3,15% | ADGSVPTDGAPK |
| ACS89933 | GH122 | Alpha-glucosidase [Thermococcus sibiricus MM 739]. | 94,90% | 1 | 2 | 2 | 1,80% | QFLMVMQMK |
| AFL86699 | GH125 | hypothetical protein Terro_0350 [Terriglobus roseus DSM 18391]. | 97,80% | 1 | 1 | 1 | 1,62% | VIAGLIHR |
| AGE22473 | GH127 | AraNlike protein [Geobacillus sp GHH01]. | 94,00% | 1 | 1 | 2 | 2,18% | GDQVELVFPMPVER |
| ACV59873 | GH13 | alpha amylase catalytic region [Alicyclobacillus acidocaldarius subsp acidocaldarius DSM 446]. | 92,10% | 1 | 1 | 1 | 1,04% | QLAWLK |
| ADW67288 | GH13 | alpha amylase catalytic region [Granulicella tundricola MP5ACTX9]. | 94,90% | 1 | 1 | 1 | 1,32% | GATPELMR |
| ADX47038 | GH13 | Protein of unknown function DUF3416 [Acidovorax avenae subsp avenae ATCC 19860]. | 94,90% | 1 | 1 | 1 | 0,97% | VAGEAVR |
| AGL19678 | GH13 | TreY [Actinoplanes sp N902109]. | 92,50% | 1 | 1 | 1 | 1,19% | GGVTVVATR |
| CAJ38414 | GH13 | 1,4-alpha-glucan branching enzyme [Anaerobranca gottschalkii]. | 92,80% | 1 | 1 | 1 | 1,47% | ENSIYVYQR |
| CCC99271 | GH13 | glycosyl hydrolase glycogen debranching enzyme [Azospirillum brasilense Sp245]. | 91,90% | 1 | 1 | 1 | 1,25% | FDLASVLSR |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AEE53364 | GH130 | glycosidase related protein [Haliscobenobacter hydroxsis DSM 1100]. | 92,10% | 1 | 1 | 1 | 2,31% | QYHTVVEVK |
| AFV13072 | GH130 | glycosylase like protein [Thermacetogenium phaeum DSM 12270]. | 100,00% | 1 | 1 | 1 | 0,59% | TPDICEK |
| CAG61092 | GH132 | unnamed protein product [Candida glabrata]. | 93,60% | 1 | 1 | 1 | 2,11% | LSQGVAICR |
| AAK76508 | GH14 | putative beta-amylase [Arabidopsis thaliana]. | 94,60% | 1 | 1 | 1 | 1,65% | LISEAGLK |
| AEB00844 | GH14 | beta-amylase 8 [Hordeum vulgare subsp vulgare]. | 92,20% | 1 | 1 | 1 | 2,37% | LDDASLSQVVR |
| AFQ33617 | GH14 | beta-amylase 5 [Citrus trifoliata]. | 93,80% | 1 | 1 | 1 | 2,50% | ENIEVAGENALS |
| AAY38020 | GH15 | Glycoside hydrolase family 15 [Pseudomonas syringae pv syringae B728a]. | 92,20% | 1 | 1 | 1 | 2,98% | LCGLCNDVGLLAEQYDPK |
| ABX02195 | GH15 | glycoside hydrolase 15related [Methanococcus maripaludis C6]. | 94,90% | 1 | 1 | 1 | 3,22% | DGGYIAVALSLCGQSSVVEK |
| AEL24471 | GH15 | glycoside hydrolase 15related protein [Cyclobacterium marinum DSM 745]. | 96,50% | 1 | 1 | 1 | 1,01% | FVESEK |
| AEP13039 | GH15 | Glucosylase like glycosyl hydrolase [Candidatus Chloracidobacterium thermophilum B]. | 99,00% | 1 | 1 | 1 | 1,40% | LGIVPPDDPK |
| AFT85110 | GH15 | glycoside hydrolase family 15 [Burkholderia phenoliruptrix BR3459a]. | 98,60% | 1 | 1 | 1 | 1,92% | TVAEFTVGEGER |
| AGL85924 | GH15 | glycosyl hydrolase family protein [Pseudomonas protegens CHA0]. | 97,60% | 1 | 1 | 1 | 1,48% | VVSGQATLR |
| CAB42080 | GH15 | conserved hypothetical protein partial [Streptomyces coelicolor A32]. | 92,60% | 1 | 1 | 1 | 1,47% | IDDTVEAWR |
| ACF44156 | GH16 | glycoside hydrolase family 16 [Pelodictyon phaeocliathratiforme BU1]. | 94,90% | 1 | 1 | 1 | 1,24% | TVEEIIR |
| CCA37346 | GH17 | glucan 1,3-beta-glucosidase [Komagataella pastoris CBS 7435]. | 94,90% | 1 | 1 | 1 | 1,93% | MIFNLK |
| CCC98700 | GH17 | glycosyl transferase [Azospirillum brasilense Sp245]. | 94,90% | 1 | 1 | 6 | 1,47% | FLPGATELTVGQK |
| ABK89520 | GH18 | Chitinase glycosyl hydrolase family 18 [Francisella tularensis subsp novicida strain U112]. | 94,90% | 1 | 1 | 1 | 0,92% | SLINAVAK |
| ABU78538 | GH18 | hypothetical protein ESA_03317 [Cronobacter sakazakii ATCC BAA894]. | 98,30% | 1 | 1 | 1 | 1,29% | LFMSLDTPR |
| AAB64045 | GH19 | putative endochitinase [Arabidopsis thaliana]. | 93,40% | 1 | 1 | 1 | 4,98% | QAFIAAAESFAAYK |
| AFJ47021 | GH19 | hypothetical protein EBL_c19290 [Shimwellia blattae DSM 4481 = NBRC 105725]. | 94,90% | 1 | 1 | 1 | 0,91% | ITVSMLR |
| AGA89238 | GH19 | putative chitinase [Thioflavococcus mobilis 8321]. | 94,90% | 1 | 1 | 1 | 2,83% | ALNELER |
| AAO75568 | GH2 | beta-galactosidase [Bacteroides thetaiotaomicron VPI5482]. | 94,90% | 1 | 1 | 1 | 0,95% | VYEGNILK |
| ABR41450 | GH2 | glycoside hydrolase family 2 candidate beta-glycosidase [Bacteroides vulgatus ATCC 8482]. | 97,20% | 1 | 1 | 1 | 1,62% | VNLSNGTGELR |
| ACL75354 | GH2 | glycoside hydrolase family 2 sugar binding [Clostridium cellulolyticum H10]. | 98,90% | 1 | 1 | 1 | 0,85% | LLLNTASK |
| ACL76021 | GH2 | glycoside hydrolase family 2 TIM barrel [Clostridium cellulolyticum H10]. | 90,50% | 1 | 1 | 1 | 1,00% | ASVYYK |
| ACU75707 | GH2 | Betagalactosidase [Catenuispora acidiphila DSM 44928]. | 92,20% | 1 | 1 | 1 | 1,41% | TWLSLDGWR |
| ADB62765 | GH2 | glycoside hydrolase family 2 sugar binding protein [Haloterrigena turkmenica DSM 5511]. | 98,50% | 1 | 1 | 1 | 0,38% | SDSQQR |
| ADL34436 | GH2 | glycoside hydrolase family 2 [Butyrivibrio proteoclasticus B316]. | 94,90% | 1 | 1 | 1 | 0,77% | VLDNEAR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ADL53745 | GH2 | glycoside hydrolase family 2 TIM barrel [Clostridium cellulovorans 743B]. | 98,70% | 1 | 1 | 1 | 0,79% | ISGNIEGR |
| AEM74685 | GH2 | glycoside hydrolase family 2 sugar binding [Caldicellulosiruptor lactoaceticus 6A]. | 94,90% | 1 | 1 | 1 | 1,44% | VESISDELAIVK |
| AFV83763 | GH2 | betagalactosidase [Alteromonas macleodii AltDE1]. | 94,80% | 1 | 1 | 1 | 1,05% | EFEAHTLK |
| BAL88793 | GH2 | putative glycosyl hydrolase [Actinoplanes missouriensis 431]. | 98,20% | 1 | 1 | 1 | 1,79% | GPQEIVGVTDVTPDQPK |
| EDY95407 | GH2 | glycosyl hydrolase family 2 sugar binding domain protein [Bacteroides plebeius DSM 17135]. | 92,90% | 1 | 1 | 1 | 1,06% | WYVGAGIYR |
| EEV37821 | GH2 | hypothetical protein ECBG_00090 [Enterococcus casseliflavus EC20]. | 98,20% | 1 | 1 | 1 | 1,50% | SGENQLTLR |
| AAN23923 | GH20 | possible beta-hexosaminidase A [Bifidobacterium longum NCC2705]. | 92,90% | 1 | 1 | 1 | 1,27% | IQTIASQGR |
| ACD05945 | GH20 | Glycoside hydrolase family 20 catalytic core [Akkermansia muciniphila ATCC BAA835]. | 99,30% | 1 | 1 | 1 | 1,46% | ELVSSIGNAPR |
| ADG98113 | GH20 | Glycoside hydrolase family 20 catalytic core [Segniliparus rotundus DSM 44985]. | 96,10% | 1 | 1 | 1 | 1,34% | HSELGVK |
| ADV47647 | GH20 | Beta-N-acetylhexosaminidase [Cellulophaga algicola DSM 14237]. | 98,30% | 1 | 1 | 1 | 2,60% | LTSVSGNTEVGGGK |
| ADY30156 | GH20 | Beta-N-acetylhexosaminidase [Cellulophaga lytica DSM 7489]. | 97,80% | 1 | 1 | 1 | 0,92% | MTPEQEK |
| BAA92145 | GH20 | beta-N-acetylglucosaminidase [Aeromonas sp 10S24]. | 91,30% | 1 | 1 | 1 | 1,20% | AASGSGITLK |
| CAH07887 | GH20 | putative hydrolase exported protein [Bacteroides fragilis NCTC 9343]. | 94,90% | 1 | 1 | 1 | 2,48% | DIIDEVSALFPSR |
| CBK66445 | GH20 | N-acetyl-beta-hexosaminidase [Bacteroides xylanisolvens XB1A]. | 99,80% | 1 | 1 | 2 | 0,71% | IGKPMR |
| ABM34743 | GH23 | Lytic transglycosylase catalytic [Acidovorax citrulli AAC001]. | 100,00% | 1 | 1 | 1 | 1,06% | LAAEANR |
| ADP31640 | GH23 | Yjbj [Bacillus atrophaeus 1942]. | 92,20% | 1 | 1 | 1 | 4,24% | GDFSIDSAIK |
| ADP31755 | GH23 | hypothetical protein BATR1942_04005 [Bacillus atrophaeus 1942]. | 99,00% | 1 | 1 | 1 | 0,65% | MLLSQAASK |
| CBS90914 | GH23 | protein of unknown function [Azospirillum lipoferum 4B]. | 94,10% | 1 | 1 | 1 | 1,56% | QIQLER |
| ADF63948 | GH24 | phage lysozyme [Enterobacter cloacae subsp cloacae ATCC 13047]. | 91,10% | 1 | 1 | 1 | 4,14% | MNPSIVK |
| ADH87371 | GH24 | glycoside hydrolase family 24 [Starkeya novella DSM 506]. | 92,80% | 1 | 1 | 1 | 5,61% | QACDAQTAWNR |
| CAP41335 | GH24 | phage related lysozyme [Bordetella petrii]. | 93,10% | 1 | 1 | 1 | 4,11% | INELTR |
| ACL29153 | GH25 | 1,4-beta-N-acetylmuramidase [Bifidobacterium animalis subsp lactis AD011]. | 91,70% | 1 | 1 | 1 | 2,06% | NINWQK |
| ADL33290 | GH25 | lysozyme Lvc25E [Butyrivibrio proteoclasticus B316]. | 94,60% | 1 | 1 | 1 | 0,74% | GDGISAAQR |
| AEE72870 | GH25 | lysozyme M1 [Propionibacterium acnes 266]. | 92,90% | 1 | 1 | 1 | 3,42% | FATNETATPR |
| AGL64698 | GH25 | Prophage Lp1 protein 52 endolysin [Lactobacillus plantarum subsp plantarum P8]. | 94,90% | 1 | 1 | 1 | 1,61% | QTIELGVNLK |
| CCC15544 | GH25 | lysin [Lactobacillus pentosus IG1]. | 94,90% | 1 | 1 | 1 | 1,41% | SGNFYR |
| AAK78518 | GH26 | Beta-mannanase ManB contains ChWrepeats [Clostridium acetobutylicum ATCC 824]. | 97,70% | 1 | 1 | 1 | 2,57% | DGELAGTTGQSLR |
| AAT36643 | GH26 | mannase precursor [uncultured organism]. | 96,90% | 1 | 1 | 1 | 1,88% | QLWQLTR |
| ADO75030 | GH26 | Glycoside hydrolase family 26 [Stigmatella aurantiaca DW431]. | 92,80% | 1 | 1 | 1 | 1,73% | MLAENDK |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AFU69689 | GH26 | endo-1,3-beta-xylanase [Psychroflexus torquis ATCC 700755]. | 89,70% | 1 | 1 | 1 | 2,01% | FEPEDGK |
| EAA61784 | GH26 | hypothetical protein AN74132 [Aspergillus nidulans FGSC A4]. | 92,50% | 1 | 1 | 1 | 3,10% | DIDAIAEQLAR |
| CBK97479 | GH27 | Alpha-galactosidase [Eubacterium siraeum 703]. | 92,00% | 1 | 1 | 1 | 3,03% | QAADAFIENGLK |
| CBX98481 | GH27 | similar to alpha-galactosidase precursor [Leptosphaeria maculans JN3]. | 93,60% | 1 | 1 | 1 | 2,17% | VMSDALNATGR |
| ADG11181 | GH28 | glycoside hydrolase family protein [Caulobacter segnis ATCC 21756]. | 92,00% | 1 | 1 | 1 | 2,15% | LGTESNGGFR |
| AEU35935 | GH29 | glycoside hydrolase family 29 alpha-L-fucosidase [Granulicella mallensis MP5ACTX8]. | 94,90% | 1 | 1 | 1 | 1,25% | IQENILK |
| ABF41379 | GH3 | Beta-glucosidase [Candidatus Koribacter versatilis Ellin345]. | 93,50% | 1 | 1 | 2 | 1,55% | LSIGSTQPQETQAK |
| ABF42247 | GH3 | Beta-glucosidase [Candidatus Koribacter versatilis Ellin345]. | 98,20% | 1 | 1 | 1 | 1,14% | ESEIDTALVR |
| ABW02138 | GH3 | glycoside hydrolase family 3 protein [Caldivirga maquilgensis IC167]. | 97,90% | 1 | 1 | 1 | 1,41% | TTVDTIVDER |
| ACO32765 | GH3 | beta-glucosidase [Acidobacterium capsulatum ATCC 51196]. | 94,90% | 1 | 1 | 1 | 1,43% | NDGHLLPLNPAR |
| ADM27757 | GH3 | glycoside hydrolase family 3 domain protein [Ignisphaera aggregans DSM 17230]. | 93,20% | 1 | 1 | 1 | 1,06% | QIMAIGSR |
| ADN36290 | GH3 | glycoside hydrolase family 3 domain protein [Methanoplanus petrolearius DSM 11571]. | 92,50% | 1 | 1 | 1 | 2,51% | IAENATGLIK |
| ADQ68885 | GH3 | beta-glucosidase like glycosyl hydrolase [Halogeometricum borinquense DSM 11551]. | 94,10% | 1 | 1 | 1 | 1,08% | LAAETANR |
| ADW71213 | GH3 | glycoside hydrolase family 3 domain protein [Granulicella tundricola MP5ACTX9]. | 98,30% | 1 | 1 | 1 | 0,90% | LLLQDTLR |
| AEI12184 | GH3 | glycoside hydrolase family 3 domain protein [Cellvibrio gilvus ATCC 13127]. | 92,00% | 1 | 1 | 1 | 1,46% | EAAAAGSVLLR |
| Q8VZR2 | GH3 | Alpha-L-arabinofuranosidase 2 [Arabidopsis thaliana]. | 98,00% | 1 | 1 | 1 | 1,04% | DMETSWR |
| ACL28346 | GH30 | glycoside hydrolase family 30 protein [Bifidobacterium animalis subsp lactis AD011]. | 92,90% | 1 | 1 | 1 | 1,38% | TPPEMTR |
| ABG84624 | GH31 | glycosyl hydrolase family 31 [Clostridium perfringens ATCC 13124]. | 94,90% | 1 | 1 | 1 | 0,84% | IMQTIK |
| ABP51590 | GH31 | Alpha-glucosidase [Pyrobaculum arsenaticum DSM 13514]. | 94,70% | 1 | 1 | 1 | 1,02% | QFTWDLK |
| ABU79235 | GH31 | hypothetical protein ESA_04054 [Cronobacter sakazakii ATCC BAA894]. | 92,30% | 1 | 1 | 1 | 1,63% | IALTQAAVSER |
| AEE94503 | GH31 | alpha-glucosidase [Acidianus hospitalis W1]. | 94,50% | 1 | 1 | 1 | 1,62% | EVLGSTPIELR |
| AFS70328 | GH31 | Glycosyl hydrolases family 31 [Exiguobacterium antarcticum B7]. | 98,90% | 1 | 1 | 1 | 1,14% | LEDQGVLAk |
| BAD34979 | GH31 | 3-alpha-isomaltosyltransferase [Arthrobacter globiformis]. | 93,10% | 1 | 1 | 1 | 0,98% | DPMAGDNVTVR |
| WP_009776446 | GH31 | glycosylhydrolase [Janibacter sp HTCC2649]. | 94,20% | 1 | 1 | 1 | 1,10% | TPSLATGSAASVR |
| AAK00768 | GH32 | exo-inulinase [Bacillus subtilis]. | 98,70% | 1 | 1 | 1 | 2,34% | EIELATVDGFPR |
| AAL19864 | GH33 | sialidase [Salmonella enterica subsp enterica serovar Typhimurium str LT2]. | 94,90% | 1 | 1 | 1 | 1,70% | IAIYNDR |
| AEL26174 | GH33 | glycosyl hydrolase BNR repeatcontaining protein [Cyclobacterium marinum DSM 745]. | 94,90% | 1 | 1 | 1 | 2,43% | LLAPVSLHK |
| AET58792 | GH33 | hypothetical protein HPL003_10155 [Paenibacillus terrae HPL003]. | 93,80% | 1 | 1 | 1 | 2,03% | QAFWGAPR |
| AGM16014 | GH34 | neuraminidase [Influenza A virus AswinelowaA014326642013H1N1]. | 92,00% | 1 | 1 | 1 | 1,49% | GDIFVIR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ADY43870 | GH35 | Beta-galactosidase [<i>Ascaris suum</i>]. | 94,40% | 1 | 1 | 1 | 1,56% | QTYETINDFK |
| ABJ76756 | GH36 | Alpha-galactosidase [<i>Leptospira borgpetersenii</i> serovar Hardjobovis str JB197]. | 94,90% | 1 | 1 | 1 | 1,24% | DFLENILK |
| ACY14533 | GH36 | Alpha-galactosidaselike protein [<i>Haliangium ochraceum</i> DSM 14365]. | 98,00% | 1 | 1 | 1 | 1,28% | FPGGLAQVAR |
| CBL09633 | GH36 | Alpha-galactosidase [<i>Roseburia intestinalis</i> M501]. | 91,40% | 1 | 1 | 1 | 1,37% | QDSIAHNLVAK |
| AEN65223 | GH37 | glycoside hydrolase family 37 [<i>Enterobacter asburiae</i> LF7a]. | 92,50% | 1 | 1 | 1 | 1,60% | LDDGAILNR |
| ABG87610 | GH38 | glycosyl hydrolase family 38 [<i>Clostridium perfringens</i> SM101]. | 98,40% | 1 | 1 | 1 | 0,69% | LGENLVK |
| ABL90195 | GH38 | Glycoside hydrolase family 38 [<i>Mycobacterium</i> sp strain KMS]. | 99,80% | 1 | 1 | 1 | 0,51% | TVALLNR |
| ABV09382 | GH38 | glycosyl hydrolase family 38 [<i>Streptococcus gordonii</i> str Challis substr CH1]. | 94,40% | 1 | 1 | 1 | 1,14% | SGVLGPQEIR |
| ACU70268 | GH38 | glycoside hydrolase family 38 [<i>Catenulispora acidiphila</i> DSM 44928]. | 94,80% | 1 | 1 | 1 | 0,55% | ELVVALAR |
| ADY40773 | GH38 | Alpha-mannosidase 2 [<i>Ascaris suum</i>]. | 98,50% | 1 | 1 | 1 | 1,11% | AADLATNMIQK |
| CAC97353 | GH38 | lin2123 [<i>Listeria innocua</i> Clip11262]. | 93,50% | 1 | 1 | 1 | 1,07% | NYGVSLNDSK |
| CAR25074 | GH38 | KLTHOG11000p [<i>Lachancea thermotolerans</i> CBS 6340]. | 92,50% | 1 | 1 | 1 | 1,20% | GALSSSTVHEALK |
| ABJ84618 | GH39 | glycoside hydrolase family 39 [<i>Candidatus Solibacter usitatus</i> Ellin6076]. | 99,30% | 1 | 1 | 1 | 1,99% | VESSGGLGLDK |
| ABS68786 | GH39 | conserved hypothetical protein [<i>Xanthobacter autotrophicus</i> Py2]. | 93,50% | 1 | 1 | 1 | 2,28% | LTGEGEASDPR |
| ACT03721 | GH39 | transcriptional regulator AraC family [<i>Paenibacillus</i> sp JDR2]. | 94,90% | 1 | 1 | 1 | 0,72% | EEVFIK |
| ADQ80254 | GH39 | Xylan 1,4-beta-xylosidase [<i>Paludibacter propionigenes</i> WB4]. | 94,90% | 1 | 1 | 1 | 1,32% | LQVEQIK |
| AFR72105 | GH39 | putative regulatory protein [<i>Staphylococcus aureus</i> 08BA02176]. | 98,60% | 1 | 1 | 1 | 1,21% | TASQMIINK |
| AGB42757 | GH39 | beta-xylosidase [<i>Mesorhizobium australicum</i> WSM2073]. | 94,90% | 1 | 1 | 1 | 1,62% | MEGVETAK |
| ABL78908 | GH4 | Alphagalactosidase [<i>Thermofilum pendens</i> Hrk 5]. | 93,70% | 1 | 1 | 1 | 2,61% | EIAQTIGLDPAR |
| AEX05011 | GH4 | alpha-galactosidase [<i>Klebsiella oxytoca</i> KCTC 1686]. | 98,10% | 1 | 1 | 1 | 2,99% | VTFMGAGSTIFAR |
| Q1NHS6 | GH4 | Putative secreted alphagalactosidase [<i>Sphingomonas</i> sp SKA58]. | 98,80% | 1 | 1 | 1 | 1,27% | LLASSETAR |
| ABP66632 | GH42 | Betagalactosidase [<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903]. | 93,50% | 1 | 1 | 1 | 1,48% | LDEILESVNK |
| ABR38219 | GH43 | glycoside hydrolase family 43 [<i>Bacteroides vulgatus</i> ATCC 8482]. | 99,40% | 1 | 1 | 1 | 1,96% | VMPQYEGSAMAPDLLK |
| ADL52247 | GH43 | LPXTGmotif cell wall anchor domain protein [<i>Clostridium cellulovorans</i> 743B]. | 99,40% | 1 | 1 | 1 | 0,65% | LPYEVIDK |
| AEK42231 | GH43 | arabinan endo-1,5-alpha-L-arabinosidase [<i>Amycolatopsis mediterranei</i> S699]. | 97,30% | 1 | 1 | 1 | 3,44% | QGSAMAAIAGR |
| AFU70802 | GH43 | putative glycoside hydrolase [<i>Bifidobacterium asteroides</i> PRL2011]. | 98,80% | 1 | 1 | 1 | 1,34% | NTVSGSSFGPAR |
| ACX75523 | GH45 | Cellulase [<i>Fibrobacter succinogenes</i> subsp succinogenes S85]. | 92,50% | 1 | 1 | 1 | 2,11% | FNTTTQTNIK |
| AEJ00888 | GH46 | peptido glycan binding domain containing protein [<i>Nitrosomonas</i> sp Is79A3]. | 91,40% | 1 | 1 | 1 | 3,08% | LHDLVEK |
| CAJ88423 | GH46 | putative secreted chitosanase [<i>Streptomyces ambofaciens</i> ATCC 23877]. | 93,50% | 1 | 1 | 1 | 2,86% | QAQADGLR |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| CAG86907 | GH47 | unnamed protein product [Debaryomyces hansenii CBS767]. | 98,00% | 1 | 1 | 1 | 0,71% | NITTPR |
| BAL86217 | GH5 | putative glycosyl hydrolase [Actinoplanes missouriensis 431]. | 94,90% | 1 | 1 | 1 | 1,22% | ELLAGIK |
| ABD80438 | GH50 | bagarase [Saccharophagus degradans 240]. | 93,10% | 1 | 1 | 1 | 1,93% | QPGDLTPEDSLGLNR |
| ABG41153 | GH50 | Agarase [Pseudoalteromonas atlantica T6c]. | 98,40% | 1 | 1 | 1 | 1,35% | MLEMLGEQYFK |
| ACT04224 | GH51 | alpha-L-arabinofuranosidase domain protein [Paenibacillus sp JDR2]. | 99,40% | 1 | 1 | 1 | 1,04% | EVVDMTR |
| ADQ80909 | GH51 | alpha-L-arabinofuranosidase domain protein [Paludibacter propionigenes WB4]. | 89,90% | 1 | 1 | 1 | 1,76% | TLEIDDMIK |
| ADV65920 | GH51 | alpha-L-arabinofuranosidase domain protein [Deinococcus maricopensis DSM 21211]. | 99,40% | 1 | 1 | 1 | 1,93% | DLHVNHTHAK |
| ADW70746 | GH51 | alpha-L-arabinofuranosidase domain protein [Granulicella tundricola MP5ACTX9]. | 98,50% | 1 | 1 | 1 | 3,53% | FAAGDALEFPPTYDTLAR |
| AFV91139 | GH51 | Alpha-L-arabinofuranosidase [Propionibacterium acidipropionici ATCC 4875]. | 92,10% | 1 | 1 | 1 | 1,61% | ELGATAVR |
| CBX25410 | GH51 | hypothetical_protein [Oryza glaberrima]. | 97,90% | 1 | 1 | 1 | 1,76% | GSADSTWGSVR |
| Q5WL16 | GH51 | Alpha-L-arabinofuranosidase [Bacillus clausii strain KSMK16]. | 94,70% | 1 | 1 | 1 | 1,19% | TLEMDR |
| AFK86458 | GH52 | glycoside hydrolase family 52 [Thermoanaerobacterium saccharolyticum JWSLYS485]. | 92,30% | 1 | 1 | 1 | 1,32% | IFGEDEIER |
| AEA20827 | GH53 | glycosyl hydrolase family 53 [Prevotella denticola F0289]. | 98,60% | 1 | 1 | 1 | 3,44% | NGVPTELMALLK |
| ADO44667 | GH57 | glycoside hydrolase family 57 [Hydrogenobacter thermophilus TK6]. | 99,20% | 1 | 1 | 3 | 0,89% | LLILIK |
| ADP97474 | GH57 | glycoside hydrolase family 57 [Marinobacter adhaerens HP15]. | 94,90% | 1 | 1 | 1 | 3,13% | DAVGDLVHHMENIAGQAK |
| ADU96944 | GH57 | glycoside hydrolase family 57 [Thermovibrio ammonificans HB1]. | 94,90% | 1 | 1 | 4 | 1,61% | ELSDLIGFVYK |
| AAS02597 | GH6 | CelA [Mycobacterium avium subsp paratuberculosis K10]. | 94,90% | 1 | 1 | 1 | 3,34% | MTFSAAGAVAR |
| ACY20379 | GH6 | Cellulase [Gordonia bronchialis DSM 43247]. | 96,70% | 1 | 1 | 1 | 1,94% | YVVDTSR |
| CAL66155 | GH63 | trehalase like protein [Gramella forsetii KT0803]. | 99,30% | 1 | 1 | 1 | 1,41% | SYFLVK |
| EAA27909 | GH64 | hypothetical protein [Neurospora crassa]. | 98,90% | 1 | 1 | 1 | 1,34% | LAAALNR |
| Q2QLS3 | GH64 | Alpha-amylase trypsin inhibitor putative expressed [Oryza sativa subsp japonica]. | 99,80% | 1 | 1 | 1 | 4,26% | GPSCAVDITAR |
| ABG83429 | GH65 | glycosyl hydrolase family 65 [Clostridium perfringens ATCC 13124]. | 98,10% | 1 | 1 | 1 | 0,77% | DEIEIR |
| ACD90563 | GH65 | beta-phosphoglucomutase family hydrolase [Chlorobium limicola DSM 245]. | 92,10% | 1 | 1 | 1 | 0,86% | ETICGLGNR |
| ADL03489 | GH65 | beta-phospho-glucomutase [Clostridium saccharolyticum WM1]. | 94,40% | 1 | 1 | 1 | 0,81% | ENLEVSER |
| ADO77388 | GH65 | glycoside hydrolase family 65 central catalytic [Halanaerobium praevalens DSM 2228]. | 98,60% | 1 | 1 | 1 | 0,90% | EIIWEDR |
| EEV38923 | GH65 | hypothetical protein ECBG_01192 [Enterococcus casseliflavus EC20]. | 93,40% | 1 | 1 | 1 | 0,91% | MNTILSK |
| ADE82923 | GH67 | putative alpha-glucuronidase [Prevotella ruminicola 23]. | 96,40% | 1 | 1 | 1 | 2,03% | VTGPECIAAQELR |
| CBK52811 | GH7 | cellobiohydrolase I partial [uncultured fungus]. | 91,90% | 1 | 1 | 1 | 8,14% | FVTQNSSGSNVGSR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|--|
| CCA94937 | GH7 | putative glycoside hydrolase family 7 [uncultured eukaryote]. | 94,00% | 1 | 1 | 1 | 2,16% | EFAVTIDSSR |
| AGC00414 | GH70 | glucosyl transferaseS [Streptococcus mutans]. | 94,40% | 1 | 1 | 1 | 1,23% | QAEAAQTASSQALTQAK |
| BAA14241 | GH70 | glucosyltransferase precursor [Streptococcus sobrinus]. | 97,80% | 1 | 1 | 1 | 0,75% | DGAAVTGAQTIR |
| AEV75389 | GH71 | Glycosyl hydrolase family 71 [Mycobacterium rhodesiae NBB3]. | 93,70% | 1 | 1 | 1 | 2,70% | IMLMPDMHGELK |
| ACQ69424 | GH73 | Lysozyme subfamily 2 [Exiguobacterium sp AT1b]. | 94,60% | 1 | 1 | 1 | 1,74% | LQLLER |
| ADE39837 | GH73 | Bax protein hydrolyses peptidoglycan [Candidatus Puniceispirillum marinum IMCC1322]. | 90,30% | 1 | 1 | 1 | 1,92% | KPASPAPK |
| AFZ03402 | GH73 | cell wall hydrolaseautolysin [Calothrix sp PCC 6303]. | 93,60% | 1 | 1 | 2 | 2,01% | AYASTEALK |
| AGK55374 | GH73 | Peptidase M23M37 family protein [Bacillus sp 1NLA3E]. | 93,60% | 1 | 1 | 1 | 4,50% | MNIGADWLAR |
| CCO12350 | GH73 | mannosyl glyco endo-beta-N-acetyl-glucosaminidase family protein [Carnobacterium maltaromaticum LMA28]. | 99,60% | 1 | 1 | 1 | 0,76% | LMGQAAINLER |
| EAA49293 | GH76 | hypothetical protein MG009514 [Magnaporthe grisea 7015]. | 94,90% | 1 | 1 | 1 | 1,24% | MVNLVVK |
| ABS27268 | GH77 | conserved hypothetical protein [Anaeromyxobacter sp Fw1095]. | 92,20% | 1 | 1 | 1 | 1,27% | ALAEAAGR |
| ADE14271 | GH77 | 4-alpha-glucano transferase [Nitrosococcus halophilus Nc 4]. | 92,50% | 1 | 1 | 1 | 1,24% | EEQVLK |
| ADH85763 | GH77 | 4-alpha-glucano transferase [Desulfurivibrio alkaliphilus AHT2]. | 98,90% | 1 | 1 | 1 | 1,73% | EPAAAEQR |
| AEW75631 | GH77 | 4-alpha-glucano transferase [Enterobacter cloacae EcWSU1]. | 99,10% | 1 | 1 | 1 | 1,15% | MLADVGAR |
| AFM00967 | GH77 | 4-alpha-glucano transferase [Desulfitobacterium dehalogenans ATCC 51507]. | 94,90% | 1 | 1 | 1 | 0,86% | FIQDNQEWLK |
| AGL02728 | GH77 | 4-alpha-glucano transferase [Desulfotomaculum gibsoniae DSM 7213]. | 92,30% | 1 | 1 | 1 | 1,02% | QLPLPGK |
| ABG60675 | GH8 | CHU large protein bglycosidase related glycoside hydrolase family 8 protein [Cytophaga hutchinsonii ATCC 33406]. | 99,30% | 1 | 1 | 1 | 1,67% | VVVTTSIGCNAGPGISNEVA LNIVSSMPSVSITGAPAVQ CSGLPVR |
| ACR32438 | GH8 | Cellulase [Burkholderia glumae BGR1]. | 94,90% | 1 | 1 | 1 | 1,46% | ILGWIR |
| BAA32084 | GH80 | chitosanase [Mitsuaria chitosanitabida]. | 97,30% | 1 | 1 | 1 | 9,21% | SNATNADMLAWGVQVTG VTAGNFQADDTALDPYPSK |
| AAV03316 | GH81 | Sequence 15273 from patent US 6747137 [Unknown]. | 93,60% | 1 | 1 | 1 | 0,91% | LNQALFDPK |
| ABX31208 | GH84 | Hyalurononglucosaminidase [Petrogoga mobilis SJ95]. | 94,90% | 1 | 1 | 1 | 1,60% | NFEESFIK |
| ADU14370 | GH84 | Beta-N-acetyl hexosaminidase [Asticcacaulis excentricus CB 48]. | 94,90% | 1 | 1 | 1 | 1,98% | FAAAQGLAAEASK |
| CAJ06605 | GH85 | glycosyl hydrolaselike protein [Leishmania major strain Friedlin]. | 94,50% | 1 | 1 | 2 | 1,26% | AGTPTMAASAAAQAASPR |
| CBZ30033 | GH85 | glycosyl hydrolaselike protein [Leishmania mexicana MHOMGT2001U1103]. | 91,90% | 1 | 1 | 1 | 0,70% | ESGGPLFSIR |
| ACT01236 | GH88 | glycosyl hydrolase family 88 [Paenibacillus sp JDR2]. | 99,50% | 2 | 2 | 2 | 5,32% | DDPINGGR ENMNILQAWGPK |
| AEB75453 | GH88 | putative glucuronyl hydrolase [Clostridium botulinum BKT015925]. | 99,20% | 1 | 1 | 1 | 2,23% | QEVEVAIQK |
| AAB91971 | GH9 | putative glucanase [Arabidopsis thaliana]. | 94,90% | 1 | 1 | 1 | 1,52% | GASIASFK |
| ABN51814 | GH9 | glycoside hydrolase family 9 [Clostridium thermocellum ATCC 27405]. | 94,90% | 1 | 1 | 3 | 1,49% | QADYALGSSGR |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ACT60135 | GH9 | glycoside hydrolase family 9 [Hirschia baltica ATCC 49814]. | 93,00% | 1 | 1 | 1 | 1,23% | TASEAILK |
| BAA96207 | GH9 | unnamed protein product [Oryza sativa japonica cultivargroup]. | 92,40% | 1 | 1 | 1 | 1,27% | GASIVSIR |
| BAL82172 | GH9 | putative endoglucanase [Selenomonas ruminantium subsp lactilytica TAM6421]. | 99,50% | 1 | 1 | 1 | 1,01% | YLEENK |
| BAM03196 | GH9 | putative glycoside hydrolase [Phycisphaera mikurensis NBRC 102666]. | 98,10% | 1 | 1 | 1 | 0,78% | LQTSEGGVR |
| ABQ05385 | GH92 | Glycoside hydrolase family 92 [Flavobacterium johnsoniae UW101]. | 92,40% | 1 | 1 | 1 | 1,03% | QLDKPFLK |
| ACQ81248 | GH92 | alpha-1,2-mannosidase [Beutenbergia cavernae DSM 12333]. | 92,00% | 1 | 1 | 1 | 0,86% | ENVYVQSLR |
| ADW69225 | GH92 | alpha-1,2-mannosidase [Granulicella tundricola MP5ACTX9]. | 98,20% | 1 | 1 | 1 | 1,30% | VPDFVNTLIR |
| AEV83179 | GH92 | alpha-1,2-mannosidase [Actinoplanes sp SE50110]. | 92,00% | 1 | 1 | 1 | 1,13% | IELSGGDR |
| AFK83607 | GH92 | putative mannosidase Man3 [Cellulosimicrobium cellulans]. | 99,00% | 1 | 1 | 1 | 1,30% | GATSAAPTANGYVQR |
| BAM23220 | GH92 | antigenic cell wall surface anchor protein family homolog [Streptococcus intermedius JTH08]. | 92,10% | 1 | 1 | 1 | 1,59% | FDIVEQGEADR |
| CBW21308 | GH92 | putative alpha-1,2-mannosidase precursor [Bacteroides fragilis 638R]. | 90,40% | 1 | 1 | 1 | 1,19% | ENLYIESLR |
| CBX95121 | GH92 | hypothetical protein LEMA_P1153601 [Leptosphaeria maculans JN3]. | 94,40% | 1 | 1 | 1 | 0,76% | MLGENQR |
| ABA81222 | GH94 | Glycosyl transferase [Rhodobacter sphaeroides 241]. | 99,40% | 1 | 1 | 1 | 0,33% | DLTLGIGER |
| ACV62532 | GH94 | glycosyl transferase 36 [Desulfotomaculum acetoxidans DSM 771]. | 94,80% | 1 | 1 | 2 | 0,40% | VVEGYVQLQPR |
| ADL50417 | GH94 | glycosyl transferase 36 [Clostridium cellulovorans 743B]. | 92,10% | 1 | 1 | 1 | 0,28% | QIEYSSTR |
| ADO77439 | GH94 | chitobiose phosphorylase [Halanaerobium praevalens DSM 2228]. | 93,00% | 1 | 1 | 1 | 1,25% | EFNLAAGEEK |
| AGA64013 | GH94 | LOW QUALITY PROTEIN: Cyclic beta12glucan synthase [Liberibacter crescens BT1]. | 97,40% | 1 | 1 | 1 | 0,28% | NLLSIGYK |
| BAH05588 | GH94 | hypothetical protein [Clostridium kluyveri NBRC 12016]. | 98,20% | 1 | 1 | 1 | 0,21% | IIFEIK |
| BAH99468 | GH94 | cyclic beta 12 glucan synthetase [Acetobacter pasteurianus IFO 328301]. | 92,30% | 1 | 1 | 1 | 1,21% | VEIVFLLGEAENEAEAR |
| BAI76389 | GH94 | glycosyltransferase [Azospirillum sp B510]. | 99,80% | 1 | 1 | 1 | 0,67% | IDSIAQSWAVLSGAADPER |
| BAK56901 | GH94 | hypothetical protein SFBM_1138 [Candidatus Arthromitus sp SFBmouseJapan]. | 99,20% | 1 | 1 | 1 | 0,81% | SGVPYLIDK |
| CCE24192 | GH94 | putative phosphorylase carbohydrate binding [Methylomicrobium alcaliphilum 20Z]. | 99,20% | 1 | 1 | 1 | 0,28% | IPHGADTK |
| CCF21339 | GH94 | Protein ndvB [Rhizobium sp]. | 99,80% | 1 | 1 | 1 | 0,35% | VAVEGFLGIR |
| Q126H3 | GH94 | Cyclic beta 1,2 glucan synthetase [Polaromonas sp strain JS666 ATCC BAA500]. | 97,90% | 1 | 1 | 1 | 0,64% | QLAEGSASAFVPASPD MR |
| Q97LF2 | GH94 | Cyclic beta 1,2 glucan synthetase [Clostridium acetobutylicum strain ATCC 824 DSM 792 JCM 1419 LMG 5710 VKM B1787]. | 92,60% | 1 | 1 | 1 | 0,42% | DEGIMTSHFIER |
| ACE83895 | GH95 | alpha-L-fucosidase putative afc95A [Cellvibrio japonicus Ueda107]. | 90,80% | 1 | 1 | 1 | 1,02% | DGSISPEK |

| Nº de Acesso | Familia CAZy | Nome da Proteína [organismo] | ID % | Nº UP | Nº US | Nº TS | Seq. coverage (%) | Peptide sequence |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AEV97756 | GH95 | trehalosemaltose hydrolase [Niaestella koreensis GR2010]. | 92,40% | 1 | 1 | 1 | 1,04% | TITVVPGR |
| AFC30926 | GH95 | alpha-L-fucosidase [Paenibacillus mucilaginosus 3016]. | 94,90% | 1 | 1 | 1 | 2,16% | QLEAEEAAQMGLMSIPK |
| AGA57684 | GH95 | hypothetical protein Theco_1545 [Thermobacillus composti KWC4]. | 99,10% | 1 | 1 | 1 | 1,07% | IFDGVADVR |
| ACU92184 | GH97 | alpha-glucosidase [Capnocytophaga ochracea DSM 7271]. | 97,30% | 1 | 1 | 1 | 1,51% | YVNPNNNSQIR |
| AEV99441 | GH97 | Glycoside hydrolase 97 [Niaestella koreensis GR2010]. | 99,70% | 1 | 1 | 2 | 1,59% | ATSINNNQSR |
| CAP50362 | GH97 | exported putative alphasglucosidase [Xanthomonas campestris pv campestris]. | 90,80% | 1 | 1 | 1 | 0,92% | LEIELR |
| AAN01207 | GT1 | putative glycosyl transferase [Lechevalieria aerocolonigenes]. | 92,20% | 1 | 1 | 1 | 1,66% | AVLDEPR |
| ABF42785 | GT1 | macrolide glycosyl transferase like protein [Candidatus Koribacter versatilis Ellin345]. | 93,50% | 1 | 1 | 2 | 3,02% | LLDGLLER |
| ABK58687 | GT1 | PdmQ [Actinomadura hibisca]. | 92,40% | 1 | 1 | 2 | 1,84% | LGGGEGQR |
| ABR55288 | GT1 | Glycosyltransferase 28 domain [Methanococcus vannielii SB]. | 96,10% | 1 | 1 | 1 | 1,66% | ELDIIR |
| BAL45953 | GT1 | putative macrolide glycosyltransferase [Bacillus licheniformis]. | 97,00% | 1 | 1 | 1 | 2,22% | QSLGELSER |
| CAL55763 | GT10 | alpha-3-fucosyltransferase ISS [Ostreococcus tauri]. | 93,60% | 1 | 1 | 1 | 1,60% | VAGVMNR |
| AAC26306 | GT14 | Hypothetical protein H41C033 [Caenorhabditis elegans]. | 92,50% | 1 | 1 | 1 | 1,19% | NEIDQR |
| AAK94757 | GT14 | GLY1 [Caenorhabditis elegans]. | 92,70% | 1 | 1 | 1 | 3,60% | ALNGSMNTDVSTFEVDR |
| CAG82612 | GT15 | YALI0B01672p [Yarrowia lipolytica CLIB122]. | 92,50% | 1 | 1 | 1 | 1,63% | EAVLAMR |
| BAD36053 | GT16 | putative beta-1,2-N-acetylglucosaminyl transferase II [Oryza sativa Japonica Group]. | 92,30% | 1 | 1 | 1 | 2,51% | QSALSLSLSQR |
| ACE06134 | GT19 | hypothetical protein Aasi_0753 [Candidatus Amoebophilus asiaticus 5a2]. | 93,70% | 1 | 1 | 1 | 2,43% | QLASYESIR |
| BAH38633 | GT19 | lipid A disaccharide synthase [Gemmatimonas aurantiaca T27]. | 93,10% | 1 | 1 | 1 | 3,47% | AASILPFEEALLR |
| CAX49184 | GT19 | lipid A disaccharide synthase [Neisseria meningitidis 8013]. | 94,60% | 1 | 1 | 1 | 1,82% | LAEILQR |
| ABN07461 | GT2 | glycosyl transferase family 2 [Methanocorpusculum labreanum Z]. | 94,90% | 1 | 1 | 1 | 2,59% | EQILELEK |
| AFO56973 | GT2 | Glycosyl transferase family 2 [Natriema sp J72]. | 92,40% | 1 | 1 | 1 | 2,43% | QFQAEQLPK |
| BAI61024 | GT2 | hypothetical protein [Methanocella paludicola SANAE]. | 96,50% | 1 | 1 | 1 | 0,72% | SLNIEMDK |
| CAJ35687 | GT2 | glycosyl transferase family 2 [Methanocella arvoryzae MRE50]. | 94,40% | 1 | 1 | 1 | 1,63% | YWIJK |
| ACV81189 | GT20 | trehalosephosphatase [Nakamurella multipartita DSM 44233]. | 98,70% | 1 | 1 | 1 | 1,61% | LELQQITAEYPK |
| ADO41265 | GT20 | alpha-alpha-trehalose-phosphate synthase [Ketogulonicigenium vulgare Y25]. | 95,70% | 1 | 1 | 1 | 1,72% | EVAELAGR |
| AEA59818 | GT21 | Glycosyl transferase group 2 family protein [Burkholderia gladioli BSR3]. | 94,90% | 1 | 1 | 2 | 2,56% | VSNLNIAGR |
| AAF48079 | GT23 | alpha-1,6-fucosyltransferase [Drosophila melanogaster]. | 94,00% | 1 | 1 | 1 | 1,13% | MASVSTR |
| CCC47222 | GT24 | putative UDPglucose:glycoprotein glucosyltransferase [Trypanosoma vivax Y486]. | 91,30% | 1 | 1 | 1 | 0,36% | EQFTEK |
| CCD25201 | GT24 | hypothetical protein NDAI_0E03840 [Naumovozyma dairenensis CBS 421]. | 99,00% | 1 | 1 | 1 | 0,58% | IIHNGSMK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------------|
| XP_362816 | GT25 | hypothetical protein MGG_08232 [Magnaporthe oryzae 7015]. | 92,00% | 1 | 1 | 1 | 3,05% | FDAMAIQAHLK |
| AFI04665 | GT25 | lipopolysaccharide biosynthesis protein [Helicobacter cetorum MIT 007128]. | 93,80% | 1 | 2 | 4 | 2,29% | NDMFLK |
| AAH74234 | GT27 | MGC83963 protein [Xenopus laevis]. | 97,30% | 1 | 1 | 1 | 1,26% | HLEVAVK |
| AGA25615 | GT27 | RHS repeatassociated core domain protein [Singulisphaera acidiphila DSM 18658]. | 99,00% | 1 | 1 | 1 | 0,44% | QNTVAETLGR |
| ABJ85296 | GT28 | Mono-galactosyl di-acylglycerol synthase [Candidatus Solibacter usitatus Elin6076]. | 92,90% | 1 | 1 | 1 | 3,19% | EMGAGMVLPDFR |
| ABQ29292 | GT28 | glycosyl transferase [Acidiphilium cryptum JF5]. | 93,60% | 1 | 1 | 1 | 3,26% | LAAMAEAAASAR |
| ACA54603 | GT28 | Glycosyl Transferase [Clostridium botulinum A3 str Loch Maree]. | 93,90% | 1 | 1 | 1 | 1,98% | VINEIVR |
| ACU54176 | GT28 | glycosyl transferase [Acidimicrobium ferrooxidans DSM 10331]. | 89,90% | 1 | 1 | 1 | 2,29% | LAEAGIVR |
| ABO37165 | GT29 | ST8 alpha-N-acetylneuraminide alpha28sialyltransferase 11 [Strongylocentrotus purpuratus]. | 92,40% | 1 | 1 | 1 | 3,85% | MELLALETLSK |
| CAG29195 | GT29 | alpha-2,3-sialyltransferase [Takifugu rubripes]. | 97,40% | 1 | 1 | 1 | 2,77% | EETAFLQK |
| CCO20362 | GT29 | glycosyl transferase family 29 protein sialyltransferase family protein [Bathycoccus prasinos]. | 93,50% | 1 | 1 | 1 | 1,16% | KPSCALVGNSR |
| AAM07034 | GT3 | predicted protein [Methanosarcina acetivorans C2A]. | 96,60% | 1 | 1 | 1 | 2,85% | ITDMNGLAPLSTDGELK |
| AFV22988 | GT3 | hypothetical protein Mpsy_0779 [Methanobolus psychrophilus R15]. | 93,60% | 1 | 1 | 1 | 1,16% | LLANAGR |
| ACT94413 | GT30 | ThreedeoxyDmannooctulosonicacid transferase domain protein [Dyadobacter fermentans DSM 18053]. | 92,30% | 1 | 1 | 1 | 2,39% | FQEAVDLIEK |
| ADL56804 | GT30 | Glycosyl Transferase [Gallionella capsiferiformans ES2]. | 92,30% | 1 | 1 | 1 | 1,92% | AFLAastr |
| AFE09168 | GT30 | 3deoxyDmannooctulosonicacid transferase [Coralococcus coralloides DSM 2259]. | 93,00% | 1 | 1 | 3 | 2,84% | QSGASVVMTNGR |
| AFX99576 | GT30 | Glycosyl transferases [Candidatus Endolissoclinum patella L2]. | 90,00% | 1 | 1 | 1 | 2,36% | EVNDAGSLAK |
| CCG99435 | GT30 | 3deoxyDmannooctulosonicacid transferase [Fibrella aestuarina BUZ 2]. | 94,90% | 1 | 1 | 1 | 4,15% | VMGIVDGLALQGMNGA A |
| ABG81487 | GT31 | coatomer protein complex subunit gamma 1 [Bos taurus]. | 94,90% | 1 | 1 | 2 | 1,95% | SIATLAITLLK |
| AFP63269 | GT31 | hypothetical protein [Musca domestica]. | 93,50% | 1 | 1 | 1 | 0,95% | AIDEEIK |
| BAD17751 | GT31 | putative avr9 elicitor response protein [Oryza sativa Japonica Group]. | 94,50% | 1 | 1 | 1 | 1,75% | STLELLR |
| ADO46406 | GT32 | hypothetical protein Entcl_0128 [Enterobacter lignolyticus SCF1]. | 96,30% | 1 | 1 | 1 | 4,19% | SVVDAQPVMGNQR |
| AEO32168 | GT32 | hypothetical protein [Amblyomma maculatum]. | 99,00% | 1 | 1 | 1 | 4,58% | GXIYLDLDVIVLK |
| CAI34541 | GT32 | putative glycosyl transferase [Streptococcus pneumoniae]. | 93,60% | 1 | 1 | 1 | 2,48% | NQMLELEK |
| AEN82277 | GT34 | AT3G62720like protein partial [Capsella grandiflora]. | 93,10% | 1 | 1 | 1 | 13,50% | DRPAFEADDQSAMVYLLAX ER |
| CCF58764 | GT34 | hypothetical protein KAFR_0F01670 [Kazachstania africana CBS 2517]. | 92,40% | 1 | 1 | 1 | 2,61% | DIPVMAGSNVR |
| ABS23705 | GT35 | glycogen starch alpha-glucan phosphorylase [Bacillus cytotoxicus NVH 39198]. | 92,70% | 1 | 1 | 1 | 0,87% | EIGLSLK |
| ACX75684 | GT35 | glycogen starch alpha-glucan phosphorylase [Fibrobacter succinogenes subsp succinogenes S85]. | 99,90% | 1 | 2 | 3 | 3,03% | IIPAADLSEQISTAGTEASGT GNMK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ADN00381 | GT35 | Glycogen phosphorylase [Dickeya dadantii 3937]. | 92,30% | 1 | 1 | 1 | 0,87% | EPDEWTR |
| BAI92994 | GT35 | glycogen phosphorylase [Arthrospira platensis NIES39]. | 94,90% | 1 | 1 | 1 | 0,94% | VLSENIK |
| BAD35174 | GT37 | putative galactoside 2alphaL-fucosyltransferase [Oryza sativa Japonica Group]. | 98,20% | 1 | 1 | 1 | 1,59% | DAVFYLLSR |
| AEA46631 | GT4 | glycosyl transferase group 1 [Archaeoglobus veneficus SNP6]. | 93,20% | 1 | 1 | 1 | 2,08% | EVINALSK |
| CAL78844 | GT41 | putative O-linked N-acetylglucosamine transferase SPINDLY family TPR domain protein [Bradyrhizobium sp ORS 278]. | 98,60% | 1 | 1 | 1 | 1,08% | GLASIALR |
| CAI93182 | GT43 | glycosyl transferase [Sorghum bicolor]. | 94,90% | 1 | 1 | 1 | 2,65% | AWPHDPSELLK |
| AAB07008 | GT47 | EXT2 [Homo sapiens]. | 98,50% | 1 | 1 | 1 | 1,53% | ETAQAMAQLSR |
| BAF15530 | GT47 | Os04g0574100 [Oryza sativa Japonica Group]. | 93,80% | 1 | 1 | 1 | 1,48% | AGNVTVR |
| BAB90325 | GT48 | putative glucan synthase [Oryza sativa japonica cultivargroup]. | 93,90% | 1 | 1 | 1 | 1,16% | SILLDPR |
| CAH01680 | GT48 | KLLA0C14069p [Kluyveromyces lactis]. | 94,90% | 1 | 1 | 1 | 0,45% | TLSEIGAK |
| XP_327157 | GT48 | hypothetical protein [Neurospora crassa OR74A]. | 92,90% | 1 | 1 | 1 | 0,61% | TDNLSPYTPGVK |
| ACE84974 | GT5 | ADPglucose:starch glucosyltransferase gt5A [Cellvibrio japonicus Ueda107]. | 93,70% | 1 | 1 | 1 | 1,87% | QLQINAMNR |
| ADI25655 | GT5 | glycogenstarch synthase ADPglucose type [Geobacillus sp C56T3]. | 92,10% | 1 | 1 | 1 | 1,44% | QPSIWER |
| ADU27513 | GT5 | glycogenstarch synthase ADPglucose type [Ethanoligenens harbinense YUAN3]. | 92,30% | 1 | 1 | 1 | 2,75% | QFNAHDMIGAISR |
| AEC16078 | GT5 | glycogen synthase [Gallibacterium anatis UMN179]. | 94,90% | 1 | 1 | 1 | 3,78% | ATGGLADTVVDTTPETQK |
| CAB49000 | GT5 | glgA glycogen synthase [Pyrococcus abyssi GE5]. | 92,10% | 1 | 1 | 1 | 2,06% | VFGEVQVK |
| CAQ80993 | GT5 | glycogen synthase [Aliivibrio salmonicida LF1238]. | 92,20% | 1 | 1 | 1 | 1,87% | SGGLADVAK |
| AAK88438 | GT51 | AGR_C_4930p [Agrobacterium tumefaciens str C58]. | 94,90% | 1 | 1 | 1 | 2,62% | LAGLIER |
| ACM37832 | GT51 | penicillin binding protein [Agrobacterium vitis S4]. | 96,70% | 1 | 1 | 1 | 1,29% | AGIVSAQDR |
| CAE77830 | GT51 | penicillin-binding protein 1A [Bdellovibrio bacteriovorus HD100]. | 98,90% | 1 | 1 | 1 | 0,78% | KPSEALK |
| AEA47330 | GT55 | mannosyl-3-phosphoglycerate synthase [Archaeoglobus veneficus SNP6]. | 97,60% | 1 | 1 | 1 | 1,55% | ILEELR |
| ABG89267 | GT61 | beta-1,2-xylosyl transferase isoform 2 [Lemna minor]. | 92,60% | 1 | 1 | 1 | 1,77% | GAVHVHTMR |
| AFN27720 | GT61 | glycosyltransferase partial [Coris gaimard]. | 92,10% | 1 | 1 | 1 | 3,10% | EGASSEDEK |
| AEO55174 | GT62 | glycosyl transferase family 62 protein [Myceliophthora thermophila ATCC 42464]. | 94,20% | 1 | 1 | 1 | 1,76% | AELQNMATR |
| CCE90185 | GT62 | hypothetical protein TDEL_0B00560 [Torulaspora delbrueckii]. | 92,00% | 1 | 1 | 1 | 2,95% | FEQEVSMMLDSTVEK |
| CBZ32808 | GT67 | phosphoglycan beta-1,3-galactosyl transferase putative [Leishmania donovani]. | 94,90% | 1 | 1 | 1 | 1,10% | SVLQAGGNVAEGAAR |
| AAI63087 | GT68 | Protein Ofucosyltransferase 2 [Danio rerio]. | 94,20% | 1 | 1 | 1 | 2,25% | GGPYLGVHLR |
| AAA50619 | GT7 | Hypothetical protein PAR24a [Caenorhabditis elegans]. | 94,30% | 1 | 1 | 1 | 0,87% | LAEQEAR |
| ADY43071 | GT7 | Chondroitin sulfate synthase 2 partial [Ascaris suum]. | 92,20% | 1 | 1 | 1 | 1,46% | VDVDEVLAAR |
| ACM39033 | GT70 | conserved hypothetical protein [Agrobacterium vitis S4]. | 94,50% | 1 | 1 | 1 | 1,83% | IGAFLPK |
| CAR23943 | GT71 | KLTH0F03938p [Lachancea thermotolerans CBS 6340]. | 96,90% | 1 | 1 | 1 | 1,60% | AICSDNPIR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| CAR25446 | GT71 | KLTH0G19690p [Lachanea thermotolerans CBS 6340]. | 94,40% | 1 | 1 | 1 | 0,00% | LLWINGGLR |
| CAX44723 | GT71 | alpha-1,3-mannosyl transferase putative [Candida dubliniensis CD36]. | 93,60% | 1 | 1 | 1 | 1,06% | ELLHDDEK |
| CCD56913 | GT71 | glycosyl transferase family 71 protein [Botryotinia fuckeliana T4]. | 92,00% | 1 | 1 | 1 | 1,54% | MLASSLSK |
| CCE72863 | GT71 | Piso0_000464 [Millerozyma farinosa CBS 7064]. | 94,60% | 1 | 1 | 1 | 1,15% | ADEDGDLFK |
| CCH60420 | GT71 | hypothetical protein TBLA_OC06250 [Tetrapisispora blattae CBS 6284]. | 98,90% | 1 | 1 | 1 | 0,95% | DALTLFR |
| CAR25639 | GT76 | ZYRO0A04642p [Zygosaccharomyces rouxii]. | 93,70% | 1 | 1 | 1 | 2,11% | FVSVADSSK |
| ABV79280 | GT8 | Lipopolysaccharide 1,2-glucosyl transferase RfaJ [Rickettsia bellii OSU 85389]. | 94,50% | 1 | 1 | 1 | 1,13% | YEEAAVEK |
| CBA75603 | GT8 | lipopolysaccharide 1,2-glucosyl transferase [Arsenophonus nasoniae]. | 99,20% | 1 | 1 | 1 | 1,80% | SLELVR |
| ABM72556 | GT83 | 4-amino-4-deoxy-L-arabinose transferase [Prochlorococcus marinus str MIT 9515]. | 94,90% | 1 | 1 | 1 | 0,99% | LFINSK |
| ADZ90791 | GT83 | glycosyl transferase family 39 [Marinomonas mediterranea MMB1]. | 92,50% | 1 | 1 | 1 | 1,12% | LFQPNK |
| ABJ87848 | GT84 | glycosyl transferase 36 [Candidatus Solibacter usitatus Ellin6076]. | 93,90% | 1 | 1 | 1 | 0,61% | EDGAVALR |
| CAE80185 | GT9 | heptosyl transferase [Bdellovibrio bacteriovorus HD100]. | 94,90% | 1 | 1 | 1 | 2,30% | AGLSGLIK |
| CBA33803 | GT9 | hypothetical protein Csp_B20870 [Curvibacter putative symbiont of Hydra magnipapillata]. | 93,70% | 1 | 1 | 1 | 2,08% | IGGGFALK |
| EGJ50555 | GT9 | glycosyl transferase family 9 [Desulfovibrio africanus str Walvis Bay]. | 92,70% | 1 | 1 | 1 | 2,53% | VLGSMLLFEMEAR |
| AAH93594 | GT90 | KDEL LysAspGluLeu containing 2 [Rattus norvegicus]. | 94,90% | 1 | 1 | 1 | 1,57% | NDGTFLVR |
| CAA96643 | GT92 | Protein F28H77 [Caenorhabditis elegans]. | 94,30% | 1 | 1 | 1 | 3,37% | FINPVQIMQEMPSR |
| ADG73654 | GT94 | putative glycosyl transferase [Cellulomonas flavigena DSM 20109]. | 98,60% | 1 | 1 | 1 | 1,66% | TAVALAR |
| Q15WM9 | not found! | Choline dehydrogenase and related flavoproteinslike protein [Pseudoalteromonas atlantica strain T6c ATCC BAA1087]. | 94,90% | 1 | 1 | 1 | 1,32% | INANIVK |
| AAM34295 | not found! | esterase [Acinetobacter lwoffii]. | 97,20% | 1 | 1 | 1 | 1,92% | NTIEIAK |
| ACJ00751 | not found! | Transferase putative [Rhodospirillum centenum strain ATCC 51521 SW]. | 99,90% | 1 | 1 | 1 | 2,17% | LFTTLVR |
| AAK68253 | not found! | Hypothetical protein C46C111a [Caenorhabditis elegans]. | 99,60% | 1 | 1 | 1 | 2,01% | IVMVGDSAGGNLIMSVNLR |
| Q4WQV0 | not found! | Glutathione Stransferase putative [Neosartorya fumigata strain ATCC MYA4609 Af293 CBS 101355 FGSC A1100]. | 92,60% | 1 | 1 | 1 | 4,12% | TSASGLSDITR |
| Q7UNV4 | not found! | Putative uncharacterized protein [Rhodopirellula baltica strain SH1]. | 99,90% | 1 | 1 | 1 | 0,16% | MVNGQVINNR |
| Q8IK83 | not found! | LCCL domain containing protein CCP1 [Plasmodium falciparum isolate 3D7]. | 99,70% | 2 | 2 | 2 | 1,11% | MQTLFEK VIIIITNNNEDK |
| Q8XKW3 | not found! | Hyaluronidase [Clostridium perfringens strain 13 Type A]. | 99,40% | 1 | 1 | 1 | 0,86% | EMVDVGIATK |
| Q97U60 | not found! | Glycolate oxidase glcD subunit GlcD [Sulfolobus solfataricus strain ATCC 35092 DSM 1617 JCM 11322 P2]. | 94,60% | 1 | 1 | 1 | 3,21% | VGGVPSGEHGIGIEK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AFK47751 | PL1 | unknown [Lotus japonicus]. | 94,10% | 1 | 1 | 1 | 1,79% | VTIAFNR |
| CCP20740 | PL1 | pectate lyase [Bacillus amyloliquefaciens subsp plantarum UCMB5036]. | 92,20% | 1 | 1 | 1 | 3,33% | SAAYPFYAWGAGR |
| AEU36651 | PL10 | pectate lyase [Granulicella mallensis MP5ACTX8]. | 93,90% | 1 | 1 | 3 | 1,99% | NLDMSGAPR |
| AFN74600 | PL12 | heparinase III protein [Melioribacter roseus P3M2]. | 98,60% | 1 | 1 | 1 | 0,92% | MFDAIVK |
| CBX91396 | PL3 | hypothetical protein LEMA_P0690401 [Leptosphaeria maculans JN3]. | 93,00% | 1 | 1 | 1 | 6,87% | AGLPVPASK |
| A1C995 | PL4 | Probable rhamnogalacturonate lyase A [Aspergillus clavatus strain ATCC 1007 CBS 51365 DSM 816 NCTC 3887 NRRL 1]. | 90,70% | 1 | 1 | 1 | 1,33% | NISGSVK |
| EAA59400 | PL4 | hypothetical protein AN41392 [Aspergillus nidulans FGSC A4]. | 92,20% | 1 | 1 | 1 | 0,67% | FSVAVAK |
| AFU99271 | PL6 | Poly-beta-D-mannuronate lyase [Simidiua agarivorans SA1 = DSM 21679]. | 99,90% | 1 | 1 | 1 | 0,82% | NNQFER |
| AAA25049 | PL7 | alginate lyase [Klebsiella pneumoniae]. | 92,10% | 1 | 1 | 1 | 2,28% | LTVTLMR |
| ADV42606 | PL8 | Chondroitin AC lyase [Bacteroides helcogenes P 36108]. | 99,80% | 1 | 1 | 1 | 1,77% | FISPDQYSSAIK |
| AFR73846 | PL8 | hyaluronate lyase precursor 1 [Staphylococcus aureus 08BA02176]. | 98,40% | 1 | 1 | 1 | 1,12% | NPVTSVENR |
| AGM07581 | PL8 | hyaluronate lyase [Amycolatopsis orientalis HCCB10007]. | 98,40% | 1 | 1 | 1 | 2,18% | STISIVSGAMLGDAER |
| Q82ZM8 | PL8 | Polysaccharide lyase family 8 [Enterococcus faecalis strain ATCC 700802 V583]. | 97,60% | 1 | 1 | 1 | 0,44% | LIQEQK |
| ADQ05619 | PL9 | Pectate disaccharidelyase [Caldicellulosiruptor owensensis OL]. | 94,40% | 1 | 1 | 1 | 0,53% | EEMIMVVMR |

Tabela 8: Lista de proteínas encontradas nas análises metaproteômicas da comunidade MCS , o banco de dados utilizado foi o banco do CAZy.

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|---|
| B9LP48 | AA NC | Catalase peroxidase [Halorubrum lacusprofundi strain ATCC 49239 DSM 5036 JCM 8891 ACAM 34]. | 97,60% | 1 | 1 | 1 | 1,40% | TAWASASTFR |
| O68601 | AA NC | Dissimilatory coppercontaining nitrite reductase [Alcaligenes xylosoxydans xylosoxydans]. | 100,00% | 3 | 3 | 3 | 11,10% | LTVNPNPGEQATLR VGALTGANALTAK VGETVLLIHSQANR |
| Q5KFA5 | AA5 | Glyoxal oxidase putative [Cryptococcus neoformans var neoformans serotype D strain JEC21 ATCC MYA565]. | 94,70% | 1 | 1 | 1 | 1,27% | AGVSSIPR |
| Q89HF0 | AA7 | Blr6041 protein [Bradyrhizobium japonicum strain USDA 110]. | 92,20% | 1 | 1 | 1 | 0,81% | EADEIAAR |
| AEI40312 | CBM16 | parallel betahelix repeatcontaining protein [Paenibacillus mucilaginosus KNP414]. | 97,10% | 1 | 1 | 1 | 1,69% | LGADVGGVVK |
| BAL89256 | CBM2 | putative glycosyl hydrolase [Actinoplanes missouriensis 431]. | 92,10% | 1 | 1 | 1 | 2,77% | NVSWNGALAPR |
| CBK97477 | CBM23 | Endo-beta-mannanase [Eubacterium siraeum 703]. | 92,40% | 1 | 1 | 1 | 0,93% | NNVNTIVSDK |
| ADZ82224 | CBM3 | Cellulase [Clostridium lentocellum DSM 5427]. | 92,10% | 1 | 1 | 1 | 1,03% | ENVNSSNNR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ACI19833 | CBM32 | F58 type C domain protein [Dictyoglomus thermophilum H612]. | 91,90% | 1 | 1 | 1 | 1,31% | HTVENVSIK |
| AEN08307 | CBM32 | APHP domain protein [Streptomyces sp SirexAEE]. | 99,20% | 1 | 1 | 1 | 0,99% | EQDATAFTGLAAAK |
| Q4PGC6 | CBM4 | Putative uncharacterized protein [Ustilago maydis strain 521 FGSC 9021]. | 94,70% | 1 | 1 | 1 | 1,49% | AFQDNESNSR |
| ADL52313 | CBM46 | LPXTGmotif cell wall anchor domain protein [Clostridium cellulovorans 743B]. | 92,60% | 1 | 1 | 1 | 0,75% | DGEITLK |
| Q7VJR0 | CBM5 | Flagellar hook protein FlgE [Helicobacter hepaticus strain ATCC 51449 3B1]. | 93,80% | 1 | 1 | 2 | 1,21% | AGLDLEVSR |
| AGL19933 | CBM6 | putative glycoside hydrolase [Actinoplanes sp N902109]. | 94,70% | 1 | 1 | 2 | 0,85% | VGDLAGEASGR |
| CAN90868 | CE NC | unnamed protein product [Sorangium cellulosum So ce56] | 92,30% | 1 | 1 | 1 | 7,50% | IDLVLVNPWSDAALR |
| YP_001641159 | CE NC | Glutathione Stransferase domain [Methylobacterium extorquens strain PA1]. | 94,20% | 1 | 1 | 1 | 6,10% | EALGDALAEAVIR |
| AEE48406 | CE1 | esterase [Haliscomenobacter hydrossis DSM 1100]. | 94,70% | 1 | 1 | 1 | 1,96% | QNIAMFK |
| CCG53982 | CE1 | Phospho-lipase Carboxylesterase precursor [Flavobacterium indicum GPTSA1009]. | 93,90% | 1 | 1 | 1 | 3,95% | SLSGTDINR |
| AEN13068 | CE12 | conserved hypothetical protein [Streptomyces sp SirexAEE]. | 94,10% | 1 | 1 | 1 | 2,60% | TGAAPAIKAPDR |
| ACX90963 | CE14 | LmbE family protein [Sulfolobus solfataricus 982]. | 93,20% | 1 | 1 | 1 | 4,15% | ILGIDEVK |
| AEW00642 | CE14 | LmbE family protein [Niastellia koreensis GR2010]. | 94,70% | 1 | 1 | 1 | 5,71% | GGIIDLQGELGTR |
| AGF56136 | CE14 | 1D-myo-inositol 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase MshB [Clostridium saccharoperbutylacetonicum N14HMT]. | 94,70% | 1 | 1 | 2 | 2,43% | IEITHR |
| ACB75551 | CE15 | hypothetical protein Oter_2268 [Opitutus terrae PB901]. | 91,80% | 1 | 1 | 1 | 0,98% | FENGTR |
| AAX55266 | CE5 | cutinase [Phytophthora infestans]. | 91,90% | 1 | 1 | 1 | 4,61% | AAAQAAALVK |
| CAZ97661 | CE6 | Carbohydrate esterase family CE6 Alphagalactosidase family GH27 [Zobellia galactanivorans]. | 93,70% | 1 | 1 | 1 | 0,92% | IEIDIDVK |
| ABE85704 | CE8 | Pectinesterase inhibitor [Medicago truncatula]. | 93,90% | 1 | 1 | 1 | 1,36% | AATDNVIK |
| ADU33260 | CE8 | pectin methyl esterase [Sitophilus oryzae]. | 94,30% | 1 | 1 | 1 | 1,85% | HGLSLAR |
| AEO67543 | CE8 | glycoside hydrolase family 28 protein [Thielavia terrestris NRRL 8126]. | 92,20% | 1 | 1 | 1 | 1,43% | NVTGTADPSAR |
| AEY65184 | CE8 | pectin methyl esterase [Clostridium sp BNL1100]. | 93,30% | 1 | 1 | 1 | 1,44% | LTSSASPK |
| BAK05168 | CE8 | predicted protein [Hordeum vulgare subsp vulgare]. | 94,70% | 1 | 1 | 2 | 1,65% | TATVAVVGK |
| CAP49460 | CE8 | exported pectinesterase [Xanthomonas campestris pv campestris]. | 94,70% | 1 | 1 | 1 | 4,31% | TVQAAIDAAVQGGK |
| CAW41285 | CE8 | unnamed protein product [Glycine max]. | 92,70% | 1 | 1 | 1 | 1,99% | SASSSDSELIEK |
| EAA57400 | GH 72 | hypothetical protein MG083704 [Magnaporthe grisea 7015]. | 92,90% | 1 | 1 | 1 | 1,31% | LGTNVIR |
| AAD28227 | GH1 | phospho-beta-glucosidase BglB [Enterococcus faecium]. | 98,30% | 1 | 1 | 1 | 1,68% | QGSYQVPR |
| AEB06286 | GH1 | 6-phospho beta-glucosidase [Coriobacterium glomerans PW2]. | 92,40% | 1 | 1 | 1 | 1,91% | NNESIAVER |
| ABN48479 | GH10 | endo-xylanase [Aspergillus fumigatus]. | 92,30% | 1 | 1 | 1 | 4,31% | QAASINQAFSTHGK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AFK83737 | GH10 | xylanase A [uncultured bacterium 35A20]. | 93,30% | 1 | 1 | 1 | 1,89% | DLASMVPAK |
| AGL15291 | GH10 | endo-1,4-beta-xylanase [Actinoplanes sp N902109]. | 91,80% | 1 | 1 | 1 | 2,37% | IVSQATSMGAR |
| ACA37625 | GH101 | endo-alpha-N-acetyl-galactosaminidase [Streptococcus pneumoniae Hungary19A6]. | 94,70% | 1 | 1 | 2 | 0,57% | LSSIVSDQGK |
| ABM60519 | GH102 | MLTA domain protein [Verminephrobacter eiseniae EF012]. | 94,70% | 1 | 1 | 1 | 2,30% | ISPMTSTLLR |
| CAR44556 | GH102 | membrane bound lytic murein transglycosylase [Proteus mirabilis HI4320]. | 94,70% | 1 | 1 | 1 | 2,75% | GASAIPLIAK |
| ACS57857 | GH103 | lytic murein transglycosylase [Rhizobium leguminosarum bv trifolii WSM1325]. | 93,30% | 1 | 1 | 1 | 2,21% | AAISAVQSR |
| ADG92984 | GH103 | membranebound lytic murein transglycosylase [Arcobacter nitrofigilis DSM 7299]. | 91,80% | 1 | 1 | 1 | 3,99% | MSGWQMGPVAVK |
| ABI87408 | GH104 | glycoside hydrolase family 24 [Burkholderia ambifaria AMMD]. | 92,60% | 1 | 1 | 1 | 4,67% | IYQAQMK |
| AFZ32114 | GH104 | glycoside hydrolase family 24 [Gloeocapsa sp PCC 7428]. | 91,50% | 1 | 1 | 1 | 3,27% | MLQEELK |
| CAK28349 | GH104 | Muramidase [Synechococcus sp RCC307]. | 94,70% | 1 | 1 | 1 | 4,11% | LGLGDFGPR |
| AAZ17388 | GH11 | endo-1,4-beta-xylanase [Bacillus amyloliquefaciens]. | 100,00% | 4 | 4 | 8 | 19,70% | GWTTGSPFR |
| | | | | | | | | SDGGTYDIYTTTR |
| | | | | | | | | TTFTQYWSVR |
| | | | | | | | | YNAPSIDGDR |
| AEH10296 | GH113 | hypothetical protein FsymDg_2979 [Frankia symbiont of Datisca glomerata]. | 92,20% | 1 | 1 | 1 | 1,82% | AAEQTIR |
| AEG48522 | GH114 | TM1410 hypothetical related protein [Sphingobium chlorophenolicum L1]. | 91,80% | 1 | 1 | 1 | 2,61% | NASGSSLK |
| AFR96657 | GH114 | endo alpha-1,4-polygalactosaminidase precursor [Cryptococcus neoformans var grubii H99]. | 93,10% | 1 | 1 | 1 | 1,96% | ISSFVAK |
| Q09563 | GH114 | Kelch repeat and BTB domaincontaining protein F47D127 [Caenorhabditis elegans]. | 99,30% | 1 | 1 | 1 | 2,07% | MWQALPDMAEAK |
| ACY47429 | GH115 | conserved hypothetical protein [Rhodothermus marinus DSM 4252]. | 97,80% | 1 | 1 | 1 | 0,82% | QVSAGEAR |
| BAD78876 | GH116 | hypothetical protein [Synechococcus elongatus PCC 6301]. | 97,80% | 1 | 1 | 1 | 1,00% | LAGAIAVR |
| AAW76376 | GH121 | conserved hypothetical protein [Xanthomonas oryzae pv oryzae KACC 10331]. | 92,10% | 1 | 1 | 1 | 0,49% | LLGSALR |
| AEK23688 | GH123 | Hypothetical protein Ccan_15720 [Capnocytophaga canimorsus Cc5]. | 94,70% | 1 | 1 | 1 | 1,68% | LPFEVQLQR |
| ABG39436 | GH127 | protein of unknown function DUF1680 [Pseudoalteromonas atlantica T6c]. | 94,70% | 1 | 1 | 1 | 1,95% | VTGQTNWLDLSIK |
| AEB44830 | GH127 | hypothetical protein VAB18032_18640 [Verrucosipora maris AB18032]. | 94,70% | 1 | 1 | 1 | 1,23% | VDAGRPPR |
| ABK03410 | GH13 | alpha amylase catalytic region [Arthrobacter sp FB24]. | 92,70% | 1 | 1 | 1 | 1,72% | IEAAGVDQAAR |
| ACB23595 | GH13 | maltotoligosyltrehalose synthase [Methylobacterium radiotolerans JCM 2831]. | 92,20% | 1 | 1 | 1 | 0,61% | LAAALSAEAR |
| ACL67608 | GH13 | alpha amylase catalytic region [Anaeromyxobacter dehalogenans CP1]. | 97,60% | 1 | 1 | 1 | 1,70% | MPPMPFAPR |
| ACV28521 | GH13 | alpha-alpha-phospho-trehalase [Anaerococcus prevotii DSM 20548]. | 97,90% | 1 | 1 | 1 | 1,63% | FDVLNVIGK |
| ACV59873 | GH13 | alpha amylase catalytic region [Alicyclobacillus acidocaldarius subsp acidocaldarius DSM 446]. | 93,50% | 1 | 1 | 2 | 1,04% | QLAWLK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|------------------------------------|
| ACV59878 | GH13 | alpha amylase catalytic region [Alicyclobacillus acidocaldarius subsp acidocaldarius DSM 446]. | 99,00% | 1 | 1 | 1 | 0,62% | NAVIDFFR |
| ADX47038 | GH13 | Protein of unknown function DUF3416 [Acidovorax avenae subsp avenae ATCC 19860]. | 92,40% | 1 | 1 | 1 | 0,97% | VAGEAVR |
| AEM46812 | GH13 | malto oligosyl trehalose synthase [Acidithiobacillus ferrovarans S53]. | 97,60% | 1 | 1 | 1 | 0,58% | ETELPGNGER |
| BAL91363 | GH13 | putative glycosyl hydrolase [Actinoplanes missouriensis 431]. | 93,20% | 1 | 1 | 1 | 1,51% | IDVANMTGR |
| CBS88422 | GH13 | glycosyl hydrolase glycogen debranching enzyme [Azospirillum lipoferum 4B]. | 94,70% | 1 | 1 | 1 | 0,98% | LMPDNPR |
| BAM93102 | GH130 | hypothetical protein S58_71370 [Bradyrhizobium oligotrophicum S58]. | 93,20% | 1 | 1 | 1 | 2,33% | MSPLQGSAAR |
| CCF12061 | GH130 | hypothetical protein [Paenibacillus sp Q8]. | 94,40% | 1 | 1 | 1 | 2,43% | MTVVDTK |
| ACB24299 | GH15 | glycoside hydrolase 15related [Methylobacterium radiotolerans JCM 2831]. | 94,70% | 1 | 1 | 1 | 1,43% | DEAGALIER |
| ACM57091 | GH15 | glycoside hydrolase 15related [Halorubrum lacusprofundi ATCC 49239]. | 94,70% | 1 | 1 | 1 | 0,50% | LDAALDAK |
| ADY02480 | GH15 | glycoside hydrolase 15related protein [Vulcanisaeta moutnovskia 76828]. | 94,70% | 1 | 1 | 2 | 1,59% | VLMVNGGLAR |
| AEI04010 | GH15 | glycosyl hydrolase family protein [Oligotropha carboxidovorans OM4]. | 93,10% | 1 | 1 | 1 | 1,50% | GDAGVEMR |
| ACO94508 | GH16 | BgIS27 [Streptomyces sp S27]. | 92,60% | 1 | 1 | 1 | 3,09% | VTSGGDSGGGGGGR |
| CCA25481 | GH16 | PREDICTED: hypothetical protein [Albugo laibachii Nc14]. | 98,90% | 1 | 1 | 1 | 0,43% | LIFSDEFNVSR |
| ADX73086 | GH18 | hypothetical protein Asphe3_19310 [Arthrobacter phenanthrenivorans Sphe3]. | 94,40% | 1 | 1 | 1 | 2,30% | ALAALFAK |
| CBG91178 | GH18 | putative polysaccharide degrading enzyme [Citrobacter rodentium ICC168]. | 94,20% | 1 | 1 | 1 | 1,59% | LDALGLQSDK |
| AAF64475 | GH19 | chitinase 2 [Cucumis melo]. | 98,80% | 1 | 1 | 1 | 2,94% | LGFATTGK |
| AAO75565 | GH2 | beta-mannosidase precursor [Bacteroides thetaiotaomicron VPI5482]. | 93,40% | 1 | 1 | 1 | 0,81% | ILNLTGK |
| ABU77081 | GH2 | hypothetical protein ESA_01827 [Cronobacter sakazakii ATCC BAA894]. | 92,60% | 1 | 1 | 1 | 0,00% | LAATGSTPQR |
| ACU60661 | GH2 | glycoside hydrolase family 2 sugar binding [Chitinophaga pinensis DSM 2588]. | 94,50% | 1 | 1 | 1 | 0,99% | NPLSEVNLK |
| ACV50538 | GH2 | glycoside hydrolase family 2 sugar binding [Atopobium parvulum DSM 20469]. | 96,90% | 1 | 1 | 1 | 0,73% | TWAQFDIEASVAR |
| ADB61987 | GH2 | Beta-galactosidase [Haloterrigena turkmenica DSM 5511]. | 97,20% | 1 | 1 | 1 | 0,93% | ELVTDPLGR |
| AEM74685 | GH2 | glycoside hydrolase family 2 sugar binding [Caldicellulosiruptor lactoaceticus 6A]. | 93,60% | 1 | 1 | 1 | 1,44% | VESISDELAIVK |
| ABQ26767 | GH20 | glycoside hydrolase family 20 [Geobacter uraniireducens Rf4]. | 92,00% | 1 | 1 | 1 | 0,73% | LSGDVFK |
| ACD05956 | GH20 | Beta-N-acetylhexosaminidase [Akkermansia muciniphila ATCC BAA835]. | 93,40% | 1 | 1 | 1 | 1,28% | DVVDELK |
| AEV87390 | GH20 | beta-hexosaminidase [Actinoplanes sp SE50110]. | 93,70% | 1 | 1 | 1 | 2,58% | MNEPLTSIVPRPAR |
| CCG99496 | GH20 | Beta-N-acetylhexosaminidase [Fibrella aestuarina BUZ 2]. | 94,00% | 1 | 1 | 1 | 1,58% | MValseaiwtar |
| 350715 | GH22 | lysozyme DL3 [Unknown]. | 94,70% | 2 | 2 | 2 | 24,00% | IVSDGDGMNAWVAWR NTDGSTYGLIQINSR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AAA48943 | GH22 | lysozyme [Gallus gallus]. | 100,00% | 3 | 3 | 4 | 23,80% | FESNFNTQATNR |
| | | | | | | | | HGLDNYR |
| | | | | | | | | NTDGSTDYGLIQNSR |
| AGK52048 | GH22 | lysozyme M1 [Bacillus sp 1NLA3E]. | 90,50% | 1 | 1 | 1 | 2,28% | AGSVDVVVK |
| ABI56371 | GH23 | Lytic transglycosylase catalytic [Alkalilimnicola ehrlichii MLHE1]. | 93,90% | 1 | 1 | 1 | 1,42% | VWNEELR |
| AEB24436 | GH23 | YqbO [Bacillus amyloliquefaciens TA208]. | 94,20% | 1 | 1 | 1 | 0,71% | YAGALETIAMK |
| AEO99558 | GH23 | Lytic transglycosylase catalytic [Bacillus coagulans 36D1]. | 94,70% | 1 | 1 | 1 | 6,06% | STISQLASLMELEALR |
| AGH89002 | GH23 | putative transglycosylase [Acinetobacter baumannii]. | 99,30% | 1 | 1 | 1 | 2,64% | LETQSACKPVK |
| ABU76879 | GH24 | hypothetical protein ESA_01625 [Cronobacter sakazakii ATCC BAA894]. | 94,70% | 1 | 1 | 2 | 4,32% | QIDPHIK |
| AEU37292 | GH24 | hypothetical protein Acix8_2989 [Granulicella mallensis MP5ACTX8]. | 93,70% | 1 | 1 | 2 | 3,42% | MGLGQMLR |
| ABM07147 | GH25 | putative lysozyme like protein [Arthrobacter aurescens TC1]. | 100,00% | 1 | 1 | 1 | 1,42% | GYGAPISLGTGWK |
| AEX45245 | GH25 | putative hydrolase [Corynebacterium diphtheriae 241]. | 94,70% | 1 | 1 | 1 | 2,13% | QAANSGLK |
| BAG30070 | GH25 | putative lysozyme [Kocuria rhizophila DC2201]. | 94,40% | 1 | 1 | 1 | 1,46% | NTSSIGGK |
| ADB42489 | GH26 | Mannan endo-1,4-betamannosidase [Spirosoma linguale DSM 74]. | 94,30% | 1 | 1 | 1 | 1,20% | SSEDIK |
| ADU14234 | GH27 | Alpha galactosidase [Asticcacaulis excentricus CB 48]. | 94,70% | 1 | 1 | 1 | 2,40% | EIADALVSSGLR |
| AAD25666 | GH28 | putative polygalacturonase [Arabidopsis thaliana]. | 98,40% | 1 | 1 | 1 | 2,48% | NIELIDINIK |
| ADZ81208 | GH29 | Alpha-L-fucosidase [Sphingobacterium sp 21]. | 92,20% | 1 | 1 | 1 | 2,23% | IADLANGYALK |
| ACU73743 | GH3 | glycoside hydrolase family 3 domain protein [Catenulispora acidiphila DSM 44928]. | 92,10% | 1 | 1 | 1 | 0,91% | NAVTDGTYTLR |
| AEV83323 | GH3 | beta-glucosidase [Actinoplanes sp SE50110]. | 94,70% | 1 | 1 | 1 | 0,73% | LGIAVFK |
| CBW21397 | GH3 | putative beta-glucosidase [Bacteroides fragilis 638R]. | 93,00% | 1 | 1 | 1 | 1,84% | NADAVVMVMGGSSAR |
| ADL05173 | GH30 | Cna B domain protein [Clostridium saccharolyticum WM1]. | 99,90% | 1 | 1 | 1 | 0,32% | ASEGDIINK |
| ABL84125 | GH31 | glycoside hydrolase family 31 [Nocardioides sp JS614]. | 94,70% | 1 | 1 | 1 | 1,75% | IGVTADPAVVAVR |
| ABQ07434 | GH31 | Candidate alpha-glycosidase Glycoside hydrolase family 31 [Flavobacterium johnsoniae UW101]. | 94,70% | 1 | 1 | 1 | 0,84% | NAVITIK |
| ADY36155 | GH31 | glycoside hydrolase family 31 [Bacteroides salanitronis DSM 18170]. | 98,00% | 1 | 1 | 1 | 1,25% | SILDNGPEYR |
| ACD22897 | GH32 | sucrose6phosphate hydrolase [Clostridium botulinum B str Eklund 17B]. | 92,00% | 1 | 1 | 1 | 1,65% | NNMELGGK |
| ADE68624 | GH32 | levanase [Bacillus megaterium QM B1551]. | 99,40% | 1 | 1 | 1 | 0,67% | NLMVTENK |
| AEG44848 | GH32 | Glycosyl hydrolase family 32 domain protein [Isoptericola variabilis 225]. | 99,00% | 1 | 1 | 1 | 0,49% | LLGGTK |
| AFC63922 | GH32 | glycoside hydrolase family 32 [Enterococcus faecium Aus0004]. | 90,70% | 1 | 1 | 1 | 1,48% | EEVTLVK |
| AAQ15986 | GH33 | trans sialidase putative [Trypanosoma brucei brucei strain 9274 GUTat101]. | 92,50% | 1 | 1 | 1 | 1,14% | ESSVLIWK |
| CCO19653 | GH33 | hypothetical protein PTSG_00864 [Bathycoccus prasinos]. | 97,40% | 1 | 1 | 1 | 1,97% | LSLAVSSGDGK |
| AAE35903 | GH35 | Sequence 2 from patent US 5981831 [Unknown]. | 98,80% | 1 | 1 | 1 | 0,82% | IVNIMK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| CAA20078 | GH35 | beta-galactosidase [Streptomyces coelicolor A32]. | 94,70% | 1 | 1 | 2 | 2,18% | ESASFTVTGTGAR |
| ABQ89756 | GH36 | glycoside hydrolase clan GHD [Roseiflexus sp RS1]. | 97,80% | 1 | 1 | 1 | 1,17% | ANVDVAAR |
| ADK30283 | GH36 | alpha-galactosidase [Brachyspira pilosicoli 951000]. | 91,00% | 1 | 1 | 1 | 0,99% | GLGDMSR |
| AGA84969 | GH37 | neutral trehalase [Pseudomonas stutzeri RCH2]. | 94,10% | 1 | 1 | 1 | 2,43% | YGHDALALEIEER |
| ACQ81244 | GH38 | glycoside hydrolase family 38 [Beutenbergia cavernae DSM 12333]. | 94,10% | 1 | 1 | 1 | 0,57% | EFFLTQK |
| ACX64156 | GH38 | glycosyl hydrolase 38 domain protein [Paenibacillus sp Y412MC10]. | 94,10% | 1 | 1 | 1 | 0,77% | EMLEYINR |
| ACL23353 | GH39 | conserved hypothetical protein [Chloroflexus aggregans DSM 9485]. | 94,70% | 1 | 1 | 1 | 0,79% | AVNELVK |
| AGL18049 | GH4 | alpha-galactosidase [Actinoplanes sp N902109]. | 94,70% | 1 | 1 | 1 | 2,09% | LETAEAIAR |
| ACM30171 | GH42 | beta-D-galactosidase protein [Agrobacterium radiobacter K84]. | 91,90% | 1 | 1 | 1 | 1,53% | LTGVTVEDFGR |
| ADI05003 | GH42 | putative beta-galactosidase [Streptomyces bingchengensis BCW1]. | 91,80% | 1 | 1 | 1 | 0,88% | MTVNSR |
| AEL24942 | GH43 | beta galactosidase I [Cyclobacterium marinum DSM 745]. | 92,00% | 1 | 1 | 1 | 3,44% | SLLTSPAGGDK |
| AGF58063 | GH43 | xylosidase arabinosidase XylB [Clostridium saccharoperbutylacetonicum N14HMT]. | 93,90% | 1 | 1 | 1 | 1,47% | GNFIITAK |
| CBK73885 | GH43 | Beta-xylosidase [Butyrivibrio fibrisolvens 164]. | 99,00% | 1 | 1 | 1 | 0,45% | IDMSGFDLTK |
| CAP96076 | GH47 | Pc21g11790 [Penicillium chrysogenum Wisconsin 541255]. | 92,30% | 1 | 1 | 1 | 1,08% | LSMVQDVR |
| CAR26380 | GH47 | ZYRO0B08360p [Zygosaccharomyces rouxii]. | 94,60% | 1 | 1 | 1 | 1,29% | LLAGILK |
| CCE65388 | GH47 | hypothetical protein TPHA_OL00320 [Tetrapisispora phaffii CBS 4417]. | 93,50% | 1 | 1 | 1 | 1,76% | ILDSFIENSK |
| AAQ77850 | GH5 | Sequence 2 from patent US 6329187 [Unknown]. | 93,90% | 1 | 1 | 1 | 1,81% | NWEDMLLQIK |
| AEI88911 | GH5 | glycosyl hydrolase family 5 protein [Candidatus Midichloria mitochondrii IricVA]. | 92,50% | 1 | 2 | 2 | 1,56% | IFQIIK |
| AEK40905 | GH5 | endo-glycosyl ceramidase [Amycolatopsis mediterranei S699]. | 94,70% | 1 | 1 | 1 | 1,53% | GFNVSGETK |
| AAW74571 | GH51 | alpha-L-arabinosidase [Xanthomonas oryzae pv oryzae KACC 10331]. | 94,60% | 1 | 1 | 1 | 1,52% | AVAPAVFK |
| BAM02957 | GH52 | beta-xylosidase [Phycisphaera mikurensis NBRC 102666]. | 99,20% | 1 | 1 | 1 | 1,60% | FPGETEADGSPR |
| ACS34213 | GH57 | Glycosyl hydrolase family 57 possibly alpha-amylase [Thermococcus gammatolerans EJ3]. | 97,60% | 1 | 1 | 1 | 1,17% | FEETGEK |
| ADV62137 | GH57 | 4-alpha-glucano transferase [Isosphaera pallida ATCC 43644]. | 92,00% | 1 | 1 | 1 | 1,07% | LIYDTPR |
| AFL75815 | GH57 | alpha-amylase alpha-mannosidase [Thiocystis violascens DSM 198]. | 92,40% | 1 | 1 | 1 | 2,00% | AEADLTASR |
| ACF11278 | GH65 | Kojibiose phosphorylase [Chlorobaculum parvum NCIB 8327]. | 97,60% | 1 | 1 | 1 | 1,01% | TGAPVEIK |
| ADH66550 | GH65 | glycoside hydrolase family 65 central catalytic [Nocardiopsis dassonvillei subsp dassonvillei DSM 43111]. | 94,70% | 1 | 1 | 1 | 1,02% | DPDQGQVR |
| CCH88085 | GH65 | Putative glycosidase [Modestobacter marinus]. | 94,30% | 1 | 1 | 1 | 0,89% | QPQTAAR |
| AAS79426 | GH70 | dextran sucrose [Leuconostoc mesenteroides]. | 97,70% | 1 | 1 | 1 | 0,62% | NFATVIDGK |
| AEJ53066 | GH70 | glucosyl transferase [Streptococcus salivarius 571]. | 94,70% | 1 | 1 | 1 | 0,63% | TVNENGAIR |

| Nº de Acesso | Familia CAZy | Nome da Proteína [organismo] | ID % | Nº UP | Nº US | Nº TS | Seq. coverage (%) | Peptide sequence |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| BAC07265 | GH70 | glucosyl transferase [Streptococcus sobrinus]. | 94,70% | 1 | 1 | 1 | 0,64% | SAALSLDNVK |
| BAF43788 | GH70 | glucosyl transferase [Streptococcus sanguinis SK1]. | 99,70% | 1 | 1 | 1 | 0,44% | QGAI AER |
| CBW23803 | GH76 | putative membrane attached hydrolase [Bacteroides fragilis 638R]. | 94,70% | 1 | 1 | 1 | 5,26% | ATQTLDSL YANYSVSGTCLL R |
| ABK75749 | GH77 | 4-alpha-glucanotransferase [Mycobacterium smegmatis str MC2 155]. | 92,90% | 1 | 1 | 1 | 1,23% | VAALAAHDR |
| ACT05921 | GH78 | alpha-L-rhamnosidase [Dickeya zeae Ech1591]. | 94,70% | 1 | 1 | 1 | 1,36% | SEFYLSQASALK |
| ABI86817 | GH8 | Cellulase [Burkholderia ambifaria AMMD]. | 92,90% | 1 | 1 | 1 | 1,72% | MAWAFAR |
| Q1YUJ0 | GH81 | Putative uncharacterized protein [gamma proteobacterium HTCC2207]. | 92,40% | 1 | 1 | 1 | 0,73% | VAELAAIAR |
| BAM62822 | GH83 | hemagglutininneuraminidase partial [Human parainfluenza virus 4a]. | 94,70% | 1 | 1 | 1 | 1,51% | EGGTCTVNR |
| CAP51686 | GH84 | Putative membrane protein [Xanthomonas campestris pv campestris]. | 93,50% | 1 | 1 | 1 | 1,38% | EPYPAATLK |
| Q8XNP3 | GH85 | Endo-beta-N-acetyl-glucosaminidase [Clostridium perfringens strain 13 Type A]. | 91,80% | 1 | 1 | 1 | 0,97% | DANGNFPIVDK |
| ACP49076 | GH88 | glycosyl hydrolase family 88 [Sulfolobus islandicus YN1551]. | 94,70% | 1 | 1 | 1 | 1,96% | ITDAIDK |
| ADY51417 | GH88 | glycosyl hydrolase family 88 [Pedobacter saltans DSM 12145]. | 99,00% | 1 | 1 | 1 | 1,50% | FIFSNK |
| AGJ89148 | GH88 | Unsaturated glucuronyl hydrolase [Raoultella ornithinolytica B6]. | 92,60% | 1 | 1 | 1 | 1,77% | NMVELSR |
| ABK39743 | GH9 | endo-chitinase [Aeromonas hydrophila subsp hydrophila ATCC 7966]. | 97,90% | 1 | 1 | 1 | 2,08% | QGGGMAIAALAR |
| ADL51173 | GH9 | LPXTG motif cell wall anchor domain protein [Clostridium cellulovorans 743B]. | 93,60% | 1 | 1 | 1 | 0,78% | SGDIGDIK |
| AEN70809 | GH9 | endo-1,4-beta-glucanase [Gossypium hirsutum subsp latifolium]. | 94,70% | 1 | 1 | 1 | 1,78% | LVGAMVAGPDK |
| AAO77306 | GH92 | alpha-1,2-mannosidase [Bacteroides thetaiotaomicron VPI5482]. | 93,10% | 1 | 1 | 1 | 1,19% | GIQTVDAEK |
| CCD45752 | GH92 | glycoside hydrolase family 92 protein [Botryotinia fuckeliana T4]. | 91,30% | 1 | 1 | 1 | 1,02% | VGISWISK |
| ABJ07582 | GH94 | glycosyl transferase 36 [Rhodospseudomonas palustris BisA53]. | 99,60% | 1 | 1 | 1 | 0,28% | ASGTDVIK |
| ADH90203 | GH94 | glycosyl transferase 36 [Starkeya novella DSM 506]. | 97,60% | 1 | 1 | 1 | 0,28% | LLTAGDSR |
| AEG00628 | GH94 | glycosyl transferase 36 [Methylomonas methanica MC09]. | 94,70% | 1 | 1 | 1 | 0,28% | VLLIENLR |
| CCE99088 | GH94 | K13688 cyclic beta-1,2-glucan synthetase [Sinorhizobium fredii HH103]. | 97,30% | 1 | 1 | 1 | 0,39% | QPTGVAPDDAR |
| AEK23108 | GH95 | Alpha-1,2-fucosidase 2 [Capnocytophaga canimorsus Cc5]. | 93,00% | 1 | 1 | 1 | 1,36% | EAEIIAQTAGK |
| AEU37001 | GH95 | Alpha-L-fucosidase [Granulicella mallensis MP5ACTX8]. | 92,50% | 1 | 1 | 1 | 0,89% | GGLSFHR |
| ADK15979 | GT1 | putative glycosyl transferase [Clostridium ljungdahlii DSM 13528]. | 91,90% | 1 | 1 | 1 | 2,49% | WNANNIQSAVK |
| AGM10421 | GT1 | glycosyl transferase [Amycolatopsis orientalis HCCB10007]. | 94,70% | 1 | 1 | 1 | 2,18% | AAIDGLAR |
| CAQ00470 | GT1 | hypothetical protein CMS0349 [Clavibacter michiganensis subsp sepedonicus]. | 94,20% | 1 | 1 | 1 | 2,12% | TLSSAAPPR |
| CAH04970 | GT10 | alpha-3-fucosyl transferase [Ciona savignyi]. | 93,50% | 1 | 1 | 1 | 1,58% | MIQLIK |

| Nº de Acesso | Familia CAZy | Nome da Proteína [organismo] | ID % | Nº UP | Nº US | Nº TS | Seq. coverage (%) | Peptide sequence |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AAY74589 | GT18 | Sequence 4 from patent US 6897052 [Unknown]. | 94,70% | 1 | 1 | 1 | 0,54% | TNVSGSIR |
| ACE06134 | GT19 | hypothetical protein Aasi_0753 [Candidatus Amoebophilus asiaticus 5a2]. | 99,40% | 1 | 1 | 1 | 1,62% | ELIQEK |
| AEM23121 | GT19 | putative lipid A disaccharide synthase [Brachyspira intermedia PWSA]. | 93,10% | 1 | 1 | 1 | 2,93% | MLNDMLSSNIR |
| AFI04277 | GT19 | Lipid-A-disaccharide synthase [Helicobacter cetorum MIT 007128]. | 92,80% | 1 | 1 | 1 | 1,68% | IGFILR |
| CCH00776 | GT19 | lipid-A-disaccharide synthase [Fibrella aestuarina BUZ 2]. | 94,70% | 1 | 1 | 1 | 1,62% | ATEAVAR |
| ADJ15508 | GT2 | glycosyl transferase family 2 [Halalkalicoccus jeotgali B3]. | 94,70% | 1 | 1 | 1 | 1,83% | IDGNALR |
| ADY41481 | GT2 | N-acetyl-galactos aminyl transferase 6 [Ascaris suum]. | 94,70% | 1 | 1 | 1 | 0,65% | WAIQLK |
| ADZ09154 | GT2 | glycosyl transferase family 2 [Methanobacterium sp AL21]. | 93,10% | 1 | 1 | 1 | 0,89% | IGGFDESLK |
| AEF97062 | GT2 | glycosyl transferase family 2 [Methanoterris igneus Kol 5]. | 98,20% | 1 | 1 | 2 | 1,69% | KPYLLK |
| AFY35674 | GT2 | glycosyl transferase family 2 [Calothrix sp PCC 7507]. | 93,20% | 1 | 1 | 1 | 0,74% | IGEPSTVLIR |
| BAI62619 | GT2 | putative glycosyl transferase [Methanocella paludicola SANAE]. | 93,30% | 1 | 1 | 1 | 4,44% | MLDMATIISGVK |
| AAZ98579 | GT20 | glucosyl glycerol phosphate synthase [Thiobacillus denitrificans ATCC 25259]. | 94,70% | 1 | 1 | 1 | 1,35% | IIVSIGR |
| ADE12748 | GT20 | Alpha, alpha-trehalose phosphate synthase UDPforming [Sideroxydans lithotrophicus ES1]. | 92,20% | 1 | 1 | 1 | 0,81% | FHAIER |
| ADG99367 | GT20 | Alpha-alpha-trehalose-phosphate synthase UDP forming [Segniliparus rotundus DSM 44985]. | 94,30% | 1 | 1 | 1 | 1,38% | AVEEILK |
| ADY55086 | GT20 | HAD-superfamily hydrolase subfamily IIB [Syntrophobotulus glycolicus DSM 8271]. | 92,80% | 1 | 1 | 1 | 1,22% | MPGSLLEEK |
| AFG37273 | GT20 | UDP-forming alpha-alpha-trehalose phosphate synthase [Spirochaeta africana DSM 8902]. | 94,50% | 1 | 1 | 1 | 1,49% | VTLTMIVAPSR |
| AAK73021 | GT21 | ceramide glucosyl transferase [Gossypium arboreum]. | 93,20% | 1 | 1 | 1 | 1,35% | FTDLGGK |
| CAF32096 | GT22 | glycosyl transferase putative [Aspergillus fumigatus]. | 92,40% | 1 | 1 | 1 | 1,44% | TSEFYTIR |
| AAF26036 | GT24 | putative UDPglucose:glycoprotein glucosyltransferase 10120091134 [Arabidopsis thaliana]. | 94,10% | 1 | 1 | 1 | 0,36% | QEHTLK |
| AAI42351 | GT25 | GLT25D1 protein [Bos taurus]. | 94,00% | 1 | 1 | 2 | 1,12% | EWLVAVK |
| ABL70841 | GT25 | glycosyl transferase family 25 [Paracoccus denitrificans PD1222]. | 92,20% | 1 | 1 | 1 | 4,20% | QGAETVSTHR |
| ACO68229 | GT25 | glycosyl transferase family 25 protein [Micromonas sp RCC299]. | 93,90% | 1 | 1 | 1 | 2,94% | MMNVPVSR |
| ACT48627 | GT25 | glycosyl transferase family 25 [Methylothera mobilis JLW8]. | 92,70% | 1 | 1 | 1 | 3,20% | FSAVSFEK |
| ABF44080 | GT26 | glycosyl transferase WecBTagACpsF family [Deinococcus geothermalis DSM 11300]. | 91,90% | 1 | 1 | 1 | 2,99% | IAGMEASK |
| AAO34932 | GT28 | 1,2-diacylglycerol 3glucosyltransferase [Clostridium tetani E88]. | 94,70% | 1 | 1 | 1 | 2,06% | NIYNLILK |
| ABV62098 | GT28 | beta-N-acetylglucosaminyl transferase [Bacillus pumilus SAFR032]. | 94,70% | 1 | 1 | 1 | 1,65% | VINEIK |
| ADZ82781 | GT28 | UDP-N-acetylglucosamine-N-acetylmuramyl pentapeptide pyrophosphorylundecaprenol N-acetylglucosamine transferase [Clostridium lentocellum DSM 5427]. | 94,70% | 1 | 1 | 1 | 2,56% | GVYTGSPIR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|--------------------------|
| AEO08574 | GT28 | undecaprenyldiphospho- muramoylpentapeptide beta-N- acetylglucosaminyltransferase[Buchner a aphidicola str Ak Acyrthosiphon kondoii]. | 92,70% | 1 | 1 | 1 | 1,98% | IAGITNR |
| AFM21517 | GT28 | UDPN-acetyl-glucosamine:LPS Nacetylglucosamine transferase [Anaerobaculum mobile DSM 13181]. | 93,30% | 1 | 1 | 1 | 2,19% | QSIIEVAK |
| CBW73839 | GT28 | N-acetylglucosamine transferase EC 241227 [Burkholderia rhizoxinica HKI 454]. | 99,20% | 1 | 1 | 1 | 2,39% | ASLACVAAR |
| AEE12699 | GT3 | Glycogen starch synthase [Porphyromonas asaccharolytica DSM 20707]. | 92,00% | 1 | 1 | 1 | 1,11% | IGEFLR |
| ABB32573 | GT30 | CMP-3-deoxy-D-manno-octulosonate-- lipid A tetraacyldisaccharide 3-deoxy-D- manno-octulosonate transferase[Geobacter metallireducens GS15]. | 94,20% | 1 | 1 | 1 | 3,21% | GGVLLVDTIGELMK |
| ABQ04758 | GT30 | Glycosyltransferase family 30 [Flavobacterium johnsoniae UW101]. | 94,70% | 1 | 1 | 3 | 1,96% | GATNTIMK |
| ADP85473 | GT30 | Three deoxy-D-mannooctulosonicacid transferase domaincontaining protein [Desulfovibrio vulgaris RCH1]. | 99,00% | 1 | 1 | 1 | 1,89% | ADGAGMPPVR |
| AEI78134 | GT30 | 3-deoxy-D-mannooctulosonic acid transferase KdtA [Cupriavidus necator N1]. | 93,50% | 1 | 1 | 1 | 2,56% | TAASVLADPAR |
| AAC60100 | GT31 | radical fringe [Gallus gallus]. | 93,30% | 1 | 1 | 1 | 2,42% | MGPPGGS AK |
| ABK24735 | GT31 | unknown [Picea sitchensis]. | 94,70% | 1 | 1 | 1 | 5,88% | ISTIMLSMAAVMASIYVAG R |
| ACG24856 | GT31 | galactosyl transferase transferring hexosyl groups [Zea mays]. | 94,70% | 1 | 1 | 1 | 2,87% | DGEAAVMVSQFMVELVGT K |
| ACP98530 | GT31 | Sequence 30748 from patent US 7504490 [Unknown]. | 92,90% | 1 | 1 | 1 | 1,99% | IFALVEPDER |
| BAJ89142 | GT31 | predicted protein [Hordeum vulgare subsp vulgare]. | 93,20% | 1 | 1 | 1 | 2,76% | SGPVLAQEGVK |
| ABN47391 | GT32 | EpsO [Lactococcus lactis subsp cremoris]. | 94,70% | 1 | 1 | 1 | 2,55% | NYDVNKK |
| AGK52658 | GT32 | mannosyl transferase OCH1like enzyme [Bacillus sp 1NLA3E]. | 92,10% | 1 | 1 | 1 | 8,61% | FNENGTNNELTNVAIITEILK |
| ABV15883 | GT35 | hypothetical protein CKO_04838 [Citrobacter koseri ATCC BAA895]. | 99,40% | 1 | 1 | 1 | 0,88% | LAVVHDK |
| ACB40238 | GT35 | glycosyl transferase group 1 [Pyrobaculum neutrophilum V24Sta]. | 94,70% | 1 | 1 | 1 | 1,45% | ALGVEAR |
| ACD51074 | GT35 | glycogen phosphorylase [Clostridium botulinum E3 str Alaska E43]. | 94,70% | 1 | 1 | 1 | 0,86% | IMDIYNR |
| AGC67926 | GT35 | glycogen phosphorylase GIGP [Clostridium stercorarium subsp stercorarium DSM 8532]. | 94,70% | 1 | 1 | 2 | 0,70% | ASLIFR |
| AGJ61955 | GT35 | Glucan phosphorylase [Sulfolobus islandicus LAL141]. | 97,50% | 1 | 1 | 1 | 1,76% | LINEIEDK |
| AGH41934 | GT39 | glycosyl transferase family 39 [Bifidobacterium thermophilum RBL67]. | 92,70% | 1 | 1 | 1 | 1,69% | MVLSSVNSGR |
| CAR27316 | GT39 | ZYRO0C14036p [Zygosaccharomyces rouxii]. | 94,20% | 1 | 1 | 1 | 0,91% | CKPLIR |
| ADB63892 | GT4 | glycosyl transferase group 1 [Haloterrigena turkmenica DSM 5511]. | 94,50% | 1 | 1 | 1 | 2,48% | LAGVDPSR |
| AEG18294 | GT4 | Phosphatidy linositol alpha-mannosyl transferase [Methanobacterium sp SWAN1]. | 92,30% | 1 | 1 | 1 | 2,27% | LGGIPDIVK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ABQ37575 | GT41 | putative O linked N-acetylglucosamine transferase SPINDLY family [Bradyrhizobium sp BTAi1]. | 97,70% | 1 | 1 | 1 | 1,08% | GLASIILR |
| CCD00780 | GT41 | protein of unknown function [Azospirillum brasilense Sp245]. | 93,60% | 1 | 1 | 1 | 15,20% | MAGVEGCVAR |
| AAF61913 | GT47 | multiple exo stoses like 1 protein [Mus musculus]. | 97,30% | 1 | 1 | 1 | 0,90% | EMLPSR |
| AED94135 | GT47 | Exo stosin family protein [Arabidopsis thaliana]. | 93,90% | 1 | 1 | 1 | 1,46% | DSEMLSAR |
| ABE81542 | GT48 | putative callose synthase catalytic subunitrelated [Medicago truncatula]. | 93,60% | 1 | 1 | 1 | 1,65% | ESGAFSR |
| ACS36253 | GT48 | glucan synthase like 7 [Hordeum vulgare]. | 93,30% | 1 | 1 | 2 | 0,62% | GQEDGAALMK |
| ADX07311 | GT48 | putative 1,3-beta-glucan synthase [Flammulina velutipes]. | 97,70% | 1 | 1 | 1 | 0,25% | IVLNDK |
| ACA17488 | GT5 | glycogenstarch synthase ADPglucose type [Methylobacterium sp 446]. | 99,40% | 1 | 1 | 1 | 2,55% | TGGLGEVAAALPR |
| ABW26724 | GT51 | penicillin binding protein 1A putative [Acarochloris marina MBIC11017]. | 93,70% | 1 | 1 | 1 | 1,58% | FDQAALVSLDPR |
| ACL42333 | GT51 | glycosyl transferase family 51 [Arthrobacter chlorophenolicus A6]. | 94,70% | 1 | 1 | 1 | 1,04% | LGITSLPR |
| ADM37076 | GT51 | penicillin binding protein 1 [Bacillus subtilis subsp spizizenii str W23]. | 92,20% | 1 | 1 | 1 | 1,12% | EAVSAQGR |
| ADV42290 | GT51 | glycosyl transferase family 51 [Bacteroides helcogenes P 36108]. | 93,90% | 1 | 1 | 1 | 0,90% | NIFLNR |
| AEN91517 | GT51 | Membrane carboxypeptidase Penicillinbinding proteinlike protein [Bacillus megaterium WSH002]. | 99,70% | 1 | 1 | 1 | 1,55% | SGGVLSVMGGR |
| AGH94377 | GT51 | penicillin binding protein 1A [Bdellovibrio exovoror JSS]. | 94,70% | 1 | 1 | 1 | 1,14% | DILLAIQMEK |
| BAI71171 | GT51 | penicillin binding protein 1A [Azospirillum sp B510]. | 94,70% | 1 | 1 | 1 | 1,04% | LMAEFAAER |
| ACY19589 | GT53 | cell wall arabinan synthesis protein [Gordonia bronchialis DSM 43247]. | 94,70% | 1 | 1 | 2 | 0,55% | VGLFIR |
| ACY19590 | GT53 | cell wall arabinan synthesis protein [Gordonia bronchialis DSM 43247]. | 92,40% | 1 | 1 | 1 | 0,62% | SVEVVNR |
| BAD56627 | GT53 | putative arabinosyltransferase [Nocardia farcinica IFM 10152]. | 94,70% | 1 | 1 | 1 | 0,55% | LDVVLR |
| AFP59373 | GT54 | N-A-cetylglucosaminyl transferase [Musca domestica]. | 94,70% | 1 | 1 | 1 | 1,99% | ITLNDSLER |
| AAQ15697 | GT60 | hypothetical protein conserved [Trypanosoma brucei brucei strain 9274 GUTat101]. | 97,60% | 1 | 1 | 1 | 1,07% | FAGLDPVR |
| CAN72290 | GT61 | hypothetical protein VITISV_007350 [Vitis vinifera]. | 99,90% | 1 | 1 | 1 | 1,36% | ESSLADK |
| ADX82476 | GT66 | oligosaccharyl transferase STT3 subunit [Sulfolobus islandicus HVE104]. | 91,30% | 1 | 1 | 1 | 0,79% | EIVDSR |
| AFS81830 | GT66 | d-olichyldiphosphooligosaccharideprotein glycotransferase [Candidatus Nitrosopumilus sp AR2]. | 94,50% | 1 | 1 | 1 | 1,40% | LPSDGDGPLR |
| AFH32889 | GT7 | chondroitin sulfate glucuronyl transferase [Macaca mulatta]. | 94,70% | 1 | 1 | 1 | 1,30% | LSNLEGLGGR |
| CAR22176 | GT71 | KLTHOC08976p [Lachancea thermotolerans CBS 6340]. | 94,70% | 1 | 1 | 1 | 1,14% | ALALLVSR |
| CCK72377 | GT71 | hypothetical protein KNAG_0J02990 [Kazachstania naganishii CBS 8797]. | 91,20% | 1 | 1 | 1 | 1,24% | MNGVLIPIK |
| EAA29160 | GT71 | predicted protein [Neurospora crassa OR74A]. | 93,60% | 1 | 1 | 1 | 2,26% | ASSESGMIVLDK |
| ADE03063 | GT75 | conserved hypothetical protein [Haloferax volcanii DS2]. | 91,80% | 1 | 1 | 1 | 3,38% | DWLDCLDALAAPR |

| Nº de Acesso | Familia CAZy | Nome da Proteína [organismo] | ID % | Nº UP | Nº US | Nº TS | Seq. coverage (%) | Peptide sequence |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| ACO61836 | GT77 | glycosyl transferase family 77 protein [Micromonas sp RCC299]. | 94,70% | 1 | 1 | 1 | 1,06% | NALALAR |
| ACO69274 | GT77 | glycosyl transferase family 77 protein [Micromonas sp RCC299]. | 92,60% | 1 | 1 | 1 | 1,07% | EAASDAIAK |
| CAL52578 | GT77 | unnamed protein product [Ostreococcus tauri]. | 98,70% | 1 | 1 | 1 | 1,50% | VLESGAPLSTGLK |
| ADN91571 | GT8 | Hypothetical protein CJM1_1381 [Campylobacter jejuni subsp jejuni M1]. | 92,60% | 1 | 1 | 1 | 1,62% | NNLQNAK |
| ADZ50891 | GT8 | putative lipopolysaccharide biosynthesis protein [Helicobacter pylori 2018]. | 92,20% | 1 | 1 | 1 | 3,49% | MVEDLSAENVAK |
| CBY86707 | GT8 | putative glycosyl transferase pseudogene [Haemophilus influenzae F3047]. | 92,40% | 1 | 1 | 1 | 2,43% | MIVLSEIR |
| AEK39560 | GT81 | putative glucosyl3phosphoglycerate synthase [Amycolatopsis mediterranei S699]. | 92,00% | 1 | 1 | 1 | 3,20% | QILGTALDR |
| ACY18108 | GT83 | 4-amino-4-deoxy-L-arabinose transferase and glycosyl transferase [Haliangium ochraceum DSM 14365]. | 93,20% | 1 | 1 | 1 | 0,87% | IAAAENR |
| AFZ61496 | GT83 | mannosyl transferase family protein [Pseudomonas fluorescens]. | 93,90% | 1 | 1 | 1 | 1,44% | ADAVDTR |
| ADL49744 | GT87 | hypothetical protein Micau_6251 [Micromonospora aurantiaca ATCC 27029]. | 94,70% | 1 | 1 | 1 | 3,80% | VGAADMTANQSLAGLLAR |
| BAC73197 | GT87 | hypothetical protein SAV_5485 [Streptomyces avermitilis MA4680]. | 92,40% | 1 | 1 | 1 | 2,23% | AGGDLYTLR |
| ABB24650 | GT9 | Lipopolysaccharide heptosyltransferase II [Chlorobium luteolum DSM 273]. | 91,60% | 1 | 1 | 1 | 2,23% | VEDLTGK |
| CBS90254 | GT9 | putative Glycosyl transferase family 9 [Azospirillum lipoferum 4B]. | 92,50% | 1 | 1 | 1 | 3,57% | LTGQQGALQGAR |
| CCG01410 | GT9 | ADP heptose:LPS heptosyltransferase [Blastococcus saxobsidens DD2]. | 93,80% | 1 | 1 | 1 | 2,69% | ALSAAVAGSR |
| A9VUM0 | not found! | hypothetical protein, partial [Bacillus cereus] | 92,30% | 1 | 1 | 1 | 1,07% | ATITNMEK |
| P87061 | not found! | Tip elongation aberrant protein 1 [Schizosaccharomyces pombe strain 972 ATCC 24843]. | 98,20% | 1 | 1 | 1 | 0,61% | GFGIFEK |
| Q7Q9L3 | not found! | AGAP005170PA [Anopheles gambiae]. | 92,70% | 1 | 1 | 1 | 1,11% | DAVSTVK |
| Q8S5V1 | not found! | Putative potassium proton antiporterlike protein [Oryza sativa subsp japonica]. | 93,10% | 1 | 1 | 1 | 1,68% | ATMLEAYNR |
| Q8WYK1 | not found! | Contactin associated proteinlike 5 [Homo sapiens]. | 92,80% | 1 | 1 | 1 | 0,61% | EPLTNAVR |
| WP_022098775 | not found! | putative uncharacterized protein [Eubacterium eligens CAG:72] | 94,70% | 1 | 1 | 1 | 2,68% | LMAADNK |
| YP_002496522 | not found! | AsnC family transcriptional regulator [Methylobacterium nodulans ORS 2060] | 92,40% | 1 | 1 | 1 | 0,43% | DATSPAVR |
| AAA24920 | PL13 | heparinase [Pedobacter heparinus]. | 93,60% | 1 | 1 | 1 | 1,82% | TSTIAYK |
| AGE48774 | PL14 | hypothetical protein PBCVAP110A_306R [Paramecium bursaria Chlorella virus AP110A]. | 98,80% | 1 | 1 | 1 | 1,18% | ESGSIVK |
| AEV65451 | PL7 | Alginate lyase precursor [Pseudomonas fluorescens F113]. | 97,50% | 1 | 1 | 1 | 4,04% | ETNSDGTLR |
| P37369 | SOD | Superoxide dismutase Fe [Bordetella pertussis strain Tohama I ATCC BAA589 NCTC 13251]. | 99,50% | 1 | 1 | 3 | 7,81% | SAAGNFGSGWTWLVK |
| P00448 | SOD A | Superoxide dismutase Mn [Escherichia coli strain K12]. | 94,70% | 1 | 1 | 2 | 2,91% | RPDYIK |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| WP_003080507 | SOD C | superoxide dismutase [Comamonas testosteroni] | 94,70% | 1 | 1 | 2 | 12,70% | AVIVHAQPDDYTTQPTGNAGAR |
| XP_321476 | SODC | AGAP001623PA [Anopheles gambiae]. | 99,40% | 1 | 1 | 1 | 0,69% | LSDAYEAR |

Tabela 9: Lista de proteínas encontradas nas análises metaproteômicas da comunidade MCF, o banco de dados utilizado foi o banco do CAZy.

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AAK44448 | CE NC | esterase lipase putative [Mycobacterium tuberculosis CDC1551]. | 94,90% | 1 | 1 | 1 | 3,30% | AYLGDADAR |
| ACU07564 | CE1 | hypothetical protein FIC_01114 [Flavobacteriaceae bacterium 351910]. | 94,90% | 1 | 1 | 1 | 1,97% | LSEELQNQTLEK |
| CCG53982 | CE1 | Phospho-lipase Carboxyl-esterase precursor [Flavobacterium indicum GPTSA1009]. | 94,90% | 1 | 1 | 1 | 3,95% | SLSGTDINR |
| AAL44669 | CE4 | conserved hypothetical protein [Agrobacterium tumefaciens str C58]. | 94,90% | 1 | 1 | 1 | 1,20% | EMAEKPLGK |
| ACD05322 | CE4 | polysaccharide deacetylase [Akkermansia muciniphila ATCC BAA835]. | 94,90% | 1 | 1 | 1 | 2,50% | IPGVTVNR |
| ADF40961 | CE4 | probable sporulation protein polysaccharide deacetylase family [Bacillus megaterium DSM 319]. | 94,90% | 1 | 1 | 1 | 1,84% | MLEVELK |
| ADU86930 | CE6 | putative acetyl xylan esterase [uncultured bacterium]. | 94,90% | 1 | 1 | 1 | 2,42% | FQMLACTADFINPK |
| BAB59827 | GH1 | beta-glycosidase [Thermoplasma volcanium GSS1]. | 94,90% | 1 | 1 | 1 | 2,08% | VENNFPEDFK |
| AFZ83895 | GH103 | peptidoglycan binding protein domaincontaining protein 1 [Candidatus Kinetoplastibacterium blastocritidii ex Strigomonas culicis]. | 94,90% | 1 | 1 | 1 | 2,24% | EEGILILK |
| AFZ32114 | GH104 | glycoside hydrolase family 24 [Gloeocapsa sp PCC 7428]. | 94,90% | 1 | 2 | 7 | 3,27% | MLQEELK |
| ACS97297 | GH13 | alpha-amylase 1,4-alpha-D-glucan glucanohydrolase [Aggregatibacter aphrophilus NJ8700]. | 94,90% | 1 | 1 | 3 | 1,02% | NDNSYGR |
| ACU39879 | GH13 | alpha amylase catalytic region [Actinosynnema mirum DSM 43827]. | 94,90% | 1 | 1 | 1 | 1,52% | MSPEGIGLDR |
| ACV59873 | GH13 | alpha amylase catalytic region [Alicyclobacillus acidocaldarius subsp acidocaldarius DSM 446]. | 94,90% | 1 | 1 | 4 | 1,04% | QLAWLK |
| ADW71079 | GH13 | alpha-1,6-glucosidase pullulanasetype [Granulicella tundricola MP5ACTX9]. | 94,90% | 1 | 1 | 1 | 0,87% | ASLNDSSAIR |
| AAK24253 | GH15 | gluco-amylase [Caulobacter crescentus CB15]. | 94,90% | 1 | 1 | 1 | 1,17% | DGDALAALK |
| ADL48466 | GH15 | trehalose phosphatase [Micromonospora aurantiaca ATCC 27029]. | 94,90% | 1 | 1 | 2 | 1,15% | VLENMVEAVR |
| ACO94508 | GH16 | BglS27 [Streptomyces sp S27]. | 94,90% | 1 | 1 | 5 | 3,09% | VTSGGDSGGGGGGR |
| AAC16474 | GH17 | putative beta-1,3-endoglucanase [Unknown]. | 94,90% | 1 | 1 | 1 | 1,19% | VVQMLK |
| BAB86263 | GH17 | putative glucan endo13betaDglucosidase [Oryza sativa japonica cultivargroup]. | 94,90% | 1 | 1 | 1 | 1,68% | TTVEATR |
| CAK39383 | GH17 | unnamed protein product [Aspergillus niger]. | 94,90% | 1 | 1 | 1 | 1,06% | NLKPQVK |
| ABG68450 | GH2 | beta-galactosidase [Escherichia coli 536]. | 99,80% | 2 | 2 | 2 | 2,15% | FNDDFSR |
| | | | | | | | | VNWVWGLGQENYPR |

| <u>Nº de Acesso</u> | <u>Familia CAZy</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|---|
| AEK23668 | GH2 | Lactase [Capnocytophaga canimorsus Cc5]. | 94,90% | 1 | 1 | 1 | 0,67% | TSGWNGMVK |
| AEK23864 | GH2 | Lactase [Capnocytophaga canimorsus Cc5]. | 94,90% | 1 | 1 | 1 | 0,77% | QLPTDFFK |
| AEM74685 | GH2 | glycoside hydrolase family 2 sugar binding [Caldicellulosiruptor lactoaceticus 6A]. | 94,90% | 1 | 1 | 2 | 1,44% | VESISDELAIVK |
| AGH73894 | GH2 | Beta galactosidase beta-glucuronidase [Edwardsiella tarda C07087]. | 94,90% | 1 | 1 | 1 | 0,96% | QATALVR |
| CBW22258 | GH2 | beta-mannosidase [Bacteroides fragilis 638R]. | 98,40% | 1 | 1 | 1 | 1,05% | ITITSPEIK |
| ABE96866 | GH20 | beta-N-acetyl-hexosaminidase [uncultured bacterium]. | 94,90% | 1 | 1 | 2 | 3,13% | QILAASNSDPQOSIK |
| ABU25236 | GH20 | putative beta-N-acetyl-hexosaminidase [Mycoplasma crocodyli MP145]. | 99,40% | 1 | 1 | 1 | 0,40% | DIEDTK |
| AGL14870 | GH20 | N-acetyl-beta-hexosaminidase [Actinoplanes sp N902109]. | 94,90% | 1 | 1 | 1 | 2,05% | IEAADEGTAR |
| AAA48943 | GH22 | lysozyme [Gallus gallus]. | 94,90% | 1 | 1 | 1 | 8,16% | FESNFNTQATNR |
| AAB41205 | GH22 | lysozyme c precursor [Gorilla gorilla]. | 100,00% | 2 | 2 | 2 | 14,20% | ATNYNAGDR STDYGIFQINSR |
| ACU03376 | GH28 | glycoside hydrolase family 28 [Pedobacter heparinus DSM 2366]. | 94,90% | 1 | 1 | 1 | 1,44% | NALLQFTK |
| ADF51358 | GH28 | glycoside hydrolase family protein [Zunongwangia profunda SMA87]. | 94,90% | 1 | 1 | 1 | 2,13% | FNDIDLEVSK |
| ADG20215 | GH28 | glycoside hydrolase family protein [Burkholderia sp CCGE1002]. | 94,90% | 1 | 1 | 1 | 4,09% | AANATLNLGANVSLIGSS MAHGSGGILNTSSGAG K |
| CBJ43054 | GH28 | Exo-polygalacturonosidase signal peptide protein [Ralstonia solanacearum CFBP2957]. | 94,90% | 1 | 1 | 1 | 1,87% | GDGATLNTAVIQK |
| ACU73544 | GH31 | glycoside hydrolase family 31 [Catenulispora acidiphila DSM 44928]. | 94,90% | 1 | 1 | 1 | 1,82% | SAEQVVAGGVPVMR |
| ACS59102 | GH32 | Glycosyl hydrolase family 32 domain protein [Rhizobium leguminosarum bv trifolii WSM1325]. | 94,90% | 1 | 1 | 1 | 1,41% | IPSFAAAR |
| ADF57367 | GH32 | beta-fructosidase [Roseburia inulinivorans DSM 16841]. | 94,90% | 1 | 1 | 4 | 1,78% | AENMLEIDR |
| AFI29216 | GH32 | glycoside hydrolase family protein [Bacillus sp JS]. | 94,90% | 1 | 1 | 1 | 2,04% | SGNVTFNPAFDTGK |
| ABJ81207 | GH33 | Exoalphasialidase [Candidatus Solibacter usitatus Ellin6076]. | 94,90% | 1 | 1 | 1 | 3,91% | NSASDTGDIDVLLR |
| AAO78902 | GH36 | possible alpha-galactosidase [Bacteroides thetaiotaomicron VPI5482]. | 94,90% | 1 | 1 | 1 | 2,31% | SADSSCDSILAFNWEK |
| ADO56725 | GH36 | Glycoside hydrolase clan GHD [Paenibacillus polymyxa SC2]. | 94,90% | 1 | 1 | 1 | 1,81% | MENGQLEHLYGK |
| BAJ75041 | GH36 | alpha-galactosidase [Microbacterium testaceum StLB037]. | 94,90% | 1 | 1 | 1 | 2,79% | GLVWGVHLSWNGAV SAER |
| ACP19537 | GH38 | alpha mannosidase [Cadophora finlandica]. | 94,90% | 1 | 1 | 1 | 2,06% | EVIPEGGK |
| AEM82111 | GH38 | glycoside hydrolase family 38 [Streptomyces violaceusniger Tu 4113]. | 97,70% | 1 | 1 | 1 | 0,67% | LEGSVPK |
| AFR92825 | GH38 | alpha-mannosidase [Cryptococcus neoformans var grubii H99]. | 94,90% | 1 | 1 | 1 | 0,60% | IASYDK |
| AFY69254 | GH38 | Alpha-mannosidase [Pseudanabaena sp PCC 7367]. | 94,90% | 1 | 1 | 1 | 0,90% | MNQAIQTQLK |
| BAB59351 | GH38 | alpha-mannosidase [Thermoplasma volcanium GSS1]. | 94,90% | 1 | 1 | 1 | 0,69% | LNNNNIK |
| ADI09723 | GH39 | glycoside hydrolase family 39 [Streptomyces bingchenggensis BCW1]. | 94,90% | 1 | 1 | 1 | 1,28% | NIDPELK |
| ACQ79575 | GH42 | Beta-galactosidase [Beutenbergia cavernae DSM 12333]. | 94,90% | 1 | 1 | 1 | 1,03% | LGDALQR |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|---|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AFL03954 | GH43 | putative endo-1,4-beta-xylanase [Bifidobacterium bifidum BGN4]. | 94,90% | 1 | 1 | 1 | 2,16% | VVISVVAER |
| AAX70137 | GH47 | mannosyl oligosaccharide 12alphamannosidase IB putative [Trypanosoma brucei]. | 94,90% | 1 | 1 | 1 | 1,53% | EIALAINLR |
| AAT47714 | GH5 | mannanase [Alicyclobacillus acidocaldarius]. | 94,90% | 1 | 1 | 1 | 2,23% | IVLSDGGQWEK |
| ACU73181 | GH51 | Alpha-N-arabinofuranosidase [Catenulispora acidiphila DSM 44928]. | 94,90% | 1 | 1 | 1 | 2,98% | LANVEPMMAVNLGTR |
| AA47668 | GH53 | arabinogalactan endo-1,4-beta-galactosidase [Xanthomonas campestris pv campestris str 8004]. | 97,70% | 1 | 1 | 1 | 2,70% | AMLGNLITR |
| ADZ82428 | GH53 | Arabinogalactan endo-1,4-beta-galactosidase [Clostridium lentocellum DSM 5427]. | 94,90% | 1 | 1 | 1 | 0,71% | VLLNGQMK |
| AAZ98013 | GH57 | glycosyl hydrolase family 57 [Thiobacillus denitrificans ATCC 25259]. | 94,90% | 1 | 1 | 1 | 1,92% | SGASEEDIQTK |
| AGG87862 | GH65 | beta-phospho-glucomutase [Rhodanobacter sp 2APBS1]. | 94,90% | 1 | 1 | 1 | 0,69% | LGDTPVR |
| ACA59379 | GH77 | 4-alpha-glucanotransferase [Candidatus Desulforudis audaxviator MP104C]. | 94,90% | 1 | 1 | 1 | 1,16% | HWSGDLDR |
| AEK41123 | GH77 | 4-alpha-glucano transferase [Amycolatopsis mediterranei S699]. | 94,90% | 1 | 1 | 1 | 1,23% | MPEEIPAR |
| CBL01340 | GH77 | 4-alpha-glucanotransferase [Faecalibacterium prausnitzii SL33]. | 94,90% | 1 | 1 | 1 | 1,33% | TLAAQIK |
| AGK94629 | GH78 | Alfa-L-rhamnosidase [Lactobacillus acidophilus La14]. | 94,90% | 1 | 1 | 1 | 0,75% | DFELANK |
| CAB53318 | GH78 | putative hydrolase [Streptomyces coelicolor A32]. | 94,90% | 1 | 1 | 1 | 0,79% | GAAEGTVVR |
| AAD38027 | GH9 | beta 1,4-endoglucanase [Cherax quadricarinatus]. | 94,90% | 1 | 1 | 1 | 1,49% | MTGILLK |
| ADZ85666 | GH94 | cellobiose phosphorylase-like protein [Clostridium lentocellum DSM 5427]. | 94,90% | 1 | 1 | 1 | 1,22% | YLYDEAVGGYK |
| YP_001431979 | GH94 | Glycosyl transferase 36 [Roseiflexus castenholzii strain DSM 13941 HLO8]. | 94,90% | 1 | 1 | 2 | 0,31% | LTLESGEER |
| CAB56146 | GH95 | putative large secreted protein [Streptomyces coelicolor A32]. | 94,90% | 1 | 1 | 1 | 1,24% | FTGDTGALSR |
| ADK96325 | GH97 | conserved hypothetical protein [Prevotella melaninogenica ATCC 25845]. | 94,90% | 1 | 1 | 1 | 1,49% | TVIPMAEIGK |
| AFR48754 | GT1 | Glycosyl transferase probable UDP glucuronosyltransferase [Gordonia sp KTR9]. | 94,90% | 1 | 1 | 1 | 1,74% | VVVSGLR |
| CAH04970 | GT10 | alpha-3-fucosyltransferase [Ciona savignyi]. | 94,90% | 1 | 1 | 1 | 1,58% | MIQLIK |
| CAJ42067 | GT10 | alpha-1,3-fucosyltransferase [Drosophila yakuba]. | 94,90% | 1 | 1 | 1 | 1,35% | IDSELR |
| AAW60043 | GT19 | Lipid-A-disaccharide synthase [Gluconobacter oxydans 621H]. | 94,90% | 1 | 1 | 1 | 2,41% | VNPVTATIAR |
| ABQ04459 | GT19 | Candidate lipidAdisaccharide synthase Glycosyltransferase family 19 [Flavobacterium johnsoniae UW101]. | 94,90% | 1 | 1 | 1 | 2,70% | GSAISYQIAK |
| ADI34954 | GT19 | lipid-A-disaccharide synthase [Helicobacter pylori v225d]. | 94,90% | 1 | 1 | 1 | 2,22% | FTGVFEGK |
| CAK23661 | GT20 | Glycosyl transferase of family GT20 probable trehalosephosphate synthase [Synechococcus sp WH 7803]. | 99,90% | 1 | 1 | 1 | 1,21% | IDEDTK |
| AAB05993 | GT24 | UDPGlc:Glycoprotein Glucosyltransferase [Schizosaccharomyces pombe]. | 94,90% | 1 | 1 | 2 | 0,69% | DVDIDLFIK |
| ABB32573 | GT30 | CMP3deoxyDmannooctulosonatelipid A tetraacyldisaccharide 3deoxyDmannooctulosonate transferase [Geobacter metallireducens GS15]. | 94,90% | 1 | 1 | 1 | 3,21% | GGVLLVDITIGELMK |

| <u>Nº de Acesso</u> | <u>Familia CAZY</u> | <u>Nome da Proteína [organismo]</u> | <u>ID %</u> | <u>Nº UP</u> | <u>Nº US</u> | <u>Nº TS</u> | <u>Seq. coverage (%)</u> | <u>Peptide sequence</u> |
|---------------------|---------------------|--|-------------|--------------|--------------|--------------|--------------------------|-------------------------|
| AEB75065 | GT35 | glycogen starch alpha-glucan phosphorylase [Clostridium botulinum BKT015925]. | 94,90% | 1 | 1 | 1 | 0,86% | NGGYNAR |
| CAY80859 | GT39 | Pmt4p [Saccharomyces cerevisiae EC1118]. | 94,90% | 1 | 1 | 1 | 0,79% | ISDLEK |
| AAZ69224 | GT4 | glucosyl transferase [Methanosarcina barkeri str Fusaro]. | 94,90% | 1 | 1 | 1 | 2,75% | DMAESSGVEK |
| BAF11040 | GT47 | Os03g0174300 [Oryza sativa Japonica Group]. | 94,90% | 1 | 1 | 1 | 1,91% | LAAEFGSTPDK |
| BAB89139 | GT48 | OJ1029_F044 [Oryza sativa japonica cultivargroup]. | 94,90% | 1 | 1 | 1 | 0,37% | VLANTLK |
| ADZ90369 | GT5 | Glycogen synthase [Marinomonas mediterranea MMB1]. | 94,90% | 1 | 1 | 1 | 2,52% | MIMPAYNSILDK |
| AGH38509 | GT5 | Glycogen synthase [Bibersteinia trehalosi USDAARSUSMARC192]. | 94,90% | 1 | 1 | 1 | 1,19% | ETIENR |
| AAS40096 | GT51 | penicillinbinding protein 1 [Bacillus cereus ATCC 10987]. | 99,40% | 1 | 1 | 1 | 1,88% | EEGAQGAFLLMDNR |
| ACL42333 | GT51 | glycosyl transferase family 51 [Arthrobacter chlorophenicus A6]. | 94,90% | 1 | 1 | 1 | 1,04% | LGITSLPR |
| CAM60289 | GT53 | Probable arabinosyltransferase C [Mycobacterium abscessus]. | 94,90% | 1 | 1 | 1 | 1,29% | DQTLVGEIQANALK |
| CBZ23085 | GT60 | conserved hypothetical protein [Leishmania mexicana MHOMGT2001U1103]. | 94,90% | 1 | 1 | 1 | 1,34% | GFVGVSSAATER |
| AAB05994 | GT66 | putative transmembrane protein precursor [Homo sapiens]. | 99,70% | 1 | 1 | 1 | 1,70% | ENDYYTPTGEFR |
| CAP95716 | GT71 | Pc21g08190 [Penicillium chrysogenum Wisconsin 541255]. | 94,90% | 1 | 1 | 1 | 1,61% | MSALRPGR |
| CCH04789 | GT9 | hypothetical protein NH44784_008081 [Achromobacter xylooxidans NH447841996]. | 94,90% | 1 | 1 | 1 | 2,52% | VAILMAPR |
| Q7UQX3 | not found! | Putative uncharacterized protein [Rhodopirellula baltica strain SH1]. | 94,90% | 1 | 1 | 1 | 0,82% | GEVLSFDQR |
| XP_002500607 | not found! | Inward rectifier K+ channel family transporter [Micromonas sp strain RCC299 NOUM17]. | 94,90% | 1 | 1 | 1 | 2,10% | SGMGAGHSDGSAAAMK |
| ADZ77057 | PL22 | hypothetical protein Sph21_0475 [Sphingobacterium sp 21]. | 94,90% | 1 | 1 | 1 | 2,14% | IQGEIVHAK |