

UNIVERSIDADE FEDERAL DO PARANÁ

MARCELO SCARDUELLI

PROSPECÇÃO DE GENES DE ENZIMAS HIDROLÍTICAS DE BACTÉRIAS
ISOLADAS DE MANGUEZAL PARANAENSE

CURITIBA

2018

MARCELO SCARDUELLI

PROSPECÇÃO DE GENES DE ENZIMAS HIDROLÍTICAS DE BACTÉRIAS
ISOLADAS DE MANGUEZAL PARANAENSE

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Ciências – Bioquímica, pelo programa de pós-graduação em Ciências – Bioquímica do departamento de Bioquímica e Biologia Molecular, do setor de Ciências Biológicas da Universidade Federal do Paraná.

Orientador: Prof. Dr. Emanuel Maltempi de Souza

Co-orientador: Prof. Dr. Luciano Fernandes Huergo

CURITIBA

2018



MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
(BIOQUÍMICA)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS (BIOQUÍMICA) da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **MARCELO SCARDUELLI** intitulada: **PROSPECÇÃO DE GENES DE ENZIMAS HIDROLÍTICAS DE BACTÉRIAS ISOLADAS DE MANGUEZAL PARANAENSE**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 26 de Junho de 2018.

EMANUEL MALTEMPI DE SOUZA
Presidente da Banca Examinadora

ROSELI WASSEM
Avaliador Externo

LEDA SATIE CHUBATSU
Avaliador Interno

CAROLINA WEIGERT GALVAO
Avaliador Externo

NADIA KRIEGER
Avaliador Interno

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Telma Terezinha Stresser de Assis –CRB/9-944)

Scarduelli, Marcelo

Prospecção de genes de enzimas hidrolíticas de bactérias isoladas de manguezal paranaense. / Marcelo Scarduelli. – Curitiba, 2018.
66 p.: il. ; 30cm.

Orientador: Emanuel Maltempi de Souza

Co-orientador: Luciano Fernandes Huergo

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Ciências - Bioquímica.

1. Manguezais. 2. Enzimas. 3. Microorganismos. I. Título. II. Souza, Emanuel Maltempi de. III. Huergo, Luciano Fernandes. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Ciências - Bioquímica.

CDD (20. ed.) 574.1925

AGRADECIMENTOS

Agradeço primeiro a Deus, pelas oportunidades e imensa misericórdia.

Agradeço com sinceridade aos Professores Emanuel M. de Souza e Luciano F. Huergo pela orientação e ensinamentos, estando sempre dispostos a me ajudar, já há quase 7 anos.

A todos os professores e demais integrantes do Departamento de Bioquímica e Biologia Molecular pelas discussões, dicas e convívio.

À minha esposa, Cleidi Tokarski Scarduelli, pelo carinho, amor, apoio, compreensão e ensinamentos, juntamente com meu filho Vini. Amo muito vocês.

À minha família, em especial aos meus pais Valdir e Graça Scarduelli por todo carinho, educação e fé depositada em mim, e aos meus irmãos Bruno e Daniel Scarduelli, pelas horas de descontração e apoio. A distância pode ser longa, mas nossos corações estão sempre pertos. Amo muito vocês.

Aos amigos dos laboratórios 271 e 272 pelo convívio motivador e animado, em especial para Edileusa Cristina Marques Gerhardt, Rafaela Perez, Heloísa Bruna Sanchuki Soligo, Fernanda Gravina, Alysson Henrique Urbanski, Ana Goerdt, Gillize Araujo, Thiago Estefano Rodrigues, Esther Esteves, Heloísa Soligo Sanchuki, Rocio DelPilar Cuaspa Ropaín, Paloma Bonatto, Leilane Caline e Vanessa Kessler Chicora.

Aos técnicos Valter de Baura, Roseli Prado e Bruna pelo auxílio oferecido constantemente.

À coordenação do curso de pós-graduação em Ciências – Bioquímica, e ao suporte financeiro da CAPES, CNPq, INCT e Fundação Araucária.

“E conhecereis a verdade, e a verdade vos libertará.”
João 8:32.

RESUMO

O manguezal é um bioma costeiro que apresenta fauna e flora adaptadas a inundações periódicas, causando grandes variações de salinidade. Dentre os estudos sobre manguezais no mundo, aproximadamente 20% estão relacionados com manguezais brasileiros, e destes apenas 1% apresentam ensaios relacionados a identificação ou caracterização de enzimas hidrolíticas presentes nesses ambientes. Assim, este projeto de doutorado teve como objetivo identificar e caracterizar enzimas com potencial aplicação biotecnológica a partir de genes presentes em microrganismos do manguezal da baía de Paranaguá, do estado do Paraná. Amostras desse bioma foram coletadas e semeadas em meios seletivos específicos, permitindo o isolamento de 17 bactérias. Dentre elas, o isolado XL-2 teve seu genoma totalmente sequenciado e foi classificado como *Bacillus safensis* BRM1. Seu genoma apresenta 3,74 Mb e conteúdo GC de 41,8%. Dentre os 3.983 genes preditos, 23 podem estar relacionados com genes que codificam para celulasas e hemicelulasas, evidenciando o potencial biotecnológico presente desta estirpe. Além disso, outro isolado (XS-14, *Pseudomonas sp.*) teve seu genoma parcialmente sequenciado, onde foi possível identificar uma esterase (Est1) com atividade em substratos de cadeia curta e em níveis elevados de NaCl, sendo parcialmente caracterizada. Dessa forma, o estudo realizado nesse período de doutorado permitiu contribuir para o conhecimento sobre os manguezais brasileiros.

Palavras-chave: *manguezal, Bacillus safensis* BRM1, *enzimas*

ABSTRACT

Mangrove is a coastal biome with fauna and flora adapted to periodic flooding, causing great variations of salinity. Among the mangrove studies in the world, approximately 20% are related to Brazilian mangroves, and only 1% of these have been analyzed for identification or characterization of hydrolytic enzymes. Thus, this project had the objective of identifying and characterizing enzymes with potential biotechnological application from genes microorganisms of the Paranaguá bay mangrove, in the state of Paraná. Samples of the soil from this biome were collected and microorganisms were seeded into specific selective media, allowing the isolation of 17 bacteria. Among them, the XL-2 isolate had its genome fully sequenced and, after deposition in GenBank, was classified as *Bacillus safensis* BRM1. Its genome has 3.74 Mb and 41.8% of GC content. Among the 3,983 predicted genes, 23 may be related to genes encoding cellulases and hemicellulases, evidencing the biotechnological potential present in this strain. Another isolate (XS-14, *Pseudomonas* sp.) had its genome partially sequenced, where it was possible to identify an esterase (Est1) with activity on short-chain substrates and high levels of NaCl, being partially characterized. Thus, the study conducted in this doctoral period contributed to the knowledge on Brazilian mangroves.

Key-words: *mangrove*, *Bacillus safensis* BRM1, *enzymes*.

LISTA DE FIGURAS

Figura 1 – Foto ilustrativa do manguezal.....	16
Figura 2 – Distribuição de manguezais e respectiva diversidade de espécies vegetais	17
Figura 3 – Número de artigos publicados sobre manguezal segundo o Pubmed	18
Figura 4 – Localização das coletas de material de manguezal paranaense	54
Figura 5 – Exemplos de halos de degradação formados por isolados.....	56

CAPÍTULO I

Figure 1. Map of the <i>Bacillus safensis</i> BRM1 genome	30
Figure 2. Heatmap showing Average Nucleotide Identity between the 12 <i>Bacillus</i> strain genomes deposited in NCBI database	31
Figure 3. COG classification from <i>B. safensis</i> BRM1 genes	32
Figure 4. Halo hydrolysis formed after 4 days at 30°C of cultivation.	34
Figure 5. <i>Bacillus safensis</i> BRM1 growth in CSE medium.....	35

CAPÍTULO II

Figure 1. Est1 amino acids sequence	46
Figure 2. Phylogenetic tree of aligned sequences from different lipases/esterases families.....	48
Figure 3. Three conserved motifs identified into Est1 protein sequence	49
Figure 4. Phylogenetic tree of aligned protein sequences obtained after BLASTp assay using Est1 sequence as query.	50

LISTA DE TABELAS

Tabela 1 – Artigos relacionados a pesquisas sobre manguezais brasileiros	18
Tabela 2 – Enzimas e suas utilizações na indústria	22
Tabela 3 – composição dos meios seletivos	55
Tabela 4 – Quantidade de isolados apresentando halos de degradação	57
Tabela 5 – Identificação de isolados baseado na identidade da sequência de genes de RNA ribossômico 16S	57
Tabela 6 – Enzimas selecionadas para estudo	60

CAPÍTULO 1

Table 1. Genome features of <i>Bacillus safensis</i> BRM1.....	30
Table 2. Genes for putative cellulases and hemicellulases in <i>Bacillus safensis</i> BRM1 genome.....	33
Table S1. Raw data from ANI matrix analysis among the <i>Bacillus</i>	39
Table S2. Clusters of orthologous group analysis for <i>Bacillus safensis</i> BRM1.	40
Table S3. Putative cellulases and hemicellulases in <i>Bacillus safensis</i> BRM1.....	41

CAPÍTULO 2

Table S1. Sequence alignment between Est1 and others similar proteins.....	52
--	----

LISTA DE SIGLAS

ANI	- <i>Average Nucleotide Identity</i>
CAZY	- <i>Carbohydrate-Active enZYmes database</i>
CMC	- carboximetilcelulose
DNS	- ácido dinitrosalicílico
IUBMB	- <i>International Union of Biochemistry and Molecular Biology</i>
MMA	- Ministério do Meio Ambiente

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVO	15
2.1	OBJETIVOS ESPECÍFICOS.....	15
3	REVISÃO BIBLIOGRÁFICA	16
3.1	MANGUEZAL: UM AMBIENTE ESPECIAL	16
3.2	ENZIMAS UTILIZADAS EM PROCESSOS INDUSTRIAIS	21
3.3	ENZIMAS OBTIDAS DE AMOSTRAS DE MANGUEZAL NÃO BRASILEIRO .24	
4	CAPÍTULO I	26
5	CAPÍTULO II	42
6	ATIVIDADES REALIZADAS	54
6.1	ISOLAMENTO DE MICRORGANISMOS.....	54
6.2	MEIOS E CONDIÇÕES DE CULTIVO	55
6.3	OBTENÇÃO DE ISOLADOS	55
6.4	SEQUENCIAMENTO GENÔMICO	58
6.5	SELEÇÃO DE ENZIMAS.....	59
7	CONCLUSÃO	61
	REFERÊNCIAS	62

1 INTRODUÇÃO

Para a manutenção da vida de seres vivos, é definido duas condições fundamentais: o ser vivo deve ser capaz de gerar descendentes e deve conseguir catalisar reações químicas de maneira eficiente e seletiva, ou seja, essas reações precisam ocorrer na velocidade necessitada pelo organismo (em geral, muito superior daquela que ocorreria naturalmente) e agir em substratos específicos para cada situação metabólica. As biomoléculas responsáveis pela segunda condição são conhecidas como ribozimas e enzimas (NELSON; COX, 2002).

De caráter proteico (exceto para as ribozimas ou RNAs catalíticos), as enzimas participam da grande maioria das reações presentes nos organismos, apresentando então diversas funções: transferência de elétrons ou grupos químicos, quebra ou formação de ligações químicas, transporte de compostos. Além disso, características especiais também podem ser encontradas de acordo com o habitat do organismo de estudo: resistência a altas temperaturas ou pH, resistência a solventes ou elevada pressão osmótica por exemplo.

Por serem capazes de aumentar a velocidade de reações químicas e de maneira específica, as enzimas podem ser utilizadas também para catalisar reações químicas industriais. Com o seu uso, inúmeros processos podem ser realizados de maneira mais rápida e sob condições mais brandas (temperaturas e pressão menores) quando comparados com processos químicos tradicionais. Devido à especificidade das enzimas, produtos indesejáveis, formados em reações secundárias nos processos não enzimáticos, puderam ser reduzidos ou até mesmo eliminados. Isso proporcionou um aumento no rendimento e facilitou posteriores etapas de purificação do produto desejado, acarretando em diminuição de custos de produção e de impacto ambiental. Esses benefícios podem ser comprovados pelo crescimento das vendas de enzimas industriais pelo mundo. Em 2014 e 2015, o mercado de enzimas movimentou US\$ 4,6 e 4,9 bilhões, respectivamente. Estima-se que em 2021 esse valor alcance US\$ 6,3 bilhões (CHARNOCK; MCCLEARY, 2006; BCC RESEARCH, 2017).

Os microrganismos são a principal fonte genética e produtora de enzimas em escala industrial. Bactérias, fungos e leveduras apresentam enzimas versáteis com diversas estruturas e funções, sendo aplicáveis em diversos processos industriais. As principais enzimas microbianas utilizadas atualmente são: amilases, proteases,

pectinases, lipases, xilanases, celulasas e lacases. Quando comparadas com as de origem animal ou de planta, as enzimas microbianas apresentam diversas vantagens, como possuírem maiores atividade e estabilidade, principalmente em condições de fermentação. Além disso, apresentam maior rendimento em sua produção e facilidade na modificação e otimização através de manipulação genética e técnicas moleculares (engenharia proteica) (LIU; KOKARE, 2017; ILLANES et al., 2012; GURUNG et al., 2013).

A descoberta de novas enzimas microbianas pode ser feita através do uso de técnicas como a metagenômica, *genome mining* e isolamento de microrganismos em meios específicos. A primeira técnica é descrita como análises funcional e de sequência de genomas microbianos obtidos em uma amostra ambiental, na qual o termo metagenoma engloba todos os genomas ali presentes. De forma resumida, essa técnica consiste em purificar e clonar todo o DNA presente em uma amostra, seguido de sequenciamento e análise das sequências e/ou prospecção por clones funcionais. Já a técnica de *genome mining* consiste em identificar bioativos/metabólitos secundários em dados genéticos, normalmente obtidos por métodos de metagenômica. Essa busca é realizada através do uso de ferramentas de bioinformática. Por fim, a técnica de isolamento consiste em coletar amostras de ambientes de interesse, cujos microrganismos ali presentes são semeados em meios de cultivo específicos. A composição desses meios pode ser formulados apresentando uma pressão seletiva, ou seja, podem conter substratos ou condições que favoreçam o crescimento de microrganismos com características especiais (HANDELSMAN et al., 1998; ALETI; SESSITSCH; BRADER, 2015).

Visando encontrar microrganismos de interesse, a exploração da diversidade microbiana em ambientes extremos tem atraído a atenção de pesquisadores, devido a capacidade destes microrganismos em sobreviver em condições extremas, como temperaturas (de -2 a 110 °C), pressão, radiação, salinidade (2 a 5M NaCl) e pH (abaixo de 2 ou acima de 9). Microrganismos extremófilos apresentam grande interesse na busca de enzimas capazes de suportar as condições ambientais adversas presentes em seu habitat de origem (ADRIO; DEMAIN, 2014).

Assim, o presente projeto objetivou o isolamento de microrganismos de amostras de manguezal paranaense, com a finalidade de encontrar enzimas hidrolíticas que apresentem características relevantes para um possível uso em escala industrial.

2 OBJETIVO

Realizar a prospecção de genes que codifiquem enzimas com potencial aplicação biotecnológica a partir do sequenciamento de genoma de microrganismos do manguezal da Baía de Paranaguá, do estado do Paraná.

2.1 OBJETIVOS ESPECÍFICOS

- Selecionar, através de cultivo em meios seletivos, microrganismos presentes em amostra de manguezal;
- Classificar taxonomicamente os isolados encontrados através de sequenciamento do gene que codifica o RNA ribossômico 16S;
- Sequenciar o genoma de microrganismos que apresentarem atividade hidrolítica em diferentes substratos ou meios de cultivo específicos;
- Buscar no genoma obtido genes que codificam enzimas com potencial interesse industrial;
- Clonar, expressar, purificar e caracterizar as enzimas selecionadas.

3 REVISÃO BIBLIOGRÁFICA

3.1 MANGUEZAL: UM AMBIENTE ESPECIAL

O manguezal (Figura 1) é um ecossistema costeiro que apresenta fauna e flora adaptadas a inundações periódicas causadas pelas marés. Essas mudanças levam a grandes variações de salinidade, alcançando níveis de até 44 g/L (750 mM) de cloreto de sódio (MACIEL, 1991; DIELE; SMITH, 2006), sendo considerado um ambiente extremo. Presente em regiões tropicais, os manguezais possuem condições propícias para alimentação, proteção e reprodução de diversas espécies animais. Segundo o Ministério do Meio Ambiente (MMA), o Brasil apresenta cerca de 12 mil km² de manguezal em quase todo o litoral brasileiro, indo desde o Oiapoque, no Amapá, até a Laguna em Santa Catarina (SCHAEFFER-NOVELLI, 1995; DIEGUES, 2002).

Figura 1 – Foto ilustrativa do manguezal

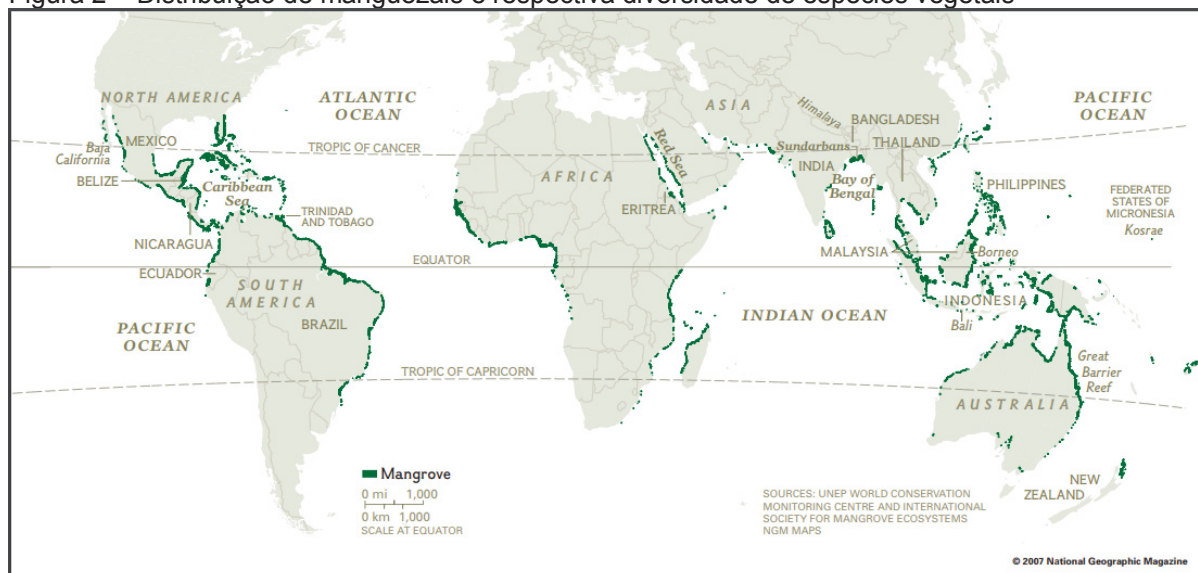


FONTE: Athila Bertoncini – MMA, BRASIL.

Esse ecossistema também é encontrado em diversas partes do mundo (Figura 2), atingindo em torno de 150 mil km² (15 milhões de hectares). Os manguezais estão localizados principalmente entre os trópicos e subtropicais terrestres, tendo seu desenvolvimento mais elevado entre as latitudes 25°N e 25°S. Como as espécies vegetais presentes nos manguezais não estão adaptadas a condições muito frias, sua distribuição latitudinal fica restringida, tendo baixa diversidade. Estima-se que haja aproximadamente 16 famílias e 40 a 50 espécies. Porém, em se tratando de espécies microbianas, esse ecossistema apresenta

elevada biodiversidade, das quais apenas 5% das espécies estão descritas (BLANQUET et al., 1996, THATOI et al., 2013).

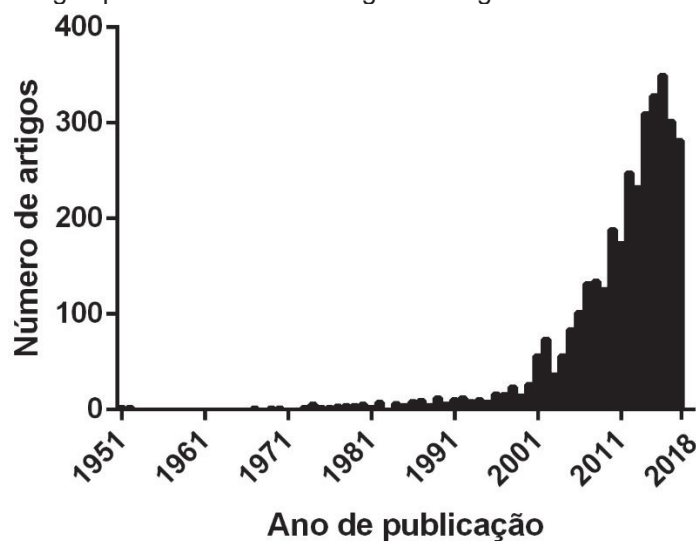
Figura 2 – Distribuição de manguezais e respectiva diversidade de espécies vegetais



FONTE: (NATIONAL GEOGRAPHIC MAGAZINE, 2007).

Com o auxílio da plataforma de pesquisa PubMed/NCBI (*National Center for Biotechnology Information*, EUA), utilizando o termo “*mangrove*” na pesquisa por artigos científicos indexados, foram encontrados 3.435 artigos no período de 1951 a 2018 (Figura 3). A partir dos resultados, nota-se que houve um rápido crescimento nos estudos sobre esse ambiente a partir de 2005. Esse aumento pode estar relacionado com o desenvolvimento de plataformas de sequenciamento de DNA em larga escala (*high-throughput*), o que permitiu o desenvolvimento da metagenômica. Inúmeras amostras ambientais passaram a ser coletadas e tiveram seus materiais genéticos purificados e depois sequenciados nos novos sequenciadores de DNA da época: 454 *Sequencer*, da Life Sciences, e SOLiD, da Agencourt (HUTCHISON, 2007).

Figura 3 – Número de artigos publicados sobre manguezal segundo o Pubmed



FONTE: o autor (2018).

Dentre esses artigos, os que estão relacionados com manguezais brasileiros somam 422 trabalhos (de 1979 a 2018), relacionados a estudos ambientais, diversidade biológica, análises de metais pesados, bioindicadores ou impactos antropológicos. Desses, apenas 14 apresentam alguma análise relacionada a enzimas com potencial biotecnológico até a presente data (Tabela 1).

Tabela 1 – Artigos relacionados a pesquisas sobre manguezais brasileiros

continua

N.	LOCAL DO MANGUEZAL	TIPO DE ESTUDO	ENZIMAS ESTUDADAS	REFERÊNCIA
1	Pontal do Paraná (Paraná)	Construção de biblioteca metagenômica	Lipase	COUTO et al. (2010)
2	Cananéia e Bertioga (São Paulo), com contaminação de petróleo	Construção de biblioteca metagenômica e análise de mapas metabólicos.	<i>Sem busca por enzimas</i>	ANDREOTE et al. (2012)
3	Arquipélago de Abrolhos (Bahia)	Construção de biblioteca metagenômica e análises de biodiversidade e de mapas metabólicos.	<i>Sem busca por enzimas</i>	BRUCE et al. (2012)
4	Bertioga (São Paulo), com contaminação de petróleo	<i>Draft</i> de genoma de um <i>Bacillus pumilus</i> (estirpe CCMA-560). Presença de biosurfactante	<i>Sem busca por enzimas</i>	DOMINGOS et al. (2013)

Tabela 2 – Artigos relacionados a pesquisas sobre manguezais brasileiros

N.	LOCAL DO MANGUEZAL	TIPO DE ESTUDO	ENZIMAS ESTUDADAS	conclusão
				REFERÊNCIA
5	Ilha Grande (Rio de Janeiro) e em Porto Seguro (Bahia)	Isolamento de microrganismos em meio de cultura seletivo e construção de biblioteca metagenômica	Celulases	THOMPSON et al. (2013)
6	Cananéia e Bertioga (São Paulo)	Isolamento de microrganismos em meio de cultura e posterior análise de atividade enzimática	Amilase, esterase, lipase, protease e endoglucanase	CASTRO et al. (2014)
7	Ilha do Cardoso (São Paulo)	Isolamento de microrganismos em meio de cultura e posterior análise de atividade enzimática	Endoglucanases e exoglucanases	JÚNIOR et al. (2014)
8	Cananéia e Bertioga (São Paulo)	Isolamento de microrganismos em meio de cultura e posterior análise de atividade enzimática	Endoglucanases	SÁ et al. (2014)
9	Cananéia e Bertioga (São Paulo)	Análise metagenômica de amostras de solo e identificação de enzimas por similaridade de sequência	Lipases, esterases, epóxido-hidrolases (EHs), desalogenase e hidroxinitrila liases	JIMÉNEZ et al. (2015)
10	Região de São Sebastião (São Paulo)	Isolamento de fungos presentes em esponjas e algas marinhas	Xilanase	DOS SANTOS et al. (2016)
11	Cananéia e Bertioga (São Paulo), com contaminação de petróleo	Construção de biblioteca metagenômica e análise de mapas metabólicos.	Dioxigenases que hidroxilam anéis aromáticos (ARHD, do inglês <i>Aromatic-ring-hydroxylating dioxygenases</i>).	DE SOUSA et al. (2017)
12	Bertioga (São Paulo)	Construção de biblioteca metagenômica de amostra de solo e prospecção funcional de clones	Identificação de enzima com atividade β -N-acetilglucosaminidase, β -glucosidase e β -1,4-endoglucanase	SOARES et al. (2017)
13	Coruripe (Alagoas)	Análise de atividade de colinesterase presente em ostras de manguezal (potencial uso como biomarcador)	Colinesterases	DE SOUZA et al. (2018)
14	São Paulo	Seleção de fungos endofíticos isolados de plantas de manguezal para produção de enzimas hidrolíticas	Celulases e hemicelulases	MAROLDI et al. (2018)

FONTE: o autor (2018).

Segundo a pesquisa realizada anteriormente, o trabalho de COUTO e colaboradores (2010) foi um dos primeiros a realizar estudos sobre o manguezal brasileiro. A partir de uma amostra do litoral do Paraná (Pontal do Paraná), esses autores construíram uma biblioteca metagenômica e, a partir de *screening* funcional, encontraram um gene que codificava para uma esterase. Após ensaios de atividade, os pesquisadores encontraram que essa enzima alcançou atividade de até 0.87 U/mg de proteína em pH 8,0 e 30°C (substrato utilizado p-nitrofenil-decanoato, C10).

Já THOMPSON e colaboradores (2013) coletaram amostras de manguezal da cidade de Ilha Grande (Rio de Janeiro) e de Porto Seguro (Bahia). Depois de purificado todo o material genético ali presente, os autores construíram duas bibliotecas metagenômicas. Em paralelo, parte das amostras foram inoculadas em meio seletivo contendo bagaço de cana como principal fonte de carbono, selecionando possíveis microrganismos produtores de celulases e hemicelulases. Como resultado, houve o isolamento de diversas bactérias produtoras de enzimas (no cultivo seletivo) e a identificação de diversas enzimas (nas bibliotecas metagenômicas), mas sem nenhuma quantificação de atividade enzimática.

Os estudos de CASTRO e colaboradores (2014) mostraram o isolamento de microrganismos endofíticos de plantas presentes nos manguezais das cidades de Cananéia e Bertioga (São Paulo). Vários isolados apresentaram halos de degradação quando cultivados em meios seletivos. Não houve a purificação de enzimas neste trabalho e as atividades enzimáticas para amilase, esterase, lipase, protease e endoglucanase foram quantificadas apenas por medição de halos de degradação. A mesma estratégia de quantificação de atividade foi realizada no trabalho de JÚNIOR e colaboradores (2014) e MAROLDI e colaboradores (2018).

Para DOS SANTOS e colaboradores (2016), o foco de estudo foi em fungos produtores de xilanases. Foram coletadas espécies diferentes de algas marinhas, sendo uma delas presente em manguezal do estado de São Paulo. Os autores quantificaram a atividade de xilanases extracelulares presente no meio de cultivo utilizando o método de DNS (ácido dinitrosalicílico), de quantificação de açúcar redutor (MILLER, 1959). O valor máximo de atividade encontrada foi de 711,43 U/mL de proteína.

Por fim, SOARES et al. (2017), através de construção de biblioteca metagenômica a partir de amostra de manguezal de Bertioga (São Paulo), identificou uma enzima multifuncional denominada β -N-acetilhexosaminidase que pode

hidrolisar celulose e quitina. Em ensaios de atividade enzimática, os pesquisadores alcançaram atividades de até 74 mmol/min.mL pH 4,5 e 30°C, utilizando substratos sintéticos.

Os outros trabalhos mostrados na Tabela 1 não realizaram ensaios de atividade enzimática ou mesmo purificação de enzimas. Assim, nota-se que ainda há um grande potencial em estudar outras áreas de manguezais brasileiros na busca por enzimas com aplicabilidade industrial.

3.2 ENZIMAS UTILIZADAS EM PROCESSOS INDUSTRIAIS

O mercado de enzimas industriais é dominado por duas multinacionais responsáveis por produzir, em 2012, em torno de 75% de todas as enzimas utilizadas comercialmente no mundo (LI et al., 2012). A líder há vários anos é a dinamarquesa Novozymes, tendo apenas ela uma fatia de 48% das vendas mundiais nos anos de 2014, 2015 e 2016. Em seguida vem a norte-americana DuPont, com aproximadamente 19% (NOVOZYMES, 2014, 2015, 2016).

As enzimas são utilizadas em vários segmentos industriais, como nutrição animal, indústria de alimentos e bebidas, produção de cosméticos e medicamentos, além de estarem presentes em produtos para laboratórios de pesquisa (*Research and Development tools*). Dentre as quase 4000 enzimas conhecidas, aproximadamente 200 tipos, que apresentam origem microbiana, são utilizados comercialmente (LI et al., 2012). Alguns exemplos estão mostrados na Tabela 2 abaixo.

Tabela 2 – Enzimas e suas utilizações na indústria

continua

ÁREAS		ENZIMA	BENEFÍCIOS TÉCNICOS
Aplicações técnicas	Indústria de papel e polpa celulósica	Amilases	Clivagem de moléculas de amido para reduzir a viscosidade em etapas de pigmentação
		Lipases	Remoção de tinta e resinas em processos de obtenção de polpa celulósica
		Celulases	Promove maciez e flexibilidade em fibras de celulose
		Mananases	Degradação de glucomananas residuais para promover brancura no papel
		Lacases	Branqueamento da polpa celulósica
		β -xilanasas	Aumento da eficiência de processos de branqueamento da polpa celulósica
Aplicações técnicas	Indústria têxtil	Amilases	Redimensionamento de tecidos sem efeitos prejudiciais
		Celulases	Remoção de “bolinhas” e microfibrilas para dar uma aparência mais suave e brilhante em tecidos; remoção de tintura em jeans para dar aspecto desgastado
		Pectinases	Desestabilização da parede extracelular vegetal, aumentando a extração das fibras celulósicas no algodão
		Lacases, glicoseoxidases	Agentes de limpeza em processos de branqueamento
Aplicações técnicas	Lava-roupas	Proteases	Hidrólise de manchas de carácter proteico em tecidos
		Lipases	Decomposição de manchas de óleos e gorduras em tecidos, principalmente em colarinhos e punhos
		Amilases	Remoção de resíduos de amido, que podem englobar sujidades (ex. “encardido”)
		Celulases	Modificação da estrutura da fibra de celulose para facilitar tingimento e garantir suavidade no algodão
Alimentos	Indústria de laticínios	Quimosina, lipases, lisozimas	Produção de queijos
		β -galactosidades, lactases	Hidrólise da lactose do leite, em glucose e galactose, em produtos para intolerantes à lactose

Tabela 2 – Enzimas e suas utilizações na indústria

		conclusão	
Alimentos	Panificação	<p>α-amilases</p> <p>β-xilanases</p> <p>Oxidoredutases</p> <p>Lipases</p> <p>Proteases</p>	<p>Degradação do amido na farinha, aumentando a disponibilidade de glucose para o fermento, e controle de volume do pão</p> <p>Estabilidade da massa</p> <p>Aumento da força do glúten, permitindo massas com hidratação maior, mais resistentes e que tendem a crescer mais</p> <p>Aumento da estabilidade das moléculas de gás dentro da massa</p> <p>Redução de proteínas em farinhas</p>
	Indústria de sucos	<p>Amilases, glucoamilases</p> <p>Pectinases</p> <p>Celulases, hemicelulases</p> <p>Naringinase e limoninase</p>	<p>Quebra do amido em glucose, aumentando a doçura. Atua também na clarificação de sucos, principalmente no de maçã</p> <p>Degradação de pectinas, proporcionando aumento da extração de sucos</p> <p>Diminuição de viscosidade</p> <p>Degradam componentes que causam amargor em sucos cítricos</p>
Cosméticos	Oxidases, peroxidases, polifenol oxidases		Tingimento de cabelos
	Dissulfeto isomerases, transglutaminases		Modelagem de cabelos
	Papaína, bromelaína, subtilisina		Proporciona efeito <i>peeling</i> na pele
	Amiloglucosidases, glicose oxidases		Presente em formulações de cremes e enxaguantes bucais

FONTE: Modificado de LI et al., (2012).

Diante de diferentes funções e tipos de enzimas, a partir de 1956 foi estabelecida uma forma de classificá-las de acordo com suas funções e substratos que atuam, denominado número EC (do inglês, *Enzyme Commission number*). Essa classificação, criada pelo IUBMB (do inglês, *International Union of Biochemistry and Molecular Biology*), apresenta 4 números separados por pontos (EC W.X.Y.Z), onde W, X e Y referem-se, respectivamente, à classe (funções), subclasse e sub-sub-classe (relativas a seu mecanismo de reação), e o número Z é específico para cada enzima. O banco de dados *Enzyme*, do Instituto Suíço de Bioinformática (do inglês, *SIB*), armazena a nomenclatura de todas as enzimas já estudadas até o momento,

seguindo as recomendações descritas pelo IUBMB. A última atualização realizada, datada no dia 20 de dezembro de 2017, apresentou 6115 classificações diferentes.

Porém estudos revelaram que essa classificação pode não ser a adequada para distinguir certas enzimas, principalmente glicosil-hidrolases. HENRISSAT (1991) percebeu que esse tipo de classificação não evidenciava propriedades estruturais específicas de certas enzimas, deixando de distinguir enzimas que sofreram evolução divergente (mudança na especificidade enzima-substrato) ou convergente (diferenças estruturais catalisando a mesma reação no mesmo substrato). Além disso, o mesmo autor notou que enzimas que degradassem vários tipos de substratos seriam mais difíceis de serem classificadas. Dessa forma, o autor criou a classificação das glicosil-hidrolases por famílias, baseada na similaridade de sequência de aminoácidos e, conseqüentemente, na similaridade de suas estruturas proteicas e possíveis funções.

Seguindo essa nomenclatura, o banco de dados CAZY (do inglês, *Carbohydrate-Active enZymes database*) apresenta uma ótima fonte de informação sobre as variadas famílias de glicosil-hidrolases. Até o momento, a classificação vai da família 1 (enzimas com estrutura de barril (β/α)₈) até a família 148 (atividade β -1,3-glucanase em glucanas β -1,3-1,4).

3.3 ENZIMAS OBTIDAS DE AMOSTRAS DE MANGUEZAL NÃO BRASILEIRO

Como os manguezais apresentam condições físico-químicas especiais, os microrganismos presentes nestes ambientes contêm enzimas com adaptações que permitem a manutenção de suas estruturas e atividades enzimáticas diante de tais condições. Dessa forma, estudos sobre os manguezais podem revelar novas enzimas com características especiais.

Como exemplo, pode-se citar o trabalho de MAI e colaboradores (2014), que conseguiram identificar uma endoglucanase com halotolerância e resistência a solventes orgânicos. Essa enzima foi encontrada em uma biblioteca metagenômica, construída a partir de amostra de manguezal da cidade de Sanya, China. Durante os ensaios de atividade, os autores constataram que a endoglucanase teve sua atividade mantida em concentrações de NaCl variando de 0 a 1,5 M (em torno de 16,3 U/mg de enzima), característico de enzimas halotolerantes. Além disso, foi visto

também que a atividade da enzima diminuiu pouco na presença de solventes orgânicos (30% em média). Dessa forma, os autores acreditam que esta enzima apresenta potencial de uso em processos de degradação de biomassa lignocelulósica, com condições aumentadas de sal ou em presença de solventes orgânicos.

Outra enzima que apresentou características especiais foi descrita por GAO e colaboradores em 2010. Esses autores encontraram um gene para a enzima endo- β -1,4-glucanase no genoma de *Vibrio* sp. G21, originado de solo de manguezal de Xiamen, China. Após sua expressão e purificação, eles identificaram atividade da enzima mesmo na concentração de 4 M de NaCl, mantendo o valor próximo à atividade na ausência de sal (em torno de 629 ± 12 U/mg⁻¹). Além disso, foi visto que a enzima se encontrava ativa e com atividade próximo da máxima em valores de pH variando de 5 a 10,5, apresentando potencial em aplicações industriais.

GHANEM e colaboradores (2000) encontraram uma lipase originada de *Bacillus alcalophilus*, cujo isolamento foi feito a partir de sedimentos de manguezal da Baía de Tubli, Estado de Bahrain. A enzima apresentou atividade máxima a 60°C e pH 10,6, além de demonstrar termoestabilidade considerável: 60% de sua atividade foi mantida após incubação a 80°C por 20 minutos.

Assim, esses estudos suportam a ideia de realizar estudos com amostras de manguezal a fim de se obter enzimas versáteis e com possível aplicabilidade industrial.

4 CAPÍTULO I

The complete genome sequence of *Bacillus safensis* BRM1 isolated from Brazilian mangrove sediment: a potential source of biomass converting enzymes

Marcelo Scarduelli¹, Dieval Guizelini², Paulo da Cunha Lana³, Rodrigo Luis Alves Cardoso¹, Denny Marcel Seccon¹, Lucélia Donatti³, Valter Antônio de Baura¹, Fábio de Oliveira Pedrosa¹, Luciano Fernandes Huergo¹, Emanuel Maltempi de Souza¹

1 – Programa de Pós-Graduação em Ciências – Bioquímica, Departamento de Bioquímica e Biologia Molecular, UFPR Curitiba, PR, Brazil

2 - Setor de Educação Profissional e Tecnológica da UFPR, Curitiba, PR, Brazil

3 - Departamento de Biologia Celular da UFPR, Curitiba, PR, Brazil

ABSTRACT

A strain able to grow in xylan as sole carbon source was isolated from Brazilian mangrove sediment. In order to identify genes encoding biomass conversion enzymes, the genome of this bacterium was sequenced and showed 99% nucleotide identity to the *Bacillus safensis* genome. The isolated strain was named *Bacillus safensis* BRM1, its genome consists of a circular chromosome of 3.74 Mb with a GC content of 41.8%. The total predicted protein-coding genes was 3,699 and 18 rRNAs and 91 tRNAs genes were found. A total of 23 genes encoding putative cellulases or hemicellulases were identified. These data support that *B. safensis* BRM1 is an interesting candidate for the prospection of enzymes that can be applied in the conversion of cellulolitic biomass to biofuel.

Keywords: *Bacillus safensis*, xylanases, mangrove.

INTRODUCTION

The genus *Bacillus* is comprised of aerobic Gram-positive rod-shaped bacteria. Species of this genus can be found in many different habitats, such as soil, water, food, diseased animals (*B. anthracis* or *B. larvae*), hot springs and plants (FALKOW et al., 2006). In terms of biotechnological applications, these bacteria can be a source of a variety of hydrolytic enzymes such as amylases and proteases which are used in the industry worldwide (KUNST et al., 1997). Many *Bacillus* strains are able to produce cellulases and hemicellulases that may be applied in biofuel production, some of these enzymes showed particular interesting properties such as wide range of pH stability and thermo-stability (ITO, 1997; KHASIN; ALCHANATI; SHOHAM, 1993; RASTOGI et al., 2010). In the same way, the *Bacillus safensis* species has demonstrated the potential of producing enzymes with different biotechnology applications, including inulinases, laccases, galactosidases, oxidoreductases and keratinases (SINGH; SINGH; YADAV, 2013; SINGH et al., 2014; NATH et al., 2013; DA FONSECA et al., 2015; LATEEF et al., 2015).

Searching for new enzymes with biotechnology potential, we collected a sample of a Brazilian mangrove, located at Paranaguá bay, Paraná (25°29'S 48°27'W). The microorganisms from the sample were isolated after growth in selective medium. Then, a strain capable to growth on xylan as sole carbon source, and in LB (Luria-Bertani) medium containing 5% NaCl, was isolated and named BRM1 (*Brazilian Mangrove*). In order to identify genes encoding biomass conversion enzymes, the genome of this strain was sequenced.

MATERIALS AND METHODS

GENOME, ANNOTATION AND PHYLOGENETICS ANALYSIS

Genomic DNA was purified using phenol-chloroform extraction of cells lysed with lysozyme and SDS. Genome sequencing of BRM1 was performed using an Illumina MiSeq Sequencer. Library construction was done with Nextera DNA Library Prep Kit, and sequencing generated 3,185,036 end-paired reads (~448 Mb), yielding a 120-fold genome coverage. The data was assembled using CLC Genomics

Workbench 6.5.1 (GNERRE et al., 2011) and Roche GS De Novo Assembler (Newbler, release 2.8). Contigs were ordered in one scaffold using jContigSort program (GUIZELINI et al., 2011) and the gaps were closed with FGAP program (PIRO et al., 2014) and *in house* scripts.

Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (ANGIUOLI et al., 2008), on the fly during the genome deposit process at GenBank. Annotation was also performed using RAST genome annotation service (AZIZ et al., 2008). The Average Nucleotide Identity matrix was performed using the specific tool at Kostas Lab website (RODRIGUEZ-R; KONSTANTINIDIS, 2016) and the results were plotted as a heatmap using the Heatmapper tool (BABICKI et al., 2016).

The carbohydrate-active enzymes coding genes in the strain BRM1 genome were identified using at the CAZY online site (CANTAREL et al., 2009) using the Genbank accession number. The ribosomal sequences were identified using RNAmmer (LAGESEN et al., 2007) and ARAGORN (LASLETT; CANBACK, 2004).

The *Bacillus safensis* BRM1 genome map was generated using DNAPlotter (CARVER et al., 2009). Clusters of orthologous group (COGs) analysis were performed using the EggNog 4.5.1 software (HUERTA-CEPAS et al., 2017). PSORTb v3.0 were used to predict the protein subcellular localization (YU et al., 2010, table S4).

The 16S rRNA sequence alignments were performed using MUSCLE (EDGAR, 2004) and the phylogenetic tree was generated using MEGA6 software (TAMURA et al., 2013). The tree was constructed by neighbor-joining method using 1000 bootstraps. The 16S rRNA sequences length analyzed were 1550 nucleotides long. The whole-genome phylogenetic tree was created by NCBI general information ("Database resources of the National Center for Biotechnology Information", 2016).

GROWTH CONDITIONS

The culture medium used for the detection of enzymatic activity had the following composition: 1.6g/L of K_2HPO_4 , 0.2g/L of KH_2PO_4 , 0,2 g/L of $MgSO_4 \cdot 7H_2O$, 0.01g/L of $FeSO_4 \cdot 7H_2O$, 0.2 g/L of yeast extract, 0.2 g/L casamino acid, 1.2×10^{-2} g/L of $MnSO_4 \cdot H_2O$, 0.02 g/L of $CaCl_2 \cdot 2H_2O$, 1.4×10^{-2} g/L of H_3BO_3 , 4.0×10^{-4} g/L of $CuSO_4 \cdot 5H_2O$, 1.2×10^{-3} g/L $ZnSO_4 \cdot 7H_2O$, 0.01 g/L of $Na_2MoO_4 \cdot 2H_2O$, 20 g/L of NaCl,

15 g/L of agar and 5 g/L of xylan or carboxymethyl cellulose (CMC). The plates were maintained at 30°C for 4 days. The presence of CMC hydrolytic activity was verified by the formation of hydrolytic halo using the red congo staining protocol (CARDER, 1986). For the xylanase activity, the halo could be identified directly without the need of a staining step.

For enzymatic assay in liquid medium, the strain was cultivated aerobically at rotator shaker at 30°C 120 rpm. The medium composition was the same as described by COMMICHAU and collaborators (2008) but diluted 1:1 (water:medium). In some experiments the medium was complemented with a carbon source using 0.5% (w/v) of glucose (control), cellobiose, xylan or carboxymethyl cellulose (CMC). A 300 µL aliquot was removed daily from each cultivation flask and the cell density was monitored by measuring the optical density at 600 nm.

RESULTS

GENOME ANALYSIS

The genome of BRM1 consists of one circular chromosome with 3,740,842 pb (Figure 1 and Table 1) and a GC content of 41.80%, similar to others 5 complete genomes of *Bacillus safensis* strains in NCBI database (range: 41.36 to 41.80% GC content). A total of eight rRNA operons were identified: eight 5 S rRNAs, eight 16 S rRNAs and eight 23 S rRNAs. There were 91 tRNAs, covering all 20 amino acids. The number of coding sequences (CDS) was 3.699, on average 837 nt in size. The content of non-coding regions was 10.82%.

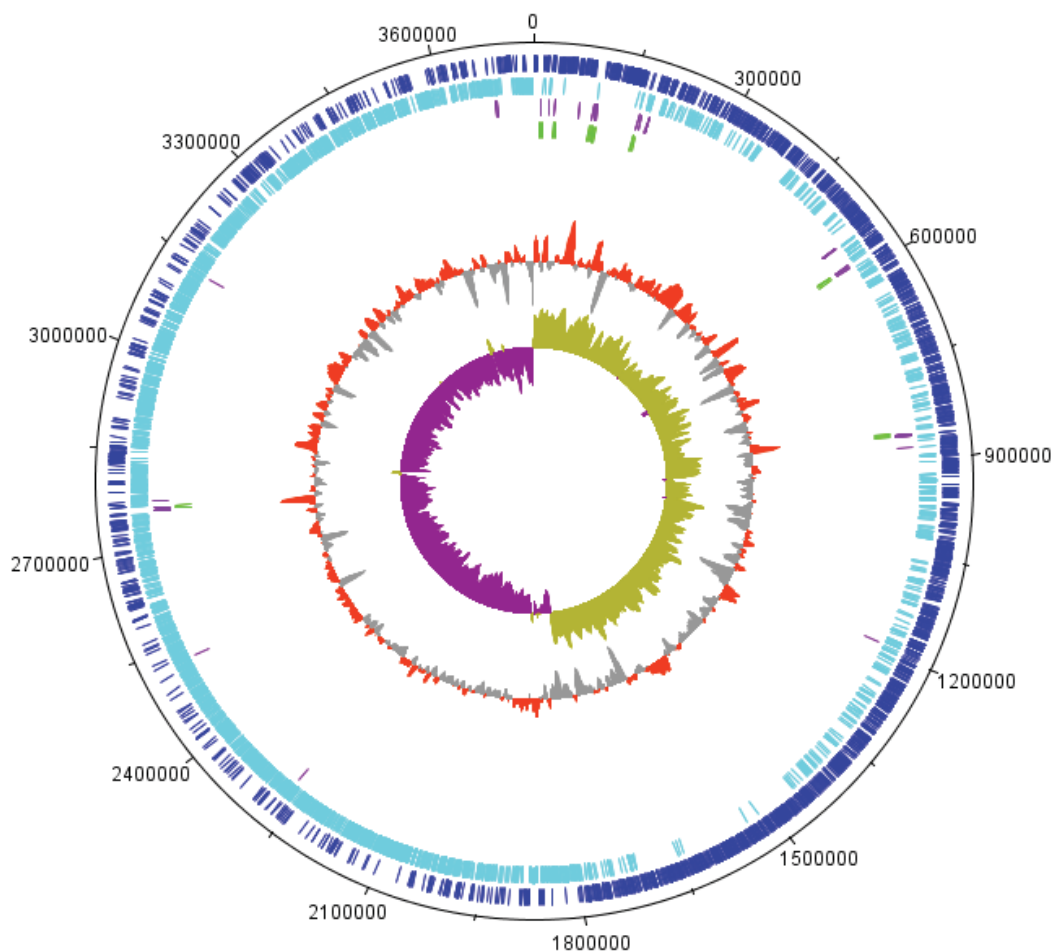


Figure 1. Map of the *Bacillus safensis* BRM1 genome. From the outer to inner circles: ORFs on the forward strand (in dark blue) and reverse strand (light blue), tRNA genes in both strands (purple), rRNA genes in both strands (green), GC percentage plot (above average in orange, below average in grey), GC skew (lime green above average, purple below average).

Table 1. Genome features of *Bacillus safensis* BRM1.

Features	Value
Genome size (bp)	3,740,842
Chromosome number	1
G+C content (%)	41.8%
Total predicted ORFs	3,699
Plasmid	0
rRNAs	24 (8 5S rRNA, 8 16S rRNA and 8 23S rRNA)
tRNAs	91
NCBI accession number	CP018100

PHYLOGENETIC ANALYSIS

Analysis of similarity was performed using matrix methodology, supported by Average Nucleotide Identity tool (RODRIGUEZ-R; KONSTANTINIDIS, 2016). The results were plotted as a heatmap (Figure 2 and Table S1). The identity varies between 77% to 97.5% among 14 complete *Bacillus* genomes deposited in NCBI database, previously selected due to their higher identity of the 16S rRNA alignment with BRM1. The *B. safensis* strain U14-5 genome showed the highest identity with BRM1 nucleotides (97.5%). After the comparison with BRM1 genome, we could not find the enzymes chitinase and ketoglutarate semialdehyde dehydrogenase, the DNA repair methyltransferase, the two phage proteins and the Clp protease-like protein, maintaining exclusivity in BRM1 genome.

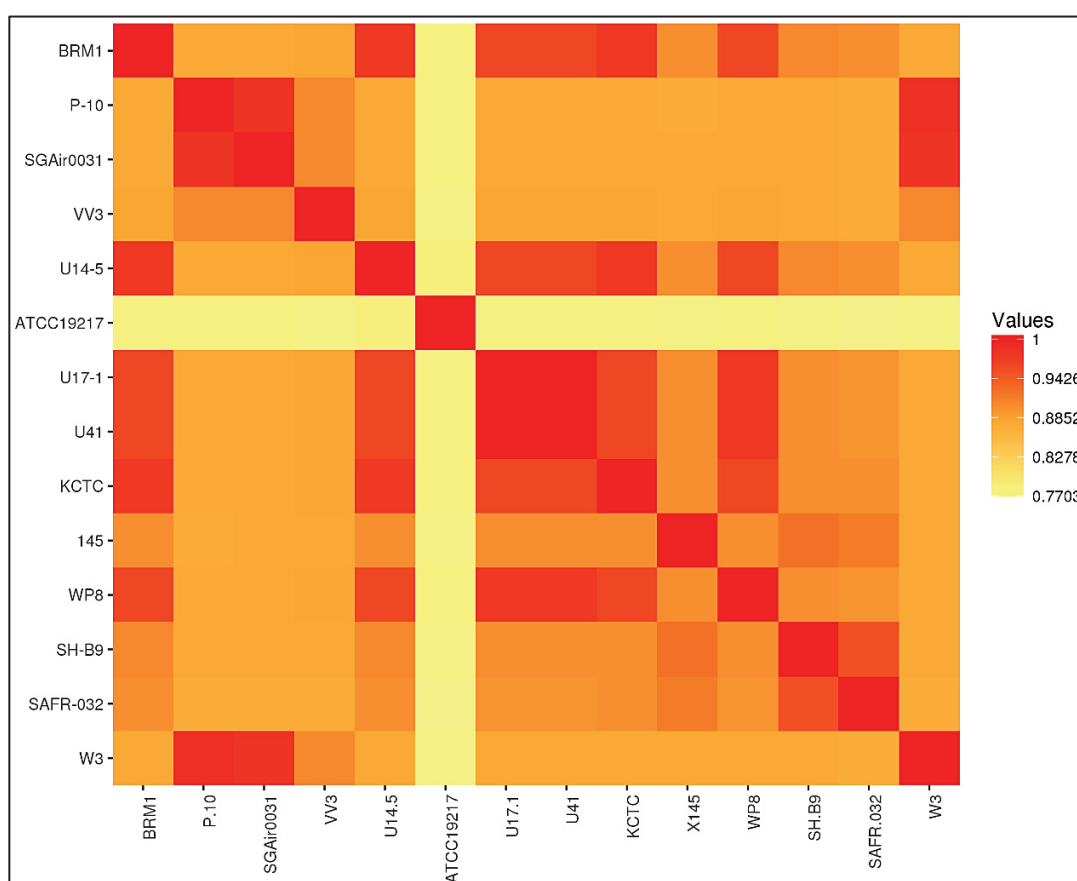


Figure 2. Heatmap showing Average Nucleotide Identity between the 12 *Bacillus* strain genomes deposited in NCBI database. The Genbank accession numbers were follow: *B. safensis* BRM1 (CP018100), *B. altitudinis* P-10 (CP024204.1), *B. altitudinis* SGAir0031 (CP022319.1), *B. xiamenensis* VV3 (CP017786.1), *B. safensis* U14-5 (CP015607.1), *B. subtilis* ATCC 19217 (CP009749.1), *B. safensis* U17-1 (CP015611.1), *B. safensis* U41 (CP015610.1), *B. safensis* KCTC (CP018197.1), *B. pumilus* 145 (CP027116.1), *Bacillus* sp. WP8 (CP010075.1), *B. pumilus* SH-B9 (CP011007.1), *B. pumilus* SAFR-032 (CP000813.4) and *B. pumilus* W3 (CP011150.1).

ANNOTATION AND PROTEINS ANALYSIS

To visualize the genes function in *B. safensis* BRM1, we performed a cluster of orthologous group (COGs) analysis, which classify the proteins/ORFs into specific categories related to different aspects of physiology and cellular metabolism, using the EggNog 4.5.1 program (HUERTA-CEPAS et al., 2017). The categories distribution is depicted in Figure 3 and in supplementary material (Table S2). Genes related to transcription (K, 9.22%), amino acid (E, 9.09%) and carbohydrate metabolism (G, 7.05%) reach 25.3% of total protein coding regions identified. Protein belonging to unknown functions represented 26.6% of the sequences. Other categories, such as translation (5.53%), cell wall/membrane (5.72%), ion transport and metabolism (5.47%) proteins, also showed higher values.

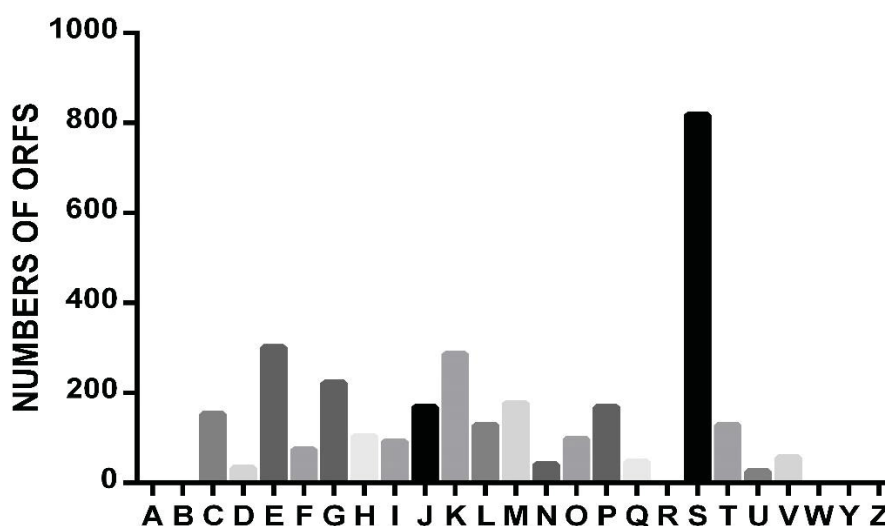


Figure 3. COG classification from *B. safensis* BRM1 genes. The 25 categories are: [A] RNA processing and modification, [B] Chromatin structure and dynamics, [C] Energy production and conversion, [D] Cell cycle control, cell division, chromosome partitioning, [E] Amino acid transport and metabolism, [F] Nucleotide transport and metabolism, [G] Carbohydrate transport and metabolism, [H] Coenzyme transport and metabolism, [I] Lipid transport and metabolism, [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair, [M] Cell wall/membrane/envelope biogenesis, [N] Cell motility, [O] Post-translational modification, protein turnover, and chaperones, [P] Inorganic ion transport and metabolism, [Q] Secondary metabolites biosynthesis, transport, and catabolism, [R] General function prediction only, [S] Function unknown, [T] Signal transduction mechanisms, [U] Intracellular trafficking, secretion, and vesicular transport, [V] Defense mechanisms, [W] Extracellular structures, [Y] Nuclear structure, [Z] Cytoskeleton.

BIOMASS HYDROLYTIC ENZYMES

According to carbohydrate-active enzyme (CAZy) site (CANTAREL et al., 2009), the BRM1 genome had 35 genes that could encode biomass degradation enzymes (Table 2 and Table S3). They were classified into 18 different families of glycosyl hydrolases (GHs) and carbohydrate esterases (CEs). After signal peptide analysis, 12 enzymes are predicted to be extracellular, showing the biomass conversion potential in this strain.

Table 2. Genes for putative cellulases and hemicellulases in *Bacillus safensis* BRM1 genome.

Probable activity	Conserved domains (# of genes found)
<i>Cellulases</i>	
Endo/exoglucanases	GH 5 (1), 9 (1), 48 (1), 51(1)
β -Glucosidase	GH 1 (6), 3 (1) or 42 (1)
<i>Hemicellulases</i>	
Xylanases or β -Xylosidases	GH 10 (1), 11 (1), 16 (1), 30 (1), 43 (2)
α -galactosidases	GH 4 (4)
Galactanases	GH 53 (1)
Acetyl xylan esterase	CE 4 (9), 6 (1), 7 (1), 12 (1)
(Total: 35 genes for putative hydrolases)	

GROWTH ON SPECIFIC SUBSTRATES

The BRM1 strain was isolated during screening steps, when a Brazilian mangrove sample was spreaded in solid media containing CMC or xylan as a sole carbon source were used (as described in Material and Methods section). This strain is capable to develop under 2% (m/m) of NaCl and the hydrolysis halo could be seen after 4 days of growth at 30°C (Figure 4).

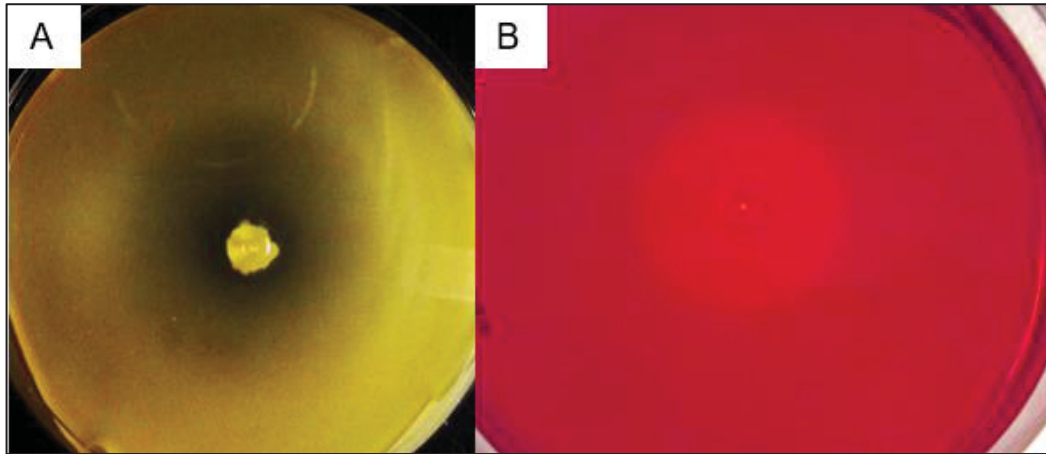


Figure 4. Halo hydrolysis formed after 4 days at 30°C of cultivation. (A) plate containing xylan as a carbon source, indicating xylanase activity. (B) plate with CMC as a sole carbon source, showing CMCcase activity. Here, the halo was revealed after Red Congo stain.

In order to confirm the hydrolytic potential, the BRM1 strain was cultured aerobically in liquid medium supplemented with glucose (control), cellobiose, xylan or carboxymethyl cellulose (CMC). Growth analysis was performed using the minimal medium CSE described by COMMICHAU et al. (2008). The use of CSE media provided better homogeneity during cell growth, without the appearance of biofilm and bacteria flocculation that was observed using LB or M9 medium (data not shown).

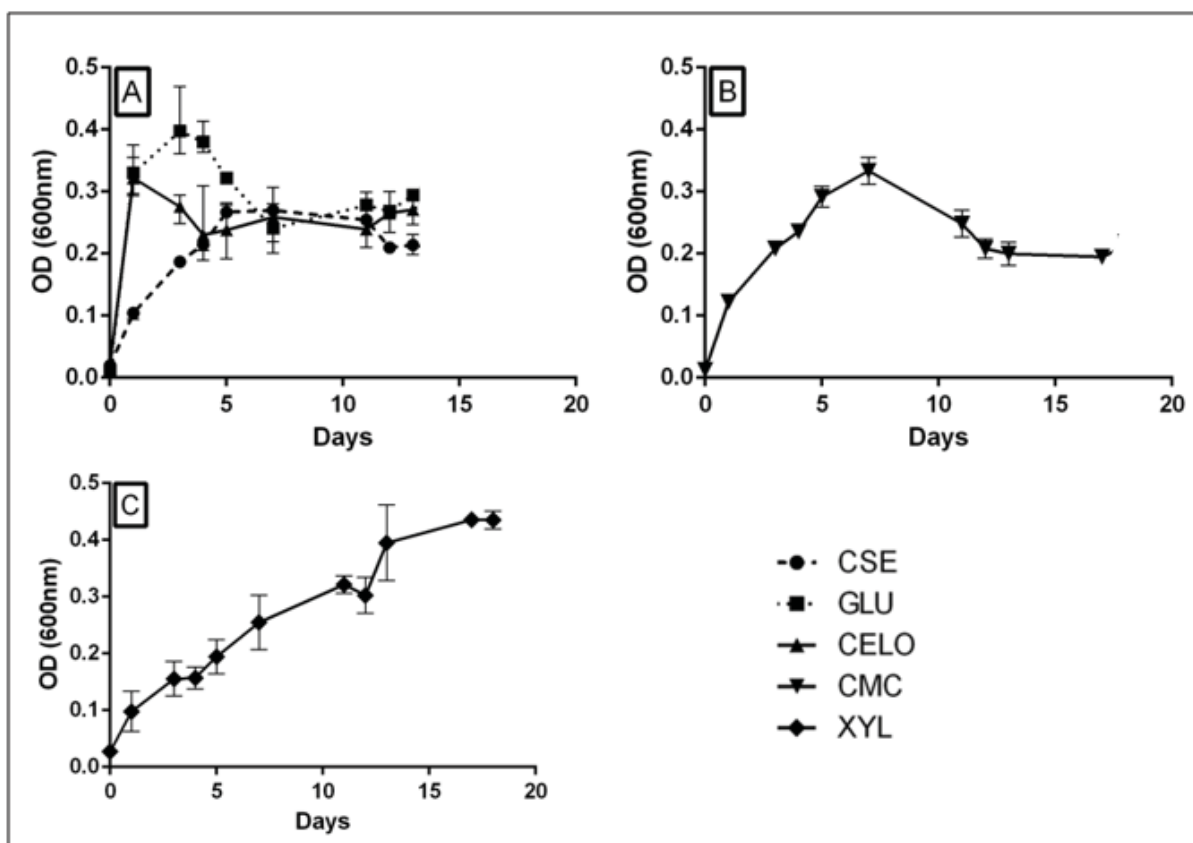


Figure 5. *Bacillus safensis* BRM1 growth in CSE medium supplemented with cellobiose (A), CMC (B) and xylan (C), in a concentration of 0.5% (m/v). The controls were the CSE medium without carbon source (●) and with glucose at 0.5% (m/v, ■). The curve related to growth in cellobiose is marked with triangle (▲). For growth in CMC and xylan, the points were marked with inverted triangle (▼) and diamond (◆), respectively. The growth was performed in triplicate (n=3).

The growth curves in figure 5A indicate that BRM1 was capable to grow in CSE minimal medium until day 5, probably consuming the glutamate and succinate present in the broth. After this, the strain reaches the stationary phase, indicating nutrients depletion. For cellobiose, the bacteria used it similarly to glucose, showing high growth rate during the first days until day 7 and after that we suppose that occurred the initial consumption of other nutrients (glutamate and succinate). On the other hand, the strain was capable of using xylan, growing until day 17 and it did not show the stationary phase like the other conditions. The CMC conditions did not reveal difference when compared to the controls (CSE medium only and glucose supplementation), indicating a metabolic interference.

DISCUSSION

Here we report the genome of *B. safensis* BRM1, the fifth complete genome of *Bacillus safensis* deposited in the *Genbank* database (up to now). The BRM1 strain was isolated from the Brazilian mangrove soil, at Paranagua Bay, which is an environment with variation in salinity, water level and nutrient availability. The other strains with complete genome described were also found in saline environments: Antarctic Lake and Chinese ocean. The BRM1 strain genome encodes the enzymatic machinery to degrade different carbon sources, including xylan, cellobiose and CMC. RAST analysis showed that the xylan utilization pathway is complete in the BRM1 strain genome: xylose isomerase (XylA), xylulose kinase (XylB), xylose-responsive transcription regulator (XylR), xyloside transporter (XynT) and three xylan hydrolases (putatives XynA, XynB and/or XynC). These genes explain the ability of this strain to grow in medium using xylan as a carbon source (figures 4 and 5).

At least 12 enzymes potentially involved in the degradation of CMC/cellulose could be identified in the genome of BRM1, these enzymes could allow the efficient degradation of cellulose to cellobiose and then to D-glucose, as seen at growth curves assays. We also identified four ABC type membrane transport proteins which could involve in the uptake of cellobiose. The growth assay revealed CMC degradation only in solid medium. Since BRM1 demonstrated hydrolysis halo in solid medium plate containing CMC, we suggest that the aeration could interfere in CMCase (or cellulase) activity, blocking the expression of hydrolases from this type of substrate.

CONCLUSION

The complete genome of *Bacillus safensis* BRM1 revealed that this strain may have biotechnological potential due the identified cellulases and hemicellulases. Growth assays showed that the strain was capable to degrade cellobiose, CMC and xylan under certain growth conditions, which support the existence of active enzyme machinery for biomass degradation. Future studies should be conducted to evaluate the activities of these enzymes separately and the possible usage in biomass-to-biofuel conversion processes.

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The genome information of *Bacillus safensis* BRM1 was deposited in NCBI under a GenBank accession number CP018100.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Science and Technology of Nitrogen Fixation/ CNPq/MCT, Fundação Araucária and Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

- ANGIUOLI, S. V. et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. **Omics: A Journal of Integrative Biology**, v. 12, n. 2, p. 137–141, jun. 2008.
- AZIZ, R. K. et al. The RAST Server: Rapid Annotations using Subsystems Technology. **BMC Genomics**, v. 9, p. 75, 8 fev. 2008.
- BABICKI, S. et al. Heatmapper: web-enabled heat mapping for all. **Nucleic Acids Research**, v. 44, n. W1, p. W147–W153, 8 jul. 2016.
- CANTAREL, B. L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research**, v. 37, n. Database issue, p. D233-238, jan. 2009.
- CARDER, J. H. Detection and quantitation of cellulase by Congo red staining of substrates in a cup-plate diffusion assay. **Analytical Biochemistry**, v. 153, n. 1, p. 75–79, 15 fev. 1986.
- CARVER, T. et al. DNAPlotter: circular and linear interactive genome visualization. **Bioinformatics**, v. 25, n. 1, p. 119–120, 1 jan. 2009.
- COMMICHAU, F. M. et al. Glutamate metabolism in *Bacillus subtilis*: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. **Journal of Bacteriology**, v. 190, n. 10, p. 3557–3564, maio 2008.
- DA FONSECA, F. S. A. et al. Identification of oxidoreductases from the petroleum *Bacillus safensis* strain. **Biotechnology Reports**, v. 8, p. 152–159, 1 dez. 2015.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.
- FALKOW, S. et al. **The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria**. [s.l.] Springer Science & Business Media, 2006.
- GNERRE, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. **Proceedings of the National Academy of Sciences**, v. 108, n. 4, p. 1513–1518, 25 jan. 2011.

- GUIZELINI, D. et al. jContigSort: a new computer application for contigs ordering. In: **7th International Conference of The Brazilian Association for Bioinformatics and Computational Biology: abstract book**. Florianópolis: [s.n.].
- HUERTA-CEPAS, J. et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. **Molecular Biology and Evolution**, v. 34, n. 8, p. 2115–2122, 1 ago. 2017.
- ITO, S. Alkaline cellulases from alkaliphilic *Bacillus*: Enzymatic properties, genetics, and application to detergents. **Extremophiles**, v. 1, n. 2, p. 61–66, maio 1997.
- KHASIN, A.; ALCHANATI, I.; SHOHAM, Y. Purification and characterization of a thermostable xylanase from *Bacillus stearothermophilus* T-6. **Applied and Environmental Microbiology**, v. 59, n. 6, p. 1725–1730, 6 jan. 1993.
- LAGESEN, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. **Nucleic Acids Research**, v. 35, n. 9, p. 3100–3108, 5 jan. 2007.
- LASLETT, D.; CANBACK, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, v. 32, n. 1, p. 11–16, 2004.
- LATEEF, A. et al. Green synthesis of silver nanoparticles using keratinase obtained from a strain of *Bacillus safensis* LAU 13. **International Nano Letters**, v. 5, n. 1, p. 29–35, 1 mar. 2015.
- NATH, A. et al. Purification and Characterization of β -Galactosidase Synthesized from *Bacillus safensis* (JUCHE 1). **Industrial & Engineering Chemistry Research**, v. 52, n. 33, p. 11663–11672, 21 ago. 2013.
- PIRO, V. C. et al. FGAP: an automated gap closing tool. **BMC Research Notes**, v. 7, n. 1, p. 1–5, 2014.
- RASTOGI, G. et al. Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. **Bioresource Technology**, v. 101, n. 22, p. 8798–8806, nov. 2010.
- RODRIGUEZ-R, L. M.; KONSTANTINIDIS, K. T. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. 27 mar. 2016.
- SINGH, D. et al. Molecular Docking of Laccase Protein from *Bacillus Safensis* DSKK5 Isolated from Earthworm gut: A Novel Method to Study dye Decolorization Potential. **Water, Air, & Soil Pollution**, v. 225, n. 11, p. 2175, 1 nov. 2014.
- SINGH, R. S.; SINGH, R. P.; YADAV, M. Molecular and biochemical characterization of a new endoinulinase producing bacterial strain of *Bacillus safensis* AS-08. **Biologia**, v. 68, n. 6, p. 1028–1033, 1 dez. 2013.
- TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, dez. 2013.
- YU, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **Bioinformatics**, v. 26, n. 13, p. 1608–1615, 1 jul. 2010.

SUPPLEMENTARY MATERIAL

Table S1. Raw data from ANI matrix analysis among the *Bacillus* deposited genomes (with plasmids if exist), where values closer to 1 indicate greater genome identity. In order: *B. pumilus* 145 (CP027116.1), *B. subtilis* ATCC 19217 (CP009749.1), *B. safensis* BRM1 (CP018100), *B. safensis* KCTC with plasmid1 (CP018197.1), *B. altitudinis* P-10 (CP024204.1), *B. pumilus* SAFR-032 (CP000813.4), *B. altitudinis* SGair0031 with plasmid1 (CP022319.1), *B. pumilus* SH-B9 with plasmid1 (CP011007.1), *B. safensis* U14-5 with plasmid1 and 2 (CP015607.1), *B. safensis* U17-1 (CP015611.1), *B. safensis* U41 (CP015610.1), *B. xiamenensis* VV3 (CP017786.1), *B. pumilus* W3 (CP011150.1) and *Bacillus* sp. WP8 (CP010075.1).

	145	ATCC 19217	BRM1	KCTC+p1	P-10	SAFR-032	SGair0031+p1	SH-B9+p1	U14-5+p1+p2	U17-1	U41	VV3	W3	WP8
145	1.000000	0.779222	0.901513	0.901604	0.875882	0.916894	0.876593	0.921958	0.902487	0.901369	0.900797	0.878882	0.876743	0.900439
ATCC 19217	0.779222	1.000000	0.782503	0.782877	0.782626	0.770348	0.782859	0.775531	0.785638	0.780652	0.779673	0.777217	0.775798	0.781545
BRM1	0.901513	0.782503	1.000000	0.973070	0.879438	0.898965	0.879523	0.903775	0.975625	0.959749	0.959888	0.883499	0.878762	0.958992
KCTC+p1	0.901604	0.782877	0.973070	1.000000	0.879265	0.899239	0.880598	0.902816	0.974392	0.960854	0.960918	0.883685	0.878047	0.960240
P-10	0.875882	0.782626	0.879438	0.879265	1.000000	0.873315	0.980254	0.877342	0.879750	0.878855	0.878658	0.905796	0.984046	0.878875
SAFR-032	0.916894	0.770348	0.898965	0.899239	0.873315	1.000000	0.873954	0.953929	0.899810	0.898417	0.897817	0.874100	0.874101	0.898137
SGair0031+p1	0.876593	0.782859	0.879523	0.880598	0.980254	0.873954	1.000000	0.879206	0.880349	0.879063	0.877642	0.906511	0.980353	0.879244
SH-B9+p1	0.921958	0.775531	0.903775	0.902816	0.877342	0.953929	0.879206	1.000000	0.904246	0.902655	0.902511	0.878174	0.877286	0.902348
U14-5+p1+p2	0.902487	0.785638	0.975625	0.974392	0.879750	0.899810	0.880349	0.904246	1.000000	0.961192	0.961278	0.883971	0.878734	0.959536
U17-1	0.901369	0.780652	0.959749	0.960854	0.878855	0.898417	0.879063	0.902655	0.961192	1.000000	0.999992	0.882472	0.878200	0.974153
U41	0.900797	0.779673	0.959888	0.960918	0.878658	0.897817	0.877642	0.902511	0.961278	0.999992	1.000000	0.882408	0.878768	0.973989
VV3	0.878882	0.777217	0.883499	0.883685	0.905796	0.874100	0.906511	0.878174	0.883971	0.882472	0.882408	1.000000	0.905962	0.882047
W3	0.876743	0.775798	0.878762	0.878047	0.984046	0.874101	0.980353	0.877286	0.878734	0.878200	0.878768	0.905962	1.000000	0.878145
WP8	0.900439	0.781545	0.958992	0.960240	0.878875	0.898137	0.879244	0.902348	0.959536	0.974153	0.973989	0.882047	0.878145	1.000000

Table S2. Clusters of orthologous group (COG) analysis for *Bacillus safensis* BRM1.

COG category	Code	Total	Total (%)
Information storage and processing			
RNA processing and modification	A	0	0,0%
Chromatin structure and dynamics	B	0	0,0%
Translation, ribosomal structure and biogenesis	J	169	5.1%
Transcription	K	286	8.6%
Replication, recombination and repair	L	129	3.9%
Cellular processes and signaling			
Cell cycle control, cell division, chromosome partitioning	D	33	1.0%
Cell wall/membrane/envelope biogenesis	M	177	5.3%
Cell motility	N	41	1.2%
Posttranslational modification, protein turnover, chaperones	O	98	2.9%
Signal transduction mechanisms	T	129	3.9%
Intracellular trafficking, secretion, and vesicular transport	U	25	0.8%
Defense mechanisms	V	56	1.7%
Extracellular structures	W	0	0.0%
Nuclear structure	Y	0	0.0%
Cytoskeleton	Z	0	0.0%
Metabolism			
Energy production and conversion	C	153	4.6%
Amino acid transport and metabolism	E	301	9.0%
Nucleotide transport and metabolism	F	74	2.2%
Carbohydrate transport and metabolism	G	222	6.7%
Coenzyme transport and metabolism	H	104	3.1%
Lipid transport and metabolism	I	92	2.8%
Inorganic ion transport and metabolism	P	169	5.1%
Secondary metabolites biosynthesis, transport and catabolism	Q	47	1.4%
Poorly characterized			
General function prediction only	R	0	0,00%
Function unknown	S	818	24.6%
Total of ORFs Annotated		3,326	100%
ORFs unclassified		373	

Table S3. Putative cellulases and hemicellulases in *Bacillus safensis* BRM1, according to CAZY database (CANTAREL et al., 2009).

Gene ID	Conserved domain	Localization prediction ¹
ARD54928.1	GH1	Cytoplasmic
ARD58015.1	GH1	Cytoplasmic
ARD56702.1	GH1	Cytoplasmic
ARD58075.1	GH1	Cytoplasmic
ARD57987.1	GH1	Cytoplasmic
ARD5091.1	GH1	Cytoplasmic
ARD54868.1	GH3	Cytoplasmic
ARD55290.1	GH4	Cytoplasmic
ARD55295.1	GH4	Cytoplasmic
ARD56363.1	GH4	Cytoplasmic
ARD57955.1	GH4	Cytoplasmic
ARD56165.1	GH5	Extracellular
ARD56162.1	GH9+CBM3	Extracellular
ARD56164.1	GH10	Extracellular
ARD56433.1	GH11	Extracellular
ARD55167.1	GH16	Extracellular
ARD56360.1	GH30	Extracellular
ARD58058.1	GH42	Cytoplasmic
ARD56361.1	GH43	Cytoplasmic
ARD56437.1	GH43	<i>Unknown</i>
ARD56163.1	GH48	Extracellular
ARD57051.1	GH51	Cytoplasmic
ARD58057.1	GH53	Membrane-ligated
ARD54864.1	CE4	Extracellular
ARD55371.1	CE4	Cytoplasmic
ARD55521.1	CE4	Cytoplasmic
ARD56186.1	CE4	Extracellular
ARD56538.1	CE4	Cytoplasmic
ARD56877.1	CE4	Cytoplasmic
ARD56884.1	CE4	Extracellular
ARD57608.1	CE4	Extracellular
ARD57984.1	CE4	Membrane-ligated
ARD55769.1	CE6	Membrane-ligated
ARD55018.1	CE7	Cytoplasmic
ARD55391.1	CE12	Extracellular

¹ – Subcellular localization prediction by PSORTb v3.0 (YU et al., 2010).

GH: glycosyl hydrolase family. CE: carbohydrate esterase family. CBM: carbohydrate-binding module.

5 CAPÍTULO II

Identification of a novel esterase family after identification of an esterase from a marine *Pseudomonas sp.*

Marcelo Scarduelli¹, Fábio de Oliveira Pedrosa¹, Luciano Fernandes Huergo¹,
Emanuel Maltempi de Souza¹

1 – Programa de Pós-Graduação em Ciências – Bioquímica, Departamento de
Bioquímica e Biologia Molecular, UFPR Curitiba, PR, Brazil

ABSTRACT

An isolate, with similarity with *Pseudomonas sp.*, was identified during a study with a Brazilian mangrove sample. After DNA sequence, a gene encoding a putative esterase was identified and named Est1. Analysis of the Est1 amino acid sequence revealed an alpha/beta hydrolase conserved domain and the pentapeptide motif commonly found within the lipase/esterase families. On the other hand, Est1 does not have other signature domains typically present in other esterases. Recombinant Est1 was able to hydrolyze short alkyl chain pNP substrates *in vitro*. Other proteins similar to Est1 were found at the GenBank database and these sequences revealed two conserved motifs not previously described within esterases. Our data support a novel group of enzymes within the esterase family.

Keywords: esterase, *Pseudomonas sp.*, esterase family.

INTRODUCTION

Esterases are enzymes that act on ester bonds. These enzymes have the ability to catalyze esterification, transesterification (alcoholysis, interesterification and acidolysis), aminolysis (amide synthesis) and lactonization. The reactions of hydrolysis (direct reaction) or synthesis (reverse reaction) depend on the amount of water or organic solvent present in the reaction mixture (VERMA; AZMI; KANWAR, 2008).

Also called carboxylesterases, esterases have the number EC 3.1.1.3 and EC 3.1.1.1, respectively, according to the nomenclature established by IUBMB. This classification refers to the class of hydrolases (3) acting on ester bonds (3.1), more specifically on ester bonds in carboxylic acids (EC 3.1.1). Lipases and esterases are distinguished accordingly to the length of the hydrocarbonic chain of the target substrate. Lipases use long chain acyl esters as substrates. On the other hand, esterases act on smaller substrates typically containing less than 10 carbon atoms in their structure (JAEGER; DIJKSTRA; REETZ, 1999; KANG et al., 2006).

Microbial carboxylesterases have molecular mass between 20 and 60 kDa and contain a catalytic triad usually consisted of serine (Ser), histidine (His) and aspartate (Asp). The presence of a lid, which hides the active site, is frequent in these enzymes. The lid may expose or hide the substrate binding site accordingly to the interactions within the lipid/water interface. Both lipases and esterases belong to the α/β hydrolases family (OLLIS et al., 1992), with a conserved structural motif composed of β -sheets and intercalated α -helices. The serine residue belonging to the catalytic triad is found within a highly conserved Gly-X-Ser-X-Gly pentapeptide, where "X" may represent any residue (JAEGER; DIJKSTRA; REETZ, 1999).

Esterases have a number of biotechnological applications, including the production of drugs, production of biodiesel through the transesterification of triacylglycerols with short alcohols, and in the production of dairy products (ISO et al., 2001; KHEADR; VUILLEMARD; EL-DEEB, 2003).

Here we report a novel esterase identified from an isolate, a *Pseudomonas* sp., that was found in Brazilian mangrove sample. This enzyme shows the pentapeptide motif, but no other conserved domain of esterases families as described by ARPIGNY and JAEGER (1999).

MATERIAL AND METHODS

EST1 SEQUENCE ANALYSIS

Comparison of amino acids sequence were performed using BlastP (ALTSCHUL et al., 1997). PSORT program (YU et al., 2010) was used for the prediction of protein localization in cells. Structure model for Est1 was suggested by Phyre2 algorithm (KELLEY et al., 2015), using the intensive modeling mode. Lipases families sequences were retrieved from GenBank accordingly to the sequence motifs described by Arpigny and Jaeger (1999). Sequences were aligned using MUSCLE (EDGAR, 2004). Conserved sequence motifs were identified using WebLogo (CROOKS et al., 2004).

CLONING, EXPRESSION AND PURIFICATION OF EST1

The Est1 gene was amplified by PCR using the primers est1-F (5'-AAGGATTCCCATATGCGCCACGA-3') and est1-R (5'-CTCGTAGGGGATCCTGGTTA-3'), containing restriction enzyme sites (underlined) for *Nde*I and *Bam*HI, respectively. The template used was a genomic DNA from *Pseudomonas sp.* strain isolated from Brazilian mangrove using selective medium. The PCR product was digested with *Nde*I and *Bam*HI and ligated into the expression vector pET-28a (Novagen) and introduced into *E. coli* BL21(λDE3) (Novagen). The transformants were cultivated in LB medium containing 50 µg/mL of kanamycin at 37°C and 120 rpm. Protein expression was performed at 37°C by adding 0.3 mM IPTG to an OD₆₀₀ ~ 0.6 culture, following by 3 h induction. Induced cells were harvested and lysed by sonication in lysis buffer (50 mM Tris-HCl pH 8.0 and 50 mM KCl). Cell extracts were centrifuged at 30,000 g for 30 min, and the soluble fraction was loaded first onto a HiTrap Heparin HP affinity chromatography column (1 mL), to eliminate adsorbed contaminant proteins, then onto a 1 mL HiTrap Chelating-Ni²⁺ chromatography column (GE Healthcare). Proteins were eluted in lysis buffer for Heparin Sepharose chromatography. The non-ligated fraction obtained were eluted in Chelating-Ni²⁺ column in an imidazole gradient. Fractions of 1 ml were collected and analyzed by 12.5% SDS-PAGE (LAEMMLI, 1970).

PROTEIN QUANTIFICATION

The total protein concentration was estimated using Bradford method reagent (BioRad) according to the manufacturer's protocol. Bovine serum albumin (BSA) was used as standard (BRADFORD, 1976).

ENZYMATIC ASSAY

The esterase activity was determined by a kinetic assay using p-nitrophenyl (pNP) monoacyl esters (Sigma Aldrich) as substrates and the release of p-nitrophenol was monitored at 410 nm in iMark™ microplate reader (Bio-Rad). The standard reaction conditions were 43.5 mM Tris-HCl pH 7.5, 10 µg of purified Est1, 1 mM of substrate pNP monoacyl esters, in a final volume of 250 µl. The temperature used was 37°C and 10 minutes of reaction time. One unit of esterase activity was defined as the amount of enzyme which liberated 1 µmol of fatty acids per minute per mg of enzyme.

RESULTS AND DISCUSSION

SEQUENCE ANALYSIS

During an isolation with selective medium of microorganisms belonging to a Brazilian mangrove sample, we discovered the isolate XS-14, which has similarity to a *Pseudomonas sp.* strain PrPI085 (91% of identity with 16S rRNA gene). This strain was capable to grow in medium containing xylan as carbon source and with high saline concentration (up to 5 g of NaCl per 100 mL of medium).

We have acquired partial sequences of the genome of XS-14, which was found the esterase *est1*. It encodes for a 316 amino acid protein with calculated mass weight of 34.9 kDa, a theoretical pI value of 6.26 and it is predicted as a cytoplasmatic protein. BlastP analysis shows high identity (91-99%) with alpha/beta hydrolases from different *Pseudomonas* species: *P. oleovorans*, *P. mendocina*, *P. toyotomiensis*, *P. alcaliphila*, *P. pseudoalcaligenes*. The Est1 enzyme shows a alpha/beta hydrolases superfamily domain, which is associated with proteases, lipases, peroxidases, esterases, epoxide hydrolases and dehalogenases. Est1 is

considered a esterase because its sequence has the pentapeptide GHSMG, common in this type of enzymes (GX SXG module, where S is a catalytic triad member). Figure 1 shows the Est1 amino acids residues sequence. A structure model for Est1 was built, which the catalytic triad residues showed to be next to each other spatially.

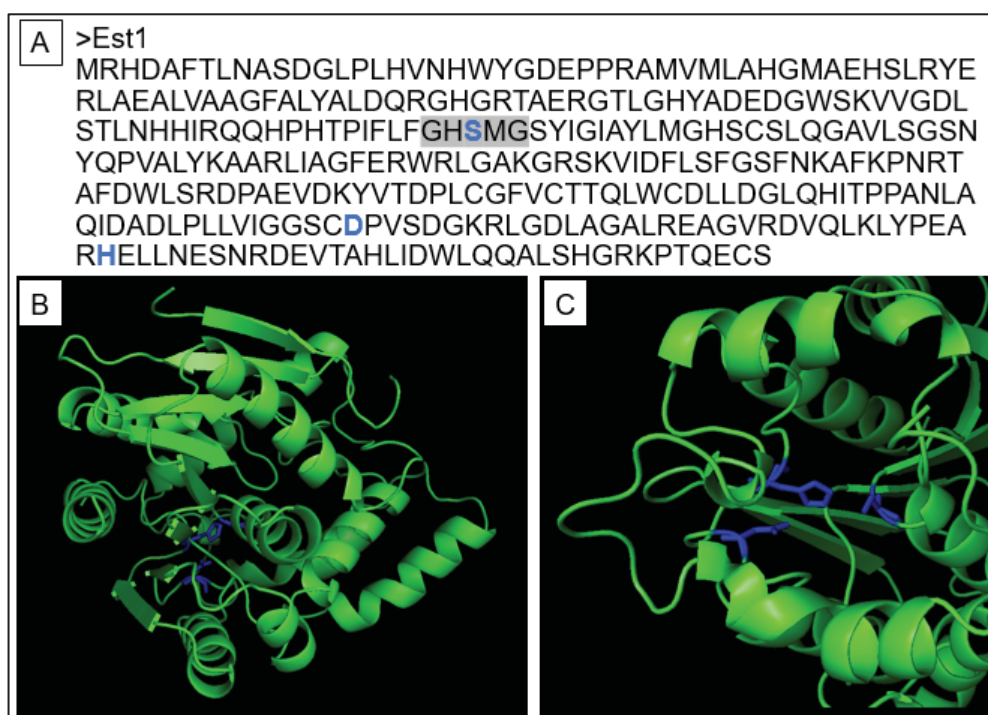


Figure 1. (A) Est1 amino acids sequence. The pentapeptide (GHSMG) is marked in gray and the catalytic triad are in blue (Ser¹¹³, Asp²⁴⁹ and His²⁸²). (B) A structure model for Est1, generated by Phyre2, with 100% of confidence, 96% of sequence coverage and 23% of amino acids identity to a highly thermostable carboxyl esterase from *Bacillus coagulans* (pdb id: c5oluA). (C) The catalytic triad is evidenced in blue.

Arpigny and Jaeger (1999) suggested that bacterial carboxylesterases can be classified into families according to their conserved amino acid sequences and biochemical properties. True lipases, which are capable to hydrolyze long-chain triacylglycerides, are members of the Family I. Family II contains enzymes that do not exhibit the conventional pentapeptide (GHSMG), but present the GD SL motif. Family III shows identity to isoforms of the human PAF-AH (Platelet-activating factor acetylhydrolase). Enzymes with similarity to mammalian HSL (hormone-sensitive lipase) are classified in Family IV. Family V includes various bacterial non-lipolytic enzymes (epoxide hydrolases, dehalogenases and haloperoxidases), with the alpha/beta hydrolases fold. Families VI and VII present enzymes with esterase

activities, and a small (family VI) or a large (VII) amino acids sequence. Finally, esterases or lipases that possess the SXXK motif are classified in the Family VIII.

In this way, protein sequences selected by the Arpigny and Jaeger were aligned along with the Est1 sequence. As result, a similarity tree was generated and it is shown in Figure 2. Est1 showed more similarity with an esterase from *B. subtilis* (30% of identity), which is a member of Family I, subfamily 4 that do not show any conserved regions. Comparing to another families, the Est1 sequence does not show other characteristic conserved regions such as the LVHGL, KVNLIHGSXG, NDGLV sequences of this family, only the conventional pentapeptide GHSMG. Other motifs, like GDSL (Family II), NXAXGGXT (Family II), VGGND (Family III), LFXDXXHPT (Family V), HGGGF or GDSAGGXLA (Family VI) were also not found.

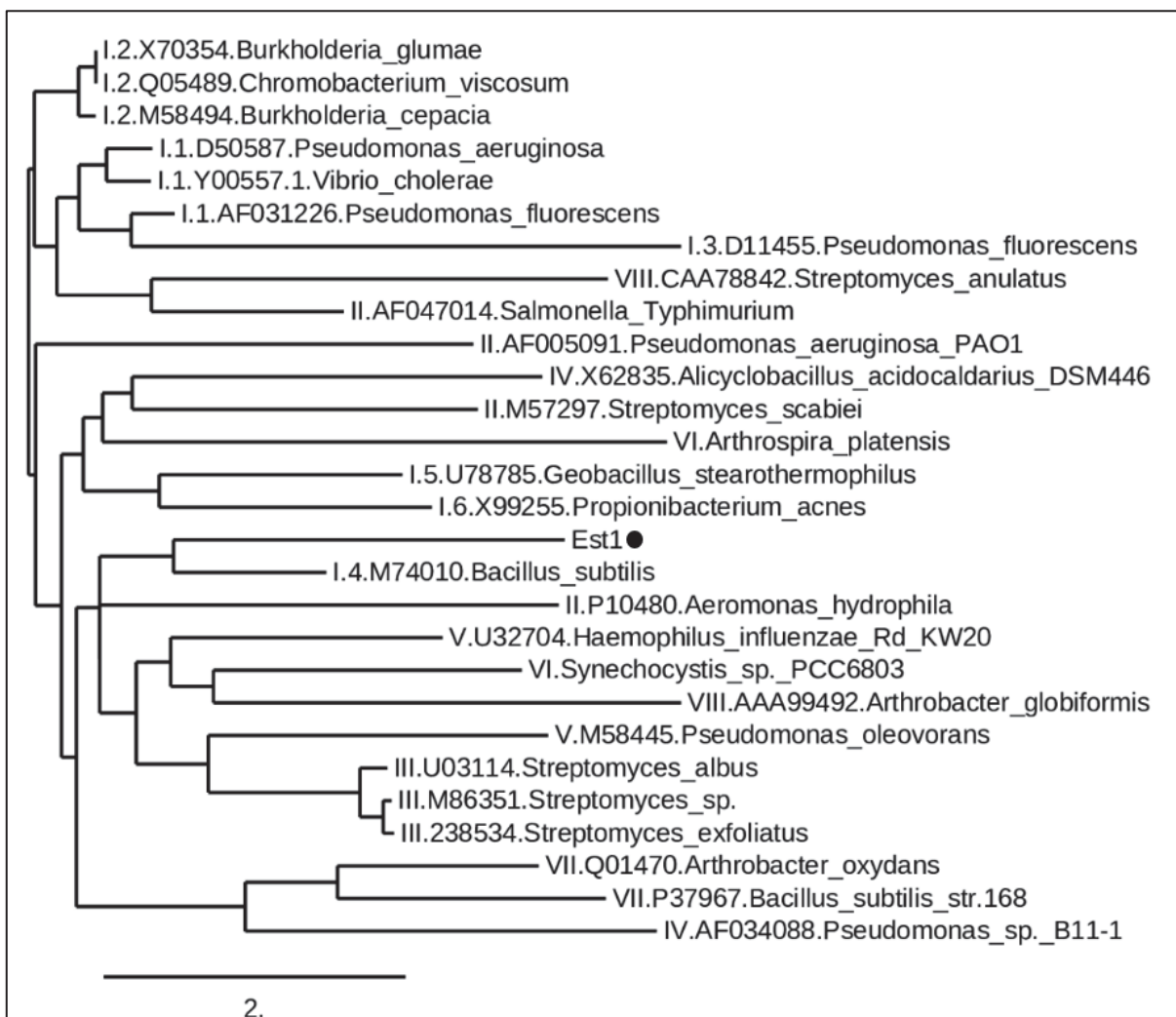


Figure 2. Phylogenetic tree of aligned sequences from different lipases/esterases families described by Arpigny and Jaeger (1999). The tree was created using Maximum-likelihood (ML) tree construction method. For each sequence the family number is indicated followed by accession number and species.

A NOVEL ESTERASE FAMILY?

Est1 showed the pentapeptide common in esterases sequences (Figure 1), however other conserved motifs typically present, described by Arpigny and Jaeger (1999), could not be detected in Est1. Analyzing other protein sequences similar to Est1 derived from other organisms (selected after a BLAST search using Est1 as query), we noticed the presence of two long conserved sequence motifs in proteins related to the Est1 sequence which are: HGMAEHXXRY and DXRGHGXTA besides the conventional pentapeptide sequence motif (Figure 3). Others small sequences also show conservation different from that regions described by Arpigny and Jaeger (Figure S1).

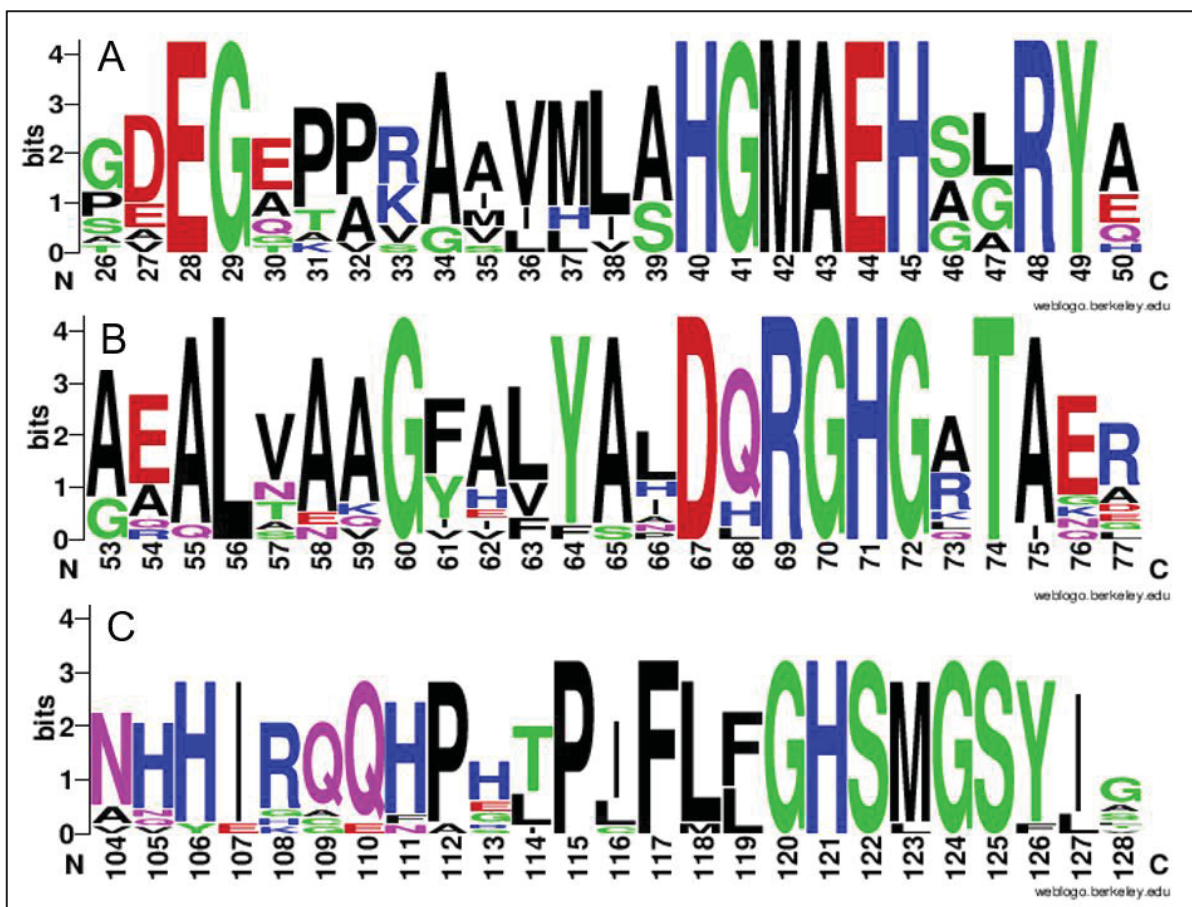


Figure 3. Three conserved motifs identified into Est1 protein sequence after comparison with the following proteins which presented alpha/beta hydrolases, similar to Est1: *Pseudomonas kuykendallii* (WP_090228553.1), *Acinetobacter baumannii* (SCY34300.1), *Enterobacter cloacae* (SAJ33106.1), *Stenotrophomonas rhizophila* (KWW14884.1), *Minicystis rosea* (APR74903.1), *Thiopseudomonas denitrificans* (WP_101497631.1), *Pseudomonas sp. ZH-FAD* (WP_100547479.1), *Pseudomonas toytomiensis* (WP_059391877.1), *Pseudomonas alcaliphila* (WP_064494389.1), *Gammaproteobacteria bacterium HGW-12* (PKM28184.1), *Endozoicomonas numazuensis* (WP_081869874.1), *Endozoicomonas elysicola* (WP_020581351.1), *Laceyella sediminis* (WP_106342408.1), *Moritella marina* (WP_019440375.1) and alpha/beta hydrolase common in *Pseudomonas* (MULTISPECIES - WP_017677741.1). (A) conserved region near the protein N-terminal – **HGMAEHXXRY** when X is an amino acid of any kind. (B) second motif, **DXRGHGXTA**. (C) the pentapeptide motif with two additional conserved residues, **GHXGXS**. Colors of the amino acids are related to chemical properties: polar (green, G,S,T,Y,C), neutral (purple, Q,N), basic (blue, K,R,H), acidic (red, D,E) and hydrophobic (black, A,V,L,I,P,W,F, M).

The HGMAEHXXRY and DXRGHGXTA motifs have not been described before and showed conservation among different species (Figure 4 and Figure S1). So, we speculate that Est1 and the related proteins analyzed could be members of a novel esterase family.

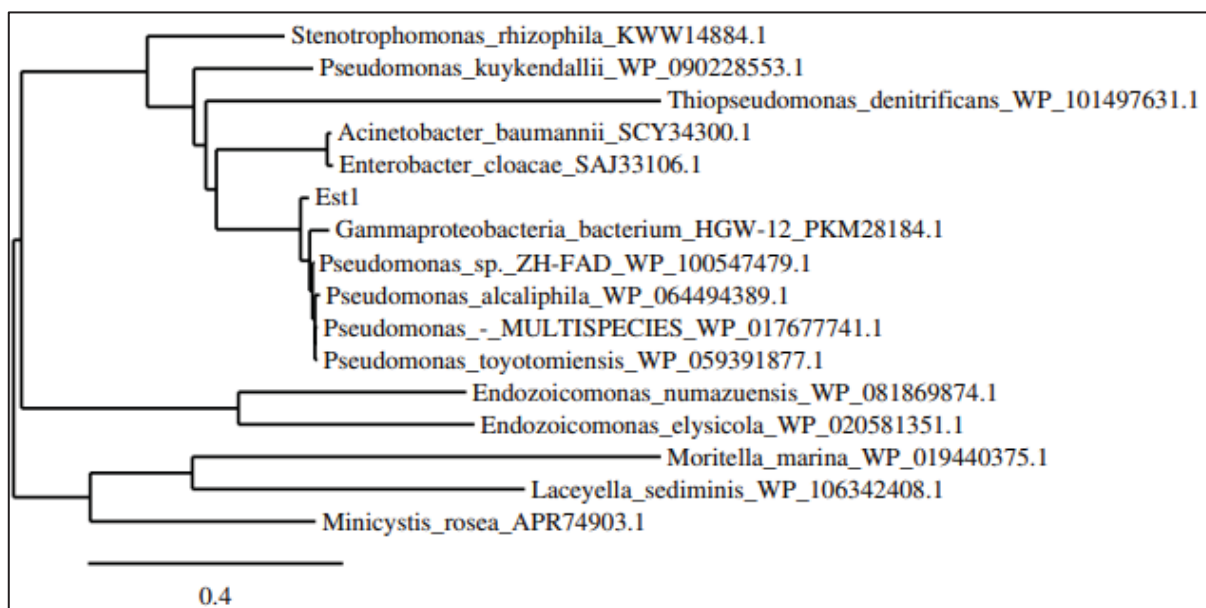


Figure 4. Phylogenetic tree of aligned protein sequences obtained after BLASTp assay using Est1 sequence as query. All the sequences have the conventional pentapeptide of esterases, alpha/beta hydrolases conserved domain and the two motifs HGMAEHXXRY and DXRGHGXTA. The specie and accession number of protein sequence were described in each leaf.

CONCLUSIONS

Est1, a cytoplasmatic esterase enzyme, was from *Pseudomonas sp.* genome. Phylogenetic analyses showed that Est1 is related to Family I of bacterial esterases, but do not present any described conserved domains except the common pentapeptide region. Sequences of proteins similar to Est1 are found in the GenBank database, and most of them are classified as alpha/beta hydrolases, without any enzymatic characterization. Our results suggest that Est1 and the similar enzymes could be members of a new esterase family, with two new conserved domains: HGMAEHXXRY and DXRGHGXTA, where X stands for any amino acid. Further studies must be made to confirm the existence of this new family, evaluating the activity of the other enzymes which are similar to Est1.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Science and Technology of Nitrogen Fixation/ CNPq/MCT, Fundação Araucária and CAPES.

REFERENCE

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res**, v. 25, 1997.

ARPIGNY, J. L.; JAEGER, K. E. Bacterial lipolytic enzymes: classification and properties. **The Biochemical Journal**, v. 343 Pt 1, p. 177–183, 1 out. 1999.

BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, p. 248–254, 7 maio 1976.

CROOKS, G. E. et al. WebLogo: a sequence logo generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, jun. 2004.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.

ISO, M. et al. Production of biodiesel fuel from triglycerides and alcohol using immobilized lipase. **Journal of Molecular Catalysis B: Enzymatic**, v. 16, n. 1, p. 53–58, 20 nov. 2001.

JAEGER, K. E.; DIJKSTRA, B. W.; REETZ, M. T. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. **Annual Review of Microbiology**, v. 53, p. 315–351, 1999.

KANG, H.-Y. et al. MELDB: A database for microbial esterases and lipases. **FEBS Letters**, v. 580, n. 11, p. 2736–2740, 15 maio 2006.

KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, v. 10, n. 6, p. 845–858, jun. 2015.

KHEADR, E. E.; VUILLEMARD, J. C.; EL-DEEB, S. A. Impact of liposome-encapsulated enzyme cocktails on cheddar cheese ripening. **Food Research International**, v. 36, n. 3, p. 241–252, 1 jan. 2003.

LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, ago. 1970.

VERMA, M. L.; AZMI, W.; KANWAR, S. S. Microbial lipases: at the interface of aqueous and non-aqueous media. A review. **Acta Microbiologica Et Immunologica Hungarica**, v. 55, n. 3, p. 265–294, set. 2008.

YU, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **Bioinformatics**, v. 26, n. 13, p. 1608–1615, 1 jul. 2010.

M.marina
L.sediminis
E.numazuensis
E.elysicola
T.denitrificans
M.rosea
S.rhizophila
A.baumannii
E.cloacae
P.kuykendallii
GammaproteobachGW-12
Est1
P.alcaliphila
P.toyotomiensis
P.sp.ZH-FAD
P.MULTI

FLDFVVSFGAFNSKFKPRTASDWLSRDPVQVDKYINDDYCGFPCSPQFWLDFMSGLISIS
LMNALSFGNFNKPFRPARTFFDWLSRDPQEVVDKYVQDPLCGEVPPTANFFCDLLTGKQMD
LIKQLTFSSYNRQFKPNRTEFDWLSRNPQAVDDYISDTLCGFPCSNQLWHLDFKGLLHIG
VIHQLTFFAGYNNNFKNRTEFDWLSRNPESVNRVYTDPLCGFRCNSQLWYDLFCGLQEIC
LLHALVFRPIQRSIQPQQSIHDWLSSDPAAVQQYLADPMC GFMCNTQLWLDVLDGLQRIT
LLTSLSFDDFNKAFKPNRTEFDWLSRDEAEVDKYIADPLCGFMVTTQLWVDLLDGTAEMS
LIEWLSFGSFNKAFFKPNRTEFDWLSRDPAEVDLYVSDPLCGFRCNSQLWLDLLQGLAQIS
LIDFLSFGSFNKAFFKPNRTEFDWLSRDPQEVVDKYVADPLCGFRCNSQLWVDLLDGLADIT
LIDFLSFGSFNKAFFKPNRTEFDWLSRDPQEVVDKYVADPLCGFRCNSQLWVDLLDGLADIT
LIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDAVYVADPLCGFRCNTNRLWMDLLDGLQDIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLLEGMQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
VIDFLSFGSFNKAFFKPNRTEFDWLSRDPAEVDKYVTDPLCGFVCTTQLWCDLDDGLQHIT
.: : * : :.* : **** : *: * : * * : .: : * .: * :

M.marina
L.sediminis
E.numazuensis
E.elysicola
T.denitrificans
M.rosea
S.rhizophila
A.baumannii
E.cloacae
P.kuykendallii
GammaproteobachGW-12
Est1
P.alcaliphila
P.toyotomiensis
P.sp.ZH-FAD
P.MULTI

QKSEIAKIPNQLPISIFSGDKDPVGLQKGVLALAQHITTINGCENVHFKLYPDGRHEMLN
EVEGLARIPNELPIFLLSGDKDPVQNGKGVKRTYQQFKQIGIKDVVLLKLYPDGRHEMLN
TISNLKRIEPELPLVLMGGEKDPVSAPNG-QQKLTAKLRKAGLKNVTLISLYPEGRHEMLN
SENRLKISGQLPVLILGGDQDPVSAPNG-LERLHKAFIRSGHSNTTLMLYPDARHELPH
PMQSMSQIHHQLPLFLFSGEQDPLHQGKR-LDLALRESGQSRVSVKLYPGRHEPLN
DPAGQALIPKDLPIYIFAGSEDPVGGKTRGLQQLVAAYRAAGIKDVSHKLYPGRHEMLN
QPSNLAQIDPNLPLMVLIGGECDPVSAGKR-LTHLADALRATGNRHVQLRVYPEARHEVLN
PPTHLRQIDADLPLLIIGGERDPVSQGKR-LGDLDALRGAGLRQVTLKTYPEARHELPH
PPTHLRQIDAGLPLLVIGGERDPVSQGRR-LGDLDALRGAGLRQVTLKTYPEARHELPH
PLEKLAQIDARLPLLVIGGECDPVSDGER-LEDLAEAMQRAGLRKVLKLYPDARHELPH
PPTNLAQIDADLPLLVIGGSRDPVSDGKR-LGDLAGALREAGVHDVQLKIYPEARHELLN
PPANLAQIDADLPLLVIGGSCDPVSDGKR-LGDLAGALREAGVRDVQLKIYPEARHELLN
PPRNLAQIDADLPLLVIGGSRDPVSDGKR-LGDLAGALREAGVRDVQLKIYPEARHELLN
PPSNLAQIDADLPLLVIGGSRDPVSDGKR-LGDLAGALREAGVRDVQLKIYPEARHELLN
PPSNLAQIDADLPLLVIGGSRDPVSDGKR-LGDLAGALREAGVRDVQLKIYPEARHELLN
PPSNLAQIDADLPLLVIGGSRDPVSDGKR-LSDLAGALREAGVRDVQLKIYPEARHELLN
* ** : :.* ** : * . ** *** : :

M.marina
L.sediminis
E.numazuensis
E.elysicola
T.denitrificans
M.rosea
S.rhizophila
A.baumannii
E.cloacae
P.kuykendallii
GammaproteobachGW-12
Est1
P.alcaliphila
P.toyotomiensis
P.sp.ZH-FAD
P.MULTI

ESNASEVFKDVTDWVTAKVS-----
ETNKEEVYRDILAWLDTRVARTSGHSEM-----
EMNSQQVINRLLQWINQTAMSRCECPAEYTPVVTES
ETNSSEVINQLCTWLKKS-----
ERHRARVFTDI IAWIRQQCD-GLP-----
ETNRDEVTRDVI AWLDAHAG-----
ETNRDEVTADILGWLEQALALGRP-----
ESNRDAVTQDLIDWLEQALRHRRDH-----
ESNRDAVTQDLIDWLEQALRHRRDH-----
ETNRDEVTAFLIDWLQQALDSSAK-----
ESNRDEVTAHLIDWLQQALSQGRSPTKECP-----
ESNRDEVTAHLIDWLQQALSHGRK-----
ESNRDEVTAHLIDWLQQALSHGRSPIKECP-----
ESNRDEVTAHLIDWLQQALSHGRSPIKECP-----
ESNRDEVTAHLIDWLQQALSHGRSPIKECP-----
ESNRDEVTAHLIDWLQQALSHGRS-----
* : * : * :

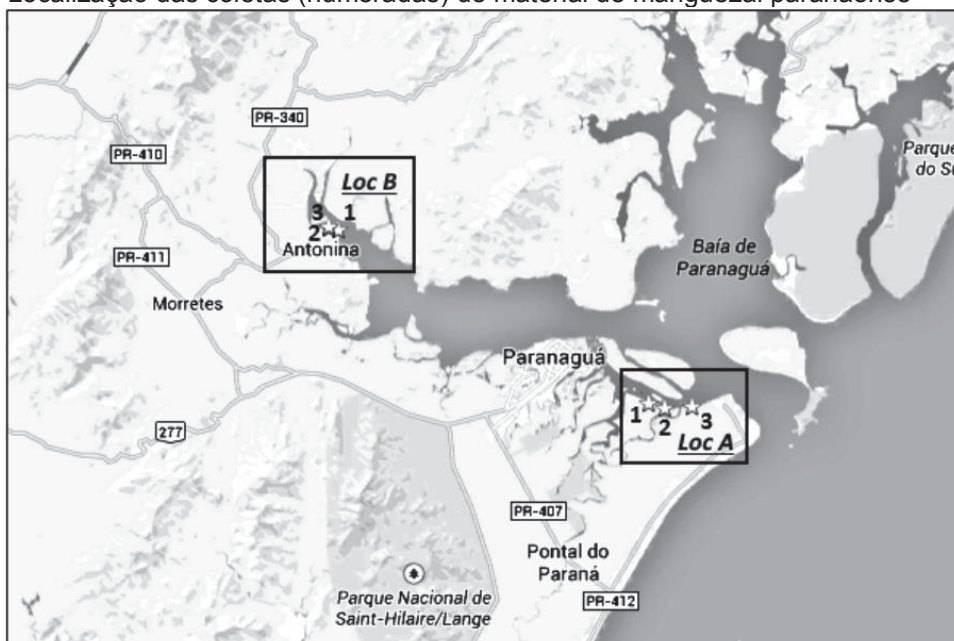
6 ATIVIDADES REALIZADAS

Nesta seção será descrita, de forma resumida, os demais dados obtidos, durante o período de doutorado.

6.1 ISOLAMENTO DE MICRORGANISMOS

A amostra estudada neste projeto foi coletada pelo doutor Denny Marcel Seccon e colaboradores, em fevereiro de 2014. A coleta foi feita na baía de Paranaguá, litoral do estado do Paraná, em pontos e períodos diferentes (figura 4). As coordenadas dos pontos de coleta são as seguintes: LocB, 1 - 25°25'02.5"S 48°42'29.9"W, 2 - 25°24'59.8"S 48°42'57.2"W, 3 - 25°30'31.4"S 48°28'07.4"W, LocA, 1 - 25°33'04.5"S 48°26'35.2"W, 2 - 25°33'19.1"S 48°25'46.1"W, 3 - 25°33'15.6"S 48°24'20.2"W.

Figura 4 – Localização das coletas (numeradas) de material de mangueza paranaense



FONTE: (SECCON, 2016)

Optou-se em não se trabalhar com cada amostra em isolado, e sim em uma mistura delas. Dessa forma, porções iguais de cada uma das amostras coletadas foram reunidas e o homogeneizado formado foi utilizado como fonte de microrganismos para os ensaios deste projeto.

6.2 MEIOS E CONDIÇÕES DE CULTIVO

Os meios de cultura utilizados para a seleção de isolados com interesse biotecnológico estão descritos na Tabela 3. As formulações foram feitas pela doutora Ana Paula Andrezza (dados não publicados) baseadas em meios descritos por HANKIN e ANAGNOSTAKIS (1977), AMORE e colaboradores (2012) e YAN e colaboradores (2013). Estes meios foram formulados com o intuito de facilitar o surgimento de isolados que apresentassem enzimas de interesse, uma vez que as fontes de carbono ou nitrogênio escolhidos não são de fácil utilização, exigindo uma maquinaria enzimática mínima para torná-las assimiláveis metabolicamente.

Tabela 3 – composição dos meios seletivos

COMPOSIÇÃO BASE		
K ₂ HPO ₄ : 1,6 g/L	Extrato de levedura: 0,2 g/L	H ₃ BO ₃ : 1,4.10 ⁻² g/L
KH ₂ PO ₄ : 0,2 g/L	Casamino ácidos: 0,2 g/L	CuSO ₄ .5H ₂ O: 4,0.10 ⁻⁴ g/L
MgSO ₄ .7H ₂ O: 0,2 g/L	MnSO ₄ .H ₂ O: 1,2 .10 ⁻² g/L	ZnSO ₄ .7H ₂ O: 1,2.10 ⁻³ g/L
FeSO ₄ .7H ₂ O: 0,01 g/L	CaCl ₂ .2H ₂ O: 0,02 g/L	Na ₂ MoO ₄ .2H ₂ O: 0,01 g/L
NaCl: 20g/L (2%) ou 50g/L (5%)		Ágar bacteriológico*: 15 g/L
<i>Meio para seleção de:</i>	<i>Fonte de carbono adicionada</i>	<i>Fonte de nitrogênio adicionada</i>
AMILASES	Amido de milho: 5 g/L	(NH ₄) ₂ SO ₄ : 1 g/L
CELULASES	CMC: 5 g/L	(NH ₄) ₂ SO ₄ : 1 g/L
LIPASES	Tributirina: 10 g/L	(NH ₄) ₂ SO ₄ : 1 g/L
PROTEASES	Glucose: 4 g/L	Caseína: 3 g/L
XILANASES	Xilana: 5 g/L	(NH ₄) ₂ SO ₄ : 1 g/L

FONTES: ANDREAZZA (não publicado).

LEGENDA: *apenas para meio sólido.

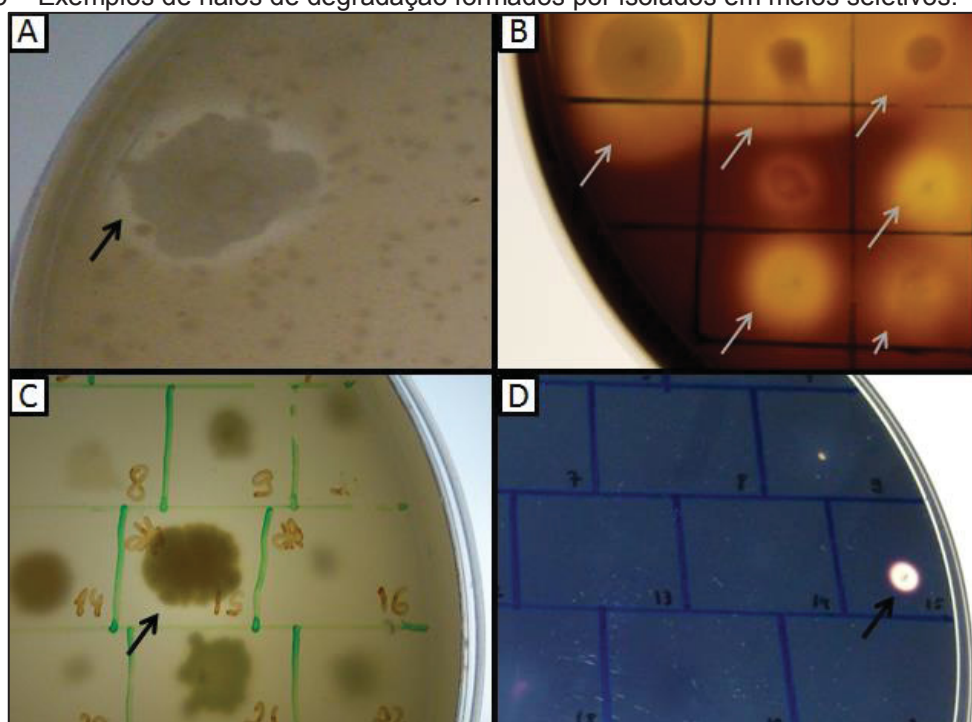
6.3 OBTENÇÃO DE ISOLADOS

Após a homogeneização da amostra, 1g do solo de manguezal foi adicionado em 5 mL de água destilada contendo 2 ou 5% de NaCl. Como a salinidade do manguezal varia periodicamente, optou-se por estabelecer essas duas condições de cultivo. Essas soluções foram diluídas nas proporções 1:5, 1:25 e 1:125 (em salina 2 ou 5%) e plaqueadas nos meios seletivos contendo 2 e 5% de NaCl, sendo mantidas

a 30°C por 48 horas. A identificação de possíveis isolados produtores de enzimas foi feita através da formação de halos de degradação. Posteriormente, as colônias foram reisoladas no mesmo meio de cultura de origem para garantir a pureza dos microrganismos e se estes ainda apresentavam halos de degradação após crescimento.

Após as etapas de cultivo, houve o crescimento de centenas de microrganismos, porém apenas parte desses isolados apresentaram halos de degradação característicos (Figura 5). Após repiques para avaliação de pureza e de formação de halos, esses isolados foram estocados em glicerol 50% ou em *storage milk* a -80°C.

Figura 5 – Exemplos de halos de degradação formados por isolados em meios seletivos.



FONTE: o autor (2018).

LEGENDA: setas indicam os halos de degradação. (A) placa contendo meio para proteases (5 g/L de caseína). (B) e (D) placas contendo meio para celulases (5 g/L de CMC), após coloração com iodo (revelador de halos). (C) placa contendo meio para xilanases (5 g/L de xilana).

O número de isolados selecionados, que mantiveram halos de degradação após repiques, está mostrado na Tabela 4.

Tabela 4 – Quantidade de isolados apresentando halos de degradação

MEIO SELETIVO	Nº DE ISOLADOS
Amilase	4 ¹
CMCase ²	4 ¹
Protease	2 ¹
Lipase	1
Xilanase	7 ¹ e 2 ³
TOTAL	20

LEGENDA: (1) o meio seletivo apresentava 5% de NaCl. (2) Como utiliza-se a carboximetilcelulose (CMC), muitos artigos utilizam o termo CMCases ao invés de celulases. (3) isolados obtidos após cultivo em meio seletivo líquido contendo 2% de NaCl.

Em seguida, os isolados selecionados tiveram seus genes de RNA ribossômico 16S amplificados e sequenciados.

Tabela 5 – Identificação de isolados baseado na identidade da sequência de genes de RNA ribossômico 16S com sequências do banco de dados RDP (*Ribosomal DatabAse Project*, WANG *et al.*, 2007).

ID ISOLADO	Nº de bases sequenciadas	Identidade	Maior similaridade	Atividade detectada
AM-1 A	906 pb	96%	<i>Leeuwnhoekiella marinoflava</i>	Amilase e protease
AM-1 B	959 pb	98%	<i>Marinomonas sp.</i>	Amilase
AM-8	1390 pb	98%	<i>Bacillus sp.</i>	Amilase
AM-9	1210 pb	99%	<i>Exiguobacterium indicum</i>	Amilase CMCase e xilanase
AM-16 B	1399 pb	99%	<i>Bacillus sp.</i>	Amilase
AM-19	881 pb	100%	<i>Bacillus hwajinpoensis</i>	Amilase
AM-25	1410 pb	99%	<i>Bacillus sp.</i>	Amilase
Lip1	534 pb	100%	<i>Bacillus megaterium</i>	Lipase
PR-1 A	873 pb	100	<i>Achromobacter cholinophagum</i>	Protease
PR-2	1398 pb	100%	<i>Bacillus subtilis</i>	Amilase, protease, CMCase e xilanase
XL-2*	1382 pb	96%	<i>Bacillus sp.</i>	Xilanase, CMCase e amilase
XS-3	1415 pb	99%	<i>Pseudoalteromonas sp.</i>	Amilase, xilanase e protease

Tabela 5 – Identificação de isolados baseado na identidade da sequência de genes de RNA ribossômico 16S com sequências do banco de dados RDP (*Ribosomal Database Project*, WANG *et al.*, 2007).

ID ISOLADO	Nº de bases sequenciadas	Identidade	Maior similaridade	conclusão Atividade detectada
XS-12	1400 pb	99%	<i>Pseudoalteromonas sp.</i>	Protease e xilanase
XS-13	1401 pb	94%	<i>Pseudoalteromonas sp.</i>	Amilase, protease e xilanase
XS-14	704 pb	97%	<i>Echinicola shivajiensis</i>	xilanase
XS-15	700 pb	97%	<i>Isoptericola sp.</i>	Amilase e xilanase
XS-16	934 pb	92%	<i>Pseudomonas pseudoalcaligenes</i>	Xilanase
XS-30	1397 pb	99%	<i>Pseudoalteromonas sp.</i>	CMCase, protease e xilanase

*isolado XL-2, após sequenciamento genômico, foi classificado como *Bacillus safensis* BRM1.

6.4 SEQUENCIAMENTO GENÔMICO

O isolado XL-2 foi submetido a sequenciamento genômico depois de mostrar atividade contra amido, CMC e xilana, após crescimento em placas contendo estes substratos. Maiores detalhes e resultados estão descritos na seção 3, onde consta o primeiro manuscrito obtido.

Inicialmente, os dados de sequenciamento gerados pelo equipamento Illumina MiSeq foram analisados através da plataforma CLC Genomics Workbench 6.5.1 (CLC bio, Aarhus, Dinamarca), originando uma montagem contendo 16 contigs. Com o auxílio do professor Dieval Guizelini, do setor de Educação Profissional e Tecnológica da UFPR – Curitiba, as sequências geradas foram reanalisadas utilizando os programas Velvet 1.2.07 (ZERBINO; BIRNEY, 2008) e Masurca 2.3.2 (ZIMIN *et al.*, 2013). Os contigs formados foram ordenados com o jContigSort (GUIZELINI *et al.*, 2011) e alguns gaps fechados com o FGAP (PIRO *et al.*, 2014).

Feito isso, obteve-se um contig único contendo 3,72 Gb, com 41,7% de conteúdo G+C, e 5 *gaps*. Depois, o doutor Rodrigo Luis Alves Cardoso, utilizando scripts *in house*, conseguiu fechar esses *gaps*, concluindo assim o sequenciamento genômico. A anotação foi feita pelo próprio NCBI, durante o processo de depósito do genoma, resultando em 3.699 sequências codificadoras. O isolado XL-2 foi

descrito como sendo uma espécie de *Bacillus safensis*, onde recebeu a denominação BRM1 (*Brazilian Mangrove*).

Outros dois isolados, XS-14 e XS-15 passaram pela tentativa de sequenciamento genômico, mas foi falho. Acredita-se que o material genético não estava com a qualidade exigida pelo sequenciador ou houve a contaminação de outro organismo, uma vez que análises das poucas sequências geradas pelo equipamento mostrou a presença de *Pseudomonas sp.*, não condizendo com o esperado para esses dois isolados (*Isoptricola sp.* e *Echinocola sp.*, respectivamente). Como as estirpes XS-14 e XS-15 conseguem crescer em meio contendo 2 e 5% de NaCl, acredita-se que são originadas do manguezal.

6.5 SELEÇÃO DE ENZIMAS

As sequências obtidas durante o sequenciamento dos isolados XL-2, XS-14 e XS-15 foram analisadas com a intenção de encontrar enzimas com algum uso industrial. A tabela a seguir (Tabela 6) mostra quais foram as enzimas selecionadas para estudo nesse projeto. Como descrito no primeiro manuscrito (seção 3), o isolado XL-2 teve todo seu genoma sequenciado. Já para XS-14 e XS-15, a quantidade de *reads* obtidos permitiu sequenciar apenas 109.379 e 195.509 pares de bases, respectivamente. Mesmo assim, foram selecionadas 7 enzimas neste material que não apresentavam similaridade com outras enzimas já descritas na literatura. Infelizmente, por falhas técnicas e/ou de manipulação, houve a clonagem apenas do gene referente a enzima Est1, que permitiu a escrita do segundo manuscrito.

Tabela 6 – Enzimas selecionadas para estudo em sequências genéticas de 3 isolados de manguezal paranaense

Isolado	Gene	Possível atividade	Tamanho (pb)	Massa molecular (kDa)	Domínios conservados	pI
	MSXIIA	Xilanase	1250	47,9	Domínio catalítico de α -amilase	6,2
	MSAmiA	Amilase	1156	40,9	Superfamília das zinco-peptidases	5,4
XL2	MSProtA	Protease	2431	86,5	Superfamília 8 das peptidases	6,8
	MSCeIA	Endoglucanase	1851	69,6	Superfamília GH9 + CBM 3	5,1
	MSCeIB	Exoglucanase	2106	78,6	Superfamília GH48	4,9
	X14Prot1	Protease	908	31,1	Superfamília das zinco-peptidases	6,4
XS14	X14Lip1	Lipase	988	34,9	Superfamília das esterases/lipases	6,3
	X14Lip2	Lipase	991	35,0	Superfamília das esterases/lipases	6,6
	X15Lip1	Lipase	1129	41,4	Superfamília das esterases/lipases	7,2
XS15	X15Lip2	Lipase	964	33,5	Superfamília das esterases/lipases	9,0
	X15Lip3	Lipase	1078	36,9	Superfamília das esterases/lipases	9,9
	X15Ami1	Amilase	1892	67,3	Superfamília GH15	5,3

7 CONCLUSÃO

O presente projeto pôde contribuir com mais informações biotecnológicas relativas ao manguezal, um ecossistema complexo e ainda pouco estudado. Amostras desse bioma foram coletadas da baía de Paranaguá e permitiram o isolamento de 20 isolados. Dentre eles, o isolado XL-2 (*Bacillus safensis* BRM1) teve seu genoma totalmente sequenciado, permitindo seu depósito no GenBank. Assim, houve dados suficientes para escrever o primeiro manuscrito. Já o segundo manuscrito foi conseguido a partir da análise da esterase Est1, identificada no genoma parcial de outro isolado (XS-14). Além disso, sua sequência permitiu a identificação de novas regiões conservadas, sugerindo a criação de uma nova família de esterases.

De forma geral, os objetivos propostos pelo trabalho foram alcançados. Realizou-se o isolamento de microrganismos presentes em amostra de manguezal através do cultivo em meios seletivos, ocorrendo posterior classificação taxonômica através do gene de RNA ribossômico 16S. Um dos isolados teve seu genoma totalmente sequenciado (XL-2), onde foi possível identificar enzimas com potencial interesse industrial. Apesar de inúmeras clonagens feitas, apenas a esterase Est1 pôde ser clonada e expressa de maneira satisfatória.

O presente trabalho propiciou grande conhecimento ao aluno, uma vez que várias áreas foram abordadas: microbiologia básica, cultivo seletivo, análises genômicas e de bioinformática, *genome mining*, expressão de proteínas e análises de atividade enzimática. Os resultados obtidos proporcionaram a escrita de dois manuscritos que logo serão publicados, beneficiando o programa de pós-graduação que permitiu a realização deste trabalho.

REFERÊNCIAS

- ADRIO, J. L.; DEMAIN, A. L. Microbial Enzymes: Tools for Biotechnological Processes. **Biomolecules**, v. 4, n. 1, p. 117–139, 16 jan. 2014.
- ALETI, G.; SESSITSCH, A.; BRADER, G. Genome mining: Prediction of lipopeptides and polyketides from Bacillus and related Firmicutes. **Computational and Structural Biotechnology Journal**, v. 13, p. 192–203, 24 mar. 2015.
- ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res**, v. 25, 1997.
- AMORE, A. et al. Cloning and recombinant expression of a cellulase from the cellulolytic strain Streptomyces sp. G12 isolated from compost. **Microbial Cell Factories**, v. 11, p. 164, 26 dez. 2012.
- ANDREOTE, F. D. et al. The Microbiome of Brazilian Mangrove Sediments as Revealed by Metagenomics. **PLoS ONE**, v. 7, n. 6, 21 jun. 2012.
- ANGIUOLI, S. V. et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. **Omics: A Journal of Integrative Biology**, v. 12, n. 2, p. 137–141, jun. 2008.
- ARPIGNY, J. L.; JAEGER, K. E. Bacterial lipolytic enzymes: classification and properties. **The Biochemical Journal**, v. 343 Pt 1, p. 177–183, 1 out. 1999.
- AZIZ, R. K. et al. The RAST Server: Rapid Annotations using Subsystems Technology. **BMC Genomics**, v. 9, p. 75, 8 fev. 2008.
- BABCOCK, L. W. **Method of producing fermentable sugars and alcohol from wood**, 26 abr. 1932. Disponível em: <<https://patents.google.com/patent/US1855464A/en>>. Acesso em: 13 abr. 2018
- BABICKI, S. et al. Heatmapper: web-enabled heat mapping for all. **Nucleic Acids Research**, v. 44, n. W1, p. W147–W153, 8 jul. 2016.
- BCC RESEARCH. **Global Markets for Enzymes in Industrial Applications**. Estados Unidos: [s.n.].
- BLANQUET, R. et al. **Mangrove Ecology: A Manual for a Field Course**. Washington, DC, USA: Smithsonian Institution, 1996.
- BON, E. P. S.; FERRARA, M. A. **Bioethanol production via enzymatic hydrolysis of cellulosic biomass**, 2007. Disponível em: <<http://www.fao.org/biotech/docs/bon.pdf>>. Acesso em: 3 out. 2018
- BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, p. 248–254, 7 maio 1976.

BRUCE, T. et al. Abrolhos Bank Reef Health Evaluated by Means of Water Quality, Microbial Diversity, Benthic Cover, and Fish Biomass Data. **PLoS ONE**, v. 7, n. 6, 5 jun. 2012.

CANTAREL, B. L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research**, v. 37, n. Database issue, p. D233-238, jan. 2009.

CARDER, J. H. Detection and quantitation of cellulase by Congo red staining of substrates in a cup-plate diffusion assay. **Analytical Biochemistry**, v. 153, n. 1, p. 75–79, 15 fev. 1986.

CARVER, T. et al. DNAPlotter: circular and linear interactive genome visualization. **Bioinformatics**, v. 25, n. 1, p. 119–120, 1 jan. 2009.

CASTRO, R. A. et al. Isolation and enzyme bioprospection of endophytic bacteria associated with plants of Brazilian mangrove ecosystem. **SpringerPlus**, v. 3, 28 jul. 2014.

CAZYPEDIA. **CAZYPedia**. Disponível em: <http://www.cazypedia.org/index.php/Main_Page>. Acesso em: 15 jan. 2018.

CENSONI, L. **Dinâmica molecular e redes complexas no estudo da difusão térmica em xilanases da família 11** Universidade de São Paulo, , 2013. . Acesso em: 23 jan. 2018

CHARNOCK, S. J.; MCCLEARY, B. V. Enzymes: Industrial and Analytical Applications. p. 5, 2006.

COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **FEMS Microbiology Reviews**, v. 29, n. 1, p. 3–23, 1 jan. 2005.

COMMICHAU, F. M. et al. Glutamate metabolism in *Bacillus subtilis*: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. **Journal of Bacteriology**, v. 190, n. 10, p. 3557–3564, maio 2008.

CROOKS, G. E. et al. WebLogo: a sequence logo generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, jun. 2004.

DA FONSECA, F. S. A. et al. Identification of oxidoreductases from the petroleum *Bacillus safensis* strain. **Biotechnology Reports**, v. 8, p. 152–159, 1 dez. 2015.

Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, v. 44, n. Database issue, p. D7–D19, 4 jan. 2016.

DE SOUZA, P. R. et al. Acetylcholinesterase of mangrove oyster *Crassostrea rhizophorae*: A highly thermostable enzyme with promising features for estuarine biomonitoring. **Aquatic Toxicology**, v. 197, p. 109–121, 1 abr. 2018.

- DIEGUES, A. C. Povos e Águas - Inventário de áreas úmidas brasileiras. n. 2, p. 15–18, 2002.
- DIELE, K.; SIMITH, D. Salinity tolerance of northern Brazilian mangrove crab larvae, *Ucides cordatus* (Ocypodidae): Necessity for larval export? **Estuarine, Coastal and Shelf Science**, v. 68, p. 600–608, 1 jul. 2006.
- DOMINGOS, D. F. et al. Draft Genome Sequence of *Bacillus pumilus* CCMA-560, Isolated from an Oil-Contaminated Mangrove Swamp. **Genome Announcements**, v. 1, n. 5, 12 set. 2013.
- DOS SANTOS, J. A. et al. Marine-derived fungus *Aspergillus* cf. *tubingensis* LAMA1 31: a new genetic resource for xylanase production. **AMB Express**, v. 6, 24 mar. 2016.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.
- FALKOW, S. et al. **The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria**. [s.l.] Springer Science & Business Media, 2006.
- FOODY, B. E.; FOODY, K. J. Development of an integrated system for producing ethanol from biomass. **Energy from biomass and wastes XIV**, 1991.
- GAO, Z. et al. A novel salt-tolerant endo-beta-1,4-glucanase Cel5A in *Vibrio* sp. G21 isolated from mangrove soil. **Applied Microbiology and Biotechnology**, v. 87, n. 4, p. 1373–1382, jul. 2010.
- GHANEM, E. H.; AL-SAYED, H. A.; SALEH, K. M. An alkalophilic thermostable lipase produced by a new isolate of *Bacillus alcalophilus*. **World Journal of Microbiology and Biotechnology**, v. 16, n. 5, p. 459–464, 1 jul. 2000.
- GNERRE, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. **Proceedings of the National Academy of Sciences**, v. 108, n. 4, p. 1513–1518, 25 jan. 2011.
- GUIZELINI, D. et al. jContigSort: a new computer application for contigs ordering. In: **7th International Conference of The Brazilian Association for Bioinformatics and Computacional Biology: abstract book**. Florianópolis: [s.n.].
- GURUNG, N. et al. **A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond**. Research article. Disponível em: <<https://www.hindawi.com/journals/bmri/2013/329121/>>. Acesso em: 4 jan. 2018.
- HANKIN, L.; ANAGNOSTAKIS, S. L. Solid media containing carboxymethylcellulose to detect CX cellulose activity of micro-organisms. **Journal of General Microbiology**, v. 98, n. 1, p. 109–115, jan. 1977.
- HENRISSAT, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal**, v. 280, n. Pt 2, p. 309–316, 1 dez. 1991.

HUERTA-CEPAS, J. et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. **Molecular Biology and Evolution**, v. 34, n. 8, p. 2115–2122, 1 ago. 2017.

HUTCHISON, C. A. DNA sequencing: bench to bedside and beyond†. **Nucleic Acids Research**, v. 35, n. 18, p. 6227–6237, set. 2007.

ILLANES, A. et al. Recent trends in biocatalysis engineering. **Bioresource Technology**, v. 115, p. 48–57, jul. 2012.

ISO, M. et al. Production of biodiesel fuel from triglycerides and alcohol using immobilized lipase. **Journal of Molecular Catalysis B: Enzymatic**, v. 16, n. 1, p. 53–58, 20 nov. 2001.

ITO, S. Alkaline cellulases from alkaliphilic Bacillus: Enzymatic properties, genetics, and application to detergents. **Extremophiles**, v. 1, n. 2, p. 61–66, maio 1997.

JAEGER, K. E.; DIJKSTRA, B. W.; REETZ, M. T. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. **Annual Review of Microbiology**, v. 53, p. 315–351, 1999.

JIMÉNEZ, D. J. et al. Compositional profile of α/β -hydrolase fold proteins in mangrove soil metagenomes: prevalence of epoxide hydrolases and haloalkane dehalogenases in oil-contaminated sites. **Microbial Biotechnology**, v. 8, n. 3, p. 604–613, maio 2015.

JÚNIOR, F. L. S. et al. Endo- and exoglucanase activities in bacteria from mangrove sediment. **Brazilian Journal of Microbiology**, v. 44, n. 3, p. 969–976, 15 jan. 2014.

KANG, H.-Y. et al. MELDB: A database for microbial esterases and lipases. **FEBS Letters**, v. 580, n. 11, p. 2736–2740, 15 maio 2006.

KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, v. 10, n. 6, p. 845–858, jun. 2015.

KHASIN, A.; ALCHANATI, I.; SHOHAM, Y. Purification and characterization of a thermostable xylanase from *Bacillus stearothermophilus* T-6. **Applied and Environmental Microbiology**, v. 59, n. 6, p. 1725–1730, 6 jan. 1993.

KHEADR, E. E.; VUILLEMARD, J. C.; EL-DEEB, S. A. Impact of liposome-encapsulated enzyme cocktails on cheddar cheese ripening. **Food Research International**, v. 36, n. 3, p. 241–252, 1 jan. 2003.

KUNST, F. et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. **Nature**, v. 390, n. 6657, p. 249–256, 20 nov. 1997.

LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, ago. 1970.

LAGESSEN, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. **Nucleic Acids Research**, v. 35, n. 9, p. 3100–3108, 5 jan. 2007.

LASLETT, D.; CANBACK, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, v. 32, n. 1, p. 11–16, 2004.

LATEEF, A. et al. Green synthesis of silver nanoparticles using keratinase obtained from a strain of *Bacillus safensis* LAU 13. **International Nano Letters**, v. 5, n. 1, p. 29–35, 1 mar. 2015.

LI, S. et al. TECHNOLOGY PROSPECTING ON ENZYMES: APPLICATION, MARKETING AND ENGINEERING. **Computational and Structural Biotechnology Journal**, v. 2, n. 3, p. e201209017, 1 set. 2012.

LIU, X.; KOKARE, C. Chapter 11 - Microbial Enzymes of Use in Industry. In: BRAHMACHARI, G. (Ed.). . **Biotechnology of Microbial Enzymes**. [s.l.] Academic Press, 2017. p. 267–298.

MACIEL, N. C. Alguns aspectos da ecologia do manguezal. n. 3, p. 9–37, 1991.

MAI, Z. et al. Cloning and characterization of a novel GH44 family endoglucanase from mangrove soil metagenomic library. **Biotechnology Letters**, v. 36, n. 8, p. 1701–1709, ago. 2014.

MAROLDI, M. M. C. et al. Potential of Mangrove-Associated Endophytic Fungi for Production of Carbohydrases with High Saccharification Efficiency. **Applied Biochemistry and Biotechnology**, v. 184, n. 3, p. 806–820, mar. 2018.

MASUDA, T. et al. Atomic resolution structure of serine protease proteinase K at ambient temperature. **Scientific Reports**, v. 7, p. 45604, 31 mar. 2017.

MILLER, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. **Analytical Chemistry**, v. 31, n. 3, p. 426–428, 1 mar. 1959.

NATH, A. et al. Purification and Characterization of β -Galactosidase Synthesized from *Bacillus safensis* (JUCHE 1). **Industrial & Engineering Chemistry Research**, v. 52, n. 33, p. 11663–11672, 21 ago. 2013.

National Geographic Magazine. [s.l: s.n.].

NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. Disponível em: <<http://www.macmillanlearning.com/catalog/Product/lehningerprinciplesofbiochemistry-seventhedition-nelson>>. Acesso em: 4 jan. 2018.

NOVOZYMES. **Novozymes Annual Report 2014**. Disponível em: <https://s21.q4cdn.com/655485906/files/doc_financials/annual_english/NovozymesReport2014.pdf>. Acesso em: 12 jan. 2018.

NOVOZYMES. **Novozymes Annual Report 2015**. Disponível em: <https://s21.q4cdn.com/655485906/files/doc_financials/annual_english/NovozymesReport2015.pdf>. Acesso em: 12 jan. 2018.

NOVOZYMES. **Novozymes Annual Report 2016**. Disponível em: <<https://report2016.novozymes.com/>>. Acesso em: 12 jan. 2018.

OLLIS, D. L. et al. The alpha/beta hydrolase fold. **Protein Engineering**, v. 5, n. 3, p. 197–211, abr. 1992.

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, n. 4, p. 404–408, abr. 2009.

PARSIEGLA, G. et al. Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. **Biochemistry**, v. 39, n. 37, p. 11238–11246, 19 set. 2000.

PIRO, V. C. et al. FGAP: an automated gap closing tool. **BMC Research Notes**, v. 7, n. 1, p. 1–5, 2014.

RASTOGI, G. et al. Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. **Bioresource Technology**, v. 101, n. 22, p. 8798–8806, nov. 2010.

RODRIGUEZ-R, L. M.; KONSTANTINIDIS, K. T. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. 27 mar. 2016.

SÁ, A. L. B. et al. Screening of endoglucanase-producing bacteria in the saline rhizosphere of *Rhizophora mangle*. **Brazilian Journal of Microbiology**, v. 45, n. 1, p. 193–197, 19 maio 2014.

SCHAEFFER-NOVELLI, Y. Manguezal: ecossistema entre a terra e o mar. 1995.

SHELLER, H. V.; ULVSKOV, P. Hemicelluloses. **Annual Review of Plant Biology**, v. 61, p. 263–289, 2010.

SECCON, D. M. **Análise metagenômica do microbioma de mangue do litoral do Paraná** Tese de doutorado, , 2016. Disponível em: <<http://acervodigital.ufpr.br/bitstream/handle/1884/45951/R%20-%20T%20-%20DENNY%20MARCEL%20SECCON.pdf?sequence=1&isAllowed=y>>. Acesso em: 26 jan. 2018

SINGH, D. et al. Molecular Docking of Laccase Protein from *Bacillus Safensis* DSKK5 Isolated from Earthworm gut: A Novel Method to Study dye Decolorization Potential. **Water, Air, & Soil Pollution**, v. 225, n. 11, p. 2175, 1 nov. 2014.

SINGH, R. S.; SINGH, R. P.; YADAV, M. Molecular and biochemical characterization of a new endoinulinase producing bacterial strain of *Bacillus safensis* AS-08. **Biologia**, v. 68, n. 6, p. 1028–1033, 1 dez. 2013.

SOARES, F. L. et al. A Novel Multifunctional β -N-Acetylhexosaminidase Revealed through Metagenomics of an Oil-Spilled Mangrove. **Bioengineering**, v. 4, n. 3, 9 jul. 2017.

TAHERZADEH, M. J.; KARIMI, K. ENZYMATIC-BASED HYDROLYSIS PROCESSES FOR ETHANOL. **BioResources**, v. 2, n. 4, p. 707–738, 1 nov. 2007.

TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, dez. 2013.

THATOI, H. et al. Biodiversity and biotechnological potential of microorganisms from mangrove ecosystems: a review. **Annals of Microbiology**, v. 63, n. 1, p. 1–19, mar. 2013.

THOMPSON, C. E. et al. A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. **AMB Express**, v. 3, p. 65, 26 out. 2013.

U.S. DOE. U.S. Department of Energy Bioenergy Research Centers. p. 32, 2018.

VERMA, M. L.; AZMI, W.; KANWAR, S. S. Microbial lipases: at the interface of aqueous and non-aqueous media. A review. **Acta Microbiologica Et Immunologica Hungarica**, v. 55, n. 3, p. 265–294, set. 2008.

WAKARCHUK, W. W. et al. Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. **Protein Science: A Publication of the Protein Society**, v. 3, n. 3, p. 467–475, mar. 1994.

YAN, X. et al. Discovery of (hemi-) cellulase genes in a metagenomic library from a biogas digester using 454 pyrosequencing. **Applied Microbiology and Biotechnology**, v. 97, n. 18, p. 8173–8182, set. 2013.

YU, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **Bioinformatics**, v. 26, n. 13, p. 1608–1615, 1 jul. 2010.

ZAMOST, B. L.; NIELSEN, H. K.; STARNES, R. L. Thermostable enzymes for industrial applications. **Journal of Industrial Microbiology**, v. 8, n. 2, p. 71–81, 1 set. 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, 5 jan. 2008.

ZIMIN, A. V. et al. The MaSuRCA genome assembler. **Bioinformatics (Oxford, England)**, v. 29, n. 21, p. 2669–2677, 1 nov. 2013.