

HELISSON FAORO

**PROSPECÇÃO METAGENÔMICA DE BIOCATALISADORES DA
MICROBIOTA DE SOLOS DA FLORESTA ATLÂNTICA PARANAENSE**

Tese apresentada ao Curso de Pós-graduação em Bioquímica da Universidade Federal do Paraná como requerimento parcial para a obtenção do grau de Doutor em Ciências-Bioquímica.

Orientador: Prof. Fábio de Oliveira Pedrosa

Coorientadora: Prof^a. Liu Un Rigo

CURITIBA

2010

**PROSPECÇÃO METAGENÔMICA DE BIOCATALISADORES DA
MICROBIOTA DE SOLOS DA FLORESTA ATLÂNTICA PARANAENSE**

HELISSON FAORO

**PROSPECÇÃO METAGENÔMICA DE BIOCATALISADORES DA
MICROBIOTA DE SOLOS DA FLORESTA ATLÂNTICA PARANAENSE**

Tese apresentada ao Curso de Pós-graduação em Bioquímica da Universidade Federal do Paraná como requerimento parcial para a obtenção do grau de Doutor em Ciências-Bioquímica.

Orientador: Prof. Fábio de Oliveira Pedrosa

Coorientadora: Prof^a. Liu Un Rigo

CURITIBA

2010

TERMO DE APROVAÇÃO

HELISSON FAORO

PROSPECÇÃO METAGENÔMICA DE BIOCATALISADORES DA MICROBIOTA DE SOLOS DA FLORESTA ATLÂNTICA PARANAENSE

Tese aprovada como requisito parcial para a obtenção do grau de Doutor no Programa de Pós-Graduação em Ciências-Bioquímica, Setor de Ciências Biológicas da Universidade Federal do Paraná, pela seguinte banca examinadora

Prof. Dr. Fábio de Oliveira Pedrosa (Orientador)

Departamento de Bioquímica e Biologia Molecular - UFPR

Prof^a. Dr^a. Luciane Passaglia

Universidade Federal do Rio Grande do Sul - UFRGS

Prof. Dr. Ricardo Kruger

Universidade de Brasília - UnB

Prof^a. Dr^a. Rose Adele Monteiro

Departamento de Bioquímica e Biologia Molecular – UFPR

Prof. Dr. Marcelo Müller dos Sandos

Departamento de Bioquímica e Biologia Molecular - UFPR

*Aos meus pais
À Lysangela e Bianca*

AGRADECIMENTOS

Aos orientadores Prof. Fábio de Oliveira Pedrosa e Prof. Liu Un Rigo pela oportunidade de realizar esse trabalho, pela orientação e principalmente pela amizade, que vem desde 2001 quando ingressei na Iniciação Científica. Obrigado por toda a experiência transmitida durante esses anos fundamental para o meu crescimento como pesquisador e para a minha carreira científica.

Ao Professor Emanuel Maltempi de Souza pelas críticas e sugestões que muito contribuíram para esse trabalho.

Ao Professor Leonardo Magalhães Cruz pelo auxílio em Bioinformática na montagem das sequências de DNA.

À Professora Leda Satie Chubatsu pelas sugestões nos experimentos de expressão e purificação de proteínas e por ter me emprestado sua coleção de DVDs de Jornada nas Estrelas (*live long and prosper*)

À Professora Rose Adele Monteiro pela amizade e pelo auxílio em várias etapas desse trabalho.

Ao Professor Marcelo Müller e ao Arnaldo Glogauer pelo auxílio nos experimentos de caracterização enzimática essenciais para a conclusão desse trabalho.

Às Professoras Maria Berenice Steffens, Roseli Wassem e ao Professor Luciano Huergo pelo auxílio e sugestões.

Ao Marco Antônio Kadowaki pelo auxílio na etapa de purificação de proteínas.

Ao Valter pela amizade e auxílio na coleta dos solos e em várias outras etapas que ajudaram a concretizar esse trabalho.

À Roseli Prado e à Dona Julieta pelo auxílio técnico e pelo carinho com o qual que sempre me trataram.

Ao pessoal do grupo de discussão de Biodiversidade e Metagenômica.

Aos amigos do Anexo, Michelle, Anelis, Viviane, Marcelo Batista, Sarah, Bruno, Maria Augusta, Eduardo e Adriano com quem passei a maior parte do tempo e muitos momentos divertidos.

Aos demais colegas de laboratório

À Coordenação do curso de Pós-Graduação em Bioquímica e Biologia Molecular e ao CNPq pelo auxílio financeiro

Aos meus amigos de infância Leandro Adiers (o Coeio), Marlon Ribeiro (o Pexe) e Vander Faria (o Fala-fina) pelos momentos agradáveis e descontraídos que passamos e pelos mais de 20 anos de amizade.

Aos novos amigos que fiz em Curitiba Lilian, Marcus, Giovana e Leonardo.

A minha esposa Lysangela pelo amor, carinho e paciência durante essa etapa das nossas vidas. Pelo incentivo nos momentos difíceis em que tudo parecia dar errado, fundamental para chegar até esse momento e principalmente pela nossa filha Bianca.

Ao meu pai Helio Faoro, à minha mãe Ana Carneiro Faoro e à minha irmã Mariana Faoro pelo amor incondicional, incentivo, atenção e paciência dedicados em todos os momentos. Apesar da distância, vocês estão em meus pensamentos todos os dias.

“Se, a princípio, a idéia não é absurda, então não há esperança para ela”

Albert Einstein

SUMÁRIO

TERMO DE APROVAÇÃO	iii
LISTA DE FIGURAS.....	xi
LISTA DE TABELAS	xiii
LISTA DE ABREVIATURAS.....	xiv
LISTA DE SÍMBOLOS	xvii
RESUMO.....	xviii
ABSTRACT	xix
1. INTRODUÇÃO	2
2. REVISÃO BIBLIOGRÁFICA	5
2.1 A Biotecnologia e o desenvolvimento da sociedade humana.....	5
2.2 Biotecnologia e Biodiversidade	7
2.3 Metagenômica.....	11
2.4 Biocatalisadores.....	17
2.4.1 Amilases	21
2.4.2 Proteases	22
2.4.3 Lipases e esterases.....	22
2.5 Lipases	23
2.6 Potencial biotecnológico associado ao solo da Floresta Atlântica Paranaense. 	27
3. OBJETIVOS	31
3.1 Objetivo geral	31
3.2 Objetivos específicos	31
4. MATERIAL E MÉTODOS	33
4.1 Microrganismos, vetores e plasmídeos.....	33
4.2 Material geral	35
4.2.1 Meios de cultivo.....	35
4.2.2 Antibióticos.....	36
4.2.3 Reagentes e material de consumo.....	36
4.3 Coleta das amostras de solo e extração de DNA total.....	36
4.4 Manipulação de DNA.....	39
4.4.1 Purificação de DNA	39
4.4.2 Sequenciamento de DNA.....	41
4.4.3 Preparo de células eletrocompetentes e transformação bacteriana.....	42

4.4.4 Construção das bibliotecas metagenômicas	43
4.4.4.1 Reparo e seleção dos fragmentos de DNA ambiental entre 40 e 50 kb	43
4.4.4.2 Ligação dos fragmentos de DNA no vetor e empacotamento dos fosmídeos em fago lambda.....	44
4.4.4.3 Transformação de <i>Escherichia coli</i> EPI300 por infecção com fago lambda.	45
4.4.4.4 Coleta e armazenamento dos clones transformantes	45
4.4.5 Construção de sub-bibliotecas	46
4.4.5.1 Extração de DNA fosmidial	46
4.4.5.2 Nebulização	46
4.4.5.3 Reparo e fosforilação das extremidades dos fragmentos de DNA nebulizado	47
4.4.5.4 Seleção de fragmentos para construção das sub-bibliotecas.....	47
4.4.5.5 Ligação dos fragmentos de DNA no vetor pUC18	47
4.4.5.6 Coleta de clones transformantes	48
4.4.6 Inativação de genes por mutagênese aleatória com transposon.....	48
4.4.7 Amplificação e clonagem dos genes que codificam lipases	49
4.5 Bioprospecção das bibliotecas metagenômicas por clones de interesse	51
4.6 Manipulação de proteínas	54
4.6.1 Eletroforese em gel de poliacrilamida em condição desnaturante.....	54
4.6.2 Superexpressão de proteínas	55
4.6.3 Purificação da lipase LP001ORF27	56
4.6.4 Dosagem da concentração de proteína e determinação da pureza	56
4.7 Caracterização da lipase LP001ORF27	57
4.7.1 Teste de especificidade de substrato da lipase LP001ORF27.....	58
4.7.2 Efeito de íons metálicos, quelantes de metais e detergentes sobre a atividade da lipase LP001ORF27	58
4.7.3 Efeito do pH sobre atividade da lipase LP001ORF27.....	59
4.7.4 Efeito da temperatura sobre a atividade da lipase LP001ORF27	60
4.8 Análises <i>in silico</i>	61
4.8.1 Análise de biodiversidade bacteriana de amostras de solo da Floresta Atlântica paranaense	61
4.8.2 Montagem das sequências contíguas a partir do sequenciamento de DNA dos subclones	62
4.8.3 Análise de sequências	64

4.8.4 Análises filogenéticas	65
5. RESULTADOS	67
5.1 Seleção das amostras de solo e construção de bibliotecas metagenômicas ...	68
5.1.1 Análise da diversidade bacteriana das amostras de solo da Floresta Atlântica paranaense.	68
5.1.2 Coleta das amostras de solo	76
5.1.3 Purificação de DNA total das amostras de solo	79
5.1.4 Construção das bibliotecas metagenômicas	81
5.2 Prospecção de enzimas.....	85
5.2.1 Identificação de clones com atividade esterase/lipase	85
5.2.2 Identificação de clones com atividade amilase e protease.....	91
5.2.3 Identificação de clones com fenótipos diferenciados	94
5.3 Sequenciamento do inserto de DNA dos clones de interesse e análise de sequências.....	98
5.3.1 Construção de sub-bibliotecas	98
5.3.2 Sequenciamento de DNA e montagem de contigs.....	98
5.3.3 Caracterização genética dos fasmídeos MAF1LP001	104
5.3.4 Caracterização genética do fasmídeo MAF1LP018	107
5.3.5 Caracterização genética do fasmídeo MAF1LP090	109
5.3.6 Identificação dos genes responsáveis pela atividade lipolítica	112
5.3.7 Análises de sequência e filogenética das lipases	115
5.3.7.1 Lipase LP001ORF27	115
5.3.7.2 Lipase LP018ORF16	120
5.3.7.3 Lipase LP090ORF24	125
5.3.8 Análise de sequência da α -amilase	129
5.4 Clonagem, expressão, purificação e caracterização das lipases	132
5.4.1 Amplificação e clonagem dos genes que codificam lipases	132
5.4.2 Expressão de proteínas LP001ORF27, LP090ORF24, LP018ORF15 e LP018ORF16.....	135
5.4.3 Purificação da proteína LP001ORF27	142
5.4.4 Caracterização da lipase LP001ORF27	146
6. DISCUSSÃO	155
6.1 Análises das comunidades bacterianas do solo da Floresta Atlântica Paranaense	155
6.2 Bibliotecas metagenômica	156

6.3 Caracterização genética dos insertos de DNA clonados em fosmídeos dos clones de interesse	157
6.4 Clone MAF1LP001.....	158
6.5 Clone MAF1LP018.....	161
6.6 Clone MAF1LP090.....	163
6.7. Considerações finais.....	165
7. CONCLUSÕES.....	167
8. REFERÊNCIAS BIBLIOGRÁFICAS	169
9. ANEXOS.....	194
9.1 Anexo 1: Mapa do vetor pCC2FOS.....	194
9.2 Anexo 2: Mapa do vetor pUC18	195
9.3 Anexo 3: Mapa do vetor pET28a.....	196
9.4 Anexo 4: Modelagem molecular da lipase LP001ORF27.....	197
9.5 Anexo 5: Modelomolecular da lipase LP001ORF27	198
9.6 Anexo 6: Identificação e análise de clones resistentes a antibióticos	199
9.7 Anexo 7: Artigo aceito para publicação na revista <i>Applied and Environmental Microbiology</i>	214

LISTA DE FIGURAS

FIGURA 1: Biotecnologia e micro-organismos – da fermentação à Metagenômica.....	6
FIGURA 2: Árvore filogenética do domínio Bacteria proposta por WOESE (1987)	9
FIGURA 3: Árvore filogenética do domínio Bacteria proposta por RAAPÉ e GIOVANNONI (2003)	10
FIGURA 4: Purificação de DNA e construção de bibliotecas metagenômicas	15
FIGURA 5: Prospecção de bibliotecas metagenômicas	16
FIGURA 6: Distribuição dos tipos de enzimas utilizadas na indústria biotecnológica	19
FIGURA 7: O dobramento $\alpha\beta$ -hidrolase e os resíduos da tríade catalítica.....	20
FIGURA 8: Reações catalisadas por lipases	26
FIGURA 9: Filos bacterianos identificados no solo da Floresta Atlântica Paranaense.....	29
FIGURA 10: Método de coleta das amostras de solo da Floresta Atlântica Paranaense	38
FIGURA 11: Reação de hidrólise do p-Nitrofenol éster catalisada por lipases e esterase.	57
FIGURA 12: Comparação entre as sequências dos grupos de alta altitude e de baixa altitude pelo programa LibShuff	72
FIGURA 13: Variação do índice Shannon entre os grupos de alta altitude e de baixa altitude gerado pelo programa DOTUR.	73
FIGURA 14: Curva de rarefação para os grupos de alta e baixa altitude gerada pelo programa DOTUR	74
FIGURA 15: Número estimado de OTUs para os grupos de alta e baixa altitude gerados pelo programa DOTUR.	75
FIGURA 16: Pontos de coleta	77
FIGURA 17: Sítios de coleta das amostras de solo.....	78
FIGURA 18: Purificação de DNA das amostras de solo da Floresta Atlântica paranaense	80
FIGURA 19: Seleção de fragmentos de DNA para construção das bibliotecas metagenômicas..	83
FIGURA 20: Prospecção por clones com atividade esterase/lipase em tributirina.....	87
FIGURA 21: Identificação de clones com atividade esterase/lipase em tricaprilina.	88
FIGURA 22: Identificação de clones com atividade lipase em trioleína	89
FIGURA 23: Prospecção por clones com atividade amilase	92
FIGURA 24: Prospecção por clones com atividade Protease.....	93
FIGURA 25: Clones com fenótipo diferenciado identificados na biblioteca MAF1.	95
FIGURA 26: Análise da sequência contígua obtida a partir da sub-biblioteca SCMAF3LP050	100
FIGURA 27: <i>Scaffold</i> obtido para o clone MAF1LP001	101
FIGURA 28: <i>Scaffold</i> obtido para o clone MAF1LP018	102

FIGURA 29: <i>Scaffold</i> obtido para o clone MAF1LP090	103
FIGURA 30: Representação esquemática da distribuição gênica e mutagênese por inserção de transposon.....	114
FIGURA 31: Análise filogenética da proteína LP001ORF27.....	117
FIGURA 32: Análise filogenética da proteína LP001ORF27 e lipases da família I.....	118
FIGURA 33: Alinhamento múltiplo de sequências de lipases das subfamílias I.5 e I.6, LP001ORF27 e proteínas fúngicas relacionadas	119
FIGURA 34: Análise filogenética da proteína LP018ORF16.....	122
FIGURA 35: Alinhamento múltiplo de sequências da lipase LP018ORF16 e proteínas relacionadas.	123
FIGURA 36: Análise filogenética da proteína LP018ORF15.....	124
FIGURA 37: Análise filogenética da proteína LP090ORF24.....	126
FIGURE 38: Análise filogenética da proteína LP090ORF24 e lipases da família I.....	127
FIGURA 39: Alinhamento múltiplo de sequências de lipases das subfamílias I.1 e I.2, LP090ORF24 e proteínas relacionadas	128
FIGURA 40: Análise de sequência do contig MAF1LP090 e da proteína LP090ORF27.....	130
FIGURA 41: Amplificação dos genes codificadores de lipases	133
FIGURA 42: Clonagem dos produtos de PCR no vetor de expressão pET28a	134
FIGURA 43: Teste do clones expressando lipases em meio seletivo.....	137
FIGURA 44: Teste de expressão das proteínas LP001ORF27, LP090ORF24, LP018ORF15 e LP018ORF16	138
FIGURA 45: Teste de solubilidade das proteínas LP001ORF27 e LP090ORF24.....	139
FIGURA 46: Teste de solubilidade das proteínas LP018ORF15 e LP018ORF16.....	140
FIGURA 47: Teste das frações solúvel e insolúvel da proteína LP001ORF27 em meio seletivo.	141
FIGURA 48: Purificação da proteína LP001ORF27.	143
FIGURA 49: Diálise das frações F1 e F2 da purificação de LP001ORF27	144
FIGURA 50: Teste das frações F1 e F2 de LP001ORF27 em meio seletivo agar-tricaprilina 1%	145
FIGURA 51: Especificidade de substrato da lipase LP001ORF27	148
FIGURA 52: Efeito do pH sobre a atividade de LP001ORF27.....	149
FIGURA 53: Efeito de íons e quelantes de metal sobre a atividade da lipase LP001ORF27	150
FIGURA 54: Efeito de detergentes sobre a atividade da lipase LP001ORF27.....	151
FIGURA 55: Efeito da temperatura sobre a atividade da lipase LP001ORF27	152

LISTA DE TABELAS

TABELA 1: Microrganismos	33
TABELA 2: Vetores.....	33
TABELA 3: Plasmídeos	34
TABELA 4: Iniciadores utilizados para sequenciamento de DNA.....	42
TABELA 5: Iniciadores utilizados na amplificação dos genes codificadores de lipases.....	50
TABELA 6: Composição dos meios seletivos para identificação de atividade enzimática.....	53
TABELA 7: Composição do gel SDS-PAGE utilizado para análise de proteínas.....	54
TABELA 8: Comparação entre as sequências das bibliotecas MA01 a MA10 pelo programa S-LibShuff	71
TABELA 9: Características dos pontos de coleta	76
TABELA 10: Análise das sequências das extremidades dos insertos de DNA dos clones com atividade em tricaprilina	90
TABELA 11: Clones com fenótipo diferenciado identificados na biblioteca MAF1.....	94
TABELA 12: Resultado final da prospecção das bibliotecas metagenômicas.....	96
TABELA 13: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP001.....	106
TABELA 14: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP018.....	108
TABELA 15: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP090.....	111

LISTA DE ABREVIATURAS

a.C	– antes de Cristo
AgNO₃	– nitrato de prata
Amp	– ampicilina
ATP	– trifosfato de adenosina
Ba(CH₃COO)₂	– acetato de bário
CaCl₂	– cloreto de cálcio
CdSO₄	– sulfato de cádmio
Cm	– cloranfenicol
CO₂	– dióxido de carbono
CoCl₂	– cloreto de cobalto
COG	– <i>Clusters of Orthologous Groups</i>
CsCl	– cloreto de céσιο
CTAB	– brometo de cetiltrimetilamônio
CuCl₂	– cloreto de cobre
Da	– dalton
DNA	– ácido desoxirribonucléico
dNTP	– desoxirribonucleotídeo trifosfato
D.O.	– densidade óptica
DOTUR	– <i>Distance-Based OTU and Richness</i>
DTT	– ditiotreitól
EDTA	– ácido etileno-diamino-tetracético
FeCl₃	– clorato de ferro
FeSO₃	– sulfeto de ferro
GPS	– <i>Global positioning system</i>
HCl	– ácido clorídrico
HgCl	– cloreto mercúrio
I₂	– iodo
IPTG	– Isopropil β-D-1-tiogalactopiranosídeo

Kan	– canamicina
KCl	– cloreto de potássio
kDa	– kilodalton
KI	– iodeto de potássio
LiCl	– cloreto de lítio
LED	– <i>Lipase Engineering Database</i>
Mar	– <i>multiple antibiotic resistance</i>
MarR	– repressor do operon <i>mar</i>
MFS	– <i>major facilitator superfamily</i>
MgCl₂	– cloreto de magnésio
MgSO₄	– sulfato de magnésio
MnCl₂	– cloreto de manganês
NaCl	– cloreto de sódio
NaOH	– hidróxido de sódio
NiCl₂	– cloreto de níquel
N-LS	– n lauril sarcosil
NP40	– octilfenoxipolietoxietanol
ORF	– fase de leitura aberta
OTU	– unidade taxonômica operacional
pH	– potencial hidrogeniônico
PHA	– poli-hidroxiálcanoato
PHADED	– <i>PHA Depolymerase Engineering Database</i>
pI	– ponto isoelétrico
pNP	– paranitrofenol
RbCl	– cloreto de rubídio
RDP	– <i>Ribosomal Database Project</i>
RNA	– ácido ribonucleico
RNase	– ribonuclease
rpm	– rotações por minuto
rRNA	– RNA ribossômico

SDS	– dodecil sulfato de sódio
SDS-PAGE	– eletroforese em gel de poliacrilamida desnaturante SDS
Sm	– estreptomicina
SnCl₂	– cloreto de estanho
tRNA^{SER}	– RNA transportador para serina
X-gal	– 5-bromo-4-cloro-3-indolil-β-D-galactopiranosideo
wHTH	– <i>winged helix-turn-helix</i>
ZnCl₂	– cloreto de zinco

LISTA DE SÍMBOLOS

cm	– centímetro
g	– grama
km	– quilometro
kpb	– kilo pares de base
kPa	– kilopascal
kV	– kilovolts
M	– molar
Mb	– mega base
mL	– mililitro
mM	– milimolar
mm	– milímetro
ng	– nanograma
nm	– nanômetro
nM	– nanomolar
pb	– pares de bases
pmol	– picomol
V	– volts
U	– unidade
x g	– gravidade
µg	– micrograma
µL	– microlitro
µM	– micromolar
%	– porcento
U\$	– dólares
°C	– graus Celsius

RESUMO

A diversidade bacteriana do solo e seu potencial biotecnológico são pouco explorados. Dados indicam que 1 g de solo contém mais de 10 bilhões de micro-organismos distribuídos em milhares de espécies. Entretanto, a grande maioria dessas espécies são desconhecidas devido a sua inabilidade de se multiplicar nos meios de cultivo tradicionalmente usados em laboratório. A Metagenômica tornou possível o acesso a essa vasta diversidade genética permitindo a análise direta do DNA de uma comunidade bacteriana e levando à descoberta de novos biocatalisadores. Nesse trabalho são descritas a construção de três bibliotecas metagenômicas, a prospecção por novos biocatalisadores e a caracterização de uma nova lipase identificada nessas bibliotecas. As bibliotecas foram construídas a partir de amostras de solo da Floresta Atlântica Paranaense coletadas em diferentes altitudes. O DNA das amostras de solo foi purificado, reparado e clonado no vetor pCC2FOS. Os fósídeos recombinantes foram transformados na estirpe EPI300 de *Escherichia coli* gerando as bibliotecas metagenômicas MAF1, MAF2 e MAF3 com 34.560, 29.280 e 36.288 clones, respectivamente. Todos os clones foram analisados quanto à atividade triacil-hidrolásica em tributirina (315 clones positivos), tricaprilina (10 clones positivos) e trioleína (3 clones positivos). Os clones da biblioteca MAF1 também foram analisados quanto à atividade de protease (460 clones positivos) e atividade de amilase (4 clones positivos). O DNA inserto dos clones com atividade lipolítica MAF1LP001, MAF1LP018 e MAF1LP090 foram sequenciados, o que permitiu identificar 80 ORFs/genes. A lipase LP001ORF27 foi identificada como membro da família I das lipases e similaridade próxima à subfamília termofílica I.5. A enzima purificada apresenta atividade contra diferentes para-nitrofenil-monoacilesteres (pNP monoacilesteres) com atividade máxima (7 U/mg) contra pNP decanoato. LP001ORF27 apresentou atividade do pH 5,0 ao 11,0, com atividade ótima em pH 7,0 e não apresentou dependência de metais. A temperatura ótima de atividade ocorreu na faixa de 50-60°C e a enzima apresentou ativação térmica de 80% após incubação por 1 hora a 50°C. A lipase LP018ORF16 não apresentou similaridade com sequências de lipases conhecidas e a análise filogenética sugeriu que ela constitui uma nova família. Essa enzima também demonstrou dependência da proteína LP018ORF15, provavelmente uma chaperona, para atividade. LP090ORF24 foi identificada como membro da família I das lipases, próxima a subfamília I.2, e também apresenta similaridade com uma poli-hidroxialcanoato (PHA) depolimerase. Próximo a LP090ORF24 foi identificado um gene que codifica uma α -amilase. Finalmente, a análise da sequência dessa proteína sugeriu que ela contém um domínio de ciclodextrina glicosiltransferase (CGTase). Os resultados obtidos nesse trabalho confirmam o potencial da Metagenômica para a descoberta de novas enzimas de importância biotecnológica.

ABSTRACT

The bacterial soil diversity and its biotechnological potential are poorly explored. Data indicated that 1 g of soil contain more than 10 billion microorganisms distributed in thousand of species. However, less than 1% of these species is known because of its inability to grow in traditional culture medium. Metagenomics has made possible to access the vast genetic resources hidden in uncultivable microorganisms genomes since it allows direct analysis of a whole bacterial community DNA and can lead to the discovery of new biocatalysts. This work describe the construction and screening for biocatalysts from three metagenomic libraries from Brazilian Atlantic Forest soil and the characterization of a new lipase. The libraries were constructed from soil samples collected at different altitude levels and with different bacterial community along the Graciosa Road. The total DNA of these samples was extracted, end-repaired and cloned in pCC2Fos fosmid. The recombinant fosmids were transformed into *Escherichia coli* EPI300 strain yielding the metagenomic libraries MAF1, MAF2 and MAF3 with 34,560, 29,208 and 36,192 clones. All clones were screened for lipase activity on tributyrin (315 positive clones), tricapyrylin (10 positive clones) and triolein (3 positive clones). Clones from MAF1 library were also screened for protease activity (460 positive clones) and amylase activity (4 positive clones). The DNA insert of the triolein degrading clones MAF1LP001 and MAF1LP018 and tricapyrylin degrading clone MAF1LP090 were sequenced. A total of 80 genes were identified. Lipase LP001ORF27 was identified as member of family I of lipases close to the thermophilic subfamily I.5. The purified enzyme presented activity in a broad range of substrates with the highest activity against p-Nitrophenyl decanoate. LP001ORF27 showed activity in pH from 5.0 to 11.0, with an optimum at 7.0, and any dependence of metal ions was observed for its activity. The optimum temperature was 50-60°C and the enzyme had a thermal activation of 80% after 1 hour at 50°C. Lipase LP018ORF16 showed no similarity with previous lipase sequence and phylogenic analysis suggested that it constitute a new family. This enzyme also showed dependence of LP018ORF15, probably a foldase, for its lipolytic activity. LP090ORF24 was identified as family I lipase near to subfamily I.2 and share sequence similarity with a poly-beta-hydroxyalkanoate (PHA) depolymerase enzyme. Near to LP090ORF24 was also identified a α -amylase coding gene sequence. Finally, protein sequence analyses suggest that this enzyme has a cyclodextrin glycosyltransferase (CGTase) domain. The obtained results confirmed the potential of the metagenomic approach in the discover of new enzymes of biotechnological importance.

CAPÍTULO 1

INTRODUÇÃO

1. INTRODUÇÃO

A diversidade microbiana do solo e o potencial biotecnológico a ela associada são pouco conhecidos. Dados estatísticos indicam que 1 g de solo pode conter cerca de 10 bilhões de microrganismos compreendendo milhares de espécies diferentes. No entanto, a grande maioria desses microrganismos são desconhecidos devido a sua inabilidade em se multiplicar nos meios de cultivo tradicionalmente usados em laboratório. Para superar os limites impostos pelos meios de cultivo, uma nova abordagem, denominada Metagenômica, vem sendo aplicada. A Metagenômica permite o acesso direto ao DNA de uma comunidade bacteriana presente em um determinado ambiente (metagenoma) de modo independente de cultivo. Pela análise de bactérias recombinantes que expressam genes provindos do metagenoma é possível acessar um grande número de novas enzimas ou mesmo vias metabólicas.

Atualmente existe uma grande demanda por novas enzimas e moléculas bioativas para aplicação em diversos setores da indústria. Nesse cenário, a Metagenômica surge como a melhor ferramenta de prospecção e com potencial para revolucionar o campo da biotecnologia. Enzimas com uma enantio e régio seletividade superior à das enzimas atualmente utilizadas podem gerar inúmeros benefícios econômicos e, também, aperfeiçoar os processos industriais, diminuindo o desperdício e a contaminação do meio ambiente. Além disso, há também uma grande gama de moléculas bioativas produzidas por microrganismos ainda não descobertas que podem ser usadas como agentes terapêuticos. A perspectiva de uma vasta diversidade microbiana ainda desconhecida somada à sua imensa diversidade genética justifica o uso das técnicas de Metagenômica para prospecção de novas enzimas e biocompostos a partir de amostras ambientais. Assim sendo, o objetivo desse trabalho foi construir bibliotecas metagenômicas de amostras de solo da Floresta Atlântica Paranaense e realizar a prospecção por novos biocatalisadores.

CAPÍTULO 2

REVISÃO

BIBLIOGRÁFICA

2. REVISÃO BIBLIOGRÁFICA

2.1 A Biotecnologia e o desenvolvimento da sociedade humana

2.2 Biotecnologia e Biodiversidade

2.3 Metagenômica

2.4 Biocatalisadores

2.5 Lipases

2.6 Potencial biotecnológico associado ao solo da Floresta Atlântica Paranaense

2. REVISÃO BIBLIOGRÁFICA

2.1 A Biotecnologia e o desenvolvimento da sociedade humana

A história da Biotecnologia se mistura com a história da humanidade. Os micro-organismos vêm sendo explorados pelo ser humano há muito tempo. Processos fermentativos que utilizam micro-organismos começaram a ser usados há milênios. Relatos históricos indicam que a fermentação de cereais para produção de bebidas alcoólicas já era praticada pelos sumérios e babilônios no ano 6 mil a.C. Por volta do ano 2 mil a.C os egípcios passaram a utilizar fermento também para fabricação de pão além da cerveja (BUD, 1993). Entretanto, os agentes que realizavam o processo fermentativo eram desconhecidos e assim permaneceram até o século 17 quando o pesquisador Anton van Leeuwenhock, utilizando um microscópio criado por ele mesmo, descreveu a existência de organismos microscópicos invisíveis ao olho nu. Mas foi somente em 1876 que Louis Pasteur demonstrou que os micro-organismos eram os responsáveis pelos processos fermentativos e que micro-organismos diferentes fermentavam compostos diferentes e podiam viver e se multiplicar na ausência de ar. Por essa época também Robert Koch desenvolveu o meio sólido com ágar como método para isolar culturas puras de micro-organismos que vira a se tornar uma ferramenta fundamental para a genética bacteriana (KELLER e ZENGLER, 2004). Em 1897, por sua vez, Eduard Buchner demonstrou a conversão de açúcar em álcool na ausência de células vivas, apresentando o primeiro indício da existência das enzimas. Outro grande marco no campo foi a descoberta da penicilina por Alexander Fleming em 1929. Ele observou que um fungo contaminante (*Penicillium notatum*) que cresceu em suas placas de cultivo inibia a multiplicação da sua cultura de *Staphylococcus aureus* (DEMAIN, 2006). Mas foi a partir das décadas de 50, com descoberta da estrutura do DNA, e 70 com as técnicas de manipulação gênica que a Biotecnologia passou de fato a existir. O crescente aperfeiçoamento da Biologia Molecular e da Microbiologia conduziram ao desenvolvimento da Metagenômica como ferramenta fundamental para a Biotecnologia (Figura 1).

FIGURA 1: Biotecnologia e micro-organismos – da fermentação à Metagenômica



O surgimento da Metagenômica está diretamente relacionada ao desenvolvimento da Biotecnologia e da Biologia Molecular.

2.2 Biotecnologia e Biodiversidade

Os micro-organismos (fungos e bactérias) constituem a base da Biotecnologia, sejam como fornecedores de produtos naturais ou atuando diretamente em bioprocessos e, mais recentemente, como fornecedores de material genético. Durante muito tempo os micro-organismos foram explorados com base no cultivo e isolamento de culturas puras. Muitos antibióticos e outros produtos de aplicação farmacológica foram obtidos desse modo (DEMAIN, 2006). Ainda hoje culturas puras são utilizadas no processamento de alimentos, bebidas alcoólicas, ácidos orgânicos, álcoois, alimentos fermentados, tratamento e/ou remediação de resíduos e na fertilização de solos (HUNTER-CEVERA, 1998). A base para toda essa variedade de atividades é encontrada nas enzimas que evoluíram durante milhões de anos mantendo os organismos vivos nos mais extremos e competitivos tipos de ambientes (LORENZ e SCHLEPER, 2002). Assim sendo, o desenvolvimento das técnicas de Biologia Molecular tornou desnecessária a presença de micro-organismos em muitos processos biotecnológicos. Os genes de interesse passaram a ser clonados, superexpressos e a enzima pura utilizada no lugar do microrganismo original. Mas ainda os micro-organismos isolados continuavam sendo a principal fonte de produtos e recursos genéticos para a biotecnologia. No entanto, a descoberta de novas funções e produtos derivados de culturas puras vem diminuindo significativamente devido a uma alta taxa de reisolamento de espécies já conhecidas. Essa diminuição está diretamente relacionada com a diminuição do número de novos organismos descritos. A principal razão para o baixo número de novas espécies descritas era a necessidade de cultivo celular no processo de identificação dos microrganismos, o que na grande maioria das vezes não é possível (PACE et al., 1986).

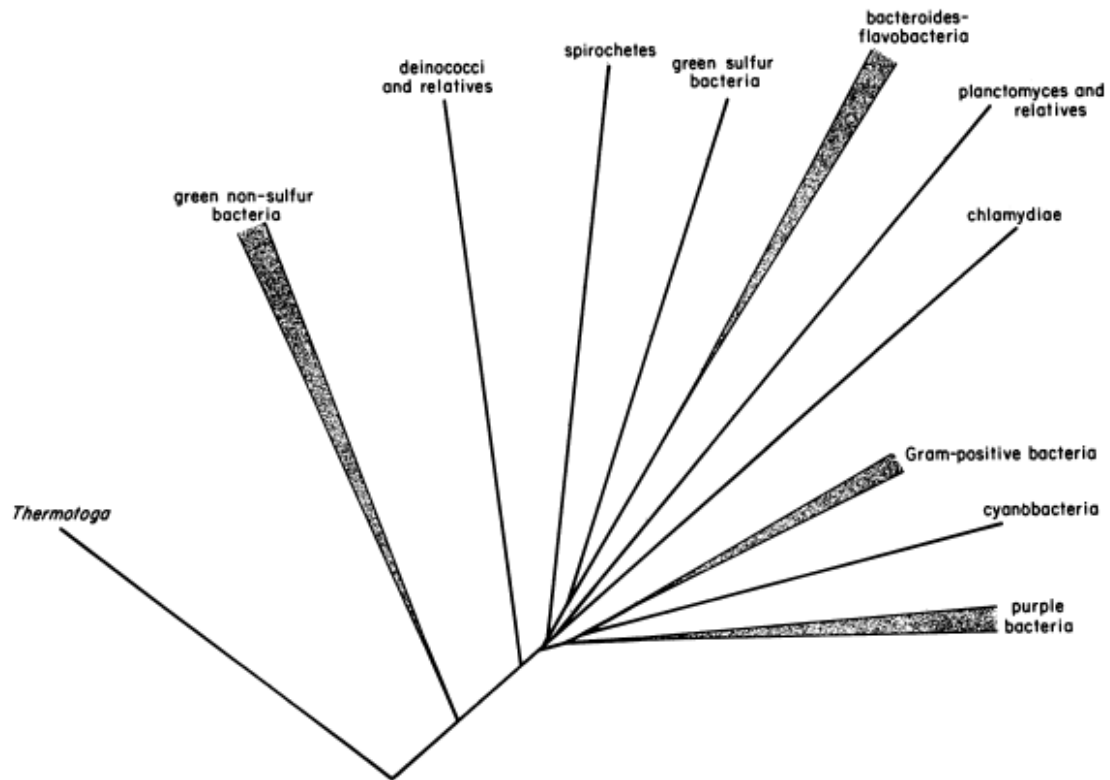
Em 1980, TORSVIK publicou um trabalho no qual apresentava um protocolo para purificação de DNA de toda a comunidade bacteriana de uma amostra de solo (TORSVIK, 1980). Através de medições da cinética de reassociação do DNA dessas amostras comparadas com a cinética do genoma puro de *Escherichia coli* sugeriu a existência de aproximadamente 5 mil genomas na amostra analisada. Essa descoberta

criou uma nova oportunidade para a obtenção de diversidade genética e deu novo fôlego à biotecnologia (TORSVIK, 1980; TORSVIK, GOKSØYR e DAAE, 1990). Em 1986, PACE utilizou iniciadores específicos para o gene 16S rRNA para caracterizar a biodiversidade de amostras de solo de modo independente de meios de cultivo. A descoberta dessa diversidade teve grande impacto também no sistema de classificação filogenética de bactérias. Na proposta inicial de Carl Woese (WOESE, 1987) o domínio Bacteria era formado por 11 filos compostos por indivíduos cultivados e caracterizados (Figura 2). Entretanto, os estudos moleculares independentes de cultivo demonstraram que o domínio Bacteria possui muito mais filos, sendo que alguns apresentam poucos ou nenhum representante cultivável (DeLONG e PACE, 2001). Recentemente RAPPÉ e GIOVANNONI (2003), baseado em sequências do gene 16S rRNA amplificadas de diversos ambientes, aumentaram esse número para 52 filos (Figura 3).

Dados estatísticos indicam que 1g de solo pode conter cerca de 10 bilhões de microrganismos compreendendo milhares de espécies diferentes, a maioria pertencendo aos domínios Bacteria e Archaea (AMANN, LUDWING e SHLEIFER, 1995). No entanto, apenas uma pequena parcela desses micro-organismos é conhecida e podem ser cultivados pelos métodos tradicionais usados em laboratório, sendo que a grande maioria permanece desconhecida à Ciência (TORSVISK, ØVREAS e THINGSTAD, 2002). Desse modo, os micro-organismos presentes no ambiente se tornaram os maiores fornecedores de diversidade genética para a Biotecnologia, criando uma fonte de novos genes e, conseqüentemente, novas enzimas e novos produtos.

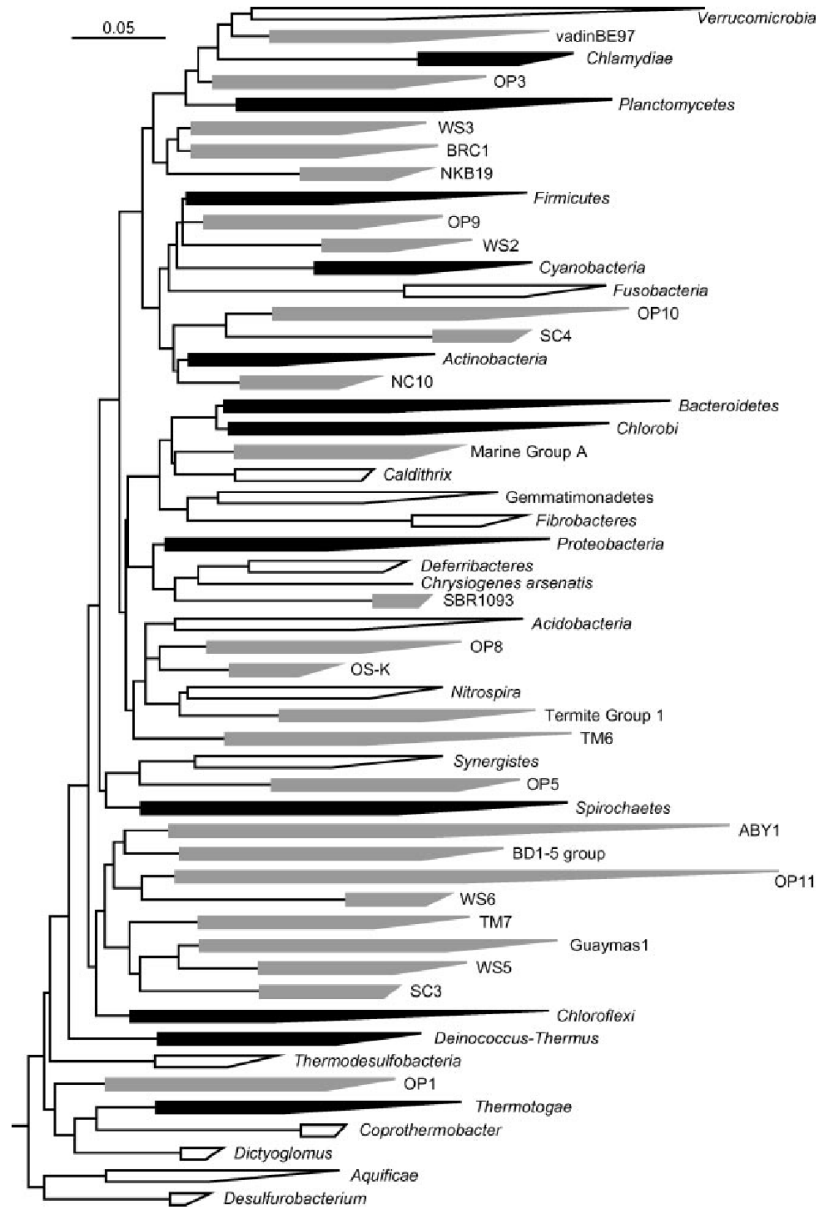
A perspectiva de uma vasta diversidade de microrganismos não-cultiváveis no solo estimulou o desenvolvimento de técnicas que permitissem o estudo dessas comunidades microbianas e do seu potencial biotecnológico independente de cultivo. Dentre essas, a Metagenômica surge como o principal método de prospecção biotecnológica.

FIGURA 2: Árvore filogenética do domínio Bacteria proposta por WOESE (1987)



FONTE: WOESE, 1987. A árvore apresenta 11 filos bacterianos. O grupo das bactérias Gram-positivas foi posteriormente dividido em *Firmicutes* e *Actinobacteria*.

FIGURA 3: Árvore filogenética do domínio Bacteria proposta por RAAPÉ e GIOVANNONI (2003)



FONTE: RAPPÉ e GIOVANNONI, 2003. A árvore apresenta 52 filos bacterianos. Setas preenchidas em preto representam os 12 filos originais (bactérias Gram positivas foram divididas em *Firmicutes* e *Actinobacteria*) descritos por WOESE (1987), setas não preenchidas os filos que possuem representantes cultivados reconhecidos desde 1987 e em cinza estão os 26 filos candidatos que não possuem representantes cultiváveis conhecidos. A barra de escala representa 0,05 mudanças nucleotídicas por posição.

2.3 Metagenômica

A Metagenômica é descrita como a análise funcional e de sequência de genomas microbianos coletivos contidos em uma amostra ambiental (HANDELSMAN et al, 1998). Segundo esse conceito, o conjunto de genomas em uma determinada amostra ambiental é denominado metagenoma. Entretanto, vários nomes já foram utilizados em diferentes trabalhos científicos para descrever essa mesma abordagem como, por exemplo, bibliotecas de DNA ambiental (STEIN et al, 1996), zoobibliotecas (HUGHES, FELBECK e STEIN, 1997), bibliotecas de eDNA (BRADY e CLARDY, 2000), genoma de comunidades (TYSON et al, 2004) e bibliotecas ambientais recombinantes (COURTOIS et al, 2003). A Metagenômica foi apresentada como uma revolução no campo da química biológica e uma mudança de paradigma no campo da biotecnologia. O método utiliza técnicas clássicas de biologia molecular (purificação de DNA, clonagem e transformação) e microbiologia (caracterização fisiológica), mas a combinação desses dois campos à luz da vasta diversidade desconhecida tornam ilimitadas as possibilidades para prospecção biotecnológica. Através da Metagenômica é possível acessar um grande número de genes que codificam novas enzimas ou enzimas mais eficientes do ponto de vista biotecnológico. Operons de genes que codificam para uma via metabólica completa também podem ser recuperados por essa técnica e usados na produção de moléculas do metabolismo secundário (DANIEL, 2004).

Apesar de o termo Metagenômica ter sido cunhado em 1998, o trabalho publicado por STEIN e colaboradores no ano de 1996 é considerado o primeiro no campo. Nesse trabalho os autores analisaram um fragmento de DNA de 40 kpb derivado de uma arqueia marinha (STEIN et al, 1996). O objetivo do trabalho foi evidenciar a presença do organismo de interesse na amostra utilizando-se PCR e iniciadores específicos para o gene 16S rRNA. Não foi realizada uma prospecção funcional com enfoque biotecnológico. A Metagenômica do solo avançou mais lentamente devido à dificuldade de se trabalhar com esse tipo de amostra complexa e à grande quantidade de contaminantes que inibem as reações enzimáticas. Somente

quatro anos mais tarde foi publicado o primeiro trabalho de Metagenômica de solo (RONDON et al, 2000). Nesse trabalho os autores abordaram todos os aspectos que hoje competem ao campo da Metagenômica realizando, além da análise filogenética, análises funcionais para detecção de enzimas celulases, protease, lipases, esterases, amilases, quitinases, queratinases e DNAses e por vias metabólicas produtoras de biomoléculas. O DNA inserto de clones de interesse foi sequenciado, algumas proteínas foram superexpressas e caracterizadas numa abordagem voltada para a biotecnologia. Devido ao grande número de análises muitos autores reconhecem o trabalho de RONDON e colaboradores (2000) como o primeiro no campo da Metagenômica.

A técnica de Metagenômica consiste basicamente na purificação e clonagem de DNA de amostras ambientais (Figura 4) seguido de prospecção por novos genes (Figura 5). Vários protocolos já foram descritos para a purificação de DNA do solo (TORSVIK, 1980; OGRAM, SAYLER e BARKAY, 1987; ZHOU, BRUNS e TIEDJE, 1996; HURT et al, 2001; LILES et al, 2008). No entanto, as metodologias podem ser divididas entre método direto e método indireto (DANIEL, 2005). A purificação de DNA pelo método direto é realizada com toda a matriz do solo. Geralmente, esse método envolve um passo de lise mecânica na qual o solo é misturado com esferas de sílica e submetido à intensa agitação. O objetivo dessa etapa é desfazer os agregados de solo, liberar o máximo possível de células em solução e também auxiliar na fragmentação da parede celular de bactérias resistentes à lise química. No método indireto de purificação de DNA as amostras de solo são submetidas a vários ciclos de lavagem e centrifugação com o objetivo de separar as células bacterianas da matriz do solo. Após se obter a suspensão de células, a purificação de DNA prossegue como uma purificação de DNA genômico tipicamente usada em culturas puras. Ambos os métodos possuem vantagens e desvantagens. O método direto, por não realizar a lavagem do solo, permite o acesso a um número maior de células e à uma maior diversidade de espécies na amostras quando comparado ao método indireto, mas, justamente por não realizar a lavagem, o DNA purificado apresenta uma quantidade maior de contaminantes. Além disso, a agitação

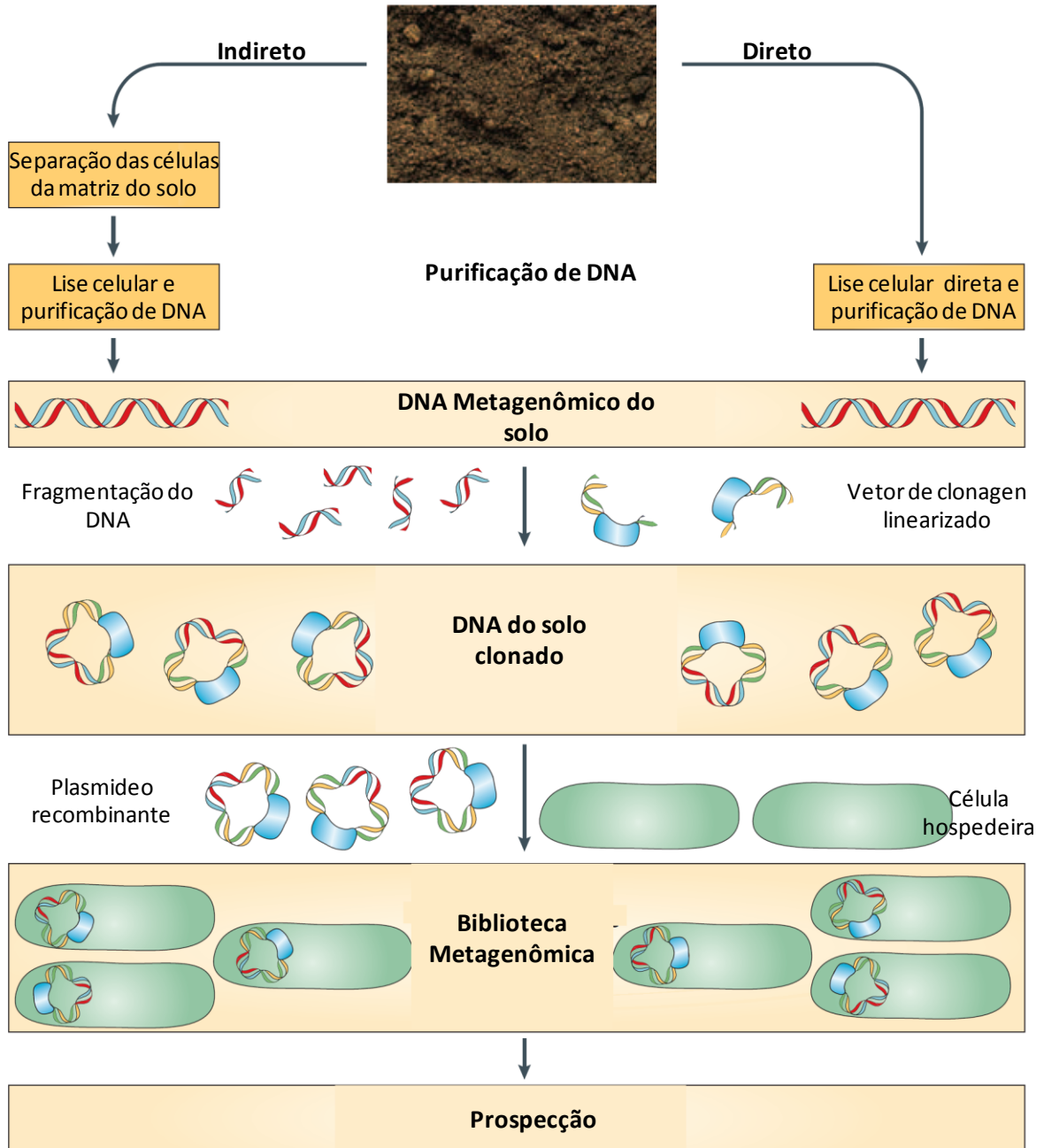
com esferas de sílica pode fragmentar em excesso o DNA e inviabilizar a construção de bibliotecas com insertos grandes de DNA. O método indireto, por realizar várias etapas de lavagem das amostras de solo e por utilizar uma lise branda, garante um DNA mais puro e de alto peso molecular, mas acredita-se que uma parte da diversidade bacteriana é perdida por não conseguir se separar da matriz do solo durante a lavagem (KOZDRÓJ e VAN ELSAS, 2001; LAKAY, BOTHA e PRIOR, 2007). Outro fator importante a ser avaliado na construção de uma biblioteca metagenômica é o tamanho de inserto a ser clonado. Bibliotecas com insertos de DNA grandes (acima de 10 kb) são propícias para identificação não somente de enzimas, mas também de vias metabólicas completas compostas por vários genes (GILLESPIE et al, 2002; MacNEIL et al, 2001; GINOLHAC et al, 2004). No entanto, a obtenção e clonagem de um DNA de alto peso molecular são mais difíceis quando comparadas à obtenção e clonagem de um DNA de baixo peso molecular. Ainda, os vetores utilizados geralmente são de baixo número de cópia, o que pode dificultar a identificação do fenótipo. Por sua vez, a construção de bibliotecas com insertos pequenos de DNA (> 10 kb) é relativamente de fácil execução e os vetores utilizados são de alto número de cópias. Mas o tamanho de inserto pequeno torna praticamente impossível a recuperação de vias metabólicas e implica a análise de muito mais clones para se identificar uma atividade de interesse (HENNE et al, 1999; BRADY e CLARDY, 2000; KNIETSCH et al, 2003).

A análise, ou prospecção, de bibliotecas metagenômicas pode ser feita através de dois métodos: o método baseado na função e o método baseado na sequência (SCHLOSS e HANDELSMAN, 2003) (Figura 5). O método baseado na função identifica clones que expressam uma enzima ou uma via metabólica desejada (RONDON et al, 2000; VOGET et al, 2003; JEONG et al, 2009). A identificação do fenótipo é realizada em meio de cultivo sólido contendo um substrato seletivo para a função de interesse (meio contendo amido para detecção de amilases, por exemplo). Clones que estejam expressando a enzima formam um halo de hidrólise em volta da colônia. Outro modo de realizar a prospecção funcional é através da adição de indicadores de pH no meio seletivo. Nesse caso, o produto de interesse é detectado

através da mudança de pH em volta da colônia. O método baseado na sequência faz uso de regiões conservadas do DNA para se desenhar sondas de hibridização ou oligonucleotídeos iniciadores para reações de PCR e, dessa forma, identificar clones que possuam os genes de interesse. Outra abordagem utilizada é o sequenciamento das extremidades dos insertos de DNA clonados ou mesmo o sequenciamento de todo o DNA metagenômico de uma determinada amostra (VENTER et al, 2004; SCHLÜTER et al, 2008; MORIMOTO e FUJII, 2009). Por essa metodologia, a identificação é realizada por comparação das sequências obtidas com os bancos de dados internacionais. Assim como os dois métodos de purificação de DNA, esses dois métodos possuem vantagens e desvantagens. O método baseado em função identifica rapidamente clones expressando enzimas funcionais e que potencialmente poderão ser aplicados em processos biotecnológicos. No entanto, o sistema de expressão heterólogo pode não funcionar para maioria das proteínas. Isso acaba limitando a prospecção aos genes reconhecidos pelo sistema de expressão de bactéria hospedeira. O método baseado em sequência, por sua vez, não é limitado ao sistema de expressão da célula hospedeira e consegue atingir um maior nível de diversidade genética. Porém, a identificação de um determinado gene não garante que este será expresso heterologamente nem que a proteína codificada possua a função anotada no banco de dados.

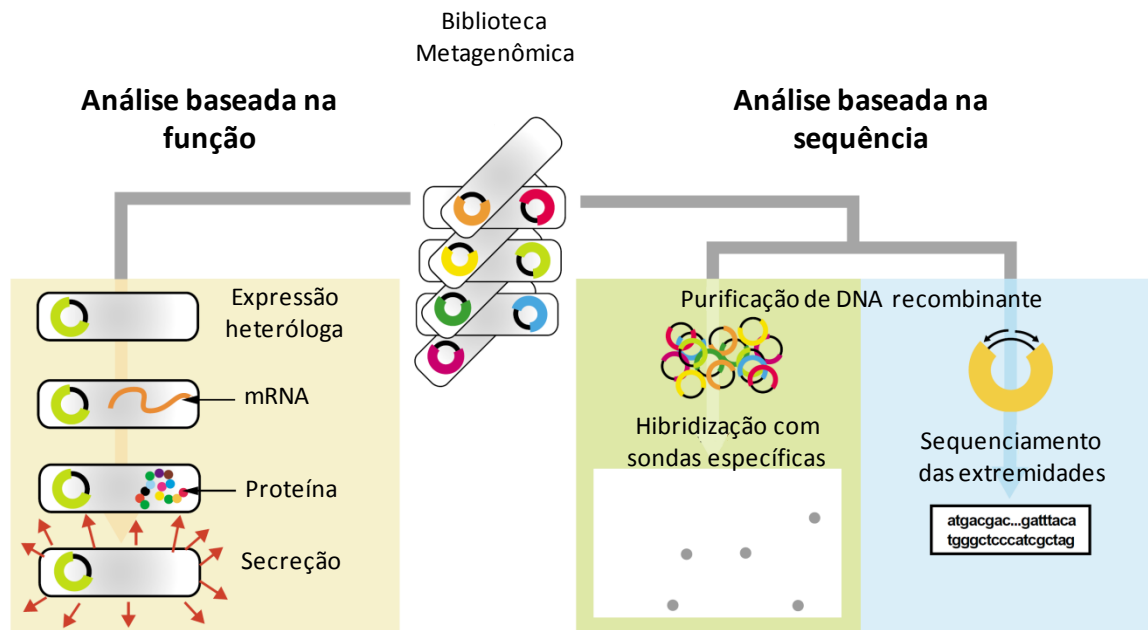
Baseadas nessas técnicas descritas várias bibliotecas metagenômicas foram construídas a partir dos mais distintos tipos de amostras, como solo (RONDON et al., 2000), água (ELEND et al., 2006), biofilmes (SCHMEISSER et al., 2003), rumem bovino (FERRER et al., 2005b), boca humana (DIAZ-TORRES et al., 2003), intestino de cupim (WARNECKE et al, 2007), entre outros, tendo-se como principal objetivo o isolamento de novas enzimas.

FIGURA 4: Purificação de DNA e construção de bibliotecas metagenômicas



FONTE: Adaptado de DANIEL, 2005. Fluxograma mostrando as principais etapas para a construção de bibliotecas metagenômicas a partir de amostras de solo. A purificação do DNA pode ser realizada pelo método indireto, através da separação das células da matriz do solo, ou pelo método direto. O DNA purificado é fragmentado, ligado em um vetor linearizado e transformado em uma estirpe hospedeira originando a biblioteca. Por fim, diferentes estratégias de prospecção podem ser aplicadas para identificação de clones de interesse.

FIGURA 5: Prospecção de bibliotecas metagenômicas



FONTE: Adaptado a partir de SCHLOSS e HANDELSMAN, 2003. A prospecção de bibliotecas metagenômicas pode ser realizada através do método baseado na função ou do método baseado na sequência.

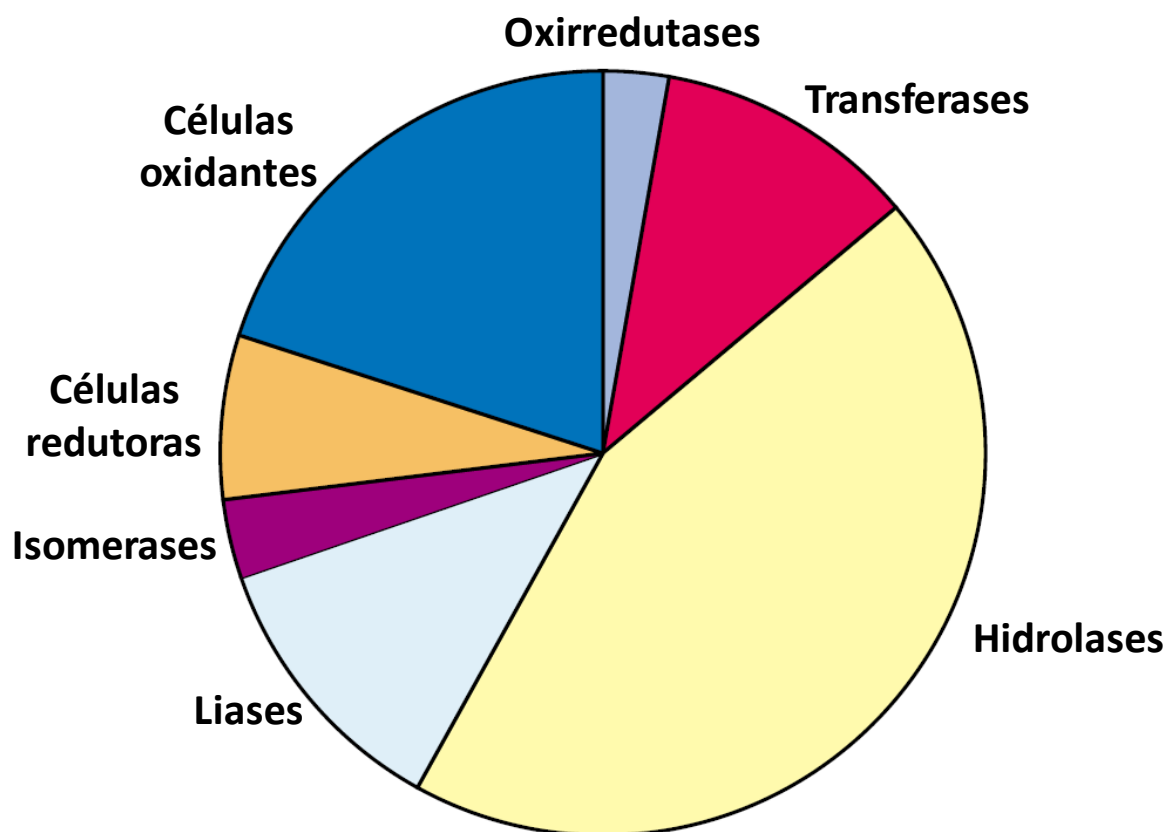
2.4 Biocatalisadores

Os catalisadores são substâncias que aumentam a velocidade de reações químicas sem que sejam consumidos durante a reação. As enzimas são catalisadores biológicos capazes de interagir com uma grande variedade de moléculas complexas e substratos. A sua aplicação no processamento de materiais biológicos apresenta muitas vantagens em relação à síntese química como condições de reação mais brandas com menos gasto de energia e sustentabilidade ambiental (FERRER, MARTÍNEZ-ABARCA E GOLYSHIN, 2005). Essas características levaram a um grande aumento na demanda industrial por novas enzimas para melhorar um processo já existente ou mesmo estabelecer novas formas de executar um processo antigo (SCHMID et al, 2001). Por isso, o potencial econômico associado aos biocatalisadores também é muito grande. O mercado global de enzimas é estimado em 2,3 bilhões de dólares por ano divididos em detergentes (U\$789 milhões), aplicações na indústria alimentícia (U\$634 milhões), agricultura (U\$376 milhões), processamento têxtil (U\$237 milhões) e aplicações em papel e couro e produção de reagentes químicos (U\$ 222 milhões) (LORENZ e ECK, 2005).

Dentre as diferentes classes de enzimas, as hidrolases são de grande interesse e respondem por quase metade do mercado de enzimas (Figura 6). Hidrolase é um nome genérico dado a enzimas que catalisam a clivagem de ligações numa reação com a água. A estrutura das hidrolases é composta por uma região central por folhas β e α -hélices o que levou à criação da família $\alpha\beta$ -hidrolase (HEIKINHEIMO et al, 1999). A organização canônica do dobramento $\alpha\beta$ -hidrolase é descrito como consistindo de 8 folhas β paralelas (com exceção da folha β 2 que é antiparalela) cercado dos dois lados por α -hélices (Figura 7). Além do dobramento, essas enzimas compartilham também as características do sítio ativo formado por uma tríade catalítica altamente conservada: um nucleófilo (serina, cisteína ou ácido aspártico), posicionado após a folha β 5, um resíduo ácido geralmente, posicionado após a folha β 7, e um absolutamente conservado resíduo de histidina, localizado após a última folha β (Figura 7) (NARDINI e DIJKSTRA, 1999).

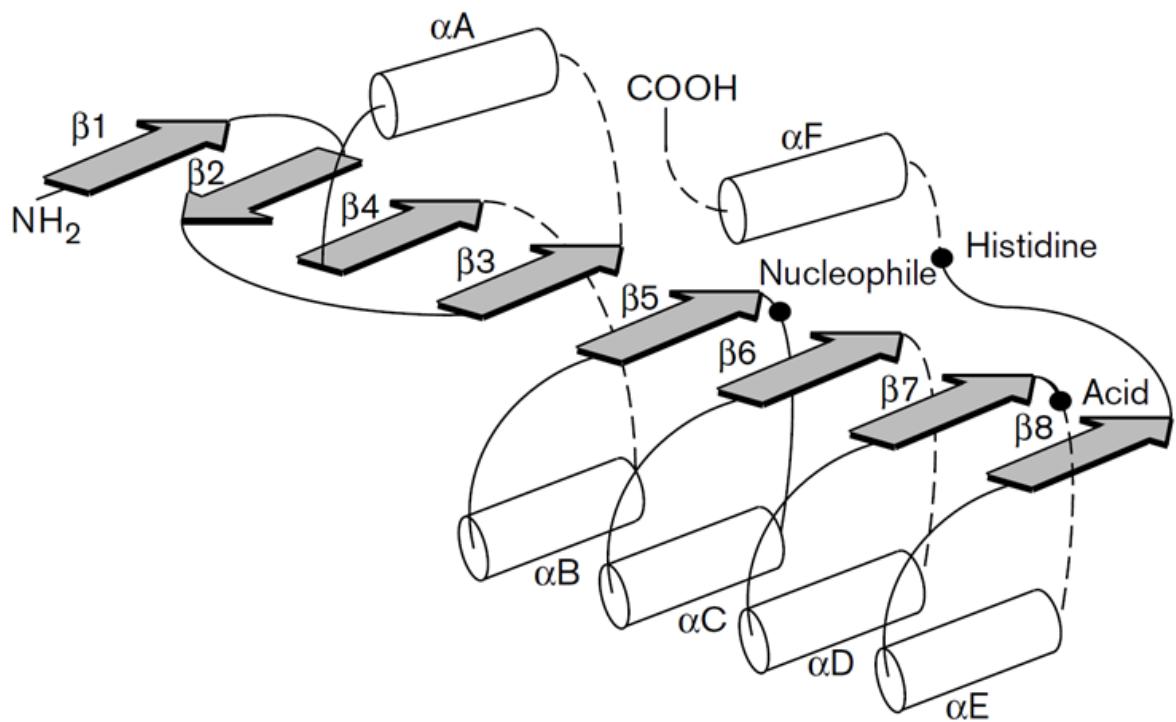
A família das $\alpha\beta$ -hidrolases possui um vasto leque de aplicações e uma alta especificidade de substratos. Além disso, muitas hidrolases apresentam outras características interessantes, como independência de cofatores, termoestabilidade, atividade em pH extremo e estabilidade em solventes orgânicos (BORNSCHEUER e KAZLAUSKAS, 2006). Amilases, celulases, proteases, esterases e lipases estão entre as classes de $\alpha\beta$ -hidrolases mais investigadas.

FIGURA 6: Distribuição dos tipos de enzimas utilizadas na indústria biotecnológica



FONTE: Adaptado de STRAATHOF, PANKE e SCHMID, 2002. As hidrolases representam quase 50% das enzimas utilizadas em processos de biotransformação.

FIGURA 7: O dobramento $\alpha\beta$ -hidrolase e os resíduos da tríade catalítica



FONTE: NARDINI e DIJKSTRA, 1999. As α -hélices são representadas pelos cilindros e as folhas β pelas setas planas. As posições topológicas dos resíduos que compõem o sítio ativo são representadas pelos círculos sólidos. O resíduo nucleófilo está depois da fita $\beta 5$, o resíduo de aspartato/glutamato está depois da folha $\beta 7$ e o resíduo de histidina está na volta entre a folha $\beta 8$ e hélice αF .

2.4.1 Amilases

As amilases são enzimas capazes de hidrolisar ligações glicosídicas entre resíduos adjacentes de glucose. Essas enzimas podem ser divididas em duas categorias: endoamilases e exoamilases. As endoamilases (α -amilases) realizam a hidrólise no interior do polímero de forma aleatória, criando oligossacarídeos de vários tamanhos. As exoamilases (β -amilases) realizam a hidrólise de ligações glicosídicas a partir das extremidades não redutoras resultando em produtos pequenos. Tanto a α quanto a β -amilase atuam somente sobre ligações glicosídicas α -1,4. Para a hidrólise de ligações do tipo α -1,6 é necessário a ação de outra enzima (amiloglucosidade) (GUPTA et al., 2003). Na Biotecnologia, as amilases são aplicadas em processos fermentativos, indústria têxtil e de papel. Além disso, o uso de enzimas no tratamento do amido produz xaropes de glucose com propriedades físico-químicas bem definidas e menos escurecidos.

A identificação de clones com atividades amilolíticas em bibliotecas metagenômicas não é muito comum. Há poucos trabalhos de triagem de amilases e apenas 4 enzimas foram caracterizadas quanto suas propriedades bioquímicas (RICHARDSON et al, 2002; YUN et al, 2004; FERRER et al, 2007, SHARMA, KHAN e QAZI, 2009). Dentre essas, duas se destacam. FERRER e colaboradores identificaram uma α -amilase em uma biblioteca metagenômica construída a partir da comunidade bacteriana do rúmen de vacas. Essa enzima apresenta uma atividade extremamente alta de 5000 U/mg de proteína a 70°C e também é ativa em uma ampla faixa de pH (FERRER et al, 2007). SHARMA e colaboradores descreveram a caracterização de uma amilase psicrófila com atividade máxima a 40°C, mas que é capaz de manter mais de 90% da atividade máxima a 10°C, uma característica única entre as amilases caracterizadas até o momento (SHARMA, KHAN e QAZI, 2009).

2.4.2 Proteases

As proteases são enzimas que catalisam a quebra das ligações peptídicas entre os aminoácidos. Essas enzimas podem ser classificadas em diferentes grupos dependendo do pH no qual são ativas e das características dos grupos do sítio ativo. Pelo pH as proteases podem ser classificadas em acídicas, neutras e alcalinas e pelas características do sítio ativo em metalo, aspártico, cisteína, sulfidril ou serina proteases (GUPTA, BEG e LORENZ, 2002). As proteases são a classe de hidrolases com maior representatividade no mercado de enzimas, contribuindo com cerca de 2/3 do volume total de enzimas comercializadas (WONG, 2010). Elas são utilizadas principalmente como constituintes de detergentes enzimáticos e menor grau nas indústrias têxtil, alimentícia e farmacêutica. A principal fonte de enzimas proteolíticas são bactérias do gênero *Bacillus*. Apesar do grande apelo comercial, em Metagenômica as proteases são a classe de enzimas com menos trabalhos descritos. A única protease caracterizada, obtida por metagenômica, foi descrita por LEE e colaboradores (2007). Essa enzima foi identificada em uma biblioteca metagenômica construída a partir de DNA extraído de sedimentos de mar profundo. Ela apresenta uma atividade ótima a 50°C e também é capaz de hidrolisar fibrina.

2.4.3 Lipases e esterases

As lipases e esterases são as classes de hidrolases com o maior número de trabalhos descritos em metagenômica, com mais de 80 artigos (SCHMEISSER, STEELE e STREIT, 2007). O nível de caracterização da enzima em cada trabalho varia desde a identificação de clones lipolíticos até a determinação da estrutura tridimensional da enzima. Uma descrição mais detalhada dessa classe de hidrolases será apresentada no tópico seguinte.

2.5 Lipases

Lipases podem ser definidas como carboxilesterases que catalisam a hidrólise de acilgliceróis de cadeia longa. Entretanto, não há uma definição estrita para o termo “cadeia longa”. Enzimas capazes de hidrolisar glicerolesters com uma cadeia carbônica de 10 átomos podem ser consideradas lipases (EC 3.1.1.1) e enzimas que atuam em glicerolesters com uma cadeia < 10 átomos são classificadas como esterases (EC 3.1.1.3) (JAEGER, DJKSTRA e REETZ, 1999). Contudo, a maioria das lipases é capaz de hidrolisar glicerolesters de cadeias curtas. Outra característica que difere a maioria das lipase das esterases é o fenômeno de ativação interfacial, que se refere ao intenso aumento na atividade da enzima na interface lipídio-água de substratos emulsificados ou micelares, disparado por um rearranjo estrutural do sítio ativo. Na ausência da interface, o sítio ativo apolar da enzima esta coberto por uma tampa (*lid*) e a enzima esta na conformação fechada. Já na presença de substâncias hidrofóbicas, a tampa é deslocada, expondo uma grande superfície hidrofóbica e tornando os resíduos catalíticos acessíveis ao substrato (conformação aberta) (VERGER, 1997). No entanto, nem todas as lipases apresentam ativação interfacial, permanecendo a primeira definição como a mais abrangente. As lipases catalisam não apenas a hidrólise de acilgliceróis, mas também a transesterificação e interesterificação (Figura 8) e podem ser aplicadas na produção de detergentes e biocombustíveis, obtenção de enantiômeros específicos, indústria de alimentos e farmacêutica.

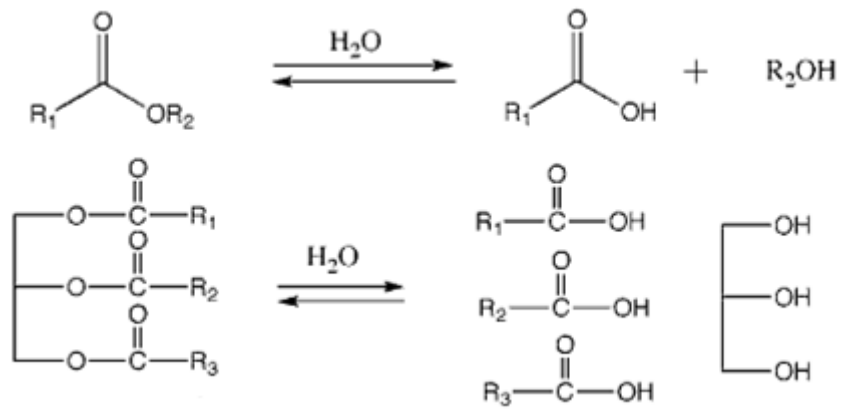
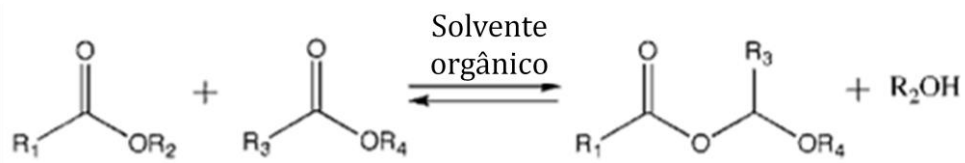
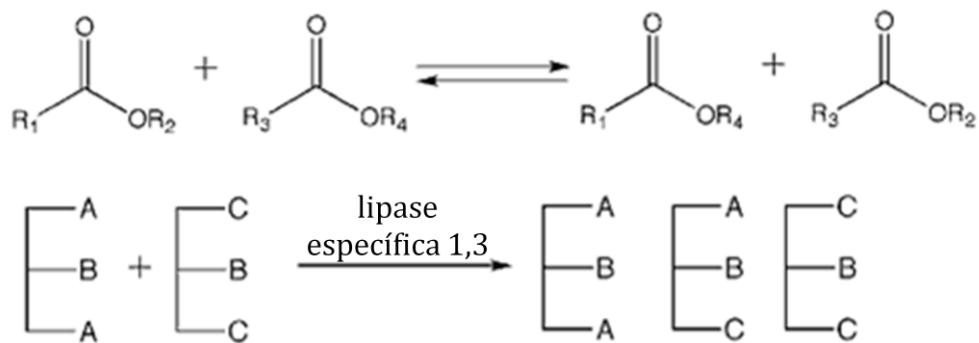
As lipases e as esterases podem apresentar diferentes especificidades de substratos, mas ambas pertencem à família das $\alpha\beta$ -hidrolases e, como tais, o sítio ativo é formado por uma tríade catalítica. No caso das lipases e esterases, a tríade é composta pelos resíduos de serina, aspartato e histidina. O resíduo de serina está no centro de um pentapeptídeo conservado, geralmente composto pelos resíduos G-X-S-X-G, onde G é glicina, X é qualquer aminoácido e S é a serina catalítica. Variações dessa estrutura incluem a substituição do primeiro resíduo de glicina por uma alanina, comum em lipases de organismos do gênero *Bacillus*, ou a reestruturação completa do

sítio ativo com a serina catalítica fazendo parte do motivo G-D-S-L, onde G é glicina, D é aspartato, S é serina e L é leucina (JAEGER, DIJKSTRA e REETZ, 1999).

Uma extensa análise filogenética dividiu as lipases e esterases em 8 principais famílias baseadas na conservação de sequência e propriedades biológicas da enzima (ARPIGNY e JAEGER, 1999). A família I, por ser a mais diversa, foi dividida inicialmente em 6 subfamílias e mais tarde recebeu uma nova subfamília (JAEGER e EGGERT, 2002). Lipases das subfamílias I.1 e I.2 são secretadas pelo sistema de secreção do tipo 2, enquanto as da I.3 são secretadas pelo sistema de secreção do tipo 1 (ANGKAWIDJAJA e KANAYA, 2006). A subfamília I.4 compreendem as lipases de *Bacillus*. Enzimas caracterizadas como termofílicas foram agrupadas na subfamília I.5. A família II de lipases compreende aquelas enzimas onde a serina catalítica faz parte do motivo GDSL. Enzimas da família III apresentam um dobramento $\alpha\beta$ -hidrolase canônico e semelhança com a PAF acetil hidrolase humana. A família IV apresenta grande semelhança com lipases sensíveis a hormônio (HSL) de mamíferos. As lipases agrupadas na família V são originadas de organismos mesofílicos, termofílicos e psicrófilos e apresentam similaridade com outras enzimas não-lipolíticas. As famílias VI e VII são formadas por lipases semelhantes à fosfolipase e acetilcolinesterase eucariótica, respectivamente. Por fim, a família VIII compreende as lipases semelhantes à β -lactamases. A Metagenômica também tem contribuindo para a identificação de novas famílias de lipases. Análises filogenéticas da lipase LipG demonstraram que ela não pertence a nenhum dos grupos previamente descritos, levando os autores a sugerirem uma nova família (LEE et al, 2006). Além da baixa similaridade de sequências, a fenda oxianion nessa família de lipases é formada pelos resíduos de arginina e glicina, uma característica incomum em lipases bacterianas.

As lipases são a classe de hidrolases mais procuradas em bibliotecas metagenômicas. Várias novas enzimas têm sido identificadas a partir de bibliotecas metagenômicas construídas a partir de diferentes ambientes. Isso ocorre devido à ampla distribuição dessa classe de enzima entre os micro-organismos e também à relativa facilidade dos métodos de prospecção. Utilizando-se a Metagenômica como ferramenta foram identificadas mais de 80 novas lipases e esterases, mas poucas

dessas foram caracterizadas (STEELE et al, 2009). A caracterização dessas novas enzimas tem revelado características interessantes para aplicação em processos biotecnológicos. LI e colaboradores identificaram uma lipase capaz de degradar o inseticida piretróide aplicado em plantações e sugere seu uso na biorremediação de ambientes contaminados (LI, WANG e LIU, 2008). FERRER e colaboradores (2005) identificaram um esterase enantiosseletiva capaz de hidrolisar preferencialmente o composto (S)-solketal, usado como molécula de construção de vários fármacos, com um rendimento muito superior ao das enzimas usualmente utilizadas (FERRER et al, 2005a). Enzimas capazes de atuar em pH alcalino e em solventes orgânicos também tem sido identificadas (ELEND et al, 2006; WU e SU, 2009). Termoestabilidade é outra característica importante. MEILLEUR e colaboradores identificaram uma enzima que, além de atuar em pH alcalino (10,5) também apresenta ótimo de atividade à temperatura a 60°C e não tem redução da atividade após 4 horas a 90°C (MEILLEUR et al, 2009). Recentemente, uma atenção especial tem sido dada a lipases que apresentem alta atividade a baixa temperatura, ditas psicrófilas, com grande potencial de aplicação na indústria de detergentes. A importância dessa característica reside no fato de que não seria necessário gastar energia para aquecer o sistema de lavagem. ELEND e colaboradores (2007) identificaram uma lipase em uma biblioteca metagenômica construída a partir de solos contaminados com óleo. A caracterização dessa enzima mostrou que ela mantém 16% de atividade a -5°C (ELEND et al, 2007). Outras duas enzimas que conseguem manter mais de 40% da atividade a 5°C e com baixa energia de ativação já foram identificadas (JEON et al 2009; KIM et al, 2009). Além disso, a baixa similaridade entre sequência de aminoácidos dessas lipases com aquelas depositadas no bando de dados GenBank levou os autores a sugerirem que se tratam de novas famílias dentro da classificação de ARPIGNY e JAEGER (1999).

FIGURA 8: Reações catalisadas por lipases
1) Hidrólise

2) Transesterificação

3) Interesterificação


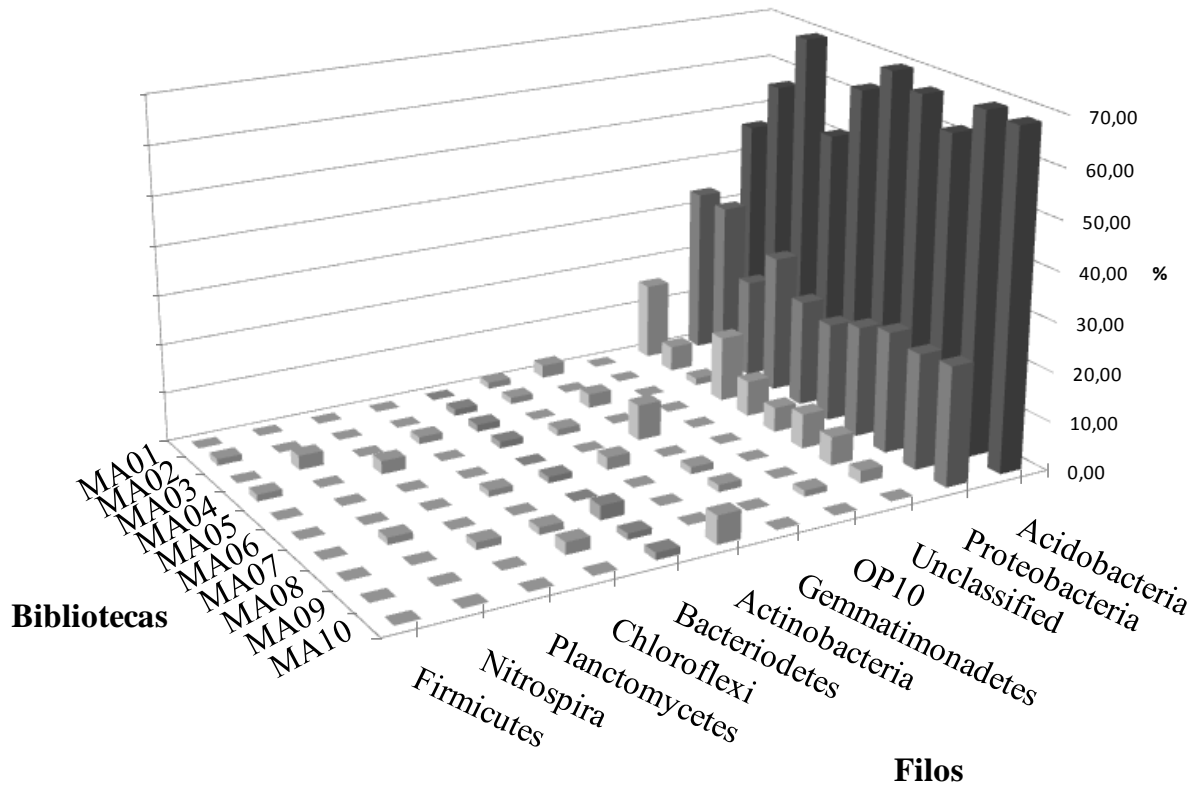
2.6 Potencial biotecnológico associado ao solo da Floresta Atlântica Paranaense.

A Floresta Atlântica brasileira é um dos 25 centros de biodiversidade reconhecidos no mundo (MYERS et al., 2000). A Serra do Mar abriga os principais remanescentes da floresta que recobria a costa leste brasileira, desde o Rio Grande do Norte ao Rio Grande do Sul. A Floresta Atlântica paranaense é a porção mais preservada dessa floresta no Brasil e também a maior área de preservação do estado com aproximadamente 500 mil hectares (CÂMARA, 2005). Um estudo anterior nesse ambiente, baseado na identificação de bactérias pela sequência do gene marcador 16S rRNA, revelou a enorme riqueza de espécies bacterianas nos solos da Floresta Atlântica paranaense (FAORO, 2006). A comparação das sequências parciais do gene 16S rRNA mostrou as amostras de solo eram dominadas por bactérias do filo *Acidobacteria* e *Proteobacteria* (Figura 9)

O potencial biotecnológico associado a essa riqueza de organismos também é muito grande. Considerando apenas alguns organismos que obtiveram altos índices de similaridade na comparação com o banco de dados podemos citar a α -proteobactéria *Novosphingobium* que degrada contaminantes da água de biorreatores, dentre estes 2,4,6-triclorofenol, 2,3,4,6-tetraclorofenol e pentaclorofenol (TIIROLA et al., 2002); a β -proteobactéria *Archangium* que produz o ácido gefirônico, composto antibiótico inibidor da síntese proteica em eucariotos (SASSE et al., 1995); e a também β -proteobactéria *Schlegelella*, que produz e degrada poli-hidroxicarboxilatos, como poli-3-hidroxi-butirato, utilizando-os como fonte de carbono (ELBANNA et al., 2003). As mixobactérias, pertencente à classe δ -*Proteobacteria*, foram outro grupo importante encontrado nos solos da Floresta Atlântica. Essas bactérias são responsáveis pela produção de uma imensa variedade de metabólitos secundários com atividade biológica, incluindo policetídeos e peptídeos não ribossômicos (REICHENBACH e HÖFLE, 1993). Dentre alguns gêneros encontrados estão o já mencionado *Archangium* e *Haliangium*, que produz um antifúngico de largo espectro, denominado aliangicina (FUDOU et al., 2001). Essas características tornam essas amostras de solo

propícias para a construção de uma biblioteca metagenômica e para a prospecção de novos genes e biomoléculas.

FIGURA 9: Filos bacterianos identificados no solo da Floresta Atlântica Paranaense.



A diversidade bacteriana do solo da Floresta Atlântica Paranaense foi investigada através da amplificação, clonagem e sequenciamento parcial do gene 16S rRNA. Foram analisadas 10 amostras de solo denominadas MA01 a MA10 coletadas em diferentes níveis de altitude. As sequências do gene 16S rRNA foram comparadas com o banco de dados RDP para a identificação da espécie.

CAPÍTULO 3

OBJETIVOS

3. OBJETIVOS

3.1 Objetivo geral

Realizar a prospecção por novos biocatalisadores em bibliotecas metagenômicas construídas a partir de amostras de solo da Floresta Atlântica paranaense.

3.2 Objetivos específicos

- Realizar a análise da comunidade bacteriana de amostras de solo da Floresta Atlântica Paranaense;
- Coletar 3 amostras de solo da Floresta Atlântica Paranaense;
- Isolar o DNA total das 3 amostras de solo da Floresta Atlântica Paranaense;
- Obter fragmentos de DNA entre 30 e 40 mil pares de bases;
- Construir 3 bibliotecas metagenômicas das amostras de solo em fosmídeo pCC2FOS;
- Realizar a prospecção metagenômica por Esterases, Lipases, Amilases e Proteases;
- Sequenciar integralmente o inserto de DNA de clones de interesse;
- Amplificar e clonar genes de interesse identificados;
- Superexpressar e purificar as enzimas identificadas;
- Caracterizar cineticamente as enzimas purificadas.

CAPÍTULO 4

MATERIAL E

MÉTODOS

4. MATERIAL E MÉTODOS

4.1 Microrganismos, vetores e plasmídeos

As bibliotecas metagenômicas foram construídas em vetor fosmídeo pCC2FOS (Anexo 1) tendo como hospedeira a estirpe de *E. coli* EPI300. O vetor pUC18 (Anexo 2) e a estirpe hospedeira de *E. coli* TOP10 foram utilizadas na construção de sub-bibliotecas para sequenciamento de fosmídeos de interesse. Por fim o vetor pET28a (Anexo 3) e a estirpe hospedeira de *E. coli* BL21 Rosetta foram utilizados para clonagem e superexpressão dos genes codificadores de enzimas de interesse. As descrições das estirpes de *E. coli* e dos vetores utilizados são apresentadas nas Tabelas 1 e 2. As descrições dos plasmídeos construídos são apresentadas na Tabela 3.

TABELA 1: Microrganismos

Estirpes de <i>Escherichia coli</i>	Fenótipo e/ou Genótipo	Referência
EPI300	[F- mcrA $\Delta(mrr-hsdRMS-mcrBC)$ (StrR) $\phi 80dlacZ\Delta M15$ $\Delta lacX74$ <i>recA1</i> <i>endA1</i> <i>araD139</i> $\Delta(ara, leu)7697$ <i>galU</i> <i>galK</i> λ - <i>rpsL</i> <i>nupG</i> <i>trfA</i> <i>tonA</i> <i>dhfr</i>]	EPICENTRE
TOP10	F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\Phi 80lacZ\Delta M15$ $\Delta lacX74$ <i>recA1</i> <i>araD139</i> $\Delta(ara leu) 7697$ <i>galU</i> <i>galK</i> <i>rpsL</i> (StrR) <i>endA1</i> <i>nupG</i>	Invitrogen
Rosetta (DE3)pLYSs	F ⁺ <i>ompT</i> <i>hsdS_B</i> (R _B ⁻ m _B ⁻) <i>gal</i> <i>dcm</i> λ (DE3 [<i>lacI</i> <i>lacUV5-T7</i> gene 1 <i>ind1</i> <i>sam7</i> <i>nin5</i>]) pLysSRARE (Cam ^R)	Novagen

TABELA 2: Vetores

Vetores	Fenótipo e/ou Genótipo	Referência
pCC2FOS	<i>lacZ</i> , Sítio Cos, <i>OriV</i> , <i>Ori2</i> , Cm ^R	EPICENTRE
pUC18	<i>lacZ</i> , Amp ^R	Invitrogen
pET28a	Promotor T7, Terminador T7, cauda His, Kan ^R	Novagen

TABELA 3: Plasmídeos

Clones em vetor pCC2FOS	Fenótipo e/ou Genótipo	Referência
MAF1 001-360/A01-H12	Clones contendo fragmentos de DNA purificados da amostra de solo MAF1	Este trabalho
MAF2 001-305/A01-H12	Clones contendo fragmentos de DNA purificados da amostra de solo MAF2	Este trabalho
MAF3 001-367/A01-H12	Clones contendo fragmentos de DNA purificados da amostra de solo MAF3	Este trabalho
Clones em vetor pUC18	Fenótipo e/ou Genótipo	Referência
SCMAF1LP001 P1-P4/A01-H12	Subclones contendo fragmentos de DNA do fósídeo MAF1LP001	Este trabalho
SCMAF1LP018 P1-P4/A01-H12	Subclones contendo fragmentos de DNA do fósídeo MAF1LP018	Este trabalho
SCMAF1LP018-8kb P1-P2/A01-H12	Subclones contendo fragmentos de 8 kb de DNA do fósídeo MAF1LP018	Este trabalho
SCMAF1LP090 P1-P4/A01-H12	Subclones contendo fragmentos de DNA do fósídeo MAF1LP090	Este trabalho
SCMAF3LP050 P1-P4/A01-H12	Subclones contendo fragmentos de DNA do fósídeo MAF3LP050	Este trabalho
Clones em vetor pET28a	Fenótipo e/ou Genótipo	Referência
pETLP001ORF27	Gene LP001ORF27 clonado no vetor pET28a	Este trabalho
pETLP090ORF24	Gene LP090ORF24 clonado no vetor pET28a	Este trabalho
pETLP018ORF15	Gene LP018ORF15 clonado no vetor pET28a	Este trabalho
pETLP018ORF16	Gene LP018ORF16 clonado no vetor pET28a	Este trabalho

4.2 Material geral

4.2.1 Meios de cultivo

Os meios de cultivo utilizados para a multiplicação das estirpes de *E. coli*, bem como sua composição, são apresentados abaixo. Os meios de cultivos foram preparados segundos SAMBROOK, FRITSCH e MANIATIS (1989) e esterilizados em autoclave a 120°C por 20 minutos antes do uso.

Meio Luria Bertani (LB):

- Triptona 10 g/L
- Extrato de levedura 5 g/L
- NaCl 10 g/L

O meio LB ágar (LA) foi obtido pela adição de 15 g/L de ágar ao meio LB.

Meio Terrific Broth (TB)

- Triptona 12 g/L
- Extrato de levedura 24 g/L
- Glicerol 4 ml/L

Meio SOB

- Triptona 20 g/L
- Extrato de levedura 5 g/L
- NaCl 0,584 g/L
- KCl 0,186 g/L

Meio SOC

- Triptona 20 g/L
- Extrato de levedura 5 g/L
- NaCl 0,6 g/L

- KCl 0,19 g/L
- MgCl₂ 0,94 g/L
- MgSO₄ 1,2 g/L
- Glucose 3,6 g/L

4.2.2 Antibióticos

Os antibióticos utilizados para multiplicação, manutenção e seleção das estirpes de *E. coli* e dos plasmídeos utilizados foram: cloranfenicol 12,5 µg/mL (Cm^{12,5}), cloranfenicol 30 µg/mL (Cm³⁰), ampicilina 250 µg/ml (Amp²⁵⁰), estreptomicina 40 µg/mL (Sm⁴⁰) e canamicina 50 µg/mL (Km⁵⁰). Ampicilina, estreptomicina e canamicina foram dissolvidas em água ultrapura para as respectivas concentrações estoques de 250, 40 e 100 mg/mL. O cloranfenicol foi dissolvido em etanol 95% (m/v) para a concentração estoque de 30 mg/mL.

4.2.3 Reagentes e material de consumo

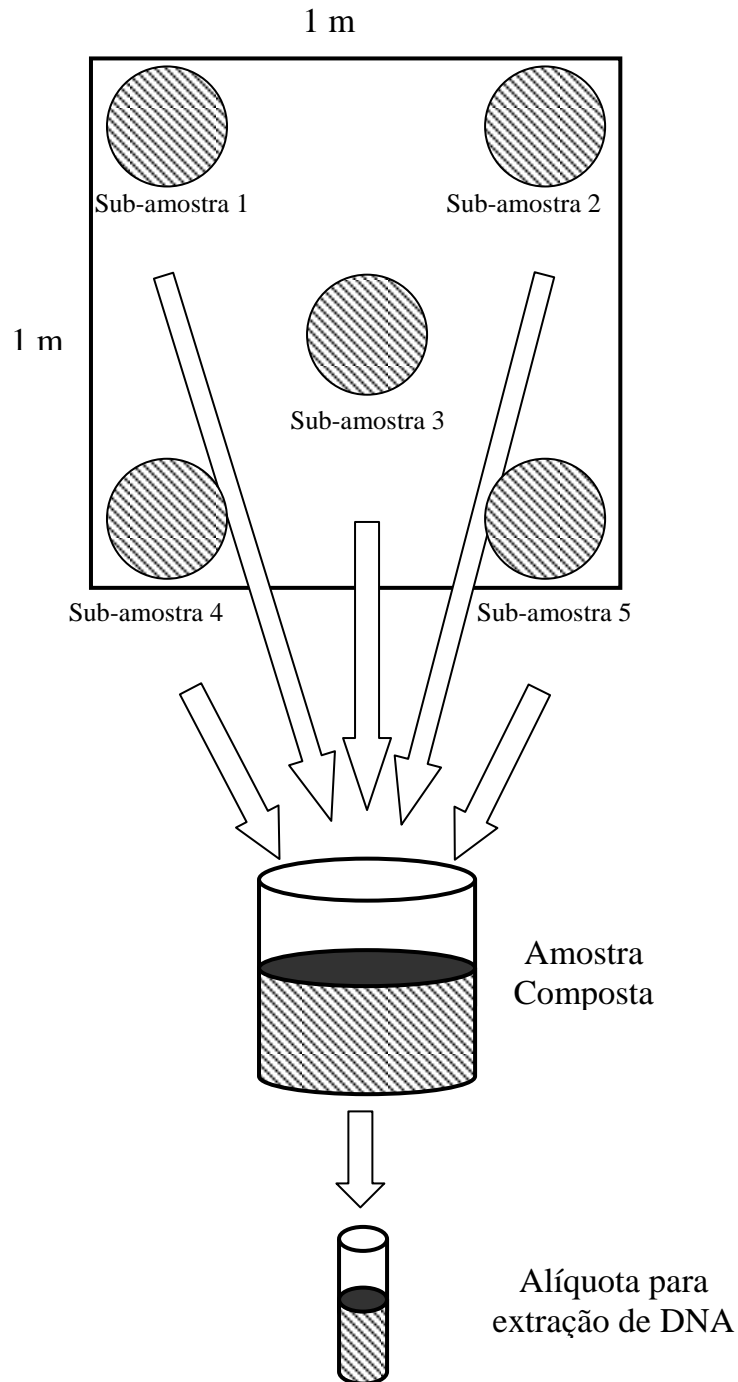
Enzimas de restrição e modificadoras de DNA, RNase, tampões e sais utilizados para preparar as soluções utilizadas na manipulação do DNA foram todas de alto grau de pureza (grau biologia molecular) obtidas de diferentes fornecedores.

4.3 Coleta das amostras de solo e extração de DNA total

A coleta de amostras de solo (MAF1, 2 e 3) foi realizada em uma área de 1 m². Essa área foi superficialmente limpa para retirada da matéria orgânica em decomposição e em seguida foram retiradas 5 subamostras de solo de uma profundidade entre 0 e 20 cm com o auxílio de ferramentas de jardinagem. As 5 subamostras foram unidas em um recipiente e misturadas originando a amostra composta que foi conservada em gelo até o retorno ao laboratório. A partir dessa amostra composta (Figura 10) foram retiradas alíquotas de 10 g de solo para extração do DNA.

Depois de cada coleta as ferramentas eram lavadas em água corrente e desinfetadas com álcool 70% para evitar contaminações de um solo para outro.

FIGURA 10: Método de coleta das amostras de solo da Floresta Atlântica Paranaense



Em cada ponto de coleta foi estabelecido uma área de 1 m^2 . Dessa área foram retiradas 5 subamostras de solo que deram origem à amostra compostas. A partir da amostra composta foi retirada a alíquota de 10 g de solo utilizada na extração de DNA.

4.4 Manipulação de DNA

4.4.1 Purificação de DNA

A purificação de DNA das amostras de solo foram realizadas utilizando o kit *PowerMax Soil DNA Isolation Kit* (MO BIO Laboratories), a partir de 10 g de solo, seguindo-se as recomendações do fabricante. As células bacterianas foram lisadas quimicamente com solução contendo SDS e NaOH e mecanicamente por agitação na presença de esferas de sílica visando máxima obtenção de DNA. Em seguida, o material foi centrifugado, o sobrenadante coletado e tratado com solução removedora de inibidores (ácidos húmicos). Por fim, o DNA foi aderido a uma membrana de sílica, lavado duas vezes e eluído em água ultrapura. O DNA purificado foi quantificado utilizando-se o sistema Qubit da Invitrogen.

A purificação de DNA fosmidial foi realizada a partir de culturas de bactérias multiplicadas em meio líquido. As bactérias contendo os fósmites selecionados foram inoculadas em frascos de cultivo de 25 ml contendo 5 ml de meio Terrific Broth suplementado com 12,5 µg/mL de cloranfenicol para seleção e 0,001% de L-arabinose para indução do número de cópias do fósmito (WILD, HRADECNA e SZYBALSKEY, 2002). Os clones bacterianos foram cultivados a 180 rpm a 37°C por aproximadamente 16 horas. A extração do DNA foi realizada em tubos plásticos de 1,5 mL pelo método de lise alcalina (SAMBROOK, FRITSCH e MANIATIS, 1989) utilizando 3 mL de cultura.

A purificação de DNA plasmidial em tubo plástico de 1,5 mL foi realizada a partir de 1,5 mL de cultura multiplicada em frasco de 25 mL contendo 5 mL de meio líquido LB suplementado com o antibiótico, de acordo com a resistência do plasmídeo. Os clones bacterianos foram cultivados a 180 rpm a 37°C por aproximadamente 16 horas. A extração do DNA foi realizada pelo método de lise alcalina (SAMBROOK, FRITSCH e MANIATIS, 1989).

A purificação de DNA plasmidial em larga escala foi realizada em placas de 96 poços. Para crescimento bacteriano foram utilizados blocos de 96 poços com

capacidade de 2 mL por poço. Em cada poço da placa foi colocado 1,25 mL de meio Terrific Broth contendo 250 µg/mL de ampicilina. As colônias foram inoculadas com o auxílio de repicador de 96 pinos e a placa selada com adesivo perfurado para permitir a aeração. As culturas foram incubadas durante 20 horas em agitador a 37°C e 180 rpm. O bloco contendo as culturas crescidas foi centrifugado por 7 minutos a 4000 rpm para sedimentar as células. O meio de cultivo foi retirado e o precipitado de células lavado com 180 µL de GET (glucose 50 mmol/L, EDTA 10 mmol/L pH 8 e Tris-HCl 25 mmol/L pH 8). Após o descarte do sobrenadante, foi adicionado 80 µL de GET contendo 2 µg/mL de RNase. As células foram suspensas por agitação e a suspensão transferida para uma microplaca (polipropileno) de 96 poços, com capacidade para 250 µL e fundo em V. A cada poço da placa foram adicionados 80 µL de solução de lise (NaOH 0,2 mol/L e SDS 1%). A placa foi selada, misturada por inversão e incubada em temperatura ambiente por 5 min. A seguir, acrescentou-se a cada poço 80 µL de acetato de potássio (3 mol/L pH 5,2) gelado e misturou-se por inversão. A placa foi deixada 10 minutos à temperatura ambiente e, então, incubada aberta em estufa a 90°C por 30 minutos. Após esse tempo a placa foi resfriada em banho de gelo por 15 minutos e centrifugada mais 15 minutos (4000 rpm, 4°C). O sobrenadante foi, então, coletado e transferido para uma placa-filtro de 96 poços Millipore (MAGV N22), fixada sobre uma placa de fundo V de 250 µL e o conjunto centrifugado (sem a tampa) por 5 minutos a 4000 rpm. Ao filtrado que passou para a microplaca adiciona-se 0,6 volumes de isopropanol e em seguida centrifugou-se por 45 minutos a 4000 rpm. O precipitado de DNA foi então lavado com 150 µL de etanol 70% e seco em estufa a 37°C por uma hora. Em seguida o DNA foi dissolvido em 30 µL de água ultrapura.

Em casos particulares que necessitavam um grau maior de pureza do DNA a extração foi realizada com o kit QIAprep da QIAGEN segundo instruções do fabricante.

4.4.2 Sequenciamento de DNA

O sequenciamento de DNA foi realizado pelo método de terminação de cadeia, cujo princípio foi descrito por SANGER, NICKLEN e COULSON (1977), utilizando-se dideoxynucleotídeos fluorescentes, separação em eletroforese em gel de poliacrilamida vertical ou capilar e detecção do fluoróforo após excitação com laser em sequenciador automático ABI377 (sequenciamento em tubo de amostra individual) ou MegaBACE1000 (sequenciamento em placa de 96 amostras).

O sequenciamento das extremidades do inserto de DNA fosmidial foi realizado em tubos plásticos de 200 μ L. O sistema de reação foi constituído de aproximadamente 350 ng de DNA de fita dupla, 5 pmol de iniciador (Tabela 2), 4 μ L de reativo *DYEnamic™ ET Terminator mix* (GE Health Care), que contém os nucleotídeos marcados e não marcados e a enzima DNA polimerase (Thermo Sequenase II), e água ultrapura suficiente para 10 μ L. A reação de sequenciamento foi conduzida em termociclador *Eppendorf Master Cycler Gradient*, com as seguintes condições: 1 ciclo de 95°C por 1 minuto seguidos de 35 ciclos de 94°C por 30 segundos, 50°C por 45 segundos finalizando com 60°C por 90 segundos. A seguir o produto das reações foi purificado adicionando-se a cada poço 10 μ L de água ultrapura, 2 μ L de acetato de amônio 7,5 mol/L e 3 volumes de etanol absoluto. Após 20 minutos de incubação em temperatura ambiente o DNA foi coletado por centrifugação a 13400 rpm por 15 minutos, lavado com 100 μ L de etanol 70% e seco em estufa à 37°C. O DNA seco e purificado foi dissolvido em tampão de corrida (formamida, EDTA pH 8,00 25 mM e Blue dextran 50 mg/mL) e aplicado no sequenciador automático.

O sequenciamento das extremidades do inserto de DNA plasmidial foi realizado em tubos individuais de 200 μ L ou em placas de 96 amostras utilizando-se aproximadamente 350 ng de DNA fita dupla, 3,25 pmol de iniciador (Tabela 4), 3 μ L de reativo *DYEnamic™ ET Terminator mix* e água ultrapura suficiente para 7,5 μ L. A reação de sequenciamento foi conduzida em termociclador *Eppendorf Master Cycler Gradient* com as seguintes condições: 1 ciclo de 95°C por 1 minuto seguidos de 30

ciclos de 94°C por 30 segundos e 62°C por 2 minutos. A seguir o produto das reações foi purificado adicionando-se a cada reação 12,5 µL de água ultrapura, 2 µL de acetato de amônio 7,5 mol/L e 3 volumes de etanol absoluto. O DNA foi coletado por centrifugação a 13.400 rpm por 15 minutos (tubo) ou 4.000 rpm por 45 minutos (placa), lavado com 1 ml ou 150 µL de etanol 70% e seco em estufa à 37°C. O DNA seco e purificado foi dissolvido em tampão de corrida (tubo) ou em água (placa) e aplicado no sequenciador automático. O sequenciamento de DNA a partir de transposon foi realizado pelo mesmo método do sequenciamento plasmidial em tubo.

TABELA 4: Iniciadores utilizados para sequenciamento de DNA.

Vetor	Iniciador	Sequência	Referência
pUC18	Universal	5'GACGTTGTAAAACGACGGCCAGT	Invitrogen
	Reverso	5'TTTCACACAGGAAACAGCTATGAC	Invitrogen
pCC2FOS	Direto	5'GTACAACGACACCTAGAC	EPICENTRE
	Reverso	5'CAGGAAACAGCCTAGGAA	EPICENTRE
pET28a	T7promotor	5'TAATACGACTCACTATAGGG	Novagen
	T7terminador	5'GCTAGTTATTGCTCAGCGG	Novagen
Transposon	KAN-2 FP1	5'ACCTACAACAAAGCTCTCATCAACC	EPICENTRE
	KAN-2 RP1	5'GCAATGTAACATCAGAGATTTTGAG	EPICENTRE

4.4.3 Preparo de células eletrocompetentes e transformação bacteriana

As estirpes de *E. coli* utilizadas para transformação foram cultivadas em 200 ml de meio SOB até D.O._{600nm} entre 0,6 e 0,8. Em seguida, a suspensão de células foi incubada em banho de gelo por 30 minutos. Posteriormente, as células foram coletadas por centrifugação em tubos tipo Falcon de 50 ml a 5.000 rpm por 5 minutos a 4°C em centrífuga *Eppendorf 5804R* utilizando o rotor de ângulo fixo, também *Eppendorf*, modelo F 34-6-38. O precipitado de células foi lavado 3 vezes com glicerol 10% gelado e suspenso em 1 mL da mesma solução.

A eletrotransformação foi realizada com alíquotas de 200 µL de células eletrocompetentes utilizando-se 1 µL de DNA, proveniente de reação de ligação,

reação de inserção de transposon ou fosmídeo purificado, ou 0,5 μ L de DNA proveniente de plasmídeo purificado. A mistura foi transferida para uma cubeta de eletroporação de 0,2 cm (BioRad) previamente resfriada e colocada na câmara de eletroporação. As células foram submetidas a um choque elétrico de 2,5 kV utilizando o eletroporador *GenePulser II* (BioRad). Às células transformadas de *E. coli* foram adicionados 800 μ L de meio SOC. A solução foi transferida para um frasco de 10 mL estéril e incubadas por 40 minutos em agitação a 180 rpm a 37°C. Após o tempo de recuperação, as amostras foram plaqueadas em meio sólido LA contendo o antibiótico para as respectivas marcas de resistência. No caso da construção de sub-bibliotecas também foi adicionado às placas 30 μ g/ml de X-gal (5-bromo-4-cloro-3-indolil- β -D-galactopiranosídeo) como substrato para a enzima β -galactosidase. As placas foram incubadas na estufa à 37°C.

4.4.4 Construção das bibliotecas metagenômicas

As bibliotecas metagenômicas foram construídas em vetor fosmídeo pCC2FOS utilizando o kit *CopyControl™ HTP Fosmid Library Production* da EPICENTRE segundo recomendações do fabricante. As bibliotecas foram construídas em 4 etapas: (1°) Reparo e seleção dos fragmentos de DNA ambiental entre 40 e 50 kb, (2°) Ligação dos fragmentos de DNA no vetor pCC2FOS e empacotamento dos fosmídeos em fago lambda, (3°) Transformação de *E. coli* EPI300 por infecção com fago lambda e (4°) Coleta e armazenagem dos clones transformantes. A descrição de cada etapa é apresentada a seguir.

4.4.4.1 Reparo e seleção dos fragmentos de DNA ambiental entre 40 e 50kb

Aproximadamente 20 μ g do DNA total extraído de cada amostra de solo (MAF1, 2 e 3) foram submetidas, individualmente, a uma reação de reparo e fosforilação das extremidades dos fragmentos com *End Repair Mix Enzyme* (EPICENTRE) durante 45 minutos a 30°C. Após esse período, as enzimas foram

inativadas por calor durante 20 minutos a 80°C. O produto dessa reação foi aplicado em um gel de agarose de baixo ponto de fusão e submetido por 26 horas a 30 V em tampão TAE (Tris acetato 40 mM e EDTA 1 mM). A faixa do gel correspondente ao marcador de peso molecular 5 kp (Gibco) foi excisada e corada com brometo de etídio. A posição da banda de 40 kb do marcador de peso molecular foi identificada por exposição da faixa corada à luz UV e marcada. Essa faixa de gel correspondente ao marcador foi posicionada ao lado do gel contendo o DNA metagenômico e atuou como guia para a seleção dos fragmentos de interesse entre 40 e 50 kb.

A purificação dos fragmentos de DNA do gel de agarose foi realizada utilizando o método do fenol. A fatia de gel contendo os fragmentos de DNA de tamanho desejado foi fundida em banho de água quente a 65°C e o volume de banda corrigido para 500 µL com água ultrapura. A esse volume foi adicionado NaCl, Tris-HCl pH8,00 e EDTA pH 8,00 nas concentrações finais de 250 mM, 100 mM e 1 mM, respectivamente. Em seguida, foi adicionado 1 volume de fenol puro equilibrado e a solução misturada cuidadosamente por inversão para evitar a quebra mecânica do DNA. Essa solução foi centrifugada 15 minutos e o sobrenadante foi coletado para um tubo novo. O DNA foi precipitado com 1 volume de isopropanol, lavado 2 vezes com etanol 80%, seco em estufa a 37°C e dissolvido em 10 µL de água ultrapura.

4.4.4.2 Ligação dos fragmentos de DNA no vetor e empacotamento dos fósmeios em fago lambda

Cerca de 250 ng do DNA purificado foram misturadas a um sistema de ligação contendo 500 ng de vetor fósmeio pCC2FOS (Cm^R), 1 µL de tampão de ligação *Fast Link* 10X, 1 µL de ATP 10 mM, 2 U de enzima ligase *Fast Link* e água ultrapura para um volume final de 10 µL. O sistema de ligação foi incubado em temperatura ambiente durante 16 horas e, no dia seguinte, a enzima foi inativada por calor 20 minutos a 80°C. Ao sistema de ligação foram adicionados 25 µL do extrato de empacotamento, também fornecido com o kit *CopyControl™ HTP Fosmid Library Production*, e a reação foi incubada durante 90 minutos a 30°C. Após completar esse

período, mais 25 μL do extrato de empacotamento foram adicionados à reação e o sistema foi incubado por outros 90 minutos a 30°C. Ao final da segunda incubação foram adicionados ao sistema 940 μL de tampão de diluição do fago (Tris-HCl pH 8,3 10 mM, NaCl 100 mM, MgCl_2 10 mM) e 25 μL de clorofórmio. O sistema foi misturado cuidadosamente por inversão e estocado a 4°C.

4.4.4.3 Transformação de *E. coli* EPI300 por infecção com fago lambda.

O preparo do pré-inóculo de *E. coli* EPI300 foi feito inoculando uma colônia em 5 mL de meio líquido LB suplementado com 10 mM de MgSO_4 e incubada a 37°C em agitador durante 16 horas a 180 rpm. No dia seguinte, todo o volume do pré-inóculo foi transferido para um Erlenmeyer de 250 mL contendo 50 mL de meio líquido LB suplementado com 10 mM de MgSO_4 . A cultura foi novamente incubada a 37°C em agitador orbital até atingir uma $\text{D.O.}_{600\text{nm}}$ entre 0,8-1,0. Para a infecção, 100 μL dessa cultura foram misturados a 1 μL do sistema de empacotamento e incubados a 37°C por 20 minutos. Em seguida, foram adicionados mais 400 μL de meio LB ao sistema de infecção passando por outra incubação a 37°C por 20 minutos. Por fim, as células infectadas foram espalhadas em placas de petri grandes (150 mm x 15 mm) contendo meio sólido LA e 12,5 $\mu\text{g/mL}$ de cloranfenicol. Esse processo foi repetido até o término do sistema de empacotamento.

4.4.4.4 Coleta e armazenamento dos clones transformantes

A coleta dos clones transformantes foi feita manualmente com auxílio de palitos de madeira estéreis ou automaticamente com auxílio do robô GeneTac (*Genomic Solutions*). As bactérias coletadas foram transferidas para uma placa de 96 poços contendo 150 μL de meio líquido Terrific Broth mais 12,5 $\mu\text{g/mL}$ de cloranfenicol e incubadas em estufa a 37°C durante a noite. Às culturas multiplicadas

foram adicionados 150 μL de glicerol 50%. (m/v) As placas foram seladas com adesivos plásticos, agitadas em aparelho tipo vortex e armazenadas a -24°C .

4.4.5 Construção de sub-bibliotecas

Clones de interesse foram submetidos à um processo de subclonagem com o objetivo de reduzir o tamanho do DNA inserto possibilitando, desse modo, o sequenciamento completo do inserto de DNA ambiental clonado em fosmídeo.

4.4.5.1 Extração de DNA fosmidial

A purificação de DNA fosmidial para subclonagem foi realizado pelo método de lise alcalina a partir de 20 mL de cultura seguindo os procedimentos descritos no item 4.4.1.

4.4.5.2 Nebulização

O DNA fosmidial purificado foi nebulizado para a obtenção de fragmentos de DNA de tamanhos entre 10 e 3 kb. O DNA alvo foi adicionado em um sistema de nebulização contendo 600 μL de glicerol 50%, 100 μL de acetato de sódio 3 M e água ultrapura para um volume total de 1000 μL . O procedimento foi realizado em uma câmara de nebulização utilizando-se gás argônio à pressão de 100 kPa por 3 segundos. Após a nebulização, o DNA total foi dividido entre três tubos de 1,5 mL e precipitado com um volume de isopropanol durante uma hora à temperatura ambiente. Após completo esse tempo, a solução foi centrifugada 20 minutos a 13.400 rpm, o sobrenadante foi descartado, o precipitado foi lavado 2x com etanol 70%, seco em estufa a 37°C e suspenso em água ultrapura.

4.4.5.3 Reparo e fosforilação das extremidades dos fragmentos de DNA nebulizado

O DNA nebulizado foi submetido a uma reação de reparo e fosforilação das extremidades dos fragmentos gerando, dessa maneira, fragmentos com extremidades cegas fosforiladas na ponta 5' permitindo a sua ligação a um vetor de clonagem. O sistema de reparo foi composto pelo DNA alvo, 4 µL de tampão de reação *React 2* 10X (Invitrogen), 1 µL de dNTP 5 mM, 1 µL de ATP 100 mM, 10 U da enzima polinucleotídeo quinase (Fermentas), 5 U de T4 DNA polimerase (Invitrogen), 10 U de Klenow (Fermentas) e água ultrapura para um volume total de 40 µL. A reação de reparo foi conduzida em termociclador *Eppendorf Master Cycler Gradient* sob as seguintes condições: 12°C por 40 minutos, 37°C por 45 minutos finalizando com 75°C por 10 minutos.

4.4.5.4 Seleção de fragmentos para construção das sub-bibliotecas

O DNA produto da reação de reparo foi aplicado em um gel de agarose de baixo ponto de fusão e corrido por 14 horas a 12 V em tampão TAE 1X (40 mM Tris-acetato e 1 mM de EDTA). A faixa do gel correspondente ao marcador de peso molecular *1kb gene ruler* (Fermentas) foi excisada e corada com brometo de etídio atuando como guia para a seleção do tamanho dos fragmentos de interesse (entre 3 e 5 kb). A purificação dos fragmentos de DNA do gel de agarose foi realizada utilizando o kit *GFX PCR DNA and Gel Band Purification* (GE).

4.4.5.5 Ligação dos fragmentos de DNA no vetor pUC18

Os fragmentos de DNA purificados da agarose de baixo ponto de fusão foram ligados no vetor pUC18. A reação de ligação foi feita utilizando-se 10 ng de vetor pUC18, linearizado com a enzima *SmaI* e desfosforilado, aproximadamente 30 ng

DNA purificado, 1 U de enzima T4 DNA ligase (USB), tampão de ligação diluído 10 vezes (660 mM Tris-HCl pH 7.6, 66 mM MgCl₂, 100 mM DTT, 660 μM ATP) e água ultrapura estéril para um volume final de 30 μL. O sistema de ligação foi incubado em temperatura ambiente durante 24 horas.

4.4.5.6 Coleta de clones transformantes

O produto da reação de ligação foi transformado por eletroporação na estirpe TOP10 de *E. coli* (item 4.4.3) e as células plaqueadas em meio sólido LA contendo 250 μg/mL de ampicilina e 30 μg/mL de X-gal. As colônias transformantes (brancas) foram coletadas aleatoriamente com o auxílio de palitos de madeira e organizadas em placas de petri contendo meio sólido LA e ampicilina. Após o crescimento em meio sólido as bactérias foram inoculadas em placas do tipo ELISA de 96 poços preenchidas com 100 μL de meio líquido Terrific Broth contendo ampicilina 250 μg/mL e incubadas durante a noite em estufa 37°C. No dia seguinte foram adicionados 100 μL de glicerol 87% estéril às culturas. As placas foram agitadas, seladas com adesivo plástico e armazenadas em congelador a -24°C.

4.4.6 Inativação de genes por mutagênese aleatória com transposon

Todos os clones das sub-bibliotecas fosmidiais foram testados em placas de meio sólido contendo 1% de tributirina para a identificação de clones positivos. Subclones positivos de interesse foram submetidos à mutagênese aleatória mediada por transposon para confirmação do gene que codifica a enzima lipase. A mutagênese foi realizada utilizando o kit de inserção de transposon EZ::TN <KAN-2> da EPICENTRE de acordo com as instruções do fabricante. O sistema de reação foi composto por aproximadamente 500 ng de DNA dupla fita, 1 μL de tampão de reação (Tris-acetato pH 7,5 0,5 M, acetato de potássio 1,5 M, acetato de magnésio 0,1 M e espermidina 0,04 M) concentrado 10X, 0,3 μL de transposase (1 U/μL), 0,3 μL de

transposon (0,1 pmol/ μ L) e água ultrapura para um volume final de 10 μ L. A reação foi incubada a 37°C durante 4 horas. Após esse período foi adicionado 1 μ L de solução de parada (SDS 1% m/v) e a reação foi incubada a 70°C durante 20 minutos. O sistema de reação (1 μ L) foi transformado em *E. coli* estirpe TOP10 e os transformantes foram selecionados em placas de meio sólido LA contendo 50 μ g/mL de canamicina. Os clones coletados nessa etapa foram testados em meio sólido contendo 1% de tributirina (m/v) para seleção de clones que não fossem capazes de produzir halo de hidrólise indicando que a inserção ocorreu no gene da lipase. Os sítios de inserção foram determinados por sequenciamento de DNA com os iniciadores KAN-2 FP-1 e RP-1 (Tabela 4) fornecidos junto com o kit. O método de sequenciamento utilizado foi o mesmo descrito para sequenciamento de plasmídeos.

A técnica de inserção de transposon também foi utilizada para auxiliar a obtenção de sequências entre contigs provenientes do sequenciamento tipo *shotgun*. Plasmídeos cuja sequência cobria integralmente os buracos foram submetidos à reação de transposon e sequenciados integralmente como descrito acima. Essas sequências foram alinhadas às anteriores pelo programa Phred-Phrap-Consed.

4.4.7 Amplificação e clonagem dos genes que codificam lipases

Iniciadores utilizados para a amplificação e clonagem dos genes que codificam as enzimas lipases foram desenhados utilizando o programa *PrimeSelect* disponível no pacote de programas *LaserGene* (DNASTAR, Madison, WI, USA). As sequências dos iniciadores construídos são apresentadas na Tabela 5. Em todos os iniciadores foram inseridos sítios para enzimas de restrição com o objetivo de facilitar a clonagem do gene amplificado em um vetor de expressão.

Os genes codificadores de lipases foram amplificados por PCR (*Polymerase Chain Reaction*) em um sistema contendo 1 U de Taq DNA polimerase de alta fidelidade PlatinumTMPfx, 5 pmol de cada iniciador descrito acima, 200 μ M de dNTP, 5 μ L de tampão de reação (Tris-HCl pH 8,4 e KCl 500 mM) concentrado 10X, 1 μ L de MgSO₄ 50 mM, aproximadamente 10 ng de DNA molde (fosmídeo) e água

ultrapura para um volume total de 50 μ L. A reação foi conduzida em termociclador *Eppendorf Master Cycler Gradiente* com o seguinte programa: 1 ciclo de 94°C por 30 segundos seguidos de 35 ciclos sequenciais de 94°C 30 segundo, 50°C 45 segundos e 68°C 2 minutos finalizando com um ciclo de 68°C por 10 minutos.

O produto de PCR foi purificado em coluna de sílica utilizando-se o kit *GFX PCR DNA and Gel Band Purification* (GE Healthcare). Em seguida, o DNA foi digerido com as enzimas de restrição do sítio mutagênico inserido. As endonucleases foram inativadas por calor 20 minutos a 80°C e o DNA amplificado foi novamente purificado utilizando kit *GFX PCR DNA and Gel Band Purification* (GE). Por fim, o produto de PCR foi ligado no vetor de expressão pET28a digerido com as mesmas enzimas e purificado da mesma maneira. A reação de ligação foi feita utilizando-se 10 ng de vetor, aproximadamente 50 ng do produto de PCR, 5 U de enzima T4 DNA ligase (Fermentas), tampão de reação diluído 10 vezes (Tris-HCl 400 mM, MgCl₂ 100 mM, DTT 100 mM, ATP 5 mM) e água ultrapura estéril para um volume final de 10 μ L. A reação foi incubada a temperatura ambiente por 24 horas.

TABELA 5: Iniciadores utilizados na amplificação dos genes codificadores de lipases

Gene	Iniciador	Sequência	Sítio de restrição
LP001ORF27	LipA01For	5'GATGCCCATATGAAACCTA	<i>NdeI</i>
	LipA01Rev	5'GTGGGATCCTCAAAGCGT	<i>BamHI</i>
LP090ORF24	LipB12For	5'AATAGCATATGGCCCCTTCG	<i>NdeI</i>
	LipB12Rev	5'AGGATCCCTCACCGACTGC	<i>BamHI</i>
LP018ORF15	PepB03For	5'GTGCTAGCATGCCAGAC	<i>NheI</i>
	PepB03Rev	5'AGGAATTCCTATAACCGCC	<i>EcoRI</i>
LP018ORF16	HipB03For	5'ATTCAGCATATGAGTTTGT	<i>NdeI</i>
	HipB03Rev	5'ATGTGGATCCATTACGAGT	<i>BamHI</i>

4.5 Bioprospecção das bibliotecas metagenômicas por clones de interesse

A identificação de clones com fenótipos de interesse foi realizada em meio de cultivo sólido específico para cada situação. Clones que apresentaram mudança morfológica ou produção de compostos coloridos em qualquer tipo de meio também foram selecionados. A composição dos meios seletivos para cada atividade enzimática é apresentada na Tabela 6.

A detecção de atividade esterase foi realizada em meio sólido LA contendo 1% de tributirina. A atividade foi evidenciada pelo aparecimento de um halo claro em volta da colônia após 3 dias de incubação a 37°C. A atividade esterase/lipase também foi avaliada em meio sólido contendo 1% de tricaprilina, pelo aparecimento de um halo claro em volta da colônia após 7 dias de incubação a 37°C, ou em meio sólido contendo 1% de trioleína pelo aparecimento de um halo precipitado de cálcio-ácido graxo em volta da colônia após 7 dias de incubação a 37°C.

A detecção de atividade amilase foi realizada em meio sólido com 2% de amido solúvel e coloração com solução de iodo (KI 50 g/L e I₂ 2,03 g/L) após 7 dias de incubação a 37°C (YUN et al, 2004). A amilose (amido não ramificado com ligações glicosídicas α 1-4) apresenta uma estrutura em hélice capaz de reter moléculas de iodo formando um complexo azul escuro. Quando a bactéria produz amilase ela irá degradar o amido a sua volta e impedir a formação desse complexo produzindo um halo claro em volta da colônia.

A atividade protease foi avaliada pela formação de um halo translúcido em volta das colônias crescidas em meio sólido LA contendo 2% (m/v) de leite desnatado sem gordura após incubação por 7 dias a 37°C (LÄMMLE et al, 2006).

Para o preparo dos meios de esterase, lipase e amilase todos os componentes foram pesados, dissolvidos em água destilada, misturados com auxílio de mixer elétrico durante 1 minuto e autoclavados. O meio para detecção de protease foi feito diluindo-se 10 vezes uma solução de leite 20% previamente autoclavada em meio LA fundido para se obter o meio LA leite 2%.

O vetor pCC2FOS usado para construir as bibliotecas possui duas origens de replicação: *Ori2* e *OriV*. Quando o fósforo é replicado a partir da *Ori2* ele é mantido a uma cópia por célula garantindo a sua estabilidade. A *OriV*, por sua vez, é uma origem de replicação induzível e a replicação a partir dela pode aumentar o número de cópias do fósforo para cerca de 200 por célula. A indução da replicação por *OriV* é dependente da proteína de replicação TrfA. Na estirpe de *E. coli* EPI300, usada como hospedeira, o gene *trfA* está sob controle do promotor p_{BAD} e da proteína regulatória do metabolismo de arabinose AraC. Quando L-arabinose está presente ela se liga à proteína AraC causando uma mudança conformacional que permite sua ligação ao promotor p_{BAD} e a ativação da transcrição dos genes adjacentes. O gene *trfA*, sob controle desse promotor, é transcrito e traduzido na proteína TrfA que permite a replicação do fósforo a partir da *OriV*. Esse sistema é capaz de aumentar o número de cópias do fósforo de 1 para até 200 permitindo maior rendimento de DNA (WILD, HRADECNA e SZYBALSKEY, 2002). Em todos os meios de análise foi adicionado 0,01% (m/v) de L-arabinose com o objetivo de aumentar o número de cópias por célula do fósforo e, conseqüentemente, o número moléculas de proteínas expressas, facilitando a identificação do fenótipo.

TABELA 6: Composição dos meios seletivos para identificação de atividade enzimática

COMPONENTES	ATIVIDADE ENZIMÁTICA				
	Esterase/Lipase	Lipase1	Lipase2	Amilase	Protease
Extrato de levedura (3 g/L)	sim	sim	sim	sim	sim
Peptona (5 g/L)	sim	sim	sim	sim	sim
Agar (15 g/L)	sim	sim	sim	sim	sim
Goma Arábica (10 g/L)	sim	sim	sim	-	-
Tributirina (10 mL/L)	sim	-	-	-	-
Tricaprilina (10 mL/L) ^a	-	sim	-	-	-
Trioleína (10 mL/L)	-	-	sim	-	-
Amido solúvel (20 g/L)	-	-	-	sim	-
Leite 20% (100 mL/L)	-	-	-	-	sim
L-arabinose (0,1 g/L)	sim	sim	sim	sim	sim

4.6 Manipulação de proteínas

4.6.1 Eletroforese em gel de poliacrilamida em condição desnaturante

A eletroforese em gel de poliacrilamida, que utiliza o SDS como agente desnaturante de proteína, é denominada SDS-PAGE (*SDS Polyacrylamide Gel Electrophoresis*) e é capaz de separar proteínas de acordo com sua massa molecular. A desnaturação é promovida pela ligação do SDS à proteína formando um complexo proteína-SDS de carga negativa. As análises eletroforéticas de proteínas em gel SDS-PAGE foram realizadas utilizando o sistema de minigéis *Mini-PROTEAN* da BioRad. A composição dos géis utilizados é apresentada na Tabela 7. Para a aplicação no gel as amostras de proteínas foram misturadas com tampão de amostra (Tris-HCl pH 6,8 100 mM, glicerol 20% (m/v), SDS 4% (m/v), azul de bromotimol 0,2% (m/v) e β -mercaptoetanol 2 mM) na proporção 1:1 e aquecidas a 95°C por 5 minutos. As eletroforeses foram conduzidas a 180 V durante 50-60 minutos em tampão de corrida (Tris-base 25 mM, glicina 250 mM e SDS 0,1%). Depois de finalizada a corrida, o gel foi corado em solução Coomassie Blue R-250 2,5% (USB Corporation) durante 30 minutos. Em seguida, o gel foi lavado em água ultrapura e descorado em solução descorante (água/metanol/ácido acético nas porcentagens 4,5/4,5/1,0 v/v/v).

TABELA 7: Composição do gel SDS-PAGE utilizado para análise de proteínas

Componente	Gel de empilhamento	Gel de separação
Tris-HCl pH 6,5 0,5M	1,3 mL	-
Tris-HCl pH 8,8 1,5M	-	2,5 mL
SDS 10%	50 μ L	100 μ L
Acrilamida 40%	487,5 μ L	3 mL
APS 10%	25 μ L	50 μ L
TEMED	10 μ L	10 μ L
Água ultrapura	3,162 mL	4,35 mL

4.6.2 Superexpressão de proteínas

Os plasmídeos recombinantes contendo os genes codificadores de lipases ligados no vetor de expressão pET28a foram transformados na estirpe Rosetta de *E. coli*. Os transformantes foram inoculados em meio LB e multiplicados a 37°C até uma D.O._{600nm} de 0,5-0,6. Nesse ponto foi adicionado ao meio de cultivo o indutor de expressão IPTG (Isopropil β-D-1-tiogalactopiranosídeo) na concentração final de 0,5 mM. Para o teste de expressão 10 mL de cultura foram induzidos durante 3 horas a 37°C e a 180 rpm. Em seguida as células foram coletadas por centrifugação a 5.000 rpm durante 5 minutos em centrífuga refrigerada a 4°C. O sobrenadante foi removido e as células ressuspensas e lisadas por sonicação em tampão de lise (Tris-HCl pH 8,0 150 mM, NaCl 500 mM e glicerol 10%). A expressão das proteínas de interesse foi verificada em gel SDS-PAGE.

A solubilidade das proteínas superexpressas também foi verificada. Nesse teste foram utilizados 10 mL de cultura induzidos à 18°C durante 14 horas a 120 rpm. As baixas temperatura e agitação têm como objetivo diminuir as taxas de transcrição e tradução e diminuir a quantidade de proteína superexpressa por unidade de tempo, aumentando as chances de a proteína ser enovelada corretamente e não agregar/precipitar. O baixo nível de expressão nessas condições foi compensado pelo aumento do tempo de indução. Em seguida, as células foram coletadas por centrifugação e lisadas por sonicação em tampão de lise (Tris-HCl pH 8,0 150 mM, NaCl 500 mM, glicerol 10%, imidazol 10 mM e β-mercaptoetanol 10 mM) na presença ou ausência do detergente TritonX-100 a 0,5% (m/v). Os extratos celulares resultantes foram centrifugados a 30.000 x g por 40 minutos para separação das frações solúvel e insolúvel. A expressão e a quantidade de proteínas nas frações solúvel e insolúvel foram avaliadas em gel SDS-PAGE. Paralelamente, alíquotas das frações solúvel e insolúvel foram testadas em meio sólido agar-tributirina, agar-tricaprilina e agar-trioleína para verificar se as lipases estavam ativas e quais frações apresentavam maior atividade.

4.6.3 Purificação da lipase LP001ORF27

A purificação da lipase LP001ORF27 foi realizada a partir de 800 mL de cultura induzida a 18°C durante 14 horas a 120 rpm. A cultura foi dividida em 4 tubos plásticos de 200 mL e centrifugada a 5000 rpm em centrífuga refrigerada a 4°C por 10 minutos. Os precipitados celulares foram suspensos em tampão de lise (Tris-HCl pH 8,0 150 mM, NaCl 500 mM, glicerol 10%, imidazol 10 mM e β -mercaptoetanol 10 mM) sem detergente para um volume total de 50 ml e lisados por sonicação. O extrato celular resultante foi centrifugado a 30.000 x g por 40 minutos para separação das frações solúvel e insolúvel. Os 50 mL da fração solúvel foram carregados em uma coluna cromatográfica de níquel pré-empacotada *HisTrap™ chelating* e a cromatografia conduzida no HPLC Akta. A proteína LP001ORF27 possui uma cauda com 6 resíduos de histidina na região N-terminal que favorece sua interação com o níquel da coluna. Enquanto as demais proteínas são eluídas, LP001ORF27 fica retida na coluna. Após toda a solução de proteínas ter passado pela coluna, a proteína de interesse foi eluída com gradiente crescente de imidazol. As frações de proteínas obtidas foram analisadas em gel SDS-PAGE e dialisadas durante a noite a 4°C em tampão de diálise (Tris-HCl pH 8,0 150 mM, NaCl 500 mM, glicerol 10% e β -mercaptoetanol 10 mM). Alíquotas das proteínas dialisadas foram novamente analisadas em gel SDS-PAGE e testadas em meio sólido agar-tributirina, agar-tricaprilina e agar-trioleína para avaliação da atividade.

4.6.4 Dosagem da concentração de proteína e determinação da pureza

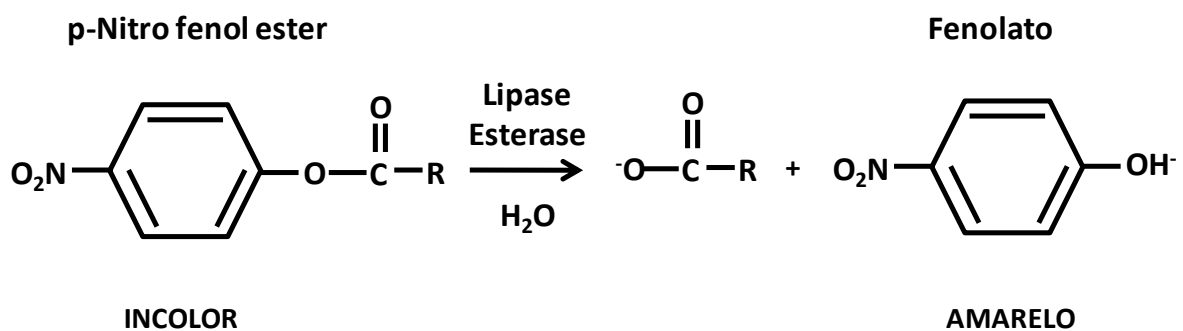
A concentração de proteínas totais foi determinada espectrofotometricamente em equipamento NanoDrop (Thermo Scientific) a 280 nm utilizando a massa molecular e o coeficiente de extinção molar da proteína LP001ORF27 fusionada com a cauda de histidina. Esses parâmetros foram obtidos submetendo a sequência de aminoácidos da proteína LP001ORF27 mais os aminoácidos que compõe a cauda de histidina à análise pelo programa Protparam (GASTEIGER et al, 2005). A pureza da

da proteína purificada foi determinada pelo software LabWorks utilizando 1 µg de proteína total.

4.7 Caracterização da lipase LP001ORF27

A caracterização enzimática da lipase LP001ORF27 foi realizada utilizando substratos artificiais, tais como ésteres de p-nitrofenila. A hidrólise da ligação éster entre a cadeia acila e o p-nitro fenol libera fenolato que, em pH neutro a básico, possui cor amarela, e pode ser medida pelo aumento da absorbância a 410 nm (Figura 11). A atividade específica foi medida em U/mg de proteína, sendo que uma unidade é a quantidade de enzima necessária para produzir 1 µmol de produto por minuto de reação. A atividade enzimática foi determinada em modo cinético por regressão linear da reta gerada a partir do aumento da absorbância proporcionada pela hidrólise do substrato (WINKLER e STUCKMANN, 1979).

FIGURA 11: Reação de hidrólise do p-Nitrofenol éster catalisada por lipases e esterase.



R representa radicais acil

4.7.1 Teste de especificidade de substrato da lipase LP001ORF27

A especificidade de substrato foi testada utilizando pNP monoacil esters com diferentes comprimentos de cadeia acila. Foram utilizados pNP acetato (C2), pNP butirato (C4), pNP hexanoato (C6), pNP decanoato (C10), pNP laurato (C12), pNP miristato (C14), pNP palmitato (C16) e pNP oleato (C18). Os substratos foram preparados na concentração estoque de 20 mM dissolvidos em solução isopropanol/acetonitrila (80/20; v/v). O sistema de reação foi composto por tampão HEPES pH 7,0 50 mM, Triton-X 100 0,3%, pNP 1 mM (um sistema diferente para cada substrato) e 50 nM de enzima purificada para um volume final de 250 μ L. O ensaio foi monitorado durante 30 minutos a 30°C no leitor de microplacas TECAN M200. Para cada substrato foram realizadas 5 reações diferentes. Como controles foram utilizados os sistemas de reação nas mesmas condições sem presença de enzima.

A atividade da enzima purificada também foi testada contra os triacilgliceróis tributirina, tricaprilina e trioleína pela titulação de ácidos graxos liberados após hidrólise utilizando o método do pH estático (*pH Stat*). As emulsões de óleo foram preparadas em solução de Tris-HCl pH 7,0 2,5 mM, NaCl 100 mM e goma arábica 1% (m/v). Antes do início da medida de titulação, o pH da emulsão foi corrigido para pH 7,0 pela adição de NaOH 0,01 mM e a reação iniciada pela adição de 50 μ g de enzima purificada. A titulação dos ácidos graxos foi monitorada no equipamento pH Stat Titrino por 5 minutos a 30°C. Os resultados representam a média de duas medições independentes.

4.7.2 Efeito de íons metálicos, quelantes de metais e detergentes sobre a atividade da lipase LP001ORF27

O efeito de íons metálicos na atividade enzimática foi determinado para Na⁺, K⁺, Ca⁺², Mn⁺², Mg⁺², Zn⁺², Fe⁺², Fe⁺³, Ni⁺, Cd⁺², Ba⁺², Ag⁺, Hg⁺, Li⁺, Sn⁺², Cs⁺, Co⁺², Cu⁺² e Rb⁺. Também foi testado o efeito de quelantes de metais EDTA e EGTA, e dos detergentes docecil sulfato de sódio (SDS), Tween20, Tween40, Tween80, N-lauril

sarcosil (N-LS), Nonidet P40 (NP40), TritonX-100 e brometo de cetiltrimetilamônio (CTAB). Íons metálicos, quelantes de metais e ânions foram testados na concentração de 1 e 10 mM. Detergentes foram testados na concentração de 0,1 e 1% (m/v). Para esses ensaios foi utilizado pNP butirato como substrato. O sistema de reação foi composto por tampão Hepes pH 7,0 50 mM, Triton-X 100 0,3%, pNP 1 mM, 50 nM de enzima purificada e o elemento a ser avaliado para um volume final de 250 μ L. O ensaio foi monitorado durante 30 minutos a 30°C no leitor de placas tipo ELISA TECAN M200. Para cada elemento foram realizadas 5 reações diferentes. Como controle foi utilizado um sistema de reação nas mesmas condições substituindo o elemento avaliado por água.

4.7.3 Efeito do pH sobre atividade da lipase LP001ORF27

O pH ótimo para atividade da enzima foi determinado medindo a atividade enzimática em soluções tampão com diferentes valores de pH. Foram utilizados os tampões acetato pH 3,0, 4,0, 5,0 e 5,5, MES (ácido 2-N-[morfolino] etano sulfônico) pH 5,5, 6,0 e 7,0, Hepes pH 7,0 e 7,5 e glicina pH 7,5, 8,0, 9,0, 10,0 e 11,0. O substrato utilizado foi o pNP butirato. Os sistemas de reações individuais foram compostos por cada tampão em diferentes pH na concentração de 50 mM, Triton-X 100 0,3%, pNP 1 mM e 50 nM de enzima purificada para um volume final de 250 μ L. O ensaio foi monitorado durante 30 minutos a 30°C no leitor de microplacas tipo TECAN M200. Como em pH ácido o pNP é incolor, o aumento da absorbância foi monitorado a 348 nm considerando o ponto isobéptico para o pNP. O ponto isobéptico se refere a um comprimento de onda específico onde espécies químicas diferentes possuem a mesma absorvidade molar, nesse caso em especial, as formas ácida e básica do pNP. Para todos os valores de pH foram realizadas 5 reações diferentes. Como controles foram utilizados sistemas de reação nas mesmas condições sem a presença de enzima.

4.7.4 Efeito da temperatura sobre a atividade da lipase LP001ORF27

O efeito da temperatura na atividade enzimática foi avaliado pela determinação da temperatura ótima de atividade e pela termoestabilidade. A temperatura ótima foi determinada medindo a atividade nas temperaturas de 20, 30, 40, 50, 60 e 70°C. Como substrato foi utilizado pNP butirato. Os sistemas de reações foram compostos por tampão Hepes pH 7,0 50 mM, pNP 2 mM e 600 nM de enzima purificada para um volume final de 1 mL. A reação foi conduzida em cubeta de quartzo (1 mL) e o aumento da absorbância foi monitorado em espectrofotômetro Shimadzu 2450 com controlador de temperatura termoelétrico (CPS 240A) acoplado. Para todas as temperaturas testadas o tampão Hepes foi pré-incubado na cela do espectrofotômetro durante 10 minutos para estabilização da temperatura. Após esse tempo, o substrato foi adicionado à cubeta, misturado por inversão 5 vezes, seguido da adição da enzima e nova mistura por inversão 5 vezes. A reação foi monitorada durante 10 minutos e para cada temperatura foram realizadas duas medições. Como controle, para cada temperatura, foi utilizado o mesmo sistema e método sem a presença da enzima. O Triton-X 100, em temperaturas iguais ou superiores a 40°C, torna-se opaco sendo por isso retirado do sistema para não comprometer a leitura da absorbância. Para determinação da termoestabilidade, alíquotas de enzima foram pré-incubadas a 50, 60 e 70°C por 1 hora seguida da determinação da atividade, à temperatura padrão de 30°C, utilizando o mesmo sistema de reação e método descrito para temperatura ótima. A medida de atividade a 30°C sem pré-incubação foi utilizada como referência (100%) para comparação com as medidas de atividade com pré-incubação.

4.8 Análises *in silico*

4.8.1 Análise de biodiversidade bacteriana de amostras de solo da Floresta Atlântica paranaense

Sequências de DNA do gene que codifica o rRNA 16S de 10 amostras de solo da Floresta Atlântica paranaense previamente obtidas (FAORO, 2006) foram submetidas à análises de bioinformática pelos programas S-LibShuff (SCHLOSS, LARGET e HANDELSMAN, 2004), LibShuff (SINGLETON et al, 2001) e DOTUR (SCHLOSS e HANDELSMAN, 2005). Para realizar essas análises as sequências parciais de 16S rRNA foram primeiramente alinhadas usando o programa ClustalW (THOMPSON, 1994) na configuração padrão e o alinhamento gerado foi usado para construção de matrizes de distância usando o programa DNAdist (FELSENSTEIN). Ambos os programas estão disponíveis no pacote de programas BioEdit (HALL, 1999). As matrizes de distância produzidas foram usadas como arquivo de entrada para os programas de análises.

Os programas S-LibShuff e LibShuff calculam a cobertura de uma biblioteca (X) em relação a outra (Y). A cobertura da biblioteca X é calculada usando fórmula $C_X(D) = 1 - [N_X(D)/n_X]$, onde D é a distância filogenética de Jukes-Cantor determinada pelo DNAdist, $N_X(D)$ é o número de sequências únicas (*singletons*) na biblioteca X para valores individuais de D e n_X é o número total de sequências na biblioteca X. A cobertura da biblioteca X pela biblioteca Y, ou a porcentagem de sequências na biblioteca X com sequências similares na biblioteca Y, C_{XY} , é calculada pela fórmula $C_{XY}(D) = 1 - [N_{XY}(D)/n_X]$, onde $N_{XY}(D)$ é o número de sequências na biblioteca X que estão a uma distância D ou maior de todas as sequências em Y. A diferença $C_X(D) - C_{XY}(D)$ (ΔC_{XY}) representa a porcentagem de sequências em X que não são singletons e não são encontradas em Y para valores individuais de D. Quando ΔC_{XY} é elevada ao quadrado (LibShuff) ela fornece a diferença entre $C_X(D)$ e $C_{XY}(D)$ para valores determinados de D, no entanto, a integração do quadrado (S-LibShuff) fornece a diferença para todos os valores possíveis de D, criando uma estatística mais confiável.

Os programas reconstróem aleatoriamente as bibliotecas X e Y recalculam o ΔC_{XY} para em cada reconstrução. A proporção da distribuição aleatória que possui valores de ΔC_{XY} maiores que os valores de ΔC_{XY} da distribuição original representa a probabilidade das diferenças observadas entre duas bibliotecas ser devida ao acaso (valor de P). Também é possível calcular o reverso da diferença ΔC_{YX} (curva heteróloga). Duas bibliotecas são consideradas diferentes quando $P < 0,05$ e consideradas iguais quando $P > 0,05$ para as duas comparações. Por fim, o programa S-LibShuff realiza uma comparação de similaridade entre duas ou mais bibliotecas utilizando apenas um arquivo de entrada, sendo ideal para casos com múltiplas bibliotecas.

O programa DOTUR (*Distance-Based OTU and Richness*) utiliza a mesma matriz de similaridade gerada pelo DNAdist e calcula o número de OTUs (*Operational Taxonomic Unit*) para todas as distâncias filogenéticas possíveis. Uma vez que as sequências foram distribuídas em OTUs é possível calcular diversas variáveis ecológicas como o índice de diversidade Shannon, o estimador de diversidade não paramétrico ACE e a curva de rarefação. O índice Shannon fornece um indicativo da diversidade da amostra baseado na abundância e frequência de espécies na amostra. O estimador ACE fornece uma estimativa de quantas OTUs existem na amostra analisada baseado no número de sequências únicas (*singletons*) e duplas (*doubletons*). Por fim, a curva de rarefação fornece um indicativo da saturação das bibliotecas, ou seja, quantas sequências faltam para identificar todas as OTUs de uma dada amostra.

4.8.2 Montagem das sequências contíguas a partir do sequenciamento de DNA dos subclones

As sequências obtidas de cada clone foram alinhadas para reconstrução da sequência completa do inserto de DNA ambiental usando o pacote de programas Phred-Phrap-Consed. O Phred (EWING et al, 1998a; EWING e GREEN, 1998b) é um programa que realiza a identificação de bases a partir do eletroforetograma (*base-calling*) e determina a confiabilidade de cada base, levando em conta anomalias na

migração eletroforética decorrente da inserção dos terminadores e do espectro de emissão dos fluoróforos. O programa utiliza os dados de sequência contidos no arquivo do eletroforetograma processado inicialmente pelo sequenciador automático. Em seguida, é executado um procedimento em 4 fases para determinar a sequência de bases mais confiável a partir desse arquivo: (1) primeiramente é determinada a localização dos picos ideais (picos preditos) usando a idéia de que os fragmentos são regularmente espaçados ao longo da matriz no capilar e, dessa forma, estimar o número correto de bases e suas localizações em regiões onde os picos não estão bem resolvidos; (2) em seguida os picos observados são identificados e (3) alinhados com os picos preditos; (4) finalmente, os picos observados que não parearam com os picos preditos na fase anterior são analisados e, se for encontrada correspondência, a base é inserida na sequência. O programa Phred também utiliza o arquivo do eletroforetograma para determinar a probabilidade de erro (p) para cada base analisada, tendo como parâmetros o espaçamento, resolução e vizinhança do pico. Esse valor p está relacionado com um índice de qualidade (q) pela equação:

$$q = -10 \cdot \log_{10}(p)$$

dessa maneira, um valor de qualidade igual a 30 representa a probabilidade de 1 erro em 1000 bases. Quanto mais altos os valores de qualidade, menor a probabilidade de erro.

O programa Phrap (www.phrap.com) foi utilizado para realizar o alinhamento entre as sequências obtidas gerando uma sequência consenso. O programa usa os dados de qualidade fornecidos pelo Phred para posicionar as bases e calcular a confiabilidade da sequência consenso. Dentre outras características do Phrap estão: a capacidade de utilizar todas as sequências geradas sem necessidade de edição de regiões com baixa qualidade, mas significativas; faz uma combinação de dados fornecidos pelo usuário e dados computados internamente aumentando a precisão da montagem; fornece uma extensa informação sobre o alinhamento, incluindo os valores de qualidade para a sequência contig; é capaz de gerenciar grandes conjuntos de dados e a sequência do contig é um mosaico constituído das melhores regiões das leituras e não apenas um consenso. Dados anormais como deleções, leituras quiméricas e

sequências provindas do vetor são automaticamente identificadas e tratadas para impedir problemas no alinhamento.

Finalmente, o programa Consed (GORDON, ABAJIAN e GREEN, 1998) foi utilizado para visualizar, editar e finalizar o alinhamento das sequências. Para ser executado, o programa requer três tipos de arquivos: os arquivos do eletroforetograma com o perfil de sinal fluorescente, arquivos *.phd criado pelo programa Phred contendo os valores de qualidade e a posição dos picos para cada base, e os arquivos *.ace criado pelo programa Phrap e que contém os dados do alinhamento incluindo a sequência contig e os valores de qualidade. Todas as informações contidas nestes arquivos são compiladas pelo Consed e através da janela criada pelo programa é possível visualizar o eletroforetograma, examinar como as leituras estavam ordenadas e orientadas e analisar os valores de qualidade para cada base, tanto na sequência individual quanto na sequência contig. Como última etapa as sequências dos iniciadores usados para realizar o sequenciamento foram localizadas e retiradas da sequência contig restando apenas a região do 16S rDNA a ser analisada.

4.8.3 Análise de sequências

A identificação de fases de leitura aberta (*Open Reading Frames* ou ORFs), a partir dos contigs reconstruídos, foi realizada com o auxílio dos programas Artemis (RUTHERFORD, 2000) e Glimmer3 (DELCHER, 1999). O Artemis é um anotador de sequências de DNA que permite a visualização dos resultados dentro do contexto da sequência com tradução para as 6 fases. A identificação de ORFs pelo Artemis é realizada basicamente pela procura entre códons de início e parada de tradução. Já o sistema Glimmer 3 pode ser treinado com um subconjunto de sequências do DNA do organismo a ser anotado podendo, desse modo, utilizar outros parâmetros (porcentagem GC e preferência de códons, por exemplo) para identificar uma ORF. Os resultados gerados pelo Glimmer foram sobrepostos aos do Artemis, produzindo uma anotação mais confiável.

As sequências de DNA das ORFs identificadas foram comparadas com as sequências depositadas no banco de dados GenBank (<http://www.ncbi.nlm.nih.gov>) utilizando-se o programa BLASTx (ALTSCHUL et al, 1997). Sequências de aminoácidos foram comparadas utilizando-se o programa BLASTp (ALTSCHUL et al, 1997). Em ambos os casos foi utilizado o banco de dados de proteínas não redundantes (nr). Características físico-químicas das proteínas foram determinadas *in silico* pelo programa ProtParam (GASTEIGER, 2005). Domínios conservados foram investigados com base no banco de dados InterPRO (APWEILER, 2001). A presença de peptídeo sinal foi determinada pelo programa SIGNALP-2.0 (NIELSEN, 1999). A identificação de tRNAs foi realizada pelo programa tRNA Scan (LOWE, 1997).

4.8.4 Análises filogenéticas

As análises filogenéticas foram conduzidas nos programas *MegaAlign*, disponível no pacote *LaserGene* (DNASTAR, Madison, WI, USA), e MEGA versão 4 (TAMURA, 2007) utilizando as sequências de aminoácidos das enzimas lipases identificadas nos insertos de DNA dos clones MAF1LP001, 018 e 090. O programa *MegaAlign* foi utilizado para análises envolvendo todas as famílias de lipase com o objetivo de reproduzir a árvore original proposta por ARPIGNY e JAEGER (1999). O programa Mega 4 foi utilizado para outras análises mais específicas. As sequências foram alinhadas pelo programa ClustalW (THOMPSON, 1994) usando a configuração padrão. As árvores filogenéticas foram construídas pelo método Neighbor Joining usando a correção de Poisson e 1000 réplicas de bootstrap. A identificação dos resíduos que formam a tríade catalítica e a fenda oxianion para lipases foi realizada com base no banco de dados LED – *Lipase Engineering Database* (PLEISS, 2000).

CAPÍTULO 5

RESULTADOS

5. RESULTADOS

5.1 Seleção das amostras de solo e construção de bibliotecas metagenômicas

5.1.1 Análise da diversidade bacteriana das amostras de solo da Floresta Atlântica paranaense.

5.1.2 Coleta das amostras de solo

5.1.3 Extração de DNA total das amostras de solo

5.1.4 Construção das bibliotecas metagenômicas

5.1 Seleção das amostras de solo e construção de bibliotecas metagenômicas

5.1.1 Análise da diversidade bacteriana das amostras de solo da Floresta Atlântica paranaense.

Em um trabalho prévio (FAORO, 2006) a diversidade bacteriana do solo da Floresta Atlântica paranaense foi investigada com base na sequência parcial do gene que codifica o 16S rRNA. Foram analisadas 10 bibliotecas de 16S rRNA provenientes de 10 amostras de solo, denominadas MA01 a MA10, coletadas em diferentes níveis de altitude, que geraram um total de 754 sequências (número de acesso no GenBank EF135620 a EF136358 e GU071058 a GU071072). A comparação destas sequências parciais contra o banco de dados do RDP (*Ribosomal Database Project*) mostrou que as amostras de solo eram dominadas por bactérias do filo *Acidobacteria*, seguido pelo filo *Proteobacteria*. No entanto, a estrutura das comunidades e as relações de diversidade entre essas 10 amostras não foi investigada, o que poderia prejudicar a seleção das amostras para a construção das bibliotecas metagenômicas. Sendo assim, antes de realizar a coleta das amostras de solo, foi necessário realizar uma análise inicial para identificar variações na comunidade e na diversidade bacteriana do solo da Floresta Atlântica paranaense (Anexo 7).

Inicialmente, as sequências parciais de 16S rRNA foram comparadas através do programa S-LibShuff para avaliar o grau de similaridade entre as bibliotecas. A análise dos valores de P para a cobertura homóloga (ΔC_{XY}) indica que as amostras MA01 a MA05 possuem uma comunidade bacteriana similar ($P > 0,05$). O mesmo ocorre com as amostras MA07 a MA10. Já a amostra MA06 aparentemente possui uma comunidade bacteriana diferente das outras (Tabela 8). A vegetação na Floresta Atlântica é dividida em costeira (0 a 50 m de altitude), sub-montana (50 a 500 m de altitude), montana (500 a 1200 m de altitude) e alta-montana (acima de 1200 m de altitude). Considerando-se esse gradiente de vegetação e o resultado obtido com o programa S-LibShuff, as amostras foram separadas em 3 grupos principais:

Alta altitude: MA01 a MA04 – entre 900 e 800 metros acima do nível do mar

Intermediário: MA05 e MA06 – 604 e 375 metros acima do nível do mar

Baixa altitude: MA07 a MA10 – entre 160 e 30 metros acima do nível do mar

Segundo essa nova divisão, o grupo de alta altitude possui 300 sequências, o grupo de baixa altitude possui 291 sequências e o grupo intermediário 169 sequências. Como os programas utilizados nas análises seguintes são sensíveis ao número de sequências, apenas os grupos de alta e baixa altitude foram utilizados para evitar a geração de artefatos. Os dois grupos foram comparados novamente através do programa LibShuff. As análises das curvas homólogas e heterólogas (Figura 12) indicam que a comunidade bacteriana no primeiro grupo é diferente daquela do segundo grupo no intervalo de distância evolutiva entre 0,0 (100% de identidade e 0% de diferença) e 0,3 (70% de identidade e 30% de diferença). Esse resultado sugere que a diversidade genética entre esses dois grupos ocorre não apenas em níveis taxonômicos baixos (espécie), como também em altos níveis taxonômicos (filo). A separação entre esses dois grupos fica mais evidente quando analisamos a curva de tendência para o índice Shannon (Figura 13), a curva de rarefação (Figura 14) e o estimador de diversidade ACE (Figura 15) gerados pelo programa DOTUR. O grupo de alta altitude apresenta uma tendência para valores maiores do índice Shannon quando comparado ao segundo grupo (considerando-se 95% de similaridade entre as sequências de 16S rRNA). De modo semelhante, a curva de rarefação para o grupo de alta altitude está menos saturada quanto comparada ao segundo grupo, indicando que mais OTUs podem ser recuperadas do primeiro do que do segundo grupo. Por fim, o estimador de diversidade ACE confirma os resultados descritos acima, indicando que o grupo de alta altitude possui potencialmente mais espécies do que o grupo de baixa altitude.

Em conjunto, esses resultados sugerem que o grupo de alta altitude possui uma comunidade mais diversa quando comparada com o grupo de baixa altitude. No entanto, a comparação entre os dois grupos pelo programa LibShuff indicou que as espécies de bactérias que compõem as duas comunidades são diferentes, o que também pode ser esperado para o grupo intermediário (Anexo 7). Essas análises permitiram a seleção das amostras de solo mais adequadas para a construção das bibliotecas metagenômicas, já que construir 10 bibliotecas seria difícil considerando os custos e o

tempo despreendido nesse processo e, posteriormente, na análise. As amostras selecionadas foram:

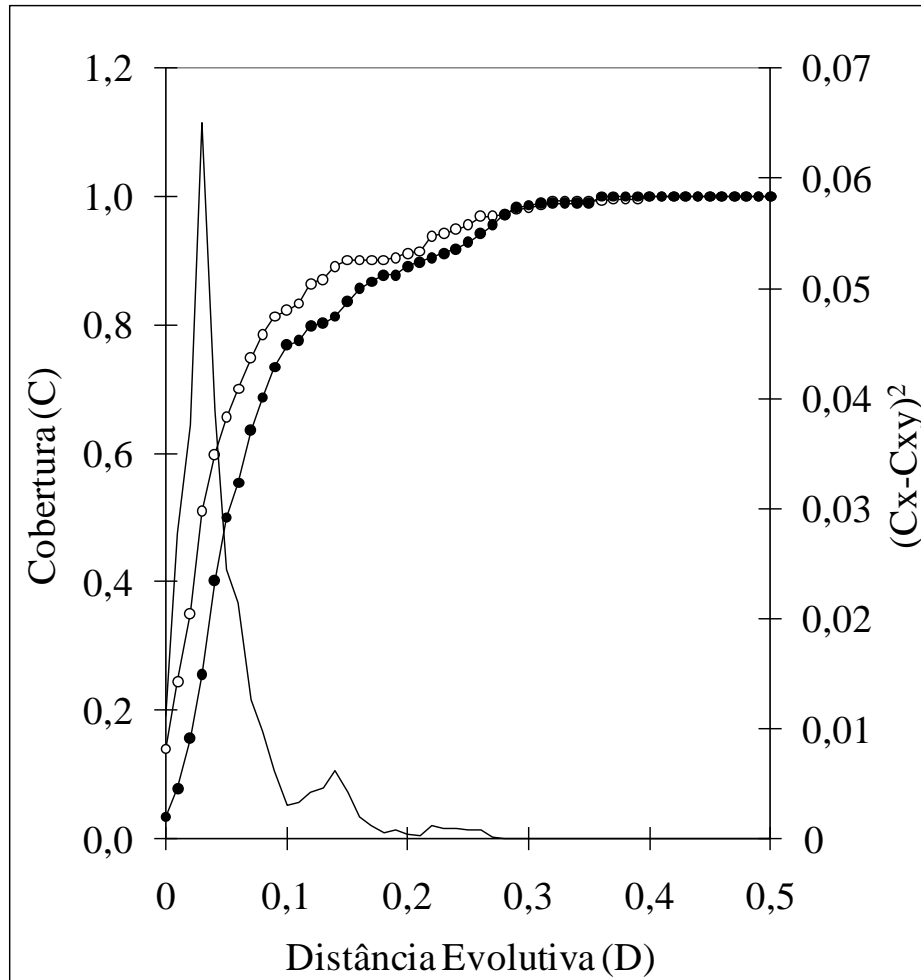
- Grupo de alta altitude: MAF1 (próxima a amostra original MA02);
- Grupo intermediário: MAF2 (próxima a amostra original MA05);
- Grupo de baixa altitude: MAF3 (próxima a amostra original MA07).

TABELA 8: Comparação entre as sequências das bibliotecas MA01 a MA10 pelo programa S-LibShuff

X	Y									
	MA01	MA02	MA03	MA04	MA05	MA06	MA07	MA08	MA09	MA10
MA01	0	0,7008	0,0017	0,5720	0,1328	0,0001	0,0072	0,0016	0,0027	0,0001
MA02	0,7275	0	0,4366	0,3047	0,2412	0,0169	0,0218	0,0123	0,0649	0,0235
MA03	0,0980	0,0968	0	0,3758	0,4719	0,1002	0,1546	0,0260	0,1093	0,0054
MA04	0,8595	0,2011	0,0801	0	0,0904	0,3712	0,0269	0,0639	0,2926	0,0203
MA05	0,1122	0,0367	0,2080	0,0608	0	0,0100	0,0004	0,0076	0,0401	0,0092
MA06	0,0001	0,0001	0,0043	0,0109	0,0032	0	0,0004	0,4532	0,0425	0,0029
MA07	0,0007	0,0005	0,0396	0,0205	0,0931	0,1343	0	0,1076	0,0117	0,0001
MA08	0,0001	0,0001	0,0009	0,0301	0,0897	0,6748	0,0040	0	0,1184	0,0002
MA09	0,0012	0,0001	0,0039	0,0226	0,4726	0,1719	0,0019	0,0395	0	0,5488
MA10	0,0001	0,0001	0,0001	0,0017	0,0175	0,0046	0,0001	0,0007	0,5284	0

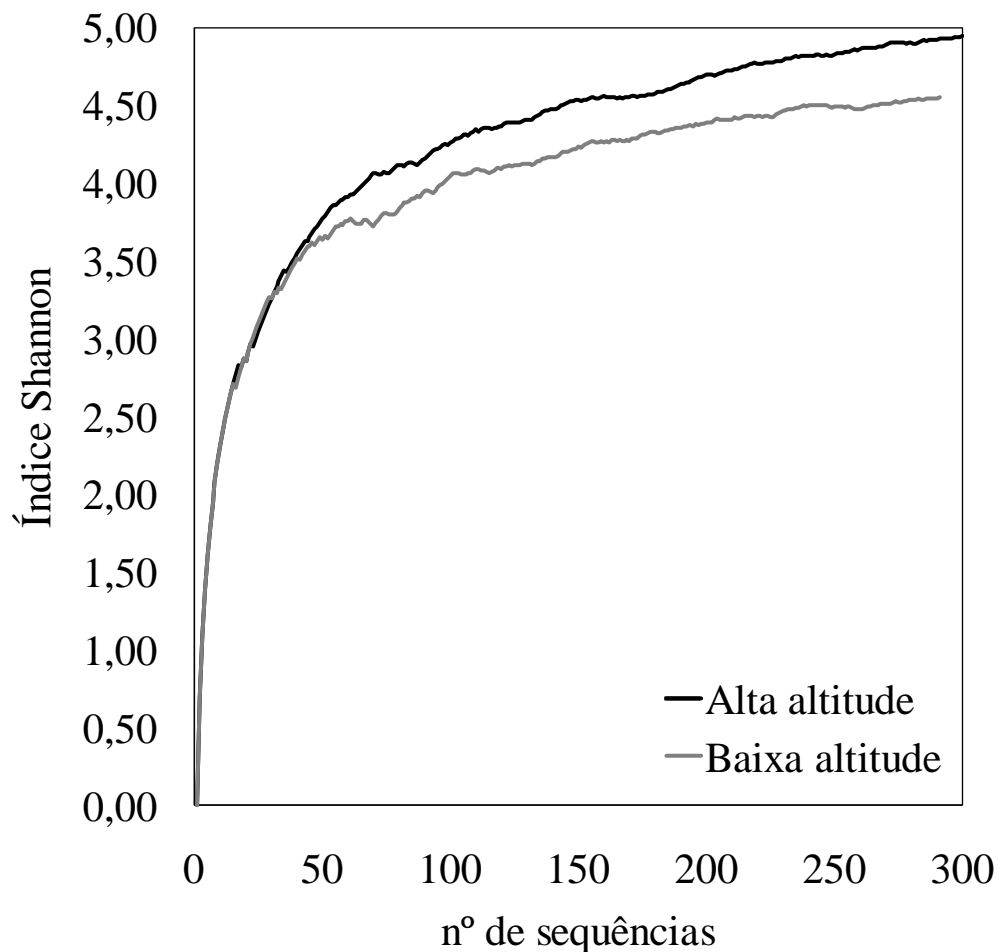
Os valores XY (acima da linha de zeros) representam os valores de P para a cobertura homóloga e os valores YX (abaixo da linha de zeros) representam os valores de P para a cobertura heteróloga. Duas bibliotecas são consideradas diferentes quando $P < 0,05$ e iguais quando $P > 0,05$ para as duas coberturas.

FIGURA 12: Comparação entre as sequências dos grupos de alta altitude e de baixa altitude pelo programa LibShuff



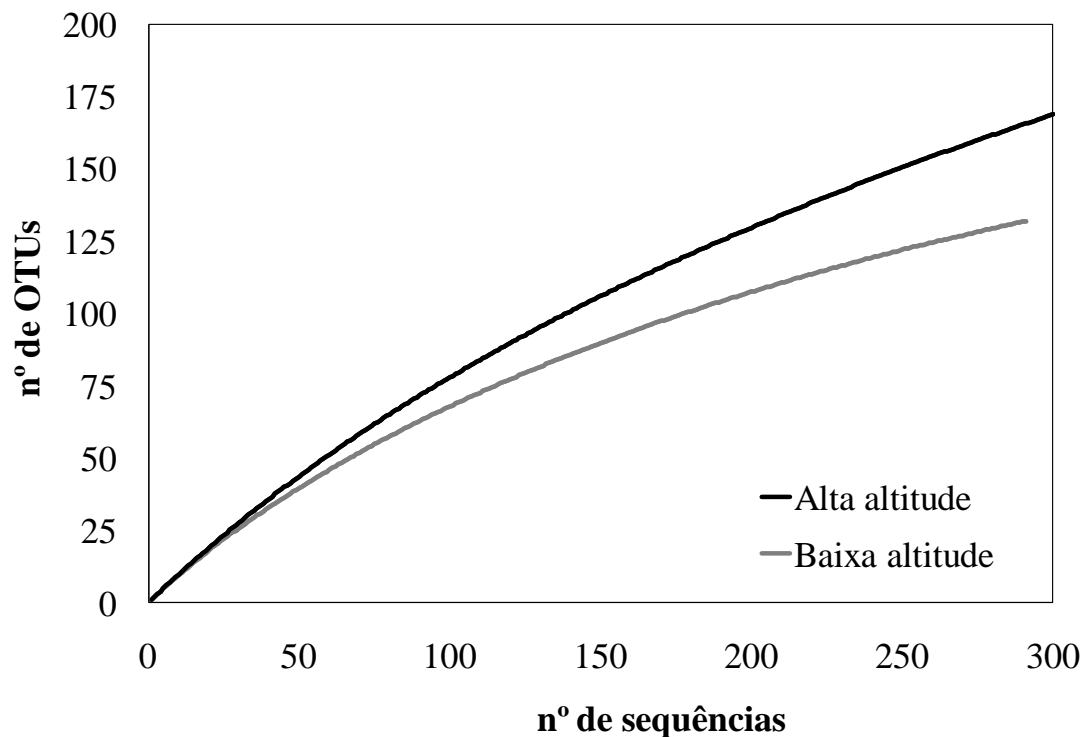
As sequências parciais de 16S rRNA dos dois grupos foram alinhadas para construção de uma matriz de identidade. A partir da matriz o programa LibShuff calcula as diferenças entre as bibliotecas. As curvas de cobertura homóloga (○) e heteróloga (●) são apresentadas. A linha sólida indica a relação $(C_X - C_{XY})^2$ para a amostra original em cada valor de D. D é igual à distância evolutiva de Jukes-Cantor determinada pelo programa DNAdist. A comparação indica que as comunidades nos grupos de alta e baixa altitude são diferentes.

FIGURA 13: Variação do índice Shannon entre os grupos de alta altitude e de baixa altitude gerado pelo programa DOTUR.



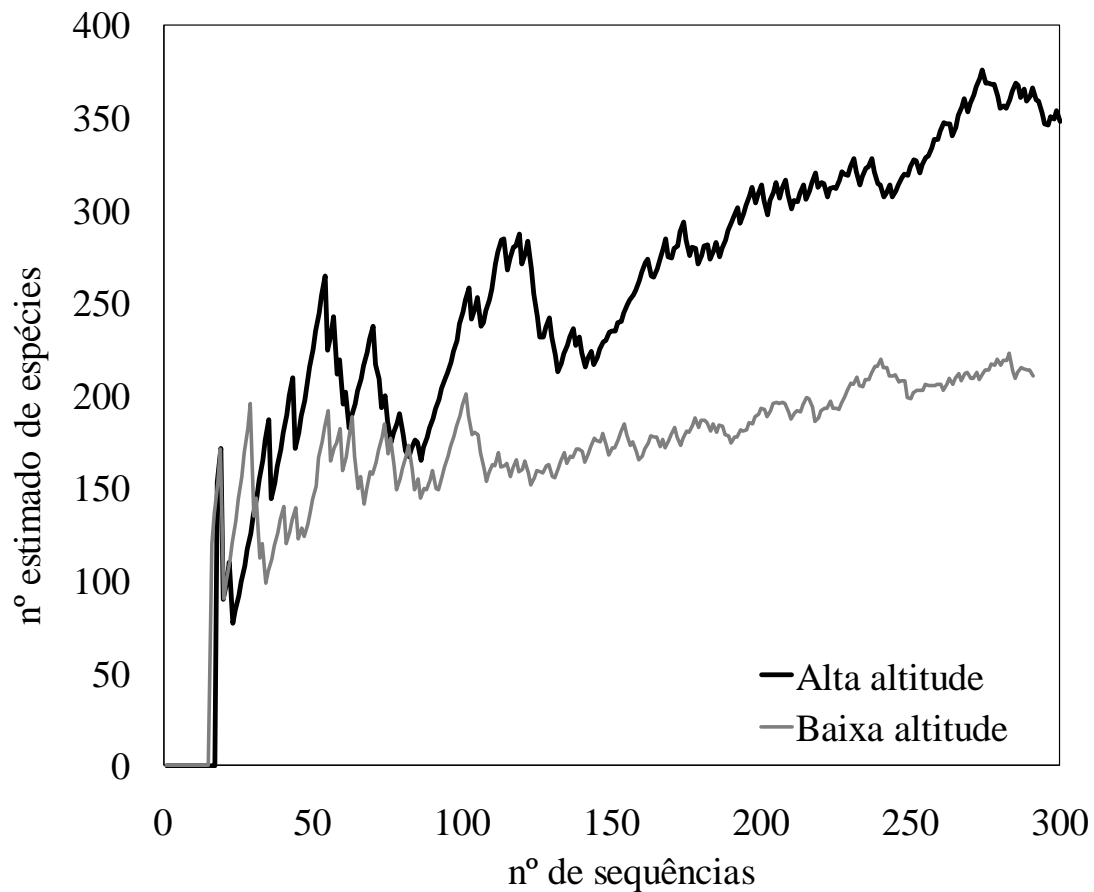
Sequências gênicas de 16S rRNA que apresentaram 95% ou mais de similaridade foram agrupadas dentro da mesma OTU (*Operational Taxonomic Unit*). O gráfico mostra a variação do índice Shannon de acordo com o número de sequências analisadas. O grupo de alta altitude possui uma diversidade maior do que o de baixa altitude.

FIGURA 14: Curva de rarefação para os grupos de alta e baixa altitude gerada pelo programa DOTUR



Sequências gênicas de 16S rRNA que apresentaram 95% ou mais de similaridade foram agrupadas dentro da mesma OTU (*Operational Taxonomic Unit*). A saturação da curva indica o limite de OTUs para a amostra. Quando a curva atinge o platô, novas OTUs não serão recuperadas mesmo com o aumento do número de sequências. O grupo de alta altitude está mais longe de atingir o platô do que o grupo de baixa altitude, indicando que o primeiro grupo possui mais OTUs.

FIGURA 15: Número estimado de OTUs para os grupos de alta e baixa altitude gerados pelo programa DOTUR.



Seqüências gênicas de 16S rRNA que apresentaram 95% ou mais de similaridade foram agrupadas dentro da mesma OTU (*Operational Taxonomic Unit*). A estimativa do número de OTUs por seqüência amostrada foi maior para o grupo de alta diversidade.

5.1.2 Coleta das amostras de solo

As amostras de solo utilizadas neste trabalho foram coletadas na porção de Mata Atlântica paranaense compreendida ao longo da rodovia PR 410, conhecida como “Estrada da Graciosa” (Figura 16A e 16B). Essa rodovia atravessa o trecho mais preservado de Mata Atlântica do país possuindo uma extensão de 28,5 km (http://www.pr.gov.br/der/malha_rod_graciosa.shtml). O portal de entrada da Estrada da Graciosa foi adotado como marco zero (km 0) e a partir dele foram coletadas as 3 amostras de solo, determinadas pela análise de comunidade, denominadas MAF1, MAF2 e MAF3 (*Metagenome of Atlantic Forest*) (Figura 17). A localização dos pontos de coleta foi determinada utilizando um aparelho GPS (*Global Positioning System*). Os dados da coleta são apresentados na Tabela 9.

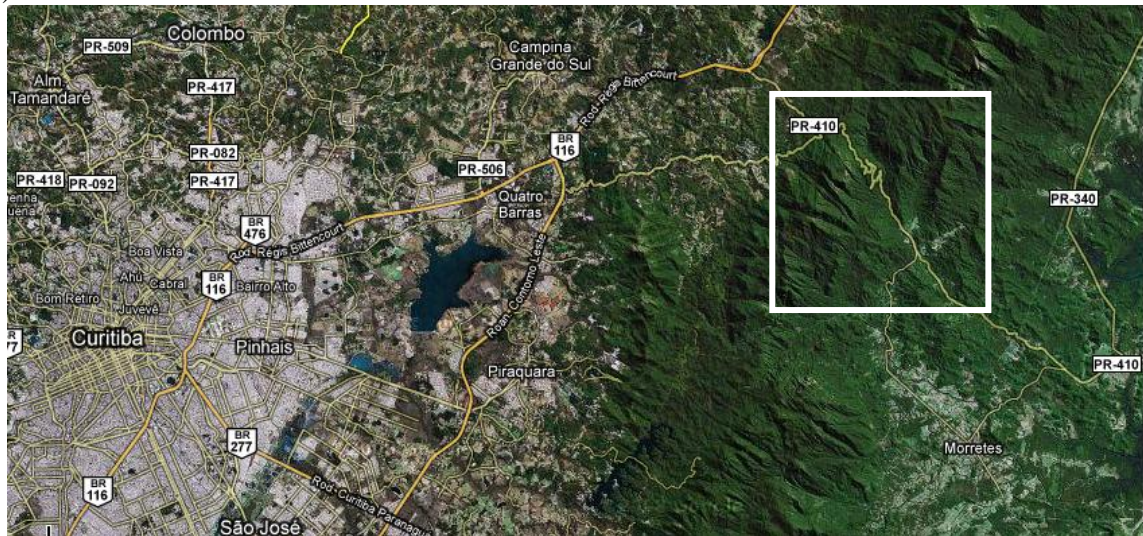
TABELA 9: Características dos pontos de coleta

Amostras	^aDistância	^bAltitude	Índice Shannon
MAF1	5,0	900	4,32
MAF2	10,0	604	3,96
MAF3	17,2	161	3,57

NOTA: ^aDistância em km a partir do portal de entrada da Estrada da Graciosa, ^bAltitude em metros em relação ao nível do mar

FIGURA 16: Pontos de coleta

(A)



Localização geográfica Serra da Graciosa (PR410).

(B)

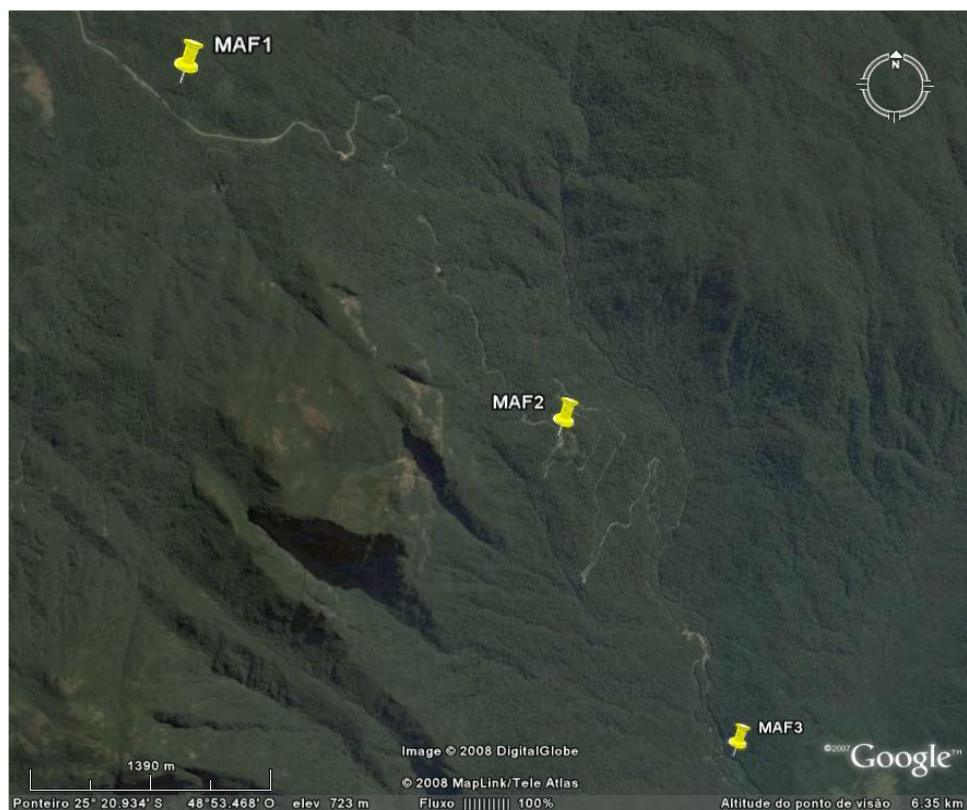
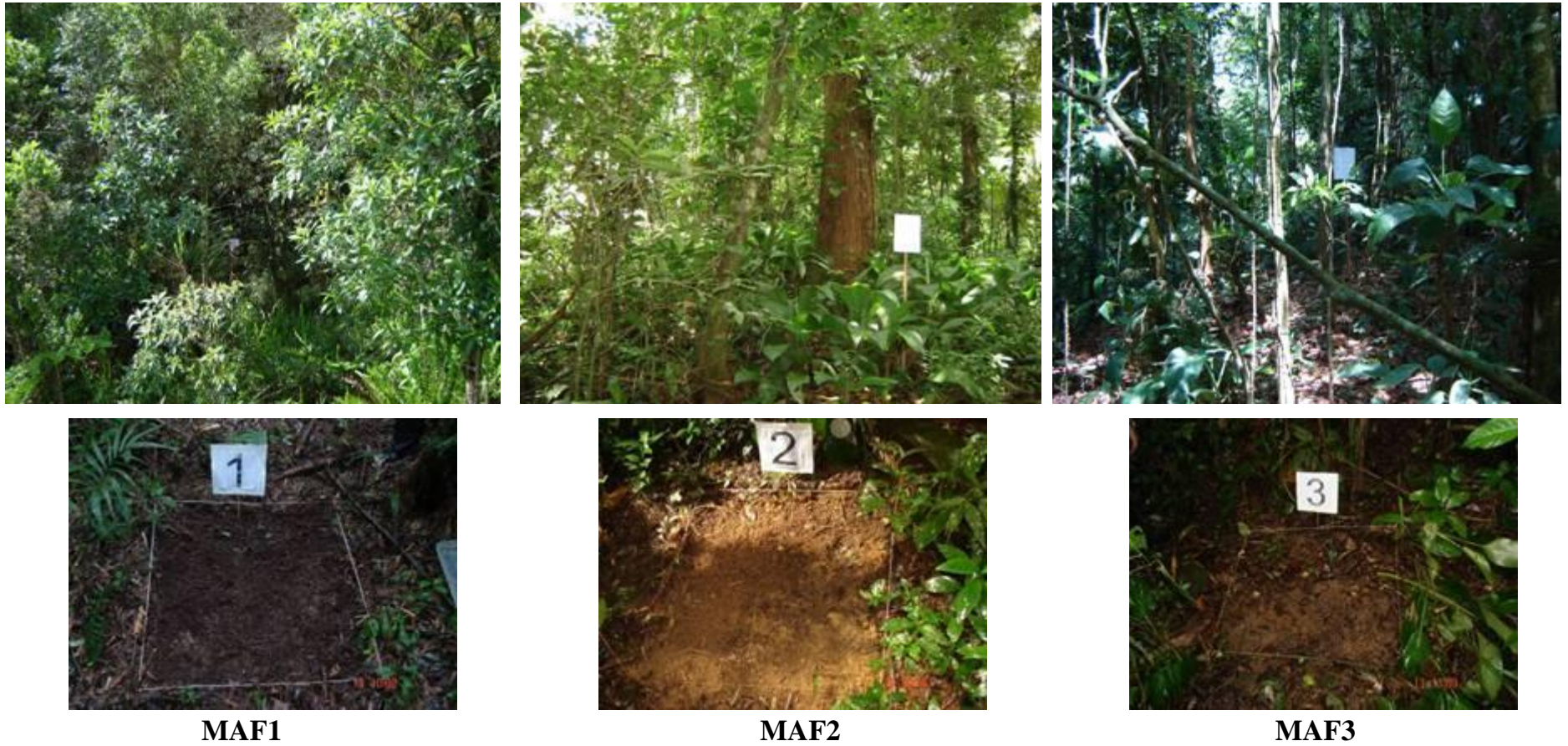


Imagem de satélite da Serra da Graciosa mostrando os pontos de coleta das amostras de solo MAF1, 2 e 3.

FIGURA 17: Sítios de coleta das amostras de solo

Em cada ponto foram coletadas 5 subamostras de solo. As subamostras foram unidas originando uma amostra composta, a partir da qual foi retirada a alíquota de 10 g utilizada para a purificação de DNA. Os painéis superiores mostram uma visão geral do local de coleta. Os painéis inferiores mostram o sítio da onde foram retirados os solos.

5.1.3 Purificação de DNA total das amostras de solo

As purificações de DNA das amostras de solo foram feitas utilizando-se o kit *PowerMax Soil DNA Isolation Kit* (MO BIO Laboratories) pelo método de purificação direto, ou seja, não houve pré-tratamentos na amostra de solo. O perfil eletroforético do DNA extraído (Figura 18) é semelhante para as três amostras de solo e indica que durante o processo de extração houve fragmentação do DNA caracterizado pelo arraste, sendo que os menores fragmentos apresentam 250 pb. Mas a presença de uma banda intensa acima da banda de 10 kpb do marcador indica que o DNA extraído apresenta alto peso molecular, o que é ideal para construção da biblioteca metagenômica.

O DNA extraído das amostras de solo foi quantificado utilizando o kit de quantificação de DNA *Qubit Fluorometer* (Invitrogen) seguindo as recomendações do fabricante. As seguintes concentrações foram obtidas:

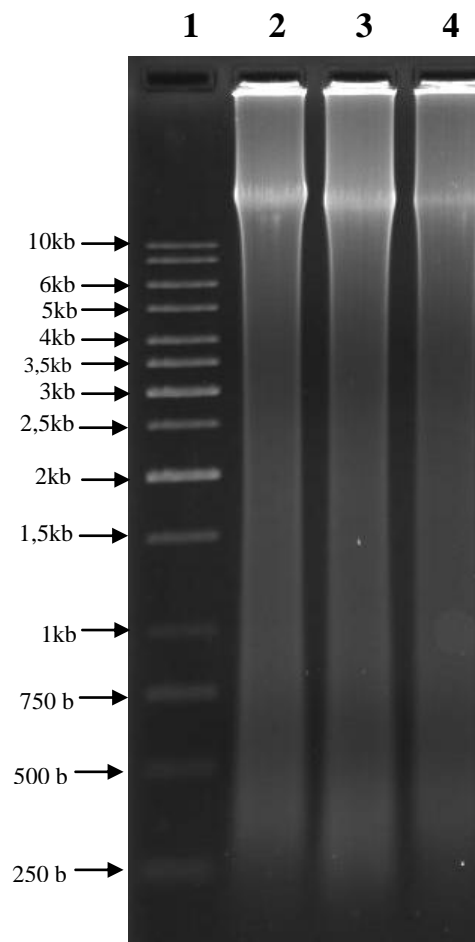
MAF1 – 539 ng/ μ l

MAF2 – 608 ng/ μ l

MAF3 – 621 ng/ μ l

O processamento das amostras ocorreu no mesmo dia da coleta para evitar alterações na composição do ecossistema, como proliferação excessiva de alguns grupos de microrganismos. Isso poderia alterar a representatividade do DNA de grupos bacterianos particulares que possuem uma divisão celular mais rápida.

FIGURA 18: Purificação de DNA das amostras de solo da Floresta Atlântica paranaense



O DNA purificado foi submetido eletroforese em gel de agarose 1%, 50 V por 2 horas em tampão TBE 1X.

- 1) Marcador de tamanho molecular *GeneRuler 1 kb DNA Ladder* (Fermentas)
- 2) DNA total MAF1
- 3) DNA total MAF2
- 4) DNA total MAF3

5.1.4 Construção das bibliotecas metagenômicas

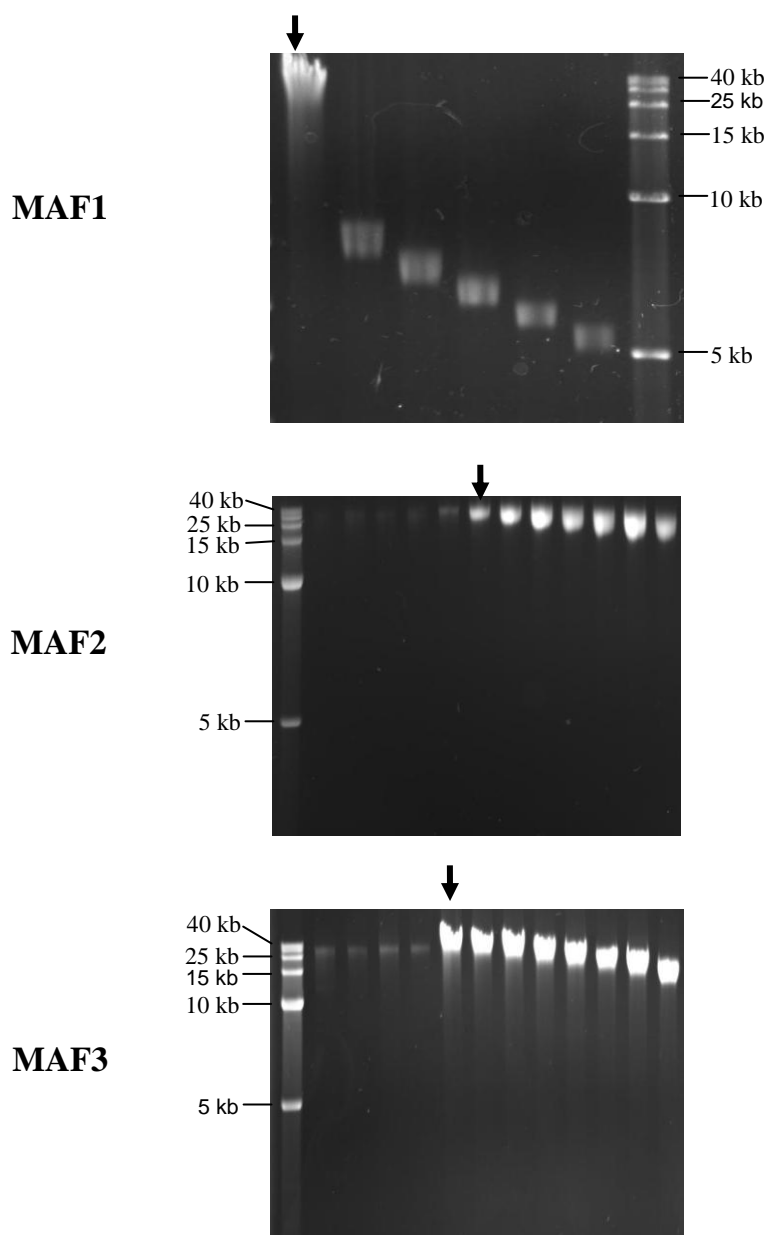
Foram construídas 3 bibliotecas metagenômicas que receberam a mesma denominação dada ao ponto de coleta: MAF1, MAF2 e MAF3. Para a construção das bibliotecas foi necessário realizar uma seleção de tamanho do DNA total com o objetivo de se obter apenas os fragmentos maiores que 40 kb. Essa seleção foi realizada em gel de agarose de baixo ponto de fusão retirando-se fatias escalonadas do gel. O DNA presente nessas fatias foi novamente purificado (Figura 19) e aquele que apresentou a banda de maior peso molecular foi ligado no vetor pCC2FOS e transformado na célula hospedeira. Esse procedimento foi repetido para o DNA obtido das 3 amostras de solo.

As bibliotecas MAF1, 2 e 3 contam com 34.560, 29.280 e 36.288 clones, respectivamente, num total de 100.128 clones armazenados individualmente em 1.043 placas de 96 poços. Todos os clones foram armazenados a -20°C em solução de glicerol 50%. O tamanho médio dos insertos clonados foi de 35 kb, avaliados por eletroforese em gel de agarose 0,7% após restrição com endonuclease *Bam*HI. O tamanho do inserto é uma característica importante em bibliotecas metagenômicas. Insertos grandes entre 30 e 40 kb podem conter vários genes, o que leva a uma diminuição no número de clones a serem analisados durante a prospecção. Além disso, bibliotecas com insertos maiores tornam possível a identificação de vias metabólicas completas (DANIEL, 2004). Além dos 100.128 clones armazenados individualmente em placas de 96 poços, também foram armazenados misturas de sistemas de infecção não plaqueados (*pools*), o que eleva o número de clones potenciais para cerca de 400 mil.

O vetor pCC2FOS usado para construir as bibliotecas contém duas origens de replicação: *Ori2* e *OriV*. A replicação a partir de *Ori2* mantém o fosmídeo a uma cópia por célula, enquanto a replicação a partir de *OriV* pode ser induzida e aumenta o número de cópias para aproximadamente 200 por célula. A replicação a partir de *OriV* esta sob o controle da proteína TrfA. A transcrição do gene que codifica a proteína TrfA esta sob o controle do promotor p_{BAD} que é ativado pela proteína reguladora do

metabolismo de arabinose AraC. O teste de indução do número de cópias do fosmídeo para os clones da biblioteca foi realizado com 0,001% de L-arabinose em meio líquido TB. Foi possível observar um aumento na quantidade de DNA nos clones que foram multiplicados em meio suplementado com L-arabinose. Sendo assim, os meios sólidos seletivos para prospecção de atividades enzimáticas foram suplementados com L-arabinose com o objetivo de aumentar, junto com a quantidade de DNA, a quantidade de proteína expressa a partir dele, o que deve tornar mais confiável a identificação de fenótipos. No entanto, a concentração de L-arabinose nas placas de meio sólido foi aumentada para 0,01% (m/v).

FIGURA 19: Seleção de fragmentos de DNA para construção das bibliotecas metagenômicas.



Eletroforese em gel de agarose 0,7% em tampão TBE 1X mostrando o DNA purificado a partir das fatias escalonadas retiradas do gel de agarose de baixo ponto de fusão. A faixa do gel correspondente ao marcador de peso molecular 5 kp (Gibco) foi excisada e corada com brometo de etídio. A posição da banda de 40 kb do marcador de peso molecular foi identificada por exposição da faixa corada à luz UV e marcada. Essa faixa de gel correspondente ao marcador foi posicionada ao lado do gel contendo o DNA metagenômico e atuou como guia para a seleção dos fragmentos de interesse entre 40 e 50 kb. O DNA utilizado para a construção de cada biblioteca esta indicado pela seta.

5. RESULTADOS

5.2 Prospecção de enzimas

5.2.1 Identificação de clones com atividade esterase/lipase

5.2.2 Identificação de clones com atividade amilase e protease

5.2.3 Identificação de clones com fenótipos diferenciados

5.2 Prospecção de enzimas

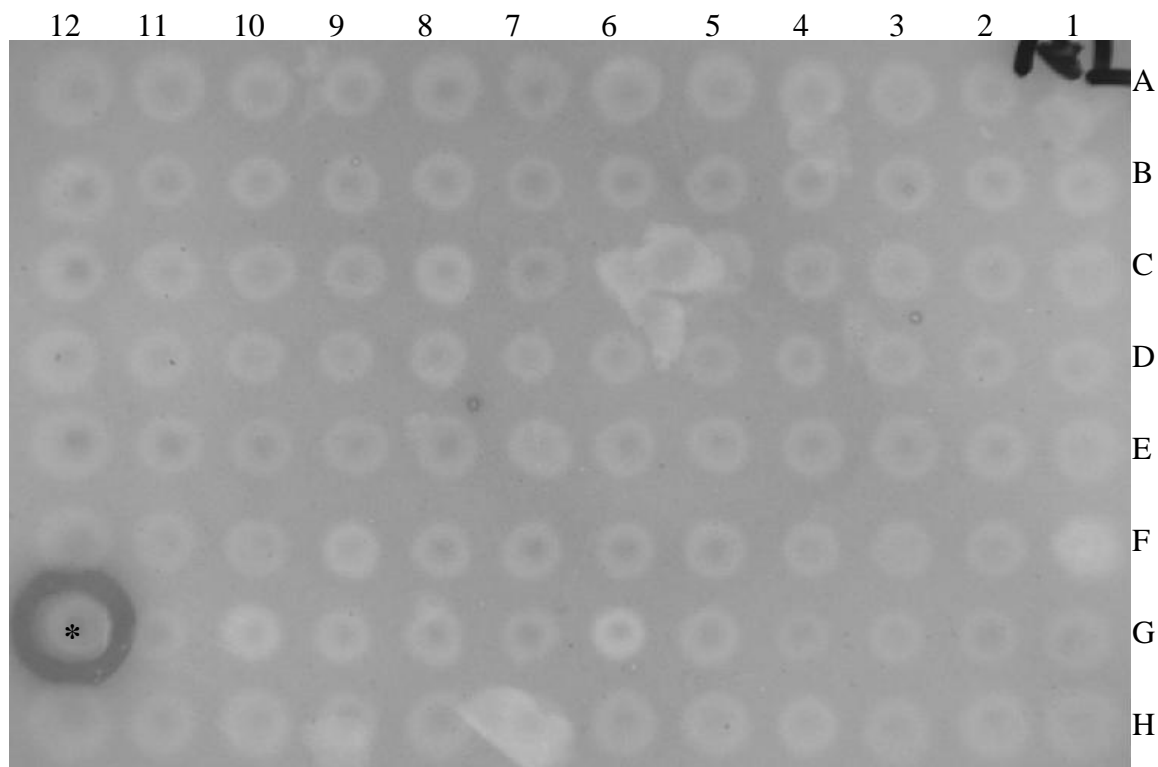
5.2.1 Identificação de clones com atividade esterase/lipase

A identificação de clones com atividade esterase/lipase foi realizada, primeiramente, em meio sólido contendo 1% de tributirina. A atividade enzimática foi evidenciada pela formação de um halo claro em volta da colônia (Figura 20). Todos os 100.128 clones foram analisados por este método e 315 clones positivos foram identificados, dos quais 109 pertencem à biblioteca MAF1, 92 à biblioteca MAF2 e 114 à biblioteca MAF3. Esses clones foram coletados, transferidos para uma nova placa de 96 poços e identificados com uma nomenclatura própria, que considera a biblioteca à qual pertencem seguidos do prefixo LP e o número correspondente à ordem em que foram identificados. Desse modo, os clones positivos para esterase/lipase da biblioteca MAF1, 2 e 3 foram nomeados MAF1LP001 a 109, MAF2LP001 a 092 e MAF3LP001 a 114.

Os 315 clones positivos em tributirina foram testados em meio sólido contendo 1% de tricaprilina e dez clones com atividade foram identificados, todos pertencentes à biblioteca MAF1. Os clones positivos em tricaprilina foram MAF1LP001, LP018, LP033, LP043, LP046, LP055, LP078, LP081, LP084 e LP090 (Figura 21). Por fim, os clones positivos em tributirina foram testados em meio sólido contendo 1% de trioleína ($18\Delta^9$). A atividade foi evidenciada pela formação de halo precipitado de cálcio-ácido graxo em torno da colônia. Três clones, denominados MAF1LP001, MAF1LP018 e MAF3LP050, apresentaram esse fenótipo (Figura 22).

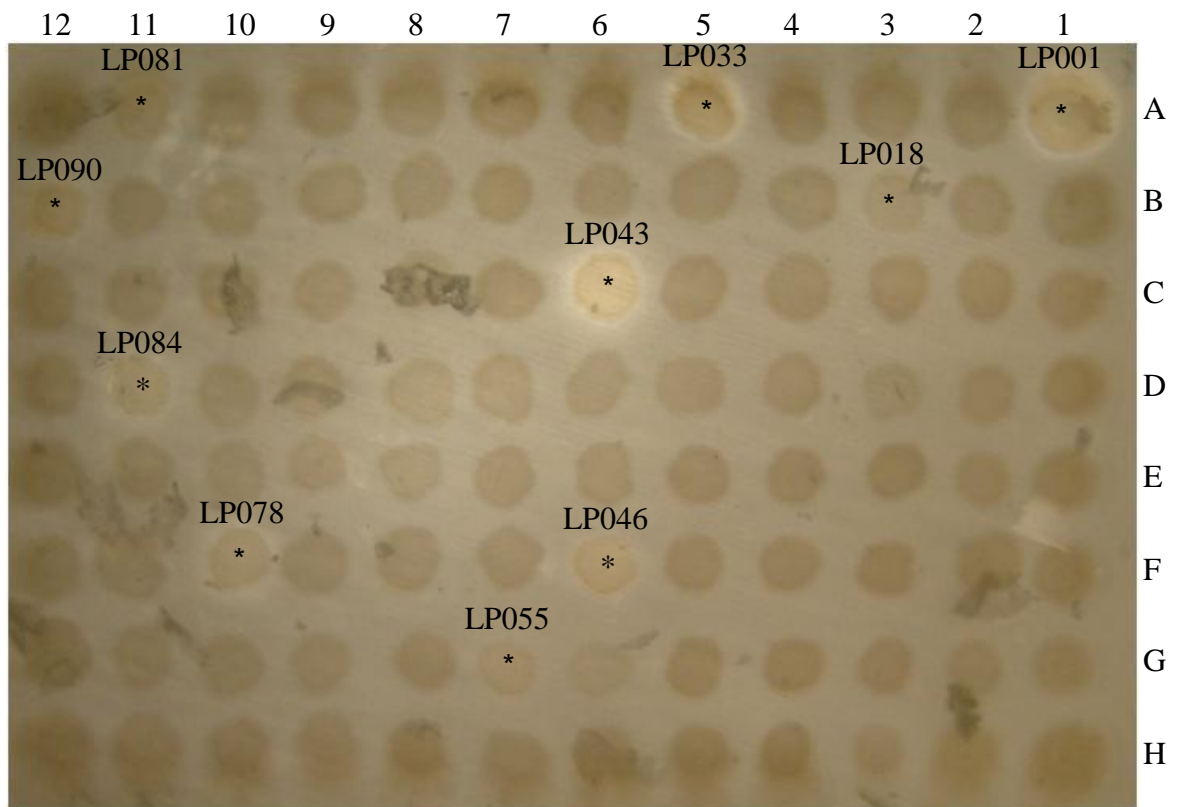
Os dez clones da biblioteca MAF1 que apresentaram atividade em meio contendo tricaprilina tiveram as extremidades de seus insertos de DNA sequenciadas. As sequências obtidas foram comparadas com as sequências depositadas no banco de dados internacional GenBank para identificar possíveis afiliações taxonômicas e metabólicas. O resultado dessa comparação mostra apenas o primeiro resultado (*best BLAST hit*) com o qual foi encontrado algum grau de semelhança (Tabela 10). Por essa metodologia foram encontradas 9 espécies diferentes. As duas que mais apareceram

foram *Acidobacteria bacterium Ellin345* e *Solibacter usitatus Ellin6076*, bactérias que colonizam o solo e que tiveram seus genomas integralmente sequenciados (WARD et al, 2009). Entretanto, essa análise de diversidade é baseada em sequências de proteínas, o que apresenta alguns inconvenientes, como o baixo número de sequências depositadas, baixa conservação de sequência, com exceção do sítio ativo, e, geralmente, sequências de organismos já descritos. Essas características justificam o fato da sequência obtida com iniciadores diferentes, a partir de um mesmo clone, resultar em espécies diferentes. A maioria das sequências submetidas à comparação apresentou similaridade com genes que codificam proteínas do metabolismo geral como hidrolases, oxirredutases, transferases e sintases. Três sequências foram identificadas como proteínas hipotéticas e para outras três sequências não foram encontradas similaridade significativa com as sequências do banco de dados, seja pela falta de uma contraparte ou por se tratar de uma região intergênica.

FIGURA 20: Prospecção por clones com atividade esterase/lipase em tributirina

Prospecção realizada na placa MAF1-114. Foi utilizado o meio seletivo para esterase descrito na Tabela 6. As colônias foram inoculadas no meio sólido contendo 1% de tributirina a partir do estoque em glicerol e crescidas por 3 dias em estufa a 37°C. O halo claro em volta da colônia na posição G12(*) indica a degradação do substrato.

FIGURA 21: Identificação de clones com atividade esterase/lipase em tricaprilina.



Os clones positivos em tributirina foram submetidos a uma nova análise, utilizando-se tricaprilina como substrato, para detecção de atividade lipase. As colônias foram inoculadas em meio sólido contendo 1% de tricaprilina (Tabela 6) a partir do estoque em glicerol e multiplicadas por 7 dias em estufa a 37°C. O halo claro em volta das colônias marcadas (*) indica a degradação do substrato. Dez clones positivos foram identificados

FIGURA 22: Identificação de clones com atividade lipase em trioleína

MAF1LP001



MAF1LP018



MAF3LP050

Os clones positivos em tributirina foram submetidos a uma nova análise, utilizando-se trioleína como substrato, para detecção de atividade lipase. As colônias foram inoculadas no meio sólido contendo 1% de trioleína (Tabela 6) a partir do estoque em glicerol e multiplicadas por 7 dias em estufa a 37°C. O halo precipitado de cálcio-ácido graxo em volta da colônia indica a atividade. Três clones positivos foram identificados.

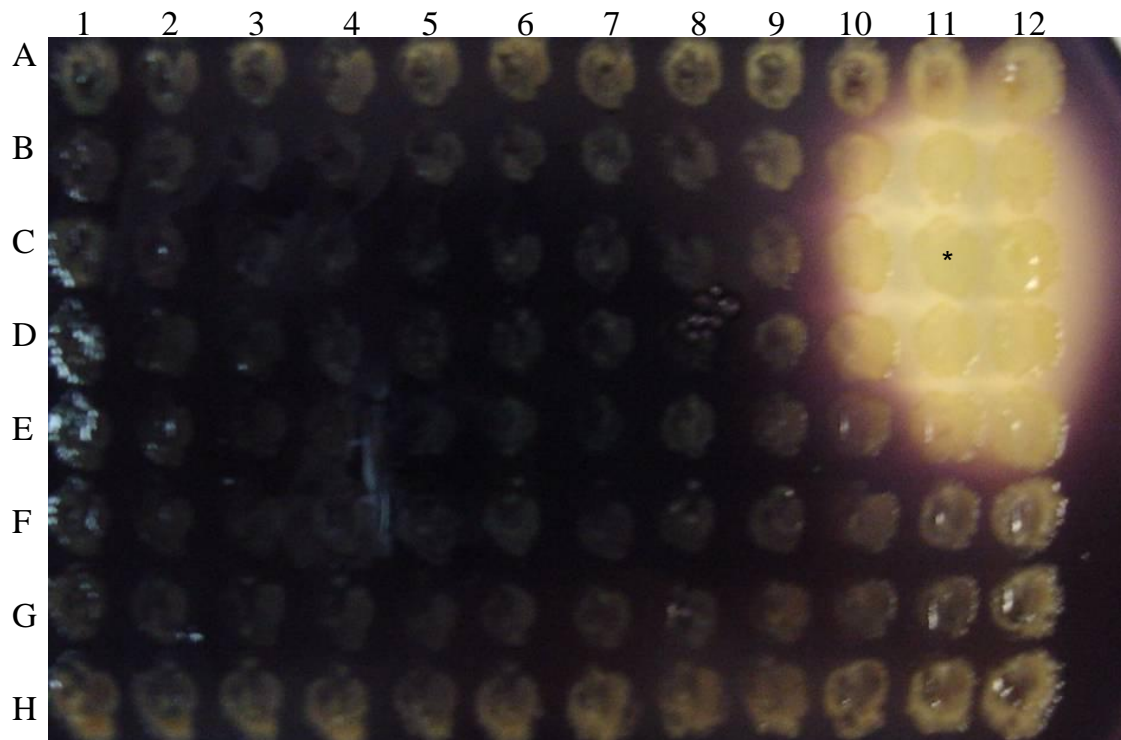
TABELA 10: Análise das sequências das extremidades dos insertos de DNA dos clones com atividade em tricaprilina

Clone	Iniciador ^a	Afiliação Taxonômica	Afiliação Metabólica	E-Value ^b	Identidade (%) ^c	Positivos (%) ^d
LP001	F	<i>Geobacillus sp. G11MC16</i>	Prolil oligopeptidase S9	3,0E-47	54%	73%
LP001	R	<i>Desulfatibacillum alkenivorans AK-01</i>	Fosfoglicerato mutase	5,0E-17	31%	51%
LP018	F	<i>Microscilla marina ATCC 23134</i>	peptidase M61	3,0E-23	32%	51%
LP018	R	ND	ND	ND	ND	ND
LP033	F	<i>Solibacter usitatus Ellin6076</i>	Proteína hipotética Acid_4341	2,0E-08	32%	47%
LP033	R	<i>Solibacter usitatus Ellin6076</i>	Proteína hipotética Acid_6572	5,0E-42	49%	58%
LP043	F	<i>Acidobacteria bacterium Ellin345</i>	Proteína semelhante a OstA	2,0E-38	34%	53%
LP043	R	<i>Acidobacteria bacterium Ellin345</i>	Proteína da família das glicosil transferases	3,0E-03	35%	56%
LP046	F	<i>Myxococcus xanthus DK 1622</i>	Fe-S oxirredutase	6,0E-73	66%	79%
LP046	R	<i>Solibacter usitatus Ellin6076</i>	Proteína hipotética Acid_3081	5,0E-86	64%	76%
LP055	F	<i>Roseiflexus castenholzii DSM 13941</i>	Beta-cetoacil sintase	1,0E-31	62%	78%
LP055	R	<i>Brevundimonas sp. BAL3</i>	Proteína da família PAS	1,3E-02	32%	56%
LP078	F	ND	ND	ND	ND	ND
LP078	R	<i>Solibacter usitatus Ellin6076</i>	Acil-CoA desidrogenase	1,0E-111	89%	95%
LP081	F	ND	ND	ND	ND	ND
LP081	R	<i>Acidobacteria bacterium Ellin345</i>	major facilitador transportador	2,0E-52	66%	83%
LP084	F	<i>Acidobacteria bacterium Ellin345</i>	Citocromo bd ubiquinol oxidase	4,0E-86	68%	77%
LP084	R	<i>Acidobacteria bacterium Ellin345</i>	peptidase M1	5,0E-17	67%	81%
LP090	F	<i>Leptothrix cholodnii SP-6</i>	Proteína de biossíntese de polissacarideo	2,0E-37	55%	71%
LP090	R	<i>Acidobacteria bacterium Ellin345</i>	Alfa amilase	7,0E-25	41%	57%

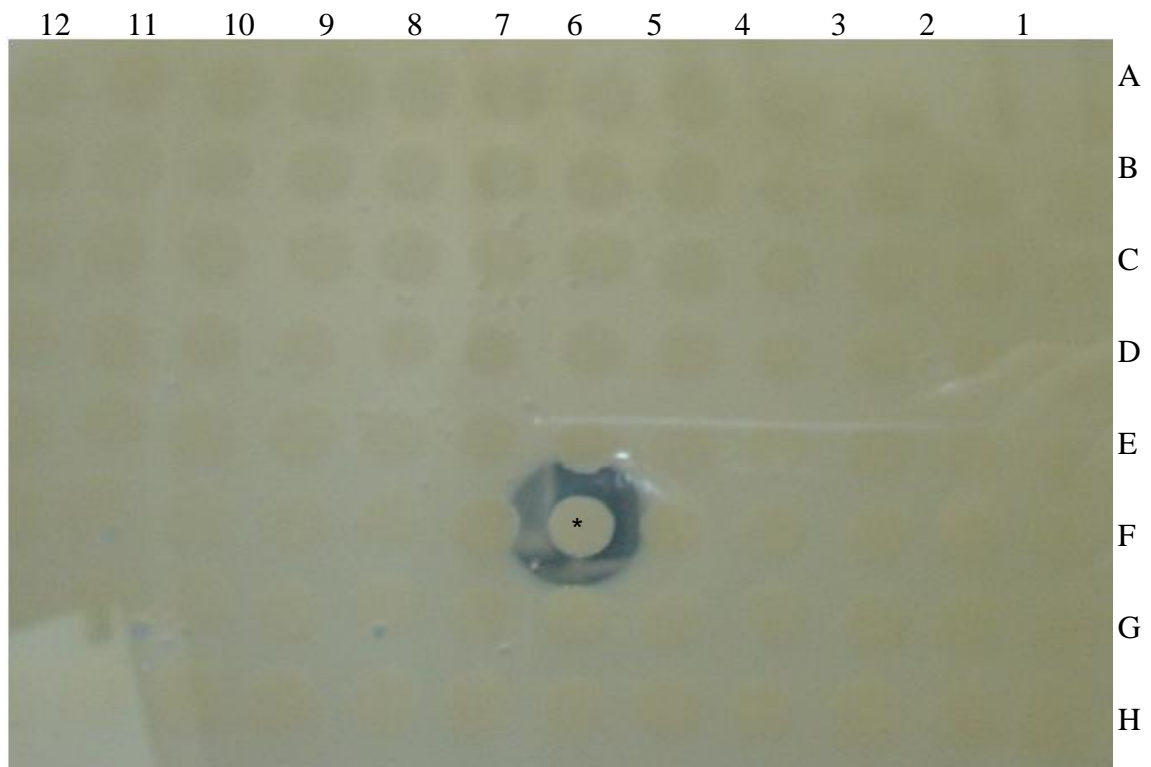
NOTA: Comparações com E-Value maior que 1 foram classificadas como ND (Não determinado). ^aF: Forward, R: Reverse. ^bValor que reflete a probabilidade da similaridade encontrada ser devido ao acaso. ^cAminoácidos iguais. ^dAminoácidos do mesmo grupo.

5.2.2 Identificação de clones com atividade amilase e protease

A prospecção por clones expressando atividade de amilase e protease foi realizada apenas para os clones da biblioteca MAF1. Para prospecção de atividade amilase foi utilizado meio seletivo contendo 1% de amido solúvel e coloração com solução de iodo (Figura 23). Para prospecção de atividade protease foi usado meio seletivo contendo 2% de leite (Figura 24). Entre os 34.560 clones que compõe a biblioteca MAF1, foram identificados 4 clones com atividade amilase e 460 clones com atividade protease. No entanto, como a prospecção por proteases foi realizada em leite e não em caseína pura é possível, que dentre os 460 clones, muitos não sejam proteases e representem outra classe de hidrolases, como as glicosil hidrolases (JONES, SUN e MARCHESI, 2007).

FIGURA 23: Prospecção por clones com atividade amilase

Prospecção realizada na placa MAF1-105. Foi utilizado o meio seletivo para amilase descrito na Tabela 6. As colônias foram inoculadas no meio sólido a partir do estoque em glicerol e crescidas por 3 dias em estufa a 37°C. Em seguida a placa foi corada com solução de iodo. O halo claro em volta da colônia na posição C11(*) indica a degradação do substrato.

FIGURA 24: Prospecção por clones com atividade Protease

Prospecção realizada na placa MAF1-100. Foi utilizado o meio seletivo para protease descrito na Tabela 6. As colônias foram inoculadas no meio sólido a partir do estoque em glicerol e crescidas por 5 dias em estufa a 37°C. O halo claro em volta da colônia na posição F06(*) indica a degradação do substrato.

5.2.3 Identificação de clones com fenótipos diferenciados

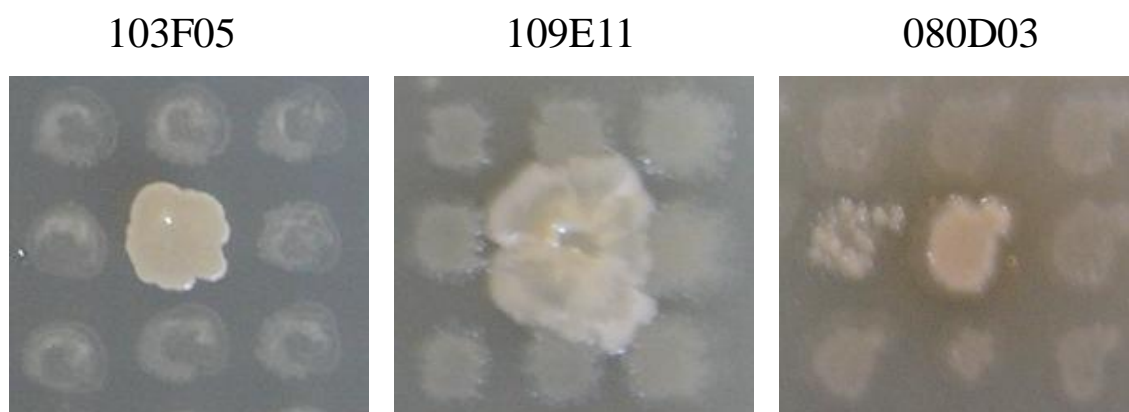
Durante a prospecção da biblioteca MAF1 por esterases, lipases, amilases e proteases, outros clones com fenótipo diferenciado foram identificados (Tabela 11). Essas alterações incluíam produção de composto marrom, alteração morfológica e produção de goma (Figura 25). A hipótese de contaminação da biblioteca por outra bactéria foi levada em consideração. No entanto, todos esses clones possuem fosmídeo com inserto de DNA, indicando que o fenótipo pode estar relacionado a alguma proteína, ou proteínas, expressas a partir do dele.

TABELA 11: Clones com fenótipo diferenciado identificados na biblioteca MAF1

Fenótipo	Clone da biblioteca MAF1^a
Produção de composto marrom escuro	013D03, 033C12, 080D03, 112A01
Produção de goma incolor	121B09
Colônia de cor diferente	025E01, 125F06, 148B07, 336D07
Colônia de aspecto cremoso	072E11, 072F06, 080C02, 103F05
Colônia de aspecto poroso	267C08
Colônia de aspecto translúcido	107D05, 227E07
Colônia esbranquiçada	283E07, 109E11

NOTA: ^aNúmero da placa e posição

FIGURA 25: Clones com fenótipo diferenciado identificados na biblioteca MAF1.



Clone 103F05 multiplicado em meio seletivo amido 1% a 37°C durante 7dias apresenta aspecto cremoso. Clone 109E11 multiplicado em meio seletivo amido 1% a 37°C durante 7 dias apresenta aspecto esbranquiçado. Clone 080D03 multiplicado em meio seletivo tributirina 1% a 37°C durante 5 dias apresenta produção de composto marrom.

O resultado completo da prospecção realizada nas bibliotecas metagenômicas esta apresentado na Tabela 12. A partir desses resultados, 5 clones foram selecionados para construção de sub-bibliotecas em vetor de clonagem pUC18 para sequenciamento integral do DNA inserto:

MAF1LP001: Atividade contra tributirina, tricaprilina e trioleína

MAF1LP018: Atividade contra tributirina, tricaprilina e trioleína

MAF1LP090: Atividade contra tributirina e tricaprilina; presença de α -amilase

MAF3LP050: Atividade contra tributirina e trioleína

TABELA 12: Resultado final da prospecção das bibliotecas metagenômicas.

	BIBLIOTECAS			
	MAF1	MAF2	MAF3	Total
Nº de clones	34.560	29.280	36.288	100.128
Tamanho médio de inserto (kpb)	35 kb	35 kb	35 kb	-
Nº total de bases (kpb)	1.209.600	1.024.800	1.270.080	3.504.480
Nº estimado de genomas ^a	302	256	317	875
Nº de clones positivos em tributirina	109	92	114	315 (0,31 ^b)
Nº de clones positivos em tricaprilina	10	0	0	10 (0,01 ^b /3,175 ^c)
Nº de clones positivos em trioleína	2	0	1	3 (0,003 ^b /0,952 ^c)
Nº de clones positivos em amido	4	nd	nd	4 (0,011 ^d)
Nº de clones positivos em leite	460	nd	nd	460 (1,332 ^d)
Nº de clones com fenótipo diferenciado	18	nd	nd	18 (0,052 ^d)

NOTA: ^aRAES et al, 2007; ^b% Considerando o número total de clones; ^c% Considerando apenas os clones positivos em tributirina; ^d% Considerando apenas os clones da biblioteca MAF1.

5. RESULTADOS

5.3 Sequenciamento do inserto de DNA dos clones de interesse e análises de sequências

5.3.1 Construção de sub-bibliotecas

5.3.2 Sequenciamento de DNA montagem de contigs.

5.3.3 Caracterização genética do fosmídeo MAF1LP001

5.3.4 Caracterização genética do fosmídeo MAF1LP018

5.3.5 Caracterização genética do fosmídeo MAF1LP090

5.3.6 Identificação dos genes codificadores de lipases através de mutagênese aleatória por inserção de transposon

5.3.7 Análises de sequência e filogenética das lipases LP001ORF27, LP018ORF16 e LP090ORF24

5.3.8 Análise de sequência α -amilase

5.3 Sequenciamento do inserto de DNA dos clones de interesse e análise de sequências

5.3.1 Construção de sub-bibliotecas

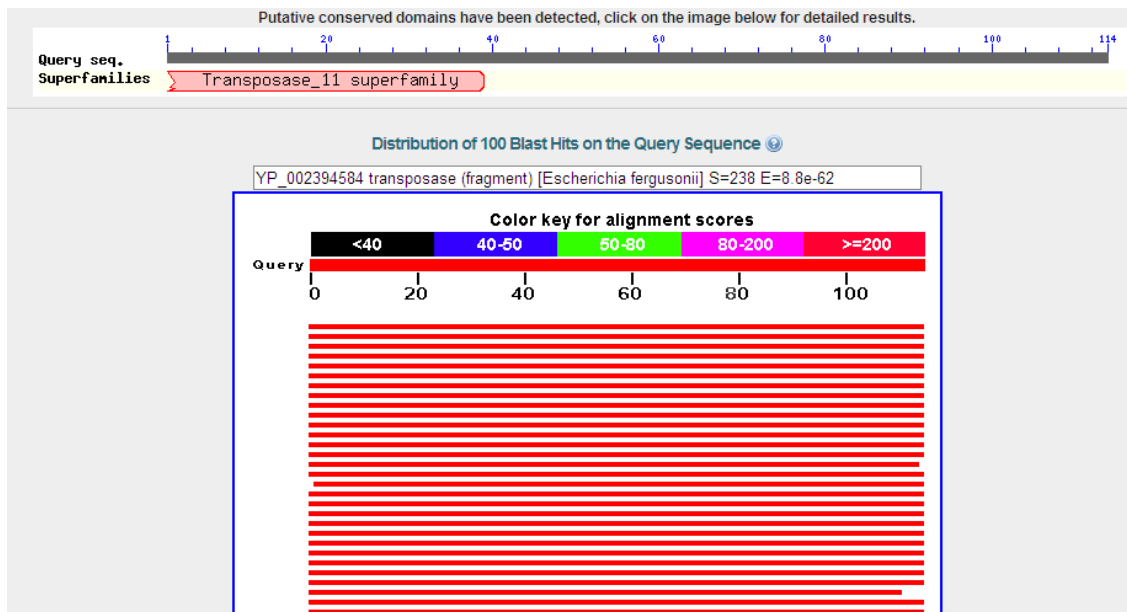
Através da prospecção enzimática foram selecionadas 5 clones de interesse para construção de sub-bibliotecas e sequenciamento integral do inserto de DNA clonado no vetor pCC2FOS. Os clones selecionados foram MAF1LP001, MAF1LP018, MAF1LP090 e MAF3LP050 (atividade esterase/lipase). O DNA fosmidial foi purificado, nebulizado, reparado, e aplicado em gel de agarose de baixo ponto de fusão, a partir do qual foi feita a seleção de fragmentos de DNA com o tamanho desejado aproximado de 3 kb. Os fragmentos de DNA purificados foram ligados no vetor pUC18 e transformados na estirpe TOP10 de *E. coli*, originando as sub-bibliotecas. Para identificação dos subclones, foi adicionado o prefixo “SC” antes do nome do clone original. Foram construídas bibliotecas com tamanho de inserto de aproximadamente 3 kb, sendo que para cada fosmídeo original foram obtidos 384 subclones (4 placas de 96 poços).

5.3.2 Sequenciamento de DNA e montagem de contigs.

Todos os clones que compunham as sub-bibliotecas SCMAF1LP001, SCMAF1LP018 e SCMAF1LP090 foram sequenciados com os iniciadores universal e reverso, cujas sequências complementares estão presentes no vetor pUC18. No entanto, após o sequenciamento de DNA da primeira placa de subclones da biblioteca SCMAF3LP050 foi identificado um novo problema: todos os clones possuíam o mesmo inserto de DNA. Todas as sequências obtidas alinharam entre si formando uma única sequência contígua de 1.000 pares de bases que apresentou similaridade com uma enzima transposase (Figura 26). O resultado de sequenciamento confirmou a hipótese levantada anteriormente sobre SCMAF3LP050 de que todos os clones possuíam o mesmo inserto de DNA. Sendo assim, essa sub-biblioteca foi descartada.

Todas as sequências de DNA obtidas foram alinhadas e editadas utilizando-se os programas Phred-Phrap-Consed. Desse modo, foi possível reconstruir parcialmente as sequências do DNA inserto dos clones MAF1LP001, MAF1LP018 e MAF1LP090. As sequências do DNA inserto dos clones MAF1LP001 (Figura 27A) e 018 (figura 28A) não estavam completas devido à existência de dois buracos (*gaps*) em sequências contíguas (contigs), ou seja, havia três principais contigs para cada clone. Baseado nas leituras obtidas com cada subclone e no iniciador usado para o sequenciamento, foi gerado um arquivo denominado *Scaffold* com um indicativo de como os contigs se organizam para remontar a sequência do fósforo original e também de possíveis clones que juntam dois contigs maiores. Baseado na análise do *Scaffold*, foi possível identificar clones cuja sequência de DNA do plasmídeo cobria os buracos em ambas as sequências. Os plasmídeos das respectivas sub-bibliotecas foram integralmente sequenciados, através da técnica de inserção de transposon, e as sequências obtidas nessa etapa foram alinhadas às anteriores, sendo suficientes para obter a sequência integral do inserto de DNA dos clones MAF1LP001 (Figura 27B) e MAF1LP018 (Figura 28B). Para o clone MAF1LP090 (Figura 29) o número de sequências produzidas não foi suficiente para o fechamento integral da sequência do DNA inserto nem para produção de um *Scaffold* confiável. Restaram ainda muitos buracos entre os contigs tornando inviável a utilização da técnica de transposon. Nesse caso, as análises foram realizadas utilizando-se as sequências parciais. O alinhamento das sequências do clone MAF1LP090 gerou 18 contigs, dos quais 8 foram utilizados para anotação (contigs 18, 17, 11, 9, 6, 5, 16 e 13). Os contigs que não foram utilizados para a anotação eram formados por poucas sequências (geralmente 2 leituras da mesma sequências), sequências de baixa qualidade ou ainda sequências de DNA da bactéria hospedeira *E. coli*.

FIGURA 26: Análise da sequência contígua obtida a partir da sub-biblioteca SCMAF3LP050



GENE ID: 7133879 transposase (fragment) [*Escherichia fergusonii*]
 Length=197 Expect = 9e-62,
 Identities = 114/114 (100%), Positives = 114/114 (100%), Gaps = 0/114 (0%)

```

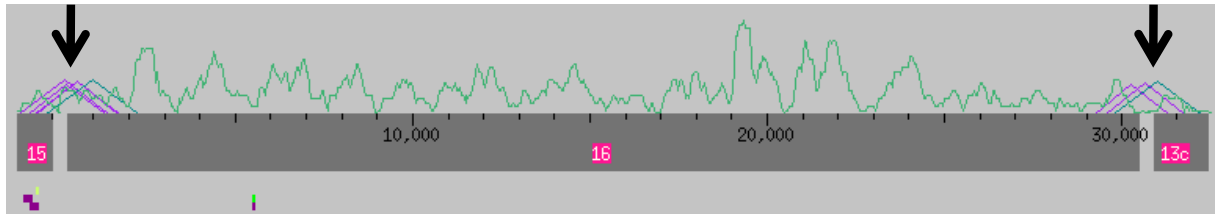
Query 1  MQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHF 60
          MQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHF
Sbjct 84  MQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHF 143

Query 61  QANTVRNRNVLSTVRLGMEVLRHSGYTTITREDLLVAATLLAQNLFTHGYALGKL 114
          QANTVRNRNVLSTVRLGMEVLRHSGYTTITREDLLVAATLLAQNLFTHGYALGKL
Sbjct 144  QANTVRNRNVLSTVRLGMEVLRHSGYTTITREDLLVAATLLAQNLFTHGYALGKL 197
    
```

As sequências obtidas a partir dos clones da sub-biblioteca SCMAF3LP050 foram alinhadas pelo pacote de programas Phred-Phrap-Consed. Possíveis sequências gênicas foram traduzidas em seus aminoácidos e comparadas com o banco de dados nr através do programa BLASTp. A busca por domínios indicou a presença de um domínio conservado da família das transposases. “Query” refere-se à sequência submetida e “Sbjct” à sequência do banco de dados. Identitites – aminoácidos iguais, Positives – aminoácidos do mesmo grupo. A comparação entre as sequências indicou 100% de identidade contra uma transposase de *E. fergusonii*.

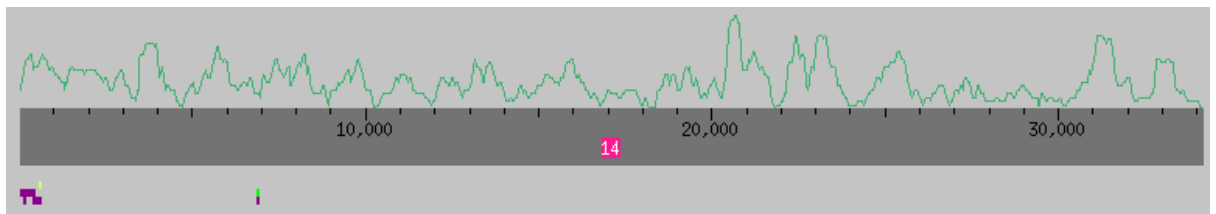
FIGURA 27: Scaffold obtido para o clone MAF1LP001

(A)



Sequência parcial com duas lacunas a serem sequenciadas. As linhas ligando dois contigs seguidos (setas pretas) indicam clones cuja sequência de DNA inserto do plasmídeo possivelmente une os contigs a que se refere. Os números destacados nas caixas rosa se referem aos contigs e os números nas réguas se referem às bases.

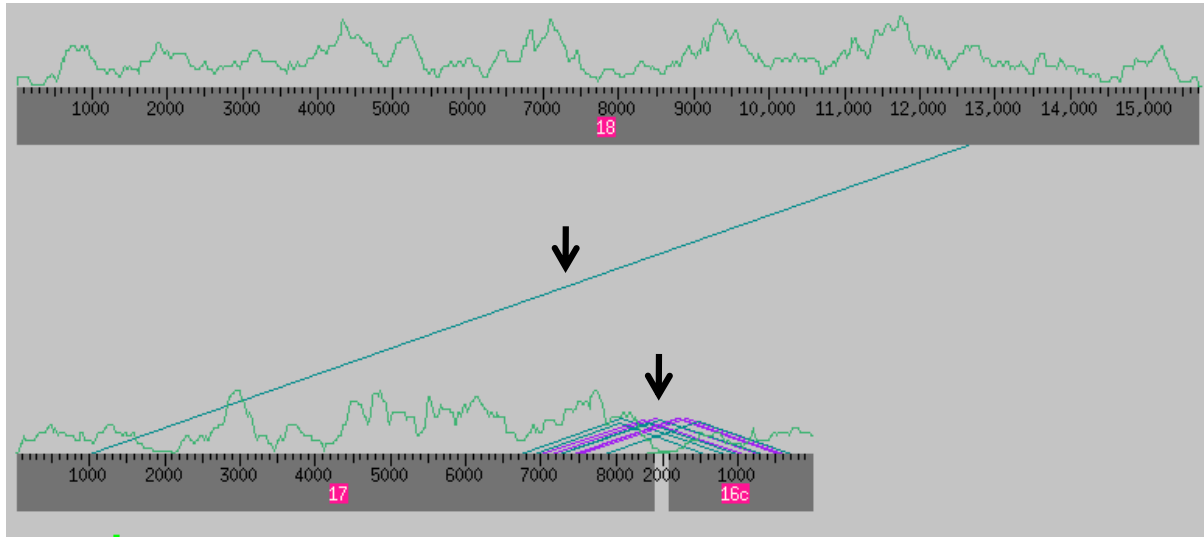
(B)



Sequência integral após o sequenciamento completo de subclones que cobriam as lacunas através da técnica de transposon. O número destacado na caixa rosa se refere ao contig e os números nas réguas se referem às bases.

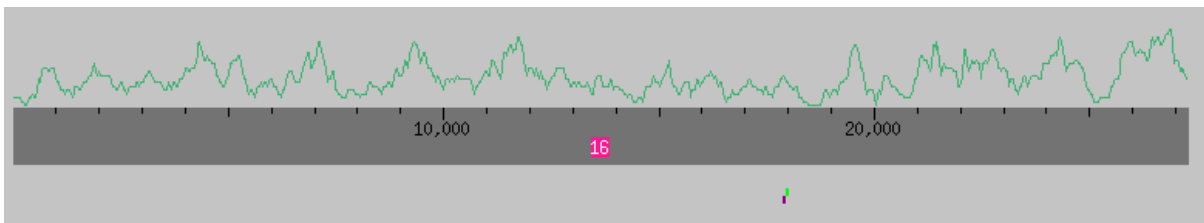
FIGURA 28: Scaffold obtido para o clone MAF1LP018

(A)



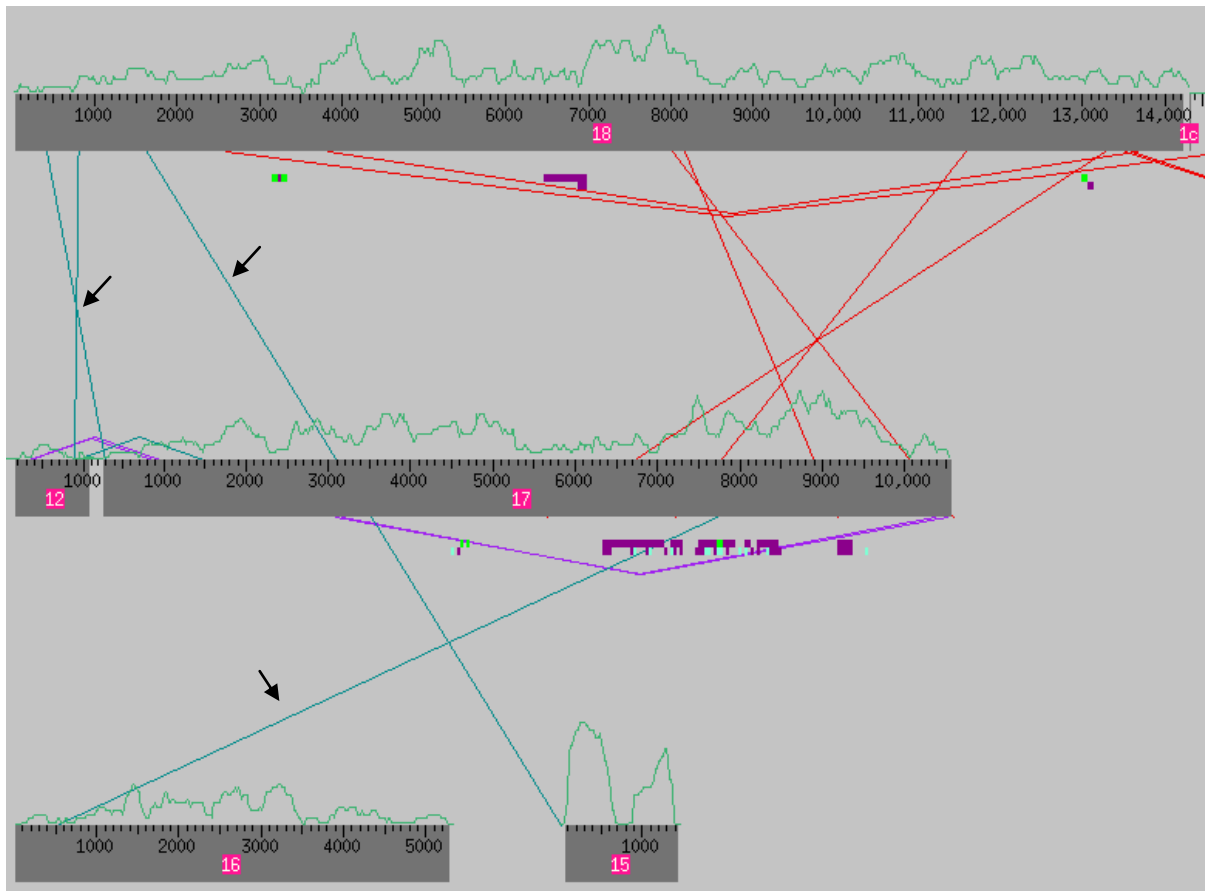
Sequência parcial com duas lacunas a serem sequenciadas. As linhas ligando dois contigs seguidos (setas pretas) indicam clones cuja sequência de DNA inserto do plasmídeo possivelmente une os contigs a que se refere. Os números destacados nas caixas rosa se referem ao contig e os números nas régua se referem às bases.

(B)



Sequência integral após o sequenciamento completo de subclones que cobriam as lacunas através da técnica de transposon. O número destacado na caixa rosa se refere ao contig e os números nas régua se referem às bases.

FIGURA 29: Scaffold obtido para o clone MAF1LP090



Os números destacados nas caixas rosa se referem ao contig. Os números nas réguas se referem às bases. As linhas ligando dois contigs seguidos (setas pretas) indicam clones cuja sequência possivelmente une os contigs a que se refere. As linhas vermelhas indicam ligações inconsistentes entre os contigs.

5.3.3 Caracterização genética dos fosmídeos MAF1LP001

O inserto de DNA presente no fosmideo MAF1LP001 possui 34.181 pares de bases com um conteúdo de G+C de 58,57%. A busca por regiões codificadoras através dos programas Artemis e Glimmer3 identificou 33 ORFs (*Open Reading Frames*). As sequências de aminoácidos dessas ORFs, traduzidas pelo próprio Artemis, foram comparadas com banco de dados GenBank através do programa BLASTp (Tabela 13). O melhor resultado de BLAST (*best BLAST hit*) para a maioria das sequências (60%) apresentou similaridade contra sequências da acidobactéria *Solibacter usitatus* Ellin6076. A ORF número 27 (LP001ORF27 a partir desse ponto), compreendida entre as bases 27.088 e 28.033 do inserto de DNA, codifica uma proteína similar a uma proteína hipotética do fungo *Neurospora crassa* estirpe OR74A. A busca por domínios conservados na proteína LP001ORF27 revelou a existência de um domínio LipA, característico de enzimas lipolíticas e a classificou como candidata responsável pelo fenótipo de degradação de tributirina. Além do primeiro resultado de BLAST apresentar similaridade com uma enzima de fungo, os outros cinco seguintes também seguiram esse padrão. A análise da sequência inserto de MAF1LP001 também revelou a existência de um operon de genes do metabolismo energético (ORFs 4 a 9, Tabela 13). Quando a sequência de aminoácidos das proteínas codificadas por esses genes foi comparada com o banco de dados, observou-se que elas possuíam similaridade com citocromos da cadeia transportadora de elétrons. Outras duas proteínas em especial chamaram a atenção nesse operon: LP001ORF7, uma subunidade do citocromo b/b6, e LP001ORF8, uma proteína com domínio Rieske e função plastoquinol-plastocianina redutase, também chamada de citocromo f. Em cianobactérias e plantas que realizam fotossíntese oxigênica essas proteínas fazem parte de um complexo maior, denominado complexo *b6f*, que possui um papel central na transferência de elétrons entre os fotossistemas II e I com bombeamento de prótons acoplado (BANIULIS et al, 2008). No entanto, a existência de acidobacterias fotossintéticas era desconhecida até recentemente quando BRYANT e colaboradores (2009) identificaram o primeiro representante. A proteína codificada pela ORF15 classificada como mandelato

racemase também pode ter uma potencial aplicação biotecnológica na separação de misturas racêmicas do ácido rac-mandélico em ácido (S)-O-acetilmandélico. A utilização da mandelato racemase de *Pseudomonas sp.* gerou um produto enantiosseletivo com 98% de pureza (STRAUSS e FABER, 1999). Por fim, a ORF17 apresentou similaridade com uma proteína carreadora 3-oxoacil. Já a busca por regiões conservadas indicou a existência de um domínio cetoacil sintase (KS). Esse domínio é característico de uma classe de enzimas denominadas policetídeo sintases responsáveis pela produção de vários compostos reconhecidamente com atividade biológica, incluindo antibióticos, imunossupressores, antiolesterolêmicos, antitumorais e antiparasíticos (CANE e WALSH, 1999). Foram encontradas, ainda, 7 proteínas hipotéticas, das quais 2 não possuem nenhum domínio conservado passível de reconhecimento pelos programas de análise. Os demais genes encontrados codificam proteínas do metabolismo geral.

TABELA 13: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP001

ORFs	GC%	Tamanho (aa)	^a Função	^a Organismo	^a Nº de acesso	^a % identidade/positivos	^a E-value	^b Domínios conservados
1	57,0	189	peptidase S9 prolyl oligopeptidase active site domain protein	<i>Geobacillus sp. G11MC16</i>	ZP_03146580	53/71	3,00E-48	sim
2	58,3	908	hypothetical protein Swit_0915 peptidase M14-like	<i>Sphingomonas wittichii RW1</i>	YP_001261420	58/74	0	sim
3	62,0	513	acetolactate synthase	<i>Thermomicrobium roseum DSM 5159</i>	YP_002523441	41/55	2,00E-96	sim
4	64,3	408	cytochrome c, class I	<i>Solibacter usitatus Ellin6076</i>	YP_820973	29/47	4,00E-43	sim
5	54,7	105	cytochrome c, class I	<i>Candidatus Koribacter versatilis Ellin345</i>	YP_591789	37/54	1,00E-07	sim
6	61,7	423	cytochrome b subunit of formate dehydrogenase-like protein	<i>Opitutus terrae PB90-1</i>	YP_001819484	34/42	3,00E-24	sim
7	57,8	365	Cytochrome b/b6 domain	<i>Chlorobium phaeobacteroides BS1</i>	YP_001960378	48/64	2,00E-80	sim
8	58,9	171	Plastoquinol-plastocyanin reductase	<i>Prosthecochloris aestuarii DSM 271</i>	YP_002016440	42/62	5,00E-26	sim
9	60,2	372	cytochrome c, class I	<i>Solibacter usitatus Ellin6076</i>	YP_829073	49/65	8,00E-105	sim
10	52,2	149	hemerythrin-like metal-binding protein	<i>Desulfatibacillum alkenivorans AK-01</i>	YP_002429217	32/57	7,00E-21	sim
11	49,5	147	hemerythrin-like metal-binding protein	<i>Desulfatibacillum alkenivorans AK-01</i>	YP_002429217	37/59	7,00E-23	sim
12	58,1	340	N-acetyl-gamma-glutamyl-phosphate reductase	<i>Solibacter usitatus Ellin6076</i>	YP_822007	66/78	6-129	sim
13	56,5	264	N-acetylglutamate kinase	<i>Solibacter usitatus Ellin6076</i>	YP_822008	55/71	9,00E-74	sim
14	57,0	323	hypothetical protein Acid_0726 MoeA-like	<i>Solibacter usitatus Ellin6076</i>	YP_822010	78/88	2,00E-148	sim
15	59,8	413	mandelate racemase/muconate lactonizing protein	<i>Solibacter usitatus Ellin6076</i>	YP_828911	63/68	2,00E-142	sim
16	62,9	239	polysaccharide deacetylase	<i>Solibacter usitatus Ellin6076</i>	YP_822029	55/70	1,00E-61	sim
17	63,0	404	3-oxoacyl-[acyl-carrier-protein] synthase II	<i>Solibacter usitatus Ellin6076</i>	YP_822027	66/79	6,00E-144	sim
18	61,1	349	hypothetical protein Acid_5547	<i>Solibacter usitatus Ellin6076</i>	YP_826769	39/60	3,00E-56	sim
19	57,5	276	diguanylate cyclase	<i>Solibacter usitatus Ellin6076</i>	YP_825871	21/43	4,00E-07	sim
20	52,2	160	diguanylate cyclase/phosphodiesterase with PAS/PAC sensor	<i>Chloroflexus sp. Y-400-fl</i>	YP_002571504	27/46	3,00E-10	sim
21	57,8	306	pentapeptide repeat-containing protein	<i>Syntrophus aciditrophicus SB</i>	YP_462974	32/45	1,00E-23	sim
22	59,8	237	two component heavy metal response transcriptional regulator	<i>Solibacter usitatus Ellin6076</i>	YP_823439	80/87	2,00E-101	sim
23	58,2	492	heavy metal sensor signal transduction histidine kinase	<i>Solibacter usitatus Ellin6076</i>	YP_823440	51/66	3,00E-136	sim
24	50,2	139	anti-sigma-factor antagonist	<i>Solibacter usitatus Ellin6076</i>	YP_825137	57/82	8,00E-42	sim
25	57,3	162	redoxin domain-containing protein	<i>Solibacter usitatus Ellin6076</i>	YP_825133	48/68	3,00E-41	sim
26	57,6	168	NADH-quinone oxidoreductase, chain I	<i>Solibacter usitatus Ellin6076</i>	YP_825132	73/85	6,00E-70	sim
27	57,7	314	hypothetical protein NCU03862	<i>Neurospora crassa OR74A</i>	XP_955855	31/42	1,00E-22	sim
28	60,8	191	hypothetical protein Acid_4087 Zinc-dependent metalloprotease like	<i>Solibacter usitatus Ellin6076</i>	YP_825336	51/63	6,00E-47	sim
29	59,3	217	serine O-acetyltransferase	<i>Solibacter usitatus Ellin6076</i>	YP_825131	73/84	2,00E-84	sim
30	61,6	170	hypothetical protein MXAN_7038	<i>Myxococcus xanthus DK 1622</i>	YP_635151	50/68	9,00E-40	não
31	60,8	225	hypothetical protein Acid_5874	<i>Solibacter usitatus Ellin6076</i>	YP_827101	57/69	7,00E-63	não
32	61,4	548	serine phosphatase	<i>Solibacter usitatus Ellin6076</i>	YP_824690	57/73	2,00E-178	sim
33	59,2	196	Phosphoglycerate mutase	<i>Desulfatibacillum alkenivorans AK-01</i>	YP_00242978	31/51	9,00E-22	sim

NOTA: ^aBaseado no melhor resultado de BLAST usando BLASTp e o banco de dados de proteínas não redundantes; ^bBaseado no banco de dados InterPRO

5.3.4 Caracterização genética do fosmídeo MAF1LP018

O inserto de DNA presente no fosmídeo MAF1LP018 possui 27.269 pares de bases com um conteúdo de G+C de 58,41%. A busca por regiões codificadoras identificou 20 ORFs e um RNA transportador para serina (tRNA^{SER} anticódon GGA) entre as ORFs 4 e 5 (Tabela 14). As sequências de aminoácidos dessas ORFs também foram comparadas com banco de dados GenBank através do programa BLASTp (Tabela 14). Novamente, a maioria das sequências (50%) apresentou similaridade contra sequências da acidobactéria *Solibacter usitatus* Ellin6076. A identificação do gene que codificava a enzima lipase não foi clara nesse primeiro momento. Nenhuma sequência apresentou similaridade contra sequências ou domínios característicos de lipases. O único resultado obtido nesse sentido foi a identificação de um motivo $\alpha\beta$ -hidrolase na proteína codificada pela ORF16, que apresentou similaridade contra uma proteína hipotética de *Nitrospira multiformis* estirpe 25196 (denominada LP018ORF16 a partir desse ponto). Aparentemente, o gene que codifica a ORF16 está em um operon com o gene que codifica para a ORF15, que codifica uma proteína similar a uma peptidase, também de *Nitrospira multiformis* estirpe 25196 (denominada LP018ORF15 a partir desse ponto). A análise funcional proposta pelo COG (*Clusters of Orthologous Groups*) classificou a proteína LP018ORF15 no grupo COG1404, que compreende subtilisinas, modificadores pós-traducionais, reciclagem de proteínas e chaperonas. A presença de chaperonas associadas a lipases é comum, portanto, é possível que esses dois genes sejam necessários para a atividade lipolítica. Essas características classificaram LP018ORF15 e ORF16 como candidatos responsáveis pelo fenótipo lipolítico do clone. No inserto de DNA de MAF1LP018 oito genes codificam proteínas hipotéticas, sendo que 4 não possuem domínios conservados identificáveis pelos programas de análises.

TABELA 14: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP018

ORFs	GC%	Tamanho (aa)	^a Função	^a Organismo	^a Nº de acesso	^a % identidade/positivos	^a E-value	^b Domínios conservados
ORF1	59.04	599	peptidase M61 domain protein	<i>Pedobacter heparinus DSM 2366</i>	YP_003093987	37/58	3,00E-118	sim
ORF2	61.64	617	hypothetical protein GobsDRAFT_08290	<i>Geodermatophilus obscurus DSM 43160</i>	ZP_03888660	26/41	3,00E-21	não
ORF3	60.40	510	hypothetical protein FBBAL38_02135	<i>Flavobacteria bacterium BAL38</i>	ZP_01733111	26/42	7,00E-32	não
ORF4	58.94	461	thiol oxidoreductase-like protein	<i>Solibacter usitatus Ellin6076</i>	YP_823769	49/62	7,00E-105	sim
tRNA ^{SER}								
ORF5	62.87	386	hypothetical protein Acid_2690	<i>Solibacter usitatus Ellin6076</i>	YP_823964	48/65	1,00E-92	sim
ORF6	60.04	306	hypothetical protein Acid_5346	<i>Solibacter usitatus Ellin6076</i>	YP_826580	49/64	1,00E-78	sim
ORF7	61.28	408	hypothetical protein BuboB_24794	<i>Burkholderia ubonensis Bu</i>	ZP_02380964	38/52	9,00E-70	não
ORF8	59.03	248	phosphoglycerate mutase	<i>Solibacter usitatus Ellin6076</i>	YP_825620	82/89	3,00E-117	sim
ORF9	58.09	427	enolase	<i>Solibacter usitatus Ellin6076</i>	YP_825619	84/91	0.0	sim
ORF10	56.63	401	glucose-1-phosphate adenyltransferase	<i>Solibacter usitatus Ellin6076</i>	YP_825618	80/90	0.0	sim
ORF11	48.01	83	SSU ribosomal protein S21P	<i>Solibacter usitatus Ellin6076</i>	YP_825617	97/97	4,00E-26	sim
ORF12	55.76	103	hypothetical protein Acid_4370	<i>Solibacter usitatus Ellin6076</i>	YP_825616	89/91	4,00E-46	não
ORF13	60.58	421	carbohydrate-selective porin OprB	<i>Solibacter usitatus Ellin6076</i>	YP_823998	66/77	2,00E-142	sim
ORF14	60.34	363	D-mannonate dehydratase	<i>Solibacter usitatus Ellin6076</i>	YP_826761	73/85	4,00E-163	sim
ORF15	58.82	594	peptidase S8 and S53, subtilisin, kexin, sedolisin	<i>Nitrosospora multiformis ATCC 25196</i>	YP_411234	47/59	3,00E-112	sim
ORF16	58.76	459	hypothetical protein Nmul_A0535	<i>Nitrosospora multiformis ATCC 25196</i>	YP_411235	38/53	1,00E-73	sim
ORF17	51.98	83	addiction module toxin, Txe/YoeB family	<i>Desulfatibacillum alkenivorans AK-01</i>	YP_002430939	72/90	2,00E-31	sim
ORF18	56.37	80	prevent-host-death family protein	<i>Alkalilimnicola ehrlichii MLHE-1</i>	YP_742621	66/85	2,00E-25	sim
ORF19	57.78	828	hypothetical protein sce5633	<i>Sorangium cellulosum 'So ce 56'</i>	YP_001616276	64/77	0.0	não
ORF20	60.69	239	peroxidase, putative	<i>Nitrosospora multiformis ATCC 25196</i>	YP_411230	49/65	2,00E-53	sim

NOTA: ^aBaseado no melhor resultado de BLAST usando BLASTp e o banco de dados de proteínas não redundantes; ^bBaseado no banco de dados InterPRO

5.3.5 Caracterização genética do fósmeo MAF1LP090

O inserto de DNA presente no fósmeo MAF1LP090 possui 34.126 bases e um conteúdo G+C de 56%. Esses parâmetros foram obtidos utilizando-se somente os contigs usados na anotação. Foram identificadas 27 ORFs e a comparação das sequências de aminoácidos com o GenBank é apresentada na Tabela 15. A maioria das sequências (37%) apresentou similaridade com proteínas da bactéria *Koribacter versatillis* Ellin345. A ORF 24, presente no contig 16, codifica uma proteína hipotética, também de *K. versatillis*, com domínio $\alpha\beta$ -hidrolase e foi selecionada como candidata responsável pelo fenótipo lipolítico do clone MAF1LP090 (denominada LP090ORF24 a partir desse ponto). Nesse fragmento de DNA também foram encontrados outros genes interessantes. As ORFs 1 a 16 provavelmente constituem uma via metabólica e codificam proteínas responsáveis pela síntese e exportação de exopolissacarídeos. A ORF 21 codifica uma proteína hipotética, que possui um domínio policetídeo ciclase Cyc2, envolvido na síntese de policetídeos e no transporte de lipídeos. A ORF 27 codifica uma α -amilase que também possui ampla aplicação nas indústrias têxtil e alimentícia.

A ORF 22 codifica uma proteína reguladora da síntese de poli-hidroxicanoato (PHA), um composto usado como matéria prima para produção de plástico biodegradável. A presença de uma proteína reguladora do acúmulo de PHA levantou a hipótese de existir, no inserto de DNA de MAF1LP090, uma proteína envolvida com síntese ou degradação de PHA. Assim sendo, as sequências de aminoácidos de todas as proteínas completas identificadas nesse inserto de DNA foram comparadas com o banco de dados PHADED (*PHA Depolymerase Engineering Database*) (KNOLL, 2009) e a ORF24, identificada como lipase, apresenta similaridade a família de PHA depolimerase intracelular com domínio lipase (*Lipase Box Family* - E-value 4e-4). PHA depolimerases são carboxilesterases e pertencem à família $\alpha\beta$ -hidrolases. No entanto, não é possível afirmar se LP090ORF24 está realmente envolvida na degradação de PHA sem uma análise funcional, diferente da atividade lipolítica que foi constatada em meio sólido seletivo.

A sequência que codifica a α -amilase (ORF27) está dividida entre os contigs 16 e 13. No contig 13 está a região N-terminal da enzima, enquanto que no contig 16 estão as regiões central e C-terminal. Ambos fragmentos apresentaram similaridade com a α -amilase de *K. versatillis* (número de acesso YP_590578) demonstrando que se trata da mesma proteína. Outras duas ORFs chamam a atenção pela baixa similaridade com sequências de proteínas depositadas no GenBank. A ORF 10 codifica uma proteína hipotética de 445 aminoácidos similar a uma proteína de *Vitis vinifera* (uva). A ORF 14 codifica uma proteína de 256 aminoácidos, similar a uma proteína hipotética de *Oryza sativa* (arroz). Em ambas as comparações, o valor de E (*E-Value*) foi alto, mas, no entanto, as duas proteínas apresentam regiões conservadas. A ORF 10 possui um domínio de proteínas da família das porinas e a ORF14 possui um peptídeo sinal e uma região transmembrana. Essas características justificam a inclusão desses genes como ORFs válidas. Foram identificadas ainda 10 proteínas hipotéticas, sendo que 7 apresentam regiões conservadas.

TABELA 15: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone MAF1LP090

ORFs	GC%	^a Tamanho (aa)	^b Função	^b Organismo	^b Nº de acesso	^b % identidade/positivos	^b E-value	^c Domínios conservados
Contig 18								
ORF1	53,1	377c	protein of unknown function DUF354	<i>Geobacter sp.</i>	ZP_0512009	59/74	3,00E-113	sim
ORF2	52,1	346c	hypothetical protein Sfum_0970	<i>Syntrophobacter fumaroxidans</i>	YP_845101	35/53	5,00E-52	sim
ORF3	55,6	400c	DegT/DnrJ/EryC1/StrS aminotransferase	<i>Syntrophobacter fumaroxidans</i>	YP_845103	66/79	7,00E-153	sim
ORF4	50	181c	hexapaptide repeat-containing transferase	<i>Syntrophobacter fumaroxidans</i>	YP_845104	72/91	4,00E-66	sim
ORF5	53	334c	oxidoreductase domain-containing protein	<i>Syntrophobacter fumaroxidans</i>	YP_845105	66/80	9,00E-134	sim
ORF6	55,3	392c	undecaprenyl-phosphate galactose phosphotransferase	<i>Syntrophobacter fumaroxidans</i>	YP_845106	44/61	2,00E-92	sim
ORF7	58	312c	capsular polysaccharide biosynthesis protein	<i>Solibacter usitatus Ellin6076</i>	YP_822924	35/58	7,00E-35	sim
ORF8	55,3	319c	general secretion pathway protein-related protein	<i>Solibacter usitatus Ellin6076</i>	YP_822701	47/63	1,00E-64	sim
ORF9	55,1	536c	lipopolysaccharide biosynthesis protein	<i>Geobacter uraniiireducens</i>	YP_001231331	31/51	1,00E-52	sim
ORF10	53,5	445c	hypothetical protein	<i>Vitis vinifera</i>	XP_002270317	26/43	7,30E-01	sim
ORF11	51,6	178c	polysaccharide export protein	<i>Stigmatella aurantiaca</i>	ZP_01461165	46/70	5,00E-37	sim
ORF12	58	460p	two component, sigma54 specific, Fis family transcriptional regulator	<i>Koribacter versatilis Ellin345</i>	YP_593208	37/55	4,00E-67	sim
Contig 17								
ORF12	52,9	107p	two component LuxR family transcriptional regulator	<i>Solibacter usitatus Ellin6076</i>	YP_823030	52/72	2,00E-05	sim
ORF13	58	87c	MerG protein	<i>Klebsiella sp.</i>	CAC82959	36/53	7,60E-01	não
ORF14	57,7	256c	hypothetical protein Osl_17916	<i>Oryza sativa</i>	CAH68208	29/37	1,10E-01	sim
ORF15	55,8	789c	TPR repeat-containing protein	<i>Solibacter usitatus Ellin6076</i>	YP_824583	26/48	2,00E-73	sim
ORF16	61,2	1768p	cell wall/surface repeat protein	<i>Paenibacillus sp.</i>	YP_003009652	27/41	2,00E-22	sim
Contig 11								
ORF16	57,9	107p	hypothetical protein Plarl_04580	<i>Paenibacillus larvae</i>	ZP_02326911	44/65	6,00E-16	sim
Contig 9								
ORF17	49,8	74c	SSU ribosomal protein S21p	<i>Solibacter usitatus Ellin6076</i>	YP_825617	100/100	4,00E-32	sim
ORF18	58,8	75p	hypothetical protein Acid_4370	<i>Solibacter usitatus Ellin6076</i>	YP_825616	43/53	0,051	não
Contig 6								
ORF19	51,4	93p	ABC spermidine/putrescine transporter, periplasmic ligand binding protein	<i>Koribacter versatilis Ellin345</i>	YP_590616	74/85	5,00E-23	não
ORF20	58,2	157p	spermidine/putrescine ABC transporter ATPase subunit	<i>Koribacter versatilis Ellin345</i>	YP_590617	55/69	1,00E-37	sim
Contig 5								
ORF21	56,5	151c	hypothetical protein sce1328	<i>Sorangium cellulosum</i>	YP_001611966	50/63	6,00E-32	sim
Contig 16								
ORF22	55,4	107p	PHA accumulation regulator DNA-binding-like protein	<i>Koribacter versatilis Ellin345</i>	YP_589430	74/85	2,00E-29	sim
ORF23	58	303c	hypothetical protein Acid345_0350	<i>Koribacter versatilis Ellin345</i>	YP_589429	46/65	2,00E-69	não
ORF24	59	261c	hypothetical protein Acid345_0349	<i>Koribacter versatilis Ellin345</i>	YP_589428	67/79	1,00E-96	sim
ORF25	53,8	230c	hypothetical protein Acid345_0360	<i>Koribacter versatilis Ellin345</i>	YP_589439	38/56	7,00E-38	sim
ORF26	57,5	159c	hypothetical protein Acid_5724	<i>Koribacter versatilis Ellin345</i>	YP_826956	49/57	2,00E-21	não
ORF27	56,9	398p	alpha amylase	<i>Koribacter versatilis Ellin345</i>	YP_590578	46/64	6,00E-103	sim
Contig 13								
ORF27	56,9	205p	alpha amylase	<i>Koribacter versatilis Ellin345</i>	YP_590578	46/59	2,00E-36	sim

NOTA: ^aSequências parciais (p) e completas (c) estão indicadas; ^bBaseado no melhor resultado de BLAST usando BLASTp e o banco de dados de proteínas não redundantes; ^cBaseado no banco de dados InterPRO

5.3.6 Identificação dos genes responsáveis pela atividade lipolítica

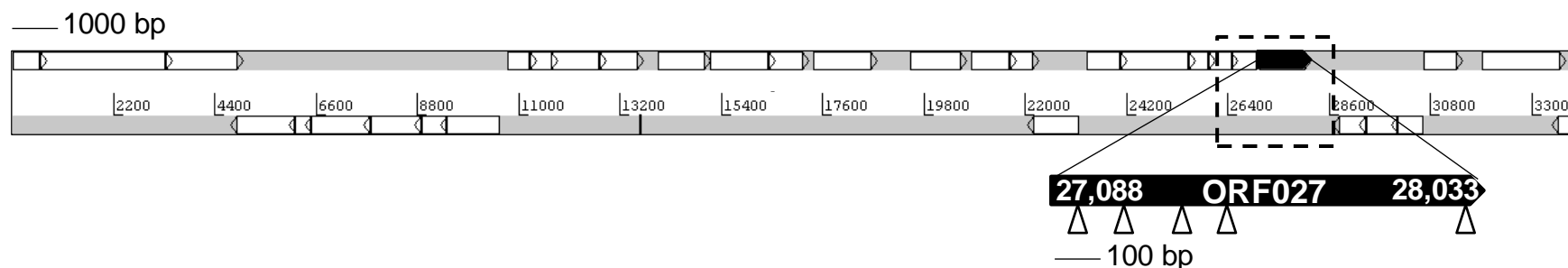
A identificação dos clones codificadores de enzimas lipases baseada unicamente na análise *in silico* não foi conclusiva. Para resolver esse problema primeiramente foi realizada a identificação de subclones em vetor pUC18 que apresentassem atividade lipolítica. Essa prospecção foi realizada inoculando as sub-bibliotecas em meio sólido contendo 1% de tributirina. A partir da sub-biblioteca SCMAF1LP001 foram recuperados 11 subclones positivos: P1G03, P1H06, P1H10, P2B09, P2D08, P2E03, P2F09, P3A09, P3G07, P3G11 e P4G03. Entretanto, nenhum clone positivo foi recuperado a partir da sub-biblioteca SCMAF1LP018 possivelmente porque as duas proteínas codificadas pelas ORFs 15 e 16 sejam necessárias para a atividade lipolítica. Então, uma nova sub-biblioteca, com 192 clones e tamanho médio de inserto de 8 kpb (SCMAF1LP018-8kb), foi construída e analisada contra o substrato tributirina. Sete subclones com atividade lipolítica foram recuperados: SCMAF1LP018-8kbP1B07, P1B11, P1E04, P1E12, P2C02, P2C08 e P2D07. Os clones das sub-bibliotecas SCMAF1LP090 também foram testados em tributirina, mas nenhum subclone positivo foi identificado. Os clones positivos das sub-bibliotecas LP001 e LP018 tiveram as extremidades dos seus insertos de DNA sequenciadas e alinhadas às respectivas sequências inserto do DNA fosmidial. Para MAF1LP001 as sequências de fato alinharam antes ou depois do gene que codifica a ORF27 indicando que essa proteína provavelmente é a responsável pela atividade lipolítica. Para MAF1LP018 as sequências também alinharam antes ou depois dos candidatos ORF15 e ORF16. Esse resultado gerou mais um indício de que os genes candidatos pré-selecionados estavam envolvidos na atividade lipolítica.

Para confirmar se os genes candidatos eram realmente os responsáveis pela atividade lipolítica foi utilizada a técnica de mutagênese aleatória por inserção de transposon. A inserção do transposon no gene que codifica a lipase iria inativá-lo e a célula perderia o fenótipo lipolítico. Pelo sequenciamento a partir das extremidades do transposon é possível mapear os sítios de inserção e identificar em qual gene o transposon foi inserido. Para realizar esse experimento foram utilizados subclones em

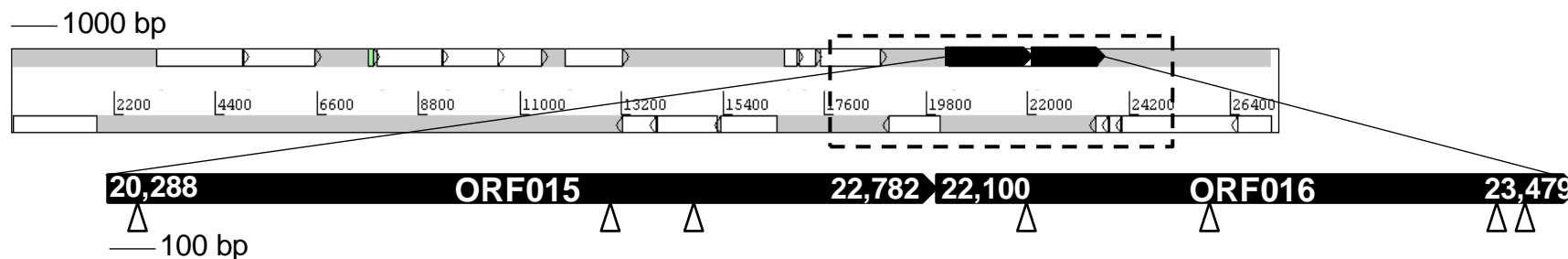
pUC com atividade lipolítica. O subclone em pUC é preferível ao clone em fosmídeo devido ao rendimento de DNA e à facilidade de manipulação. O subclone SCMAF1LP001P4G03 possui um fragmento de DNA compreendido entre as bases 26.300 a 28.840 e foi utilizado para reação de inserção de transposon. O mesmo foi feito para o subclone SCMAF1LP018-8kbP1E12 que possui um inserto de DNA compreendido entre as bases 17.590 to 24.984. A reação de inserção de transposon foi realizada *in vitro* e o plasmídeo resultante foi transformado em uma nova célula hospedeira. O DNA dos clones que perderam o fenótipo lipolítico foi purificado, sequenciados e a inserção do transposon mapeada. No subclone SCMAF1LP001P4G03 as inserções ocorreram nas bases 27.143, 27.244, 27.471, 27.310 e 27.989 confirmando a ORF27 como codificadora da lipase (Figura 30A). No subclone SCMAF1LP018-8kbP1E12 as inserções ocorreram nas bases 20.361, 21.556, 21.377, 22.277, 22.670, 23.288 e 23.348. As três primeiras inserções ocorrem na ORF15 e as três últimas na ORF16 sugerindo, novamente, que ambas as ORFs são necessárias para a atividade lipolítica (Figura 30B).

FIGURA 30: Representação esquemática da distribuição gênica e mutagênese por inserção de transposon.

(A) Inseto de DNA do clone MAF1LP001.



(B) Inseto de DNA do clone MAF1LP018



As duas fitas de DNA representadas. A numeração das ORFs inicia em 1 da esquerda para a direita independentemente da fita codificadora. A caixa pontilhada indica o fragmento de DNA no subclone positivo utilizado para a inserção de transposon (LP001SBP4G03 e LP018SB8kbP1E12). Cada Δ indica um sítio de inserção de transposon em um clone individual que perdeu a capacidade de hidrolisar tributirina.

5.3.7 Análises de sequência e filogenética das lipases

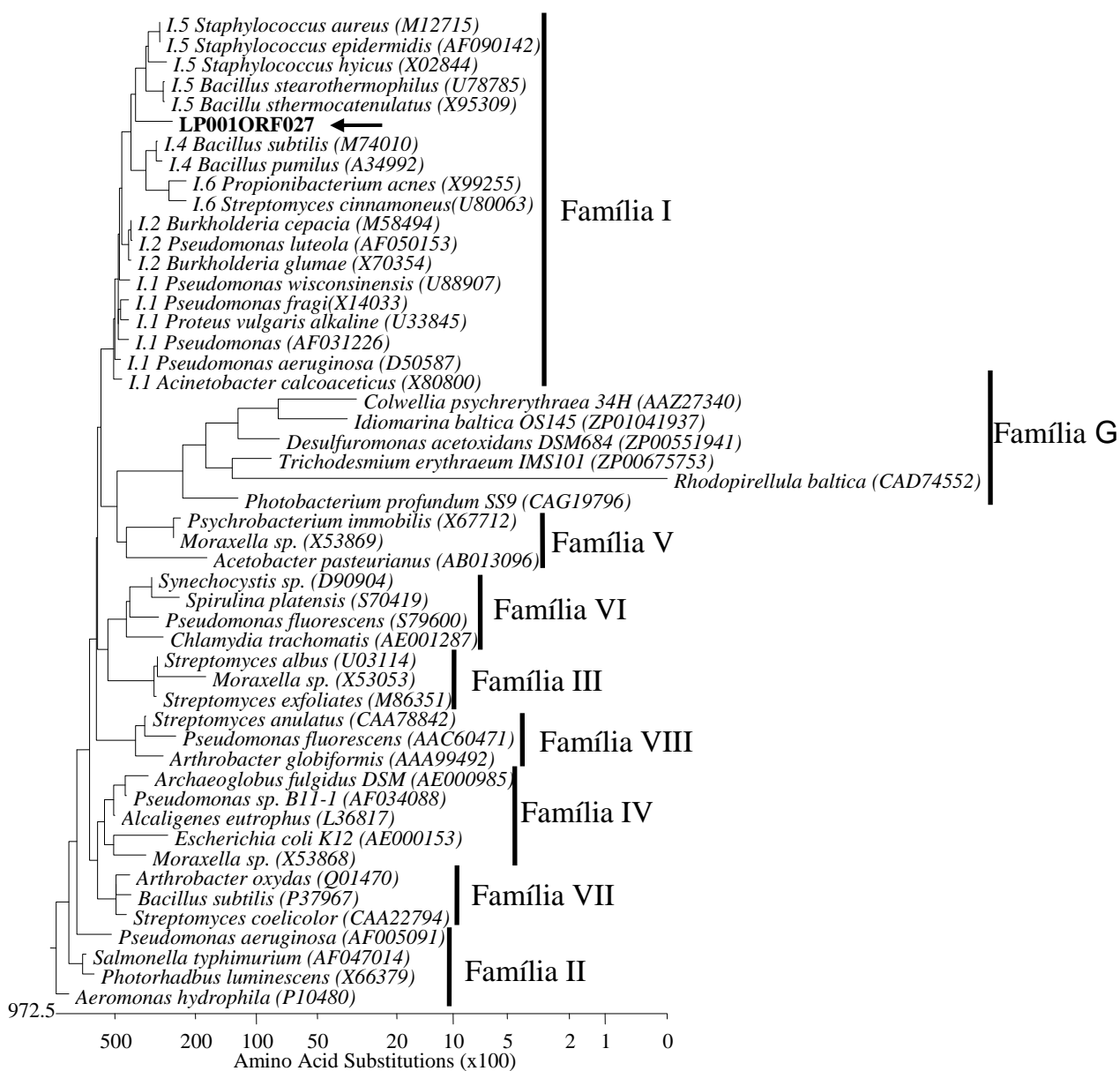
As análises filogenéticas das lipases encontradas foram realizadas baseadas na classificação das lipases bacterianas em famílias propostas por ARIGNY e JAEGER (1999) e atualizadas por JAEGER e EGGERT (2002) e LEE e colaboradores (2006). As sequências das lipases utilizadas nesses trabalhos foram recuperadas a partir do banco de dados GenBank através do seu número de acesso.

5.3.7.1 Lipase LP001ORF27

A enzima LP001ORF27 é codificada por um gene de 945 pb e possui 314 aminoácidos com uma massa molecular teórica de 33.881,9 Da. O ponto isoelétrico teórico é de 7,77 e não foram encontrados peptídeo sinal nem regiões transmembrana. A análise filogenética sugere que LP001ORF27 pertença à família I das lipases (Figura 31). A família I das lipases é composta por sete subfamílias e uma nova análise filogenética utilizando apenas as lipases das famílias I indicou que ela está próxima dos membros das subfamílias I.5 e I.6 (Figura 32). O primeiro aminoácido do pentapeptídeo conservado G-X-S-X-G foi substituído por uma alanina, comum em lipases bacterianas do gênero *Bacillus* (ARIGNY, 1999) e similar às das enzimas fúngicas usadas nas análises (Figura 33). A comparação contra o banco de dados de lipases *Lipase Engineering Database* (PLEISS, 2000) indicou que a enzima LP001ORF27 é membro da família 15.03 das $\alpha\beta$ -hidrolases e pertence à classe GX em relação à constituição da fenda oxianion (Figura 33). Como mencionado anteriormente, a enzima LP001ORF27 apresentou maior similaridade com sequências de enzimas fúngicas do que bacterianas na comparação com o GenBank. De fato, quando as sequências das enzimas fúngicas são adicionadas à análise elas formam um grupo separado, próximo às subfamílias I.5 e I.6, com LP001ORF27 no limite entre fungos e bactérias, mas mais próximas das enzimas fúngicas (Figura 32). A posição da tríade catalítica em LP001ORF27 (Anexos 4 e 5), bem como os resíduos imediatamente próximos também são semelhante à encontrada em enzimas fúngicas.

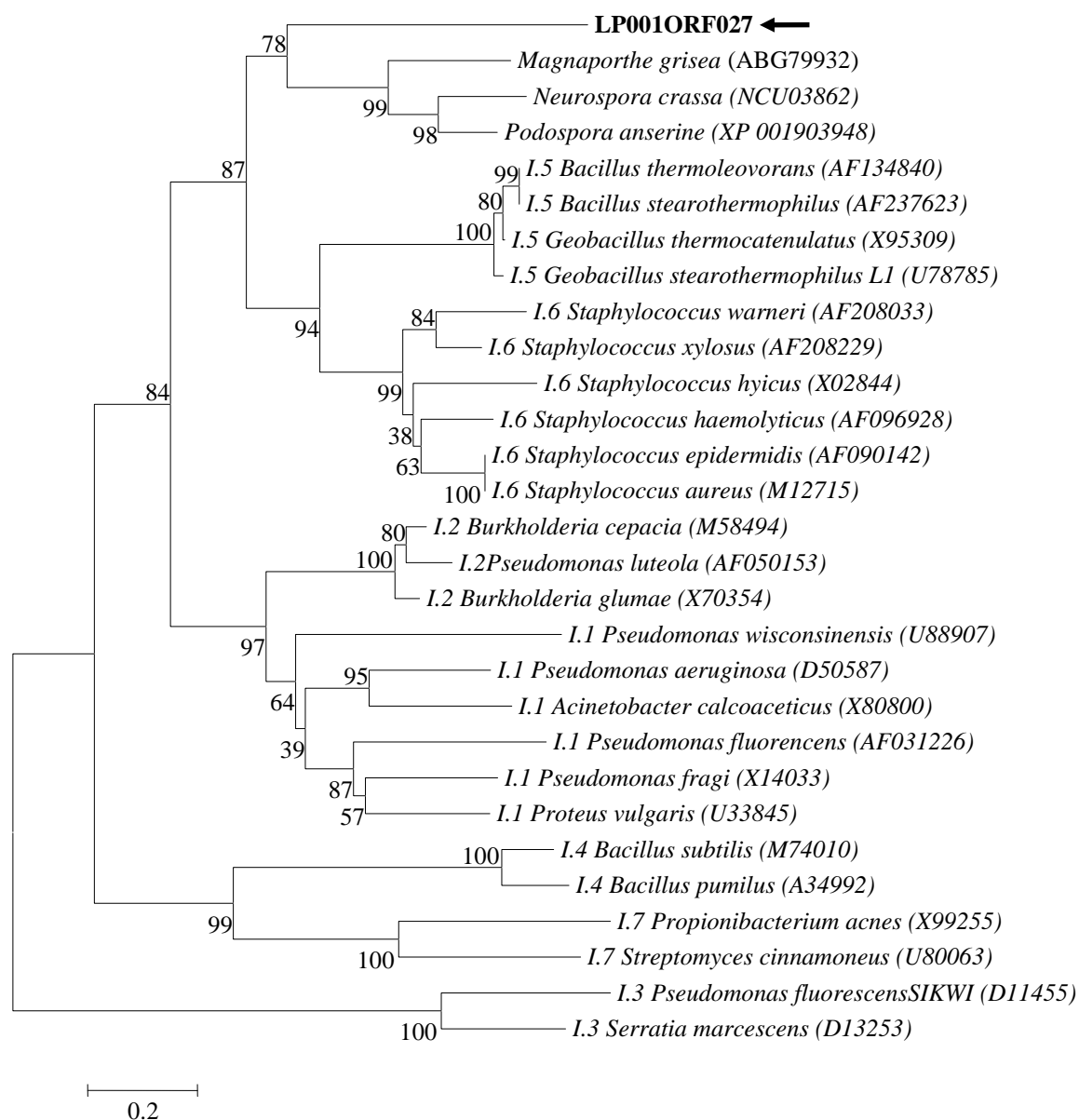
Entretanto, os resíduos que formam a fenda oxiânion em LP001ORF27 são GY (glicina-tirosina), diferentes das enzimas bacterianas (GF, glicina-fenilalanina) e fúngicas (GL, glicina-lisina).

FIGURA 31: Análise filogenética da proteína LP001ORF27



As seqüências de proteínas foram obtidas a partir do banco de dados GenBank baseados na classificação das lipases em famílias proposta por ARPIGNY e JAEGER (1999) e LEE e colaboradores (2006). O número de acesso de cada seqüência está indicado após o nome da espécie da qual a enzima foi originada. As subfamílias da família I estão indicadas. A árvore filogenética foi construída utilizando-se o programa *MegaAlign*. A posição referente à seqüência da enzima LP001ORF27 está indicada pela seta.

FIGURA 32: Análise filogenética da proteína LP001ORF27 e lipases da família I



As sequências de proteínas das lipases da família I foram obtidas a partir do banco de dados GenBank baseados na classificação proposta por ARPIGNY e JAEGER (1999) e atualizada por JAEGER e EGGERT (2000). O número de acesso de cada sequência está indicado após o nome da espécie da qual a enzima foi originada. As subfamílias estão indicadas antes do nome da espécie. A árvore filogenética foi construída utilizando-se o programa Mega4. As sequências de enzimas fúngicas foram selecionadas baseadas na comparação com o banco de dados GenBank pelo programa BLASTp. O número de bootstrap (1000 réplicas) está indicado acima dos ramos. A distância evolutiva está em unidades de número de substituições de aminoácido por posição. LP001ORF27 está próxima das subfamílias I.5 e I.6, mas é filogeneticamente mais similar com as enzimas de fungos.

FIGURA 33: Alinhamento múltiplo de seqüências de lipases das subfamílias I.5 e I.6, LP001ORF27 e proteínas fúngicas relacionadas

<i>B. stearrowthermophilus</i>	-----GRTEAASLRANDAPIVLLH GF TGWGRE-----	51
<i>B. sthermoleovorans</i>	-----GRAEAAA SRANDAPIVLLHGF TGWGRE-----	50
<i>G. thermocatenulatus</i>	-----GRTEAASPRANDAPIVLLH GF TGWGRE-----	51
<i>G. stearrowthermophilus</i>	-----GRTEAASPRANDAPIVLLH GF TGWGRE-----	51
<i>S. xylosus</i>	LKENAVATPKNQ-TQQNVEKANEQPNKA AKQGYKNQDP IIILVH GF NGFTDDINP----	382
<i>S. warneri</i>	LTKNAVAIYNNKSKAQATQPTK DQTNKVAPQQYKNHDP IIILVH GF NGFTADNGP----	374
<i>S. hyicus</i>	-----TVKAAPEAVQNPENPKNKDPFV FVH GF TG FVG--EV----	278
<i>S. aureus</i>	-----AAADAKKVRPLKANQVQPLNKYPV VFVH GF L GLVGDNAP----	322
<i>S. epidermidis</i>	-----AAADAKKVRPLKANQVQPLNKYPV VFVH GF L GLVGDNAP----	322
<i>S. haemolyticus</i>	-----QNIKPKNTDEATIKSNQYKNKYPV VLVH GF L GLVGDNAP----	348
<i>N. crassa</i>	-----ENPIVLAH GL LGF AEL KL AGS FLP	171
<i>P. anserina</i>	-----KHPV VLAH GL L GF AEL KL AGS YLP	119
<i>M. grisea</i>	-----KYPV VLAH GL L GF STLE L AGHLLP	115
LP001ORF027	-----MKP IVLAH GY L GF GNL G-----	17
	* . : : * *	
<i>B. stearrowthermophilus</i>	AKHG HAR FGRTYPGLLPELKRGGRIHII AH S QG QTARMLVSLLENGSQEEREYAKAHNV	171
<i>B. sthermoleovorans</i>	AKHG HAR FGRTYPGLLPELKRGGRIHII AH S QG QTARMLVSLLENGSQEEREYAKAHNV	170
<i>G. thermocatenulatus</i>	AKHG HAR FGRTYPGLLPELKRGGRVHII AH S QG QTARMLVSLLENGSQEEREYAKAHNV	171
<i>G. stearrowthermophilus</i>	ANDG HAR FGRTYPGLLPELKRGGRVHII AH S QG QTARMLVSLLENGSQEEREYAKAHNV	171
<i>S. xylosus</i>	EKYGH ERY GKTYEGVYKDWQPGK VHLV AH S MG QTVRQLEELLRNNGNQEIEYQKEHGG	501
<i>S. warneri</i>	AKYGH ERY GKSYAGAYRDWKPQGIHLI GH S MG QTVRLEEMLRNNGNPEEIEYQKQHGG	494
<i>S. hyicus</i>	EKYGH ERY GKTYEGVLKDWKPGH PVFI GH S MG QTI RL LEHYLRF GD KAIEAYQQQHGG	398
<i>S. aureus</i>	AKYGH ERY GKTYKGIMP NWEP GK VHLV GH S MG QTI RL MEEFLRNNGNKEEIA YH KAHGG	441
<i>S. epidermidis</i>	AKYGH ERY GKTYKGIMP NWEP GK VHLV GH S MG QTI RL MEEFLRNNGNKEEIA YH KAHGG	441
<i>S. haemolyticus</i>	AKT GH D RY GK FY Q G IMP DWEP GK KIHLI GH S MG QTI RL LEHFLRHNGNQEIE YD QKAHGG	467
<i>N. crassa</i>	SS SVE QRA EL ARGIE ARA Q-GKSVNII AH S MG GLDAR Y MIS-----QLK PAS VR--	246
<i>P. anserina</i>	SGS IE KRA AK L AQ DI E A QA Q-GKSVNV V AH S MG GLDAR Y MIS-----QLR PK GVD--	194
<i>M. grisea</i>	SAS IA VRA AK L AT Q I AD S VP S GT PVNI AH S MG GLDAR LL V S PA FAD QL S SV T AERLN--	199
LP001ORF027	PKDS V A ARA D LE A E IR AR FB GES VHII AH S MG GLDAR Y LL FP KN-----	88
	* : : : * * * * *	
<i>B. stearrowthermophilus</i>	NDGIVNTISMNGPKRGSNDRIVPYD GT L KK GVW ND MGTYN-VD H LEIIGVD----PNPSF	399
<i>B. sthermoleovorans</i>	NDGIVNTVSMNGPKRGSNDRIVPYD GT L KK GVW ND MGTYN-VD H LEIIGVD----PNPSF	398
<i>G. thermocatenulatus</i>	NDGIVNTVSMNGPKRGSNDRIVPYD GT L KK GVW ND MGTCN-VD H LEIIGVD----PNPSF	399
<i>G. stearrowthermophilus</i>	NDGIVNTISMNGPKRGSNDRIVPYD GT L KK GVW ND MGTYK-VD H LEIIGVD----PNPSF	399
<i>S. xylosus</i>	NDGLVSTISSQH PF NQA--FIEATD-EV KK GVW Q VTP IK HGW DH VDFV Q D ST DSNH PTE	718
<i>S. warneri</i>	NDGLVSVVSAQH PS NQK--YVDATD-Q I Q K GVW Q VTP V Q H D W H GFV GT Q D ENG IS E	711
<i>S. hyicus</i>	NDGLVSEISSQH PS D KN ISVDENS-EL H K GT W Q V M PT M K GW D H SDFIG ND AL D T K H SA I	616
<i>S. aureus</i>	NDG V VPV IS SL H PS N Q PF V N VT N DE P AT RR GI W Q V K PI I Q GW DH VDFIG V DF L DF K R K G A	661
<i>S. epidermidis</i>	NDG V VPV IS SL H PS N Q PF V N VT N DE P AT RR GI W Q V K PI I Q GW DH VDFIG V DF L DF K R K G A	661
<i>S. haemolyticus</i>	NDG I VPV SS SL F PT N QA AK TV G M T S P TT D KGI W Q V K P V M NG W D H LDFV GL DAT DY K R I G E	687
<i>N. crassa</i>	NDGLVSVQSS R WGTY K -----TL S GV NH LD L IN W SN R FR FT L Q R W M G RP P T-	394
<i>P. anserina</i>	NDGLVSV ES SQ W GT Y K-----TL T GV NH LD L IN W T N R I RS N L Q KL M GH P PS F	343
<i>M. grisea</i>	NDGLVSV ES SK W GT Y E-----TV V GV S HL L AL IN W T N R L G W T VR K W L Q E ET F	349
LP001ORF027	NDGLVSVAS AD W Q T S LG-----K L PA DH L G V I GW S A AS S D Y L S W Y D Q M V V N	300
	*** : * *	

O nome da espécie e o número de acesso foram os mesmos utilizados nas análises filogenéticas. *N. crassa*, *P. anserina* e *M. grisea* representam as lipases de origem fúngica. O pentapeptídeo conservado esta destacado em negrito. Os prováveis resíduos que formam a fenda oxianion estão destacados na caixa vazada e aqueles que formam a tríade catalítica estão sombreados em cinza. O posicionamento dos resíduos foi feito com base nos alinhamentos disponíveis no banco de dados LED. “*” indica aminoácidos idênticos, “:” indica uma substituição conservada e “.” indica uma substituição semi-conservada.

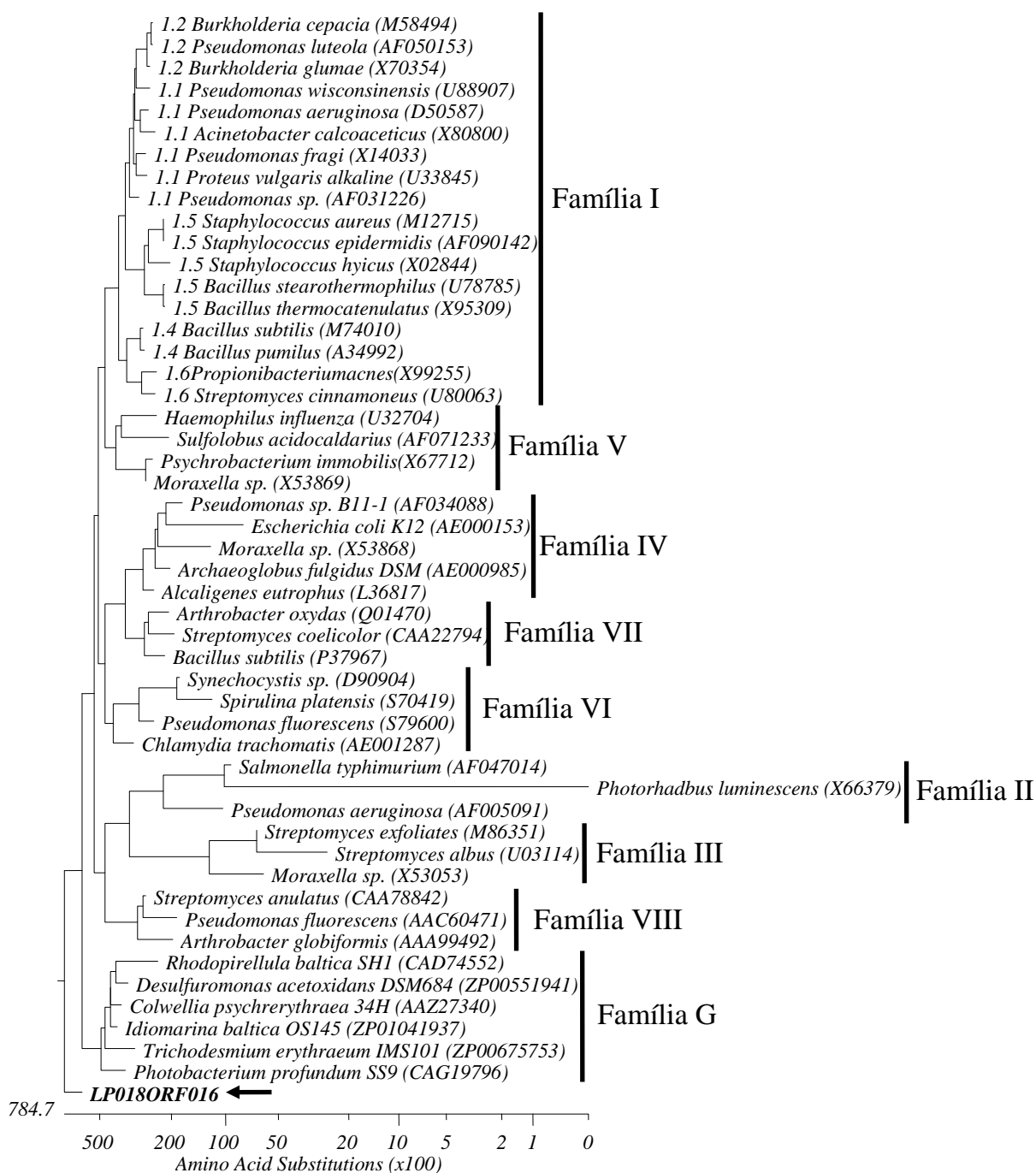
5.3.7.2 Lipase LP018ORF16

A lipase LP018ORF16 é maior que LP001ORF27. Ela é codificada por um gene de 1.380 pb e possui 459 aminoácidos. A massa molecular e o ponto isoelétrico teóricos são 48.901,5 Da e 5,62, respectivamente. Não foram encontrados peptídeo sinal nem regiões transmembrana. Ao contrário de LP001ORF27, a análise filogenética revelou que LP018ORF16 não está relacionada com nenhuma das famílias de lipases descritas (Figura 34) formando um grupo a parte. A comparação contra o banco de dados de lipases LED também não retornou resultados significativos e o tipo de dobramento $\alpha\beta$ -hidrolase não pode ser determinado, indicando que essa enzima apresenta uma conformação diferente. Como mencionado anteriormente, LP018ORF16 apresentou similaridade contra uma proteína hipotética de *N. multiformis* (nº de acesso YP411235). Além dessa, as outras três sequências mais similares também representavam proteínas hipotéticas de *Saccharopolyspora erythraea* (nº de acesso YP001107200), *Rhodococcus jostii* (nº de acesso YP707328) e *Gemmatimonas aurantiaca* (nº de acesso YP002762904), respectivamente. Como não foi encontrada relação filogenética com outras lipases, a sequência de aminoácidos de LP018ORF16 foi alinhada com sequências de aminoácidos dessas proteínas mais semelhantes, determinadas pela comparação contra o GenBank. O pentapeptídeo conservado característico está presente, indicando que se trata de uma lipase, na sua formação clássica G-X-S-X-G (Figura 35). Pela posição e conservação de resíduos de aminoácidos é possível inferir que LP018ORF16 pertence à classe GX, sendo que a fenda oxianion é composta pelos resíduos GW (glicina-triptofano). Os resíduos de aspartato e histidina, que compõem a tríade catalítica junto com a serina do pentapeptídeo, não puderam ser identificados devido à falta de relação com outras enzimas, mas pelo alinhamento gerado foi possível determinar os resíduos aspartato 380, 383, 439 e histidina 427 como candidatos a essa função.

O gene que codifica a lipase LP018ORF16 está em um operon com outro gene que codifica uma peptidase. A mutagênese por transposon realizada sugeriu que ambas as proteínas são necessárias para atividade lipolítica e que LP018 poderia estar atuando

como uma chaperona. A proteína LP018ORF15 é codificada por um gene de 1.785 pb e contém 594 resíduos de aminoácidos. Ela apresenta uma massa molecular de 64.053,7 Da e pI teórico de 5,84 semelhante a LP018ORF16 (pI 5,62). Não foram encontrados peptídeo sinal nem regiões transmembrana na sequência dessa proteína. A comparação contra o banco de dados GenBank indicou que essa proteína apresenta similaridades com sequências de enzimas peptidases e nenhuma sequência de chaperona foi encontrada na lista de proteínas similares, resultado também obtido durante a busca de domínios conservados. A relação de LP018ORF15 com chaperonas bem caracterizadas de *P. aeruginosa* (CHIHARA-SIOMI et al, 1992), *Vibrio cholerae* (OGIERMAN et al, 1997), *Burkholderia cepacia* (JØRGENSEN, SKOV e DIDERICHSEN, 1991) e *Acinetobacter calcoaceticus* (KOK et al, 1995) também foi investigada. De modo semelhante ao resultado obtido por comparação como o banco de dados, a análise filogenética indicou que LP018ORF15 está mais próxima das peptidases do que das chaperonas clássicas de lipases (Figura 36).

FIGURA 34: Análise filogenética da proteína LP018ORF16



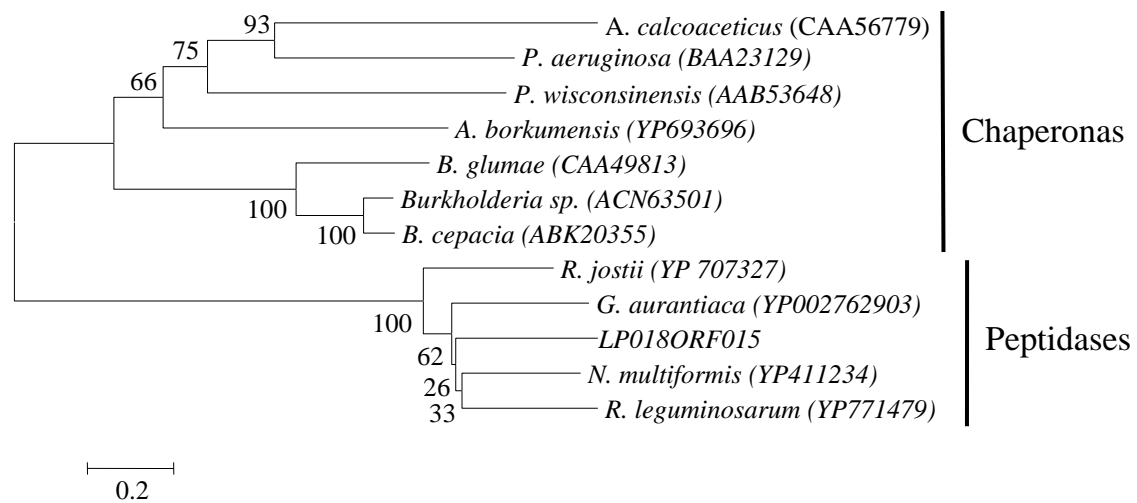
As seqüências de proteínas foram obtidas a partir do banco de dados GenBank baseados na classificação das lipases em famílias proposta por ARPIGNY e JAEGER (1999) e LEE e colaboradores (2006). O número de acesso de cada seqüência esta indicado após o nome da espécie da qual a enzima foi originada. A árvore filogenética foi construída utilizando-se o programa *MegaAlign*. A posição referente à seqüência da enzima LP018ORF16 esta indicada pela seta.

FIGURA 35: Alinhamento múltiplo de seqüências da lipase LP018ORF16 e proteínas relacionadas.

S.erythraea	-----MEFDSDAHVV-GSADPVLRL---VGDQAVTDLLVLSH GW NNDKAEALRLYR	47
R.jostii	--MITGHWDI EFTK DGLVH-DEHQVTRLL---DGLTDTTDLVLAH GW NNDKAEATLYD	54
LP018ORF016	MSFAYPYFPVQFTKDATVF-QQSDVDGLMNAVAGPAAPSDFMCH GW NNMDDATDLYS	59
N.multiformis	MNEIAGIPYVEASFDKNGQ-GPEEEVALP-----AGVTDLIIMSH GW NNKSKAEDLYN	53
G.aurantiaca	MSTLSGFDYVALPYDGEKGKPTSAKALNELLLALVKSTKATDVLLLAH GF RNDEKEASSLYE	60
	: * . :*:::.**.*: .* **	
S.erythraea	DLTSSLRAAVDSGAGPQPA-GRTIGVVAVLWPSKRFTEPGLVAAG-STALAPPIDSQIIL	105
R.jostii	TLITSLEAVGLPESGP----GSRVAVMRVLWPSKKFALES LIPGG-GAASATTANDAALL	109
LP018ORF016	ALAGLIKAQVD--ANPALA-TRKYAICGVLWPSKKFEDKDLIPSG-AAALNDAVTIEQLK	115
N.multiformis	GFFASFAAAAQPNDLs----ARKLVILGVFWPSREYDTSVAVSGTPATGGGGAALGGGNG	109
G.aurantiaca	RFLDNFRTHLTHANIAAMKDRRFVAAGVFWPSKTFYEKPKDDEG-----GTQSFDDG	113
	: : : . :*::: : :	
S.erythraea	EQIRELR--TVFP-----DPAAQRTLDFAAALVPLGLST-ADARAAFAEQMRSLLASTTLD	157
R.jostii	TALQQLEQDPVLLGADGTDPTFRADMVERAKAQIEDLDT SADARREYVLQIRNLLDPSQAH	169
LP018ORF016	SRVRDLRSLYSASEWPM DAGPAPAAAFDEIESLMDDIEDDPASQARVVLLRPLLPQDAAS	175
N.multiformis	GSIKRLEEKLDDMKEIFTEAEQKLLDEAKALLPDLEEKASVRDLFVNKMRSLDDPSAAG	169
G.aurantiaca	DARDDVRQQLAAIDDGTLSATQKKALTKAIAMFDDIEDDTDKQDAFVKLLLSVVEEADED	173
	: . . : : . : :	
S.erythraea	PEDGATDLFTLPGEELMHRLA---LPVALAP-PEGASLD-----VDAGEDMAAFDTLFG	207
R.jostii	PDDASQEFLDLDPDTLFRFTG-EVPVLLPTGPGGAAGN-----TSGGG--AFLGDLLS	220
LP018ORF016	SDDASQRFKTKTLLTNLQKPLNPPVIPPKGGAASVDPFSTEPVSGLGGAACFRDVLG	235
N.multiformis	KEDASNIFKDGGNELMKNLKAEEEDLGEDVAGGGAALPLGVSTPARAEGGAAGLVEILS	229
G.aurantiaca	PTEGLELLKAQSGSELLAKLGRPVILPTAQDGEGETAAIGTRSGNDGEEGRPLFIGGYIK	233
	: . : . * : : : : : : :	
S.erythraea	GTWSAALHLVNFTYYTMKHRAGRIGENGLAPVLADIKRARPDLVHVL GH SFGSRLAAA	267
R.jostii	GAKAAARRIANFATYYTMKTRAGAVGRGVAAVLGRARERHPQLPIHLT GH SFGGRIVTA	280
LP018ORF016	GIKSGFLNFLNYTTYMLKARAGDVGVKGVAPLIQKLRGSRGDLRIHMI GH SFGCRILSA	295
N.multiformis	GFKAAAINLLNYTTYEMKTRSGTVGKNGVAPLIDKLAS--QVERIH LI GH S FGCRVVTs	287
G.aurantiaca	SAAGRVGQLLNMTTWYTMKNRSGTVGGAGVADAVRALKT-IPGVRVHL I GH S LGGRGMAA	292
	. . . : * : * * * * * : * * * : : : * : * * * * * : :	
S.erythraea	ASKALP-ADTIT--TSLLLQAAF SHHGFAEHWKPGQ-HGHFRAIIAEKKINGPVLATHTR	323
R.jostii	TAAALPPSDAPL--SLALLQAAF SHNGLAQRFDGQH-DGAFAVLAEHRVNGPVLVTHTK	337
LP018ORF016	AINALPEQEQFRPDTVLLLQAGF SHNGFAKDDEGEVGRGVFRDVIDKKKVRGPILITHTR	355
N.multiformis	AALGSK-TDKIK--SMILLQAAF SONGFSK-----IEQGfyRAIVDNHRVKGPVLITHTP	339
G.aurantiaca	CAKALSAAPMVQADSLSLLEAAF SHYGF SKDNHGHT-RGFFRDVIDKGVVRGPLISTYSY	351
	. : : * * : * * * : : : * * : : : : * * : * :	
S.erythraea	NDRAVGIAYAIASRISGDAASLVGDATDTYGGMGRNGAQRTSEAVAATLL---PVGARY	379
R.jostii	NDLAVGVAYPLASRIAHDNAAALGDHNDPYGGMGRNGAQHTPETDPTEQLLRKVPPTHQY	397
LP018ORF016	NDKAVGTAYPIASRINGVVAALGDANDIFGGLSNGTQ-TKDTPPERTMGTLLAVGQAY	414
N.multiformis	NDRAVGLAYPLASRISGDKTMAFGDKDDLFGAMGRNGAQKMEQGEAVEGK--LLSVGGPY	397
G.aurantiaca	QDTVVGKAYSIASRLADDNVKAI GDKDDQFGGIGRNGTQRYAEATNRTLA---KVGDKY	407
	: * . * * * : * * : . . * * * : * : * * * * * : :	
S.erythraea	TWQ----PRCPHLLAD-DFIKGHADVTGREVAFAICSAMAP----	416
R.jostii	TFV----PGKVFNLKAD-ATTIDHGDVANHAVALRDLVLSR---	435
LP018ORF016	PFANGVKPSTPCNLLAD-EFIKGHSDIRKPEVAYALTVAMAAIGAS	459
N.multiformis	AFK----ANTYFNLEAS-DFIKNHNDIKGKEVGYVVRKVVATAGMG	438
G.aurantiaca	AFT----LKVNNLDGSGGLIKDHS DVTNPEVTYAVACIAAAT---	446
	: . * * . . * . * * : * . : . : .	

O nome da espécie e o número de acesso foram os mesmos utilizados nas análises filogenéticas. O pentapeptídeo conservado esta destacado em negrito. Os prováveis resíduos que formam a fenda oxianion estão destacados na caixa vazada. “*” indica aminoácidos idênticos, “:” indica uma substituição conservada e “.” indica uma substituição semi-conservada.

FIGURA 36: Análise filogenética da proteína LP018ORF15

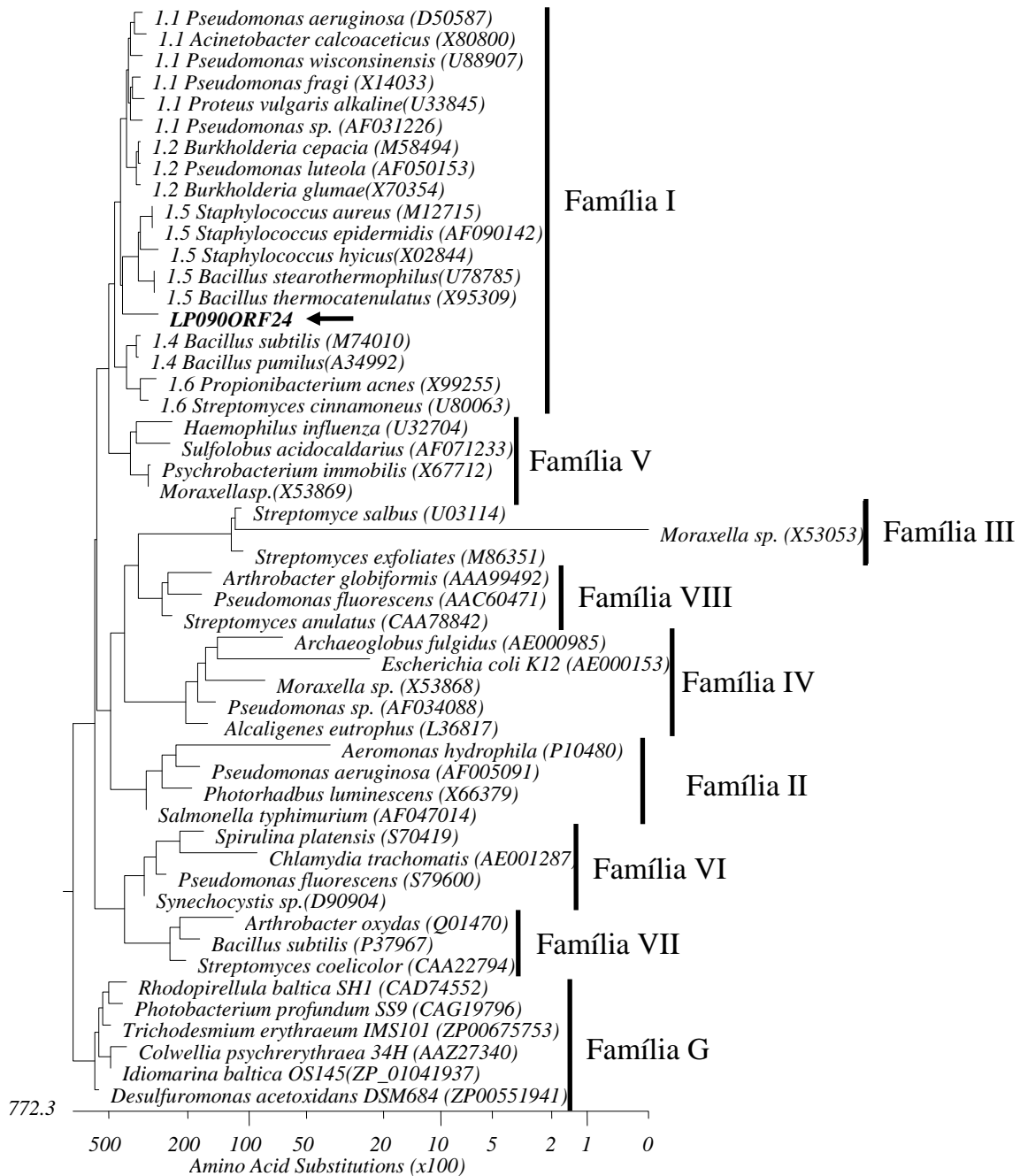


As sequências de proteínas de chaperonas associadas a lipases foram obtidas a partir do banco de dados GenBank. O número de acesso de cada sequência está indicado após o nome da espécie da qual a chaperona foi obtida. As sequências de proteínas de peptidases foram selecionadas com base na comparação de LP018ORF15 com o banco de dados GenBank e obtidas a partir do mesmo. O número de acesso de cada sequência está indicado após o nome da espécie. A árvore filogenética foi construída utilizando o programa Mega4. O número de bootstrap (1000 réplicas) está indicado acima dos ramos. A distância evolutiva está em unidades de número de substituições de aminoácido por posição.

5.3.7.3 Lipase LP090ORF24

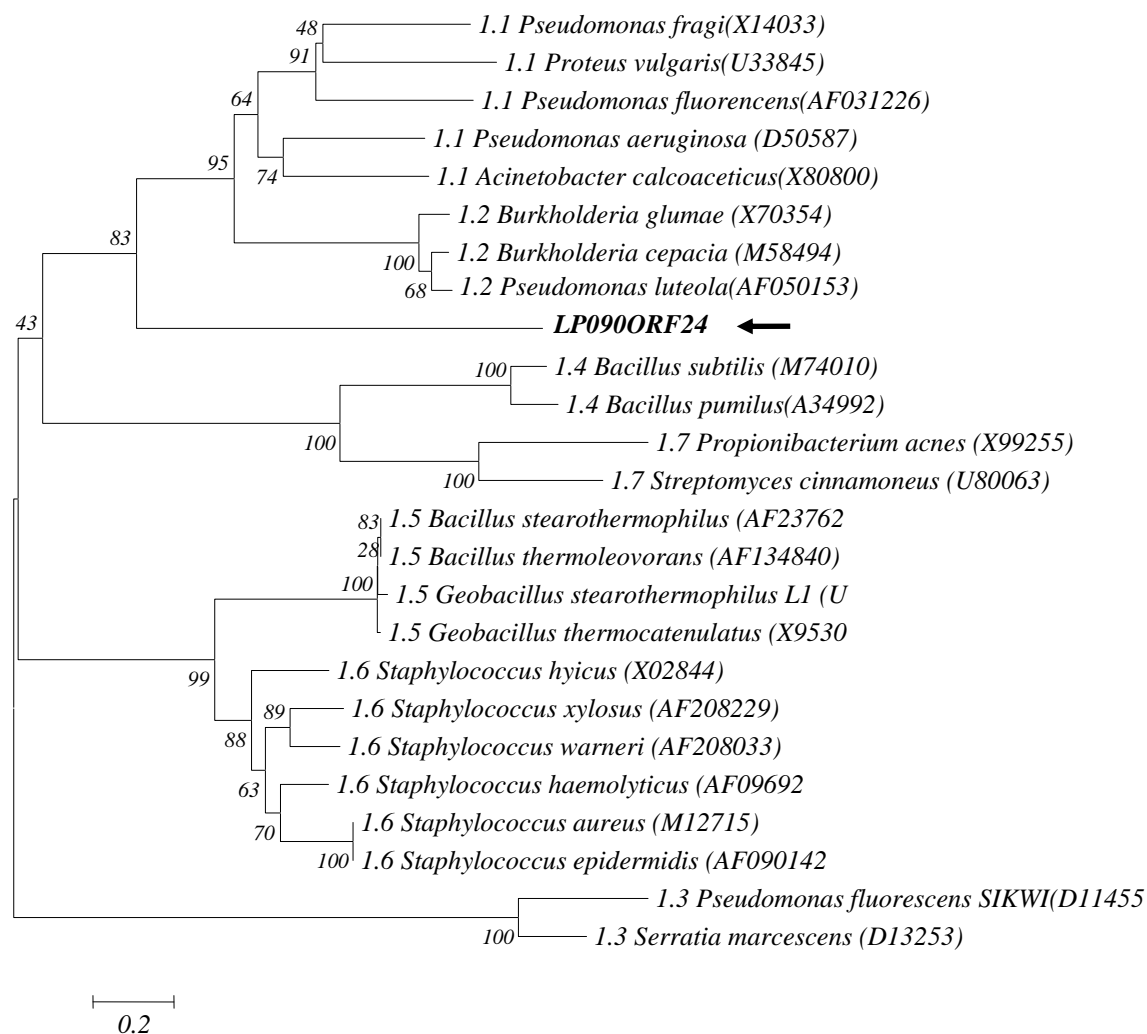
A sequência codificadora enzima MAF1LP090 foi identificada na montagem parcial das sequências de DNA desse clone no contig 16 ORF24 (a partir daqui referida com LP090ORF24). Ela possui 261 aminoácidos, sendo codificada por um gene de 786 pb. A massa molecular e ponto isoelétrico teóricos são 28,834.1 Da e 9.3, respectivamente. Não foram encontrados peptídeo sinal ou regiões transmembrana. De modo semelhante a LP001ORF27, a lipase LP090ORF24 foi classificada como membro da família I de lipase (Figura 37). Uma nova árvore filogenética foi construída utilizando apenas as lipases que compõem a família I. Essa nova análise indicou que LP090ORF24 compartilha características com as subfamílias I.1 e I.2 (Figura 38). As sequências de aminoácidos de LP090ORF24, das lipases das subfamílias I.1 e I.2 e dos dois melhores resultados da comparação contra o GenBank, proteínas hipotéticas de *Koribacter versatilis* (nº de acesso YP589428) e *Solibacter usitatus* (nº de acesso YP823062) foram alinhadas para identificação de regiões conservadas. O pentapeptídeo conservado foi identificado e também está na sua formação clássica G-X-S-X-G (Figura 39). A comparação contra o banco de dados de lipase LED mostrou que LP090ORF24 pertence à família 15.02 das $\alpha\beta$ -hidrolases e que sua fenda oxiânion é formada pelos resíduos GF (glicina-fenilalanina, Classe GX). Pelo alinhamento, também é possível inferir a posição dos resíduos que completam a tríade catalítica, aspartato 209 e histidina 232. Outra característica apresentada pelas lipases das subfamílias I.1 e I.2 é a presença de resíduos conservados de cisteína entre o aspartato e a histidina do sítio catalítico (ARPIGNY e JAEGGER, 1999). Esse resíduo também foi identificado nas proteínas LP090ORF24 (cisteína 217) e de *K. versatilis* e *S. usitatus* (Figura 39). Nas lipases das subfamílias I.1 e I.2 o resíduo de cisteína entre os resíduos catalíticos está ligado à estabilização do sítio ativo e é possível que ele desempenhe uma função semelhante em LP090ORF24. Aparentemente, ainda há outros quatro resíduos conservados de cisteína em LP090ORF24 (cisteínas 89, 178, 183 e 185), no entanto, não há contraparte para esses resíduos nas lipases das subfamílias I.1 e I.2.

FIGURA 37: Análise filogenética da proteína LP090ORF24



As sequências de proteínas foram obtidas a partir do banco de dados GenBank baseados na classificação das lipases em famílias proposta por ARPIGNY e JAEGER (1999) e LEE e colaboradores (2006). O número de acesso de cada sequência está indicado após o nome da espécie da qual a enzima foi originada. A árvore filogenética foi construída utilizando-se o programa *MegaAlign*. A posição referente à sequência da enzima LP090ORF24 está indicada pela seta.

FIGURA 38: Análise filogenética da proteína LP090ORF24 e lipases da família I



As sequências de proteínas das lipases da família I foram obtidas a partir do banco de dados GenBank baseados na classificação proposta por ARPIGNY e JAEGER (1999) e atualizada por JAEGER e EGGERT (2000). O número de acesso de cada sequência está indicado após o nome da espécie da qual a enzima foi originada. As subfamílias da família I estão indicadas. A árvore filogenética foi construída utilizando-se o programa Mega 4. As sequências de enzimas fúngicas foram selecionadas baseadas na comparação com o banco de dados GenBank pelo programa BLASTp. O número de bootstrap (1000 réplicas) está indicado acima dos ramos. A distância evolutiva está em unidades de número de substituições de aminoácido por posição.

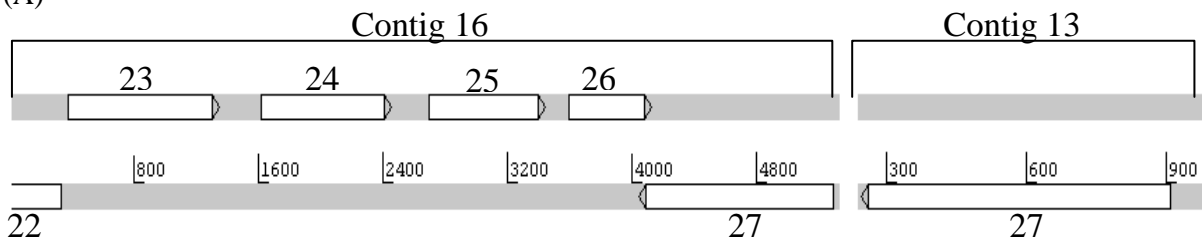
5.3.8 Análise de sequência da α -amilase

No inserto de DNA do clone MAF1LP090 foi encontrado um gene (ORF27 na Tabela 16) que codifica uma proteína similar a uma α -amilase de *K. versatillis*. No entanto, a sequência deste gene estava dividida entre os contigs 13 (região N-terminal) e 16 (região central e C-terminal) (Figura 40A). Esses dois fragmentos de sequência de proteína foram unidos, gerando um peptídeo de 603 resíduos de aminoácidos, e comparados contra o GenBank pelo programa BLASTp. Através dessa comparação foi possível identificar que ainda falta um fragmento de sequência entre os aminoácidos 193 e 200. Esse resultado foi reforçado pelo fato de que a proteína de *K. versatillis*, com a qual foi encontrada maior similaridade, possui 610 resíduos de aminoácidos, 7 a mais do que a ORF27. A análise dos cromatogramas das sequências que compõem os contigs 13 e 16 indica que esse fragmento de DNA já foi sequenciado, mas não foi alinhado devido à baixa qualidade das sequências nas extremidades dos contigs. Sendo assim, a proteína de 603 resíduos de aminoácidos gerada a partir dos dois fragmentos foi utilizada nessa análise.

A proteína LP090ORF27 possui uma massa molecular de 66.871,5 Da e um pI teórico de 5,82. Os 30 primeiros aminoácidos constituem um peptídeo sinal, sugerindo que essa proteína é exportada. Não foram identificadas regiões transmembrana. A busca por regiões conservadas indicou a existência de domínios glicosil hidrolases das famílias α -amilases e ciclomaltodextrina (Figura 40B). As α -amilases são endoamilases (EC 3.2.1.1) que hidrolisam ligações glicosídicas do tipo α -1,4 no interior da molécula de amido. Já as ciclomaltodextrina glucanotransferases (CGTase EC 2.4.2.19) são proteínas que não apenas realizam a hidrólise do amido, mas também são capazes de formar ciclodextrinas como produtos de reação através da ciclização de 6 a 8 resíduos de glucose.

FIGURA 40: Análise de sequência do contig MAF1LP090 e da proteína LP090ORF27

(A)



Anotação dos contigs MAF1LP090 16 e 13 pelo programa Artemis. O número da ORF, a posição e a orientação dos genes estão indicados. Os números na régua se referem ao número de bases. A ORF27 codifica uma enzima similar à α -amilase

(B)

SEQUENCE: [Sequence_1](#) CRC64: E841A2ECC226DF20 LENGTH: 603 aa

InterPro IPR006047 Domain	Glycosyl hydrolase, family 13, catalytic domain PF00128	Alpha-amylase
InterPro IPR013781 Domain	Glycoside hydrolase, subgroup, catalytic core G3DSA:3.20.20.80	no description
InterPro IPR015171 Domain	Cyclomaltodextrinase, N-terminal PF09087	Cyc-maltodext_N
InterPro IPR017853 Domain	Glycoside hydrolase, catalytic core SSF51445	(Trans) glycosidases
noIPR unintegrated	unintegrated	
	PTHR10357	AMYLASE
	PTHR10357:SF11	ALPHA-AMYLASE
	SignalP	signal-peptide
	SSF51011	Glycosyl hydrolase domain

Busca por regiões conservadas na LP090ORF27 pelo programa InterproScan. Domínios glicosil hidrolase, α -amilase e CGTase foram identificados.

5. RESULTADOS

5.4 Clonagem, expressão, purificação e caracterização das lipases

5.4.1 Amplificação e clonagem dos genes que codificam lipases

5.4.2 Expressão de proteínas LP001ORF27, LP018ORF15, LP018ORF16 e LP090ORF24

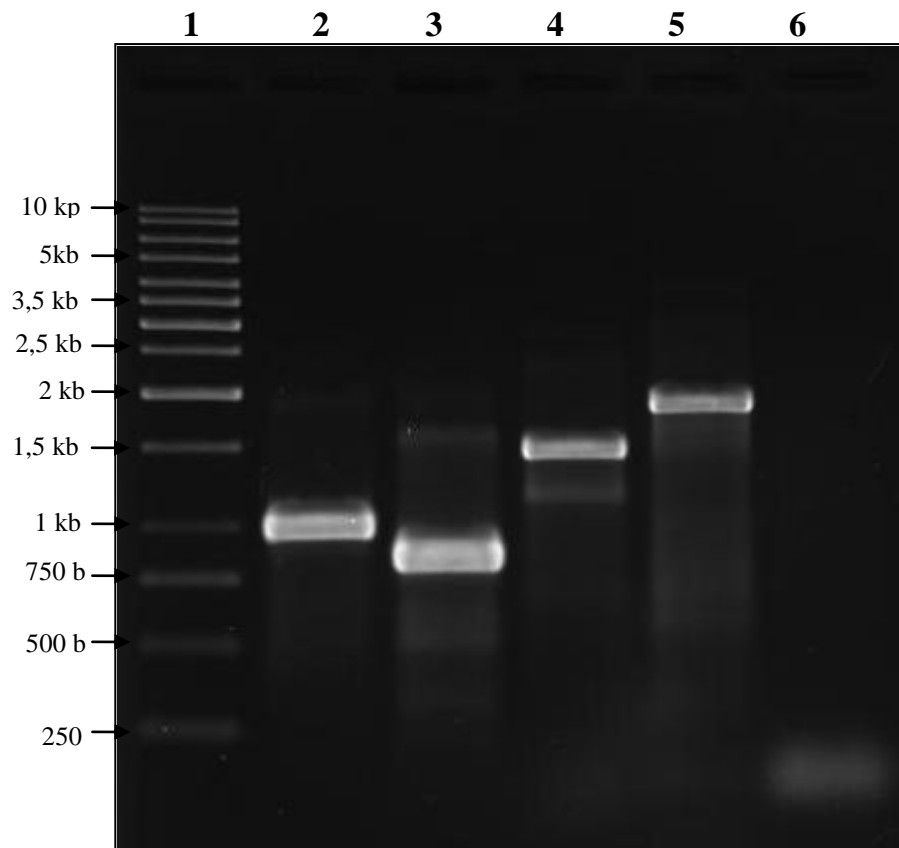
5.4.3 Purificação da proteína LP001ORF27

5.4.4 Caracterização da lipase LP001ORF27

5.4 Clonagem, expressão, purificação e caracterização das lipases

5.4.1 Amplificação e clonagem dos genes que codificam lipases

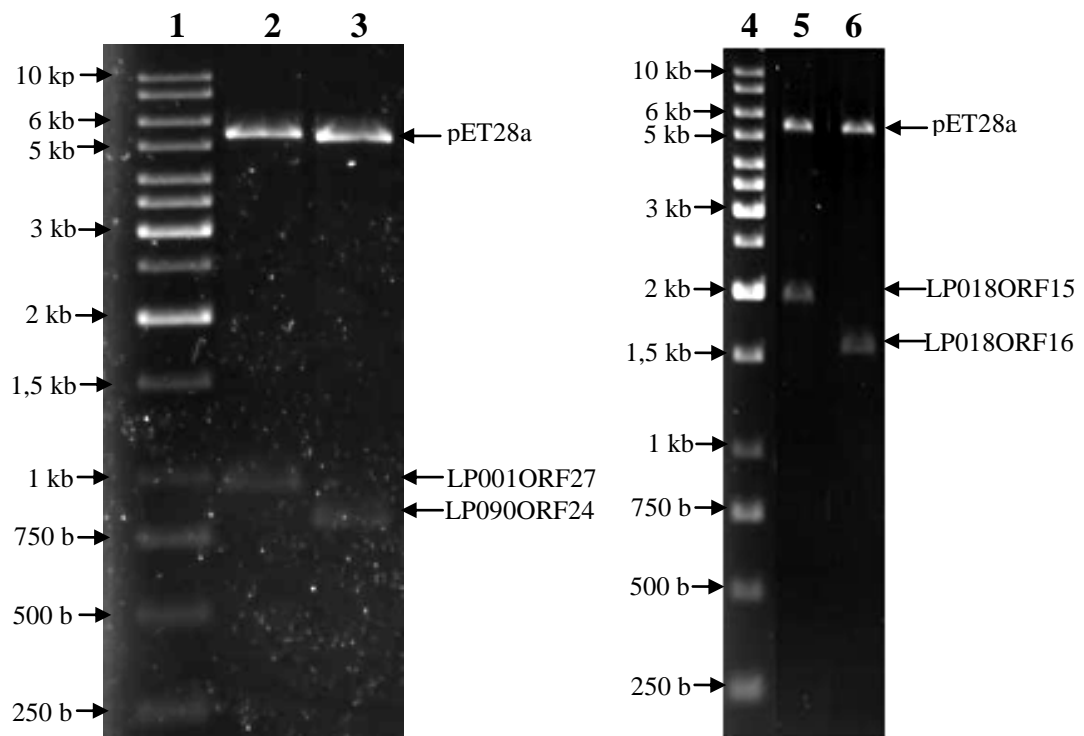
As sequências codificadoras de lipases identificadas nas etapas anteriores foram utilizadas como molde para a construção de iniciadores visando à amplificação desses genes para clonagem, expressão e purificação de proteínas. Foram construídos iniciadores para os genes codificadores das proteínas LP001ORF27, LP018ORF15, LP018ORF16 e LP090ORF24 (Tabela 5). A amplificação foi bem sucedida em todos os casos gerando fragmentos de aproximadamente 1000 pb (LP001ORF27), 1800 pb (LP018ORF15), 1400 pb (LP018ORF16) e 800 pb (LP090ORF24) (Figura 41). Os produtos de PCR foram purificados, digeridos com as respectivas endonucleases (sítios presentes nos iniciadores) e clonados no vetor de expressão pET28a digerido com as mesmas endonucleases. A identidade do gene clonado foi confirmada por perfil de restrição (Figura 42) e sequenciamento. Os novos plasmídeos foram denominados de acordo com o gene clonado: pETLP001ORF27, pETLP090ORF24, pETLP018ORF15 e pETLP018ORF16.

FIGURA 41: Amplificação dos genes codificadores de lipases

Eletroforese em gel de agarose 1% dos produtos de PCR.

- (1) Marcador de peso molecular 1kb *Gene Ruler*
- (2) Amplificado LP001ORF27 (~1000 pb)
- (3) Amplificado LP090ORF24 (~800 pb)
- (4) Amplificado LP018ORF16 (~1400 pb)
- (5) Amplificado LP018ORF15 (~1800 pb)
- (6) controle negativo da reação de amplificação sem DNA

FIGURA 42: Clonagem dos produtos de PCR no vetor de expressão pET28a



Eletroforese em gel de agarose 1%. Perfil de restrição dos plasmídeos recombinantes em vetor pET28a contendo os genes codificadores de lipases. A banda entre 5 e 6 kb corresponde ao vetor pET28a linearizado. As bandas menores correspondem aos produtos de PCR clonados disposto segundo a lista abaixo.

- (1) Marcador de peso molecular 1kb *Gene Rule*
- (2) pETLP001ORF27 digerido com as enzimas *NdeI/BamHI*
- (3) pETLP090ORF24 digerido com as enzimas *NdeI/BamHI*
- (4) Marcador de peso molecular 1kb *Gene Rule*
- (5) pETLP018ORF15 digerido com as enzimas *NheI/EcoRI*
- (6) pETLP018ORF16 digerido com as enzimas *NdeI/BamHI*

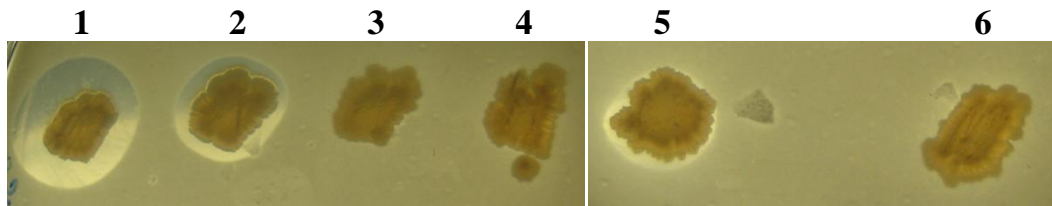
5.4.2 Expressão de proteínas LP001ORF27, LP090ORF24, LP018ORF15 e LP018ORF16

Os plasmídeos pETLP001ORF27, pETLP090ORF24, pETLP018ORF15 e pETLP018ORF16 foram transformados na estirpe hospedeira de *E. coli* BL21 Rosetta. Os transformantes obtidos foram testados quanto a sua atividade lipolítica em meio sólido contendo 1% de tributirina. As células que continham os plasmídeos pETLP001 e pETLP090 foram capazes de produzir halos de hidrólise indicando que o gene da lipase estava sendo corretamente expresso e que a enzima estava íntegra e ativa (Figura 43). No entanto, as células que receberam os plasmídeos pETLP018ORF15 e pETLP018ORF16 não produziram halos de hidrólise. Esses dois plasmídeos carregam, respectivamente, os genes LP018ORF15 e 16, que no fosmídeo original formam um operon. É possível que os dois genes sejam necessários para produzir uma enzima ativa, hipótese essa já considerada durante os experimentos de mutagênese com transposon. Sendo assim, as culturas de células contendo os plasmídeos pETLP018ORF15 e pETLP018ORF16 foram misturadas e novamente depositadas em uma placa de meio sólido contendo 1% de tributirina. Desta vez foi possível observar a presença de uma halo de hidrólise em volta da colônia (Figura 43), indicando que as proteínas estavam sendo corretamente expressas e confirmando que as duas proteínas são necessárias para a atividade lipolítica do clone MAF1LP018.

A expressão das proteínas foi avaliada no extrato celular bruto em gel desnaturante SDS antes e após indução com IPTG por 3 horas. Em todos os casos foi possível visualizar uma banda de proteína superexpressa após a indução (Figura 44) indicando que o sistema de expressão, formado pela célula hospedeira, vetor pET28a e o gene clonado, estava funcionando satisfatoriamente. O tamanho aproximado das proteínas superexpressas é 36, 31, 48 e 65 kDa para LP001ORF27, LP090ORF24, LP18ORF15 e LP018ORF16, respectivamente (Figura 44). Para a obtenção das frações solúvel e insolúvel as proteínas foram novamente superexpressas em condições mais brandas (18°C, 120 rpm durante 15 horas) para diminuir a taxa de expressão e permitir à célula hospedeira efetuar o correto dobramento das proteínas, evitando,

assim, a precipitação. Também foram testados tampões de lise com e sem detergente (Triton-X 100 a 0,5%). LP001ORF27, LP090ORF24, e LP18ORF16 ficaram quase que totalmente nas frações insolúveis em todas as condições testadas (Figura 45 e 46). LP18ORF15 foi encontrada na fração solúvel, mas sofreu clivagem evidenciada pela presença de duas bandas protéicas 30 e 35 kDa, ao invés de uma de 65 kDa (Figura 46). Considerando que LP18ORF15 é uma peptidase, possivelmente, o extenso tempo de indução e o acúmulo de proteínas superexpressas levaram a auto-hidrólise da proteína. Esse problema foi resolvido diminuindo o tempo de indução para 4 horas. Nesse caso, a proteína LP018ORF15 ficou íntegra, mas foi encontrada na fração insolúvel (dado não mostrado). Todas as frações, solúveis e insolúveis, das proteínas superexpressas foram novamente testadas em placas de meio sólido seletivo agar-tributirina 1% para avaliação da atividade lipolítica. Apesar de não ser possível visualizar quantidades significativas da proteína LP001ORF27 na fração solúvel quando aplicada no gel SDS-PAGE (Figura 45), estas frações apresentaram atividade lipolítica, tanto na presença quanto ausência de detergente (Figura 47). Os extratos solúveis de LP090ORF24 também apresentaram atividade, mas muito menor quando comparadas com LP001ORF27. Por sua vez, os extratos de LP18ORF15 e 16 não apresentaram atividade. Essa ausência pode estar relacionada com o fato da interdependência entre as duas proteínas para a atividade. A mistura dos extratos de LP18ORF15 e 16 também não apresentou atividade, possivelmente porque a clivagem inativou a proteína LP018ORF15 e ela não consegue atuar sobre LP018ORF16. Esses resultados levaram à escolha de LP001ORF27 para purificação e caracterização da enzima.

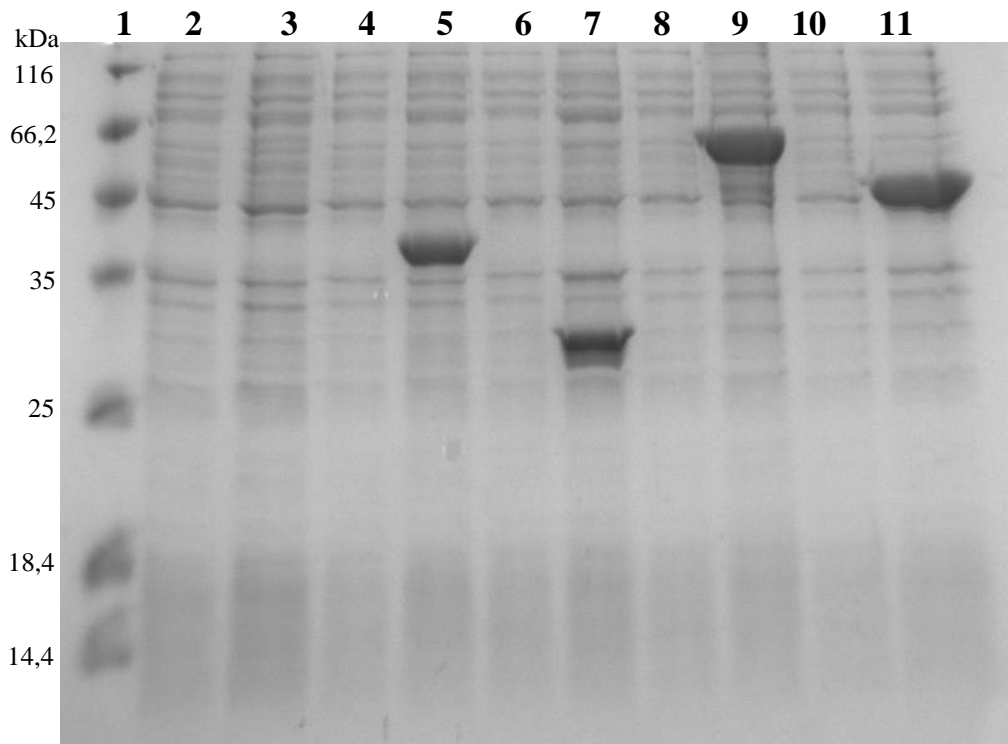
FIGURA 43: Teste do clones expressando lipases em meio seletivo.



Os transformantes contendo os genes codificadores de lipases clonados no vetor de expressão pET28a foram testados em meio LA suplementado com tributirina 1% para avaliação da atividade lipolítica. Os clones bacterianos foram crescidos em estufa a 37°C durante 7 dias.

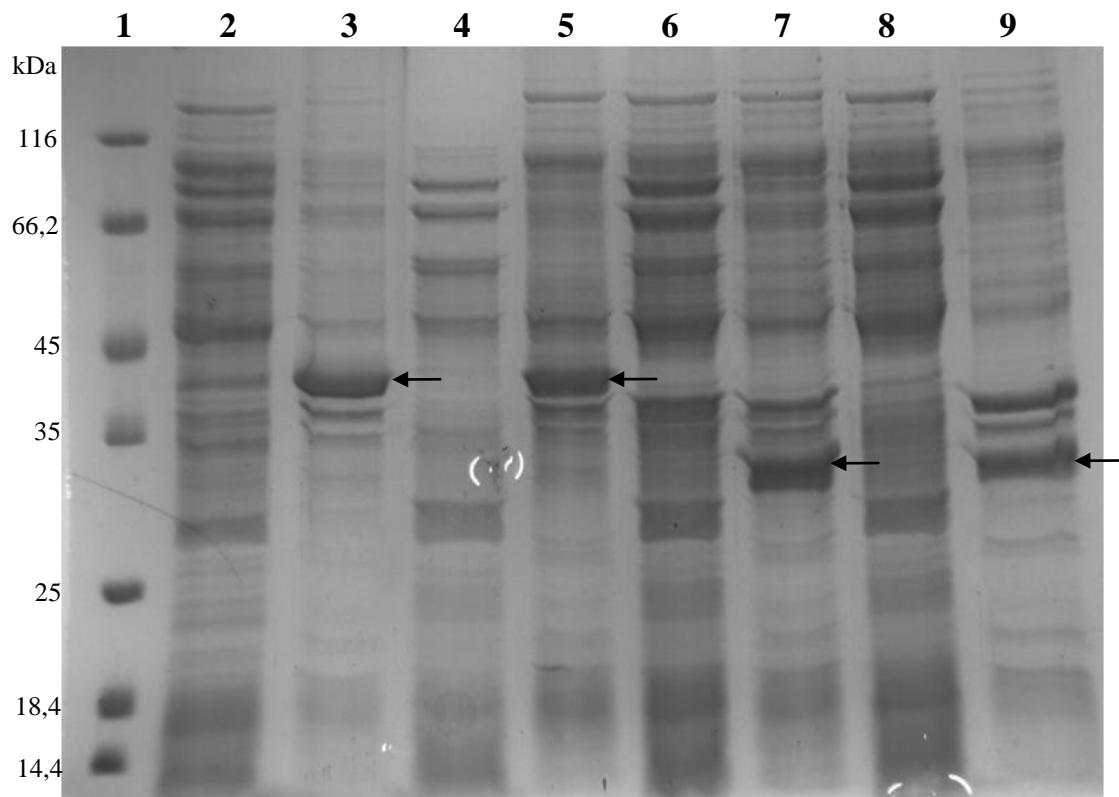
- (1) pETLP001ORF27 BL21 Rosetta (LP001ORF27)
- (2) pETLP090ORF24 BL21 Rosetta (1LP090ORF24)
- (3) pETLP018ORF15 BL21 Rosetta (LP018ORF16)
- (4) pETLP018ORF16 BL21 Rosetta (LP018ORF15)
- (5) pETLP018ORF15 BL21 Rosetta + pETLP018ORF16 BL21 Rosetta
- (6) Controle negativo pET28a BL21 Rosetta

FIGURA 44: Teste de expressão das proteínas LP001ORF27, LP090ORF24, LP018ORF15 e LP018ORF16



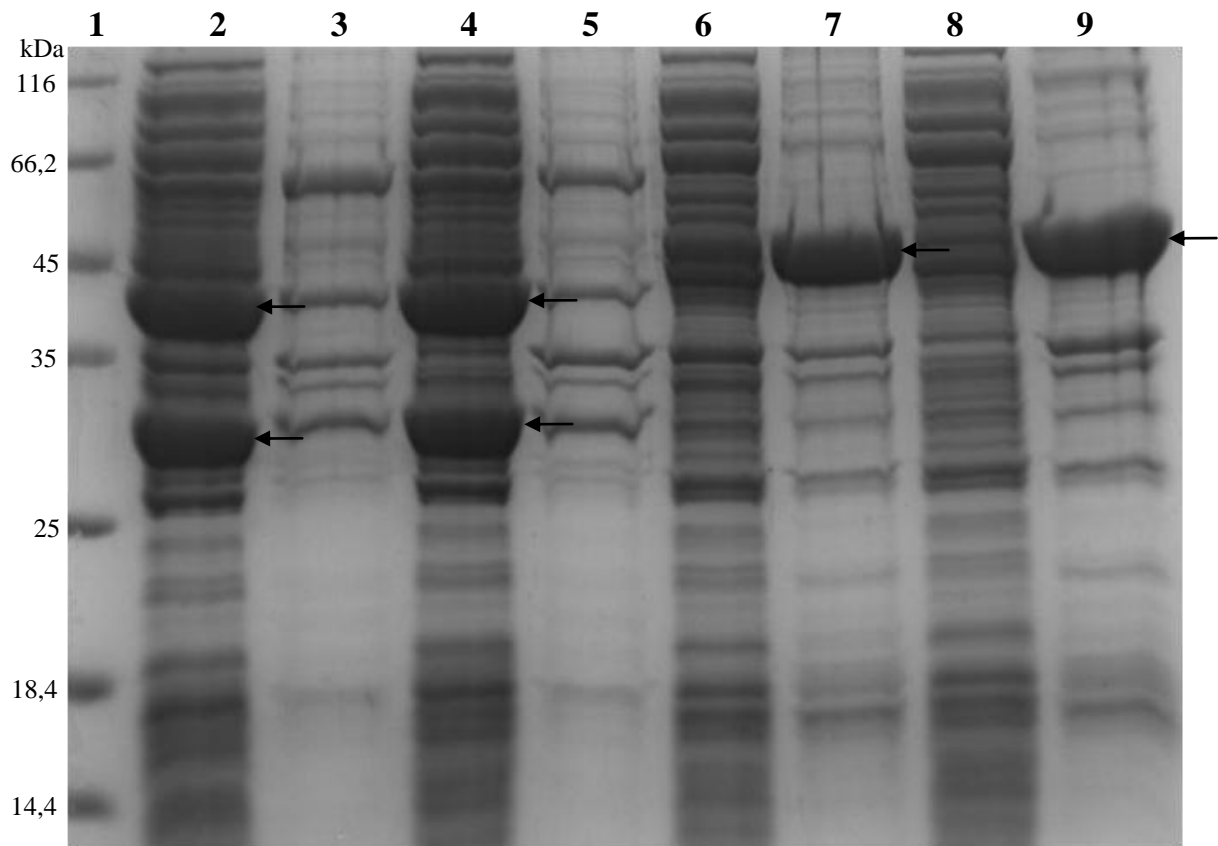
Eletroforese em Gel de SDS poliacrilamida 12% (SDS-PAGE) dos extratos celulares de *E. coli* Rosetta expressando as proteínas LP001ORF27, LP090ORF24, LP018ORF15 e LP018ORF16. Os perfis não induzidos e induzidos 3 horas com 0,5 mM de IPTG são mostrados. As proteínas foram coradas com Coomassie Blue R-250.

- (1) Marcador de massa molecular (em kDa)
- (2) Extrato do controle pET28a não induzido
- (3) Extrato do controle pET28a induzido
- (4) Extrato pETLP001ORF27 BL21 Rosetta não induzido
- (5) Extrato pETLP001ORF27 BL21 Rosetta induzido (~36 kDa)
- (6) Extrato pETLP090ORF24 BL21 Rosetta não induzido
- (7) Extrato pETLP090ORF24 BL21 Rosetta induzido (~31 kDa)
- (8) Extrato pETLP018ORF15 BL21 Rosetta não induzido
- (9) Extrato pETLP018ORF15 BL21 Rosetta induzido (~65 kDa)
- (10) Extrato pETLP018ORF16 BL21 Rosetta não induzido
- (11) Extrato pETLP018ORF16 BL21 Rosetta induzido (~48 kDa)

FIGURA 45: Teste de solubilidade das proteínas LP001ORF27 e LP090ORF24

Eletroforese em gel de SDS poliacrilamida 12% (SDS-PAGE) das frações solúvel e insolúvel, tratados e não tratados com 0,5% de TritonX-100, dos extratos celulares de *E. coli* Rosetta expressando as proteínas LP001ORF27 e LP090ORF24. Indução com 0,5 mM de IPTG, por 14 horas a 18°C e 120 rpm. As proteínas foram coradas com Coomassie Blue R-250. As setas indicam a banda correspondente às proteínas superexpressas.

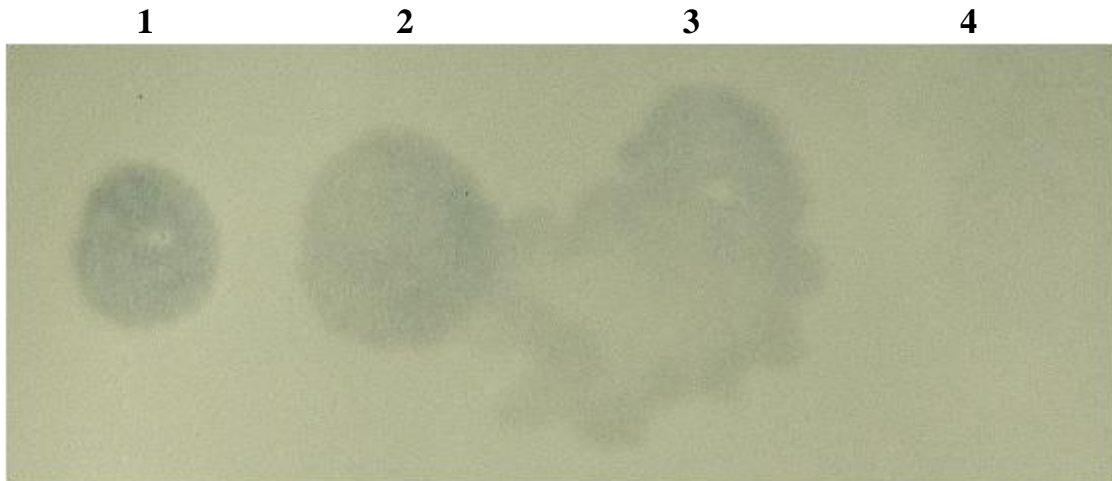
- (1) Marcador de massa molecular (em kDa)
- (2) Extrato pETLP001ORF27 BL21 Rosetta Fração solúvel sem detergente
- (3) Extrato pETLP001ORF27 BL21 Rosetta Fração insolúvel sem detergente
- (4) Extrato pETLP001ORF27 BL21 Rosetta Fração solúvel com detergente
- (5) Extrato pETLP001ORF27 BL21 Rosetta Fração insolúvel com detergente
- (6) Extrato pETLP090ORF24 BL21 Rosetta Fração solúvel sem detergente
- (7) Extrato pETLP090ORF24 BL21 Rosetta Fração insolúvel sem detergente
- (8) Extrato pETLP090ORF24 BL21 Rosetta Fração solúvel com detergente
- (9) Extrato pETLP090ORF24 BL21 Rosetta Fração insolúvel com detergente

FIGURA 46: Teste de solubilidade das proteínas LP018ORF15 e LP018ORF16

Eletroforese em gel de SDS poliacrilamida 12% (SDS-PAGE) das frações solúvel e insolúvel, tratados e não tratados com 0,5% de TritonX-100, dos extratos celulares de *E. coli* Rosetta expressando as proteínas LP018ORF15 e LP018ORF16. Indução com 0,5 mM de IPTG, por 14 horas a 18°C e 120 rpm. As proteínas foram coradas com Coomassie Blue R-250. As setas indicam a banda correspondente às proteínas superexpressas.

- (1) Marcador de massa molecular (em kDa)
- (2) Extrato pETLP018ORF15 BL21 Rosetta Fração solúvel sem detergente
- (3) Extrato pETLP018ORF15 BL21 Rosetta Fração insolúvel sem detergente
- (4) Extrato pETLP018ORF15 BL21 Rosetta Fração solúvel com detergente
- (5) Extrato pETLP018ORF15 BL21 Rosetta Fração insolúvel com detergente
- (6) Extrato pETLP018ORF16 BL21 Rosetta Fração solúvel sem detergente
- (7) Extrato pETLP018ORF16 Rosetta Fração insolúvel sem detergente
- (8) Extrato pETLP018ORF16 Rosetta Fração solúvel com detergente
- (9) Extrato pETLP018ORF16 BL21 Rosetta Fração insolúvel com detergente

FIGURA 47: Teste das frações solúvel e insolúvel da proteína LP001ORF27 em meio seletivo.

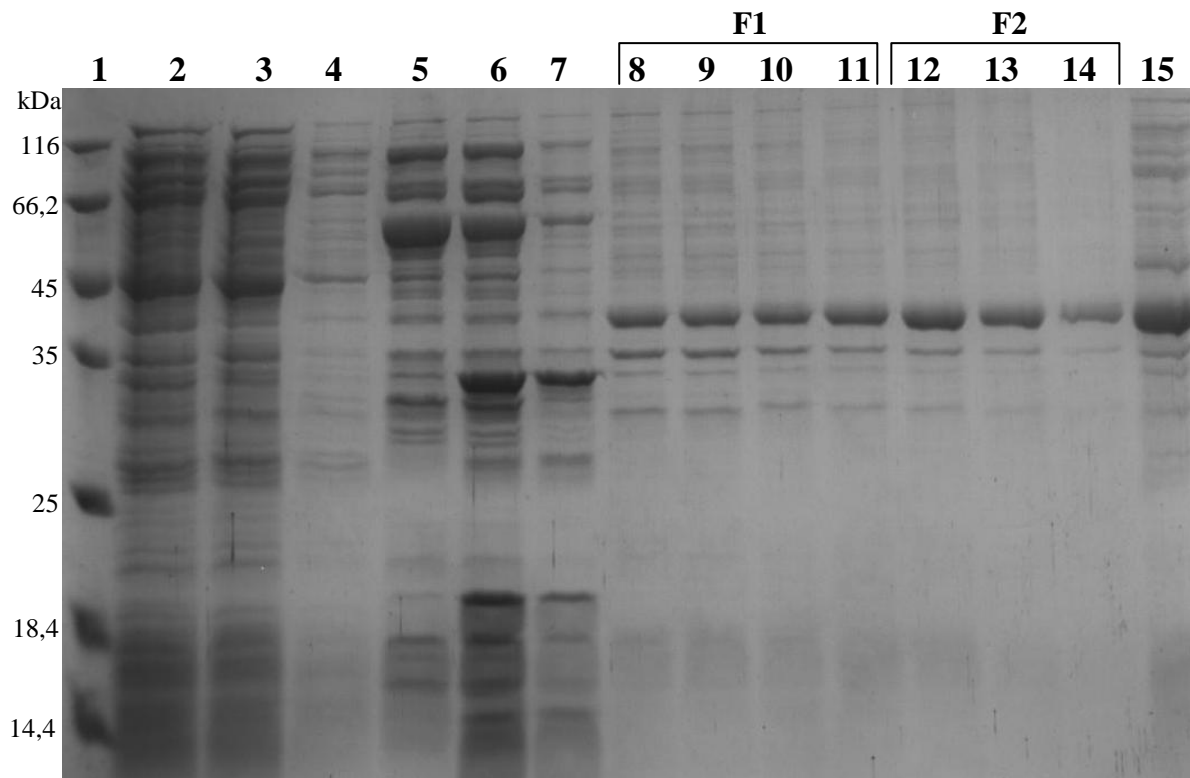


Placa de agar-tributirina 1% com halo de hidrólise produzido pelos extratos celulares de *E. coli* BL21 contendo LP001ORF27. As células foram sonicadas em tampão de lise na presença e ausência de Triton-X 100 0,5%. As frações solúveis e insolúveis em cada tratamento foram separadas por centrifugação. Uma alíquota de 5 uL de cada fração foi depositada em uma placa de meio sólido agar-tributirina para avaliação da atividade lipolítica. A placa foi incubada em estufa a 37°C durante 24 horas.

- (1) Fração solúvel sem detergente
- (2) Fração insolúvel sem detergente
- (3) Fração solúvel com detergente
- (4) Fração insolúvel com detergente

5.4.3 Purificação da proteína LP001ORF27

A proteína LP001ORF27 foi purificada pelo método cromatográfico de afinidade em coluna de níquel, através da interação entre os resíduos de histidina (His-tag) adicionados na região N-terminal da proteína e os íons metálicos de níquel da coluna. A proteína foi eluída em tampão com concentrações crescentes de imidazol, com o pico de eluição a 502 mM. Foram coletadas 24 frações protéicas de 1 ml, sendo que a proteína LP001ORF27 começou a ser eluída a partir da fração 10 (Figura 48). As alíquotas que apresentavam mais bandas contaminantes (10-17) foram reunidas na fração F1, enquanto as mais puras foram reunidas na fração F2 (18-24). As duas frações foram dialisadas em tampão contendo glicerol 50% na ausência de imidazol. Após a diálise a concentração das proteínas totais foi determinada, obtendo-se 1,06 mg/mL para F1 e 0,60 mg/mL para F2. Alíquotas de 10, 5, 2,5 e 1 μ L das frações dialisadas foram aplicadas em gel desnaturante SDS-PAGE para o cálculo de pureza da proteína (Figura 49). A pureza da proteína foi calculada, utilizando o software LabWorks, em aproximadamente 90%. As duas frações finais obtidas foram testadas em meio sólido agar-tricaprilina para avaliação da atividade. Tanto a fração F1 quanto a fração F2 foram capazes de hidrolisar a tricapriline, indicando que a enzima LP001ORF27 estava na forma ativa (Figura 50). A diferença entre os halos produzidos pelas duas frações pode estar relacionada com a quantidade de proteína em cada fração, já que em F1 há quase o dobro de proteína que em F2.

FIGURA 48: Purificação da proteína LP001ORF27.

Eletroforese em gel de SDS poliacrilamida 12% (SDS-PAGE) das frações protéicas obtidas da purificação de LP001ORF27 em coluna HisTrap carregada com níquel e eluídas em concentrações crescentes de imidazol

(1) Marcador de massa molecular (em kDa)

(2) LP001ORF27 fração solúvel

(3) LP001ORF27 injeção

(4) LP001ORF27 lavagem

(5) LP001ORF27 fração 1

(6) LP001ORF27 fração 3

(7) LP001ORF27 fração 5

(8) LP001ORF27 fração 10

(9) LP001ORF27 fração 12

(10) LP001ORF27 fração 14

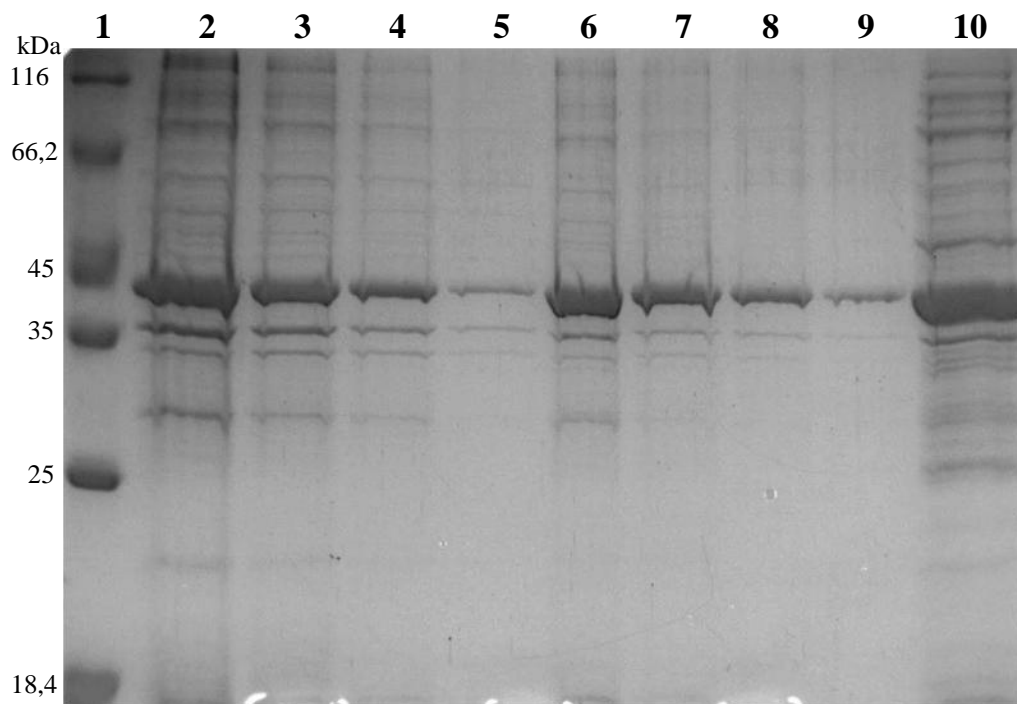
(11) LP001ORF27 fração 16

(12) LP001ORF27 fração 18

(13) LP001ORF27 fração 20

(14) LP001ORF27 fração 22

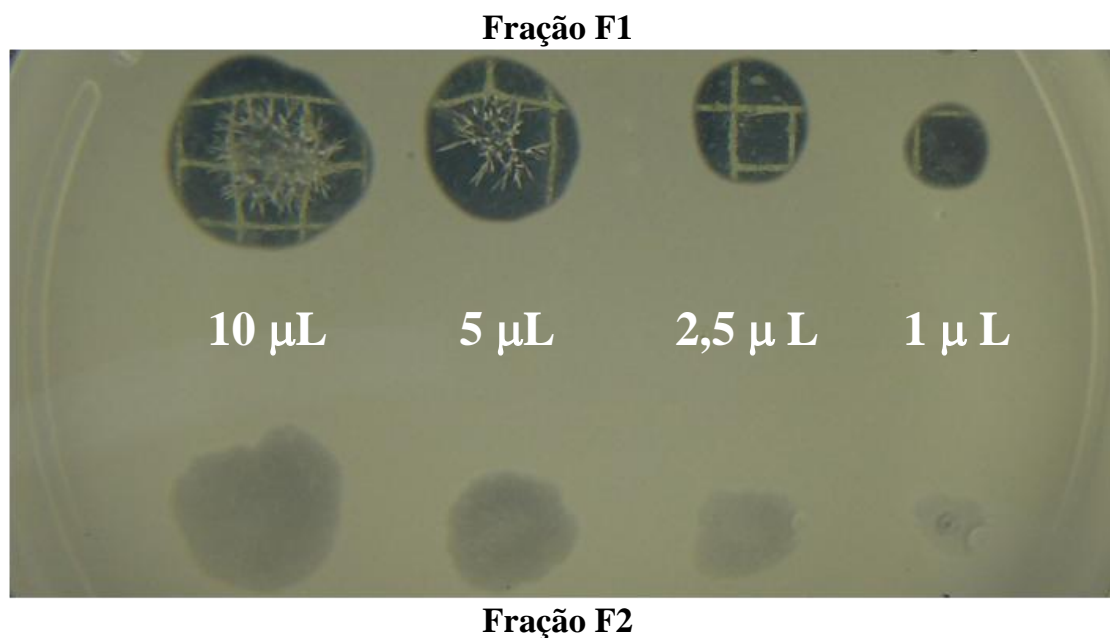
(15) Controle positivo – extrato celular bruto de *E. coli* Rosetta expressando LP001.

FIGURA 49: Diálise das frações F1 e F2 da purificação de LP001ORF27

Eletroforese em gel de SDS poliacrilamida 12% (SDS-PAGE) das frações F1 e F2 de LP001ORF27 após a diálise. Foram aplicadas alíquotas de 10, 5, 2,5 e 1 μL das duas frações. O cálculo da pureza foi feito com base em aproximadamente 1 μg de proteína (coluna 5 para F1 e coluna 8 para F2).

- (1) Marcador de massa molecular (em kDa)
- (2) LP001ORF27 fração F1 10 μL (10 μg de proteína total)
- (3) LP001ORF27 fração F1 5 μL (5 μg de proteína total)
- (4) LP001ORF27 fração F1 2,5 μL (2,5 μg de proteína total)
- (5) LP001ORF27 fração F1 1 μL (1 μg de proteína total)
- (6) LP001ORF27 fração F2 10 μL (6 μg de proteína total)
- (7) LP001ORF27 fração F2 5 μL (3 μg de proteína total)
- (8) LP001ORF27 fração F2 2,5 μL (1,5 μg de proteína total)
- (9) LP001ORF27 fração F2 1 μL (0,6 μg de proteína total)
- (10) Controle positivo – extrato celular bruto de *E. coli* Rosetta expressando LP001ORF27.

FIGURA 50: Teste das frações F1 e F2 de LP001ORF27 em meio seletivo agar-tricaprilina 1%



Halo de hidrólise produzido pela enzima LP001ORF27 purificada. Volumes de 10, 5, 2,5 e 1 µl das frações F1 e F2 foram pipetados na placa, contendo substrato tricaprilina, para avaliação da atividade. No caso da fração F1 a hidrólise foi total, produzindo um halo transparente.

5.4.4 Caracterização da lipase LP001ORF27

A caracterização enzimática da lipase LP001ORF27 foi realizada utilizando como substrato pNP butirato (exceto no teste de especificidade de substrato, no qual vários substratos foram utilizados) à temperatura constante de 30°C (exceto teste de temperatura ótima). A hidrólise da ligação éster entre a cadeia acila e o p-nitro fenol libera fenolato, que em pH neutro a básico possui cor amarela, e pode ser medida pelo aumento da absorbância a 410 nm. A atividade específica foi medida em U/mg de proteína, sendo que uma unidade é a quantidade de enzima necessária para produzir 1 μmol de produto por minuto de reação. A atividade enzimática foi determinada em modo cinético por regressão linear da reta gerada a partir do aumento da absorbância proporcionada pela hidrólise do substrato.

A especificidade quanto ao substrato foi determinada utilizando substratos artificiais pNP monoacilésters com cadeias acila compostas de 2, 4, 6, 10, 12, 14, 16 e 18 átomos de carbono. A enzima LP001ORF27 foi capaz de hidrolisar todos os substratos testados apresentando maior atividade contra pNP decanoato (C10) (Figura 51A). A atividade relativa contra pNP palmitato (C16) foi de 30%, a mesma apresentada para pNP butirato (C4). A atividade relativa contra pNP oleato (C18) representou 20% da atividade máxima em pNP decanoato. A atividade enzimática contra substratos naturais triacilgliceróis também foi avaliada. A enzima LP001ORF27 também foi capaz de hidrolisar todos os substratos testados, mas, nesse caso, a maior atividade foi observada contra tributirina (C4) (Figura 51B). As atividades relativas contra tricaprilina (C8) e trioleína (C18) foram de 70 e 10% respectivamente sugerindo que LP001ORF27 é uma lipase verdadeira.

O pH ótimo de atividade foi determinado em tampões com valores de pH variando de 3,0 a 11,0. A enzima apresentou a maior atividade em pH 7,0 (Figura 52). A partir desse ponto, a atividade rapidamente diminuiu em pH ácidos com 40% da atividade em pH 5,0 e nenhuma atividade detectada em pH 4,0 e 3,0. Em pH básicos a inativação da enzima foi mais lenta. A enzima apresentou 72, 60, e 46% de atividade,

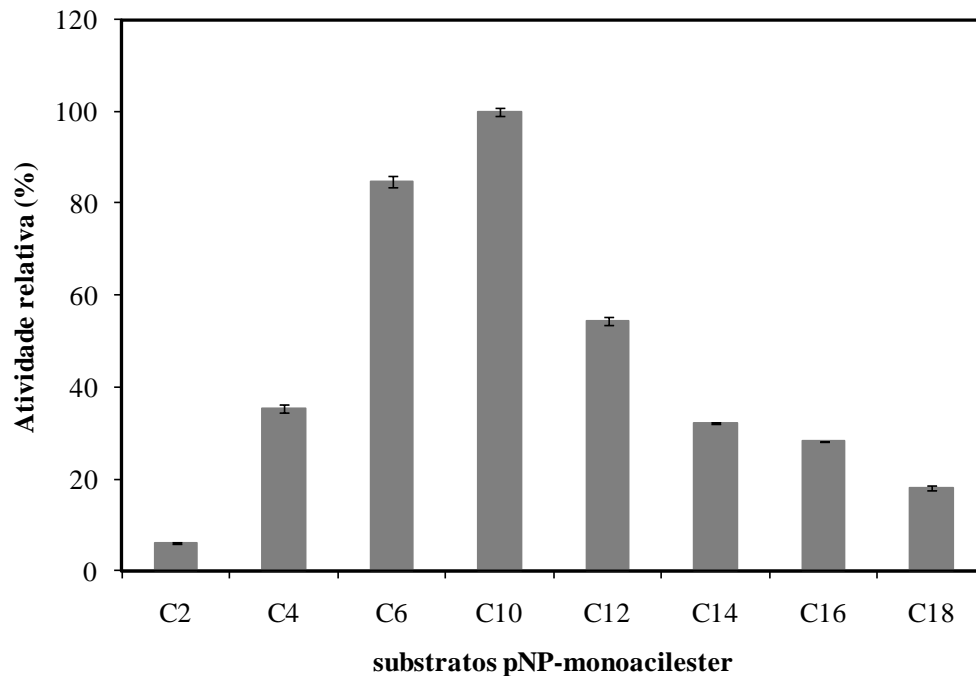
em relação àquela obtida no pH 7,0, nos valores de pH 8,0, 9,0 e 10,0, respectivamente (Figura 60), sugerindo uma tendência moderadamente alcalofílica.

A influência de íons metálicos na atividade enzimática foi avaliada nas condições de 1 e 10 mM. Em ambas as condições não houve um aumento efetivo da atividade de LP001ORF27 na presença de íons metálicos, apenas um pequeno aumento de 6 e 7% na presença de Na^+ e Hg^+ , respectivamente (Figura 53). Esse resultado foi reforçado pela pequena diminuição na atividade na presença de quelantes de metais EDTA e EGTA tanto em 1 quanto em 10 mM (Figura 61). Li^+ , K^+ , Cs^+ e Rb^+ não afetaram a enzima nas condições testadas. Zn^{+2} e Fe^{+3} também não tiveram efeito na concentração de 1 mM. O cálcio, geralmente envolvido na estabilização de lipases e aumento da atividade, reduziu a atividade enzimática em 50% em ambas as condições. Mg^{+2} , Mn^{+2} , Cd^{+2} , Fe^{+2} , Ni^+ , Cu^{+2} , Co^{+2} , Ag^+ e Sn^{+2} também reduziram a atividade enzimática. A atividade enzimática também foi avaliada na presença de 0,1 e 1% (v/v) de detergente. Detergentes aniônicos (N-LS, CTAB e SDS) inibiram fortemente a atividade de LP001ORF27 em ambas as condições testadas (Figura 54). Entretanto, detergentes não iônicos (Triton-X 100, Tween 20, Tween 40, Tween 80 e NP-40) reduziram a atividade da enzima para cerca de 70-80% a 0,1% de detergentes e para cerca de 20% a 1% de detergentes (Figura 54).

O efeito da temperatura na atividade enzimática foi avaliado de 20 a 70°C. LP001ORF27 apresentou a maior atividade a 50°C (Figura 55A) sem uma diminuição efetiva na atividade a 60°C (99,4% em relação a 50°C). A 40 e 70°C a atividade relativa diminuiu para 93 e 70%, respectivamente. A 20 e 30°C a atividade relativa diminuiu para 37 e 48%, respectivamente. Para avaliar a termoestabilidade, a enzima foi pré-incubada a 50, 60 e 70°C por 1 hora e a atividade residual foi determinada em modo cinético durante 10 minutos a 30°C. LP001ORF27 reteve 53 e 33% da atividade inicial após a incubação a 60 e 70°C, respectivamente. Entretanto, após a incubação a 50°C a enzima apresentou um aumento de 80% em sua atividade em relação ao controle não incubado (Figura 55B).

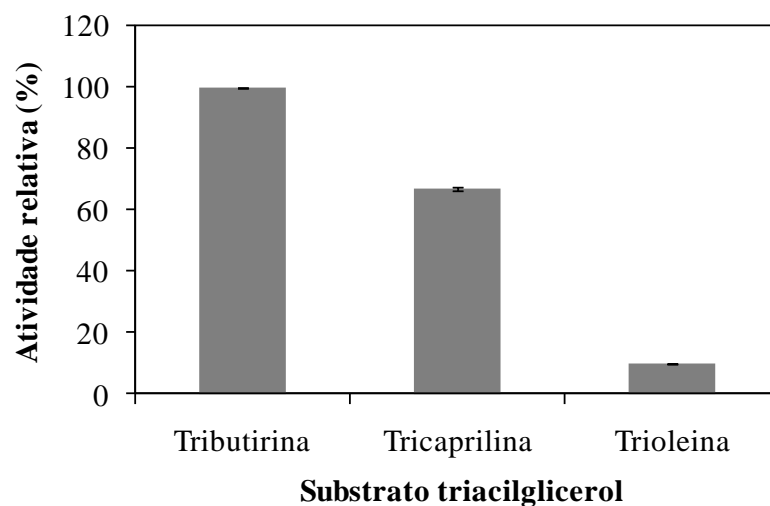
FIGURA 51: Especificidade de substrato da lipase LP001ORF27

(A) Atividade lipolítica contra diferentes substratos artificiais pNP monoacilesteres.



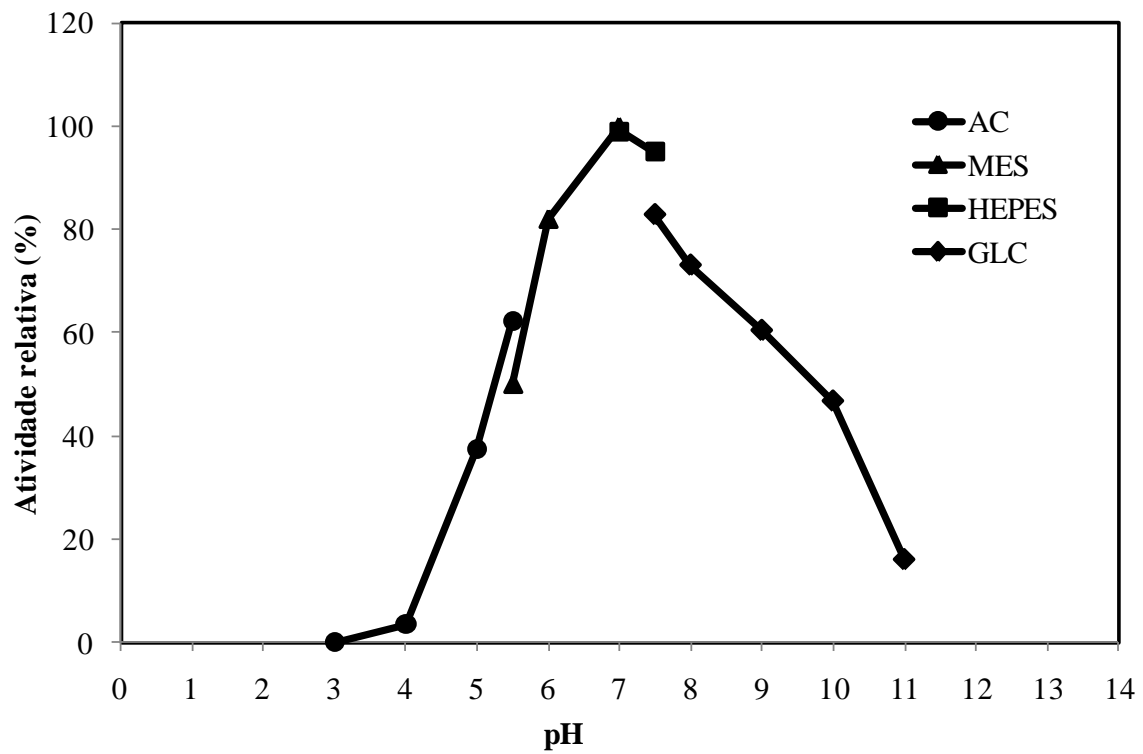
O tamanho da cadeia acil está expressa em número de átomos de carbono C. A atividade máxima contra pNP-C10 foi de 9 U/mg de proteína

(B) Atividade lipolítica contra substratos naturais triacilgliceróis.



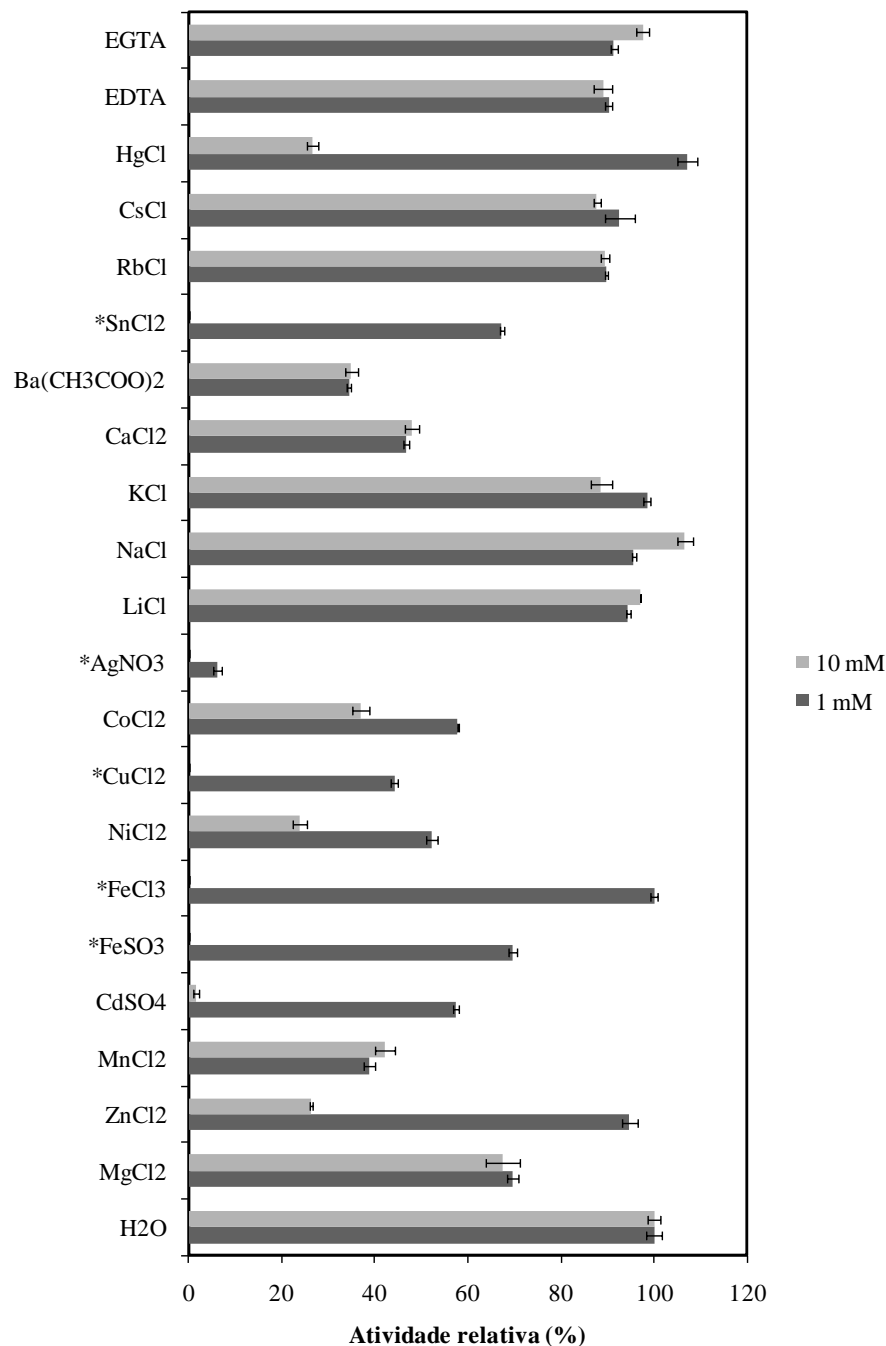
Os dados representam a média de 3 experimentos independentes. A atividade máxima contra tributirina foi de 29,2 U/mg de proteína.

FIGURA 52: Efeito do pH sobre a atividade de LP001ORF27

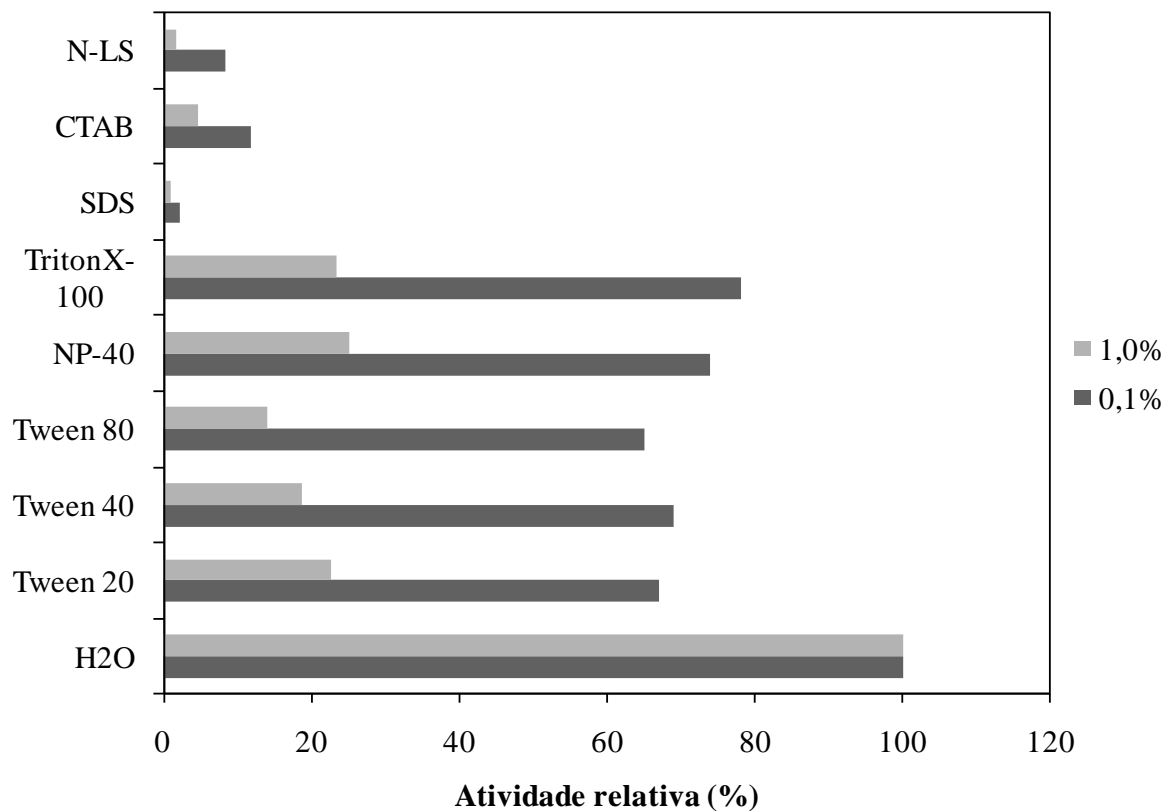


A atividade enzimática foi determinada em diferentes soluções tampão com diferentes valores de pH utilizando pNP butirato como substrato e 50 nM de enzima (monômero) a 30°C durante 30 minutos. AC – tampão acetato (pH 3,0 a 5,5); MES – tampão ácido 2-N-[morfolino] etano sulfônico (pH 5,5 a 7,0); HEPES – tampão HEPES (pH 7,0 e 7,5); GLC – tampão glicina (pH 7,5 a 11). Os dados representam a média de 3 experimentos independentes.

FIGURA 53: Efeito de íons e quelantes de metal sobre a atividade da lipase LP001ORF27



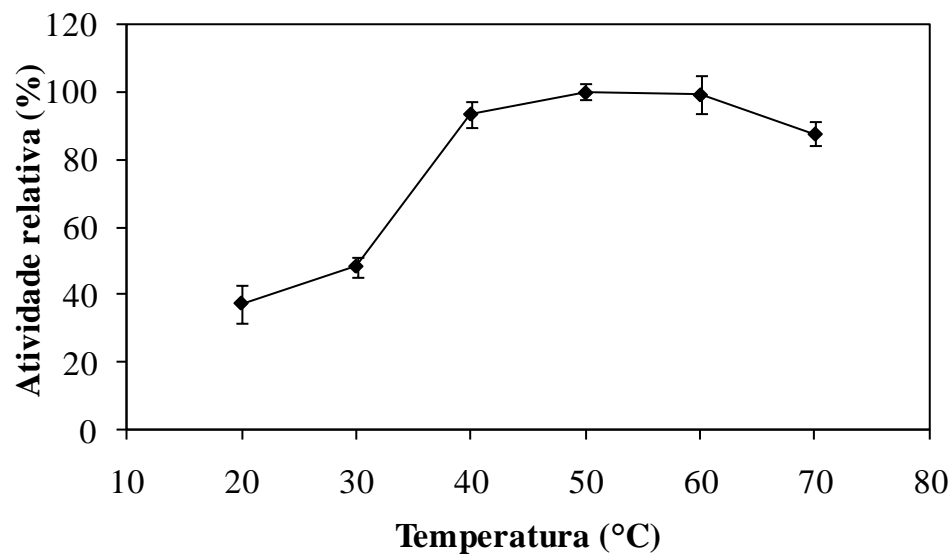
A atividade enzimática foi determinada nas concentrações de 1 e 10 mM de íons metálicos e quelantes de metais usando pNP butirato como substrato e 50 nM de enzima (monômero) a 30°C durante 30 minutos. As atividades relativas foram calculadas em comparação com a atividade controle, no qual o elemento analisado foi substituído por água. Os dados representam a média de 3 experimentos independentes. * Não determinado devido à precipitação do sal.

FIGURA 54: Efeito de detergentes sobre a atividade da lipase LP001ORF27

A atividade enzimática foi determinada nas concentrações 0,1 e 1% de detergentes usando pNP butirato como substrato e 50 nM de enzima a 30°C durante 30 minutos. As atividades relativas foram calculadas em comparação com a atividade controle no qual elemento analisado foi substituído por água. Os dados representam a média de 3 experimentos independentes.

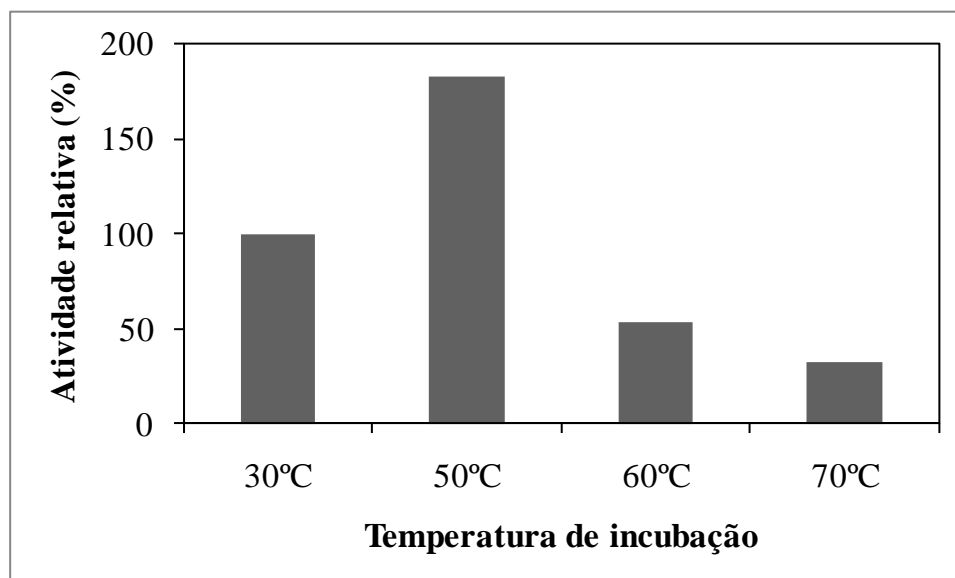
FIGURA 55: Efeito da temperatura sobre a atividade da lipase LP001ORF27

(A) Determinação da temperatura ótima.



A atividade enzimática foi determinada em diferentes temperaturas durante 5 minutos utilizando-se pNP butirato como substrato. Os dados representam a média de dois experimentos independentes.

(B) Avaliação da termoestabilidade..



A atividade enzimática foi determinada a 30°C após uma hora de incubação a 50, 60 e 70°C utilizando-se pNP butirato como substrato. As atividades relativas foram calculadas em comparação com a atividade controle a 30°C sem incubação da enzima (100%). Os dados representam a média de dois experimentos independentes.

CAPÍTULO 6

DISCUSSÃO

6. DISCUSSÃO

6.1 Análise das comunidades bacterianas do solo da Floresta Atlântica Paranaense

6.2 Bibliotecas metagenômicas

6.3 Caracterização genética dos fosmídeos

6.4 Clone MAF1LP001

6.5 Clone MAF1LP018

6.6 Clone MAF1LP090

6.7 Considerações finais

6. DISCUSSÃO

6.1 Análises das comunidades bacterianas do solo da Floresta Atlântica Paranaense

Os solos da Floresta Atlântica paranaense possuem uma vasta diversidade de espécies bacterianas representada, principalmente, pelos filos *Acidobacteria* e *Proteobacteria* (FAORO, 2006). A análise realizada com os programas S-LibShuff, LibShuff e DOTUR sugerem que é possível dividir a diversidade em dois principais grupos: alta altitude/alta diversidade (entre 900 e 800 m) e baixa altitude/baixa diversidade (entre 160 e 30 m). Essa divisão em dois grupos é coerente com a distribuição da vegetação na Serra do Mar, onde as amostras de solo foram coletadas. A Floresta Atlântica é classificada como uma floresta ombrófila densa e dividida em terras costeiras (0 a 50 metros de altitude), sub-montana (50 a 500 metros de altitude), montana (500 a 1.200 metros de altitude) e alta-montana (1.200 a 1.400), criando um gradiente de vegetação que vai desde arbustos nas terras costeiras até a floresta de montana bem desenvolvida, com árvores grandes e robustas (CÂMARA, 2005). Essa distribuição na vegetação deve estar influenciando diretamente a biodiversidade bacteriana. Apesar do domínio dos filos *Acidobacteria* e *Proteobacteria*, a comparação entre os grupos de alta e baixa altitude realizada pelo programa LibShuff indica que essas duas comunidades são diferentes considerando distâncias evolutivas de 0 (nível de espécie) até 0,3 (nível de filo) e sugere uma grande variação de espécies dentro das linhagens dominantes e entre os dois grupos. Além da vegetação é possível que os fatores físico-químicos do solo nessas duas regiões também estejam atuando na diversidade e, principalmente, na variação dentro das linhagens dominantes (YOUSSEF e ELSHAHED, 2009). Esses resultados (Anexo 7) permitiram a seleção de três amostras de solo para construção das bibliotecas metagenômicas.

6.2 Bibliotecas metagenômica

A partir das análises da estrutura da comunidade bacteriana foi possível selecionar três amostras de solo para purificação de DNA e construção das bibliotecas metagenômicas, denominadas MAF1, MAF2 e MAF3. A purificação do DNA foi realizada pelo método direto utilizando esferas de sílica sem a etapa de pré-lavagem do solo. A comparação entre os métodos direto e indireto por análise de diversidade do gene 16S rRNA em uma mesma amostra de solo mostrou que o DNA obtido pelo método direto apresenta maior diversidade (KOZDRÓJ e van ELSAS, 2001). Além disso, o método direto atinge mais células bacterianas, aumentando o rendimento de DNA na purificação. Não foram observados grandes níveis de contaminação por ácidos húmicos no DNA purificado, um problema corriqueiro quando se utiliza o método direto. Ao final da purificação, a solução de DNA estava translúcida, indicando que o kit para a purificação de DNA do solo utilizado foi eficiente. Outro problema comum quando se utiliza o método direto é a intensa fragmentação do DNA. De fato, o DNA purificado apresentava fragmentos inferiores a 250 pb, mas ainda assim a maior concentração de DNA estava na faixa acima de 20 kb, o que nos permitiu recuperar os fragmentos em torno de 40 kb para a construção das bibliotecas.

As bibliotecas metagenômicas construídas possuem ao todo 100.128 clones, com uma média de inserto de 35 kb, gerando um total aproximado de 3,5 bilhões de bases. Considerando-se o tamanho médio do genoma de uma bactéria do solo proposto por RAES e colaboradores (2007) em 4 milhões de bases, é possível estimar que as bibliotecas construídas representem cerca de 900 genomas bacterianos. Comparado à estratégia de clonagem de fragmentos relativamente pequenos (2 a 5 kb), a estratégia utilizada nesse trabalho é superior, já que permite identificar fenótipos que dependam da expressão de genes em um operon. Além disso, as dificuldades antes associadas a esse método vêm sendo superadas pelo desenvolvimento de kits específicos para a construção de bibliotecas de insertos grandes e também pelo desenvolvimento de novos vetores de clonagem com número de cópias induzível, tornando possível aumentar a expressão dos genes transcritos, a partir do inserto de DNA clonado,

facilitando a identificação do fenótipo de interesse (WILD, HRADECNA E SZYBALSKI, 2002).

O resultado da prospecção das bibliotecas metagenômicas por biocatalizadores identificou 315 esterases/lipases, possivelmente 3 lipases verdadeiras, 460 proteases e 4 amilases. Considerando-se o total de bases obtidas nas 3 bibliotecas (~3,5 bilhões de bases), uma enzima esterase/lipase foi identificada a cada 11,12 Mb ou, aproximadamente, 1 enzima a cada três genomas bacterianos. A taxa de identificação de clones positivos pode parecer baixa a primeira vista, mas foi maior do que a obtida em outros trabalhos de prospecção metagenômicas em bibliotecas de solo, que também utilizaram o fosmídeo como vetor e tributirina, como meio seletivo com 1 clone a cada 17, 18 ou 35 Mb (LORENZ e ECK, 2005). No caso das amilases essa taxa de sucesso é ainda mais baixa (1 enzima a cada 302 Mb). A maior taxa de identificação de clones foi obtida para as proteases (1 enzima a cada 2,6 Mb). No entanto, a prospecção por proteases foi realizada em leite e não em caseína pura e é possível que, dentre os 460 clones, muitos não sejam proteases e representem outra classe de hidrolases, como as glicosil hidrolases (JONES, SUN e MARCHESI, 2007). Esterases, lipases, amilases e proteases são amplamente distribuídas entre os organismos vivos e essa baixa taxa de detecção pode estar refletindo a incompatibilidade entre o sistema de expressão célula hospedeira e o DNA metagenômico. Dificuldades como esta tem levado à utilização de diferentes espécies de bactérias, que não *E. coli*, como células hospedeiras, como *Streptomyces sp.* por exemplo (WANG et al, 2000).

6.3 Caracterização genética dos insertos de DNA clonados em fosmídeos dos clones de interesse

Os insertos de DNA dos clones MAF1LP001 e MAF1LP018 foram integralmente sequenciados. Já o inserto de DNA do clone MAF1LP090 foi parcialmente sequenciado. Considerando os insertos de DNA dos 3 clones, aproximadamente 100 mil pares de bases foram sequenciados e 80 diferentes ORFs foram identificadas. A grande maioria das ORFs identificadas apresentou similaridade

com sequências de proteínas de *S. usitatus* e *K. versatilis*, ambas pertencente ao grupo *Acidobacteria*. Mas o baixo grau de identidade obtido entre as sequências de aminoácidos das ORFs identificadas e aquelas depositadas no banco de dados sugere que os fragmentos de DNA clonados em MAF1LP001, MAF1LP018 e MAF1LP090 provavelmente pertençam a um organismo do filo *Acidobacteria*, mas não à espécie *S. usitatus* e *K. versatilis* ou mesmo aos gêneros *Solibacter* e *Koribacter*. A maioria das sequências de DNA do grupo *Acidobacteria* (genes ribossomais ou do metabolismo geral) provêm de amostras ambientais e são amplamente distribuídas por todos os continentes: Ásia (MITSUI et al., 1997), Austrália (LIESACK e STACKEBRANDT 1992), Américas (BORNEMAN et al., 1996; BORNEMAN e TRIPLETT, 1997) e Europa (FELSKE et al., 1998). Esses resultados estão de acordo com aqueles obtidos anteriormente para as sequências do gene 16S rRNA, que mostravam a dominância do filo *Acidobacteria*, seguido de *Proteobacteria*, nas amostras de solo utilizadas para a construção da biblioteca (FAORO, 2006)

O recente sequenciamento do genoma de 3 organismos do filo *Acidobacteria*, um cultivável (*Acidobacterium capsulatum*) e dois não cultiváveis (justamente *Solibacter usitatus* e *Koribacter versatilis*), revelou que esse grupo compartilha características genômicas com os grupos *Proteobacteria*, *Cyanobacteria* e também algumas características presentes em fungos (WARD et al, 2009). A similaridade filogenética entre *Proteobacteria* e *Acidobacteria* pode estar contribuindo para a correta expressão gênica e tradução das proteínas identificadas na fase de prospecção, permitindo a detecção dos fenótipos.

6.4 Clone MAF1LP001

O DNA inserto do clone MAF1LP001 foi integralmente sequenciado, revelando a presença de 33 ORFs, sendo que 60% delas apresentaram similaridade com proteínas de *S. usitatus* Ellin6076. Como mencionado anteriormente, é possível que esse fragmento de DNA pertença a um organismo do filo *Acidobacteria*, mas improvável que seja de *S. usitatus*. Além da baixa identidade entre as sequências de

aminoácidos, outra evidência que corrobora essa hipótese é a presença do gene que codifica a proteína plastoquinol-plastocianina redutase em MAF1LP001 (ORF8 na Tabela 13). Uma busca no genoma de *S. usitatus* (disponível em http://genome.jgi-psf.org/cgi-bin/runAlignment?db=aci_e&advanced=1) pelo gene que codifica a proteína ORF8 não encontrou nenhum gene similar. Em organismos fotossintéticos, essa proteína faz parte de um complexo proteico maior, denominado *b6f*, que realiza a transferência de elétrons entre os fotossistemas II e I, com bombeamento de prótons acoplado (BANIULIS, 2008). Sua presença no inserto de MAF1LP001 em um operon com proteínas envolvidas na cadeia transportadora de elétrons, possivelmente as outras proteínas que compõe o complexo *b6f* (ORFs 4 a 9 na Tabela 13 e Figura 30), sugere que o organismo original possa ser uma acidobactéria fotossintética que realiza fotossíntese oxigênica. Até recentemente, apenas 5 filos bacterianos eram reconhecidamente capazes de realizar fotossíntese oxigênica: *Cyanobacteria*, *Proteobacteria*, *Chloroflexi*, *Chlorobi* e *Firmicutes*. Entretanto, análises metagenômicas de uma comunidade bacteriana associada a uma fonte termal do parque Yellowstone identificaram a primeira acidobactéria fotossintética (BRYANT et al, 2009), levando os autores a questionar se essa característica era comum dentre outros representantes do filo *Acidobacteria*. Considerando os resultados obtidos para o clone MAF1LP001, é possível que também existam acidobactérias fotossintéticas no solo da Floresta Atlântica.

A comparação de sequências mostrou que a ORF17 possui um domínio cetoacil sintase. Essas enzimas são responsáveis pela síntese de policetídeos, através de uma condensação descarboxilativa sequencial de precursores de acil-CoA, sendo a ligação carbono-carbono realizada pelo domínio Cetoacil Sintase (KS). Dentro do grupo dos policetídeos estão moléculas como o imunossupressor Rapamicina, o antibiótico Eritromicina A, o anticancerígeno Epotilona B e o antiolesterolêmico Lovastatina. Atualmente existem três tipos de policetídeo sintase (PKS) descritas. As PKSs tipo 1 são enzimas multifuncionais organizadas em módulos sendo que cada módulo possui um conjunto de atividades responsáveis por um ciclo de alongação na cadeia do policetídeo. As PKSs tipo 2 são multienzimas que possuem apenas um

modulo. Por fim, as PKSs tipo 3 são enzimas homodiméricas que apresentam apenas o domínio cetoacil sintase, não agrupadas em módulos, atuando, principalmente, como enzimas condensadoras (SHEN, 2003). Tendo em vista essa classificação, é possível que a ORF17 codifique uma PKS tipo 3.

Através da análise de sequência e pelos resultados obtidos com mutagênese por inserção de transposon foi possível identificar LP001ORF27 como o gene codificante da lipase em MAF1LP001. Diferentemente de todas as outras proteínas codificadas no inserto de MAF1LP001, a ORF27 apresentou similaridade com uma proteína de fungo (Tabela 13). Análises filogenéticas mostraram que de fato LP001ORF27 está mais próxima das proteínas fúngicas do que bacterianas. Essa relação entre proteínas de acidobactérias e proteínas fúngicas já foi abordada anteriormente e, também, já foi observada para enzimas do grupo glicosídeo hidrolase. Nesse caso, os autores atribuíram essa similaridade a um evento de transferência lateral de genes muito antigo (WARD et al, 2009).

O pentapeptídeo conservado de LP001ORF27 é similar ao encontrado em lipases de organismos do gênero *Bacillus*, no qual o primeiro aminoácido glicina foi substituído por alanina. LP001ORF27 possui maior atividade contra o substrato artificial pNP decanoato (C10) e o substrato natural triacilbutirato. No entanto, ela também apresenta atividade contra substratos contendo cadeias acila maiores (C16 e C18) e também contra trioleína. Segundo a definição de lipase proposta por VERGER (1997) e JAEGER, DJKSTRA e REETZ (1999) LP001ORF27 pode ser classificada como uma lipase verdadeira. Uma característica comum em lipases é a dependência de cálcio para sua atividade e estabilidade (GUPTA, GUPTA e RATHI, 2004). No entanto, esse não parece ser o caso para LP001ORF27. A presença de cálcio diminuiu a atividade da lipase em aproximadamente 50%. De modo contrário, a presença de 1 mM de mercúrio, reconhecidamente um forte inibidor da atividade enzimática assim como outros metais pesados, foi capaz de elevar sutilmente a atividade de LP001ORF27, mas por que isso ocorre ainda precisa ser investigado. A temperatura ótima de atividade para LP002ORF27 ficou na faixa entre 50-60°C. Diferentes trabalhos de caracterização enzimática de lipases classificam as enzimas com atividade

ótima nessa faixa de temperatura como termofílicas (LI e ZHANG, 2005; MEILLEUR et al, 2009; AHMED, RAGHAVENDRA e MADAWAR 2009; DUTTA e RAY, 2009). A lipase isolada do organismo termofílico *Bacillus stearothermophilus* (KIM et al, 1998) apresenta um perfil de atividade semelhante a LP001ORF27: maior atividade contra substratos de cadeia acila de 8 carbonos (pNP caprilato) e atividade ótima a 60°C. Desse modo, também é possível classificar LP001ORF27 como moderadamente termofílica. Quando a enzima foi pré-incubada em diferentes temperaturas antes da determinação da atividade ainda foi possível medir uma atividade residual de 53% após incubação a 60°C e de 33% após incubação a 70°C. Entretanto, após uma hora a 50°C a enzima não teve perda de atividade, ao invés disso, apresentou um aumento de 80% na sua atividade. A ativação térmica é uma característica incomum em lipases não imobilizadas em substratos sólidos. Resultados semelhantes foram obtidos para uma lipase da archaea *Sulfolobus shibatea*, a qual apresentou uma ativação térmica de 60% após 10-25 minutos a 90°C (HUDDLESTON, YALLOP e CHARALAMBOUS, 1995), e para uma lipase de *Burkholderia cepacia*, a qual apresentou ativação de 30% após 2 horas a 90°C (RATHI et al, 2000). Ainda, as análises filogenéticas indicaram que LP001ORF27 esta próxima da família I.5 formada exclusivamente por lipases termofílicas. Analisados conjuntamente, esses resultados indicam que LP001ORF27 é uma lipase termofílica de *Acidobacteria*.

6.5 Clone MAF1LP018

A proteína codificada pela LP018ORF15 apresenta domínios característicos de uma serina protease, enquanto a proteína codificada pela LP018ORF16 foi classificada como hipotética e não apresenta similaridade de sequência com nenhuma lipase conhecida. Sequências de proteínas similares à LP018ORF16 já estão disponíveis nos banco de dados, mas todas são classificadas como hipotéticas. No entanto, foi possível identificar a presença de um domínio $\alpha\beta$ -hidrolase pouco conservado (E-value 0,01). Através do alinhamento das sequências de aminoácidos de ORF16 com as proteínas hipotéticas similares disponíveis nos banco de dados também foi possível identificar o

motivo contendo o pentapeptídeo conservado característico de lipases G-X-S-X-G. A análise filogenética baseada nas famílias de lipases propostas por ARPIGNY e JAEGER (1999) mostrou que essa enzima está distante de todas as outras famílias conhecidas. Testes pré-eliminatórios indicaram que LP018ORF16 é capaz de hidrolisar substratos artificiais pNP monoacil ésteres, mas com baixa atividade (dados não mostrados) devido provavelmente à falta da chaperona. Com base nesses dados é possível inferir que todo esse grupo de proteínas hipotéticas são, de fato, proteínas com função lipolítica e que representam uma nova família na classificação das lipases. Através da metagenômica, a classificação das lipases em famílias proposta por ARPIGNY e JAEGER (1999) vem recebendo acréscimos. LEE e colaboradores (2006) identificaram uma nova família, denominada LipG, a partir de bibliotecas metagenômicas de sedimentos. As lipases que constituem essa família apresentam uma fenda oxianion formada por um resíduo de arginina seguido de um resíduo de glicina, uma característica de enzimas fúngicas. Recentemente, KIM e colaboradores (2009) também propuseram uma nova família de lipases que apresenta atividade a baixas temperaturas e pH alcalino.

O clone MAFLP018 foi selecionado devido a sua capacidade de hidrolisar trioleína, indicando a presença de um gene que codifica uma lipase no inserto de DNA clonado. A mutagenese por transposon mostrou que as proteínas codificadas pelas ORFs 15 e 16 são necessárias para a atividade lipolítica. Quando a inserção do transposon ocorreu em LP018ORF15 a atividade foi perdida. No entanto, esse pode ser um efeito polar da mutação em ORF15 sobre ORF16 indicando que, de fato, os dois genes constituem um operon. Quando a inserção do transposon ocorreu somente em LP018ORF16 a atividade também foi perdida indicando que esse gene é o responsável pela codificação da lipase.

A relação entre LP018ORF15 e 16 não foi determinada, mas pela classificação proposta pelo COG é provável que LP018ORF15 esteja atuando como uma chaperona responsável pelo processamento de LP018ORF16 em sua forma ativa. Entretanto, as comparações de sequências e filogenéticas sugerem que essa proteína está mais próxima de peptidases do que de chaperonas. É possível que, assim como

LP018ORF16 apresenta uma sequência de aminoácidos completamente distinta de outras lipases conhecidas, LP018ORF15 represente uma chaperona distinta das chaperonas clássicas de lipases. A utilização de chaperonas por lipases, que nesse caso são denominadas de “lif” (*lipase-specific foldase* – foldase específica de lipases), para obter sua conformação ativa não é incomum (JAEGER et al, 1994). Em *Pseudomonas glumae* a obtenção da forma ativa da enzima LipA é dependente da atuação de outra proteína denominada LipB. Os genes que codificam essas proteínas estão organizados em um operon *lipAB* e foi demonstrado que LipB atua como uma lif (FRENKEN et al, 2006). O mesmo tipo de organização também foi encontrada em *P. aeruginosa* (CHIHARA-SIOMI, 1992), *Vibrio cholerae* (OGIERMAN, 1997), *Burkholderia cepacia* (JØRGENSEN, SKOV e DIDERICHSEN, 1991) e em *Acinetobacter calcoaceticus* (KOK et al, 1995), sendo que nesse último organismo a ordem dos genes está invertida, com a chaperona na frente da lipase. Essa inversão na ordem também ocorre para LP018ORF15-16. A descrição de lif para lipases provinda de metagenomas não é comum. Há vários trabalhos descrevendo a obtenção de lipases a partir de metagenomas, mas nenhum cita a presença de lif associadas a elas.

6.6 Clone MAF1LP090

O clone MAF1LP090 foi selecionado devido a sua capacidade de hidrolisar tributirina e tricaprilina. A proteína codificada por LP090ORF24 foi identificada como a provável lipase que motivou o sequenciamento integral do inserto de DNA do clone MAF1LP090. A análise filogenética indicou que essa lipase pertence à família I das lipases e esta próxima das subfamílias I.1 e I.2. A subfamília I.1 inclui as lipases de *P. aeruginosa*, *P. fragi*, *P. fluorescens* C9, *P. wisconsinensis*, *V. cholerae*, *A. calcoaceticus* e *Proteus vulgaris*, enquanto a subfamília I.2 incluem as lipases de *P. luteola*, *B. glumae*, *B. cepacia*, and *Chromobacterium viscosum* (ARPIGNY e JAEGER, 1999). As lipases classificadas dentro dessas subfamílias geralmente são secretadas para o meio extracelular através de sistemas de secreção do tipo II (TISS) e possuem uma sequência sinal de endereçamento. Os genes que compõem o TISS

geralmente estão organizados em grandes operons contendo de 12 a 16 genes. Através da busca por domínios conservados nas ORFs codificadas no inserto de DNA do clone MAF1LP090 foi identificado um possível candidato relacionado ao TISS. A proteína LP090ORF8 apresenta um domínio similar ao da proteína ExaA. Em *Aeromonas hydrophila* *exaA* é co-transcrito com *exaB* a partir do operon *exaAB*. Essas duas proteínas são responsáveis pela primeira etapa de transporte da membrana interna para o espaço periplasmático e o seu operon está distante do outro operon, *exaCN*, que codifica outras proteínas que constituem o sistema de secreção nesse organismo (JAHAGIRDAR e HOWARD, 1994). A proteína LP090ORF8 apresenta ainda um domínio ATPase com motivo Walker característico da proteína GspE. Essa proteína executa uma função análoga a ExaA em *E. coli* e outras bactérias Gram negativas (MICHEL E VOULHOUX, 2009), mas não é possível realizar afirmações sobre a sua função no organismo original. Entretanto, LP090ORF24 não possui peptídeo sinal. O mais provável é que LP090ORF24 seja transportada através de um sistema não clássico de transporte que independe de peptídeo sinal.

Através do sequenciamento das extremidades do inserto de DNA do clone MAF1LP090 também foi possível identificar a presença de um gene que codifica uma proteína similar a α -amilase (LP090ORF27). Essas enzimas realizam a quebra de ligações glicosídicas do tipo α -1,4 atuando sobre amido e glicogênio. Um domínio semelhante a CGTase também foi identificado nessa proteína. As CGTases são capazes de realizar uma reação de transglicosilação a partir do amido parcialmente hidrolisado para formar ciclodextrinas de 6, 7 ou 8 moléculas de glucose. As ciclodextrinas possuem uma conformação toroidal com o exterior hidrofílico e interior hidrofóbico sendo capazes de formar complexos de inclusão com diferentes compostos, com aplicações nas indústrias farmacêutica, de cosméticos, alimentícia e têxtil (DEL VALLE, 2004). No caso dessa enzima foi identificado um peptídeo sinal na região N-terminal indicando que ela é exportada e pode ser dependente do sistema de secreção descrito acima. A presença do peptídeo sinal também é outra característica que favorece a função CGTase, já que essas enzimas atuam no meio extracelular. No

entanto, não é possível fazer mais inferências sobre a atividade dessa enzima, já que não foram realizados testes funcionais.

Analisando os dados apresentados na Tabela 16 é possível inferir que o inserto de MAF1LP090 faz parte de um grande cluster de genes responsáveis pelo metabolismo e exportação de exopolissacarídeos e também sua ligação a proteínas de membrana para formar peptidoglicanas.

6.7. Considerações finais

A grande maioria dos procariotos não responde aos meios de cultivos tradicionalmente usados em laboratório. No entanto, são considerados os maiores reservatórios de diversidade genética do planeta. Nos últimos anos a Metagenômica se tornou uma excelente ferramenta para acessar a biodiversidade genética desses organismos. Isso ficou demonstrado nesse trabalho no qual 80 genes foram identificados a partir de DNA purificado de solo da Floresta Atlântica paranaense. Dentre esses, três genes que codificam lipases e um que codifica uma amilase, com potencial aplicação biotecnológica, três genes que codificam peptidases e vários outros genes que codificam biocatalisadores do metabolismo geral da célula como oxirredutases, transferases, sintases, fosfatases e quinases. Dentre todos os genes identificados, o gene que codifica a lipase LP001ORF27 foi clonado e a proteína superexpressa e purificada, permitindo a sua caracterização quanto às propriedades bioquímicas. A especificidade de substrato e a temperatura ótima de atividade permitiram classificar LP001ORF27 como uma lipase verdadeira moderadamente termofílica.

CAPÍTULO 7

CONCLUSÕES

7. CONCLUSÕES

- A comunidade bacteriana dos solos da Floresta Atlântica paranaense em alta altitude é diferente daquela em baixa altitude,
- A quantidade de DNA metagenômico clonado nas bibliotecas representa cerca de 900 genomas bacterianos,
- A prospecção funcional por biocatalizadores identificou 315 esterases, 3 lipases verdadeiras, 4 amilases, 460 proteases, 10 clones resistentes a ácido nalidíxico e 14 clones resistentes a rifampicina,
- O sequenciamento do inserto de DNA de 4 fosmídeos guiados pela prospecção funcional levou à identificação de 100 genes,
- O fragmento de DNA clonado em MAF1LP001 provavelmente deriva de uma acidobactéria fotossintética,
- A lipase LP001ORF27 foi caracterizada quanto a sua filogenia e suas propriedades bioquímicas e pode ser classificada como uma lipase verdadeira moderadamente termofílica,
- A lipase LP018ORF16 é dependente da proteína LP018ORF15 e representa uma nova família na classificação clássica das lipases,
- A lipase LP090ORF24 compartilha características com uma enzima PHA depolimerase,

CAPÍTULO 8

REFERÊNCIAS

BIBLIOGRÁFICAS

8. REFERÊNCIAS BIBLIOGRÁFICAS

AHMED, E. H.; RAGHAVENDRA, T.; MADAMWAR, D. A thermostable alkaline lipase from isolate *Bacillus subtilis* EH 37: Characterization, partial purification, and application in organic synthesis. **Appl. Biochem. Biotechnol.**, DOI 10.1007/s12010-009-8751-4, 2009.

ALEKSHUN, M. N.; LEVY, S. B.; MEALYS, T. R.; SEATONS, B. A.; HEADS, J. F. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. **Nat. Struct. Biol.**, v. 8, n. 8, p. 710-714, 2001.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic. Acids Res.**, v. 25, p. 3389-3402, 1997.

AMANN, R. I.; LUDWING, W.; SHLEIFER, K. H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. **Microbiol. Rev.**, v. 59, n. 1, p. 143-169, 1995.

ANDREWS, J. M. Determination of minimum inhibitory concentrations. **J. Antimicrob. Chemother.**, v.48, p. 5-16, 2001.

ANGKAWIDJAJA, C.; KANAYA, S. Family I.3 lipase: bacterial lipases secreted by the type I secretion system. **Cell. Mol. Life Sci.**, v. 63, p. 2804–2817, 2006.

APWEILER, R.; ATTWOOD, T. K.; BAIROCH, A.; BATEMAN, A.; BIRNEY, E.; BISWAS, M.; BUCHER, P.; CERUTTI, L.; CORPET, F.; CRONING, M. D. R.; DURBIN, R.; FALQUET, L.; FLEISCHMANN, W.; GOUZY, J.; HERMHAKOB, H.; HULO, N.; JONASSEN, I.; KAHN, D.; KANPIN, A.; KARAVIDOPOULOU, Y.; LOPEZ, R.; MARX, B.; MULDER, N. J.; OINN, T. M.; PAGNI, M.; SEVANT, F.;

SIGRIST, C. J. A.; ZDOBNOV, E. M. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. **Nucleic Acids Res.**, v. 29, n. 1, p. 37-40, 2001.

ARPIGNY, J. L.; JAEGER, K-E. Bacterial lipolytic enzymes: classification and properties. **Biochem. J.**, v. 343, p. 177-183, 1999.

BANIULIS, D.; YAMASHITA, E.; ZHANG, H.; HASAN, S. S.; CRAMER, W. A. Structure–function of the cytochrome b6f complex. **Photochem. Photobiol.**, v. 84, p. 1349–1358, 2008.

BARBOSA, T. M.; LEVY, S. B. Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. **J Bacteriol.**, v. 182, p. 3467–3474, 2000.

BERLEMONT, R.; DELSAUTE, M.; PIPERS, D.; D'AMINCO, S.; FELLER, G.; GALLEN, M.; POWER, P. Insights into bacterial cellulose biosynthesis by functional metagenomics on Antarctic soil samples. **ISME J**, v. 3, p. 1070-1081, 2009.

BORNEMAN, J.; SKROCH, P. W.; O'SULLIVAN, K. M.; PALUS, J. Á.; RUMJANEK, N. G.; JANSEN, J. L.; NIENHUIS, J.; TRIPLET, E. W. Molecular microbial diversity of an agricultural soil in Wisconsin. **Appl. Environ. Microbiol.**, v. 62, p. 1935–1943, 1996.

BORNEMAN, J.; TRIPLET, E. W. Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. **Appl. Environ. Microbiol.**, v. 63, n. 7, p. 2647-2653, 1997.

BORNSCHEUER, U. T.; KAZLAUKAS, R. J. **Hydrolases in organic synthesis: regio- and stereoselective biotransformations**. 2º Edição, WILEY-VCH Verlag GmbH & Co. ISBN 3-527-31029-0, 2006. Acessado através do endereço eletrônico <http://books.google.com.br>

BRADY, S. F.; CLARDY, J. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. **J. Am. Chem. Soc.**, v. 122, p. 12903–12904, 2000.

BRYANT, D. A.; COSTAS, A. M. G.; MARESCA, J. A.; CHEW, AL. G. M.; KLATT, C. G.; BATESON, M. M.; TALLON, L. J.; HOSTETLER, J.; NELSON, W. C.; HEIDELBERG, J. F.; WARD, D. M. Candidatus *Chloracidobacterium thermophilum*: an aerobic phototrophic acidobacterium. **Science**, v. 317, p. 523-526, 2009.

BUD, R. **The use of life: a history of Biotechnology**. Cambridge University Press, 264p, 1993. Acessado através do endereço eletrônico <http://books.google.com.br>

CÂMARA, I. G. Breve história da conservação da Mata Atlântica. In: GALINDO-LEAL, C.; CÂMARA, I. G. **Mata Atlântica: biodiversidade, ameaças e perspectivas**. São Paulo: fundação SOS Mata Atlântica - Belo Horizonte: Conservação Internacional, 2005. 427 p

CANE, D. E.; WALSH, C. T. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. **Chem. Biol.**, v. 6, p.319–325, 1999.

CHIHARA-SIOMI, M.; YOSHIKAWA, K.; OSHIMA-HIRAYAMA, N.; YAMAMOTOA, K.; SOGABEA, Y.; NAKATANI, T.; NISHIOKA, T.; ODA, H. Purification, molecular cloning, and expression of lipase from *Pseudomonas aeruginosa*. **Arch. Biochem. Biophys.**, v. 296, n. 2, p. 505-513, 1992.

COHEN, S. P.; HÄCHLER, H.; LEVY, S. B. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in *Escherichia coli*. **J. Bacteriol.**, v. 75, p. 1484–1492, 1993.

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

DANIEL, R. The soil metagenome – a rich resource for the discovery of novel natural products. **Curr. Opin. Biotechnol.**, v. 15, p. 199-204, 2004.

DANIEL, R. The metagenomics of soil. **Nat. Rev. Microbiol.**, v. 3, p.470-478, 2005.

DANTAS, G.; SOMMER, M. O. A.; OLUWASEGUN, R. D., CHURCH, G. M. Bacteria subsisting on antibiotics. **Science**, v. 320, p. 100-103, 2008.

DAVIES, J. Inactivation of antibiotics and the dissemination of resistance genes. **Science**, v. 264, n. 5157, p. 375 – 382, 1994.

D’COSTA, V. M.; GRIFFITHS, E.; WRIGHT, G. D. Expanding the soil antibiotic resistome: exploring environmental diversity. **Curr. Opin. Microbiol.**; v. 10, p. 481-489, 2007.

DELCHER, A. L., HARMON, D., KASIF, S., WHITE, O., SALZBERG, S. L. Improved microbial gene identification with glimmer. **Nucleic Acids Res.**, v. 27, n. 23, p.4636-4641, 1999.

DeLONG, E. F.; PACE, N. A. Environmental diversity of Bacteria and Archaea. **Syst. Biol.**, v. 50, n. 4, p. 470-478, 2001.

DEL VALLE E. M. M. Cyclodextrins and their uses: a review. **Process Biochem.**, v. 39, p. 1033–1046, 2004.

DEMAIN, A. L. From natural products discovery to commercialization: a success story. **J. Ind. Microbiol. Biotechnol.**, v. 33, p.486–495, 2006.

DIAZ-TORRES, M. L.; McNAB, R.; SPRATT, D. A.; VILLEDIEU, A.; HUNT, N.; WILSON, M.; MULLANY, P. Novel tetracycline resistance determinant from the oral metagenome. **Antimicrob. Agents Chemother.**, v. 47, n. 4, p. 1430-1432, 2003.

DUAN, C-J.; XIAN, L.; ZHAO, G.C.; FENG, Y.; PANG, H.; BAI, X-L.; TANG, J-L.; MA, Q-S.; FENG, J-X. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. **J. Appl. Microbiol.**, v. 10, p. 245-256, 2009.

DUTTA, S.; RAY, L. Production and characterization of an alkaline thermostable crude lipase from an isolated strain of *Bacillus cereus* C₇. **Appl. Biochem. Biotechnol.**, v. 159, p. 142-154, 2009.

ELBANNA, K.; LÜTKE-EVERSLOH, T.; TRAPPEN, S. V.; MERGAERT, J.; SWINGS, J.; STEINBÜCHEL A. *Schlegelella thermodepolymerans* gen. nov., sp. nov., a novel thermophilic bacterium that degrades poly(3-hydroxybutyrate-co-3-mercaptopropionate). **Int. J. Syst. Evol. Microbiol.**, v.53, p. 1165-1168, 2003.

ELEND, C.; SCHMEISSER, C.; LEGGEWIE, C.; BABIAK, P.; CARBALLEIRA, J. D.; STEELE, H. L.; REYMOND, J-L.; JAEGER, K-E.; STREIT, W. R. Isolation and biochemical characterization of two novel Metagenome-derived esterases. **Appl. Environ. Microbiol.**, v. 72, n. 5, p. 3637-3645, 2006.

ELEND, C.; SCHMEISSER, C.; HOEBENREICH, H.; STEELE, H. L.; STREIT, W. R. Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. **J. Biotechnol.**, v. 130, p. 370-377, 2007.

EWING, B.; HILILIER, L.; WENDL, M. C.; GREEN, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. **Genome Res.**, v. 8, p. 175-185, 1998a.

EWING, B.; GREEN, P. Basecalling of automated sequencer traces using Phred. II. Error probabilities. **Genome Res.**, v. 8, p. 186-194, 1998b.

FAORO, H. Determinação da biodiversidade de Bacteria e Archaea da Mata Atlântica paranaense. Curitiba: **Dissertação (Mestrado em Ciências-Bioquímica)**. Setor de Ciências Biológicas, Universidade Federal do Paraná. 2006, 171f.

FELSENSTEIN, J. PHYLIP (evolution.genetics.washington.edu/phylip.htm)

FELSKE, A.; WOLTERINK, A.; van LIS, R.; AKKERMANS, A. D. L. Phylogeny of the main bacterial 16S rRNA sequences in a Drentse grassland soils (The Netherlands). **Appl. Environ. Microbiol.**, v. 54, p. 871-879, 1998.

FENG, Y.; DUAN, C-J.; PANG, H.; MO, X-C.; WU, C-F.; YU, Y.; HU, Y-L.; WEI, J.; TANG, J-L.; FENG, J-X. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. **Appl. Microbiol. Biotechnol.**, v. 75, p. 319–328, 2007.

FERRER, M.; MARTÍNEZ-ABARCA, F.; GOLYSHIN, P. N. Mining genomes and metagenomes for novel catalysts. **Curr. Opin. Biotechnol.**, v. 16, p. 588-593, 2005.

FERRER, M.; GOLYSHINA, O. V.; CHERNIKOVA, T. N.; KHACHANE, A. N.; dos SANTOS, V. A. P. M.; YAKIMOV, M. M.; TIMMIS, K. N.; GOLYSHIN, P. N. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. **Chem. Biol.**, v. 12, p. 895-904, 2005a.

FERRER, M.; GOLYSHINA, O. V.; CHERNIKOVA, T. N.; KHACHANE, A. N.; REYES-DUARTE, D.; MARTIN dos SANTOS, V. A. P.; STROMPL, C.; ELBOROUGH, K.; JARVIS, G.; NEEF, A.; YAKIMOV, M. M.; TIMMIS, K. N.; GOLYSHIN, P. N. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. **Environ. Microbiol.**, v. 7, n. 12, p. 1996-2010, 2005b.

FERRER, M.; BELOQUI, A.; GOLYSHINA, O. V.; PLOU, F. J.; NEEF, A.; CHERNIKOVA, T. N.; FERNÁNDEZ-ARROJA, L.; GHAZIL, I.; BALLESTEROS, A.; ELBOROUGH, K.; TIMMIS, K. N.; GOLYSHIN, P. N. Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome. **Biotechnol. J.**, v. 2, p. 207-213, 2007.

FRENKEN, L.G.J.; DE GROOT, A.; TOMMASSEN, J.; VERRIPS, C.T. Role of the *lipB* gene product in the folding of the secreted lipase of *Pseudomonas glumae*. **Mol. Microbiol.**, v. 9, p. 591–599, 1993.

FUDOU, R.; JOJIMA, Y.; IIZUKA, T.; YAMANAKA, S. *Haliangium ochraceum* gen. nov., sp. nov. and *Haliangium tepidum* sp. nov.: Novel moderately halophilic myxobacteria isolated from coastal saline environments. **J. Gen. Appl. Microbiol.**, v. 48, n. 2, p. 109-115, 2002.

GASTEIGER, E.; HOOGLAND, C.; GATTIKER, A.; DUVAUD, S.; WILKINS, M. R.; APPEL, R. D.; BAIROCH, A. Protein Identification and Analysis Tools on the ExPASy Server. In: WALKER, J. M. **The Proteomics Protocols Handbook**, Humana Press, p. 571-607, 2005.

GILLESPIE, D. E.; BRADY, S. F.; BETTERMANN, A. D.; CIANCIOTTO, N. P.; LILES, M. R.; RONDON, M. R.; CLARDY, J.; GOODMAN, R. M.; HANDELSMAN, J. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. **Appl. Environ. Microbiol.**, v. 68, p. 4310–4306, 2002.

GINOLHAC, A.; JARRIN, C.; GILLET, B.; ROBE, P.; PUJIC, P.; TUPHILE, K.; BERTRAND, H.; VOGEL, T. M.; PERRIERE, G.; SIMONET, P.; NALIN, R. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. **Appl. Environ. Microbiol.**, v. 70, p. 5522–5527, 2004.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Res.**, v. 8, p.195-202, 1998.

GUPTA, R.; BEG, Q. K.; LORENZ, P. Bacterial alkaline proteases: molecular approaches and industrial applications. **Appl. Microbiol. Biotechnol.**, v. 59, p. 15-32, 2002.

GUPTA, R.; MOHAPATRA, H.; GOSWAMI, V. K.; CHAUHAN, B. Microbial α -Amylases: a Biotechnological Perspective. **Process Biochem.**, p. 1-18, 2003.

GUPTA, R.; GUPTA, N.; RATHI, P. Bacterial lipases: an overview of production, purification and biochemical properties. **Appl. Microbiol. Biotechnol.**, v. 64, p. 763-781, 2004.

HALL, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symp. Ser.**, n. 41, p. 95-98, 1999.

HANDELSMAN, J.; RONDON, M. R.; BRADY, S. F.; CLARDY, J.; GOODMAN, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chem. Biol.**, v. 5, n. 10, p.245-249, 1998.

HEIKINHEIMO, P.; GOLDMAN, A.; JEFFRIES, C.; OLLIS, D. L. Of barn owls and bankers: a lush variety of α/β hydrolases. **Structure**, v. 7, p. 141-146, 1999.

HENNE, A.; DANIEL, R.; SCHMITZ, R. A.; GOTTSCHALK, G. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. **Appl. Environ. Microbiol.**, v. 65, p. 3901-3907, 1999.

HUDDLESTON, S.; YALLOP, C. A.; CHARALAMBOUS, B. M. The identification and partial characterization of a novel inducible extracellular thermostable esterase from the archaeon *Sulfolobus shibatae*. **Biochem. Biophys. Res. Commun.**, v. 216, n. 2, p. 495-500, 1995.

HUGHES, D. S.; FELBECK, H.; STEIN, J. L. A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. **Appl. Environ. Microbiol.**, v. 63, p. 3494-3498, 1997.

HUNTER-CEVERA, J. C. The value of microbial diversity. **Curr. Opin. Microbiol.**, v. 1, n. 3, p. 278-285, 1998.

HURT, R. A.; QIU, X.; WU, L.; ROH, Y.; PALUMBO, A. V.; TIEDJE, J. M.; ZHOU, J. Simultaneous recovery of RNA and DNA from soils and sediments. **Appl. Environ. Microbiol.**, v. 67, p. 4495–4503, 2001.

JAEGER, K.-E.; RANSAC, S.; DIJKSTRA, B.W.; VAN HEUVEL, S.; MISSET, O. Bacterial lipases. **FEMS. Microbiol. Rev.**, v. 15, p. 29–63, 1994.

JAEGER, K.-E.; DIJKSTRA, B. W.; REETZ, M. T. Bacterial biocatalysts: biology, tree-dimensional structures, and biotechnological applications. **Annu. Rev. Microbiol.**, v. 53, p. 315-351, 1999.

JAEGER, K. E.; EGGERT, T. Lipases for biotechnology. **Curr. Opin. Biotechnol.**, v. 13, p. 390–397, 2002.

JAHAGIRDAR, R.; HOWARD, S. P. Isolation and characterization of a second *exe* operon required for extracellular protein secretion in *Aeromonas hydrophila*. **J. Bacteriol.**, v. 176, n. 22, p. 6819-6826, 1994.

JEONG, H. J.; KIM, J-T.; KIM, Y. J.; KIM, H-K.; LEE, H. S.; KANG, S. G.; KIM, S-J.; LEE J-H. Cloning and characterization of a new cold-active lipase from a deep-sea sediment Metagenome. **Appl. Microbiol. Biotechnol.**, v. 81, p. 865–874, 2009.

JONES, B. V.; SUN, F.; MARCHESI, J. R. Using skimmed milk agar to functionally screen metagenomic library for proteases may lead to false positives. **Lett. Appl. Microbiol.**, v. 45, p. 418-420, 2007.

JØRGENSEN, S.; SKOV, K. W.; DIDERICHSEN, B. Cloning, sequence, and expression of a lipase gene from *Pseudomonas cepacia*: lipase production in heterologous hosts requires two *Pseudomonas* genes. **J. Bacteriol.**, v. 173, n. 2, p. 559-567, 1991.

KELLER, M.; ZENGLER, K. Tapping into microbial diversity. **Nature Reviews**, v. 2, p. 141-150, 2004.

KIM, H-K.; PARK, S-Y.; LEE, J-K., OH, T-K. Gene cloning and characterization of thermostable lipase from *Bacillus stearothermophilus* L1. **Biosci. Biotechnol. Biochem.**, v. 62, n. 1, 66-71, 1998.

KIM, S-J.; LEE, C-M.; HAN, B-R.; KIM, M-Y.; YEO, Y-S.; YOON, S-H.; KOO, B-S.; JUN, H-K. Characterization of a gene encoding cellulase from uncultured soil bacteria. **FEMS Microbiol. Lett.**; v. 282, p. 44–51, 2008.

KIM, E-Y.; OH, K-H.; LEE, M-H.; KANG, C-H.; OH, T-K.; YOON, J-H. Novel cold-adapted alkaline lipase from an intertidal flat Metagenome and proposal for a new family of bacterial lipases. **Appl. Microbiol. Biotechnol.**, v. 75, n. 1, p. 257–260, 2009.

KNIETSCH, A.; WASCHKOWITZ, T.; BOWIEN, S.; HENNE, A.; DANIEL, R. Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. **J. Mol. Microbiol. Biotechnol.**, v. 5, p. 46–56, 2003.

KNOLL, M.; HAMM, T.; WAGNER, F.; MARTINEZ, V.; PLEISS, J. The PHA Depolymerase Engineering Database: A systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. **BMC Bioinformatics**, v. 10, n. 89, p. 1-8, 2009.

KOK, R. G.; van THOR, J. J.; NUGTEREN-ROODZANT, I. M.; VOSMAN, B.; HELLINGWERF, K. J. Characterization of lipase-deficient mutants of *Acinetobacter calcoaceticus* BD413: identification of a periplasmic lipase chaperone essential for the production of extracellular lipase. **J. Bacteriol.**, v. 177, n. 11, p. 3295-3307, 1995.

KOZDRÓJ, J.; van ELSAS, J. D. Application of polymerase chain reaction-denaturing gradient gel electrophoresis for comparison of direct and indirect extraction methods of soil DNA used for microbial community fingerprinting. **Biol. Fertil. Soils**, v. 31, p.372-378, 2001.

LÄMMLER, K.; ZIPPER, H.; BREUER, M.; HAUER, B.; BUTA, C.; BRUNNER, H.; RUPP, S. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. **J. Biotechnol.**, p. 1-18, 2006.

LASKARIS, P.; TOLBA, S.; CALVO-BADO, L.; WELLINGTON, L. Coevolution of antibiotic production and counter-resistance in soil bacteria. **Environ. Microbiol.**, v. 12, n. 3, p. 783-796, 2010.

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

LEE, M-H.; LEE, C-H.; Oh, T-K.; Song, J. K.; Yoon, J-H. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. **Appl. Environ. Microbiol.**, v. 72, n. 11, p 7406–7409, 2006.

LEE, D-G.; JEON, J. H.; JANG, M. K.; KIM, N. Y.; LEE, J. H.; LEE, J-H.; KIM, S-J.; KIM, G-D.; LEE, S-H. Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. **Biotechnol. Lett.**, v. 29, p. 465-472, 2007.

LEWINSON, O.; ADLER, J.; SIGAL, N.; BIBI, E. Promiscuity in multidrug recognition and transport: the bacterial MFS Mdr transporters. **Mol. Microbiol.**, v. 61, n. 2, p. 277-284, 2006.

LI, H.; ZHANG, X. Characterization of thermostable lipase from thermophilic *Geobacillus sp.* TW1. **Protein Expr. Purif.**, v. 42, p. 153-159, 2005.

LI, G.; WANG, K.; LIU, H. Y. Molecular cloning and characterization of a novel pyrethroid-hydrolyzing esterase originating from the Metagenome. **Microb. Cell Fact.**, v. 7, n. 38, p. 1-10, 2008.

LIESACK, W.; STACKEBRANDT, E. Occurrence of novel groups of the domain bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. **Appl. Environ. Microbiol.**, v. 174, p. 5072–5078, 1992.

LILES, M. R.; WILLIAMSON, L. L.; RODBUMRER, J.; TORSVIK, V.; GOODMAN, R. M.; HANDELSMAN, JO. Recovery, purification and cloning of high molecular weight DNA from soil microorganisms. **Appl. Environ. Microbiol.**, v. 74, n. 10, p.3302-3305, 2008.

LIM, H. K.; CHUNG, E. J.; KIM, J-C.; CHOI, G. J.; JANG, K. S.; CHUNG, Y. R.; CHO, K. Y.; LEE, S-W. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. **Appl. Environ. Microbiol.**, v. 71, n. 121, p 7768–7777, 2005.

LINARES, J. F.; GUSTAFSSON, I.; BAQUEROS, F.; MARTINEZ, L. Antibiotics as intermicrobial signaling agents instead of weapons. **Proc. Natl. Acad. Sci. U.S.A.**, v. 103, n. 51, p. 19484-19489, 2006.

LORENZ, P.; SCHLPER, C. Metagenome – a challenging source of enzyme discovery. **J. Mol. Catal., B Enzym.**, v. 19-20, p. 13-19, 2002.

LORENZ, P.; ECK J. Metagenomics and industrial applications. **Nat. Rev. Microbiol.**, v. 3, p. 510-516, 2005.

LOWE, T. M.; EDDY, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Res.**, v. 25, n. 955-964, 1997.

LYND, L. R.; WEIMER, P. J.; van ZYL, W. H.; PRETORIUS, I. S. Microbial cellulose utilization: fundamentals and biotechnology. **Microbiol. Mol. Biol. Rev.**, v. 66, n. 3, p. 506-577, 2002.

MacNEIL, I. A.; TIONG, C. L.; MINOR, C.; AUGUST, P. R.; GROSSMAN, T. H.; LOIACONO, K. A.; LYNCH, B. A.; PHILLIPS, T.; NARULA, S.; SUNDARAMOORTHY, R.; TYLER, A.; ALDREDGE, T.; LONG, H.; GILMAN, M.; HOLT, D.; OSBURNE, M. S. Expression and isolation of antimicrobial small molecules from soil DNA libraries. **J. Mol. Microbiol. Biotechnol.**, v. 3, p. 301-308, 2001.

MARTINEZ, J. L. Antibiotics and antibiotic resistance genes in natural environments. **Science**, v. 321, p. 365-367, 2008.

MARTINEZ, J. L.; FAJARDO, A.; GARMENDIA, L.; HERNANDEZ, A.; LINARES, J. F.; MARTINEZ-SOLANO, L.; SÁNCHEZ, M. B. A global view of antibiotic resistance. **FEMS Microbiol. Rev.**, v. 33, p. 44-63, 2009.

MARSHALL, C. G.; BRADHEAD, G.; LESKIW, B. K.; WRIGHT, G. D. D-Ala-D-Ala ligases from glycopeptide antibiotic-producing organisms are highly homologous to the enterococcal vancomycinresistance ligases VanA and VanB. **Proc. Natl. Acad. Sci. U S A.**, v. 94, p. 6480-6483, 1997.

McNULTY, D. E.; CLAFFEE, B. A.; HUDDLESTON, M. J.; KANE, J. F. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. **Protein Expr Purif.**, v. 27, n. 2, p. 365-374, 2003.

MEILLEUR, C.; HUPÉ, J-F.; JUTEAU, P., SHARECK, F. Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. **J. Ind. Microbiol. Biotechnol.**, v. 36, p.853–861, 2009.

MICHEL, G. P. F.; VOULHOUX, R. Type II secretory system (T2SS) in gram-negative bacteria: a molecular nanomachine for secretion of Sec and Tat-dependent extracellular proteins. In: WOOLDRIDGE, K. **Bacterial secreted proteins – secretory mechanisms and role in pathogenesis**. Norfolk, UK: Caister Academic Press, 2009, p. 67-92. Acessado através do endereço eletrônico <http://books.google.com.br>

MITSUI, H.; GORLACH, K.; LEE, H. J.; HATTORI, R.; HATTORI, T. Incubation time and media requirements of culturable bacteria from different phylogenetic groups. **J. Microbiol. Methods**, v. 30, p. 103-110, 1997.

MORIMOTO, S.; FUJII, T. A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. **Appl Microbiol Biotechnol.**, v. 83, p. 389–396, 2009.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C.G.; DA FONSECA, G. A. B.; KENT; J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, p. 853-858, 2000.

NARDINI, M.; DIJKSTRA, B. W. α/β Hydrolase fold enzymes: the family keeps growing. **Curr. Opin. Struct. Biol.**, v. 9, p. 732-737, 1999.

NIELSEN, H.; ENGLEBRECHT, J.; BRUNAK, S.; von HEIJNE, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. **Protein Eng.**, v. 10 p. 1-6, 1997.

OGIERMAN, M. A.; FALLARINO, A.; RIESS, T.; WILLIAMS, S. G.; ATTRIDGE, S. R.; MANNING, P. A. Characterization of the *Vibrio cholera* El Tor lipase operons *lipAB* and a protease gene downstream of the *hly* region. **J. Bacteriol.**, v. 179, n. 22, p. 7072-7080, 1997.

OGRAM, A.; SAYLER, G. S.; BARKAY, T. The extraction and purification of microbial DNA from sediments. **J. Microbiol. Methods.** v. 7, p. 57-66, 1987.

PACE, N. R.; STAHL, D. A.; LANE, D. J.; OLSEN, G. J.. The analysis of natural microbial populations by ribosomal RNA sequences. **Adv. Microb. Ecol.**, v. 9, p. 1-55, 1986.

PLEISS, J.; FISCHER, M.; PEIKER, M.; THIELE C.; SCHMID, R. D. Lipase engineering database: understanding and exploiting sequence-structure-function relationships. **J Mol Catal B: Enzym.**, v. 10, p. 491-508, 2000.

RAES, J.; KORBEL, J. O.; LERCHER, M. J.; von MERING, C.; BORK, P. Prediction of effective genome size in metagenomic samples. **Genome Biol.**, v. 8, n. 1, p. 8R10.1-8R10.11, 2007.

RAPPÉ, M. S.; GIOVANONNI, S. J. The uncultured microbial majority. **Annu. Rev. Microbiol.**, v. 67, p. 369-394, 2003.

RATHI, P.; BRADDOO, S.; SAXENA, R. K.; GUPTA, R. A hyper-thermostable, alkaline lipase from *Pseudomonas sp.* with the property of thermal activation. **Biotechnol. Lett.**, v. 22, p. 495-498, 2000.

REES, H. C.; GRANT, S.; JONES, B.; GRANT, W. D.; HEAPHY, S. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. **Extremophiles**, v. 7, p. 415-421, 2003.

REICHENBACH, H.; HÖFLE, G. Production of bioactive secondary metabolites. In: DWORKIN, M.; KAISE, D. **Myxobacteria II**. Washington: American Society for Microbiology, 1993. p. 347-398.

RICHARDSON, T. H.; TAN, XUQIU.; FREY, G.; CALLEN, W.; CABELL, M.; LAM, D.; MACOMBER, J.; SHORT, J. M.; ROBERTSON, D. E.; MILLER, C. A novel, high performance enzyme for starch liquefaction: discovery and optimization of a low pH, thermostable amylase. **J. Biol. Chem.**, v. 277, n. 29, p. 26501–26507, 2002.

RIESENFELD, C. S.; GOODMAN, R. M.; HANDELSMAN, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. **Environ. Microbiol.**, v. 6, p. 981–989, 2004.

RONDON, M.; AUGUST, P.; BETTERMANN, A. D.; BRADY, S. F.; GROSSMAN, T. H.; LILES, M. R.; LOIACONO, K. A.; LYNCH, B. A.; MACNEILI, A.; MINOR, C.; TIONG, C. L.; GILMAN, M.; OSBURNE, M. S.; CLARDY, J.; HANDELSMAN, J.; GOODMAN, R. M. Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. **Appl. Environ. Microbiol.**, v. 66, n. 6, p. 2541-2547, 2000.

RUTHERFORD, K.; PARKHILL, J.; CROOK, J.; HORSNELL, T.; RICE, P.; RAJANDREAM, M. A.; BARRELL, B. Artemis: sequence visualization and annotation. **Bioinformatics**. v. 16, p. 944-945, 2000.

SAIDIJAM, M.; BENEDETTI, G.; REN, Q.; XU, Z.; HOYLE, C. J.; PALMER, S. L.; WARD, A.; BETTANEY, K. E.; SZAKONI, G.; MEULLER, J.; MORRISON, S.; POS, M. K.; BUTAYE, P.; WALRAVENS, K.; LANGTON, K.; HERBERT, R. B.; SKURRAY, R. A.; PAULSEN, I. T.; O'REILLY, J.; RUTHERFORD, N. G.; BROWN, M. H.; BILL, R. M.; HENDERSON, P. J. Microbial drug efflux proteins of the major facilitator superfamily. **Curr. Drug Targets**, v. 7, n. 7, p. 793-811, 2006.

SAMBROOK, J.; FRITSCH, E.F. MANIATIS, T. **Molecular cloning: a laboratory manual**. 2ed. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press., 1989.

SANGER, F., NICKLEN, S., COULSON, A. R. DNA sequencing with chain terminating inhibitors. **Proc. Natl. Acad. Sci. USA**, v. 74, p. 5463-5467, 1977.

SASSE, F.; STEINMETZ, H.; HOFLE, G.; REICHENBACH, H. Gephyronic acid, a novel inhibitor of eukaryotic protein synthesis from *Archangium gephyra* (myxobacteria). Production, isolation, physico-chemical and biological properties, and mechanism of action. **J. Antibiot.**, v. 48, n. 1, p. 21-5, 1995.

SCHLOSS, P. D.; HANDELSMAN, J. Biotechnological prospects from metagenomics. **Curr. Opin. Biotechnol.**, v. 14, p.303–310, 2003.

SCHLOSS, P. D.; LARGET, B. R.; HANDELSMAN, J. Integration of microbial ecology and statistics: a test to compare gene libraries. **Appl. Environ. Microbiol.**, v. 70, n. 9, p. 5485-5492, 2004.

SCHLOSS, P. D.; HANDELSMAN, J. Introducing DOTUR, a computer program for defining Operational Taxonomic Units and estimating species richness. **Appl. Environ. Microbiol.**, v. 71, n. 3, p. 1501-1506, 2005.

SCHLÜTER A.; BEKEL T.; DIAZ N. N., DONDRUP M.; EICHENLAUB R.; GARTEMANN K-H.; KRAHN I.; KRAUSE L.; KRÖMEKE H.; KRUSE O.; MUSSGNUG J. H.; NEUWEGER H.; NIEHAUS K.; PÜHLER A.; RUNTE K. J.; SZCZEPANOWSKI R.; TAUCH A.; TILKER A.; VIEHÖVER P.; GOESMANN A. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. **J. Biotechnol.**, v. 136, p.77-90, 2008.

SCHMEISSER, C.; STÖCKIGT, C.; RAASCH, C.; WINGERDER, J.; TIMMIS, K. N.; WENDEROTH, D. F.; FLEMMING, H. C.; LIESEGANG, H.; SCHMITZ, R. A.; JAEGER, K-E.; STREIT, W. D. Metagenomi survey of biofilms in drinking-water networks. **Appl. Environ. Microbiol.**; v. 69, n. 12, p. 7298-7309, 2003

SCHMEISSER, C.; STEELE, H.; STREIT, W. Metagenomics, biotechnology with non-culturable microbes. **Appl. Microbiol. Biotechnol.**, v. 75, p. 955-963, 2007.

SCHMID, A.; DORDICK, J. S.; HAUER, B.; KIENER, A.; WUBBOLTS, M.; WITHOLT, B. Industrial biocatalysis today and tomorrow. **Nature**, v. 409, p. 258–268, 2001.

SHARMA, S.; KHAN, F. G.; QAZI, G. N. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. **Appl. Microbiol. Biotechnol.**, DOI 10.1007/s00253-009-2404-y, 2009.

SHEN, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. **Curr. Opin. Chem. Biol.**, v. 7, p. 285–295, 2003

SINGLETON, D. R.; FURLONG, M. A.; RATHBUN, S. L.; WHITMAN, W. B. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. **Appl. Environ. Microbiol.**, v. 67, n. 9, p. 4374-4376, 2001.

STEELE, H. L.; JAEGER, K-E.; DANIEL, R.; STREIT, W. R. Advances in recovery of novel biocatalysts from metagenome. **J. Microbiol. Biotechnol.**, v. 16, p. 25-37, 2009.

STEIN, J. L.; MARSH, T. L.; WU, K. Y.; SHIZUYA, H., DeLONG, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. **J. Bacteriol.**, v. 178, p. 591–599, 1996.

STRAATHOF, A. J. J.; PANKE, S.; SCHMID, A. The production of fine chemicals by biotransformations. **Curr. Opin. Biotechnol.**, v. 13, p. 548-556, 2002.

STRAUSS, U. T.; FABER, K: Deracemization of (+/-)-mandelic acid using a lipase-mandelate racemase two-enzyme system. **Tetrahedron Asymmetry**, v. 10, p. 4079-4081, 1999.

TABITA, R. F.; HANSON, T. E.; SATAGOPAN, S.; WITTE, B. H.; KREEL, N. E. Phylogenetic and evolutionary relationships of rubisco and the rubisco-like proteins and the functional lessons provide by diverse molecular forms. **Phil. Trans. R. Soc. B**, v. 363, p. 2629–2640, 2008.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol. Biol. Evol.**, v. 24, p. 1596-1599, 2007.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. Clustal W: improving the sensivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucl. Acids Res.**, v. 22, p. 4673-4680, 1994.

TIIROLA, M. A.; MÄNNISTO, M. K.; PUHAKKA, J. A.; KULOMAA, M. S. Isolation and characterization of *Novosphingobium sp.* Strain MT1, a dominant polychlorophenol-degrading strain in a groundwater bioremediation system. **Appl. Environ. Microbiol.**, v. 68, n. 1, p. 173-180, 2002.

TORSVIK, V. L. Isolation of bacterial DNA from soil. **Soil Biol. Biochem.**, v. 12, p. 15-21, 1980.

TORSVIK, V.; GOKSØYR, J.; DAAE, F. L. High diversity in DNA or soil bacteria. **Appl. Environ. Microbiol.**, v. 56, n. 3, p.782-787, 1990.

TORSVIK, V.; ØVREAS, L.; THINGSTAD, T. F. Prokaryotic diversity – magnitude, dynamics and controlling factors. **Science**, v. 296, p. 1064-1066, 2002.

TYSON, G. W.; CHAPMAN, J.; HUGENHOLTZ, P.; ALLEN, E. E.; RAM, R. J.; RICHARDSON, P. M.; SOLOVYEV, V. V.; RUBIN, E. M.; ROKHSAR, D. S.; BANFIELD, J. F. Community structure and metabolism through reconstruction of microbial genomes from the environment. **Nature**, v. 428, n. 37–43, 2004.

VENTER, J. C.; REMINGTON, K.; EISEN, J. A.; WU, D.; PAULSEN, I.; NELSON, K. E.; NELSON, W.; FOUTS, D. E.; LEVY, S.; KNAP, A. H.; LOMAS, M. W.; NEALSON, K.; WHITE, O.; PETERSON, J.; HOFFAN, J.; PARSONS, R.; BADEN-TILLSON, H.; PFANNKONCH, C.; ROGERS, Y-H.; SMITH, H. O. Environmental genome shotgun sequencing of the sargasso sea. **Science**, v. 304, p. 66-74, 2004.

VERGER, R. Intervacial activation of lipases: facts and artifacts. **Trends Biotechnol.**, v. 15, p. 32-38, 1997.

VOGET, S.; LEGGEWIE, C.; UESEBECK, A.; RAASCH, C.; JAEGER, K-E.; STREIT, W. R. Prospecting for novel biocatalysts in a soil metagenome. **Appl. Environ. Microbiol.**, v. 69, p. 6235-6242, 2003.

VOGET, S.; STEELE, H. L. STREIT, W. R. Characterization of a metagenome-derived halotolerant cellulase. **J. Biotechnol.**, v. 126, p. 26-36, 2006.

WANG, G. Y.; GRAZIANI, E.; WATERS, B.; PAN, W.; LI, X.; McDERMOTT, J.; MEURER, G.; SAXENA, G.; ANDERSEN, R. J.; DAVIES, J. Novel natural products from soil DNA libraries in a streptomycete host. **Org. Lett.**, v. 2, p. 2401-2404, 2000.

WARD, N. L.; CHALLACOMBE, J. F.; JANSSEN, P. H.; HENRISSAT, B.; COUTINHO, P. M.; WU, M.; XIE, G.; HAFT, D. H.; SAIT, M.; BADGER, J.; BARABOTE, R. D.; BRADLEY, B.; BRETTIN, T. S.; BRINKAC, L. M.; BRUCE,

D.; CREASY, T.; DAUGHERTY, S. C.; DAVIDSEN, T. M.; DEBOY, R. T.; DETTER, J. C.; DODSON, R. J.; DURKIN, A. S.; GANAPATHY, A.; GWINN-GIGLIO, M.; HAN, C. S.; KHOURI, H.; KISS, H.; KOTHARI, S. P.; MADUPU, R.; NELSON, K. E.; NELSON, W. C.; PAULSEN I.; PENN, K.; REN, Q.; ROSOVITZ, M. J.; SELENGUT, J. D.; SHRIVASTAVA, S.; SULLIVAN, S. A.; TAPIA, R.; THOMPSON, L. S.; WATKINS, K. L.; YANG, Q.; YU, C.; ZAFAR, N.; ZHOU, L.; KUSKE, C. R. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. **Appl Environ Microbiol.**, v. 75, n. 7, p. 2046-2056, 2009.

WARNECKE, F.; LUGINBU, P.; IVANOVA, N.; GHASSEMIAN, M.; RICHARDSON, T. H.; STEGE, J. T.; CAYOUEITE, M.; MCHARDY, A. C.; DJORDJEVIC, G.; ABOUSHADI, N.; SOREK, R.; TRINGE, S. G.; PODAR, M.; MARTIN, H. G.; KUNIN, V.; DALEVI, D.; MADEJSKA, J.; KIRTON, E.; PLATT, D.; SZETO, E.; SALAMOV, A.; BARRY, K.; MIKHAILOVA, N.; KYRPIDES, N. C.; MATSON, E. G.; OTTESEN, E. A.; ZHANG, X.; HERNÁNDEZ, M.; MURILLO, C.; ACOSTA, L. G.; RIGOUTSOS, I.; TAMAYO, G.; GREEN, B. D.; CHANG, C.; RUBIN, E. M.; MATHUR, E. J.; ROBERTSON, D. E.; HUGENHOLTZ, P.; LEADBETTER, J. R. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. **Nature**, v. 450, p. 560-569, 2007.

WILD, J.; HRADECNA, Z.; SZYBALSKI, W. Conditionally amplifiable BACs: switching from single-copy to high-copy vector and genomic clones. **Genome Res.**, v. 12, p. 1434-1444, 2002.

WINKLER, U. K.; STUCKMANN, M. Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. **J. Bacteriol.**, v. 138, p. 663-670, 1979.

WHITE, D. G.; GOLDMAN, J. D.; DEMPSE, B.; LEVY, S. B. Role of the *acrAB* locus in organic solvent tolerance mediated by expression of *marA*, *soxS*, or *robA* in *Escherichia coli*. **J Bacteriol.**, v. 179, p. 6122–6126, 1997.

WOESE, C. R. Bacterial evolution. **Microbial Rev.**, v. 51, n. 2, p. 221-271, 1987

WONG, D. W. S. Applications of metagenomics for industrial bioproducts. In: MARCO, D. **Metagenomics: theory, methods and applications**. Norfolk, UK: Caister Academic Press, 2010, p. 141-158. Acessado através do endereço eletrônico <http://books.google.com.br>

WU, C.; SU, B. Identification of novel esterase from metagenomic library of Yangtze river. **J. Microbiol. Biotechnol.**; v. 19, n. 2, p. 187-193, 2009.

YOUSSEF, N. H.; ELSHAHED, M. S. Diversity rankings among bacterial lineages in soil. **ISME J**, v. 3, p. 305–313, 2009.

YUN, J.; SEOWON, K.; PARK, S.; YOON, H.; KIM, M-J.; HEU, S.; RYU, S. Characterization of a novel amyolytic enzyme encoded by a gene from a soil-derived metagenomic library. **Appl. Environ. Microbiol.**, v. 70, n. 12, p. 7229–7235, 2004.

ZHOU, J. M.; BRUNS, M. A.; TIEDJE, J. M. DNA recovery from soils of diverse composition. **Appl. Environ. Microbiol.**, v. 62, p. 316–322, 1996.

ZHOU, X.; SMITH, J. A.; OI, F. M.; KOEHLER, P. G.; BENNETT, G. W.; SCHARF, M. E. Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. **Gene**, v. 395, p. 29–39, 2007.

ANEXOS

9. ANEXOS

9.1 Anexo 1: Mapa do vetor pCC2FOS

9.2 Anexo 2: Mapa do vetor pUC28

9.3 Anexo 3: Mapa do vetor pET28a

9.4 Anexo 4: Modelagem molecular da lipase LP001ORF27

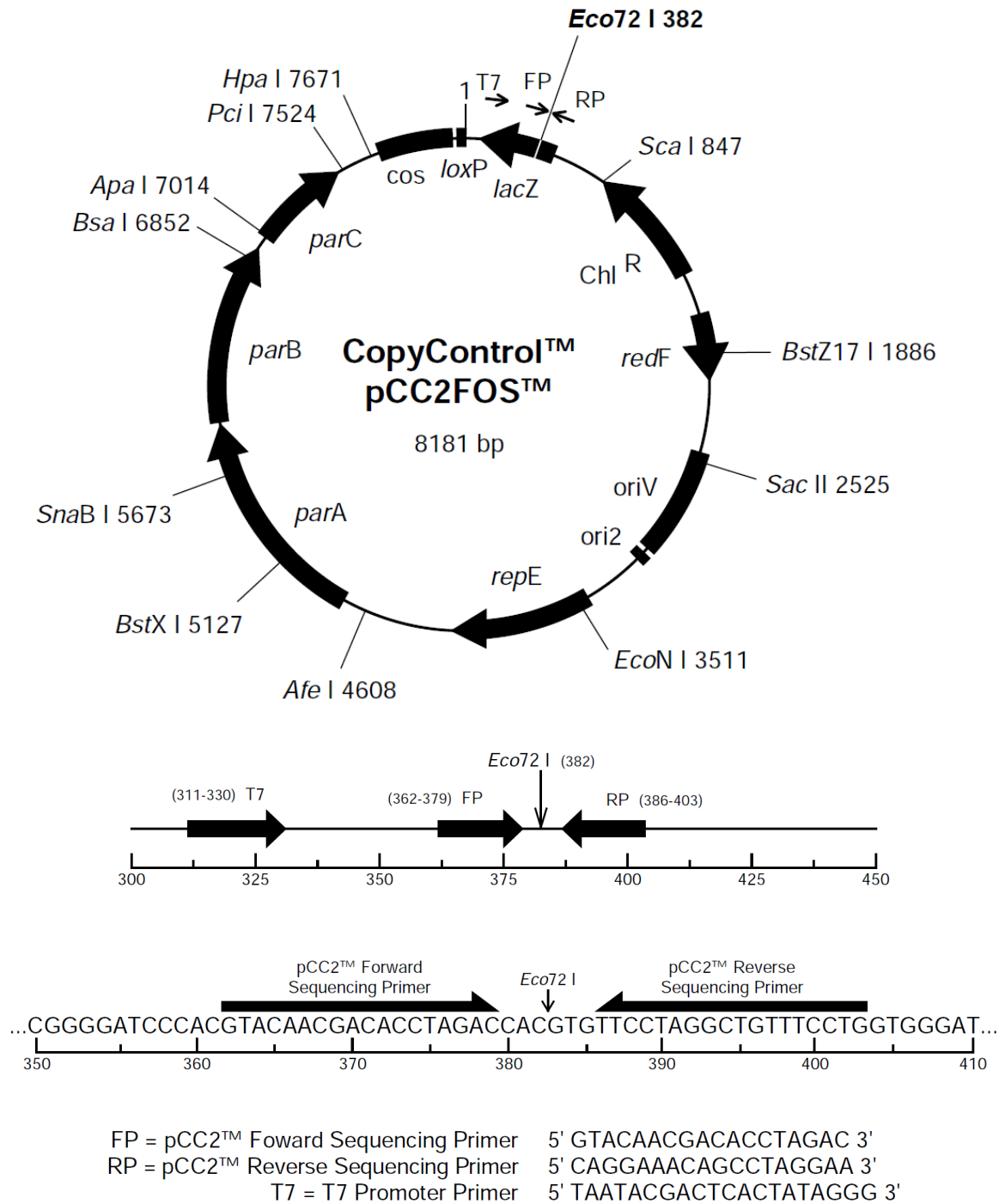
9.5 Anexo 5: Modelo molecular da lipase LP001ORF27

9.6 Anexo 6: Identificação e análise de clones resistentes a antibióticos

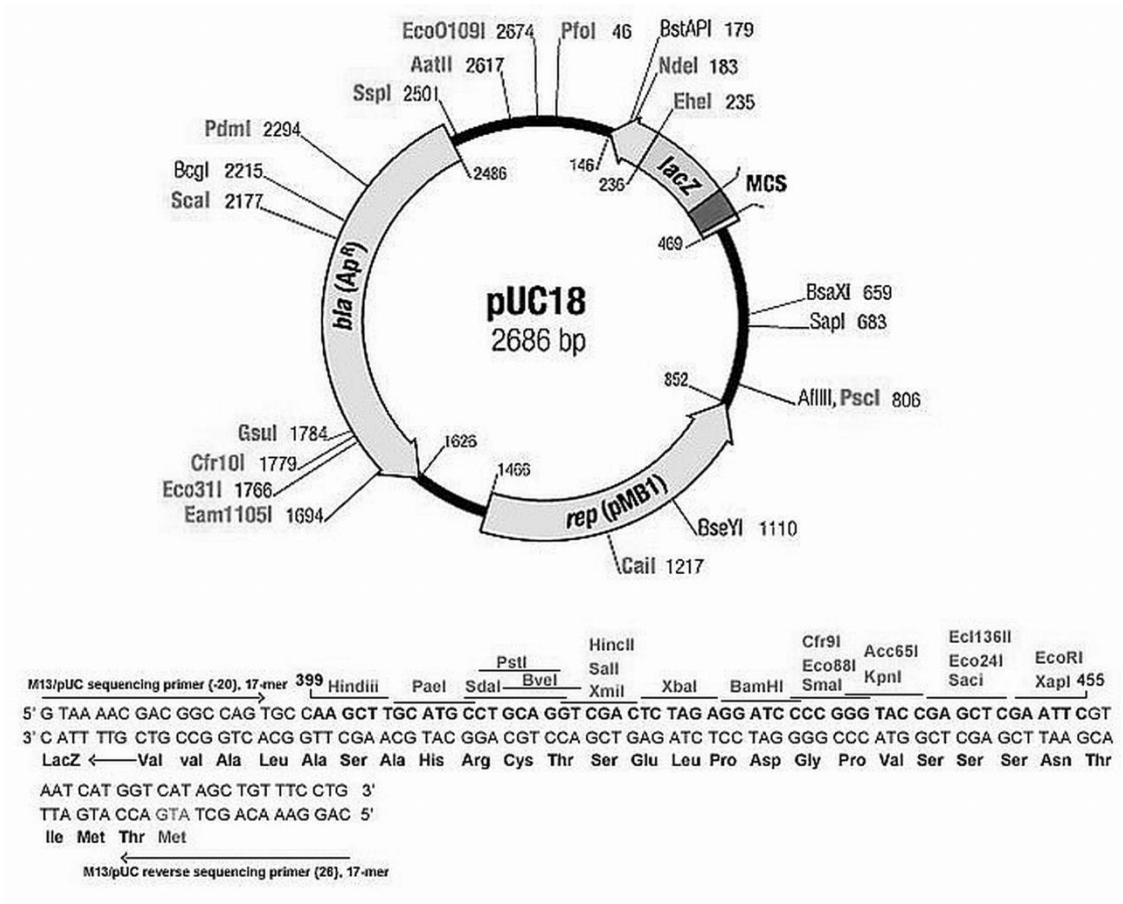
9.7 Anexo 7: Artigo aceito para publicação na revista *Applied and Environmental Microbiology*

9. ANEXOS

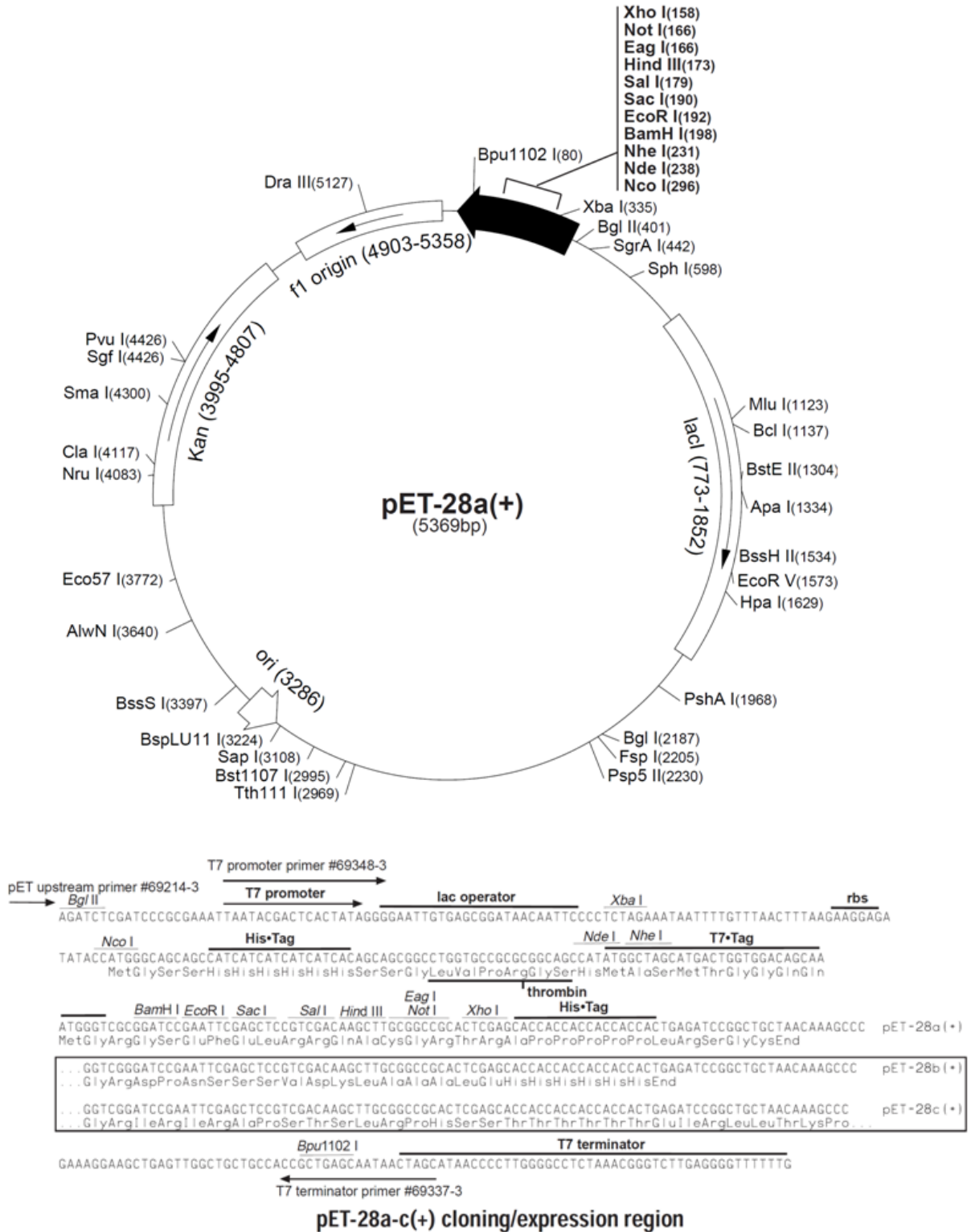
9.1 Anexo 1: Mapa do vetor pCC2FOS



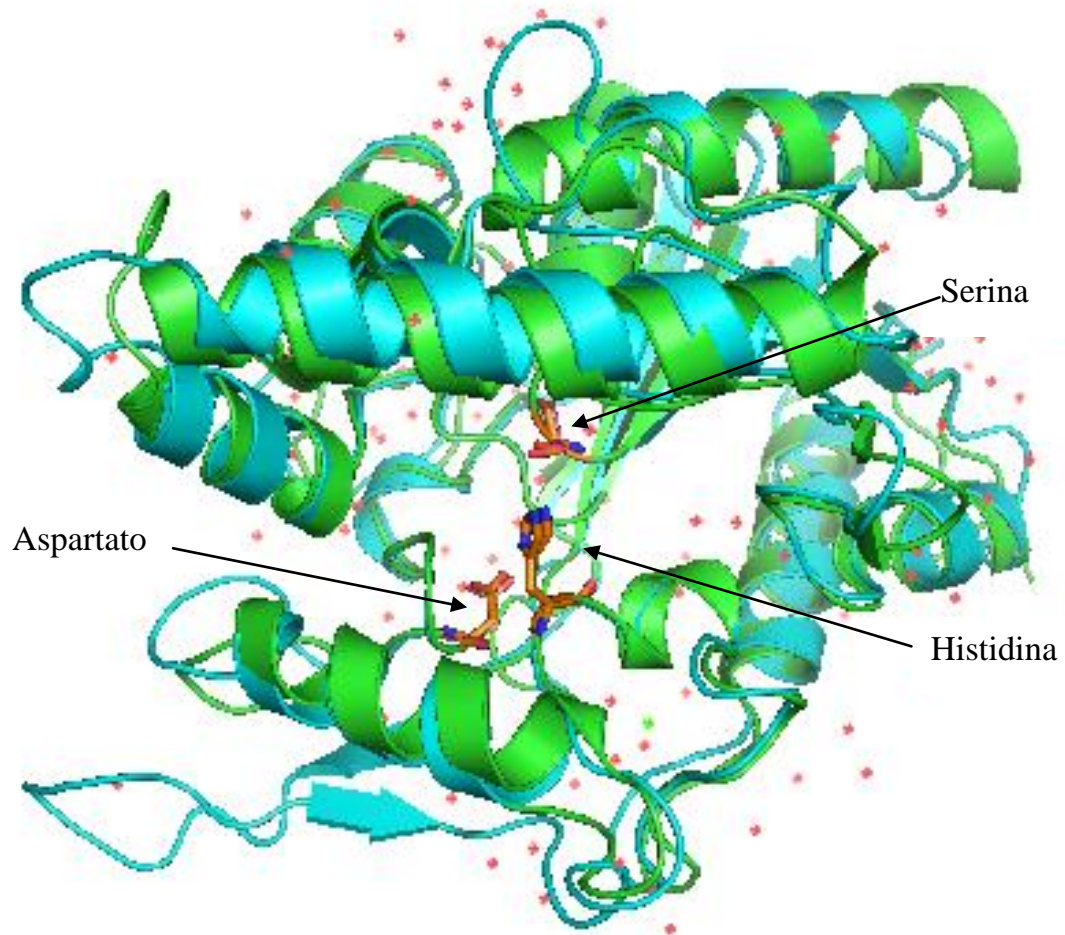
9.2 Anexo 2: Mapa do vetor pUC18



9.3 Anexo 3: Mapa do vetor pET28a

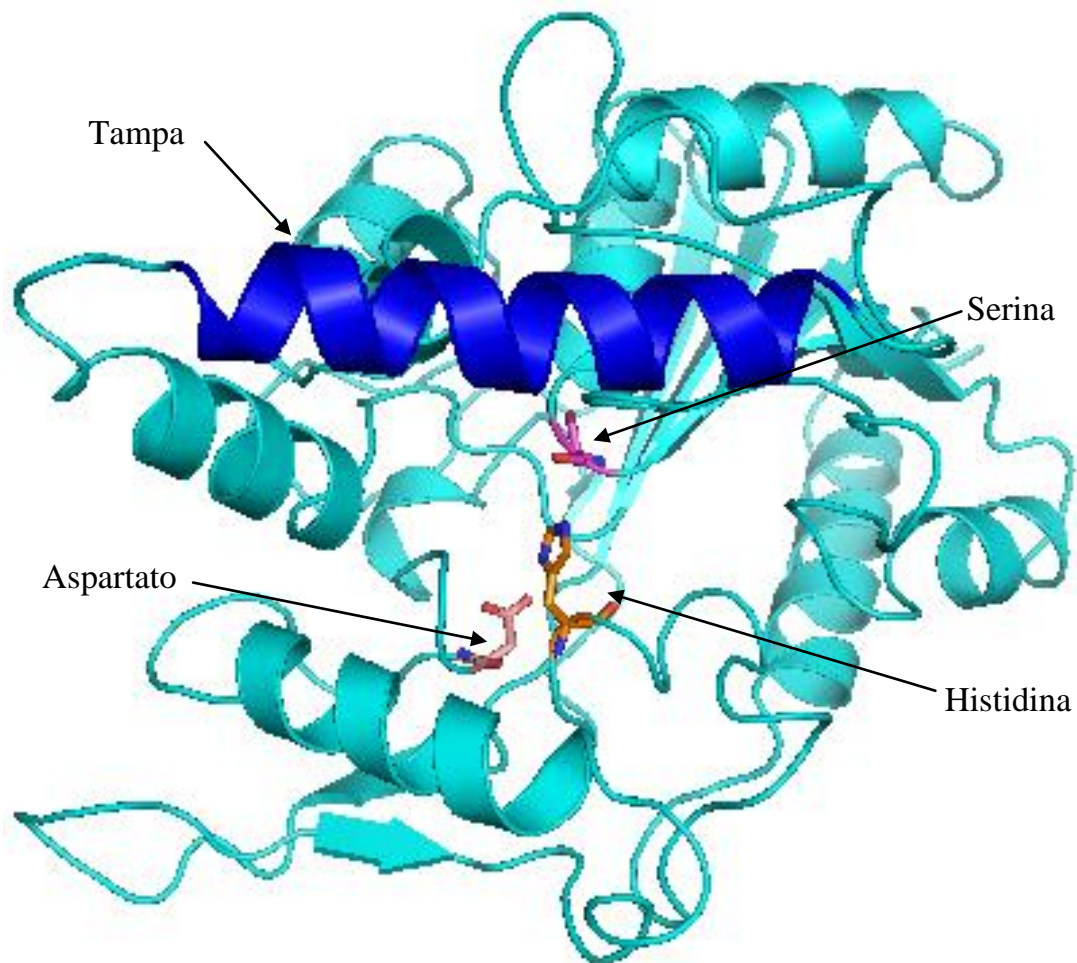


9.4 Anexo 4: Modelagem molecular da lipase LP001ORF27



Sobreposição da estruturas da lipase LP001ORF27 com a lipase de *Pseudomonas aeruginosa*. A modelagem da estrutura da lipase LP001ORF27 foi realizada com base na estrutura mais similar depositada no *Protein Data Bank* (*P. aeruginosa* - PDB ID 1EX9) utilizando o programa MODLER. A manipulação do modelo foi realizada no programa PYMOL. A estrutura da lipase LP001ORF27 é apresentada na cor ciano e enquanto a estrutura molde de *P. aeruginosa* é apresentada em verde. Os resíduos que formam a tríade catalítica estão indicados.

9.5 Anexo 5: Modelo molecular da lipase LP001ORF27



A modelagem da estrutura da lipase LP001ORF27 foi realizada com base na estrutura mais similar depositada no *Protein Data Bank* (*P. aeruginosa* - PDB ID 1EX9) utilizando o programa MODLER. A manipulação do modelo foi realizada no programa PYMOL. Os resíduos que formam a tríade catalítica estão indicados. Através dessa análise também foi possível identificar a região da tampa que cobre o sítio ativo (*lid*) destaca em azul escuro.

9.6 Anexo 6: Identificação e análise de clones resistentes a antibióticos

9.6.1 Mecanismos de resistência a antibióticos

Genes que codificam proteínas capazes de conferir ao micro-organismo algum grau de resistência a antibióticos têm grande importância médica, tanto quanto as vias de produção de antibióticos. Cada vez mais bactérias resistentes aos mais avançados tipos de antibióticos vêm surgindo em ambientes hospitalares, afetando o desenvolvimento de doenças infecciosas e aumentando os riscos de tratamento com imunossupressores (MARTINEZ et al, 2009). Atualmente, os principais mecanismos de resistência a antibióticos encontrados em bactérias são divididos em três grupos: (1) bombas de efluxo de drogas, (2) enzimas capazes de modificar a molécula de antibiótico através de fosforilação, adenililação ou acetilação e (3) enzimas hidrolíticas, como a β -lactamase, capazes de degradar o antibiótico.

Na natureza, acreditava-se que os antibióticos e os mecanismos de resistência eram utilizados pelas bactérias unicamente na competição pela sobrevivência como ataque e defesa. Recentemente, descobriu-se que em baixas concentrações os antibióticos possuem função sinalizadora mediando a interação entre as diferentes espécies bacterianas e regulando a dinâmica da comunidade (LINARES et al, 2006). Outro trabalho mostrou que bactérias do solo não apenas resistem aos antibióticos, mas também são capazes de utilizá-los como fonte carbono (DANTAS et al, 2008). Considerando a imensa diversidade das bactérias do solo, não é surpresa que mais de 80% dos antibióticos utilizados derivam de bactérias do solo como resultado direto da produção pela bactéria de um composto natural ou pela produção pelo homem de um composto semi-sintético (D’COSTA, GRIFFITHS e WRIGHT, 2007). Geralmente, um organismo que produz um tipo de antibiótico também possui meios de sobreviver à ação deletéria desse composto. Essa relação entre síntese e resistência ao antibiótico é tão próxima que é comum os genes que formam esses dois sistemas estarem organizados em operons ou muito próximos no genoma da bactéria (LASKARIS et al, 2010). Sendo assim, para a grande maioria dos antibióticos disponíveis atualmente já

existe um mecanismo de resistência igualmente disponível na natureza (MARTINEZ, 2008). Mas qual a relação entre os mecanismos de resistência de bactérias do solo e dos patógenos humanos? Sabe-se que um grande número de patógenos oportunistas humanos, como *Burkholderia*, *Pseudomonas*, *Enterobacter* e *Staphylococcus* habitam também o solo. Devido à facilidade com que a transferência lateral de genes ocorre no solo e também ao fato de que plasmídeos de resistência podem ser transferidos entre bactérias filogeneticamente distantes, é provável que muitos genes de resistência encontrados em patógenos tenham origem em bactérias do solo (DAVIES, 1994). Um exemplo clássico dessa questão foi a descoberta de um sistema de resistência a vancomicina, um antibiótico de último recurso, em um actinomiceto no solo, que era muito similar àquele descoberto em bactérias de relevância clínica (MARSHALL et al, 1997).

A imensa maioria dos trabalhos com genes de resistência ou síntese de antibióticos tem sido realizada com organismos cultiváveis. Mas considerando-se que esses organismos representam uma porcentagem ínfima do total de espécies que existem, ainda há muito que ser investigado e a Metagenômica também tem sido utilizada nesse sentido. A identificação de clones em uma biblioteca metagenômica capazes de produzir compostos coloridos levou à descoberta dos antibióticos turbomicina A e turbomicina B (GILLESPIE et al, 2002). LIM e colaboradores (2005) investigaram clones de uma biblioteca metagenômica do solo capazes de produzir um pigmento roxo e descobriram que se tratava de indirubina e seu isômero índigo azul que também apresentavam atividade antimicrobiana. Já a prospecção por genes de resistência a antibióticos em uma biblioteca metagenômica do solo levou a identificação de 9 clones resistentes a antibióticos aminoglicosídicos (resistência por modificação enzimática) e 1 resistente a tetraciclina (resistência por efluxo) (RIESENFELD, GOODMAN e HANDELSMAN, 2004). DIAZ-TORRES e colaboradores identificaram um novo gene de resistência à tetraciclina, a partir de uma biblioteca metagenômica construída com DNA extraído da comunidade bacteriana oral humana (DIAZ-TORRES et al, 2003). Trabalhos como esses citados acima demonstram o potencial da Metagenômica também nessa área de pesquisa.

9.6.2 Preparo de meio seletivo para detecção de clones resistentes a antibióticos

A identificação de clones que apresentavam resistência a antibióticos foi realizada em meio sólido LA contendo diferentes antibióticos, utilizando *pools* de células. Essa prospecção foi realizada apenas nos clones da biblioteca MAF1. Para tanto, os clones foram crescidos em placas de meio sólido LA contendo 12,5 µg/mL de cloranfenicol durante 24 horas em estufa 37°C. Após o período de crescimento os 96 clones, correspondentes a uma placa da biblioteca, foram coletados para um tubo criogênico de 5 mL, com auxílio de uma haste de madeira estéril, e a massa celular foi suspensa em 5 mL de glicerol 50%. Esse procedimento foi repetido para as 360 placas da biblioteca MAF1 originando os *pools* 1.1 a 1.360. Em seguida 500 µL de cada *pool* de placa foram misturados de 10 em 10 originando os *pools* 2.1 a 2.36. Novamente, 400 µL dos *pools* 2 foram misturados 12 a 12 originando os *pools* 3.1 a 3.3. Por fim, 1,5 mL de cada *pool* 3 foi misturado em um único tubo originando o *pool* total, contendo células representativas dos 36 mil clones da biblioteca MAF1, que foi utilizado para a detecção de resistência a antibióticos. Para realização do experimento, 100 µL do *pool* total de células foi diluído em 10 mL de meio líquido LB e recuperado por 30 minutos em estufa a 37°C. Por fim, alíquotas de 500 µL de células diluídas foram espalhadas em placas de petri contendo meio de cultivo sólido LA com os diferentes antibióticos. A preparação e a concentração para uso de cada antibiótico (Tabela 7) foi realizada utilizando como base a concentração mínima inibitória descrita para *E. coli*, como descrito por ANDREWS, 2001.

TABELA A1: Antibióticos e concentração usados para identificação de clones resistentes.

Antibiótico	Classe	Concentração utilizada ^a
Cefadroxila	Cefalosporina 1º geração	30 µg/mL
Ceftriaxona	Cefalosporina 2º geração	1 µg/mL
Cefaclor	Cefalosporina 3º geração	20 µg/mL
Eritromicina	Macrolídio	50 µg/mL
Azitromicina	Macrolídio	30 µg/mL
Benzoilmetromidazol	Outros	40 µg/mL
Norfloxacino	Quinolona	1 µg/mL
Ciprofloxacino	Quinolona	1 µg/mL
Ácido Nalidixico	Quinolona	10 µg/mL
Sulfametoxazol+Trimetoprina	Sulfonamidas	40 µg/mL
Ampicilina	Penicilina	250 µg/mL
Amoxicilina	Penicilina	25 µg/mL
Tetraciclina	Tetraciclina	10 µg/mL
Rifampicina	Rifamicina	25 µg/mL
Estreptomicina	Aminoglicosídeos	80 µg/mL
Gentamicina	Aminoglicosídeos	10 µg/mL
Neomicina	Aminoglicosídeos	25 µg/mL
Canamicina	Aminoglicosídeos	100 µg/mL
Vancomicina ^b	Glicopeptídeo	300 µg/mL
D-Cicloserina	Antitubercular	50 µg/mL

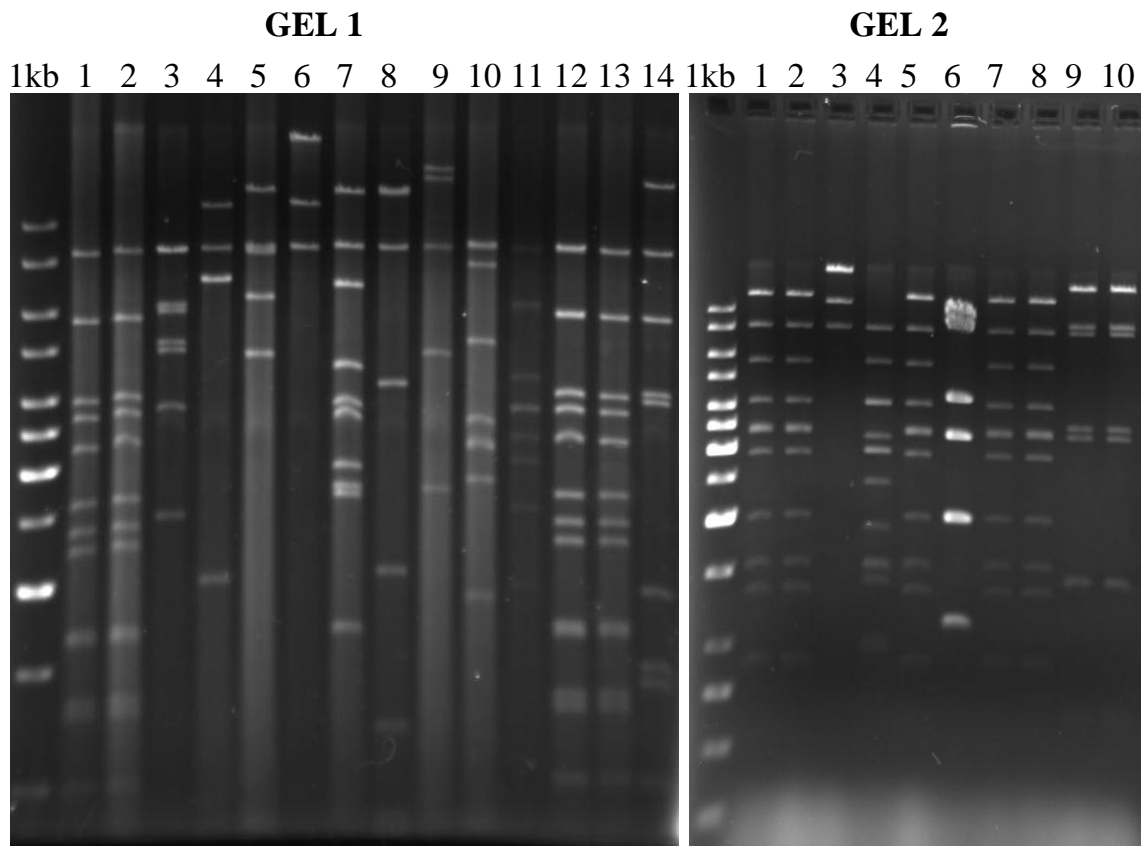
NOTA: ^a4 vezes a Concentração Mínima Inibitória para *E. coli* segundo ANDREWS, 2001;

^bConcentração mínima inibitória determinada empiricamente

9.6.3 Identificação de clones resistentes a antibióticos

Todos os clones que compõe a biblioteca MAF1 foram reunidos em um único pool celular. Essas células foram espalhadas em placas de meio sólido contendo diferentes tipos de antibióticos (Tabela A1). Os clones capazes de se multiplicar nesses meios sugerem a presença de um mecanismo de resistência ao antibiótico testado. Usando essa metodologia, foram identificados clones resistentes a rifampicina, ácido nalidíxico, ciprofloxacino, cefadroxila, norfloxacino e gentamicina. O maior número de clones resistentes foi encontrado nos antibióticos rifampicina e ácido nalidíxico, respectivamente 14 e 10 clones (respectivamente RF25-1 a RF25-14 e NAL10-1 a NAL10-10). O DNA fosmidial desses clones foi purificado e digerido com a enzima de restrição *Bam*HI. Para ambos os antibióticos foi possível notar uma repetição de perfil de restrição (Figura A1). Os clones RF25-1, 2, 12 e 13 apresentaram perfis de restrição idênticos. O mesmo aconteceu para os clones NAL10-1, 2, 5, 7 e 8 e NAL10-9 e 10. Esse resultado indica que a seleção desses clones não foi aleatória ou devida à resistência natural, reforçando a hipótese de que esses fósmeos possuem genes que codificam proteínas envolvidas na resistência a esses antibióticos. O fósmeo desses clones foi retransformado na estirpe hospedeira EPI300 para avaliação da manutenção do fenótipo. No entanto, apenas as bactérias retransformadas com o fósmeo RF25-1 foram capazes de manter o fenótipo. Esse clone foi selecionado para sequenciamento integral do inserto de DNA fosmidial, com o objetivo de se identificar o mecanismo de resistência.

FIGURA A1: Perfil de restrição do DNA fosmidial dos clones resistentes a rifampicina e ácido nalidíxico

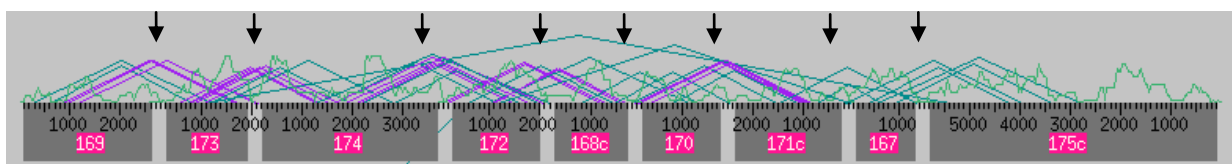


Eletroforese em gel de agarose 1% em tampão TBE 1X do DNA fosmidial dos clones resistentes a rifampicina (Gel 1) e ácido nalidíxico (Gel 2). Em ambos os casos ocorre uma repetição de padrões. Identificação das colunas. A identificação dos clones é mostrada diretamente no gel, assim como o marcador de peso molecular *1 kb Gene Ruler*,

9.6.4 Sequenciamento do inserto de DNA do clone RF25-1

O DNA fosmidial do clone RF25-1 foi purificado, nebulizado, reparado, e aplicado em gel de agarose de baixo ponto de fusão, a partir do qual foi feita a seleção de fragmentos de DNA com o tamanho desejado aproximado de 3 kb. Os fragmentos de DNA purificados foram ligados no vetor pUC18 e transformados na estirpe TOP10 de *E. coli*, originando uma sub-bibliotecas. Para identificação dos subclones, foi adicionado o prefixo “SC” antes do nome do clone original. Foi construída uma biblioteca com 384 subclones com tamanho de inserto de aproximadamente 3 kb. Todos os clones que compunham a sub-biblioteca SCRF25-1 foram sequenciados com os iniciadores universal e reverso, cujas sequências complementares estão presentes no vetor pUC18. Todas as sequências de DNA obtidas foram alinhadas e editadas utilizando-se os programas Phred-Phrap-Consed. No entanto, o número de sequências produzidas não foi suficiente para o fechamento integral da sequência do DNA inserto do clone RF25-1, nem para produção de um *Scaffold* confiável (Figura A2). Restaram ainda muitos buracos entre os contigs tornando inviável a utilização da técnica de transposon. Nesse caso, as análises foram realizadas utilizando-se as sequências parciais. Foram gerados 175 contigs, dos quais 9 foram utilizados para a anotação (contigs 169, 173, 174, 172, 168, 170, 171, 167 e 175). Os contigs que não foram utilizados para a anotação eram formados por poucas sequências (geralmente 2 leituras da mesma sequências), sequências de baixa qualidade ou ainda sequências de DNA da bactéria hospedeira *E. coli*.

FIGURA A2: Scaffold obtido para o clone RF25-1



Os números destacados nas caixas rosa se referem ao contig. Os números nas régua se referem às bases. As linhas ligando dois contigs seguidos (setas pretas) indicam clones cuja sequência possivelmente une os contigs a que se refere.

9.6.5 Caracterização genética do inserto de DNA do clone RF25-1

O inserto de DNA presente no fosmideo RF25-1 possui 25.149 pares de bases e um conteúdo G+C de 64%. De modo semelhante ao que foi feito com o fosmideo MAF1LP090, esses valores foram determinados com base somente nos contigs utilizados para a anotação. Nesse fragmento de DNA foram identificadas 24 ORFs (Tabela A2). O clone RF25-1 foi selecionado por apresentar um fenótipo de resistência à rifampicina. No entanto, a identificação dos genes que codificam as proteínas responsáveis pelo fenótipo não foi possível, mas, aparentemente, as ORFs 20 e 21 estão envolvidas nesse processo (denominadas RF25-1ORF20 e RF25-1ORF21 a partir desse ponto). A ORF 21 codifica uma proteína reguladora de transcrição similar a MarR. As proteínas da família Mar (*multiple antibiotic resistance*) estão envolvidas na resistência a múltiplos antibióticos e sua expressão é regulada pelo repressor MarR. A ORF 22, por sua vez, codifica uma proteína permease da família MFS (*Major Facilitator Superfamily*). Essa família de transportadores atua facilitando o transporte através da membrana de vários compostos incluindo íons, açúcares, peptídeos e drogas. É possível que ambas atuem juntas no mecanismo de resistência (SAIDIJAM, 2006; LEWINSON, 2006). Assim como em MAF1LP001, em RF25-1 também foram encontrados genes que codificam proteínas envolvidas na fotossíntese. As ORFs 1, 2 e 3 codificam, respectivamente, uma proteína acessória da rubisco, rubisco subunidade pequena e rubisco subunidade maior. Os organismos fotossintéticos produzem energia química na forma de ATP e NAD(P)H usando a luz do sol e fixam o CO₂ atmosférico em compostos orgânicos (açúcares) usando essa energia no ciclo de Calvin. A rubisco (ribulose 1,5 bifosfato carboxiliase oxigenase) é a enzima central que realiza a fixação do carbono atmosférico, através da carboxilação da ribulose 1,5 bifosfato para formação de duas moléculas de 3-fosfoglicerato (TABITA, 2008). A ORF 1, por sua vez, codifica um regulador transcricional (CbbX) das ORFs 2 e 3 sensíveis aos níveis de luz e oxigênio. Esse resultado sugere que o organismo original possa também ser fotossintético. Foram identificadas 4 proteínas hipotéticas, mas apenas uma não possui domínios conservados.

A sub-biblioteca SCRF25-1 foi testada em placas de meio sólido contendo rifampicina (25 µg/mL), mas nenhum subclone resistente foi recuperado. De modo semelhante ao fosmídeo MAF1LP018, é possível que as duas ORFs selecionadas como candidatas responsáveis pela resistência ao antibiótico em RF25-1 sejam necessárias para expressão do fenótipo (ORFs 20 e 21 na Tabela 17). Considerando o tamanho médio do inserto nessa sub-biblioteca de 3 kb, é possível que não haja nenhum subclone com as duas ORFs completas.

TABELA A2: Comparação e análise das ORFs codificadas a partir do inserto de DNA do clone RF25-1

ORFs	GC%	^a Tamanho (aa)	^b Função	^b Organismo	^b Nº de acesso	^b % identidade/positivos	^b E-value	^c Domínios conservados
Contig 169								
ORF1	63,5	268p	rubisco accessory protein CbbX, AAA ATPase	<i>Ralstonia eutropha H16</i>	YP_840912	78/85	1,00E-97	sim
ORF2	62,4	138c	ribulose 1,5-bisphosphate carboxylase small subunit	<i>Burkholderia xenovorans LB400</i>	YP_552889	68/81	8,00E-52	sim
ORF3	62,3	434p	Ribulose bisphosphate carboxylase large chain	<i>Xanthobacter flavus</i>	P23011	81/88	0,00E+00	sim
Contig 173								
ORF3	58,3	102p	Ribulose bisphosphate carboxylase large chain	<i>Methylibium petroleiphilum</i>	YP_001021971	57/68	6,00E-19	sim
ORF4	64,8	298c	LysR family transcriptional regulator	<i>Palaromonas naphthalenivorans</i>	YP_982209	47/65	4,00E-72	sim
ORF5	68,7	164c	hypothetical protein HNE_2878	<i>Hyphomonas neptunium</i>	YP_761556	42/54	1,00E-18	sim
Contig 174								
ORF6	63,1	225p	GMP synthase, large subunit	<i>Thioalkalivibrio sp. HL-EbGR7</i>	YP_002513110	28/44	2,70E+00	não
ORF7	63,5	623c	oligoendopeptidase F	<i>Optutus terrae PB90-1</i>	YP_001819975	46/62	1,00E-150	sim
ORF8	62,5	193c	Ycel family protein	<i>Acidovorax ebreus TPSY</i>	YP_002552505	52/69	5,00E-45	sim
ORF9	65,1	87c	hypothetical protein Mpe_A2578	<i>Methylibium petroleiphilum</i>	YP_001021766	43/58	2,00E-12	não
Contig 172								
ORF10	64,6	429c	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	<i>Palaromonas naphthalenivorans</i>	YP_980297	62/73	9,00E-142	sim
ORF11	64,5	270p	Aldehyde ferredoxin oxidoreductase	<i>Thiomonas intermedia K12</i>	YP_05500988	81/91	6,00E-131	sim
Contig 168c								
ORF12	62,8	267p	Aldehyde ferredoxin oxidoreductase	<i>Thiomonas intermedia K12</i>	YP_05500988	82/89	2,00E-125	sim
ORF13	60,1	157c	4Fe-4S ferredoxin iron-sulfur binding domain protein	<i>Thiomonas intermedia K12</i>	YP_05500989	87/94	7,00E-78	sim
ORF14	63,2	124c	putative transcriptional regulator, ModE family	<i>Acidothibacillus ferrooxidans</i>	YP_002219328	51/68	1,00E-24	sim
Contig 170								
ORF15	64,1	232c	binding-protein-dependent transport systems inner membrane component	<i>Accumulibacter phosphatis</i>	YP_003168170	70/83	6,00E-76	sim
ORF16	63,1	214c	ABC transporter related	<i>Accumulibacter phosphatis</i>	YP_003168171	51/65	8,00E-53	sim
Contig 171								
ORF17	65,4	329c	Serine O-acetyltransferase	<i>Methylibium petroleiphilum</i>	YP_001019223	65/78	4,00E-110	sim
ORF18	68,3	318c	putative molybdopterin biosynthesis-related protein	<i>Methylobacterium extorquens</i>	YP_002964075	54/66	3,00E-80	sim
Contig 167								
ORF19	65,3	455p	hypothetical protein mma_3262	<i>Janthinobacterium sp.</i>	YP_001354952	40/61	5,00E-92	sim
Contig 175								
ORF20	64,8	301p	MFS permease	<i>Bordetella petrii</i>	YP_001630812	56/71	1,00E-79	sim
ORF21	67,9	155c	MarR family transcriptional regulator	<i>Palaromonas sp.</i>	YP_550351	45/62	8,00E-25	sim
ORF22	64,9	668c	hypothetical protein H16_A1279	<i>Ralstonia eutropha H16</i>	YP_725787	62/75	0	sim
ORF23	65,1	358c	40-residue YVTN family beta-propeller repeat protein	<i>Ralstonia eutropha H16</i>	YP_002982051	54/74	1,00E-103	sim
ORF24	65,1	385c	FMN-dependent alpha-hydroxy acid dehydrogenase	<i>Burkholderia multivorans</i>	YP_001585750	78/90	0,00E+00	sim

NOTA: ^aSequências parciais (p) e completas (c) estão indicadas; ^bBaseado no melhor resultado de BLAST usando BLASTp e o banco de dados de proteínas não redundantes; ^cBaseado no banco de dados InterPRO

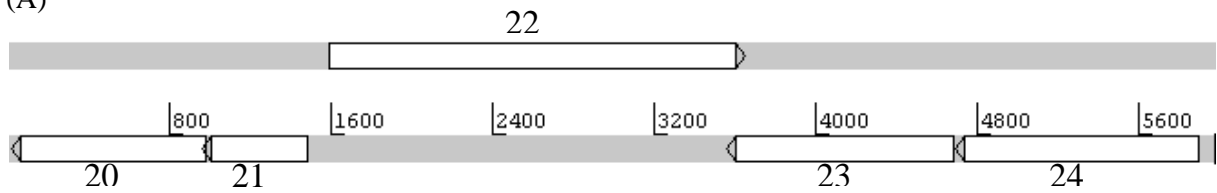
9.6.6 Análise de sequência das proteínas que conferem resistência a antibiótico

As proteínas RF25-1ORF20 e 21, presentes no contig 175 (Figura A3A), foram selecionadas como candidatas responsáveis pelo fenótipo de resistência a antibióticos. A sequência de aminoácidos da proteína RF251ORF21 está completa e apresenta similaridade com a proteína MarR, um repressor transcricional dos genes da família *mar* (*multiple antibiotic resistance*). Essa proteína possui 155 resíduos de aminoácidos, massa molecular de 17.221,7 Da e pI teórico de 10,55. Ela possui, ainda, um elevado conteúdo de arginina (12,3%), ficando atrás apenas do conteúdo de leucina (13,5%). Para *E. coli* a arginina é o aminoácido mais problemático em termos de tradução com 4 dos 6 códons usados ocorrendo a frequências abaixo de 1% (AGG, AGA, CGA e CGG). Na proteína em questão há 19 resíduos de arginina: 1 codificado por AGG, 8 por CGG e 10 por CGC. Dentre esses, o códon mais raro é CGG é que usado apenas 0,54% das vezes (MCNULTY et al, 2003). Não foram encontrados peptídeo sinal nem regiões transmembrana. A busca por domínios conservados identificou dois principais: o domínio MarR e o domínio hélice-volta-hélice na conformação alada (*winged helix-turn-helix* - wHTH) (Figura A3B). Esse domínio de ligação wHTH é formado por 2 α -hélices mais 3 folhas β (asas) organizados da forma $\beta 1-\alpha 3-\alpha 4-\beta 2-W1(asa)-\beta 3$ (ALEKSHUN et al, 2001). As proteínas Mar fazem parte de um sistema não específico de resistências a antibióticos estando a expressão sob controle do regulador MarR. Acredita-se que o antibiótico se ligue ao repressor inativando-o e permitindo a expressão dos genes *mar* responsáveis pela multirresistência. A proteína RF25-1ORF20 foi parcialmente sequenciada e codifica uma proteína similar a uma proteína transportadora de membrana da família MFS (*Major Facilitator Superfamily*). A sequência de aminoácidos recuperada possui 301 resíduos da região N-terminal, enquanto sua contraparte no banco de dados (*best BLAST hit*) possui 419 resíduos. A busca por regiões conservadas identificou, além do domínio MFS, um peptídeo sinal, formado pelos 24 primeiros aminoácidos, e 8 regiões transmembrana reforçando a hipótese dessa ORF representar uma proteína integral de membrana (Figura A3C).

Os genes que codificam as proteínas RF25-1ORF20 e 21 são transcritos na mesma orientação e estão distantes 34 pares de bases. O gene que codifica a ORF 22 é transcrito na outra orientação e esta 122 bases distante do códon de início de tradução de ORF21. Devido ao sequenciamento incompleto, não foi possível afirmar qual gene está após a ORF20 e a partir de qual fita ele é transcrito. Baseado nessa análise, somado ao fato de que não foi possível recuperar um subclone positivo em pUC e na relação de função, é possível sugerir que ORF20 e 21 estão organizados em um operon e de fato sejam os responsáveis pela resistência à rifampicina.

FIGURA A3: Análise de sequência do contig RF25-1

(A)



Anotação do contig RF25-1 175 pelo programa Artemis. O número da ORF, a posição e a orientação do gene estão indicados. Os números na régua se referem ao número de bases.

(B)

SEQUENCE: [Sequence_1](#) CRC64: 73A4C0D35CB22C4F LENGTH: 155 aa

InterPro IPR000835 Family	HTH transcriptional regulator, MarR	
	PF01047	MarR
	SM00347	HTH_MARR
	PS50995	HTH_MARR_2
InterPro IPR011991 Domain	Winged helix-turn-helix transcription repressor DNA-binding	
	G3DSA:1.10.10.10	no description

Busca por regiões conservadas na ORF21 pelo programa InterproScan. Os domínios MarR e wHTH foram identificados.

(C)

SEQUENCE: [Sequence_1](#) CRC64: 3537306343F9591C LENGTH: 301 aa

InterPro IPR011701 Domain	Major facilitator superfamily MFS-1	
	PF07690	MFS_1
	PS50850	MFS
SignalP		signal-peptide
tmhmm		transmembrane_regions

Busca por regiões conservadas na ORF20 pelo programa InterproScan. O domínio MFS, um peptídeo sinal e oito regiões transmembrana foram identificados.

9.6.7 Discussão sobre o clone RF25-1

O clone RF25-1 foi capaz de crescer em placa de meio sólido contendo rifampicina na concentração de 25 µg/ml. O perfil de restrição apresentado pelo inserto de DNA desse clone foi selecionado aleatoriamente mais de uma vez, indicando que, de fato, esse fragmento de DNA estava expressando proteínas de resistência. As ORFs 21 (repressor do sistema mar, MarR) e 20 (transportador MFS) foram identificadas como responsáveis pelo fenótipo. Em *E. coli*, o repressor MarR é codificado junto com outras duas proteínas MarA e MarB formando um operon *marRAB*, onde *marA* codifica um ativador de transcrição e *marB* codifica uma proteína de função desconhecida (COHEN, HÄCHLER e LEVY, 1993; WHITE et al, 1997). Em condições normais, MarR está ligada ao operador dos genes *mar* impedindo sua transcrição. Quando o antibiótico está presente ele se liga à MarR, formando um complexo que não é mais capaz de ligar DNA permitindo a expressão de *marA*, que irá ativar os genes responsáveis pelo fenótipo de resistência, que, no caso de *E. coli*, são mais de 60 (BARBOSA e LEVY, 2000). Entre os genes regulados por esse operon estão proteínas do metabolismo geral, reparo de DNA e tradução, resposta a estresse ácido e superóxido, resistência a antibióticos e solventes orgânicos. Essa mudança no padrão de expressão gênica resulta na diminuição do número de proteínas de influxo associadas à membrana e aumento de proteínas de efluxo (principalmente da família MFS). A superfamília de proteínas MFS é um grupo diverso de transportadores que incluem uniporter, simporter e antiportes. Essas proteínas realizam o transporte de uma grande variedade de substratos, incluindo antibióticos, e provavelmente estão atuando no fenótipo de resistência de RF25-1. Entretanto, o que chama a atenção nesse clone é que, aparentemente, o fenótipo de resistência está sendo obtido somente com a presença das ORFs 21 e 20, ao contrário dos 60 genes ativados em *E. coli*. Além disso, os outros dois genes (*marA* e *marB*) que, junto com *marR*, formam o operon clássico *marRAB* não foram encontrados, indicando um novo tipo de organização para o sistema Mar.

No inserto de DNA de RF25-1 foram encontrados dois genes que codificam proteínas similares às subunidades I e II da enzima RubisCO. Essa enzima é responsável pela fixação do carbono através do Ciclo de Calvin durante a etapa final da fotossíntese. No entanto, essa enzima já foi identificada em organismos não fotossintéticos, de modo que não é possível afirmar que o organismo original realiza a fotossíntese.

A prospecção por genes de resistência a antibióticos identificou 24 clones, 14 resistentes à rifampicina e 10 resistentes ao ácido nalidíxico. A análise do perfil de restrição dos fósmites desses clones mostrou que há repetição de padrões, indicando que a seleção não foi aleatória ou foi devido à resistência natural da bactéria hospedeira. No entanto, quando esses fósmites foram retransformados na estirpe hospedeira EPI300, apenas o fósmito do clone RF25-1 foi capaz de sustentar o fenótipo de resistência. A análise da sequência de um dos prováveis genes envolvidos no fenótipo de resistência a antibiótico (repressor MarR) mostrou que ele possui um grande número de códons para arginina, sendo alguns muito raros em *E. coli*. É possível que essa característica esteja afetando, de algum modo, a tradução desse gene em proteínas tornando instável a detecção do fenótipo.

9.7 Anexo 7: Artigo aceito para publicação na revista *Applied and Environmental Microbiology*