

**Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”**

**Elementos genéticos móveis no microbioma dos sedimentos de  
manguezais**

**Filipe Rafael Salvetti Nunes**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração:  
Microbiologia Agrícola

**Piracicaba  
2016**

**Filipe Rafael Salvetti Nunes**  
**Bacharel em Ciências Biológicas**  
**Licenciado em Ciências Biológicas**

**Elementos genéticos móveis no microbioma dos sedimentos de manguezais**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:  
Prof. Dr. **FERNANDO DINI ANDREOTE**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração:  
Microbiologia Agrícola

**Piracicaba**  
**2016**

**Dados Internacionais de Catalogação na Publicação  
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Nunes, Filipe Rafael Salvetti

Elementos genéticos móveis no microbioma dos sedimentos de manguezais / Filipe Rafael Salvetti Nunes. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

52 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Bacteriófago 2. Integrase 3. Solos 4. Evolução 5. THG I. Título

CDD 631.46  
N971e

**"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"**

## DEDICATÓRIA

Ao meu avô Pedro Salvetti pelo companheirismo e amizade, e à memória de minha avó Zilla de Barros Salvetti, por seu amor e exemplo de vida, dedico.

## AGRADECIMENTOS

À Michelle Barreto pelo carinho, apoio e incentivo,

A toda minha família, sem a qual jamais chegaria até aqui,

Ao meu orientador Fernando Dini Andreote pela aposta em minha autonomia e sábios conselhos,

À Simone Raposo Cotta pelo acompanhamento e contribuições ao longo desses dois anos,

A toda a equipe do Laboratório de Microbiologia do Solo, técnicos e alunos, com quem passei ótimos momentos,

À Fernanda, secretária do Programa de Microbiologia Agrícola, que me salvou mais de uma vez,

Ao Armando Cavalcante, “Polé” Lourenço, Luana Lira, Lucas Taniguti, Claudia Barros Monteiro e Alessandro Varani pelo suporte técnico e suas valorosas contribuições,

À Universidade de São Paulo pelo apoio institucional,

Ao CNPq pelo financiamento da bolsa,

Ao CBAB e aos colegas da Universidad Nacional de Quilmes, em Buenos Aires, pela oportunidade de minha primeira experiência internacional,

Aos meus colegas de Ponte e Curió por tantos momentos de aprendizado,

A todos que de alguma forma colaboraram para a realização desse trabalho,

Meu sincero e profundo agradecimento.



## SUMÁRIO

|  |    |
|--|----|
| RESUMO .....   | 7  |
| ABSTRACT .....   | 9  |
| SUMÁRIO.....   | 5  |
| 1 INTRODUÇÃO.....  | 11 |
| 2 REVISÃO BIBLIOGRÁFICA .....  | 13 |
| 2.1 Os manguezais.....   | 13 |
| 2.2 Elementos Genéticos Móveis .....   | 13 |
| 2.3 Bacteriófagos.....   | 15 |
| 3 OBJETIVOS E HIPÓTESES .....  | 19 |
| 3.1 Objetivo geral.....  | 19 |
| 3.2 Objetivos específicos.....   | 19 |
| 3.3 Hipóteses .....  | 19 |
| 4 MATERIAL E MÉTODOS.....  | 21 |
| 4.1 Áreas dos manguezais estudados .....   | 21 |
| 4.2 Coleta das amostras de sedimento.....  | 22 |
| 4.3 Extração de DNA e mRNA e sequenciamento das amostras de sedimento.....                                   | 22 |
| 4.4 Preparo das sequências de metagenoma e metatranscriptoma dos sedimentos para anotação .....              | 23 |
| 4.5 Anotação automática das sequências de DNA e mRNA dos sedimentos e análises funcionais no MG-RAST .....   | 23 |
| 4.6 Sequências da biblioteca de fosmídeo e montagem dos <i>contigs</i> .....                                 | 24 |
| 4.7 Predição de fagos e profagos a partir das sequências da biblioteca de fosmídeo ..                        | 24 |
| 4.8 Anotação manual de genes dos <i>contigs</i> da biblioteca de fosmídeo que contém fagos ou profagos ..... | 25 |
| 5 RESULTADOS E DISCUSSÃO .....   | 27 |
| 5.1 Abundância e expressão dos EGMs no solo .....  | 27 |
| 5.2 Anotação de bacteriófagos a partir das sequências da biblioteca de fosmídeo .....                        | 35 |
| 6 CONSIDERAÇÕES FINAIS .....   | 47 |
| REFERÊNCIAS .....  | 49 |



## RESUMO

### Elementos genéticos móveis no microbioma dos sedimentos de manguezais

Os manguezais são ecossistemas costeiros de transição formados entre o ambiente marinho e terrestre de zonas tropicais e intertropicais, e estão sujeitos ao regime de marés. Devido a tais condições, os microrganismos que vivem nos sedimentos sofrem constante pressão adaptativa. Nesse contexto, os Elementos Genéticos Móveis (EGMs) (fagos, plasmídeos e transposons) são fatores genéticos diretamente relacionados com a transferência horizontal de genes e podem facilitar processos de adaptação. Neste trabalho foram estudados os EGMs dos sedimentos de três áreas de manguezais do estado de São Paulo, duas localizadas no município de Bertioga e uma na reserva ambiental da Ilha do Cardoso, em Cananeia. Foram avaliadas a frequência e expressão dos diferentes tipos de EGMs no microbioma dos manguezais a partir de metagenomas e metatranscriptomas dos sedimentos, além da utilização de uma biblioteca de fósmites, montada a partir de DNA do sedimento de uma das áreas de Bertioga. As sequências de DNA e mRNA foram obtidas de três pontos amostrados por área, sequenciadas em equipamento Illumina HiSeq2000 e anotadas automaticamente na plataforma do MG-RAST. Para a biblioteca de fósmites, foi sequenciado um *pool* de todos os clones em equipamento Illumina HiSeq 2000. *Contigs* dessas sequências foram montados e fagos e profagos foram preditos e anotados automaticamente pelo *pipeline* VirSorter. As sequências preditas como fagos ou profagos foram anotadas manualmente. A partir da análise das sequências de metagenômica foi possível verificar que a participação dos EGMs no metabolismo total do microbioma dos manguezais corresponde de 6 a 15 % dos genes preditos, sendo que fagos e profagos correspondem a 90% dos genes associados aos EGMs. A partir dos contigs montados para as sequências da biblioteca, foram preditos e anotados automaticamente 27 fagos, dos quais nove fagos foram anotados manualmente em etapa posterior. Foram encontrados genes que codificam proteínas típicas de fagos, como capsídeo, transposase e integrase, além de genes acessórios potencialmente importantes, como relacionados ao sistema de transporte ABC e sistema de transporte tipo II. A partir da anotação manual, genes conservados de fagos foram utilizados como base para predição da linhagem viral. Foram identificados 6 tipos de vírus pertencentes à ordem Caudovirales, que infectam Proteobacteria e Firmicutes. Esse trabalho sustenta a hipótese de que os fagos infectam os grupos bacterianos mais abundantes nos manguezais e podem carregar genes importantes consigo, desempenhando papel chave na evolução das bactérias nesse ambiente.

Palavras-chave: Bacteriófago; Integrase; Solos; Evolução; THG



## ABSTRACT

### **Mobile genetic elements in mangrove sediments microbiome**

Mangroves are coastal transitional ecosystems between marine and terrestrial environments of tropical and intertropical zones and are subject to tidal regime. Due those conditions, microorganisms that lives in the sediments suffer constant adaptative pressure. In this context, Mobile Genetic Elements (MGEs) (phages, plasmids and transposons) are genetic fators directly related with horizontal gene transfer and can facilitate adaptation processes. In this work MGEs were studied in the sediments of three mangrove areas in São Paulo state, two located in Bertioga county and one in environmental reserve of Ilha do Cardoso, in Cananeaia. Frequency and expression were avaliated for diferente EGMs types in mangrove microbiome from the sediments metagenomes and metatranscriptomes, besides utilization os a fosmid library assembled from a Bertioga area sediment DNA. The sequences of DNA and mRNA were obtained from three sampled points, sequenced in Illumina HiSeq2000 equipment and automatically anotaded in MG-RAST plataform. A pool of all clone were sequenced for fosmid library in Illumina HiSeq2000 equipment. Contigs were assembled and phages and prophages were automatically predicted in VirSorter pipeline. The phages or prophages predicted sequences were manually anotaded. From the analisys os metagenomic sequences was possible to verify that the MGEs participation in total microbiome metabolism stands for 6 to 15% of predicted genes, wich 90% corresponds to phage associated genes. From contigs assembled of library sequences, 27 phages were predicted and anotaded, wich nine code for phage typical proteins, as capsid, transposase and integrase, besides accessory genes potencially important, as ABC transporter and type II transport related. Phage conserved genes were used as base for viral lineage prediction. 6 viral types belonging do Caudovirales order were identified, which infects Proteobacteria and Firmicutes. This work supports the hypothesis that phages infect the most abundant bacterial groups in mangroves and can carry important genes playing a key role in bacterial evolution in this environment.

Keywords: Bacteriophage; Integrase; Soil; Evolution; HGT



## 1 INTRODUÇÃO

Os manguezais são ecossistemas costeiros de transição formados entre o ambiente marinho e terrestre de zonas tropicais e intertropicais, e está sujeito ao regime de marés. A alta salinidade, a inundação intermitente e as influências antrópicas diversas nos manguezais são condições que mantêm uma pressão adaptativa constante sobre os microrganismos que vivem nos sedimentos. Nesse contexto, os EGMs (fagos, plasmídeos e transposons) são fatores genéticos diretamente relacionados com a transferência horizontal de genes (THG), modulando genomas, e conseqüentemente facilitando processos de adaptação dos microrganismos às alterações nas condições do ambiente.

Trabalhos recentes têm revelado aspectos da estruturação da comunidade microbiana de manguezais e de seu funcionamento, mas os mecanismos genéticos pelos quais essa comunidade adapta-se ao ambiente permanecem pouco explorados. Sabe-se que além do processo de geração de variabilidade genética por mutação, ocorre também a transferência de genes entre indivíduos de espécies diferentes, no processo conhecido como a transferência horizontal de genes (THG). Este processo é intermediado principalmente pelas atividades de conjugação, transdução e transformação, os quais são ainda pouco explorados no contexto da microbiologia ambiental.

Esse trabalho buscou avaliar a presença, abundância e expressão dos diferentes tipos de EGMs (bacteriófagos, plasmídeos e transposons) no microbioma dos sedimentos de três manguezais do estado de São Paulo a partir de dados de metagenômica e metatranscritômica. Adicionalmente, sequências de uma biblioteca de fosmídeo construída para uma das áreas foram exploradas para uma análise mais detalhada a cerca dos bacteriófagos.



## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Os manguezais

Manguezais são ecossistemas costeiros localizados entre o ambiente marinho e a restinga, nas zonas tropicais e subtropicais do planeta. O Brasil apresenta aproximadamente 12 mil ha de sua costa originalmente coberta por manguezais, apresentando a terceira maior área desse tipo de ecossistema no mundo. Os sedimentos de manguezais apresentam características físicas e químicas muito particulares devido à influência da inundação intermitente pelas marés, principalmente o alto potencial redutor e alta salinidade (SCHAEFFER-NOVELLI et al., 2000). No estado de São Paulo, as florestas de manguezais são compostas essencialmente por apenas três espécies de árvores: *Avicennia shaueriana*, *Laguncularia racemosa* e *Rhizophora mangle*, mas o manguezal abriga uma enorme diversidade de fauna, funcionando como berçário para muitas espécies de peixes e crustáceos (BERNINI; REZENDE, 2011).

Quanto aos microrganismos, uma série de trabalhos recentes já revelaram aspectos da estruturação da comunidade microbiana para fungos (FASANELLA et al., 2012), arqueias (FRANCO DIAS et al., 2012) e bactérias (DIAS et al., 2011), utilizando tanto técnicas de isolamento e cultivo quanto técnicas independentes de cultivo, como DGGE, sequenciamento e PCR quantitativo para genes específicos, e também metagenômica e metatranscriptômica (CADETE, 2014). Os grupos de bactérias predominantes nos sedimentos dos manguezais brasileiros são pertencentes às Deltaproteobacteria e Gammaproteobacteria (DIAS et al. 2010), que também foram descritas como as metabolicamente mais ativas em um estudo de metatranscriptômica para esse ambiente (CADETE, 2014).

### 2.2 Elementos Genéticos Móveis

A THG é uma das principais forças evolutivas presentes na natureza e desempenha papel fundamental na adaptação dos genomas microbianos ao ambiente. Os EGMs são os principais agentes de THG, mas seu papel na organização dos genomas e no intercâmbio dos genes entre organismos. Porém, seu papel na adaptação e resposta da microbiota do solo as diferentes condições ambientais ainda está para ser melhor descrito. EGMs são segmentos de DNA que codificam enzimas e outras proteínas capazes de mediar o movimento de DNA

dentro do genoma (mobilidade intracelular) ou entre células (mobilidade extracelular). Tradicionalmente são consideradas três formas de transferência de DNA em procariotos: conjugação, transformação e transdução (Figura 2.1) (FROST et al., 2005).

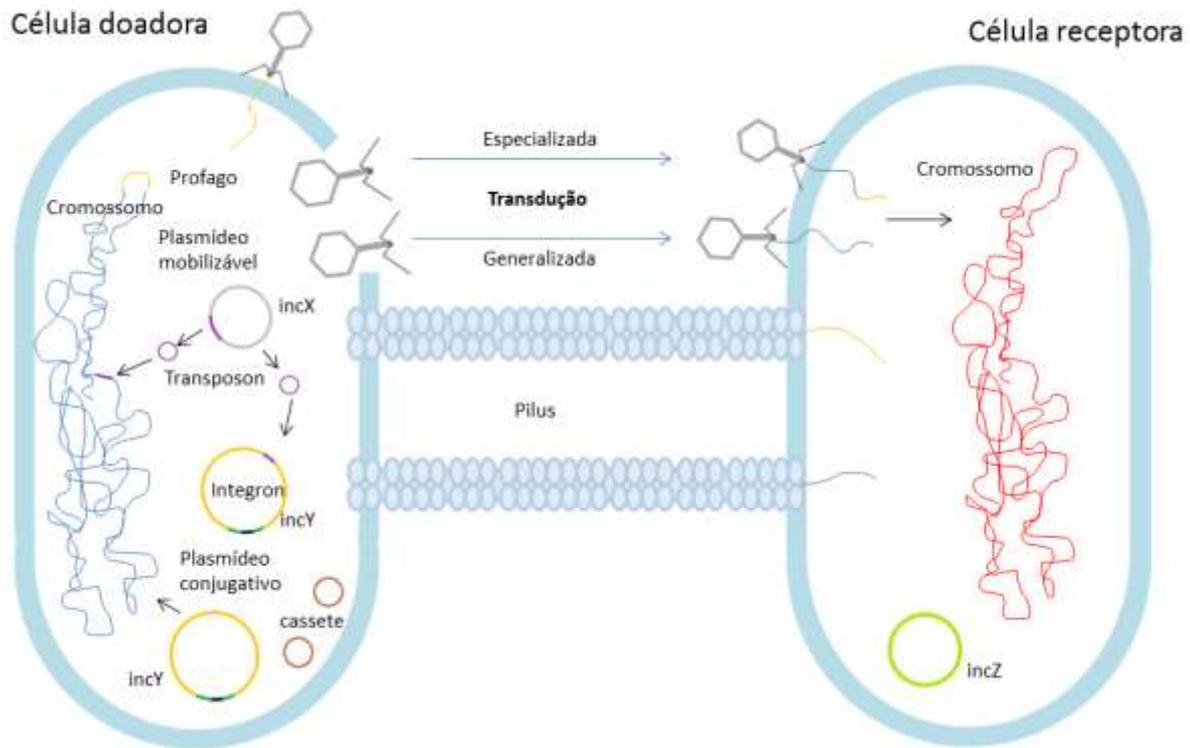


Figura 2.1 - Transferência de DNA entre células bacterianas. Transdução (1). O DNA genômico (amarelo) de um fago temperado se insere no DNA cromossomal (azul escuro) como um profago; posteriormente, se replica embalando o DNA do hospedeiro (transdução generalizada) ou com seu próprio DNA (transdução especializada), então lisa a célula, e infecta uma nova célula receptora na qual o DNA recombinaria com o cromossomo da célula hospedeira (vermelho). Conjugação (2). Plasmídeos conjugativos grandes, em baixo número de cópias (laranja) e Elementos Integrativos Conjugativos (não representados) usam o pílus (estrutura protéica) para estabelecer a conexão com a célula recipiente e transferir-se. Alternativamente, uma cópia de um pequeno plasmídeo multicópia, uma ilha genômica ou uma cópia do cromossomo completo pode ser transferido para uma nova célula, na qual esses elementos genéticos podem se inserir no cromossomo e se replicar independentemente caso sejam compatíveis com o plasmídeo já residente da nova célula (verde claro). Transposons conjugativos e plasmídeos de bactérias Gram-positivas (não representados) não utilizam pílus. Transposição (3). Transposons (rosa) integram-se em novos locais (sítios) do cromossomo ou dos plasmídeos por recombinação não-homóloga. Integrons (verde escuro) usam mecanismos similares para trocar cassetes de genes únicos (marrom). (adaptado de FROST et al., 2005).

A conjugação é a forma mais estudada de THG e está associada ao sistema de transporte do tipo IV, que cria um pílus entre duas células bacterianas pelo qual um plasmídeo é transferido da célula doadora para a receptora (ZECHNER et al., 2012). A transformação consiste na aquisição de fragmentos de DNA disponíveis no meio extracelular que são incorporados pela célula de forma artificial em condições de laboratório e acontece também

em ambientes naturais, com mudanças transitórias nas condições físicas da membrana celular causadas por oscilações nas condições do ambiente (SOUCY et al., 2015). A transdução depende da ação de bacteriófagos, e consiste na incorporação e posterior retirada de DNA viral do cromossomo bacteriano, podendo carregar alguns trechos de sequências do DNA hospedeiro adjacentes ao genoma viral. A organização dos genes de bacteriófagos é relativamente conservada e sua presença pode ser detectada por softwares de predição de vírus quando inserido no genoma microbiano (ROUX et al., 2015).

Além dos três mecanismos tradicionais de THG via EGMs, foram recentemente descritos os agentes de transferência de genes (ATGs). Os ATGs são derivados de bacteriófagos que perderam a capacidade de reconhecer sua própria sequência no genoma do hospedeiro. Acredita-se que tal inespecificidade seja devido à perda de um gene que codifica a subunidade da enzima de reconhecimento da sequência do próprio bacteriófago. Com isso, uma sequência aleatória do genoma hospedeiro é encapsulada e eventualmente exportada, sem carregar a sequência do bacteriófago (LANG et al., 2012).

### **2.3 Bacteriófagos**

Os bacteriófagos ou, simplesmente, fagos (vírus que infectam bactérias) já são conhecidos há mais de um século, e seu estudo tem trazido avanços importantes para o conhecimento sobre os vírus e outras áreas da Biologia. Por exemplo, fagos foram cruciais para o estabelecimento do dogma central da biologia molecular – informação é passada do DNA para o RNA e então à proteína – e têm desempenhado papel fundamental nos ecossistemas, condicionando a evolução e a virulência em bactérias (SALMOND; FINERAN, 2015). Além disso, as pesquisas com fagos têm desenvolvido muitas técnicas para a biotecnologia, como as enzimas de restrição, a transformação genética e a recente descoberta e exploração do sistema de resistência CRISPR Cas (Figura 2.2) (JIANG et al., 2013).

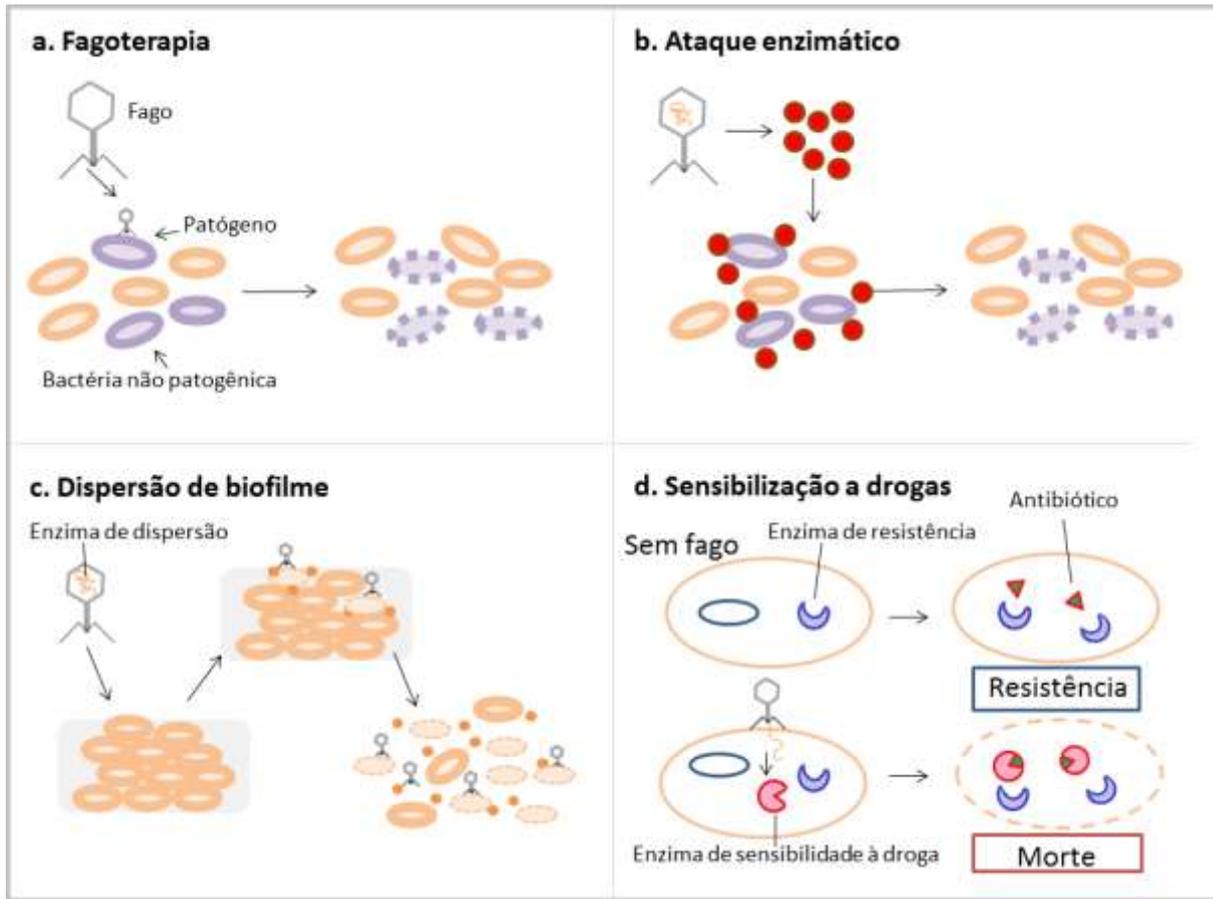


Figura 2.2 - Fagos e seus produtos têm inspirado uma série de mecanismos para o combate a bactérias. Fago terapia (a) é a utilização de bacteriófagos que reconhecem e infectam tipos específicos de bactérias patogênicas, causando sua lise. Fago enzimas (b) é a utilização de fagos que produzem enzimas para atuarem na eliminação de células, incluindo as patogênicas. Dispersão de biofilmes (c) é o emprego de fagos que produzem enzimas capazes de quebrar as estruturas da matriz do biofilme. Sensibilização às drogas (d) é a utilização de fagos para inserção de genes que codificam enzimas alvo do ataque de antibióticos em bactérias resistentes, tornando-as suscetíveis à ação do medicamento (adaptado de SALMOND; FINERAN, 2015)

Fagos contribuem para a patogenicidade e evolução de bactérias por promoção ou intensificação de múltiplos mecanismos, incluindo diferentes tipos de THGs. Por exemplo promovendo a transformação ao romper a célula do hospedeiro no ciclo lítico, o material genético da célula bacteriana fica disponível no ambiente, aumentando as chances de transformação. Também via transdução generalizada, fagos podem carregar partes do seu próprio genoma e partes do genoma da bactéria até o próximo hospedeiro. Na transdução especializada, fagos reconhecem e empacotam seu próprio genoma a partir de sua forma de profago, carregando-o até o próximo hospedeiro (SALMOND; FINERAN, 2015).

O material genético dos fagos pode ser DNA ou RNA de fita dupla (do inglês, *double-stranded* ou *ds*) ou de fita simples (*single-stranded* ou *ss*). O genoma de um fago varia em tamanho de 3,5 Kbp (fago MS2) até de 500 kbp (dsDNA do fago G de *Bacillus*) e pode

incluir nucleotídeos modificados como forma de proteção contra o ataque de enzimas de restrição do hospedeiro (ACKERMANN, 2009; SALMOND; FINERAN, 2015).

A classificação dos fagos utiliza caracteres morfológicos e genéticos e à medida que são descobertas novas formas de organização genômica, a taxonomia precisa ser periodicamente revista. Morfológicamente, os fagos podem ser caudados (*tailed*), poliédricos, filamentosos ou pleomórficos, e alguns possuem envelopes lipídicos ou lipoproteicos. Os fagos mais estudados pertencem à ordem Caudovirales (*ds DNA, tailed*), que se divide em três famílias: Myoviridae (por exemplo, fago T4), Siphoviridae (por exemplo, fago  $\lambda$ ) e Podoviridae (por exemplo, fago T7) (ACKERMANN, 2009; SALMOND; FINERAN, 2015).

Fagos podem apresentar ciclo de vida lítico ou lisogênico (Figura 2.3). Para infectar a bactéria hospedeira, o fago precisa primeiro interagir com receptores da célula hospedeira, adsorver e então injetar seu genoma. A estratégia de replicação subsequente depende se o fago é virulento ou temperado. Fagos virulentos, como o T4, são apenas capazes de replicar pelo ciclo lítico, um processo envolvendo a produção de novas progênes virais e sua liberação da célula infectada. Alternativamente, fagos temperados (como o fago  $\lambda$ ) entram no ciclo lítico ou formam uma associação estável com seu hospedeiro, chamada lisogenia. (CHATURONGAKUL; OUNJAI, 2014).

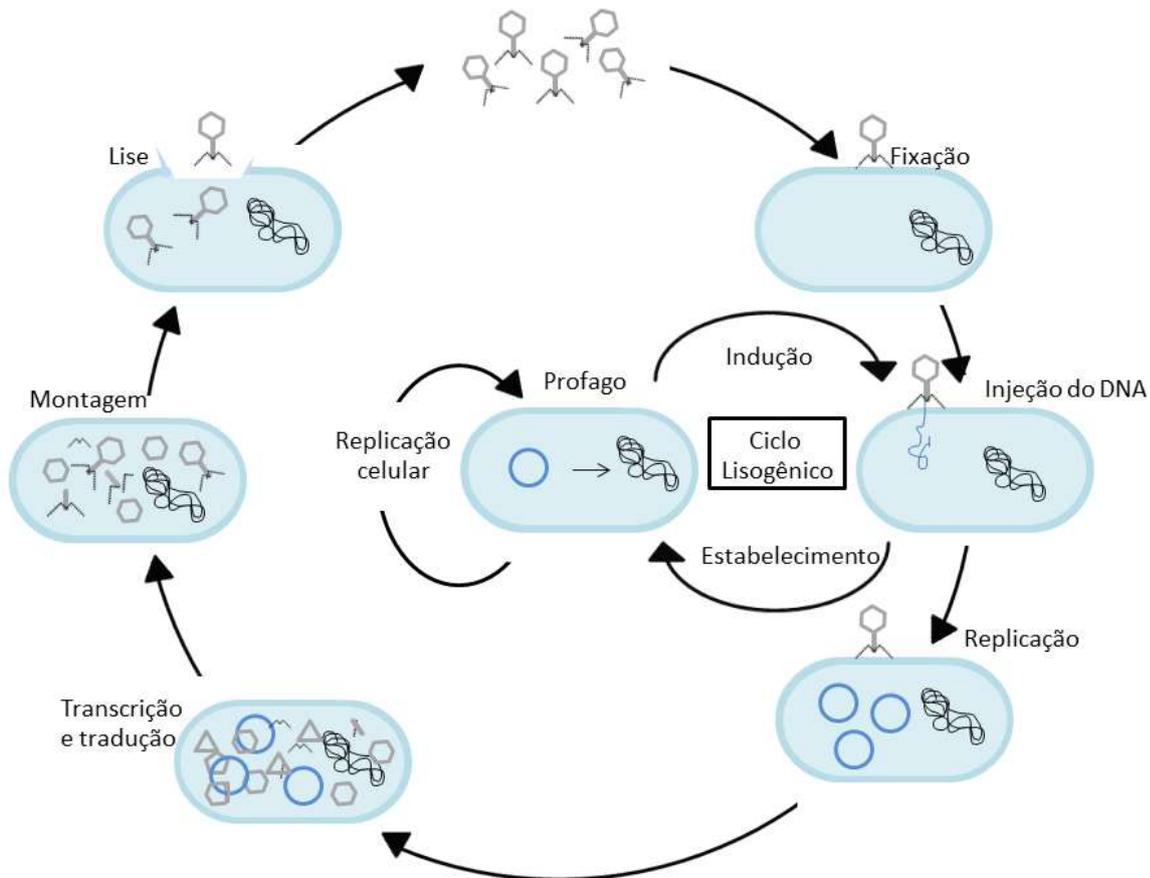


Figura 2.3 - Ilustração do ciclo de vida os bacteriófagos. Os bacteriófagos se aderem à parede da célula bacteriana e injetam seu material genético. Podem inserir-se no genoma hospedeiro até serem eventualmente induzidos à replicação (ciclo lisogênico), ou se replicarem imediatamente no interior da célula hospedeira, dando continuidade aos processos de transcrição e tradução de proteínas de fagos (ciclo lítico). Os indivíduos são montados ainda dentro da célula do hospedeiro e então rompem a célula, para infectarem novas bactérias, completando o ciclo (adaptado de SALMOND; FINERAN, 2015)

Durante a lisogenia, o genoma viral é chamado de profago e se replica de acordo com o DNA do hospedeiro. Pode permanecer em estado livre circular, semelhante a um plasmídeo (como o fago P1) ou integrado no cromossomo bacteriano (como o fago  $\lambda$ ). Sob condições de estresse, os profagos podem sair do estado lisogênico e produzir virions que são liberados da bactéria. A liberação da progênie do fago resulta na morte da célula hospedeira por lise (YOUNG, 2013; ROACH; DONOVAN, 2015), exceto para os fagos filamentosos, que podem ser liberados por secreção via membrana celular, evitando a lise celular e causando uma infecção crônica que apenas reduz o crescimento da célula do hospedeiro (RAKONJAC et al., 2011).

### **3 OBJETIVOS E HIPÓTESES**

#### **3.1 Objetivo geral**

Caracterizar os EGMs no microbioma de sedimentos de manguezais do estado de São Paulo.

#### **3.2 Objetivos específicos**

- 1) Estimar a frequência e expressão de EGMs nos sedimentos de manguezais, detalhando as categorias mais abundantes;
- 2) Anotar os bacteriófagos presentes nas sequências da biblioteca de fosmídeo e apontar para suas potenciais funções ecológicas e evolutivas com base na predição de genes.

#### **3.3 Hipóteses**

- Os EGMs mais numerosos nos manguezais são os bacteriófagos, que infectam as bactérias pertencentes aos grupos mais abundantes neste ambiente;
- As principais bactérias que compõem o microbioma do sedimento de manguezais hospedam fagos, que desempenham papel relevante na ecologia e evolução desses grupos.



## 4 MATERIAL E MÉTODOS

### 4.1 Áreas dos manguezais estudados

Foram estudados os EGMs presentes no microbioma de três áreas de manguezais do estado de São Paulo, duas localizadas no município de Bertioga (BrMgv02 e BrMgv03) e uma em Cananéia, na reserva ambiental da Ilha do Cardoso (BrMgv04). Os manguezais selecionados apresentam características peculiares e distintas entre si. Um dos manguezais é altamente contaminado com petróleo (BrMgv02), devido a um derramamento de 35 milhões de litros de óleo cru, ocorrido próximo à costa no ano de 1983, sendo que a presença de óleo nos sedimentos ainda é visível até cerca de 30 a 50 cm de profundidade. Uma segunda área de manguezal sem a presença de derramamento de óleo, porém sob a pressão de contaminantes oriundos da proximidade de tal área com o centro urbano de Bertioga (BrMgv03) foi estudada para avaliar os impactos da expansão urbana. A área preservada (BrMgv04), localizada na cidade de Cananéia foi utilizada como controle pristino, buscando esta ser um modelo da comunidade de EGMs sobre condições naturais de ocorrência dos manguezais. Quanto às características ambientais, os dois manguezais de Bertioga são banhados por uma mistura de água oriunda do mar e do rio Iriri, enquanto que o manguezal de Cananéia é banhado diretamente pelo mar. A localização e as principais características que distinguem as áreas entre si são apresentadas na Tabela 4.1. Estes manguezais já foram alvos de uma série de investigações (ANDREOTE et al., 2012; DIAS et al., 2011; FASANELLA et al., 2012; FRANCO DIAS et al., 2012; VARON-LOPEZ et al., 2014), que os dividiu com base em características de contaminação, vegetação e composição microbiológica. O contexto físico e químico e da comunidade microbiana, já caracterizados nesses trabalhos, podem trazer contribuições adicionais para o entendimento dos mecanismos genéticos de adaptação a esses ambientes.

Tabela 4.1 – Descrição e localização dos manguezais estudados no presente trabalho (adaptado de ANDREOTE et al., 2012)

| Área    | Características                        | Município | Coordenadas                | Vegetação  |
|---------|--|-----------|----------------------------|--|
| BrMgv02 | Alta contaminação com óleo             | Bertioga  | 23°53'49" S<br>46°12'28" W | Em recuperação, baixa abundância de <i>R. mangle</i>           |
| BrMgv03 | Contaminação oriunda do centro urbano  | Bertioga  | 23°54'06" S<br>45°15'03" W | Abundante, com presença de outras espécies                     |
| BrMgv04 | Ambiente pristino em reserva ambiental | Cananéia  | 25°05'02" S<br>47°57'42" W | Dominância, mas não exclusividades das espécies de manguezais* |

\*No estado de São Paulo, as florestas de manguezais são compostas essencialmente por três espécies: *Avicennia shaueriana*, *Laguncularia racemosa* e *Rhizophora mangle*

## 4.2 Coleta das amostras de sedimento

Esta etapa do trabalho foi desenvolvida em trabalho anterior do grupo, conduzido pela estudante Luana Lira Cadete durante seu projeto de mestrado (CADETE, 2014).

As amostras dos sedimentos dos manguezais foram coletadas em abril de 2013, com o auxílio de cilindros de 7 cm de diâmetro e 30 cm de profundidade, com três repetições para cada área, como descrito por Fasanella et al. (2012). Foram coletados aproximadamente 300 g por ponto, armazenados em sacos plásticos e processados no dia seguinte à coleta. Alíquotas foram feitas a partir de 5 g da parte mais interna do sedimento, minimizando o efeito de perturbação (HOLLIBAUGH et al., 2011).

## 4.3 Extração de DNA e mRNA e sequenciamento das amostras de sedimento

A extração DNA foi realizada a partir de 0,5 g de sedimento para cada amostra. O kit utilizado para extração do DNA total das amostras de sedimento foi o Power Soil DNA Extraction kit (MoBio Labs, Inc. Solana Beach, USA), como trabalho anterior (ANDREOTE et al., 2012). A extração de RNA foi realizada a partir de 2 g de sedimento por amostra, utilizando o kit RNA PowerSoil® Total RNA Isolation (MoBio Laboratories, Inc. Solana Beach, USA) seguindo as recomendações do fabricante. Além do sequenciamento do RNA total da amostra, o RNA foi tratado para remoção de parte do RNA ribossomal, enriquecendo as amostras em mRNA, que passa a representar até 48% do RNA total sequenciado (STEWART et al., 2010).

As amostras de DNA e mRNA foram preparadas como LIRA (2014) e sequenciadas em equipamento Illumina HiSeq2000. O controle de qualidade das amostras foi feito utilizando ferramentas do software CLC Genomics WorkBench v.5.1., usando como

parâmetros o limite de pontuação de qualidade = 0,03, semelhante a qualidade *Phred* 20, o número máximo de ambiguidade =1 e descarte de leituras com comprimento inferior a 50 nucleotídeos.

#### **4.4 Preparo das sequências de metagenoma e metatranscriptoma dos sedimentos para anotação**

O controle de qualidade das sequências submetidas ao MG RAST para anotação foi realizado no *software* CLC Genomics Workbench v.5.1 utilizando os parâmetros de qualidade mínima 0,01 (equivalente a *Phred* 20). Foram também descartados os *reads* com tamanho inferior a 50 nucleotídeos. O pareamento das sequências *forward* e *reverse* também foi realizado no mesmo *software*, utilizando a opção de importação de sequências vindas de plataforma Illumina.

#### **4.5 Anotação automática das sequências de DNA e mRNA dos sedimentos e análises funcionais no MG-RAST**

As sequências pareadas foram submetidas ao pipeline da plataforma Metagenomics Analysis Server (MG-RAST) (<http://metagenomics.anl.gov>) (MEYER, *et al.* 2008), na qual os dados foram submetidos a um *pipeline* adicional de remoção de artefatos de sequenciamento. Para distinguir os tipos e a frequência dos EGMs presentes nas amostras dos sedimentos dos manguezais, as sequências de DNA dos metagenomas e mRNA para os metatranscriptomas foram alinhadas localmente contra o banco de dados Subsystems no servidor do de anotação automática MG-RAST, com os seguintes parâmetros: *cutoff* máximo: 1e-5, mínimo de alinhamento: 15 pb, e mínimo de identidade: 60%. Foram realizadas análises hierárquicas funcionais das sequências referentes as 9 amostras de DNA sequenciadas (3 amostras x 3 áreas), com anotação de todas as funções metabólicas disponíveis na plataforma. Considerando os objetivos deste estudo, foram analisadas mais detalhadamente as anotações referentes aos EGMs e suas subcategorias: transposons, plasmídeos e fagos.

Para identificar padrões na distribuição dos tipos de EGMs entre as áreas, foram montados gráficos de PCoA com os genes de EGMs preditos por amostra de sedimentos para DNA e mRNA na plataforma do MG-RAST, com os parâmetros descritos no item 3.3 da metodologia. O número de hits foi normalizado por milhões de reads por amostra e a matriz de dissimilaridade Bray-Curtis e a geração do gráfico de PCoA foram feitas no *software* PAST

(HAMMER et al., 2001).

#### 4.6 Sequências da biblioteca de fosmídeo e montagem dos *contigs*

A biblioteca de fosmídeo utilizada nesse trabalho contém aproximadamente 20 mil clones com tamanho médio de inserto de 20 kpb, e foi montada em trabalho anterior, seguindo a metodologia descrita por Vasconcellos et al. (2010). Um pool com todos os clones foi sequenciado em equipamento Illumina HiSeq2000. O controle de qualidade das sequências foi realizado no *software* CLC Genomics Workbench v. 5.1 utilizando os parâmetros de qualidade mínima 0,01 (equivalente a *Phred* 20) e foram descartados os *reads* inferiores a 50 nucleotídeos. O pareamento das sequências *forward* e *reverse* também foi realizado no mesmo *software*, utilizando a opção de importação de sequências vindas de plataforma Illumina.

Para a montagem de *contigs*, foram realizados testes com diferentes parâmetros no *software* CLC Genomics Workbench v. 5.1, utilizando como variáveis fixas o tamanho da janela de leitura = 50 pb, *contig* mínimo = 500 pb, custo por troca = 2, custo por inserção = 3 e custo por deleção = 3, conforme tabela 4.2.

Tabela 4.2 - Parâmetros testados na montagem das sequências da biblioteca

| Montagem Id. | <i>K-mer</i> | Parcela de sobreposição | Similaridade |
|--------------|--------------|-------------------------|--------------|
| 1            | 24           | 0.5                     | 0.8          |
| 2            | 18           | 0.7                     | 0.8          |
| 3            | 20           | 0.7                     | 0.8          |
| 4            | 20           | 0.6                     | 0.8          |

A montagem escolhida para posterior anotação foi submetida ao teste de quimeras do *pipeline* UCHIME Chimera Check (EDGAR et al., 2011), disponível na plataforma do FunGene ([www.fungene.com](http://www.fungene.com)), utilizando os parâmetros padrões.

#### 4.7 Predição de fagos e profagos a partir das sequências da biblioteca de fosmídeo

Todos os *contigs* da montagem escolhida para as sequências da biblioteca de fosmídeo foram submetidos à plataforma do VirSorter (ROUX et al., 2015) de acordo com as instruções

da plataforma. Para predição de fagos e profagos, o VirSorter utiliza duas metodologias: a de homologia de genes e a de assinatura genômica (a segunda utilizada apenas para as sequências que não possuem genes tipicamente encontrados em fagos ou profagos). De acordo com o grau de confiabilidade da predição, o VirSorter classifica as sequências em 6 categorias: 1) Sequências que com certeza pertencem a fagos, 2) Sequências que com quase certeza pertencem a fagos, 3) Sequências que provavelmente pertencem a fagos, 4) Sequências que com certeza pertencem a profagos, 5) Sequências que com quase certeza pertencem a profagos e 6) Sequências que provavelmente pertencem a profagos. As sequências pertencentes às categorias 1 e 2 que continham genes tipicamente encontrados em fagos foram anotadas manualmente no Artemis (CARVER et al., 2012), enquanto que as sequências de todas as categorias assinaladas por assinatura genômica e as das categorias 3, 4, 5 e 6 foram anotadas automaticamente no VirSorter.

#### **4.8 Anotação manual de genes dos *contigs* da biblioteca de fosmídeo que contém fagos ou profagos**

Para os *contigs* cujas sequências continham genes típicos de fagos (categorias 1 e 2 do VirSorter) foram realizados alinhamentos de BLASTp contra as proteínas do banco nr do NCBI (<http://www.ncbi.nlm.nih.gov/>) e os genes preditos foram anotados manualmente no programa Artemis (CARVER et al., 2012) no qual foram geradas imagens dos genes anotados na ferramenta DNAPlotter (RUTHERFORD et al., 2000). Os genes de fagos foram submetidos à busca por famílias proteicas na plataforma do PFAM (<http://pfam.xfam.org/>).

#### **4.9 Predição das linhagens de bacteriófagos com base nos hospedeiros**

Os genes anotados manualmente como tipicamente pertencentes a fagos (*core genes*) foram utilizados como referência para a busca na plataforma do *GenomeNet Virus-Host DB*, que busca, com base na literatura, prever os fagos conforme já foram descritos anteriormente para determinados hospedeiros (MIHARA et al., 2016).



## 5 RESULTADOS E DISCUSSÃO

### 5.1 Abundância e expressão dos EGMs no solo

A partir do sequenciamento das amostras de sedimento em equipamento Illumina HiSeq 2000, foram obtidas aproximadamente 120 milhões de sequências somando mRNA e DNA. Após o controle de qualidade, restaram aproximadamente 72 milhões de sequências de mRNA, com média de 24 milhões reads por área amostrada e 23 milhões de sequências de DNA, com média de 7,7 milhões por área (Figura 5.1).

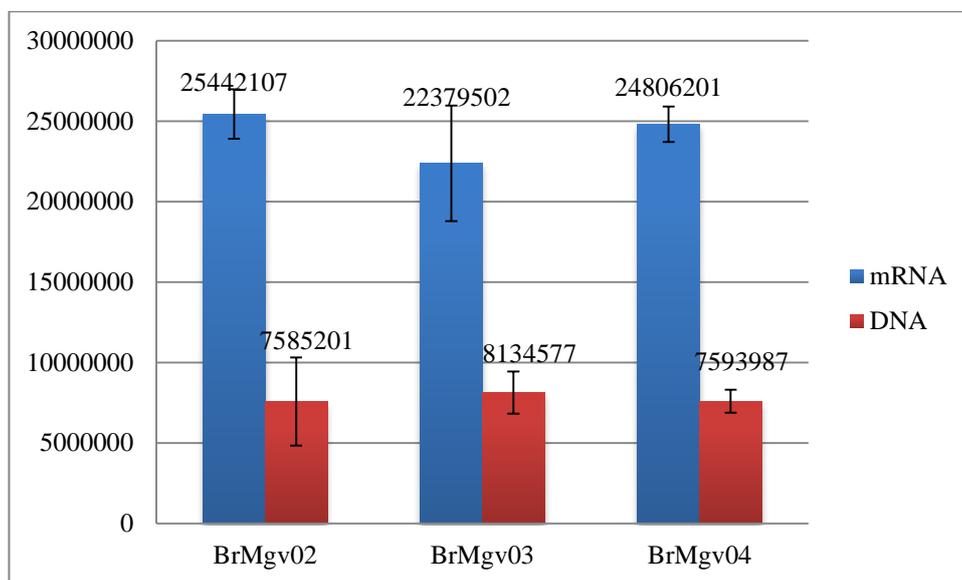


Figura 5.1 - Sequências de DNA e mRNA obtidas das amostras de sedimentos dos manguezais foram utilizadas para as análises de metagenômica e metatranscritômica, respectivamente. O sequenciamento foi realizado em equipamento Illumina HiSeq 2000 (item 3.2 da metodologia) e o controle de qualidade das sequências foi realizado no *software* CLC Genomics Workbench v.5.1, como detalhado no item 3.4

Os reads foram então submetidos à anotação automática no pipeline do MG-RAST. Os EGMs preditos pela plataforma correspondem em média a 12 % de todos os *hits* do metagenoma (Tabela 5.1). As demais categorias (88% restante) incluem manutenção e duplicação celular, ciclos biogeoquímicos como metabolismo do carbono, nitrogênio, enxofre e fósforo e produção de compostos diversos do metabolismo secundário.

Tabela 5.1 – Porcentagem do número de Hits do metagenoma relacionados a EGMs comparado ao número total de *reads* anotados nas diferentes áreas dos manguezais em triplicatas

| Manguezais | Hits das amostras (%) |                |                |
|------------|-----------------------|----------------|----------------|
|            | a                     | b              | c              |
| BrMgv02    | 14952 (5,98%)         | 49691 (11,09%) | 52656 (15,49%) |
| BrMgv03    | 49085 (15,16%)        | 44673 (13,15%) | 43601 (10,34%) |
| BrMgv04    | 48347 (14,90%)        | 43623 (10,47%) | 39702 (10,78%) |

Há grande variação entre as réplicas de cada área, em especial para a área BrMgv02, que teve variação de 5,98 % de todos os *hits* do metagenoma para a amostra BrMgv02a a 15,49 % dos *hits* para a amostra BrMgv02c.

Os EGMs foram anotados para metatranscriptoma na plataforma do MG-RAST de forma similar à análise do metagenoma. Em média foram preditos para os metatranscriptomas das áreas BrMgv02, BrMgv03 e BrMgv04 os mRNA de EGMs correspondentes respectivamente a 3,6%, 2,87% e 3,60% do total de *hits* anotados (Tabela 5.2).

Tabela 5.2 - Porcentagem do número de *hits* do metatranscriptoma relacionados a EGMs comparado ao número total de *hits* anotados nas diferentes áreas dos manguezais em triplicatas

| Manguezais | Hits das amostras (%) |               |               |
|------------|-----------------------|---------------|---------------|
|            | a                     | B             | C             |
| BrMgv02    | 38030 (3,79%)         | 34745 (3,26%) | 34512 (3,79%) |
| BrMgv03    | 22298 (3,40%)         | 25346 (1,81%) | 22545 (3,40%) |
| BrMgv04    | 62645 (3,86%)         | 55128 (3,95%) | 52814 (3,00%) |

As amostras variaram de 25346 hits (1,81 % de todos os genes expressos) para BrMgv03b até 62645 hits (3,86% de todos os genes expressos) para a amostra BrMgv04a.

Tanto para DNA quanto para mRNA, não foi possível detectar mudanças claras na abundância ou expressão dos EGMs considerados como categoria geral. Apenas para a amostra BrMgv03b foi possível detectar uma diminuição significativa na expressão relativa dos EGMs em relação a todas as amostras analisadas pelo metatranscriptoma. A abundância relativa pode variar conforme a abundância de outros genes do metabolismo primário ou secundário, ou mesmo conforme alterações na composição da comunidade microbiana. Para uma análise mais detalhada dos EGMs dos sedimentos de manguezais, é preciso observar as mudanças nos diferentes tipos de EGMs (Figura 5.2). Os bacteriófagos corresponderam a cerca de 90% de todos os genes relativos aos EGMs anotados automaticamente pela plataforma do MG-RAST para todas as repetições das três áreas.

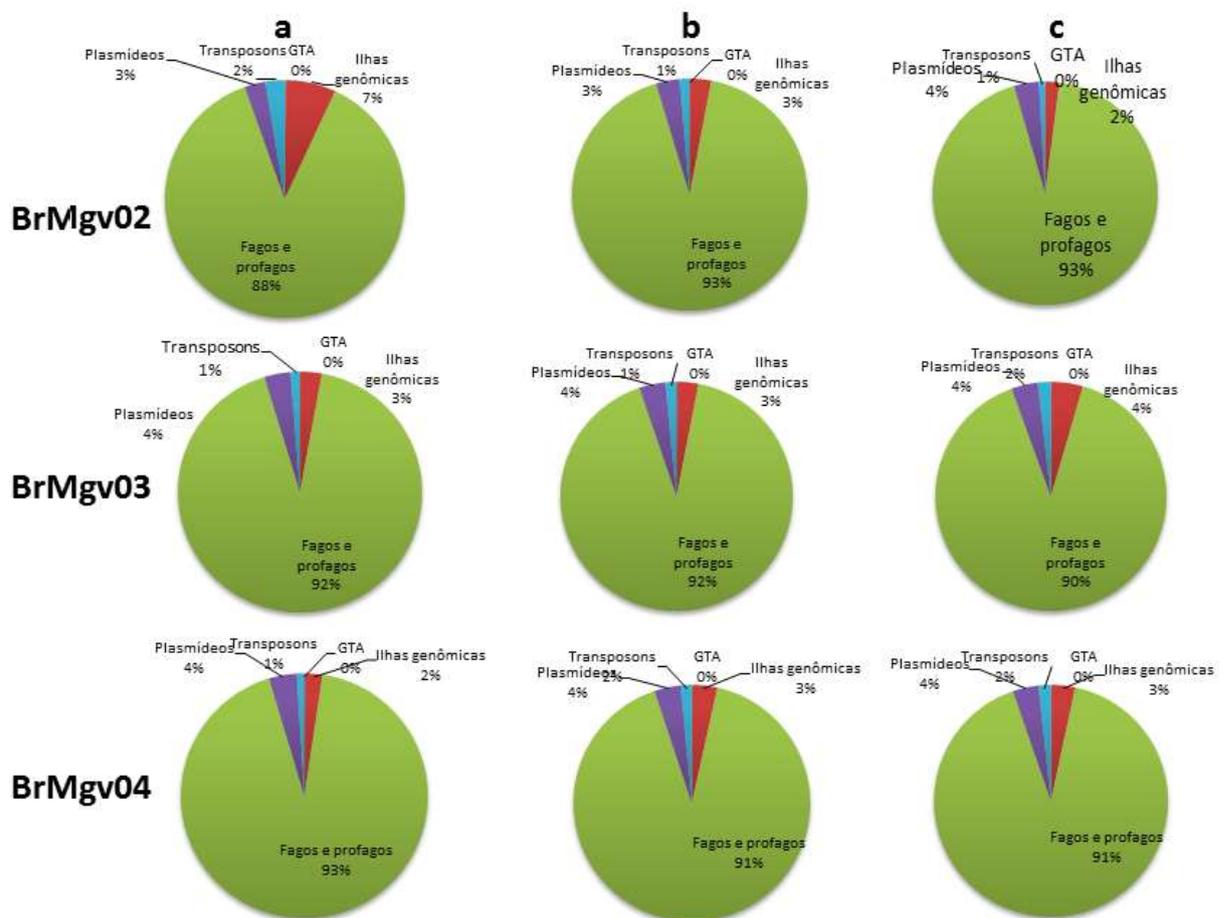


Figura 5.2 - Perfil de DNA por categoria de EGM. São mostradas as porcentagens da abundância relativa de cada categoria em relação às demais (plasmídeos, transposons, ATGs, Ilhas genômicas, fagos e profagos). Estão representadas as três áreas estudadas (BrMgv02, BrMgv03 e BrMgv04), em cada triplicata (identificadas por a, b, c). O gráfico foi montado a partir de previsões na plataforma do MG-RAST, como descrito no item 3.3 da metodologia

A dissimilaridade entre os padrões preditos para estas funções pode ser visualizada graficamente na Figura 5.3, onde a comparação entre os padrões não revelam uma separação

nos padrões de distribuição por genes ou por sua expressão. Contudo, é possível observar uma baixa variância para amostras de DNA da área BrMgv04.

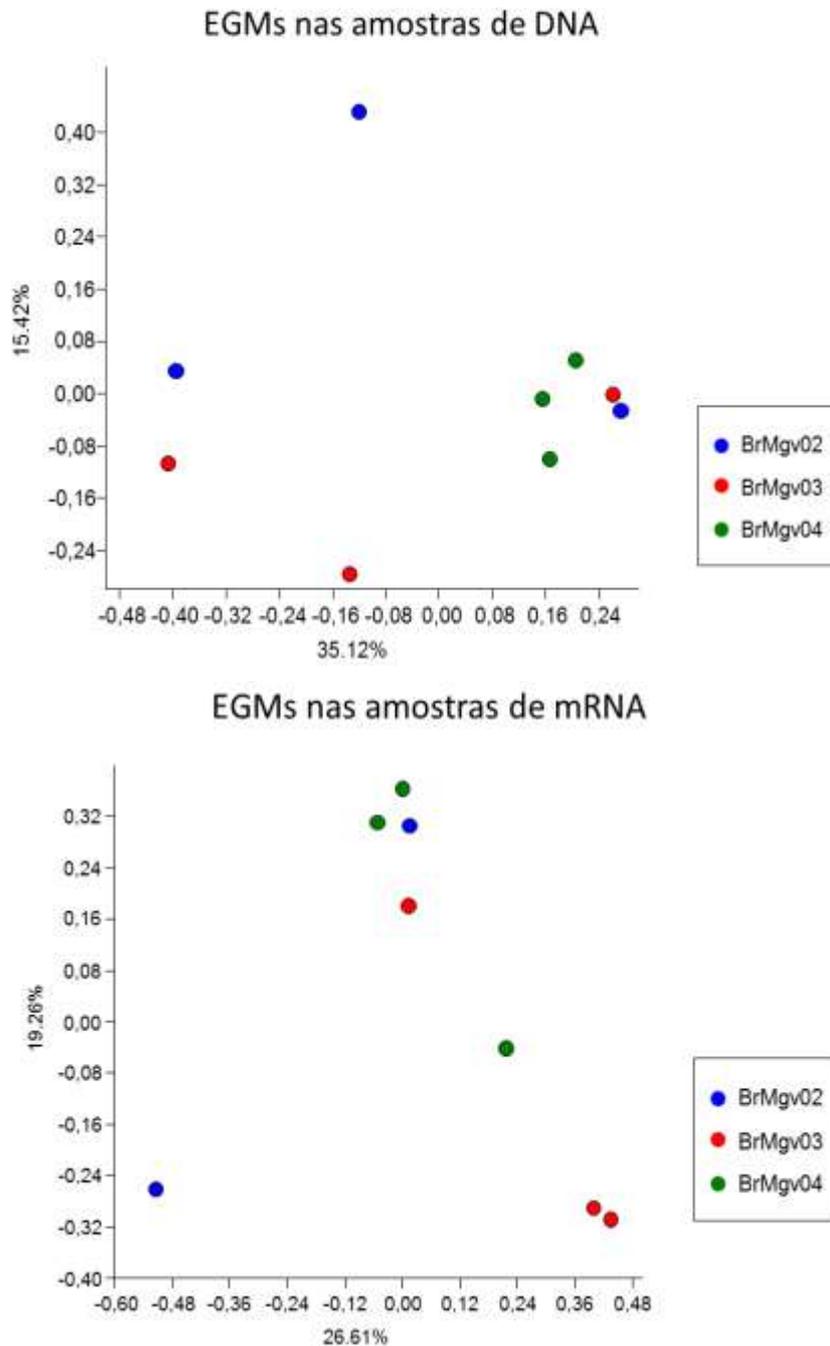


Figura 5.3 - Gráfico de PCoA com os genes de EGMs preditos por amostra de sedimentos para DNA (acima) e mRNA (abaixo) na plataforma do MG-RAST, com os parâmetros descritos no item 3.3 da metodologia. Em azul as réplicas da área BrMgv02, em vermelho as réplicas da área BrMgv03 e em verde as réplicas da área BrMgv04. O número de hits foi normalizado por milhões de reads por amostra e colocado no software PAST, utilizando a matriz de dissimilaridade de Bray-Curtis

Levando em consideração a categoria predominante de EGMs, foram observadas as abundâncias relativas dos fagos e profagos nas amostras das áreas a partir das análises do metagenoma (Tabela 5.3).

Tabela 5.3 - Abundância relativa de fagos e profagos entre os EGMs nos diferentes manguezais acessados

| Manguezais | Hits das amostras (%) |                |                |
|------------|-----------------------|----------------|----------------|
|            | a                     | b              | c              |
| BrMgv02    | 13108 (87,67%)        | 45912 (92,40%) | 49243 (93,52%) |
| BrMgv03    | 45339 (92,37%)        | 40988 (91,75%) | 39241 (90,00%) |
| BrMgv04    | 44967 (93,01%)        | 39904 (91,47%) | 36282 (91,39%) |

Quando olhamos para os bacteriófagos na perspectiva da comparação com todos os genes presentes nas amostras, encontramos uma participação significativa dessa categoria. A grande variação dos bacteriófagos entre as amostras (de 5 a 15%) explica a variação encontrada quando olhamos para os EGMs como categoria única, visto que os fagos representam aproximadamente 92 % de todos os EGMs.

A potencial atividade de expressão dos fagos e profagos foi estimada a partir dos metatranscriptomas. Em média, os fagos correspondem a 10,96% de todos os genes expressos das amostras das três áreas, sendo que os valores relativos a cada área são em média de 9,99%, 11,79% e 11,09% para BrMgv02, BrMgv03 e BrMgv04, respectivamente. (Tabela 5.4).

Tabela 5.4 – Perfil de atividade dos fagos e profagos em relação a todo o metatranscriptoma (mRNA)

| Manguezais | Hits das amostras (%) |                |                |
|------------|-----------------------|----------------|----------------|
|            | a                     | b              | c              |
| BrMgv02    | 13108 (5,24%)         | 45912 (10,25%) | 49243 (14,49%) |
| BrMgv03    | 45339 (14,00%)        | 40988 (12,07%) | 39241 (9,31%)  |
| BrMgv04    | 44967 (13,86%)        | 39904 (9,58%)  | 36282 (9,85%)  |

Foram detectados de 13108 a 49243 hits para a categoria de fagos e profagos, variando de 5,24% até 14,49% de todos os mRNA no metatranscriptoma das amostras. A ampla variação dentro da mesma área se repete para os três manguezais e não é possível determinar o que leva a tal variação tomando os EGMs como grupo único. Apesar de a variação entre as

tréplicas de um mesmo manguezal serem grandes, é possível observar um menor número de hits para a área BrMgv02 em relação à área BrMgv03 e principalmente em relação à BrMgv04. Apesar de não haver um padrão evidente na diferença da abundância de mRNA referente aos fagos entre as áreas ou mesmo entre amostras, é possível verificar diferenças significativas quando comparamos a abundância de diferentes categorias de genes para a área BrMgv03b em relação às demais amostras da mesma área e também no comparativo com as demais áreas, que apresentam um padrão no perfil de mRNA expresso para genes de fagos (Figura 5.4).

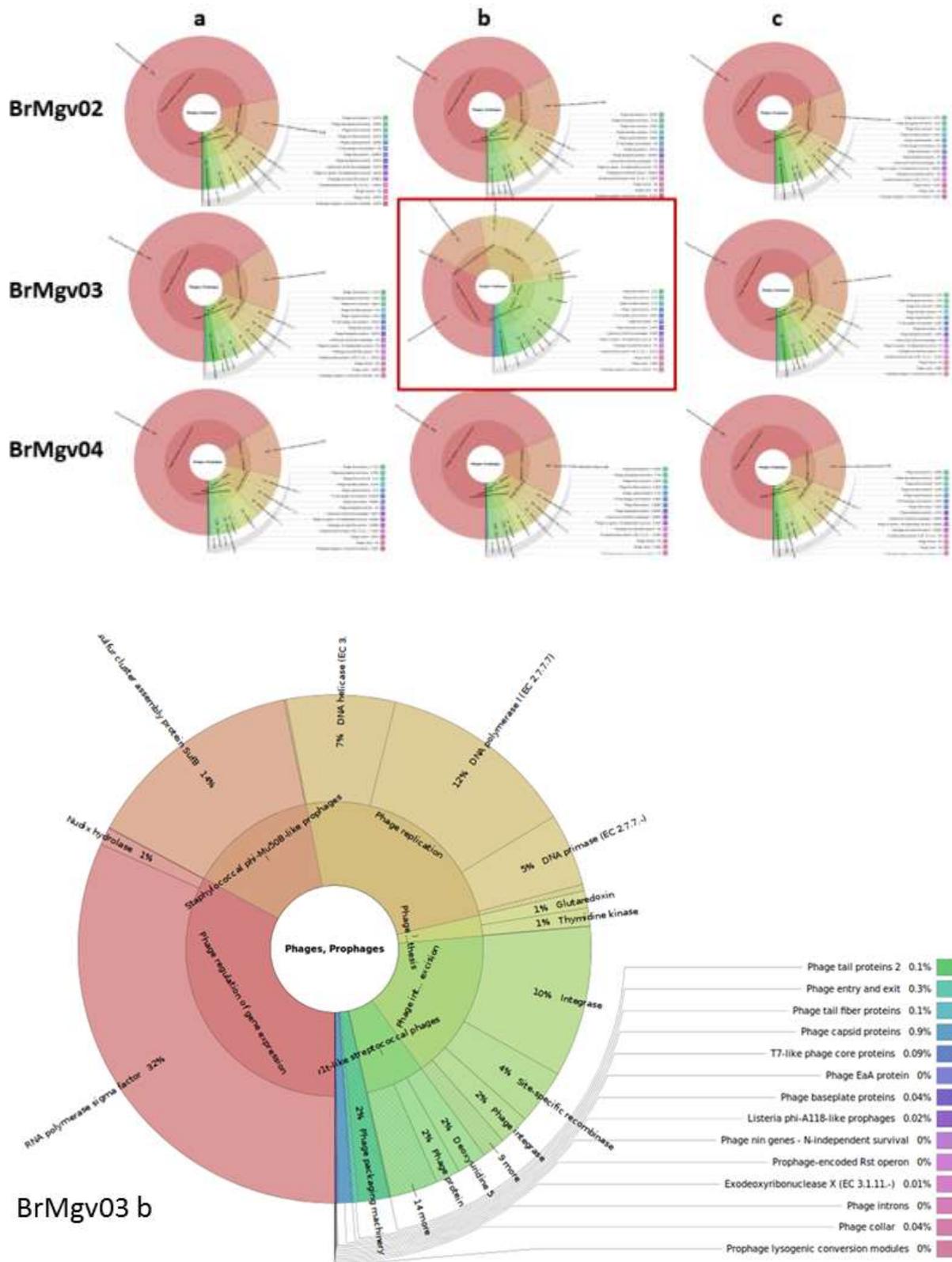


Figura 5.4 - Perfil de atividade por categoria de gene de profago. São mostradas as porcentagens da abundância relativa de cada categoria em relação às demais. Acima estão representadas as três áreas estudadas (BrMgv02, BrMgv03 e BrMgv04), em cada triplicata (identificadas por a, b, c). O gráfico foi montado a partir de predições na plataforma do MG-RAST, como descrito no item 3.3 da metodologia. Abaixo está o destaque de uma das amostras com expressão diferenciada de genes de fagos e profagos, amostra b da área BrMgv03

A categoria mais expressa em todas as amostras é a de regulação de expressão gênica dos fagos, especialmente as RNA polimerases. Para a amostra BrMgv03b houve um grande aumento na expressão de genes relacionados à replicação, como a DNA polimerase, DNA helicase e DNA primase. Também houve um aumento de uma ordem na expressão da categoria de genes relacionados com a excisão dos fagos, principalmente as integrases e recombinases que representam 12% e 4%, respectivamente, dos mRNA relativos a fagos em BrMgv03b enquanto para as outras amostras essas abundâncias não ultrapassam os 2% para essas categorias. Detalhes da expressão das integrases na Figura 5.5.

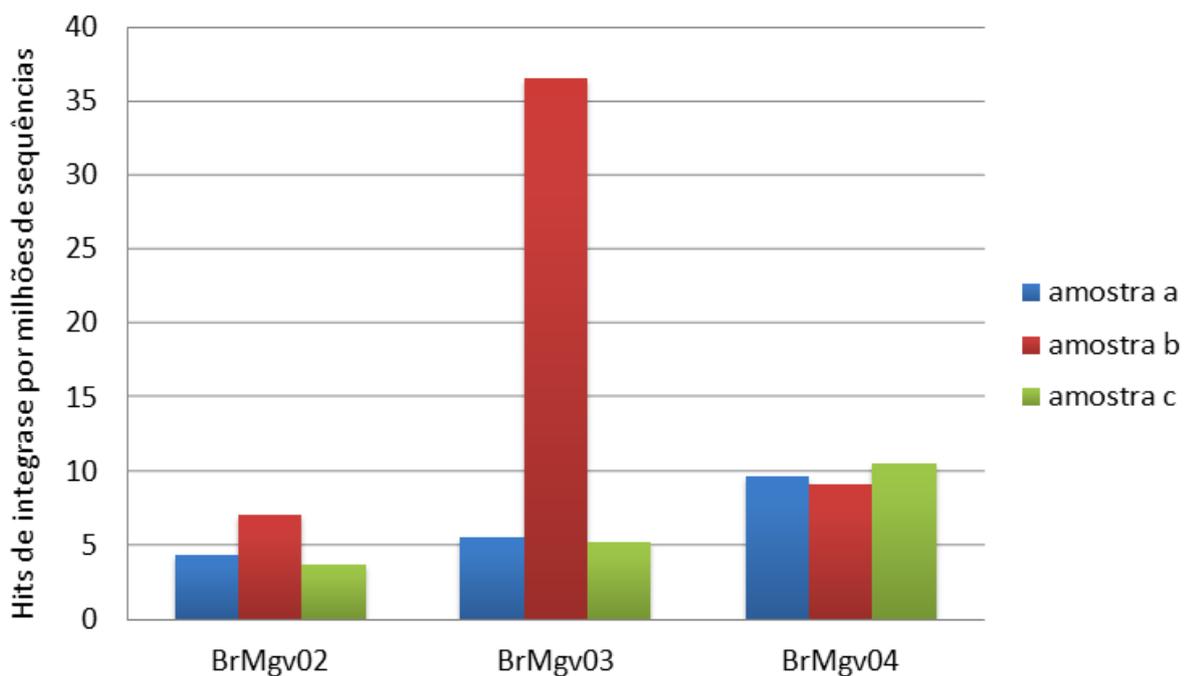


Figura 5.5 – Gráfico do perfil de mRNA dos genes de integrase nas amostras de metatranscriptoma. As áreas estão representadas por suas siglas e a, b, c correspondem às réplicas. A predição de genes foi realizada na plataforma do MG RAST com os parâmetros descritos na metodologia (item 3.3) contra o banco Subsystems. A normalização dos dados foi realizada por milhão de *reads* de acordo os resultados do sequenciamento de mRNA para cada área

A produtividade, a temperatura do ambiente e a disponibilidade de nutrientes são inversamente proporcionais à indução de lisogenia de fagos, uma vez que essa é uma estratégia de enfrentamento a condições de baixa disponibilidade de fontes de energia ou de condições ambientais adversas em geral (PAUL, 2008). O predomínio do ciclo lítico ou lisogênico depende de condições ambientais e celulares no momento da infecção. Sob condições ambientais favoráveis, no caso de uma população com dinâmica de crescimento favorecida por abundância de recursos energéticos o processo de infecção lítico tende a prevalecer. Por outro lado, caso as condições ambientais sejam desfavoráveis, a infecção toma

a via lisogênica. Quando acontece uma melhora nas condições oferecidas pelo ambiente, uma parte da população pode mudar para o ciclo lítico, o que segue a morte de parte da população de hospedeiros e liberação de fagos no ambiente (KOSKELLA; BROCKHURST, 2014). Um trabalho recente simulou e analisou uma série de modelos computacionais para a evolução de comunidades microbianas em ambientes sobre pressões seletivas de fagos e com heterogeneidade de disponibilidade de nutrientes, e encontrou que a infecção por fagos reduz as taxas de especiação e estabilizam a comunidade microbiana. A intensidade desse efeito varia conforme fatores de distribuição espacial e situação da comunidade no momento da invasão, sendo que comunidades mais bem estabelecidas e com melhores condições ambientais são mais resistentes à infecção por fagos e a maior quantidade de nutrientes disponíveis tende a um aumento na biomassa total e redução da diversidade por conta da alta competição por recursos, com alguns grupos prevalecendo sobre outros (KLIMENKO et al., 2016).

Tomando as três áreas utilizadas nesse estudo, podemos considerar que as áreas BrMgv02 e BrMgv03 possuem como fator diferenciado a alta quantidade de carbono disponibilizada pela presença de óleo (BrMgv02) e ao esgoto domiciliar e outros rejeitos devido à proximidade ao centro urbano da cidade (BrMgv03), o que pode causar mudanças no balanço nutricional. Um estudo anterior nessas mesmas áreas também encontrou diferenças maiores na área BrMgv03 quanto à composição da comunidade bacteriana e expressão de genes relacionados aos ciclos biogeoquímicos em relação às áreas BrMgv02 e BrMgv04 (CADETE, 2014). Entre as áreas estudadas é possível que BrMgv03 seja a que sofre maiores mudanças recentes nas condições do ambiente, já que o derramamento de óleo na área BrMgv02 aconteceu há 30 anos e a comunidade microbiana teve tempo suficiente para se estabilizar frente as mudanças nas condições ambientais.

## **5.2 Anotação de bacteriófagos a partir das sequências da biblioteca de fosmídeo**

Após o controle de qualidade das leituras no CLC Genomics Workbench v.5.1 com os parâmetros do item 3.6 da metodologia, foram obtidos 141,3 milhões de *reads* entre 50 e 100 pb, que foram utilizados para a montagem de *contigs* no mesmo programa. A melhor montagem obtida foi a de identificação 4 (Tabela 5.5), com detalhamento dos parâmetros na Tabela 3.2 da metodologia., que foi obtida a partir de um ajuste da montagem 3, com redução da quantidade de *reads* que não se encaixavam nos *contigs* maiores que 500 pb.

Tabela 5.5 - Montagens dos *contigs* no *software* CLC GenomicsWorkbench v.5.1., a partir das sequências do *pool* amostrado na biblioteca de fosmídeo da área BrMgv02

| Montagem<br>Id. | Quantidade de<br><i>contigs</i> | Maior<br><i>read</i> | Tamanho<br>médio | <i>Reads</i> utilizados | N-50 |
|-----------------|---------------------------------|----------------------|------------------|-------------------------|------|
| 1               | 165487                          | 23198                | 1519             | 119.580.163             | 1889 |
| 2               | 154013                          | 29985                | 1609             | 100.173.564             | 2080 |
| 3               | 123253                          | 37911                | 1738             | 91.551.757              | 2302 |
| 4*              | 122986                          | 37842                | 1741             | 101.828.917             | 2308 |

\*montagem escolhida

Os *contigs* da montagem escolhida foram então submetidos ao UCHIME – Chimera Check e não foram detectadas sequências quiméricas. A predição e a anotação automática de fagos e profagos foi realizada pela plataforma do VirSorter, conforme descrito na metodologia. O VirSorter classifica as sequências preditas de fagos e profagos em níveis de confiança decrescente que variam de 1 a 3 para fagos e de 4 a 6 para profagos, com base em genes conservados de fagos ou assinatura genômica. Para os *contigs* utilizados, foram preditos 27 fagos na categoria 2 do VirSorter, sendo 10 anotados com base em genes conservados de fagos e 17 anotados com base na assinatura genômica. A categoria 5 de profagos contou com uma única anotação, baseada em genes conservados de fagos. As categorias 1 e 4 não tiveram *contigs*, e as categorias 3 e 6 tiveram 37 fagos e 1 profagos, sendo que todos foram reconhecidos com base apenas na assinatura genômica.

Os 10 fagos preditos na categoria 2 do VirSorter com base em genes conservados de fagos foram manualmente anotados no programa Artemis. Em umas dessas sequências não foi possível confirmar a predição de genes conservados de fago. Um esquema foi montado para representar esses fagos (Figura 5.6).

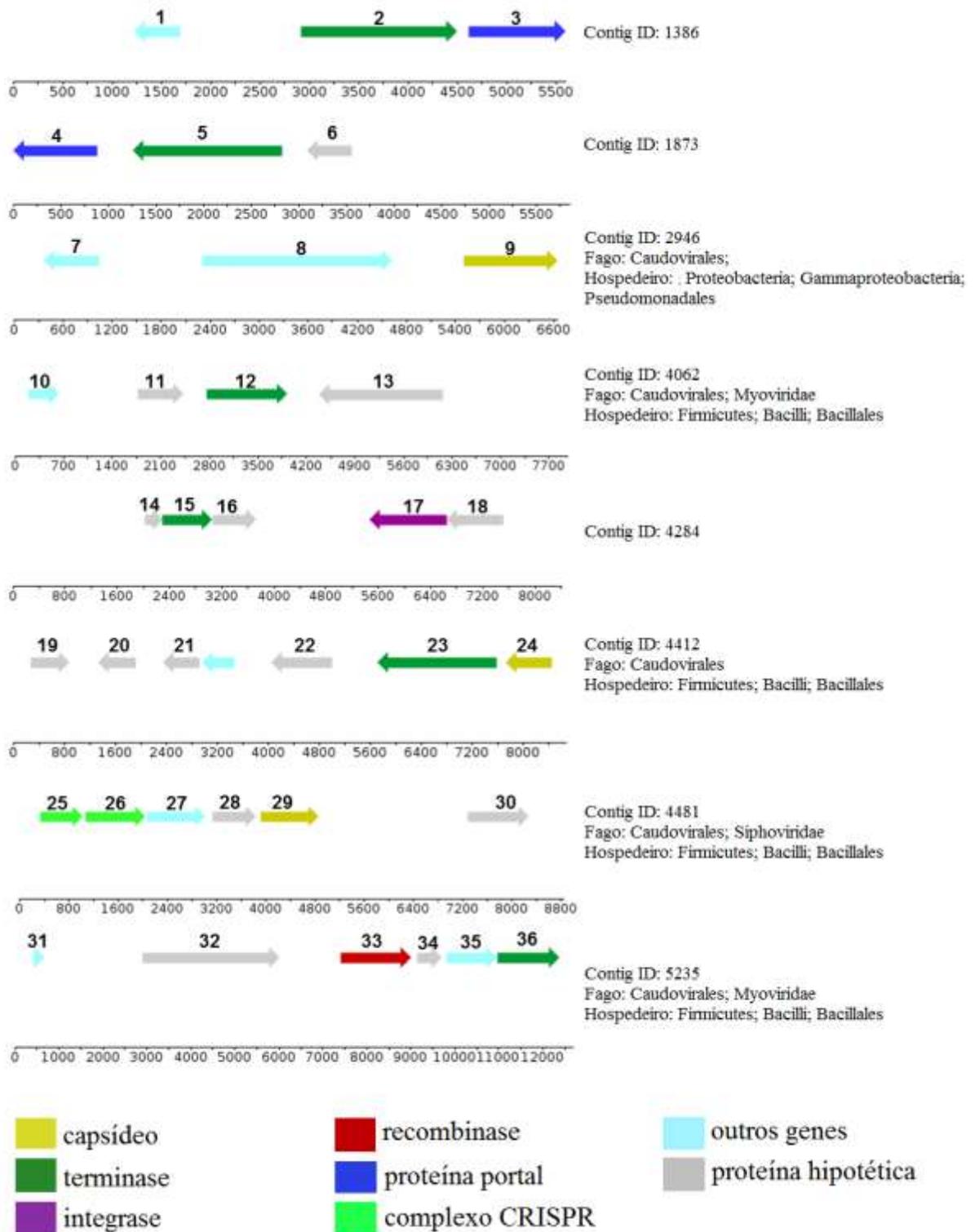


Figura 5.6 - Esquema representando os fagos anotados manualmente para as seqüências montadas da cicloteca de fosmídeo. Cada *contig* representa parte de um fago. Os detalhes da anotação de cada gene podem ser conferidos a partir dos números de identificação acima de cada gene da figura, que segue a identificação utilizada nas tabelas 4.6 e 4.7. Ao lado direito de cada *contig* está o número de identificação do *contig*, a linhagem viral e a linhagem do hospedeiro já descrito na literatura, predito no GenomeNet VirusHost Database, conforme ítem 3.9 da metodologia

Foram encontrados genes relacionados aos capsídeos (identificação 9, 24 e 29 na Figura 4.4) e dois genes associados à THG foram anotados: integrase e recombinase (respectivamente, os genes 17 e 33). As integrases e as recombinases estão associadas ao processo de excisão - integração dos fagos a outros genomas. Outros genes de importância ecológica e evolutiva também podem ser observados, como por exemplo, proteínas de transporte e secreção, e associadas ao mecanismo de defesa CRISPR Cas (Tabela 4.6).

Tabela 5.6 - Anotação manual de genes de fagos no programa Artemis. Resultados obtidos a partir de Blastp contra o banco nr do NCBI (versão disponível em fevereiro de 2016)

(continua)

| Contig id | Gene ID | Posição | Início | Fim  | Gene                                | Produto                             | Similaridade  | Família PFAM                                 |
|-----------|---------|---------|--------|------|-------------------------------------|-------------------------------------|---|--|
| 1386      | 1       | -       | 1232   | 1693 | transcriptional regulator           | transcriptional regulator           | Bacillus amyloliquefaciens, transcriptional regulator (143aa), e-value: 1e-06. 35% id in 111aa  |  |
|           | 2       | +       | 2914   | 4500 | terminase-like Family               | terminase-like Family               | Bellilinea caldifistulae, terminase-like family (512aa), e-value: 3e-69. 35% id in 516aa  |  |
|           | 3       | +       | 4621   | 5604 | phage portal protein                | phage portal protein, SPP1 Gp6-like | Anaerolinea thermolimososa, phage portal protein, SPP1 Gp6-like (424aa), e-value: 2e-54. 37% in 308aa                                   |  |
| 1873      | 4       | -       | 1      | 888  | phage portal protein, SPP1 Gp6-like | phage portal protein, SPP1 Gp6-like | Anaerolinea thermolimososa, phage portal protein, SPP1 Gp6-like (424aa), e-value: 3e-51. 35% in 299aa                                   |  |
|           | 5       | -       | 1265   | 2827 | terminase-like Family               | terminase-like Family               | Bellilinea caldifistulae, terminase-like family (512aa), e-value: 5e-81. 38% id in 522aa  |  |
|           | 6       | -       | 3097   | 3564 | hypothetical protein                | hypothetical protein                | Dehalococcoidia bacterium SM23_28_1, hypothetical protein (132aa), e-value: 1e-11. 37% id in 134aa                                      |  |
| 2946      | 7       | -       | 369    | 1040 | DNA primase                         | DNA primase                         | phage-associated bacterium endosymbiont of Mortierella elongata FMR23-6, DNA primase/helicase (757aa), e-value: 6e-143. 37% id in 793aa |  |
|           | 8       | +       | 2297   | 4630 | geranyl reductase family protein    | geranyl reductase family protein    | Nitrospira sp. 1, geranylgeranyl reductase family protein (387aa), e-value: 8e-33. 35% id 214aa   |  |
|           | 9       | +       | 5498   | 6651 | phage capsid protein                | phage capsid protein                | Acinetobacter tandoii, phage capsid protein (399aa), e-value: 1e-82. 40% id in 396aa  | Phage_capsid, Clan: CL0373, e-value: 5.7e-67 |

Tabela 5.6 - Anotação manual de genes de fagos no programa Artemis. Resultados obtidos a partir de Blastp contra o banco nr do NCBI (versão disponível em fevereiro de 2016)

(continuação)

| Contig id | Locus tag | Posição | Início | Fim  | Gene                 | Produto              | Similaridade  | Família PFAM |
|-----------|-----------|---------|--------|------|----------------------|----------------------|---|--------------|
| 4062      | 10        | +       | 199    | 651  | ABC transporter      | ABC transporter      | Gemmatimonas sp. SM23_52, ABC transporter permease (374aa), e-value: 9e-52. 64% id in 149aa |              |
|           | 11        | +       | 1776   | 2435 | hypothetical protein | Hypothetical protein | Bradyrhizobium elkanii, hypothetical protein (267aa), e-value: 4e-10. 35% id in 168aa       |              |
|           | 12        | +       | 2762   | 3937 | terminase A          | terminase A          | Bacillus fordii, terminase A (475aa), e-value: 2e-39. 30% id in 366aa                       |              |
|           | 13        | -       | 4380   | 5828 | hypothetical protein | hypothetical protein |   |              |
| 4284      | 14        | +       | 2015   | 2299 | hypothetical protein | hypothetical protein | Brevibacterium linens, hypothetical protein (162aa), e-value: 2e-27. 50% id in 88aa         |              |
|           | 15        | +       | 2296   | 3069 | terminase            | terminase            | Nitrobacter sp. Nb-311A, terminase (250aa), e-value: 3e-73. 49% id in 247aa                 |              |
|           | 16        | +       | 3062   | 3739 | hypothetical protein | hypothetical protein | Arthrobacter castelli, hypothetical protein (225aa), e-value: 5e-34. 40% id in 220aa        |              |
|           | 17        | -       | 5477   | 6649 | integrase            | integrase            | Frankia sp. EUN1f, integrase (406aa), e-value: 6e-81. 41% id in 384aa                       |              |
|           | 18        | -       | 6646   | 7509 | hypothetical protein | hypothetical protein | Salinispora pacifica, hypothetical protein (171aa), e-value: 0,003. 29% id in 160aa         |              |
| 4412      | 19        | +       | 282    | 893  | hypothetical protein | hypothetical protein | Nitrosococcus watsonii, hypothetical protein (412aa), e-value: 7e-26. 35% id in 191aa       |              |

Tabela 5.6 - Anotação manual de genes de fagos no programa Artemis. Resultados obtidos a partir de Blastp contra o banco nr do NCBI (versão disponível em fevereiro de 2016) (continuação)

| Contig id | Locus tag | Posição | Início | Fim  | Gene                           | Produto                        | Similaridade  | Família PFAM                                    |
|-----------|-----------|---------|--------|------|--------------------------------|--------------------------------|---|---|
| 4412      | 20        | -       | 1341   | 1922 | prophage protein               | Lj965 prophage protein         | Streptococcus pneumoniae, Lj965 prophage protein (169aa), e-value: 1,5. 39% id in 70aa  |   |
|           | 21        | -       | 2358   | 2933 | hypothetical protein           | hypothetical protein           | Achromobacter sp., disulfide bond formation protein DsbA (174aa), e-value: 0,20. 40% id in 68aa                                 |   |
|           | 22        | -       | 4061   | 5002 | hypothetical protein           | hypothetical protein           | Bacillus cereus, hypothetical protein (326aa), e-value: 2e-92. 48% id in 310aa  |   |
|           | 23        | -       | 5726   | 7588 | terminase                      | Terminase                      | Geobacillus thermoglucosidasius, terminase (392aa), e-value: 3e-22. 44% id in 133aa   | Peptidase_S78_2, Clan: CL0201, e-value: 3.5e-31 |
|           | 24        | -       | 7743   | 8453 | phage minor capsid protein     | phage minor capsid protein     | Cellulophaga phage phi17:1, phage minor capsid protein (290aa), e-value: 4e-08. 33% id in 134aa                                 |   |
| 4481      | 25        | +       | 337    | 1047 | CRISPR-associated protein      | CRISPR-associated protein      | Mastigocladus laminosus, CRISPR-associated protein (235aa), e-value: 2e-12. 30% id 223aa  | Cas_Cas02710, Clan: CL0236, e-value: 7.5e-07    |
|           | 26        | +       | 1082   | 2065 | CRISPR-associated protein Cas6 | CRISPR-associated protein Cas6 | Candidatus Magnetoglobus multicellularis str. Araruama, CRISPR-associated protein Cas6 (309aa), e-value: 5e-58. 32% id in 318aa | family: DUF2276, Clan: CL0362, e-value: 4.6e-19 |
|           | 27        | +       | 2077   | 3036 | epimerase                      | epimerase                      | Alicyclobacillus herbarius, epimerase (323aa), e-value: 1e-171. 74% id in 312aa   |   |
|           | 28        | +       | 3145   | 3855 | hypothetical protein           | hypothetical protein           | Ornatilinea apprima, hypothetical protein ADN00_18900 (178aa), e-value: 9e-10. 44% id in 97aa                                   |   |
|           | 29        | +       | 3929   | 4873 | major capsid protein           | major capsid protein           | Dorea formicigenerans, major capsid protein (326aa), e-value: 1e-119. 54% id in 307aa   | Phage_capsid, Clan: CL0373, e-value: 3.6e-36    |

Tabela 5.6 - Anotação manual de genes de fagos no programa Artemis. Resultados obtidos a partir de Blastp contra o banco nr do NCBI (versão disponível em fevereiro de 2016) (conclusão)

| Contig id | Locus tag | Posição | Início | Fim   | Gene                                  | Produto                               | Similaridade  | Família PFAM                           |
|-----------|-----------|---------|--------|-------|---------------------------------------|---------------------------------------|---|--|
|           | 30        | +       | 7299   | 8285  | hypothetical protein                  | hypothetical protein                  | Ornatilinea apprima, hypothetical protein ADN00_18935 (328aa), e-value: 4e-37. 29% id in 333aa              |  |
| 5235      | 31        | +       | 432    | 695   | type II secretion system protein GspD | type II secretion system protein GspD | Ralstonia sp. GA3-3, type II secretion system protein GspD (812aa), e-value: 2e-05. 40% id in 88aa          |  |
|           | 32        | +       | 2899   | 6033  | hypothetical protein                  | hypothetical protein                  | Phycisphaerae bacterium SM1_79, hypothetical protein AMJ75_02645 (1242aa), e-value: 9e-156. 40% id in 870aa |  |
|           | 33        | +       | 7400   | 9025  | recombinase                           | recombinase                           | Saccharibacillus sacchari, site-specific recombinase (532aa), e-value: 1e-28. 25% id in 533aa               | Recombinase, Clan: n/a, e-value: 9e-24 |
|           | 34        | +       | 9133   | 9708  | hypothetical protein                  | hypothetical protein                  | Achromobacter sp., hypothetical protein (214aa), e-value: 0,42. 21% id in 164aa                             |  |
|           | 35        | +       | 9804   | 10973 | ParB domain protein nuclease          | ParB domain protein nuclease          | Defluviitoga tunisiensis, ParB domain protein nuclease (417aa), e-value: 2e-92. 42% id in 403aa             |  |
|           | 36        | +       | 10973  | 12376 | terminase large subunit               | terminase large subunit               | uncultured Mediterranean phage uvMED, terminase large subunit (462aa), e-value: 7e-56. 32% id in 469aa      |  |

Foram preditas proteínas de transporte e secreção do sistema ABC e do tipo II. Um trabalho recente descreve esses genes de transporte e secreção associados aos EGMs (RICHARDS; TALBOT, 2013). Ao adquirir e tornar funcional um gene associado a determinado sistema de transporte, é possível que o hospedeiro adquira uma vantagem adaptativa no ambiente. Por exemplo, a possibilidade de absorção de um nutriente disponível que antes não poderia ser absorvido por falta de um canal de transporte apropriado.

Dois genes relacionados ao sistema CRISPR Cas foram anotados no mesmo fago (*contig* 4481, genes 25 e 26 na Tabela 4.6). A presença de genes associados a repetições CRISPR pode significar duas possibilidades: a presença desses genes é uma marca de sua entrada no genoma do hospedeiro, portanto este é *contig* representa um profago; a segunda possibilidade é que se trata de um fago não integrado ao genoma do hospedeiro que possui uma sequência que mimetiza a região de CRISPR da bactéria hospedeira, evitando o reconhecimento e eliminação do DNA viral pelo mecanismo CRISPR Cas da célula bacteriana (BLOWER et al., 2012).

Na tentativa de estimar qual seria a origem dos fagos anotados pelo VirSorter, foi utilizada a plataforma do GenomeNet VirusHost DataBase, onde é feita uma busca com base na literatura para prever os fagos conforme já foram descritos anteriormente para determinados hospedeiros. Os hospedeiros preditos para os fagos pertencem aos filos Proteobacteria (classe Gammaproteobacteria) e Firmicutes (classe Bacilli) (Tabela 5.7). Contudo, há uma limitação técnica quanto à interação entre parasita-hospedeiro, já que não é possível determinar a abrangência de espécies ou linhagens de hospedeiro ou as taxas de infecção de cada fago. Não se pretende aqui determinar quais são as linhagens ou espécies dos hospedeiros dos fagos nos manguezais, tampouco determinar as taxas de infecção, mas sim dar uma predição de quais são os potenciais grupos microbianos infectados nessas amostras, para que seja possível a extrapolação da predição para os grupos microbianos já descritos na microbiota dos solos de manguezais para as mesmas áreas.

Tabela 5.7 - Taxonomia dos fagos a seus respectivos hospedeiros, com base em genes conservados. Análise realizada na plataforma do *GenomeNet VirusHost Database* com base na predição taxonômica feita por blastp contra o banco nr do NCBI (versão disponível em 03/2016)

| Contig ID | Gene ID | Orientação | Início | Fim  | Proteína nr/np       | Família PFAM                                    | Linhagem do hospedeiro   | Linhagem do fago   |
|-----------|---------|------------|--------|------|----------------------|---|--|--|
| 2946      | 9       | +          | 5498   | 6651 | phage capsid protein | Phage_capsid, Clan: CL0373, e-value: 5.7e-67    | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter | Viruses; dsDNA viruses, no RNA stage; Caudovirales               |
| 4062      | 12      | +          | 2762   | 3937 | terminase A          | Não encontrada                                  | Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus                             | Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae   |
| 4412      | 20      | -          | 1341   | 1922 | prophage protein     | Não encontrada                                  | Bacteria; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus              | Viruses; dsDNA viruses, no RNA stage; Caudovirales               |
| 4412      | 23      | -          | 5726   | 7588 | Terminase            | Peptidase_S78_2, Clan: CL0201, e-value: 3.5e-31 | Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Geobacillus                          | Viruses; dsDNA viruses, no RNA stage; Caudovirales               |
| 4481      | 29      | +          | 3929   | 4873 | major capsid protein | Phage_capsid, Clan: CL0373, e-value: 3.6e-36    | Bacteria; Firmicutes; Bacilli; Bacillales; Paenibacillaceae; Paenibacillus                   | Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae |
| 5235      | 33      | +          | 7400   | 9025 | Recombinase          | Recombinase, Clan: n/a, e-value: 9e-24          | Bacteria; Firmicutes; Bacilli; Bacillales; Paenibacillaceae; Brevibacillus                   | Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae   |

Por serem parasitas intracelulares obrigatórios e terem certo nível de especificidade quanto ao tipo de hospedeiro, um aspecto importante a ser analisado na ecologia dos fagos é a abundância dos grupos bacterianos que podem servir como hospedeiros. Estima-se que dois terços dos genomas de Gammaproteobacterias e bactérias Gram-negativas possuem profagos (BRÜSSOW, 2004), enquanto que arqueias e grupos patogênicos intracelulares com frequência não apresentam fagos lisogênicos, embora estudo recente do viroma em sedimentos do oceano tenha encontrado um número significativo de fagos que foram descritos infectando arqueias (ENGELHARDT et al., 2015).

A microbiota dos sedimentos de manguezais brasileiros é composta principalmente pelos filos Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes e Chloroflexi (ANDREOTE et al., 2012). Um estudo realizado na reserva ambiental da Ilha do Cardoso, em Cananéia – SP, com base no sequenciamento Sanger de 16S RNAr, mostrou a dominância das classes Alphaproteobacteria, Gammaproteobacteria e Acidobacteria, e a presença em menor abundância das classes Betaproteobacteria, Deltaproteobacteria, Firmicutes, Actinobacteria e Bacteroidetes (DIAS et al., 2010). Um trabalho mais amplo, utilizando o pirosequenciamento do gene ribossomal 16S RNAr de amostras oriundas de diversas áreas de manguezais brasileiros, encontrou como padrão o predomínio das classes Deltaproteobacteria e Gammaproteobacteria (VARON-LOPEZ, 2014), que também foram descritas como as metabolicamente mais ativas em um estudo de metatranscriptômica (CADETE, 2014).

Os fagos já descritos na literatura para esses grupos pertencem à ordem Caudovirales (*ds DNA*, sem estágio de RNA), com a presença de duas famílias: Siphoviridae e Myoviridae. A ordem Caudovirales ainda possui uma terceira família, Podoviridae (International Committee on Taxonomy of Virus - <http://ictvonline.org/>), que não foi encontrada nas predições para os fagos da biblioteca de fosmídeo. A ordem Caudovirales predomina nos ambientes pelágicos (YOSHIDA et al., 2013; HOPKINS et al., 2014). Um estudo que analisou o viroma dos sedimentos oceânicos em diferentes profundidades e encontrou grande variação na participação dos Caudovirales entre as diferentes profundidades, mas as três famílias foram encontradas em todas as amostras (6 profundidades, variando de 5 a 159 m de coluna d'água), sendo que houve o predomínio de Siphoviridae a 50 m, contribuindo com 86% dos reads e também a 91 m, representando 43% das sequências virais de mRNA.

No manguezal BrMgv02, é possível que o derramamento de óleo ocorrido há três décadas já tenha acelerado o processo de co-evolução entre fagos e as bactérias hospedeiras em um momento inicial, com uma série de recrutamentos, seleção recíproca, na qual seleção

direcional leva à fixação de novas mutações de resistência nas bactérias, seguida de uma nova mutação dos fagos, restaurando a infectividade dos vírus (no inglês, *arms-race coevolution*). Porém, após 30 anos do derramamento, provavelmente o convívio entre parasitas e hospedeiros ao longo de muitas gerações pode ter desacelerado essas taxas de co-evolução, de forma que a abrangência da resistência das bactéria e da infectividade dos fagos já tenham alcançado o seu ponto máximo. A diferença encontrada nesse estudo pode ser devido às diferenças ambientais analisadas, e não ao fator de mudança recente no ambiente com o derramamento do óleo.

A presença de hidrocarbonetos em excesso na área BrMgv02 provavelmente já não exerce o papel de mudança recente no ambiente sobre as relações entre fagos e bactérias. Dessa forma, a evolução no modelo *arms-race* pode ter dado lugar a oscilações sustentadas de genótipos para bactérias e fagos com diferentes níveis de especificidade de resistência e infectividade (KOSKELLA; BROCKHURST, 2014). Por esse processo, os fagos evoluem para infectar os genótipos bacterianos mais comuns no ambiente (ou seja, as espécies ou linhagens mais abundantes), conferindo vantagem para alelos bacterianos de resistência, que rapidamente aumentam sua frequência na população, entrando em uma dinâmica flutuante de seleção entre fagos e bactérias (GANDON et al., 2008). Esse modelo de co-evolução é chamado *killing the winner* (morte ao vencedor, em português) no qual as bactérias que possuem as melhores condições de adaptação ao ambiente são alvo mais frequente de parasitas (como os fagos), que têm uma grande pressão de seleção sobre genes que conferem possibilidade de infecção de hospedeiros abundantes (WEITZ et al., 2013). O manguezal BrMgv03 por sua vez possui diferenças significativas quanto à expressão de genes de fagos, em especial de genes associados às funções de excisão e entrada em genomas (ciclo lisogênico), como as integrases e portanto podem entrar na dinâmica da *arms-race*, conforme novas fontes de carbono são adicionadas ao ambiente, provenientes do centro urbano próximo.

## 6 CONSIDERAÇÕES FINAIS

- Os EGMs mais numerosos nos manguezais são os bacteriófagos, que infectam as bactérias mais abundantes como estratégia de replicação. Genes importantes para a adaptação foram anotados para fagos parcialmente montados a partir das sequências da biblioteca de fosmídeo, como os relacionados ao sistema de defesa CRISPR Cas e ao sistema de transporte ABC e do tipo II. Esses resultados apontam para um papel importante dos bacteriófagos nos solos de manguezais na regulação da comunidade microbiana e possivelmente no carreamento de genes de importância ecológica e evolutiva, via THG. Esses dados suportam a hipótese de que os EGMs desempenham um papel chave na evolução de grupos microbianos nos manguezais.
- As principais bactérias que compõem o microbioma do sedimento de manguezais já foram descritas com profagos ou suscetíveis à infecção por fagos (Caudivirales), como as Gammaproteobacteria e Firmicutes. Os genes relacionados à expressão de EGMs, em especial de fagos, sugerem que os EGMs desempenham papel relevante na ecologia e evolução das bactérias que vivem nos sedimentos dos manguezais, podendo afetar indiretamente os ciclos biogeoquímicos e outros serviços ecossistêmicos prestados por bactérias nesse ambiente.



## REFERÊNCIAS

ACKERMANN, H.W. Phage classification and characterization. **Methods in Molecular Biology**, Totowa, v. 501, p. 127–140, 2009.

ANDREOTE, F.D.; JIMENEZ, D.J.; CHAVES, D.; DIAS, A.C.F.; LUVIZOTTO, D.M.; DINI-ANDREOTE, D.; FASANELLA, C.C.; LOPEZ, M.V.; BAENA, S.; TAKETANI, R.G.; de MELLO, I.S. The microbiome of Brazilian mangrove sediments as revealed by metagenomics. **PLoS ONE**, Oxford, v. 7, n. 6, p. e38600, 2012.

BERNINI, E.; REZENDE, C.A. Vegetation structure in a mangrove forest in Southeastern Brazil. **Pan-American Journal of Aquatic Sciences**, Rio Claro, v. 6, n. 3, p. 193-209, 2011.

BLOWER, T.R.; EVANS, T.J.; PRZYBILSKI, R.; FINERAN, P.C.; SALMOND, G.P. Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. **PLoS Genetics**, Oxford. v. 8, n. 10, p. e1003023, 2012. doi: e1003023.

BRÜSSOW, H.; CANCHAYA, C.; HARDT, W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. **Microbiology and Molecular Biology Reviews**, Washington, v. 3, n. 68, p. 560-602, 2004. doi: 10.1128/MMBR.3.560-602.2004

CADETE, L.L. **Descrição da comunidade microbiana ativa em solos de manguezais pormetagenômica e metatranscriptômica**. 2014. 84 p. Dissertação (Mestrado em Microbiologia Agrícola) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2014.

CARVER, T.; HARRIS, S.R.; BERRIMAN, M.; PARKHILL, J.; McQUILLAN, J.A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. **Bioinformatics**, Oxford, v. 28, n. 4, p. 464-469, 2012.

CHATURONGAKUL S.; OUNJAI, P. Phage-host interplay: examples from tailed phages and gram-negative bacterial pathogens. **Frontiers in Microbiology**, Lausanne, v. 5, p. 20-25, 2014.

DIAS, A.C.F., DINI-ANDREOTE, F.; TAKETANI, R.G.; TSAI, S.M.; AZEVEDO, J.L.; de MELO, I.S.; ANDREOTE, F.D. Archaeal communities in the sediments of three contrasting mangroves. **Journal of Soils and Sediments**, Landsberg. v. 11, n. 8, p. 1466–1476, 2011.

DIAS, A.C.F.; RIGONATO, J.; FIORE, M.F.; MELO, L.S.; ARAÚJO, W.L. The bacterial diversity in Brazilian non-disturbed mangrove sediment. **Antonie Van Leeuwenhoek**, Wageningen, v. 98, n. 4, p. 541–551, 2010.

EDGAR, R.C.; HAAS, B.J.; CLEMENTE, J.C.; QUINCE, C., KNIGHT, R. UCHIME improves sensitivity and speed of chimera detection. **Bioinformatics**, Oxford. v. 27, n. 16, p. 2194–2200, 2011.

ENGELHARDT, T.; ORSI, W.D.; JORGENSEN, B.B. Viral activities and life cycles in deep sub seafloor sediments. **Environmental Microbiology Reports**, Malden MA, v. 7, p. 868-873, 2015. DOI: 10.1111/1758-2229.12316.

FASANELLA, C.C.; DIAS, ACF.; RIGONATO, J.; FIORE, M.F.; SOARES JR., F.L.; MELO, I.S.; PIZZIRANI-KLEINER, A.P.; van ELSAS, J.D.; ANDREOTE, F.D. The selection exerted by oil contamination on mangrove fungal communities. **Water air and soil pollution**, Dordrecht, v. 223, n. 7, p. 4233–4243, 2012.

FRANCO DIAS, A.C.; SILVA, M.C.P.; COTTA, S.R.; DINI-ANDREOTE, F.; SOARES JR., F.L.; SALLES, J.F.; AZEVEDO, J.L.; van ELSAS, J.D.; ANDREOTE, F.D. Abundance and genetic diversity of nifH gene sequences in anthropogenically affected Brazilian mangrove sediments. **Applied and Environmental Microbiology**, Washington, v. 78, n. 22, p. 7960–7967, Nov. 2012.

FROST, L.S.; LEPLAE, R.; SUMMERS, A.; TOUSSAINT, A. Mobile genetic elements: the agents of open source evolution. **Nature Reviews Microbiology**, London, v. 3, n. 9, p. 722–732, 2005.

GANDON, S.; BUCKLING, A.; DECAESRECKER, E.; DAY, T. Host-parasite coevolution and patterns of adaptation across time and space. **Journal of Evolution Biology**, Basel, v. 21, p. 1861-1866, 2008.

HAMMER, Ø.; HARPER, D.A.T.; RYAN, P.D. PAST: Paleontological Statistics Software package for education and data analysis. **Palaeontologia Electronica**, College Station, v. 4, n. 1, p. 1-9, 2001.

HOPKINS, M.; KAILASAN, S.; COHEN, A.; ROUX, S.; RUCKER, K.P.; SHEVENELL, A.; AGBANJE-McKENNA, M.; BREITBART, M. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of partial major capsid protein. **The ISME Journal**, London, v. 8, p. 2093-3103, 2014.

JIANG, W.; BIKARD, D.; COX, D.; ZHANG, F.; MARRAFINI, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. **Nature Biotechnology**, London, v. 31, p. 233-239, Jan. 2013. DOI: 10.1038/nbt.2508.

KLIMENKO, A.I.; MATUSHKIN, Y.G.; KOLCHANOV., N.A.; LASHIN, S.A. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study. **BMC Microbiology**, London, v. 16, n. 10, p. 31-41, 2016. DOI: 10.1186/s12866-015-0620-4.

KOSKELLA, B.; BROCKHURST, M.A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. **FEMS Microbiology Reviews**, Amsterdam, v. 38, p. 916-931, 2014.

LANG, A.S.; ZHAXYBAYEVA, O.; BEATTY, T.J. Gene transfer agents: phage-like elements of genetic exchange. **Nature Reviews Microbiology**, London, v. 10, p. 472-482, July 2012.

MEYER, F.; PAARMAN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E.M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J; EDWARDS, R.A. The metagenomics RAST server: a public resource for the automatic

phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, London, v. 9, n. 1, p. 386-386, 2008. doi: 10.1186/1471-2105-9-386.

MIHARA, T.; NISHIMURA, Y.; SHIMIZU, Y.; NISHIYAMA, H.; YOSHIKAWA, G.; UERAHA, H.; HINGAMP, P.; GOTO, S.; OGATA, H. Linking virus genomes with host taxonomy. **Viruses**, Basel, v. 66, n. 8, p. 1-6, 2016. doi: 10.3390/v8030066.

PAUL, J.H. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? **The ISME Journal**, London, v. 2, p. 579-589, 2008.

RAKONJAC, J.; BENNET, N.J.; SPAGNUOLO, J.; GAGIC, D.; RUSSEL, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. **Current Issues in Molecular Biology**, Wyomondham, v.13, p. 51-76, 2011.

RICHARDS, T.A.; TALBOT, N.J. Horizontal gene transfer in osmotrophs: playing with public goods. **Nature Reviews Microbiology**, London, v. 11, p. 720-727, 2013.

ROACH, D.R.; DONOVAN, D.M. Antimicrobial bacteriophage-derived proteins and therapeutic applications. **Bacteriophage**, Austin, v. 5, n. 3, p. e1062590, 2015.

ROUX, S.; ENAULT, F.; HURWITZ, B.L.; SULLIVAN, M.B. VirSorter: mining viral signal from microbial genomic data. **PeerJ**, Corte Madera, v. 3, p. e985, 2015. doi: 10.7717/peerj.985.

ROUX, S.; ENAULT, F.; HURWITZ, B.L.; SULLIVAN, M.B. VirSorter: mining viral signal from microbial genomic data. **PeerJ**, Corte Madera, v. 3, p. e985, 2015.

RUTHERFORD, K.; PARKHILL, J.; CROOK, J.; HORSNELL, T.; RICE, P.; RAJANDREAM, M.A BARRELL, B.. Artemis: sequence visualization and annotation. **Bioinformatics**, Oxford, v. 16, n. 10, p. 944-945, 2000.

SALMOND, G.P.C.; FINERAN, P.C. A century of the phage: past, present and future. **Nature Reviews Microbiology**, London, v. 13, p. 777-786, 2015.

SCHAEFFER-NOVELLI, Y.; CINTRÓN-MOLERO, G.; SOARES, M.L.G.; DE-ROSA, T. Brazilian mangroves. **Aquatic Ecosystem Health and Management**, New York, v. 3, p. 561-570, 2000.

SOUCY, S.M.; HUANG, J.; GOGARTEN, J.P. Horizontal gene transfer: building the web of life. **Nature Reviews Genetics**, London, v. 16, p. 472-482, 2015.

STEWART, F.J.; OTTESEN, E.A.; DeLONG, E.F. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. **The ISME Journal**, London, v. 4, p. 896-907, 2010.

VARON-LOPEZ, M.; DIAS, A.C.F.; FASANELLA, C.C.; DURRER, A.; MELO, I.S.; KURAMAE, E.E.; ANDREOTE, F.D. Sulphur-oxidizing and sulphate-reducing communities in Brazilian mangrove sediments. **Environmental Microbiology**, Oxford, v. 16, n. 3, p. 845-855, 2014.

VASCONCELLOS, S.P. de; ANGOLINI, C.F.F.; GARCÍA, I.N.S.; DELLAGNEZZE, B.M.; SILVA, C.C.; MARSAIOLI, A.J.; SANTOS NETO, E.V.; de OLIVEIRA, V.M. Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. **Organic Geochemistry**, Oxford, v. 41, n. 7, p. 675–681, 2010.

WEITZ, J.S.; POISOT, T.; MEYER, J.R.; FLORES, C.O.; VALVERDE, S; SULLIVAN, M.B.; HOCKBERG, M.E. Phage-bacteria infection networks. **Triends in Microbiology**, Cambridge, v. 21, n°2, p. 82-91, 2013.

YOSHIDA, M.; TAKAKI, Y.; EITOKU, M.; NUNOURA, T.; TAKAI, K. Metagenomic analysis of viral communities in (Hado) pelagic sediments. **PLoS ONE**, Oxford, v. 8, p. e57271, 2013.

YOUNG, R. Phage lysis: do we have the whole story yet? **Current Opinion in Microbiology**, London, v. 16, p. 790–797, 2013.

ZECHNER, E.L.; LANG, S., SHILDBACH, J.F. Assembly and mechanisms of bacterial type IV secretion machines. **Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences**, London, v. 367, p. 1073-1087, 2012.