

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”
Centro de Energia Nuclear na Agricultura**

Diversidade de cianofagos em lagoas salino-alcálinas do Pantanal

Mariana B. P. F. de Oliveira

Trabalho de conclusão de curso apresentado como
parte dos requisitos para obtenção do título de
Bacharela em Ciências Biológicas.

**Piracicaba
Ano 2022**

Mariana B. P. F. de Oliveira

Diversidade de cianofagos em lagoas salino-alcálinas do Pantanal

Orientadora: **Prof. Dr^a. MARLI DE FÁTIMA FIORE**

Trabalho de conclusão de curso apresentado como
parte dos requisitos para obtenção do título de
Bacharela em Ciências Biológicas

**Piracicaba
Ano 2022**

SUMÁRIO

RESUMO.....	4
ABSTRACT	5
LISTA DE FIGURAS	6
LISTA DE TABELAS	7
1. INTRODUÇÃO	8
1.1. Objetivo geral.....	10
2. MATERIAIS E MÉTODOS	11
2.1. Área de estudo	11
2.2. Genomas de cianobactérias	12
2.3. Obtenção dos dados de metagenoma.....	12
2.4. Métricas estatísticas	16
3. Resultados	17
3.1. Genomas de cianobactérias	17
3.2. MAGs (Metagenome Assembled Genome)	20
3.3. Diversidade de famílias	27
4. Discussão.....	30
5. Conclusão	32
REFERÊNCIAS.....	33

RESUMO

Diversidade de cianofagos em lagoas salino-alcálinas do Pantanal

Bacteriófagos (fagos) são vírus que infectam bactérias com o objetivo de assumir o controle do seu metabolismo para produzir proteínas necessárias à sua replicação. Os vírus que infectam especificamente cianobactérias são chamados de cianofagos. Esses fagos são capazes de se manter em estado latente, sem expressar genes, com o seu genoma integrado ao genoma bacteriano. O objetivo deste estudo foi usar a plataforma de bioinformática Phaster para realizar a busca de cianofagos em 4 genomas de cianobactérias [*Anabaenopsis elenkinii* CCIBt3563, *Geminocystis* sp. CENA526, *Limnospira* sp. CENA597, *Limnospira* sp. CENA650] e em 23 MAGs (Metagenome Assembled Genome) recuperadas a partir do sequenciamento metagenômico das lagoas salino-alcálinas do Pantanal. O grupo de pesquisa do Laboratório de Biologia Celular e Molecular do Centro de Energia Nuclear na Agricultura da Universidade de São Paulo (CENA-USP) coletou amostras de água nestas lagoas nos anos de 2017, 2018 e 2019. As regiões que foram identificadas como pertencentes a fagos foram analisadas para a identificação das prováveis famílias destes cianofagos usando a plataforma de bioinformática NCBI BLASTP. Foram obtidos resultados para as famílias *Myoviridae*, *Siphoviridae* e *Podoviridae*, o que indica que as cianobactérias dessas lagoas salino-alcálinas podem estar sendo infectadas por cianofagos. Identificar a presença de cianofagos é importante, pois estes organismos podem ser utilizados para realizar o controle biológico de florações de cianobactérias. Quando não são controladas, essas florações podem prejudicar a qualidade de sistemas aquáticos provocando danos ambientais, econômicos e de saúde pública.

Palavras-chave: Cianofagos, Genoma, Metagenoma, Profagos, Bioinformática

ABSTRACT

Diversity of cyanophages in saline-alkaline lagoons of the Pantanal

Bacteriophages (phages) are viruses that infect bacteria with the aim of taking control of their metabolism to produce proteins necessary for their replication. Viruses that specifically infect cyanobacteria are called cyanophages. These phages are able to keep in a latent state, without expressing genes, with their genome integrated into the bacterial genome. The aim of this study was to use the Phaster bioinformatics platform to search for cyanophages in 4 genomes of cyanobacteria [*Anabaenopsis elenkinii* CCIBt3563, *Geminocystis* sp. CENA526, *Limnospira* sp. CENA597, *Limnospira* sp. CENA650] and 23 MAGs (Metagenome Assembled Genome) recovered from the metagenomic sequencing of the soda lakes of the Pantanal. The research group of the Laboratory of Cellular and Molecular Biology of the Center for Nuclear Energy in Agriculture of the University of São Paulo (CENA-USP) collected water samples in these lakes in the years 2017, 2018 and 2019. The regions that were identified as belonging phages were analyzed for the identification of likely families of these cyanophages using the NCBI BLASTP bioinformatics platform. Results were obtained for the *Myoviridae*, *Siphoviridae* and *Podoviridae* families, which indicates that the cyanobacteria of these soda lakes may be being infected by cyanophages. Identifying the presence of cyanophages is important, as these organisms can be used to carry out the biological control of cyanobacterial blooms. When not controlled, these blooms can harm the quality of aquatic systems causing environmental, economic and public health damage.

Keywords: Cyanophages, Genome, Metagenome, Prophage, Bioinformatics

LISTA DE FIGURAS

Figura 1. Distribuição geográfica das lagoas salino-alcálinas amostradas no Pantanal da Nhecolândia, MS, Brasil.....	11
---	----

LISTA DE TABELAS

Tabela 1. Anotação de sequências identificadas como pertencentes a cianofagos no genoma de cianobactérias.	20
Tabela 2. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2017).	22
Tabela 3. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2018).....	24
Tabela 4. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2019).....	27

1. INTRODUÇÃO

Os bacteriófagos ou fagos são vírus que infectam bactérias e são reconhecidamente os organismos mais abundantes presentes na biosfera e também com maior diversidade genética (Mann, 2005). Esses organismos infectam bactérias com o objetivo de controlar o seu metabolismo e utilizá-lo para produzir proteínas essenciais a sua replicação. Para infectar a bactéria, o fago se anexa a ela em um processo conhecido como adsorção, em seguida ele realiza o traslado do seu material genético para dentro da bactéria e expressa seus genes no citoplasma da célula hospedeira com o intuito de assumir o controle do metabolismo desta e produzir as proteínas de seu interesse. Após esse processo, o fago provoca a lise celular e as novas partículas celulares são lançadas para o meio extracelular num processo conhecido como ciclo lítico (Dennehy e Abedon, 2021). Também existem fagos que possuem a capacidade de infectar a bactéria hospedeira e se manter em estado latente com o seu material genético integrado ao cromossomo bacteriano, sendo denominados de profagos. Nesse caso, não ocorre a expressão dos genes. Esses fagos que se encontram dormentes acabam por serem replicados quando a célula hospedeira se divide no que é conhecido como ciclo lisogênico (You Zhou et. al., 2011). Desse modo, sequências do genoma de fagos podem ser encontradas no genoma bacteriano, chegando a representar em torno de 20% do genoma da bactéria (Hatfull et. al., 2011).

Os cianofagos são fagos que infectam especificamente cianobactérias. Eles são morfologicamente bastante semelhantes aos demais bacteriófagos existentes. Entre as particularidades que possuem em relação aos demais fagos estão, por exemplo, a necessidade de luz para que ocorra a infecção e a sua capacidade de se manter estável em pHs que variam de 4 até 11, diferente dos demais bacteriófagos que costumam se manter estáveis em pHs de variam de 5 a 8 (Saffermann e Morris, 1964; Sherman e Brown, 1978). Segundo Saffermann (1983), taxonomicamente, os cianofagos podem ser classificados como sendo pertencentes a três famílias distintas de fagos: a família *Myoviridae*, *Siphoviridae* e *Podoviridae*. A família *Myoviridae* inclui fagos com um tubo central e uma bainha contrátil separada da cabeça por um pescoço. O gênero *Cyanomyovirus* da família *Myoviridae* se hospeda em cianobactérias unicelulares dos gêneros *Anacystis* e *Synechococcus*. Morfologicamente, possui cabeça isométrica com diâmetro de 90 nm e uma cauda

estendida de 244 x 23 nm contraindo para 93 nm; os pinos de cauda curtos são fixados a uma placa de base, com 40 nm de largura. O gênero *Cyanomyovirus* é constituído por 30 proteínas estruturais e sua replicação é feita no nucleoplasma (Saffermann, 1983).

Segundo a descrição de Safferman (1983), a família *Siphoviridae* compreende fagos com caudas longas e não contráteis. O gênero *Cyanosiphovirus* da família *Siphoviridae* possui caudas não contráteis e se hospeda em cianobactérias unicelulares do gênero *Synechococcus*. Morfologicamente, possui cabeça isométrica com diâmetro de 50 nm e cauda rígida não contrátil de 140 nm de comprimento. Possui 13 proteínas estruturais (Safferman, 1983). Segundo o mesmo autor, a família *Podoviridae*, contém fagos com caudas curtas. O gênero *Cyanopodovirus* da família *Podoviridae* possui caudas curtas e se hospeda em cianobactérias filamentosas dos gêneros *Lyngbya*, *Plectonema* e *Phormidium*. Morfologicamente, possui cabeça isométrica com diâmetro de 59 nm e uma cauda 15-20 nm comprimento e 15 nm de largura. Possui 10 proteínas estruturais e multiplicam-se na região periférica (estroma virogênico) da célula hospedeira. O deslocamento lateral das lamelas fotossintéticas é a primeira alteração morfológica observada (Safferman, 1983).

Hurst (2011) relata que até o momento, a grande maioria dos fagos líticos de *Synechococcus* e *Prochlorococcus* são da família *Myoviridae*, possivelmente porque os myovirus podem ser mais fáceis de isolar do que os podovirus e siphovirus. Tal como acontece com os resultados da maioria dos estudos baseados em cultura, esses dados podem não ser representativos do padrão que existe na natureza.

Esses cianofagos têm sido estudados devido a possibilidade de utilizá-los para realizar o controle biológico de florações de cianobactérias que são nocivas para ambientes aquáticos e produzem substâncias tóxicas. Em ambientes aquáticos de com grande abundância de cianofagos nota-se que as cianobactérias não são abundantes (Safferman e Morris, 1964). Realizar o controle destas florações é fundamental, pois elas podem comprometer a qualidade de sistemas aquáticos provocando danos ambientais, econômicos e de saúde pública.

Atualmente, diversas plataformas de bioinformática estão sendo desenvolvidas com o objetivo de identificar essas sequências de fagos em genomas bacterianos e também em metagenomas ambientais. Essas plataformas atuam

realizando uma varredura do genoma bacteriano para encontrar locais abundantes em material genético viral. O método computacional para aferir a presença de fagos em genomas de bactérias que já tiveram o seu genoma sequenciado surge como uma alternativa aos métodos experimentais tradicionais de detecção de fagos que consistiam em induzir na bactéria hospedeira as condições necessárias para que liberasse fagos viáveis (Hurwitz et. al., 2017).

De acordo com You Zhou et. al. (2011), analisar sequências de cianofagos em genomas sequenciados de cianobactérias e MAGs (Metagenome Assembled Genome) utilizando o método computacional tem como vantagem em relação a outros métodos experimentais que só detectam fagos viáveis, a possibilidade de detectar a presença de profagos cujos genomas se encontram integrados a genomas bacterianos. Métodos experimentais podem ser falhos, já que nem sempre as condições necessárias para induzir a bactéria hospedeira a liberar fagos viáveis serão iguais para todos os fagos e para que não ocorra falhas é necessário saber quais são essas condições. Conhecer o genoma de cianofagos pode também nos trazer informações acerca da coevolução destes e das cianobactérias, já que estas adquirem genes dos fagos, por meio da transferência horizontal de genes, que podem ser úteis a sua adaptação e sobrevivência no meio em que habitam (You Zhou et. al., 2011).

1.1. Objetivo geral

Este trabalho tem como objetivo identificar, por meio de plataformas computacionais, sequências de prováveis cianofagos em 4 genomas de cianobactérias e de 23 MAGs provenientes de amostras de águas coletadas em seis lagoas salino-alcálinas localizadas na Reserva de São Roque no município de Aquidauana (MS) nos anos de 2017, 2018 e 2019. O segundo objetivo é identificar as famílias a que pertencem esses prováveis cianofagos.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo

As lagoas salino-alcálinas são caracterizadas pela grande quantidade de carbonatos e bicarbonatos em suas águas resultando em altamente alcalinas e frequentemente salinas. O pH dessas lagoas pode variar de 9,5 a 11 (Sorokin et al., 2014; Tindall, 1988).

Na sub-região do Pantanal sul Matogrossense, conhecida como Nhecolândia, o grupo do Laboratório de Biologia Celular e Molecular do Centro de Energia Nuclear na Agricultura da Universidade de São Paulo (CENA-USP) coletou amostras de água de lagoas salino-alcálinas na Reserva de São Roque no município de Aquidauana (MS). Essas coletas foram realizadas em três anos consecutivos: em outubro de 2017 durante o período de seca, em setembro de 2018 durante o período de chuva e em 2019 durante o período de seca. Na coleta realizada no ano 2017, o período de seca foi mais intenso do que em outros anos e a coluna d'água se encontrava bastante reduzido. Este estudo realizado por Pellegrinetti, et al. (2022) aferiu a presença de cianobactérias nessas lagoas do Pantanal.

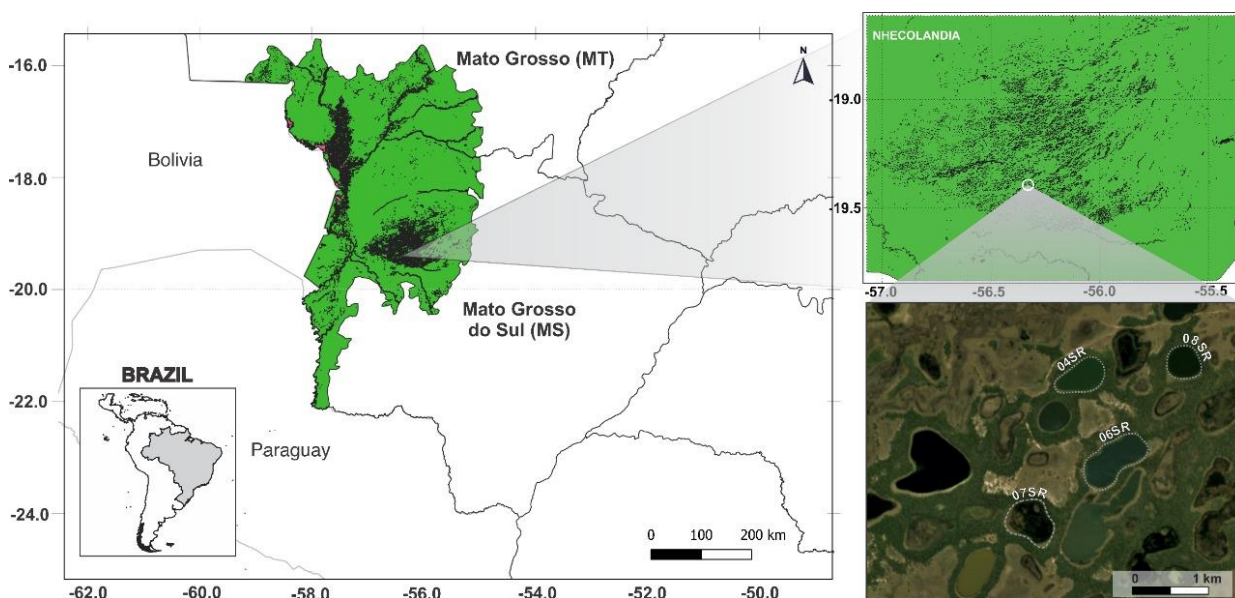


Figura 1. Distribuição geográfica das lagoas salino-alcálinas amostradas no Pantanal da Nhecolândia, MS, Brasil.

Fonte: Pellegrinetti, Thierry A. et al.; 2022.

2.2. Genomas de cianobactérias

Após a coleta, este mesmo grupo realizou em laboratório o isolamento das cianobactérias presentes nas amostras coletadas de água. Após a sua purificação (uma única cianobactéria por cultivo), o seu DNA foi extraído e enviado para sequenciamento no Centro de Genômica funcional no Departamento de Zootecnia da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP). As estirpes analisadas foram quatro: *Anabaenopsis elenkinii* CCIBt3563, *Geminocystis* sp. CENA526, *Limnospira* sp. CENA597, *Limnospira* sp. CENA650. A plataforma utilizada para realizar o sequenciamento dos genomas das cianobactérias isoladas foi a “Illumina HiSeq 2500” (Illumina, Inc., EUA).

2.3. Obtenção dos dados de metagenoma

2.3.1. Extração de DNA total

Para a obtenção dos dados de metagenoma das amostras de água coletadas, estas foram submetidas ao processo de liofilização antes de obter o seu DNA. Depois, foi utilizado o kit “PowerLyzer® PowerSoil® DNA Isolation” (MoBio Laboratories Inc., Carlsbad, CA, EUA) para realizar a extração de DNA das amostras de água liofilizadas e não filtradas. No espectrofotômetro “NanoDrop” e no fluorômetro “Qubit 2.0” (Thermo Fisher Scientific, Waltham, MA, EUA), foi realizado a quantificação desse DNA extraído. Em seguida, foi preparado em laboratório um gel de agarose 1% (m/v) em solução tampão TBE 1x (400 mM Tris, 20 mM ácido acético global, 1mM EDTA), com aplicação de 1 µL do DNA extraído com adição de 2 µL de corante fluorescente GelRed TM para avaliar a integridade do DNA que foi extraído por meio da eletroforese. Por fim, utilizou-se um fotodocumentador da marca “Kodak Gel Logic 212 – Molecular Imaging Software, v. 5,0” (Carestream Health, INC., Rochester, EUA) para visualizar o resultado (Pellegrinetti, et al., 2022).

2.3.2. Sequenciamento dos metagenomas

Para realizar o sequenciamento dos metagenomas, Pellegrinetti, et al. (2022) utilizou o kit “Nextera XT DNA Sample Preparation” (Illumina, Inc., EUA) para criar

bibliotecas metagenômicas. O Laboratório de Multiusuários Centralizado de Genômica Funcional Aplicada à Agropecuária e Agroenergia, da Universidade de São Paulo (ESALQ-USP) foi o responsável por processar e sequenciar as amostras de DNA que passaram pelos processos de purificação e fragmentação, para isso o laboratório fez uso da plataforma “Illumina HiSeq 2500” (Illumina, Inc., EUA) com o kit “HiSeq Reagent Kit V4” (Illumina, Inc., EUA). E como última etapa, foi utilizada a ferramenta “FastQC 0.10.1” para criar gráficos que permitissem avaliar a qualidade do sequenciamento metagenômico.

Em seguida, é feita a recuperação dos genomas por meio da sobreposição dos fragmentos sequenciados. Durante esse processo, busca-se analisar fragmentos que sejam similares e que se complementem. Após essas montagens, os genomas que foram recuperados pelo grupo de pesquisa por meio do sequenciamento metagenômico foram filtrados. Estabeleceu-se uma completude de sequência superior a 70% e contaminação inferior a 10% com a plataforma CheckM. A classificação taxonômica e análises filogenômicas foram realizadas utilizando a plataforma de bioinformática chamada GTDBtk que, por sua vez, usa como banco de dados o Genome Taxonomy Database (GTDB). As 23 MAGs recuperadas que foram afiliadas ao gênero *Cyanobacteria* foram selecionadas e analisadas. A descrição das MAGs utilizadas nesse trabalho é feita a seguir.

No ano de 2017, foram analisados três MAGs identificadas como pertencentes ao filo *Cyanobacteria* recuperados dos metagenomas de duas lagoas salino-alcálinas (SR04 e SR06) do Pantanal. As MAGs analisadas foram: *Nodularia* (MAG27-SR04-2017), *Arthrospira platensis* (MAG41-SR04-2017) e *Cyanobiaceae* PCC7001 (MAG17-SR06-2017).

No ano de 2018, foram analisados 10 MAGs identificadas como pertencentes ao filo *Cyanobacteria* recuperados dos metagenomas de 4 lagoas salino alcálinas (SR01, SR04, SR05 e SR08) do Pantanal. As MAGs analisadas são: *Phormidesmiaceae* (MAG11-SR01-2018), *Nodularia* (MAG19-SR01-2018), *Phormidesmiaceae* (MAG14-SR04-2018), *Arthrospira platensis* (MAG36-SR04-2018), *Nodularia* (MAG42-SR04-2018), *Cyanobiaceae* (MAG48-SR05-2018), *Phormidesmiaceae* (MAG49-SR05-2018), *Nodularia* (MAG18-SR08-2018), *Raphidiopsis brookii* (MAG45-SR08-2018) e *Snowella* (MAG56-SR08-2018).

No ano de 2019, foram analisados 10 MAGs identificados como pertencentes ao filo Cyanobacteria recuperados dos metagenomas de cinco lagoas salino-alcálicas (SR01, SR04, SR05, SR06, SR08 e SR09) do Pantanal. As MAGs analisadas são: *Phormidesmiaceae* (MAG33-SR01-2019), *Cyanobiaceae* (MAG44-SR01-2019), *Cyanobiaceae* PCC7001 (MAG12-SR04-2019), *Nodularia* (MAG19-SR04-2019), *Raphidiopsis brookii* (MAG34-SR05-2019), *Nodularia* (MAG35-SR05-2019), *Cyanobiaceae* PCC7001 (MAG19-SR06-2019), *Raphidiopsis brookii* (MAG7-SR08-2019), *Phormidesmiaceae* (MAG13-SR08-2019) e *Cyanobacteriia* (MAG26-SR09-2019).

2.3.3. PHAge Search Tool Enhanced Release: Phaster

Para averiguar a presença de sequências genômicas de cianofagos que se encontram integradas ao cromossomo bacteriano nos 4 genomas de cianobactérias e nas 23 MAGs, utilizou-se a plataforma de bioinformática Phaster. Esta plataforma mede a densidade de agrupamento de genes, identifica e anota sequências de profago que estão contidas no genoma de bactérias e identifica as proteínas que podem ajudar na identificação (You Zhou et. al., 2011). A plataforma Phaster foi escolhida para a realização deste trabalho por ser a mais popular, mais precisa e mais rápida na análise de resultados (Arndt et. al., 2016).

O Phaster utiliza dois bancos de dados: o banco de dados de fagos do National Center for Biotechnology Information (NCBI) que possui 46.407 proteínas de 598 genomas de fagos e o Prophage Database desenvolvido por Srividhya et al. (2006), que possui 159 regiões de profago e 9.061 proteínas não disponíveis no banco de fagos do NCBI. A plataforma só apresenta como resultado as proteínas cuja função em fagos é conhecida para evitar que proteínas bacterianas sejam tidas erroneamente como sendo pertencentes a fagos (You Zhou et. al., 2011). Após consultar os dois bancos de dados, a plataforma realiza o agrupamento de genes que apresentam semelhança entre si com base em um algoritmo chamado DBSCAN. Esses genes aparecem nos resultados agrupados em regiões de profagos (Arndt et. al., 2016).

Esta plataforma tem como limitação, a dificuldade de identificar novos fagos. Desse modo, a presença de muitas proteínas cuja função não é conhecida pode

apontar para um fago novo (You Zhou et. al., 2011). Também é importante ressaltar, que os resultados sempre têm um grau de incerteza quanto a extensão do profago que foi previsto pela plataforma (Arndt et al., 2019).

A plataforma aceita arquivos no formato Gen Bank e FASTA. Neste trabalho, foi processado no Phaster os quatro genomas de cianobactérias e as 23 MAGs em arquivos FASTA para procurar por prováveis regiões de profago de cianofagos. Após a identificação de uma provável região de profago pelo programa, as sequências genômicas das proteínas indicadas como pertencentes a este foram avaliadas pelo seu valor de e-value. As proteínas com valor de e-value abaixo de 0.001 tiveram as suas sequências utilizadas para a identificação da provável família do profago.

2.3.4. NCBI BLAST: Basic Local Alignment Search Tool

A plataforma de bioinformática utilizada para identificar as famílias é a NCBI BLAST. O BLAST é o software de bioinformática mais utilizado do mundo. Ele analisa sequências de proteínas, RNA e DNA possibilitando descobrir a sua função e a classificação taxonômica do organismo do qual aquela sequência genômica é originária. Para isso a sequência de uma proteína, por exemplo, é comparada as sequências proteicas dos bancos de dados do NCBI para descobrir regiões de similaridade, ou seja, é realizado o alinhamento das sequências para saber quantos aminoácidos elas possuem em comum (Stover e Cavalcanti, 2017).

Após identificar as prováveis regiões de profagos dos genomas de cianobactérias e das MAGs utilizando a plataforma de bioinformática Phaster, a próxima etapa metodológica do trabalho foi tentar identificar as prováveis famílias desses profagos utilizando as sequências genômicas das proteínas fornecidas pelo Phaster. Foram utilizadas as proteínas que apresentaram valor de e-value abaixo de 0.001.

O Blast possui cinco plataformas: BLASTN, BLASTP, BLASTX, TBLASTN, TBLASTX. A plataforma escolhida para identificar as famílias foi o BLASTP, esta plataforma é usada especificamente para consultar a sequência genômica de uma proteína que será comparada com as sequências proteicas homólogas presentes nos bancos de dados do NCBI (Stover e Cavalcanti, 2017).

Todas as sequências de proteínas das prováveis regiões de profago identificadas pelo Phaster foram processadas no BLASTP para descobrir a classificação das famílias desses cianofagos. Para avaliar a qualidade dos alinhamentos obtidos no BLASTP, foram avaliadas as seguintes métricas estatísticas: e-value, a porcentagem de query cover e porcentagem de identidade.

2.4. Métricas estatísticas

2.4.1. e-value

O e-value é útil para avaliar a qualidade e escolher o melhor alinhamento. É recomendado que ele seja menor do que 0.001. Quanto mais próximo de zero, melhor é o e-value e melhor é o alinhamento. Essa métrica indica a probabilidade de o alinhamento ocorrer ao acaso, tendo em vista um banco de dados aleatório e do mesmo tamanho do banco de dados alvo.

Segundo Stover e Cavalvanti (2017), o e-value é a métrica mais importante de ser analisada. A probabilidade da sequência genômica que está sendo analisada apresentar ancestralidade comum com o resultado aumenta, quanto mais próximo de zero for o e-value. A confiança da análise feita sobre os resultados obtidos também será maior.

Neste trabalho, foram consideradas e avaliadas as sequências genômicas de proteínas fornecidas pelo Phaster com e-value < 0.001. Os resultados fornecidos pelo BLASTP acerca da classificação das famílias das prováveis regiões de profago também foram avaliados quanto a esta métrica privilegiando sempre o valor de e-value < 0.001 e mais próximo de zero.

2.4.2. Query Cover

O query cover é uma métrica estatística que indica a porcentagem da sequência genômica que está sendo analisada que se encontra alinhada a sequência encontrada no banco de dados pela plataforma, ou seja, aponta a extensão da sequência que está alinhada, que corresponde a sequência alvo (Boratyn et. al., 2013).

Neste trabalho, foram considerados os resultados dos alinhamentos obtidos no BLASTP com a maior porcentagem de query cover possível.

2.4.3. Porcentagem de identidade

A porcentagem de identidade é uma métrica estatística que indica a proporção de matches dentro de um alinhamento em relação ao total alinhado. O seu cálculo é feito dividindo o número de letras que são idênticas pelo comprimento do alinhamento que, por sua vez, é a soma do número de letras que deram match mais o número de lacunas (Boratyn et. al., 2013).

Neste trabalho, foram considerados os resultados dos alinhamentos obtidos no BLASTP com uma porcentagem de identidade acima de 50%.

3. Resultados

3.1. Genomas de cianobactérias

3.1.1. *Anabaenopsis elenkinii* CCIBt3563

O genoma da cianobactéria *Anabaenopsis elenkinii* CCIBt3563 processado no Phaster apresentou como resultados 2 prováveis regiões de profago.

Na região 1, o processamento da sequência genômica desta proteína no BLASTP obteve um resultado para a família *Siphoviridae* com bons valores de e-value e query cover. A porcentagem de identidade de 62.80% ficou acima dos 50% estabelecidos para este trabalho, embora não seja um valor muito alto. Na oitava e nona colunas da Tabela 1, é possível observar respectivamente a função da sequência identificada e o tamanho da sequência comparada.

Na região 2, novamente foi obtido no BLASTP resultado para a família *Siphoviridae* com bons valores de e-value e query cover. A porcentagem de identidade é de 62%.

3.1.2. *Limnospira* sp. CENA597

O genoma da cianobactéria *Limnospira* sp. CENA597 processado no Phaster apresentou como resultados 8 prováveis regiões de profago. Na região 1, a sequência genômica foi processada no BLASTP obtendo um resultado para a família *Myoviridae*. Os valores de e-value, query cover e porcentagem de identidade podem ser observados na Tabela 1.

Nas regiões de profago 2, 3, 4, 5, 6 e 7, o Phaster identificou proteínas com valores de e-value menores do que 0.001. No entanto, ao processar as sequências dessas proteínas no BLASTP as porcentagens de identidade ficaram abaixo dos 50% estabelecidos para este trabalho, não sendo possível identificar ou sugerir uma classificação taxonômica e constam como não classificados na Tabela 1.

Na região 8, a sequência genômica dessa proteína obteve resultado para a família *Siphoviridae* no BLASTP. Os valores de e-value, query cover e porcentagem de identidade podem ser observados na Tabela 1.

3.1.3. *Limnospira* sp. CENA 650

O genoma da cianobactéria *Limnospira* sp. CENA 650 processado no Phaster apresentou como resultados 8 prováveis regiões de profago. Na região 1, 4 e 5, o Phaster identificou proteínas com valores de e-value menores do que 0.001. No entanto, ao processar as sequências dessas proteínas no BLASTP as porcentagens de identidade ficaram abaixo dos 50% estabelecidos para este trabalho.

Na região 2, as proteínas identificadas pelo Phaster apresentaram valores ruins de e-value. Por isso, as sequências genômicas dessas proteínas não foram processadas no BLASTP.

Na região 3, a sequência genômica da proteína obteve resultado para a família *Siphoviridae*. Na região 6, a sequência genômica da proteína obteve resultado para a família *Myoviridae* no BLASTP. Na região 7, a sequência genômica da proteína obteve resultado para a família *Podoviridae* no BLASTP. Na região 8, a sequência da proteína obteve resultado para a família *Myoviridae* no BLASTP.

Todos os valores de e-value, query cover e porcentagem de identidade podem ser observados na Tabela 1. Assim como a função da sequência identificada e o tamanho da sequência comparada.

3.1.4. *Geminocystis* sp. CENA650

O genoma da cianobactéria *Geminocystis* sp. CENA526 processado no Phaster apresentou como resultado 1 provável região de profago. A sequência genômica da proteína foi analisada no BLASTP. A família que apareceu com o score mais alto foi a *Inoviridae*, que é uma família de bacteriófagos, mas não de cianofagos especificamente. A segunda família que apareceu com o score mais alto foi a família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover. A porcentagem de identidade é de 57.19%.

Tabela 1. Anotação de sequências identificadas como pertencentes a cianofagos no genoma de cianobactérias.

Genoma de Cianobactérias	Quant. de prováveis regiões de profago	Região	Phaster e-value	Blast e-value	Query cover	% identidade	Função da sequência identificada	Tamanho da sequência comparada (Length)	Família
<i>Anabaenopsis elenkinii</i> CCIBt3563	2	1	6.29e-137	0.0	97%	62.80	TPA: S adenosylmethionine synthase	424	<i>Siphoviridae</i>
		2	1.05e-113	0.0	93%	62.00	TPA: chaperonin	540	<i>Siphoviridae</i>
<i>Limnospira sp.</i> CENA597	8	1	1.80e-22	3.0e-30	57%	66.67	Transposase	402	<i>Myoviridae</i>
		2	9.84e-20	-	-	-	-	-	Não classificado
		3	1.80e-103	-	-	-	-	-	Não classificado
		4	3.01e-39	-	-	-	-	-	Não classificado
		5	3.65e-50	-	-	-	-	-	Não classificado
		6	3.92e-38	-	-	-	-	-	Não classificado
		7	4.86e-75	-	-	-	-	-	Não classificado
		8	1.32e-35	8.0e-39	75%	60.18	TPA: endonuclease	367	<i>Siphoviridae</i>
<i>Limnospira sp.</i> CENA650	8	1	1.25e-48	-	-	-	-	-	Não classificado
		2	valores ruins de e-value	-	-	-	-	-	Não classificado
		3	5.32e-38	9.0e-48	96%	55.56	dUTPase	154	<i>Siphoviridae</i>
		4	9.55e-49	-	-	-	-	-	Não classificado
		5	3.92e-38	-	-	-	-	-	Não classificado
		6	2.26e-36	2.0e-59	96%	59.49	Hypothetical protein b3_0124	218	<i>Myoviridae</i>
		7	7.80e-30	1.0e-30	77%	51.52	Pentapeptide repeat family protein	356	<i>Podoviridae</i>
		8	1.10e-95	9.0e-95	43%	56.83	Formylglycine-generating enzyme family protein	266	<i>Myoviridae</i>
<i>Geminocystis sp.</i> CENA526	1	1	1.81e-86	0.0	98%	58.29	TPA: replicative helicase	869	<i>Inoviridae</i>
		1	1.81e-86	0.0	99%	57.19	TPA: replicative helicase	863	<i>Myoviridae</i>

Legenda: (-) Sequência analisada não apresentou resultado ou apresentou resultado com métricas estatísticas abaixo dos valores estabelecidos para este trabalho.

3.2. MAGs (Metagenome Assembled Genome)

3.2.1. MAGs 2017

A MAG *Nodularia* (MAG27-SR04-2017) processada no Phaster apresentou como resultado 3 prováveis regiões de profago. Na região 1, a sequência genômica desta proteína foi identificada pela plataforma BLASTP como sendo pertencente à

família *Myoviridae* com bons valores de e-value e de query cover. A porcentagem de identidade ficou em 51.04%, acima do estabelecido para este trabalho.

Na região 2, a sequência genômica da proteína obtida no Phaster apresentou valor de e-value igual a zero, o que é excelente. No entanto, ao ter sua sequência processada no BLASTP o resultado obtido foi de vírus de DNA não classificado. Ao processar no BLASTP, a segunda proteína com melhor e-value obtida no Phaster, a proteína com valor de e-value de 2.45×10^{-82} , foi obtido resultado para a família *Myoviridae*.

Na região 3, a sequência genômica da proteína obtida no Phaster com e-value obteve resultado para a família Siphoviridae no BLASTP. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Arthrospira platensis* (MAG41-SR04-2017) processada do Phaster apresentou como resultado 1 provável região de profago. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Siphoviridae*.

A MAG *Cyanobiaceae* PCC7001 (MAG17-SR06-2017) processada no Phaster apresentou como resultado 2 prováveis regiões de profago. As proteínas dessas 2 regiões apresentaram valores de e-value menores do que 0.001. No entanto, ao terem as suas sequências processadas no BLASTP apresentaram valores de porcentagem de identidade menores do que os 50% estabelecidos para este trabalho. E por isso, as suas famílias constam como não classificadas na Tabela 2.

Em todos os resultados apresentados, a porcentagem de identidade é superior a 50%.

Tabela 2. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2017).

MAGs	Quant. de prováveis regiões de profago	Região	Phaster e-value	Blast e-value	Query cover	% identidade	Função da sequência identificada	Tamanho da sequência comparada (Length)	Família
<i>Nodularia</i> (MAG27-SR04-2017)	3	1	4.15e-108	2.0e-117	98%	51.04	TPA: GDP-D-mannose dehydratase	344	<i>Myoviridae</i>
		2	0.0	1.0e-171	86%	54.84	AAA ATPase	524	Vírus de DNA não classificado
		2	2.45e-82	2.0e-98	96%	54.05	TPA: cysteine synthase A	310	<i>Myoviridae</i>
		3	8.00e-172	0.0	100%	53.68	TPA: DNA TOPOISOMERASE IV, B SUBUNIT	640	<i>Siphoviridae</i>
<i>Arthrospira platensis</i> (MAG41-SR04-2017)	1	1	5.32e-38	9.0e-48	96%	55.56	dUTPase	154	<i>Siphoviridae</i>
<i>Cyanobiaceae</i> PCC7001 (MAG17-SR06-2017)	2	1	4.05e-58	-	-	-	-	-	Não classificado
		2	6.05e-122	-	-	-	-	-	Não classificado

Legenda: (-) Sequência analisada não apresentou resultado ou apresentou resultado com métricas estatísticas abaixo dos valores estabelecidos para este trabalho.

3.2.2. MAGS 2018

Na Tabela 3, as MAGS *Phormidesmiaceae* (MAG14-SR04-2018) e *Cyanobiaceae* (MAG48-SR05-2018) foram processadas no Phaster e não apresentaram resultados para prováveis regiões de profagos.

As MAGs *Phormidesmiaceae* (MAG11-SR01-2018), *Phormidesmiaceae* (MAG49-SR05-2018), *Nodularia* (MAG18-SR08-2018), *Raphidiopsis brookii* (MAG45-SR08-2018) e *Snowella* (MAG56-SR08-2018) foram processadas na plataforma Phaster e apresentaram resultados para prováveis regiões de profagos. As proteínas obtidas apresentaram valores de e-value menores do que 0.001. No entanto, ao serem processadas no BLASTP não foram obtidos resultados para famílias de fagos.

A MAG *Nodularia* (MAG19-SR01-2018) processada no Phaster apresentou 1 provável região de profago. A proteína com valor de e-value igual a zero foi analisada no BLASTP e foi identificada pela plataforma como vírus bacteriano não classificado. A segunda proteína com menor valor de e-value (2.45 e-82) foi analisada no BLASTP e obteve resultado para a família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Arthrospira platensis* (MAG36-SR04-2018) processada no Phaster apresentou 1 provável região de profago. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Siphoviridae*.

A MAG *Nodularia* (MAG42-SR04-2018) foi processada no Phaster e apresentou resultados para 3 prováveis regiões de profagos. As 3 regiões apresentaram bons valores de e-value e query cover no BLASTP. As porcentagens de identidades ficaram acima dos 50% estabelecidos neste trabalho. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Myoviridae*. A sequência genômica da região 2 foi identificada como sendo pertencente à família *Siphoviridae*. A sequência genômica da região 3 identificada como sendo pertencente à família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

Em todos os resultados apresentados, a porcentagem de identidade é superior a 50%.

Tabela 3. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2018)

MAGs	Quant. de prováveis regiões de profago	Região	Phaster e-value	Blast e-value	Query cover	% identidade	Função da sequência identificada	Tamanho da sequência comparada (Length)	Família
<i>Phormidesmiaceae</i> (MAG11-SR01-2018)	1	1	6.68e-17	-	-	-	-	-	Não classificado
<i>Nodularia</i> (MAG19-SR01-2018)	1	1	0.0	1.0e-171	86%	54.84	AAA ATPase	524	Vírus bacteriano não classificado
		1	2.45e-82	1.0e-98	96%	54.05	TPA: cysteine synthase A	310	<i>Myoviridae</i>
<i>Phormidesmiaceae</i> (MAG14-SR04-2018)	Sem resultado	-	-	-	-	-	-	-	-
<i>Arthrospira platensis</i> (MAG36-SR04-2018)	1	1	5.32e-38	9.0e-48	96%	55.56	dUTPase	154	<i>Siphoviridae</i>
<i>Nodularia</i> (MAG42-SR04-2018)	3	1	4.15e-108	2.0e-117	98%	51.04	TPA: GDP-D-mannose dehydratase	344	<i>Myoviridae</i>
		2	8.00e-172	0.0	100%	53.68	TPA: DNA TOPOISOMERASE IV, B SUBUNIT	640	<i>Siphoviridae</i>
		3	2.45e-82	1.0e-98	96%	54.05	TPA: cysteine synthase A	310	<i>Myoviridae</i>
<i>Cyanobiaceae</i> (MAG48-SR05-2018)	Sem resultado	-	-	-	-	-	-	-	-
<i>Phormidesmiaceae</i> (MAG49-SR05-2018)	1	1	1.41e-13	-	-	-	-	-	Não classificado
<i>Nodularia</i> (MAG18-SR08-2018)	1	1	0.0	-	-	-	-	-	Não classificado
<i>Raphidiopsis brookii</i> (MAG45-SR08-2018)	2	1	2.23e-171	-	-	-	-	-	Não classificado
		2	3.67e-114	-	-	-	-	-	Não classificado
<i>Snowella</i> (MAG56-SR08-2018)	1	1	1.89e-76	-	-	-	-	-	Não classificado

Legenda: (-) Sequência analisada não apresentou resultado ou apresentou resultado com métricas estatísticas abaixo dos valores estabelecidos para este trabalho.

3.2.3. MAGs 2019

A MAG *Phormidesmiaceae* (MAG33-SR01-2019) foi analisada na plataforma Phaster e não apresentou resultados para a presença de regiões de profagos.

A MAG *Cyanobiaceae* (MAG44-SR01-2019) foi analisada no Phaster e apresentou como resultado 4 prováveis regiões de profagos. As proteínas da região 2 apresentaram valores ruins de e-value, menores que 0.001, e por isso não foram analisadas no BLASTP. Na região 3, apesar da proteína apresentar um bom valor de e-value, ao ser analisada no BLASTP os valores da porcentagem de identidade ficaram baixo dos 50% estabelecidos para este trabalho. Na região 1, a sequência

genômica da proteína com maior valor de e-value (7.89 e-1170) foi identificada pelo BLASTP como sendo pertencente à família *Phycodnaviridae*. A segunda proteína com maior valor de e-value (1.81e-120) foi identificada pelo BLASTP como sendo pertencente à família *Myoviridae*. Na região 4, a sequência genômica da proteína foi identificada como sendo pertencente à família *Myoviridae*. No BLASTP, as 2 regiões identificadas apresentaram bons valores de e-value e query cover como pode ser observado na Tabela 4.

A MAG *Cyanobiaceae* PCC7001 (MAG12-SR04-2019) apresentou como resultado 1 provável região de profago no Phaster. No BLASTP, a sequência genômica da proteína foi identificada como sendo pertencente à família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Nodularia* (MAG19-SR05-2019) apresentou como resultado 1 provável região de profago no Phaster. A proteína com maior valor de e-value (0.0) foi identificada pelo BLASTP como vírus de DNA não classificado. A segunda proteína com maior valor de e-value (2.45 e-82) foi identificada pelo BLASTP como sendo pertencente à família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Raphidopsis brookii* (MAG34-SR05-2019) apresentou como resultado 2 prováveis regiões de profago no Phaster. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Siphoviridae*. A sequência genômica da proteína da região 2 foi identificada como sendo pertencente à família *Myoviridae*. A proteína do Phaster apresentou bom valor de e-value. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Nodularia* (MAG35-SR05-2019) apresentou como resultado 2 prováveis regiões de profago no Phaster. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Siphoviridae*. Na região 2, a proteína com maior valor de e-value (0.0) foi identificada pelo BLASTP como vírus de DNA não classificado. A segunda proteína com maior valor de e-value (2.45 e-82) foi identificada pelo BLASTP como sendo pertencente à família *Myoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Cyanobiaceae* PCC7001 (MAG19-SR06-2019) apresentou como resultado 3 prováveis regiões de profago no Phaster. No BLASTP, a sequência genômica da proteína da região 1 e 2 foi identificada como sendo pertencente à família *Siphoviridae*. Na região 3, a sequência genômica da proteína obtida no Phaster apresentou um valor e-value menor que 0.001, no entanto, o alinhamento obtido no BLASTP apresentou porcentagem de identidade abaixo dos 50% estabelecidos neste trabalho. Por esse motivo, a região de profago consta como não classificada na Tabela 4.

A MAG *Raphidopsis brookii* (MAG7-SR08-2019) apresentou como resultado 1 provável região de profago no Phaster. No BLASTP, a sequência genômica da proteína foi identificada como sendo pertencente à família *Myoviridae*. A proteína do Phaster apresentou bom valor de e-value. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

A MAG *Phormidesmiaceae* (MAG13-SR08-2019) apresentou como resultado 1 provável região de profago no Phaster. No BLASTP, a sequência genômica da proteína foi identificada como sendo pertencente à família *Siphoviridae*.

A MAG *Cyanobacteriia* (MAG26-SR09-2019) apresentou como resultado 2 prováveis regiões de profago no Phaster. No BLASTP, a sequência genômica da proteína da região 1 foi identificada como sendo pertencente à família *Siphoviridae*. Na região 2, a sequência genômica da proteína foi identificada pelo BLASTP como sendo pertencente à família de bacteriófagos *Herelleviridae*. O segundo resultado com maior score que apareceu para esta região de profago foi o da família *Siphoviridae*. Os alinhamentos do BLASTP apresentaram bons valores de e-value e query cover.

Em todos os resultados apresentados, a porcentagem de identidade é superior a 50%.

Tabela 4. Anotação de sequências identificadas como pertencentes a cianofagos em metagenomas (Ano 2019)

MAGs	Quant. de prováveis regiões de profago	Região	Phaster e-value	Blast e-value	Query cover	% identidade	Função da sequência identificada	Tamanho da sequência comparada (Length)	Família
<i>Phormidesmiaceae</i> (MAG33-SR01-2019)	Sem resultado	-	-	-	-	-	-	-	-
<i>Cyanobiaceae</i> (MAG44-SR01-2019)	4	1	7.89e-170	7.0e-172	92%	67.24	GDP-D-mannose dehydratase	374	<i>Phycodnaviridae</i>
		1	1.81e-120	9.0e-124	88%	62.01	glucose-1-phosphate thymidyltransferase	296	<i>Myoviridae</i>
		2	Valores ruins de e-value	-	-	-	-	-	Não classificado
		3	2.77e-115	-	-	-	-	-	Não classificado
		4	6.60e-32	8.0e-100	99%	53.46	TPA: hypothetical protein	325	<i>Myoviridae</i>
<i>Cyanobiaceae</i> PCC7001 (MAG12-SR04-2019)	1	1	3.56e-113	1.0e-131	97%	62.67	Glucose-1-phosphate thymidyltransferase	296	<i>Myoviridae</i>
<i>Nodularia</i> (MAG19-SR04-2019)	1	1	0.0	1.0e-171	86%	54.84	AAA ATPase	524	Vírus bacteriano não classificado
		1	2.45e-82	1.0e-98	96%	54.05	TPA: cysteine synthase A	310	<i>Myoviridae</i>
<i>Raphidiopsis brookii</i> (MAG34-SR05-2019)	2	1	5.10e-171	0.0	99%	61.59	TPA: ATP-dependent protease ATP-binding subunit	435	<i>Siphoviridae</i>
		2	3.67e-114	2.0e-130	97%	62.72	Glucose-1-phosphate thymidyltransferase	296	<i>Myoviridae</i>
<i>Nodularia</i> (MAG35-SR05-2019)	2	1	8.00e-172	0.0	100%	53.68	TPA: DNA TOPOISOMERASE IV, B SUBUNIT	640	<i>Siphoviridae</i>
		2	0.0	3.0e-171	86%	54.84	AAA ATPase	524	Vírus de DNA não classificado
		2	2.45e-82	1.0e-98	96%	54.05	TPA: cysteine synthase A	310	<i>Myoviridae</i>
<i>Cyanobiaceae</i> PCC7001 (MAG19-SR06-2019)	3	1	6.77e-22	2.0e-43	51%	55.84	Putative recombinase	210	<i>Siphoviridae</i>
		2	8.65e-50	2.0e-102	90%	50.85	TPA: Fic family protein	385	<i>Siphoviridae</i>
		3	2.60e-58	-	-	-	-	-	Não classificado
<i>Raphidiopsis brookii</i> (MAG7-SR08-2019)	1	1	3.67e-114	2.0e-130	97%	62.72	Glucose-1-phosphate thymidyltransferase	296	<i>Myoviridae</i>
<i>Phormidesmiaceae</i> (MAG13-SR08-2019)	1	1	9.83e-17	0.0	78%	61.56	TPA: DNA-directed RNA polymerase subunit beta	1046	<i>Siphoviridae</i>
<i>Cyanobacteriia</i> (MAG26-SR09-2019)	2	1	5.22e-100	4.0e-121	97%	53.30	TPA: hypothetical protein	357	<i>Siphoviridae</i>
		2	7.25e-118	0.0	95%	61.83	TPA: chaperonin	542	<i>Herelleviridae</i>
		2	7.25e-118	0.0	95%	60.11	TPA: chaperonin	540	<i>Siphoviridae</i>

Legenda: (-) Sequência analisada não apresentou resultado ou apresentou resultado com métricas estatísticas abaixo dos valores estabelecidos para este trabalho.

3.3. Diversidade de famílias

Nos quatro genomas das cianobactérias, foram identificadas 19 prováveis regiões de profago pelo Phaster, onde destas quatro foram identificadas pelo

BLASTP como sendo pertencentes a família *Siphoviridae* com o score mais alto; três foram identificadas como sendo pertencentes a família *Myoviridae* com o score mais alto; uma região foi identificada como sendo pertencente à família *Podoviridae* e uma região de profago obteve como primeira identificação a família *Inoviridae* e como segunda identificação a família *Myoviridae*. As 10 regiões de profagos restantes não foram classificadas por conta de valores ruins nas métricas estatísticas.

Com relação as MAGs, no ano de 2017 foram identificadas 6 prováveis regiões de profago pelo Phaster nas MAGs, onde duas foram identificadas pelo BLASTP como sendo pertencentes a família *Siphoviridae* com o score mais alto, 1 foi identificada como sendo pertencente à família *Myoviridae* com o score mais alto, uma região de profago teve como primeira identificação um vírus de DNA não classificado e como segunda identificação a família *Myoviridae* e duas não foram classificadas por conta de valores ruins nas métricas estatísticas.

No ano de 2018, foram identificadas 11 prováveis regiões de profago pelo Phaster nas MAGs, 2 foram identificadas pelo BLASTP como sendo pertencentes a família *Siphoviridae* com o score mais alto; duas foram identificadas como sendo pertencentes a família *Myoviridae* com o score mais alto; uma região de profago teve como primeira identificação um vírus bacteriano não classificado e como segunda identificação a família *Myoviridae* e seis não foram classificados pois não obtiveram resultados para famílias de fagos no BLASTP.

No ano de 2019, foram identificadas 17 prováveis regiões de profago pelo Phaster nas MAGs, destas 4 foram identificadas pelo BLASTP como sendo pertencentes a família *Myoviridae* com o score mais alto, 6 foram identificadas como sendo pertencentes a família *Siphoviridae* com o score mais alto, 1 região de profago teve como primeira identificação a família *Phycodnaviridae* e como segunda identificação a família *Myoviridae*, duas regiões de profago tiveram como primeira identificação um vírus bacteriano e um vírus de DNA não classificado e como segunda identificação a família *Myoviridae*, uma região de profago teve como primeira identificação a família *Herelleviridae* e como segunda identificação a família *Siphoviridae* e três não foram classificadas por conta de valores ruins nas métricas estatísticas.

No total, foram analisadas 23 MAGs no Phaster. Em três destas MAGs não foram encontradas prováveis regiões de profago no Phaster, as MAGs são: *Phormidesmiaceae* (MAG14-SR04-2018), *Cyanobiaceae* (MAG48-SR05-2018) e *Phormidesmiaceae* (MAG33-SR01-2019). Nas outras 20 MAGs, foram encontradas 34 prováveis regiões de profagos. Destas, 10 foram identificadas no BLASTP como sendo pertencentes a família *Siphoviridae* com o score mais alto, sete foram identificadas como sendo pertencentes a família *Myoviridae*, quatro regiões de profago tiveram como primeira identificação um vírus bacteriano e um vírus de DNA não classificado e como segunda identificação a família *Myoviridae*, uma região de profago teve como primeira identificação a família *Phycodnaviridae* e como segunda identificação a família *Myoviridae*, 1 região de profago teve como primeira identificação a família *Herelleviridae* e como segunda identificação a família *Siphoviridae* e 11 não foram classificadas por conta de valores ruins nas métricas estatísticas.

No total, contabilizando os genomas das cianobactérias e as MAGs, foram encontrados pelo Phaster 53 prováveis regiões de profagos. Destas, 14 foram identificadas no BLASTP como sendo pertencentes a família *Siphoviridae* com o score mais alto, 10 foram identificadas como sendo pertencentes a família *Myoviridae*, uma região foi identificada como sendo pertencente à família *Podoviridae*, quatro regiões de profago tiveram como primeira identificação um vírus bacteriano e um vírus de DNA não classificado e como segunda identificação a família *Myoviridae*, uma região de profago teve como primeira identificação a família *Phycodnaviridae* e como segunda identificação a família *Myoviridae*, uma região de profago obteve como primeira identificação a família *Inoviridae* e como segunda identificação a família *Myoviridae*, uma região de profago teve como primeira identificação a família *Herelleviridae* e como segunda identificação a família *Siphoviridae* e 21 não foram classificadas por conta de valores ruins nas métricas estatísticas.

Os resultados obtidos no BLASTP sobre a classificação das famílias dessas prováveis regiões de profagos indicaram para as três famílias de cianofagos conhecidas e documentadas: *Siphoviridae* (*cyanosiphovirus*), *Myoviridae* (*cyanomyovirus*) e *Podoviridae* (*cyanopodovirus*).

4. Discussão

Considerando o conjunto dos resultados obtidos a partir da análise dos genomas das cianobactérias isoladas e das MAGs, é possível sugerir que as cianobactérias que habitam as lagoas salino-alcálinas do Pantanal podem estar sendo infectadas por cianofagos e que a afiliação taxonômica desses fagos sugere que eles pertençam às famílias *Siphoviridae*, *Myoviridae* e *Podoviridae*.

Um estudo realizado por Peduzzi et. al. (2014), em uma coluna de água salino-alcálica localizada no Parque Nacional do Lago Nakuru, no Quênia, rastreou a infecção de cianofagos que provocavam a alta mortalidade da cianobactéria *Arthrospira fusiforme* que é o principal alimento da ave Flamingo. Como resultado, o estudo apresentou o maior número de fagos em ambiente aquático natural já relatado. O que indica responsabilidade desses cianofagos na alta mortalidade da *Arthrospira fusiforme*. Esse trabalho é um exemplo de como os cianofagos podem ser abundantes em ambientes salino-alcálicos.

Neste trabalho, a família que ocorreu com mais frequência, quando se considera apenas o score mais alto, foi a família *Siphoviridae* (14 regiões de profagos), seguida pela família *Myoviridae* (10 regiões de profagos) e por fim pela família *Podoviridae* com apenas uma região de profago. Se considerarmos as regiões de profago que apresentaram as famílias *Myoviridae* e *Siphoviridae* como segundo score mais alto, a família *Myoviridae* passa a frente com 16 regiões de profagos e a família *Siphoviridae* fica em segundo com 15 regiões de profagos.

Segundo Whitton e Potts (2007), os cianofagos mais abundantes e que tiveram os seus genomas isolados mais vezes são os da família *Myoviridae*. Eles foram os primeiros a serem descritos como vírus que infectam cianobactérias. O segundo maior grupo de cianofagos é o da família *Podoviridae*, os cyanopodovirus. E por fim, tem-se o grupo de cianofagos da família *Siphoviridae*, os cyanosiphovirus, que tiveram os seus genomas isolados menos vezes (HUANG, Sijun et al., 2012). No entanto, apesar de os cianofagos da família *Siphoviridae* terem tido os seus genomas isolados menos vezes, a família foi um dos resultados mais comuns neste trabalho.

E apesar de os cianofagos da família *Podoviridae* terem tido os seus genomas isolados e sequenciados mais vezes, a família apareceu apenas na identificação de uma região de profago. Em um estudo conduzido em piscinas de

cultura na França, Jacquet et. al. (2013) identificou um cyanopodovirus que infecta a cianobactéria *Arthrospira platensis* pela primeira vez. Neste estudo foi observado, que o cyanopodovirus identificado provocava a lise celular apenas desta cianobactéria em específico. Assim, como esse estudo a maior parte dos cyanopodovirus identificados apresentam alta especificidade. Diferentemente, por exemplo, dos cyanomyovirus que infectam uma quantidade muito mais ampla de hospedeiros (Jacquet et. al., 2013). Neste trabalho, a única região de profago identificada como sendo pertencente à família Podoviridae é do genoma da cianobactéria *Limnospira* sp. CENA 650.

Todos os resultados descritos neste trabalho apresentaram valores razoáveis de e-value e alguns resultados apresentaram valores ótimos com e-values bastante próximos de zero ou até mesmo igual a zero. Tanto nos resultados das sequências genômicas das proteínas obtidas no Phaster, quanto nos alinhamentos obtidos no BLASTP. Segundo Stover e Cavalvanti (2017), quanto mais próximo o e-value estiver do zero, maior a probabilidade da sequência que está sendo analisada apresentar ancestralidade comum com o resultado obtido no alinhamento do BLASTP. Os valores de query cover da maioria dos resultados mostrou que a maior parte da extensão das sequências analisadas se encontrava alinhadas as sequências do banco de dados. A maior parte dos resultados apresentou valor de query cover superior a 95%.

Das três métricas analisadas, os valores da porcentagem de identidade foram os que apresentaram os valores mais inconclusivos. A maioria dos genomas de cinaobactérias e MAGs analisadas apresentaram valores de porcentagem de identidade entre 50% e 60%, poucos apresentaram valores de identidade superiores a 60%. Esses valores baixos de identidade podem indicar limitações nos bancos de dados da plataforma, já que se trata de um ambiente muito diverso e pouco explorado. Segundou You Zhou et. al. (2011), assim como muitas plataformas de anotação de sequências genômicas baseadas em bancos de dados, o Phaster também possui capacidade bastante limitada para descobrir novos fagos. A identificação de muitas proteínas com função desconhecida pode indicar que se trata de um fago novo.

No caso das MAGs *Phormidesmiaceae* (MAG11-SR01-2018), *Phormidesmiaceae* (MAG49-SR05-2018), *Nodularia* (MAG18-SR08-2018),

Raphidiopsis brookii (MAG45-SR08-2018) e *Snowella* (MAG56-SR08-2018), que após terem sido analisadas pelo Phaster tiveram regiões de profagos identificadas, no entanto, ao serem analisadas no BLASTP esta plataforma não encontrou nenhuma família de fago (nem mesmo com métricas estatísticas ruins), o problema também pode estar na limitação dos bancos de dados ou no algoritmo que identificou o fago. Segundo You Zhou et. al. (2011), ao analisar a sequência de um genoma o algoritmo DBSCAN do Phaster atribui uma densidade uniforme aos “hits” que apontam para as similaridades com fagos ao longo de toda a sequência, o que pode ser enganoso. Desse modo, o algoritmo pode cometer erros caso a distribuição dos genes seja muito irregular.

5. Conclusão

A plataforma Phaster nos fornece resultados interessantes sobre as sequências genômicas de profagos que podem estar integradas aos genomas de cianobactérias e as MAGs. E o BLAST NCBI não é a plataforma de bioinformática mais utilizada do mundo por acaso. As duas ferramentas são muito úteis para realizar uma análise dos genomas de organismos com os quais se pretende trabalhar. Para investigar com mais profundidade a classificação desses cianofagos, seria interessante em posse desses dados realizar a montagem de árvores filogenéticas. Também seria interessante realizar experimentos laboratoriais para tentar induzir a cianobactéria hospedeira a liberar cianofagos viáveis.

REFERÊNCIAS

- ARNDT, David et al. PHASTER: a better, faster version of the PHAST phage search tool. **Nucleic acids research**, v. 44, n. W1, p. W16-W21, 2016. Disponível em: <https://academic.oup.com/nar/article/44/W1/W16/2499364?login=true>
- ARNDT, David et al. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1560-1567, 2019. Disponível em: <https://academic.oup.com/bib/article/20/4/1560/4222653?login=true>
- BORATYN, Grzegorz M. et al. BLAST: a more efficient report with usability improvements. **Nucleic acids research**, v. 41, n. W1, p. W29-W33, 2013. Disponível em: <https://academic.oup.com/nar/article/41/W1/W29/1091045?login=true>
- CASJENS, Sherwood. Prophages and bacterial genomics: what have we learned so far? **Molecular microbiology**, v. 49, n. 2, p. 277-300, 2003. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2958.2003.03580.x>
- DENNEHY, John J.; ABEDON, Stephen T. Adsorption: Phage acquisition of bacteria. Bacteriophages: biology, technology, therapy, p. 93-117, 2021. Disponível em: https://link.springer.com/referenceworkentry/10.1007/978-3-319-41986-2_2
- HATFULL, Graham F.; HENDRIX, Roger W. Bacteriophages and their genomes. **Current opinion in virology**, v. 1, n. 4, p. 298-303, 2011. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1879625711000514>
- HUANG, Sijun et al. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage–host genetic exchanges. **Environmental Microbiology**, v. 14, n. 2, p. 540-558, 2012. Disponível em: <https://doi.org/10.1111/j.1462-2920.2011.02667.x>
- HURST, Christon J. et al. Defining the ecology of viruses. Microbial botanical host systems. **Stud Viral Ecol**, v. 1, p. 3-40, 2011. Disponível em: <https://onlinelibrary-wiley.ez67.periodicos.capes.gov.br/doi/book/10.1002/9781118025666>
- HURWITZ, Bonnie L. et al. Phage Hunters: computational strategies for finding phages in large-scale ‘omics datasets. **Virus research**, v. 244, p. 110-115, 2018. University of Arizona, Tucson, United States. Disponível em: <https://doi.org/10.1016/j.virusres.2017.10.019>
- HUSON, Daniel H.; MITRA, Suparna. Introduction to the analysis of environmental sequences: metagenomics with MEGAN. **Evolutionary Genomics**. Humana Press, 2012. p. 415-429. Disponível em: https://link.springer.com/protocol/10.1007/978-1-61779-585-5_17
- JACQUET, Stéphan et al. First description of a cyanophage infecting the cyanobacterium *Arthrospira platensis* (Spirulina). **Journal of applied phycology**, v. 25, n. 1, p. 195-203, 2013. Disponível em: <https://link.springer.com/article/10.1007/s10811-012-9853-x>

MANN, Nicholas H. The third age of phage. *PLoS biology*, v. 3, n. 5, p. e182, 2005. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030182>

PEDUZZI, Peter et al. The virus's tooth: cyanophages affect an African flamingo population in a bottom-up cascade. **The ISME journal**, v. 8, n. 6, p. 1346-1351, 2014. Disponível em: <https://www.nature.com/articles/ismej2013241>

PELLEGRINETTI, Thierry A. et al. Bacterial communities along environmental gradients in tropical soda lakes. **Microbial Ecology**, p. 1-12, 2022. Disponível em: <https://link.springer.com/article/10.1007/s00248-022-02086-6>

SAFFERMAN, Robert S.; MORRIS, Mary-Ellen. Growth characteristics of the blue-green algal virus LPP-1. **Journal of Bacteriology**, v. 88, n. 3, p. 771-775, 1964. Disponível em: <https://jb.asm.org/content/88/3/771>

SAFFERMAN, Robert S.; MORRIS, Mary-Ellen. Control of algae with viruses. **Journal-American Water Works Association**, v. 56, n. 9, p. 1217-1224, 1964. Disponível em: <https://awwa.onlinelibrary.wiley.com/doi/epdf/10.1002/j.1551-8833.1964.tb01324.x>

SAFFERMAN, R. S. et al. Classification and nomenclature of viruses of cyanobacteria. **Intervirolgy**, v. 19, n. 2, p. 61-66, 1983. Disponível em: <https://www.karger.com/Article/Abstract/149339>

SHERMAN, Louis A.; BROWN, R. Malcolm. Cyanophages and viruses of eukaryotic algae. In: **Newly Characterized Protist and Invertebrate Viruses**. Springer, Boston, MA, 1978. p. 145-234. Disponível: https://link.springer.com/chapter/10.1007/978-1-4684-2724-0_3

SOROKIN, Dimitry Y. et al. Microbial diversity and biogeochemical cycling in soda lakes. **Extremophiles**, v. 18, n. 5, p. 791-809, 2014. Disponível em: <https://link.springer.com/article/10.1007/s00792-014-0670-9>

SRIVIDHYA, K. V. et al. Database and comparative identification of prophages. **Intelligent Control and Automation**. Springer, Berlin, Heidelberg, 2006. p. 863-868. Disponível em: https://link.springer.com/chapter/10.1007/978-3-540-37256-1_110

STOVER, Naomi A.; CAVALCANTI, Andre RO. Using NCBI BLAST. **Current Protocols Essential Laboratory Techniques**, v. 14, n. 1, p. 11.1. 1-11.1. 34, 2017. Disponível em: <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpet.8>

TINDALL, Brian J. Prokaryotic life in the alkaline, saline, athalassic environment. **Halophilic bacteria**, v. 1, p. 31-67, 1988.

WHITTON, Brian A.; POTTS, Malcolm (Ed.). The ecology of cyanobacteria: their diversity in time and space. **Springer Science & Business Media**, 2007. Disponível em: <https://link-springer-com.ez67.periodicos.capes.gov.br/book/10.1007/0-306-46855-7>

YOU ZHOU, Yongjie Liang et al. PHAST: A Fast Phage Search Tool. **Nucleic acids research**, v. 39, n. suppl_2, p. W347-W352, 2011. Disponível em: <https://doi.org/10.1093/nar/gkr485>