Universidade Federal da Paraíba

Centro de Ciências Exatas e da Natureza

Curso de Bacharelado em Ciências Biológicas

Análise de Coevolução entre *Mimivirus* e seus virófagos *Sputnik* e *Zamilon* a partir de suas Assinaturas Genômicas

Paulo Eduardo Toscano Soares

João Pessoa - PB

Novembro - 2016

Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Curso de Bacharelado em Ciências Biológicas

Análise de Coevolução entre *Mimivirus* e seus virófagos *Sputnik* e *Zamilon* a partir de suas Assinaturas Genômicas

Paulo Eduardo Toscano Soares

Prof. Dr. Sávio Torres de Farias
Orientador

Prof. Dr. Victor Manuel Serrano-Solís
Co-orientador

Monografia apresentada ao Curso de Ciências Biológicas (Trabalho Acadêmico de conclusão de Curso), como requisito parcial à obtenção do grau de Bacharel em Ciências Biológicas

João Pessoa - PB
Novembro de 2016

Catalogação na publicação
Universidade Federal da Paraíba
Biblioteca Setorial do CCEN
Maria Teresa Macau - CRB 15/176

S676a Soares, Paulo Eduardo Toscano.
Análise de coevolução entre *Mimivirus* e seus virófagos *Sputnik* e *Zamilon* a partir de suas assinaturas genômicas / Paulo Eduardo Toscano Soares.- João Pessoa, 2016.

51p. : il.-

Monografia (Bacharelado em Ciências Biológicas) - Universidade Federal da Paraíba.
Orientador: Profº Drº Sávio Torres de Farias.

1. Vírus. 2. *Acanthamoeba Polyphaga Mimivirus*. 3. Assinaturas genômicas. I. Título.

UFPB/BS-CCEN CDU: 578.89(043.2)

Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Curso de Bacharelado em Ciências Biológicas

Paulo Eduardo Toscano Soares

Análise de Coevolução entre *Mimivirus*, *Sputnik* e *Zamilon* a partir de suas assinaturas genômicas

Monografia apresentada ao Curso de Ciências Biológicas, como requisito parcial à obtenção do grau de Bacharel em Ciências Biológicas

Data: _____

Resultado: _____

Prof. Dr. Sávio Torres de Farias
Centro de Ciências Exatas e da Natureza
Departamento de Biologia Molecular

Prof. Dr. Victor Manuel Serrano-Solís
Centro de Ciências Exatas e da Natureza
Departamento de Biologia Molecular

Prof.ª Dr. ª Thaís Gaudêncio do Rêgo
Centro de Informática
Departamento de Informática

Dedico este trabalho aos meus pais.

Agradecimentos

À minha mãe, Rosangela, e ao meu pai, Mozart, que foram os meus maiores apoiadores e incentivadores nesta jornada.

À minha namorada Mayanna, por todo carinho, apoio e incentivo ao longo do curso e da elaboração deste trabalho.

A todos professores que fizeram parte da minha formação, especialmente àqueles que me inseriram na iniciação científica, a professora Drª. Patrícia Mirella da Silva Scardua, e na docência, os professores Dr. Rivete Silva de Lima e Dr. Edson Luiz Folador pelas experiências com tutoria e monitoria, respectivamente.

A todos do Laboratório de Genética Evolutiva Paulo Leminsk, por propiciarem reuniões divertidas e com discussões produtivas que contribuíram com a minha formação.

Ao meu amigo e companheiro de curso, Victor Montenegro pelo apoio, sugestões de artigos e troca de idéias que enriqueceram este trabalho.

Ao professor Dr. Victor Serrano-Solís, por ter me orientado e me guiado em minha iniciação científica.

Ao professor Dr. Sávio Torres e à professora Dr.ª Thaís Gaudêncio, por terem me recebido no laboratório e pelas sugestões que enriqueceram este trabalho.

A todos membros da banca examinadora, por terem aceito o convite de participação da banca e por suas contribuições com este trabalho.

“Quem quer, dá um jeito. Quem não quer, dá uma desculpa.”
- Provérbio árabe

Resumo

Acanthamoeba polyphaga Mimivirus (APMV) foi descoberto em 2004, negligenciado por décadas como um parasito obrigatório do seu hospedeiro, a ameba *Acanthamoeba polyphaga*. APMV é um virus gigante, maior que pequenas bactérias, com um genoma de 1,2 megabases (Mb) de comprimento, contendo seis tRNAs, uma característica exepcional na então conhecida virosfera. A posterior descoberta de um vírus infectando APMV levou a uma mudança de paradigma na microbiologia. Estes vírus, chamados de virófagos, se replicam apenas na presença do seu vírus auxiliador. Com a finalidade de procurar padrões coevolutivos nos genomas de APMV e seus dois virófagos conhecidos, *Sputnik* e *Zamilon*, foram realizados testes paramétricos genômicos tais como a análise de Frequência de Dinucleotídeos (FDN); e três modalidades de uso de códons (UC): Uso Relativo de Códons Sinônimos (do inglês, *Relative Synonnymous Codon Usage* - RSCU), Frequência Relativa de Códons (FRC) e Frequência Relativa de Uso de Códons Sinônimos (FRUCS). Análises de correlação foram realizadas a fim de consolidar os resultados. Também foram analisados os genomas da mitocôndria *A. polyphaga*, da bactéria parasítica intracelular de *A. polyphaga* "*Candidatus* Babela massliensis" e da enterobactéria *Escherichia coli* (grupo externo) para fins de comparação. A análise de FDN permite a detecção de um padrão ou Assinatura Genômica (AG) adequada para comparações evolutivas enquanto que as análises de UC aprofundam estas comparações. As análises de FDN revelaram padrões comuns entre as AGs de APMV e seus virófagos, enquanto que as análises de UC mostraram padrões no uso de códons. Estes resultados apontam evidências de um histórico coevolutivo compartilhado por APMV e seus virófagos, além de indicar possíveis relações desse histórico com a ameba.

Palavras-chave: Vírus, Virófagos, Coevolução, Assinaturas Genômicas

Abstract

Acanthamoeba polyphaga Mimivirus (APMV) was discovered in 2004, neglected by decades as an obligatory parasite of its host, the amoeba *Acanthamoeba polyphaga*. APMV is a giant virus, greater than small bacteria, with a genome of 1.2 megabases containing six tRNAs, an exceptional feature for the known virosphere. The discovery of a virus infecting APMV led to a change of paradigm in microbiology. These viruses, named virophages, replicate only in the presence of its helper virus. In order to find coevolutive genomic patterns between APMV and its two known virophages, *Sputnik* and *Zamilon*, parametric tests were performed. Tests such as Dinucleotide Frequency (FDN) analysis; and three modalities of Codon Usage (UC) analyses: Relative Synonymous Codon Usage (RSCU), Relative Frequency of Codons (FRC) and Relative Frequency of Synonymous Codons (FRUCS). Correlation analyses were performed in order to consolidate results. Genomes of *A. polyphaga's* mitochondrion, *A. polyphaga*'s intracellular parasitic bacteria "*Candidatus* Babela massiliensis" and the enterobacteria *Escherichia coli* were also analyzed for comparison. The DNF analysis allows the detection of a pattern or Genomic Signature (GS), adequate for evolutive comparisons while the CU analyses deepens them. The DNF analysis revealed common patterns between the GS of APMV and its virophages while the CU analyses showed codon usage patterns. These results show evidences of a coevolutive history shared by APMV and its virophages, also indicating possible relationships with the amoeba.

Keywords: Virus, Virophages, Coevolution, Genomic Signatures

Lista de Figuras

Lista de Tabelas e Quadros

Listas de abreviaturas e siglas

AG: Assinatura Genômica

APMV*:* *Acanthamoeba polyphaga Mimivirus*

CDS: Sequências Codificantes (do inglês, *Coding Sequences)*

CEO: Organismo Codificante de Capsídeo (do inglês, *Capsid Encoding Organism*)

DCA: Dinâmica de Corrida Armamentícia

DNA: Ácido Desoxirribonucleico (em inglês, *Desoxyribonucleic Acid*)

DSF: Dinâmica de Seleção Flutuante

FDN: Frequência de dinucleotídeos

FR: Frequência relativa

FRC: Frequência Relativa de Códons

FRUCS: Frequência Relativa de Uso de Códons Sinônimos

GS: Assinatura Genômica (do inglês, *Genomic Signature*)

Mb: Megabases

MIMIVIRE: Elemento de Resistência a Virófagos do Mimivirus (do inglês, *Mimivirus Virophage Resistance Element)*

mRNA: RNA mensageiro

mtDNA: DNA mitocondrial

NCLDV: Vírus Nucleotcitoplasmáticos de DNA Grande (do inglês, *Nucleocytoplasmic Large DNA viruses)*

Pb: Pares de base

REO: Organismo Codificante de Ribossomo (do inglês, *Ribossome Encoding Organism*)

RNA: Ácido Ribonucleico (do inglês, *Ribonucleic Acid*)

RSCU: Uso Relativo de Códons Sinônimos (em inglês, *Relative Synonnymous Codon Usage*)

UC: Uso de Códons

Sumário

1) Introdução

As amebas são protozoários considerados fagócitos naturais (MOLINER et al., 2010). La Scola et al. (2003) publicaram o primeiro relato de um vírus gigante infectando a ameba *Acanthamoeba polyphaga*. Este vírus, que foi chamado de *Acanthamoeba polyphaga Mimivirus* (APMV), possui um capsídeo iscosaédrico com tamanho similar ao de uma pequena bactéria (~0,4 um) e um genoma de aproximadamente 1,2 megabase (RAOULT et al., 2004) – os maiores já vistos até então. Estas descobertas surpreendentes causaram o questionamento do então conceito de vírus (PEARSON, 2008; RAOULT; FORTERRE, 2008; FORTERRE, 2010).

Sabe-se que os vírus são membros abundantes presentes em todas comunidades microbiais e, no ambiente marinho, afetam a estrutura populacional e a ciclagem de nutrientes ao infectar e lisar produtores primários (YAU et al., 2011). La Scola et al. (2008) identificaram o primeiro caso de um novo tipo de vírus, os virófagos, que para se reproduzir dependem da coinfecção com outro vírus (o vírus "auxiliador") causando notável redução no *fitness* do vírus auxiliador. No caso, descobriu-se que um vírus (nomeado *Sputnik*) causava a formação de capsídeos defeituosos em *Mamavirus* (pertencentes à linhagem A de APMV) quando ambos estavam co-infectando a ameba *A. polyphaga* (LA ESCOLA et al., 2003). Posteriormente, foi descoberto outro virófago chamado *Zamilon* que causa sintomas similares ao *Sputnik* e mostrou-se linhagem-específico de linhagens B e C de APMV (GAIA et al., 2014).

Modelos de dinâmica ecológica entre amebas, vírus gigantes e virófagos foram propostos a fim de explicar o impacto dos virófagos na regulação populacional de amebas e vírus gigantes (WODARZ, 2013). Também foram descritos modelos onde virófagos atuam controlando dinâmicas algas-vírus (YAU et al., 2011). Assim, espera-se que por trás destas interações haja um histórico coevolutivo entre as linhagens de APMV e seus virófagos, e também da ameba *A. polyphaga* com estes vírus, resistentes à digestão após a fagocitose. Recentemente, propuseram a existência de um sistema de defesa gênico, chamado de Elemento de Resistência a Virófago do Mimivirus (do inglês, *Mimivirus Virophage Resistance Element* - MIMIVIRE), que confere resistência ao virófago *Zamilon* nas linhagens A de APMV. Este sistema

depende do reconhecimento de sequências curtas repetidas (15 nucleotídeos) por proteínas para conferir resistência ao virófago (LEVASSEUR et al. 2016).

Os genomas apresentam padrões na sua sequência nucleotídica, que são chamados de assinaturas genômicas (AG) (KARLIN; BURGE, 1995). No caso de interações vírus-hospedeiro, acredita-se que estes padrões podem ser adaptados para escapar de mecanismos de defesa do hospedeiro baseados no reconhecimento de padrões na sequência (GREENBAUM et al., 2008; LOBO et al., 2009), aumentar a eficiência/velocidade no processo de tradução (PLOTKIN; KUDLA, 2011; SHARP; TUOHY; MOSURSKI, 1986) e evitar a degradação do RNA mensageiro (mRNA) (DUAN, ANTEZANA, 2003; MCDOWALL; LIN-CHAO; COHEN, 1994).

Neste trabalho, procurou-se descobrir se a análise comparativa das AGs de APMV e de seus virófagos revela parte de seu histórico coevolutivo. Para isso, foram realizadas comparação de AG ao longo do genoma inteiro (obtidas através das frequências de oligonucleotídeos) e ao longo apenas das CDS dos genomas (obtidas através do uso de códons). Para fins de comparação, também foram analisados os genomas da mitocôndria de *A. polyphaga*, da bactéria parasítica de *A. polyphaga* "*Candidatus* Babela massiliensis" (PAGNIER et al., 2015) e da enterobacteria *Escherichia coli*. Além disso, presume-se que uma elevada semelhança no UC entre hospedeiro e vírus indica um maior grau de adaptação do vírus ao seu hospedeiro já que eles compartilham os mesmos recursos disponíveis no ambiente celular (ex.: ribossomos disponíveis, aminoácidos livres, tRNAs cognatos, etc).

2) Fundamentação Teórica

2.1) O genoma

O termo genoma refere-se à série completa de informações do Ácido Desoxirribonucleico (do inglês, *Desoxyribonucleic Acid – DNA*) de um organismo e ele contém informações para a tradução de todas as proteínas e moléculas de Ácido Ribonucleico (do inglês, *Ribonucleic Acid – RNA*) que o organismo irá produzir durante a sua existência (ALBERTS, 2009, p. 200). Todas as células armazenam suas informações hereditárias na forma de moléculas de DNA fita dupla, compostas por quatro monômeros (Adenina, Timina, Citosina, Guanina ou, abreviadamente, A, T,C e G, respectivamente) que estão ordenados em uma longa sequência linear que armazena toda a sua informação genética (ALBERTS, 2009, p. 2).

A informação genética está codificada nos genes. Os genes são segmentos do DNA que especificam uma cadeia polipeptídica; isto inclui regiões que antecedem e sucedem a região codificante, assim como regiões que intervém (íntrons) entre segmentos codificantes individuais (éxons) (LEWIN et al., 2008). A capacidade de formar proteínas se dá através dos processos de transcrição e tradução. Quando os genes são transcritos, eles formam os RNAs mensageiros cuja informação está armazenada na sequência de trincas de nucleotídeos (códons) que correspondem a um único aminoácido da proteína correspondente (ALBERTS, 2009, p. 6). Os RNAs mensageiros serão lidos por RNAs transportadores em uma complexa associação com os ribossomos, formando as cadeias proteicas (ALBERTS, 2009, p. 6-7). Dos 64 códons possíveis, apenas 20 aminoácidos são necessários de forma que vários códons correspondem ao mesmo aminoácido (ALBERTS, 2009, p.6). O código genético padrão para tradução de proteínas foi preservado na maioria dos organismos, vindo a sofrer mudanças nas funções dos códons em alguns genomas nucleares de eucariotos, procariotos e em organelas (mitocôndria e cloroplasto) (KUMAR; KUMARI; SHARMA, 2016).

Por não possuírem metabolismo, nem estrutura celular comum aos seres vivos, os vírus costumam não ser considerados como vivos. Raoult e Forterre (2008)

propuseram um sistema de classificação bipartido entre os organismos com base nas suas capacidades de codificar capsídeos e ribossomos a fim de enquadrar os vírus na definição de vida. Nesta proposta, os organismos seriam encaixados em organismos codificantes de capsídeos (do inglês, *capsid-encoding organisms* – CEO) ou em organismos codificantes de ribossomos (do inglês, *ribossome-encoding organisms* – REO). Assim, os vírus seriam considerados organismos vivos, enquadrados no grupo dos CEOs. Como previamente mecionado, todos os REOs (seres celulares) armazenam seu genoma exclusivamente em DNA dupla fita, enquanto que os CEOs (vírus) possuem também outras formas de armazenar seu genoma: RNA fita simples – podendo ser classificada em positiva ou negativa, DNA fita simples e DNA dupla fita a partir de RNA fita simples positiva (nos retrovírus).

Os genomas de CEOs e REOs estão em constante mudança devido à eventual ocorrência de mutações na composição nucleotídica do mesmo. Estas mutações nos genomas são responsáveis pelo processo de evolução biológica, responsável pela enorme variedade de CEOs e REOs.

2.2) Evolução

O conceito de evolução (biológica) refere-se ao processo pelo qual os organismos sofrem mudanças na forma e no comportamento ao longo das gerações (RIDLEY, 2006, p. 28). Outra definição diz que o termo evolução pertence a qualquer mudança no *pool* genético de uma população (RICKLEFS, 2010, p. 103). Estas mudanças surgem geralmente por processos estocásticos (mutações aleatórias) ou por processos determinísticos (pressões seletivas) e podem acarretar no surgimento ou perda de características (morfológicas, fisiológicas, comportamentais, etc).

As mutações resultam de qualquer mudança na sequência dos nucleotídeos que formam um gene, ou nas regiões do DNA, que controlam a expressão de um gene (RICKLEFS, 2010, p. 103). Mutações em genes requeridos para processos celulares fundamentais são frequentemente letais (ALBERTS, 2009, p. 556-557), mas ocasionalmente produzem novas características (*fenótipos*) que são mais bem ajustados ao seu ambiente local (RICKLEFS, 2010, p. 103, grifo nosso). Os indivíduos

que possuem características que os permitem sobreviver e reproduzir são ditos adaptados (RIDDLEY, 2006, p. 29).

Os indivíduos de uma população estão constantemente sujeitos a diversas pressões ambientais que acabam por selecionar indivíduos com maior aptidão através do processo de seleção natural. Seleção natural significa que alguns indivíduos da população tendem a contribuir com uma descendência maior para a próxima geração do que outros (RIDDLEY, 2006, p. 30). Estas pressões seletivas também podem ser causadas pelas interações entre populações de diferentes organismos (ERLICH; RAVEN, 1964; JANZEN, 1980) como, por exemplo, observado entre hospedeiros, vírus e virófagos.

2.3) Amebas, Mimivirus e Virófagos

As amebas de vida livre são protozoários, organismos unicelulares eucarióticos, que fagocitam microrganismos e partículas grandes (> 0,5 µm) do meio extracelular, porém não fagocitam organismos pequenos capazes de passar por filtros de 0,2 µm como a bactéria *Minibacterium massiliensis* (LA SCOLA et al., 2003; MOLINER; FOURNIER; RAOULT, 2010). La Scola et al. (2003) mostraram a ameba *Acanthamoeba polyphaga* fagocitando um vírus de tamanho gigante, até então o maior já visto, apresentando uma partícula viral icosaédrica de 0,4 µm de diâmetro revestido por fibrilas. Este vírus foi nomeado *Mimivirus* por “imitar” (mimetizar) uma bactéria em coloração gram.

Além do tamanho gigante do capsídeo, os *Mimivirus* também mostraram um genoma gigante, de aproximadamente 1,2 megabases, com mais de 1.200 Quadros de Leitura Abertos (do inglês, *Open Reading Frame* - ORF) previstos, dos quais 911 mostraram-se possíveis genes codificantes de proteínas (RAOULT et al., 2004). Legendre et al. (2011) fizeram outro sequenciamento do genoma de APMV utlizando técnicas de transcriptômica e genômica e descobriram 75 novos genes dos quais 26 eram de RNA não-codificante desconhecidos, chegando a compor um total de 1018 genes. Dentro desses genes novos, encontraram um gene codificando para uma subunidade II de RNA polimerase.

Observou-se que, durante a infecção em *Acanthamoeba polyphaga*, APMV é capaz de se reproduzir em estruturas citoplasmáticas denominadas “fábricas virais” (LA SCOLA et al., 2003), onde parece ocorrer a maioria (se é que não toda) da transcrição dos genes, com pouca ou nenhuma participação do aparato de transcrição do hospedeiro localizado no núcleo celular (LEGENDRE et al., 2011). Com base na filogenia usando apenas genes que codificam proteínas virais, o APMV foi incluído nas famílias virais dos grandes vírus nucleocitoplasmáticos (do inglês, *Nucleocytoplasmic Large DNA viruses* – NCLDV) (LA SCOLA, 2003). Os APMV hoje estão inseridos entre os NCLDV na família taxonômica *Mimiviridae,* composta por mais de 40 vírus similares a APMV*,* classificados em linhagens A, B ou C de acordo com sequência do seu gene *pol B* (Gaia et al., 2014).

Posteriormente, descobriu-se a existência de outros vírus em *A. polyphaga* que só se reproduziam na presença de APMV, os virófagos *Sputnik* (LA SCOLA et al., 2008) e *Zamilon* (GAIA et al., 2014). O termo “virófago” foi adotado porque além destes vírus dependerem da co-infecção com APMV para se reproduzir, eles causam deformações nos capsídeos de APMV e reduzem o seu *fitness* como evidenciado pela diminuição na mortalidade das amebas (LA SCOLA et al, 2009; GAIA et al., 2014). Estes virófagos se reproduzem utilizando a fábrica viral de APMV (LA SCOLA et al, 2008; GAIA et al., 2014). Os virófagos apresentam capsídeos de aproximadamente 60 nm, genomas de cerca de 18 kilobases (com 20 ORFs cada). Além disso, foi proposto que *Sputnik* é fagocitado pelas amebas junto com *Mimivirus* por estarem aderidos às suas fibrilas (LA SCOLA et al, 2008; GAIA et al., 2014).

A co-ocorrência entre seres vivos pode acarretar em fenômenos de coevolução. Um exemplo disso é o surgimento de linhagens de CEOs com maior virulência, resultantes da seleção natural por hospedeiros mais resistentes. Este mesmo fenômeno deve então se aplicar às relações observadas entre *Acanthamoeba polyphaga*, APMV e os virófagos.

2.4) Coevolução vírus-hospedeiro e suas dinâmicas

O conceito de coevolução foi inicialmente proposto para descrever as influências que as interações entre populações de planta e de insetos têm sobre a evolução de cada uma (ERLICH; RAVEN, 1964). A coevolução difusa ocorre quando uma ou ambas populações acima são representadas por um conjunto de populações que geram uma pressão seletiva como um grupo (JANZEN, 1980).

Modelos foram propostos para tentar explicar as dinâmicas coevolutivas de resistência e infectividade em populações de vírus marinhos e seus hospedeiros (MARTINY et al., 2014) e em populações de vírus gigantes e virófagos (WODARZ, 2013). No artigo de Martiny et al. (2014) os modelos analisados foram os de evolução direcional (Dinâmica de “Corrida Armamentícia” - DCA) e de evolução não-direcional (Dinâmica de Seleção Flutuante - DSF). No caso do modelo de DCA a infectividade dos vírus aumenta a medida que os hospedeiros se tornam mais resistentes e vice-versa. Já no modelo de DSF, a resistência e a virulência são, em média, constantes ao longo do tempo devido a uma alternância temporal nas frequências dos genótipos resistentes/virulentos.

Estas dinâmicas entre vírus e seus hospedeiros podem explicar as alterações nas assinaturas genômicas dos mesmos, pois sabe-se que os vírus tendem a “imitar” a assinatura genômica dos seus hospedeiros (GREENBAUM et al., 2008; LOBO et al., 2009).

2.5) Assinaturas Genômicas

As assinaturas genômicas são padrões característicos dos genomas que permitem a distinção entre espécies filogeneticamente próximas ou distantes (KARLIN; BURGE, 1995). Os genomas podem ser identificados por suas assinaturas e as dissimilaridades entre as assinaturas são usadas para estimar a distância evolucionária entre as espécies (JERNIGAN; BARAN, 2002; KARLIN; BURGE, 1995).

Os padrões podem revelar tendências nas características do genoma através da frequência da ocorrência de motivos (do inglês, *motifs*) nucleotídicos. Um exemplo

disso é a baixa representação de tetranucleotídeos palindrômicos em bactérias com genes codificando enzimas de restrição que clivam estes tetranucleotídeos (ABE et al., 2003). Motivos polinucleotídicos podem também estar envolvidos em padrões de reconhecimento no DNA por proteínas (KARLIN; BURGE, 1995; LOBO, 2009; PLOTKIN; KUDLA, 2011). Dentro dos genes, esses padrões podem se encontrar na frequência de uso dos códons, refletindo as preferências no uso de códons.

Sabe-se que o uso de nucleotídeos e oligonucleotídeos (de tamanhos 2 a 6) em fragmentos genômicos curtos de DNA são similares ao longo do genoma inteiro (DESCHAVANNE et al., 2000). Logo, o uso de sequências e/ou oligonucleotídeos de tamanhos maiores podem ser necessários para discriminar espécies filogeneticamente próximas (DESCHAVANNE et al., 2000), visto que o grau de complexidade aumenta com o tamanho da sequência/oligonucleotídeo.

Uma das vantagens de se utilizar AGs ao invés de métodos tradicionais de filogenia é que os resultados não vão variar de acordo com o conjunto de proteínas escolhidas (CAMPBELL, MRÁZEK, KARLIN, 1999). Outra vantagem é que o uso de AGs permite a comparação com indivíduos sem um ancestral comum, tornando possível comparar CEOs com REOs ou mesmo com qualquer tipo de material genético, independente de sua origem (ie.: plasmídeos, mitocôndrias, cloroplastos, etc.).

3) Objetivos

3.1) Objetivos Gerais

- Comparar **padrões coevolutivos** nos genomas de *Acanthamoeba polyphaga Mimivirus*, *Sputnik* e *Zamilon* a partir das suas assinaturas genômicas.

3.2) Objetivos Específicos

- **Detectar** padrões de uso de nucleotídeos (di-, tri- e tetranucleotídeos), assim como padrões de uso de códons (RSCU, FRUCS, FRC), nos genomas selecionados.
- **Comparar** os padrões de uso de nucleotídeos; **comparar** os padrões de uso de códons.
- **Determinar correlações** entre os três padrões oligonucleotídicos (di-, tri- e tetranucleotídeos) dos genomas; **determinar correlações** entre o uso de códons (através das FRUCS) dos genomas.
- **Comparar** as correlações obtidas entre os padrões oligonucleotídicos (di-, tri- e tetranucleotídeos) dos genomas; **comparar** as correlações obtidas entre o uso de códons dos genomas;

4) Material e Métodos

Todos os genomas e suas respectivas CDS foram baixadas no GenBank (BENSON et al., 2013) em formato FASTA (detalhes situados na Tabela 1). Dentre as várias versões de genomas de APMV, a utilizada neste estudo foi a fornecida pelo estudo de Legendre et al. (2011).

Tabela 1. Genomas analisados.

Nome do organismo	Conteúdo GC	Tamanho (pb)	Quantidade de CDS	Código de Acesso do NCBI
A. polyphaga Mimivirus	27,96%	1.181.404	913	NC_014649.1
Sputnik virophage	27,04%	18.338	21	NC_011132.1
Zamilon virophage	29,67%	17.276	20	NC_022990.1
A. polyphaga (mtDNA)	28,95%	39.215	35	KP054475.2
Candidatus Babela massiliensis	27,38%	1.118.422	983	NC_023003.1
Escherichia coli	50,69%	5.585.611	5604	CP007136.1

Fonte: Elaborada pelo autor.

Além dos genomas virais, também foram analisados os genomas da mitocôndria (mtDNA) da ameba *Acanthamoeba polyphaga*, "*Candidatus* Babela massiliensis" e *E. coli*. As mitocôndrias possuem DNA próprio e, por serem organelas indispensáveis à célula, foram analisadas podendo servir de indicadoras de adaptações à célula. Já "*Candidatus* Babela massiliensis" é uma bactéria intracelular obrigatória que, além de ter características típicas de organismos parasitas obrigatórios (ie.: genoma reduzido e capacidades metabólicas limitadas), mostra adaptações comuns aos NCLDV como, por exemplo, codificam várias *ankyrin-repeats* implicadas nas interações vírus-hospedeiro (PAGNIER et al, 2015). Assim, a comparação dos resultados dos vírus com os da mitocôndria e da bactéria "*Candidatus* Babela massiliensis" se faz necessária para avaliar indiretamente o grau de adaptação à ameba, comum a todos estes indivíduos.

4.1) Análise de Frequências Relativas de Oligonucleotídeos

Para a análise de frequências de di-, tri- e tetranucleotídeos foram utilizadas as sequências contidas nos arquivos FASTA dos genomas. Este formato contém apenas uma das duas fitas do DNA, que é o alvo da análise visto que o uso das duas fitas implicaria em uma simetria nas frequências de nucleotídeos complementares, como esperado pela Regra de paridade de Chargaff. Logo, o uso de apenas uma das fitas reflete a distribuição assimétrica dos nucleotídeos em uma das fitas do genoma obtendo-se assim AGs do organismo.

Os genomas foram analisados através de um algoritmo (Apêndice A) desenvolvido pelo autor em linguagem Python 3.5 (VAN ROSSUM; DRAKE JR, 1995), que realiza a contagem de oligonucleotídeos de vários tamanhos (nucleotídeos, dinucleotídeos, trinucleotídeos e tetranucleotídeos) e então retorna suas respectivas frequências relativas encontradas no genoma analisado. O algoritmo funciona através de uma janela deslizante (tamanho = 1) sobre a sequência fornecida no arquivo FASTA que, a cada interação, conta separadamente as frequências mono-, di-, tri- e tetranucleotídicas cabíveis à atual posição da janela na sequência e, ao final, retorna as frequências relativas observadas (Figura 1). Este algoritmo leva em consideração os três quadros de leitura esgotando assim todas as possibilidades de ocorrência de oligonucleotídeos. Caracteres diferentes de A, T, C ou G (ex.: N, U ou X), vez ou outros presentes nos arquivos FASTA, referentes a nucleotídeos não identificados no sequenciamento dos genomas, também foram contados e calculados, porém os dados não foram inclusos nas análises e resultados deste trabalho. Como estes nucleotídeos anômalos são pouco presentes nas sequências, eles não afetam a leitura das AGs dos genomas, assim como na ocorrência de *gaps* (KARLIN, 1998).

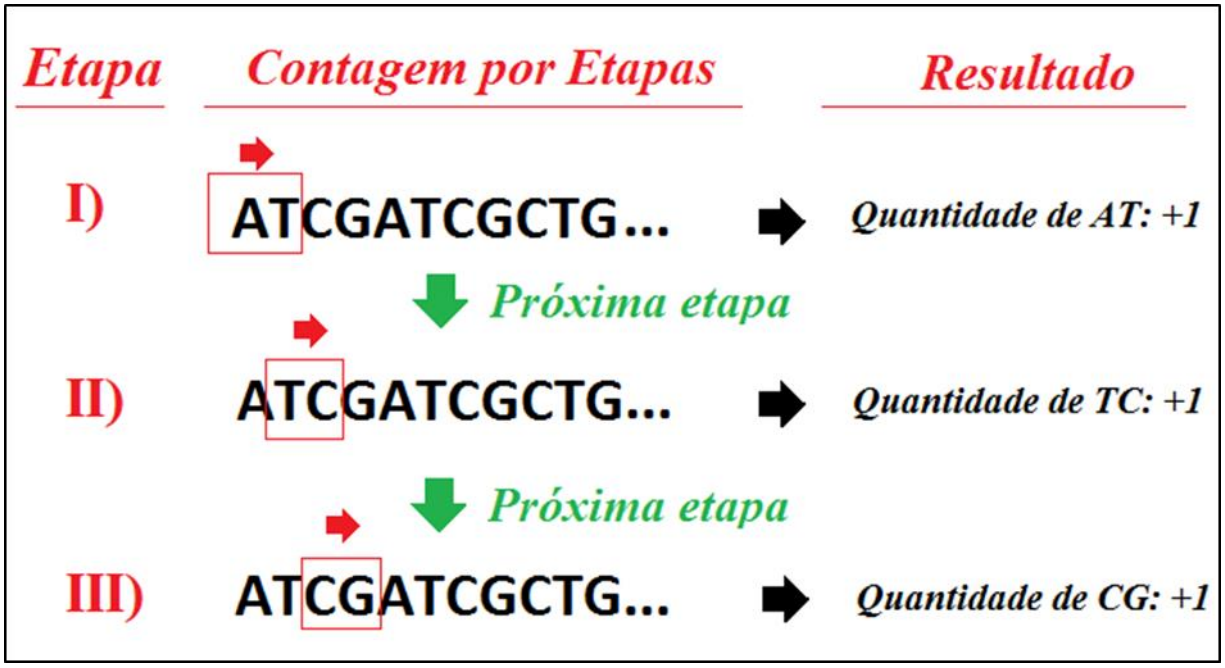
O conteúdo GC dos genomas foi obtido através do algoritmo desenvolvido pelo autor (Apêndice A). A sua obtenção consiste no cálculo da soma das frequências relativas dos nucleotídeos de G e C do genoma inteiro. Antigamente, acreditava-se haver uma correlação entre o tamanho do genoma e seu conteúdo GC, porém modelos matemáticos propostos não conseguiram achar esta correlação nos diferentes domínios da vida (MITCHELL, 2007).

Sabe-se que organismos mais proximamente relacionados mostram abundância dinucleotídica relativa muito mais parecida do que organismos

distantemente relacionados (KARLIN; BURGE, 1995; KARLIN, 1998). Já as frequências oligonucleotídicas podem ser usadas para distinguir genomas, pois elas variam significantemente entre os genomas (ABE et al., 2003).

Com as frequências de di-, tri- e tetranucleotídeos obtidas nestas análises foram realizadas, respectivamente, três análises de correlação de Pearson. As análises foram realizadas a fim de descobrir se há e qual o grau de correlação entre os padrões oligonucleotídicos (di-, tri- e tetranucleotídeos) destes três vírus e demais genomas analisados. Estas análises também serviram para avaliar a diferença no tamanho do oligonucleotídeo utilizado na comparação das assinaturas genômicas. Todas as análises de correlação foram realizadas utilizando o software PAST versão 3.13 (HAMMER; HARPER; RYAN, 2001). Os parâmetros utilizados no PAST para "estatística de correlação" foi "Linear r (Pearson)" e o "formato de tabela" foi "p(uncorr)". A figura 1 ilustra o funcionamento do algoritmo de obtenção das frequências oligonucleotídicas.

Figura 1. Modelo explicativo do algoritmo.



Fonte: Elaborada pelo autor.

4.2) Análises de Uso de Códon

Quanto às análises relativas ao uso de códons, foi utilizada a ferramenta online CAIcal (PUIGBO; BRAVO; GARCIA-VALLVE, 2008) para efetuar a contagem do total de códons nas CDS e o cálculo do RSCU (SHARP; LI, 1986; LOBO et al., 2009) de cada ORF dos genomas. A partir da contagem do total de códons das CDS dos genomas, efetuou-se o cálculo das frequências relativas de códons (FRC) e das frequências relativas de uso de códons sinônimos (FRUCS). As análises de FRC e FRUCS incluíram os códons de parada, que não são abrangidos pela análise de Uso de Códons Sinônimos Relativos (do inglês, *Relative Synonnymous Codon Usage* – RSCU) (SHARP; LI, 1986; PUIGBO; BRAVO; GARCIA-VALLVE, 2008).

O RSCU reflete a preferência de uso de códons que codificam para um aminoácido com degeneração maior que um (LOBO et al., 2009; SHARP; LI, 1986). Assim, os códons referentes a Metionina e Triptofano não são calculados por não possuírem códons sinônimos. Apesar de apresentarem códons sinônimos, os códons de parada também não são calculados no RSCU por apresentarem viéses relativos a presença de fatores de liberação (do inglês, *release factors*). O cálculo do RSCU foi realizado individualmente em todas as ORFs dos genomas exceto em *E. coli*, onde foram utilizadas 1000 CDS aleatórias (correspondendo a 17,8% do total de CDS) devido à limitações na quantidade de CDS analisadas simultaneamente pela ferramenta *online* CAICal. Estas análises permitindo verificar a ocorrência e distribuição de tendências no uso de códons sinônimos ao longo das CDS dos genomas.

Os resultados das análises de UC dos genomas foram então comparados entre si a fim de observar padrões comuns entre os genomas analisados. Estas análises são importantes pois acredita-se que o uso de códons está relacionado a um aumento na eficiência e velocidade de tradução (PLOTKIN; KUDLA, 2011; KUMAR; KUMARI; SHARMA, 2016) ou a uma correlação com o repertório (do inglês, *pool*) de tRNAs (SHARP, TUOHY, MOSURSKI, 1986; KUMAR, KUMARI, SHARMA, 2016; DUAN, ANTEZANA, 2003).

Também foram realizadas análises de Correlação de Pearson entre as frequências oligonucleotídicas dos genomas e nas FRUCS das CDS dos genomas. A

plotagem dos dados obtidos a partir das análises de FDN, FRUCS, FRC e RSCU; e a análise de correlação de FRUCS foram realizadas utilizando o *software* PAST versão 3.13 (HAMMER; HARPER; RYAN, 2001). Os parâmetros utilizados no PAST para "estatística de correlação" foi "Linear r (Pearson)" e para o "formato de tabela" foi "p(uncorr)".

5) Resultados

5.1) Frequência relativa de dinucleotídeos nos genomas

Os primeiros testes realizados foram com as frequências dinucleotídicas (FDN) utilizando o algoritmo desenvolvido pelo autor (Apêndice 1). Em geral, os CEOs analisados neste trabalho compartilham similaridades entre as suas AGs. A Figura 2 mostra a comparação entre as frequências dinucleotídicas dos genomas analisados.

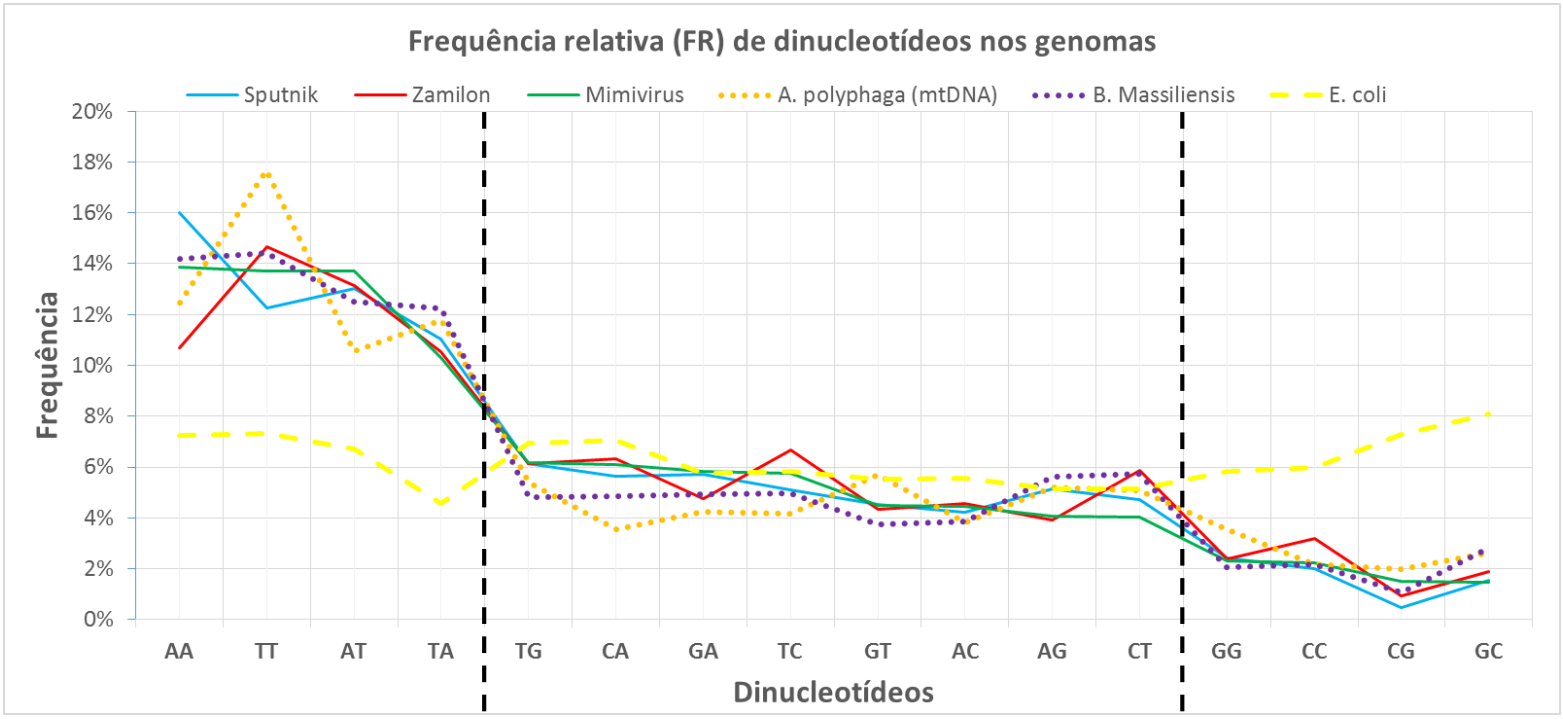


Figura 2. Padrão de distribuição de dinucleotídeos com base em suas frequências relativas. Os dinucleotídeos estão dispostos em ordem decrescente de frequência relativa, tendo como base o padrão de Mimivirus. As linhas tracejadas (cor preta) na

vertical dividem o gráfico em três regiões contendo os dinucleotídeos com frequências elevadas (esquerda), frequências intermediárias (meio) e frequências baixas (direita) observadas nos genomas intra-amebais.

Fonte: Elaborada pelo autor.

Em geral, os dinucleotídeos AA, TT, AT e TA mostraram maior frequência relativa (FR > 10%). Os dinucleotídeos AC, AG, CA, CT, GA, GT, TC, TG mostraram frequências relativas intermediárias (FR = 3,5 ~ 8%). Já GG, CC, CG e GC foram os dinucleotídeos com as menores frequências relativas (FR < 4%) observadas nos genomas virais, mitocondrial e no da bactéria "*Candidatus* B. massiliensis".

Mimivirus mostrou estabilidade nas frequências dos dinucleotídeos AT, TT e AT (FR = 13,8 ~ 13,7%), seguida por uma acentuada redução em TA (FR = 10,2%) e TG (FR = ~6%). GC se mostrou o dinucleotídeo mais evitado por *Mimivirus* (FR = 1,5%), seguido por CG (FR = 1,4%).

Sputnik teve como dinucleotídeos com maiores frequências AA, AT e TT (FR = 16%; 13% e 12,2%, respectivamente).

Zamilon apresentou maior preferência pelos dinucleotídeos TT, AT e AA (FR = 14,6%; 13,1% e 10,6%, respectivamente). Sputnik e Zamilon mostraram menor preferência pelo dinucleotídeo CG (FR = 0,4% e 0,9%, respectivamente), seguido por GC (FR = 1,5% e 1,8%, respectivamente).

A mitocôndria de *A. polyphaga* mostrou maior preferência por TT (FR > 17%) e teve CG como o dinucleotídeo menos utilizado (FR = ~2%).

A bactéria "*Candidatus* B. massiliensis" mostrou maior preferência por TT, AA e AT (FR = 14,4%; 14,2% e 12,5%, respectivamente). As menores FR observadas foram as de CG (1%) e a de GG (2%).

O genoma de *E. coli* (grupo externo) mostrou baixo uso do dinucleotídeo TA (FR < 5%) e uma acentuada preferência por GC (FR > 8%), ao contrário dos outros genomas. Já os demais dinucleotídeos apresentaram estabilidade nas suas frequências relativas (FR = 5~8%).

5.2) Correlação de Pearson entre padrões de di-, tri e tetranucleotídeos:

Após o cálculo das FDNs, foram realizados testes de Correlações de Pearson entre cada uma das frequências oligonucleotídicas obtidas (di-, tri e tetranucleotídeos) utilizando o software PAST 3.13 (HAMMER; HARPER; RYAN, 2001). O uso de oligonucleotídeos maiores mostrou progressiva redução nas correlações significativas observadas e aumento na significância estatística das mesmas, como observado na Tabela 2.

A)

Correlação entre frequências dinucleotídicas	*Candidatus* Babela massiliensis	Escherichia coli	Mimivirus	Mitocondria Apolyphaga	Sputnik	Zamilon
Candidatus Babela massiliensis		8,13E-01	8,88E-10	1,36E-08	3,25E-10	1,25E-08
Escherichia coli	0,064448		6,20E-01	7,51E-01	7,87E-01	7,98E-01
Mimivirus	0,96778	0,13437		6,73E-07	1,96E-11	1,00E-09
Mitocondria Apolyphaga	0,95211	0,086322	0,91522		2,79E-06	3,25E-07
Sputnik	0,97213	0,073335	0,98141	0,8953		2,61E-07
Zamilon	0,95271	0,069429	0,96722	0,92384	0,92626	

B)

Correlação entre frequências trinucleotídicas	*Candidatus* Babela massiliensis	Escherichia coli	Mimivirus	Mitocondria Apolyphaga	Sputnik	Zamilon
"*Candidatus* Babela massiliensis"		0,27656	2,63E-32	7,40E-30	3,70E-33	1,90E-29
Escherichia coli	0,13808		0,13296	0,17677	0,26745	0,27217
Mimivirus	0,94712	0,18984		7,89E-22	2,45E-39	7,39E-35
Mitocôndria de *A. polyphaga*	0,93624	0,17097	0,88108		5,94E-20	3,71E-23
Sputnik	0,95044	0,1407	0,96896	0,86198		6,86E-25
Zamilon	0,9342	0,13933	0,95644	0,89287	0,90644	

C)

Correlação entre frequências tetranucleotídicas	"*Candidatus* Babela massiliensis"	*Escherichia coli*	*Mimivirus*	*Mitocôndria de A. polyphaga*	*Sputnik*	*Zamilon*
Candidatus Babela massiliensis		0,0072585	1,16E-110	5,90E-103	2,09E-108	7,39E-97
Escherichia coli	0,16743		0,00056923	0,0015633	0,0048417	0,0038803
Mimivirus	0,92767	0,21392		1,42E-73	5,05E-133	4,91E-120
Mitocondria Apolyphaga	0,91634	0,1967	0,85267		2,66E-66	3,50E-73
Sputnik	0,92453	0,17558	0,95238	0,82984		1,16E-86
Zamilon	0,90606	0,17989	0,93934	0,85153	0,88579	

Tabela 2. Correlações de Pearson entre FR dinucleotídeos (A), trinucleotídeos (B) e tetranucleotídeos (C). Os valores abaixo da diagonal indicam o coeficiente de correlação linear (r). Em vermelho estão os os menores valores de *r e,* os maiores,

em verde. Acima da diagonal, estão as probabilidades bicaudais (p). O valor de corte de *p* foi escolhido com base no menor valor observado entre os genomas intra-amebais.

Fonte: Elaborada pelo autor.

As três análises de correlação (Tabela 2) foram realizadas usando as frequências relativas de di-, tri- e tetranucleotídeos observadas nos genomas, respectivamente. As análises revelaram que ao comparar as correlações dos padrões de dinucleotídeos (r^{di}), trinucleotídeos (r^{tri}) e tetranucleotídeos (r^{tetra}) das sequências observa-se uma leve redução nas correlações (r) enquanto que a significância estatística (p) aumenta consideravelmente – exceto no grupo externo (*E. coli*).

E. coli mostrou um progressivo aumento na correlação e da significância estatística com as demais sequências ao utilizar oligonucleotídeos maiores, entretanto, não mostrou valores significativos de correlação com as demais sequências em nenhum dos casos observados.

A sequência de *Mimivirus* mostrou maior correlação significativa com a *Sputnik* (r^{di} = 0,981 / r^{tri} = 0,968 / r^{tetra} = 0,952) e, em seguida, com *Zamilon* (r^{di} = 0,967 / r^{tri} = 0,956 / r^{tetra} = 0,939). Já a sua menor correlação significativa encontrada foi com a mitocôndria de *A. polyphaga* (r^{di} = 0,915 / r^{tri} = 0,881 / r^{tetra} = 0,852).

A bactéria "*Candidatus* Babela massiliensis" mostrou maior correlação com *Sputnik* nas frequências de di- e trinucleotídeos (r^{di} = 0,972 / r^{tri} = 0,950 / r^{tetra} = 0,924), porém, na frequência teranucleotídica o genoma de *Mimivirus* (r^{di} = 0,967 / r^{tri} = 0,947 / r^{tetra} = 0,927) apresentou correlação levemente maior que a correlação observada com *Sputnik*.

A mitocôndria de *A. polyphaga* mostrou maior correlação significativa com "*Candidatus* Babela massiliensis*"* (r^{di} = 0,952 / r^{tri} = 0,936 / r^{tetra} = 0,916). Sua correlação com os vírus *Mimivirus*, *Sputnik* e *Zamilon* também foi alta (r^{di} = 0,895~0,923 / r^{tri} = 0,861~0,892 / r^{tetra} = 0,829~0,852).

Por fim, ambos virófagos apresentaram maior correlação com Mimivirus, como já mencionado, chegando a ser maior do que correlação entre eles dois (r^{di} = 0,926 / r^{tri} = 0,906 / r^{tetra} = 0,885). *Sputnik* teve menor correlação com a sequência da mitocôndria de *A. polyphaga* (r^{di} = 0,895 / r^{tri} = 0,861 / r^{tetra} = 0,829).

Já *Zamilon* mostrou menor correlação com a bactéria "*Candidatus* Babela massiliensis" (r^{di} = 0,923 / r^{tri} = 0,892 / r^{tetra} = 0,851)

5.3) Análise de RSCU

O RSCU foi realizado utilizando a ferramenta *online* CAICal (PUIGBO; BRAVO; GARCIA-VALLVE, 2008) e plotado no *software* PAST 3.13. Através dele, notou-se que os genomas têm preferência por códons sinônimos terminados em nucleotídeos mais abundantes no genoma (ie.: influência direta do conteúdo GC). A figura 3 mostra a comparação entre o RSCU observado nas CDS analisadas dos genomas.

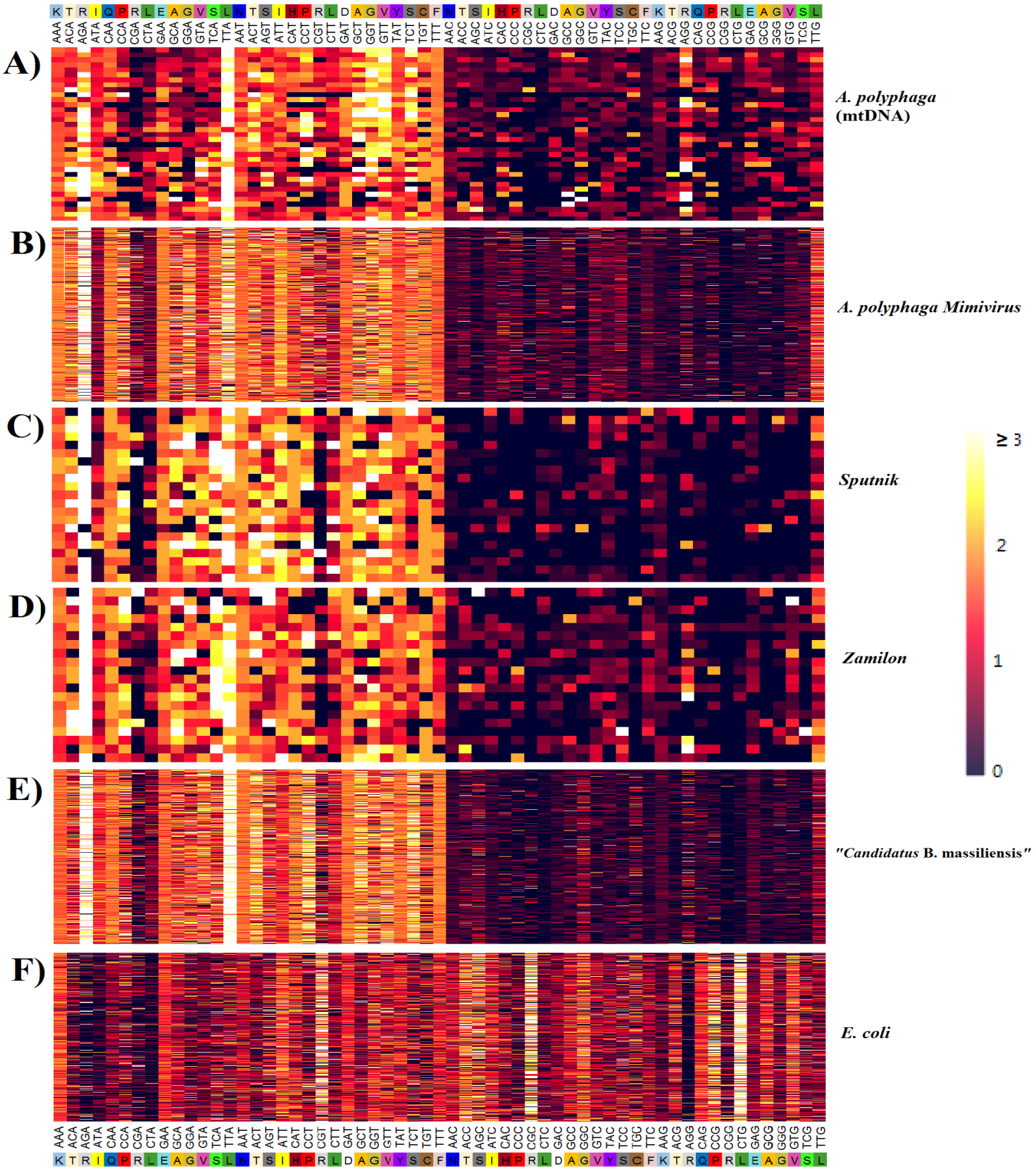


Figura 3. Análise de RSCU com as CDS dos genomas. Exceto em *Escherichia coli* onde foram utilizadas 1000 (das 5.600) CDS escolhidas ao acaso. Os códons estão ordenados por ordem alfabética de acordo com o nucleotídeo na 3ª posição. O cálculo do RSCU dos genomas foi feito utilizando a tabela de códons padrão, fornecida pela

própria ferramenta online do CAICal (PUIGBO; BRAVO; GARCIA-VALLVE, 2008). Na escala fornecida, os valores variaram entre 0 (não usado), 1 (pouco usado), 2 (moderadamente usado) e 3 ou mais (muito usado). Os códons de Metionina (ATG), Triptofano (TGG) e os códons de parada (TAA, TAG, TGA) não são incluídos nesta análise (SHARP; LI, 1986). Mais especificamente, os códons de Metionina e Triptofano não são incluídos na análise de RSCU por não apresentarem códons sinônimos e os códons de parada por apresentarem viés quanto a presença de enzimas chamadas “fatores de liberação”.

Fonte: Elaborada pelo autor.

As sequências codificantes (CDS) dos genomas aparentaram compartilhar um padrão de uso de códons, como observado pela presença de “faixas” brancas/avermelhadas ou pretas/escuras correspondentes aos códons mais utilizados ou evitados pelas CDS, respectivamente (Figura 3). Apesar disso, pode-se observar a existência de algumas CDS que mostram preferências por códons poucos utilizados como, por exemplo, apenas uma das 21 CDS de *Zamilon* mostrou uma utilização moderada (próximo de 2, na escala fornecida) pelo códon GCG.

O RSCU de *E. coli* mostrou um padrão diferente dos demais genomas, mostrando preferência elevada por códons terminados em C e G. Dentre estes códons terminados em C e G, os preferidos foram CTG (leucina), CCG (prolina) e CGC (arginina). Dentro dos códons ricos em A e T, o mais utilizado foi CGT (arginina). Os códons AGA (arginina) e TTA (leucina) apresentaram-se pouco utilizados em relação aos demais genomas.

O RSCU das sequências codificantes de *Mimivirus*, *Zamilon*, *S*putnik, “*Candidatus* B. massiliensis” e da mitocôndria de *A. polyphaga*, mostraram uma clara preferência por códons terminados em A e T (primeira metade) e aversão por códons terminados em C e G (segunda metade). Dentre os códons ricos em A e T, AGA (arginina) e TTA (leucina) foram os mais usados em todos estes genomas. Em geral, padrão de cores das faixas das CDS de *Mimivirus*, *Sputnik* e *Zamilon* mostrou-se bastante parecido entre si (Figura 3).

A bactéria “*Candidatus* B. massiliensis” mostrou padrão de uso de códons similar ao dos vírus e da mitocôndria (AGA, TTA). Também mostrou preferência por dois códons evitados por vírus e preferidos por *E. coli*: GGC (glicina) e TGC (cisteína).

O códon GGC (glicina) mostrou-se evitado apenas pelos vírus. Já o códon TTG (leucina) mostrou preferência relativamente similar em todos os genomas. Códons como AGG (arginina), CTA (leucina), AGA (lisina) e TTA (lisina) mostraram-se evitados apenas em *E. coli*.

5.4) Frequência relativa de uso de códons sinônimos (FRUCS)

A análise de FRUCS, diferentemente da análise de RSCU, retornou valores não-normalizados e incluiu os códons de parada. Os resultados reforçaram o resultado do RSCU, mostrando que a escolha de códons preferidos depende do conteúdo GC e que os virófagos apresentaram maior similaridade com APMV do que os REOs intra-amebais. A figura 4 ilustra o uso de códons pelos genomas, incluindo os códons de parada (que não são incluídos na análise de RSCU).

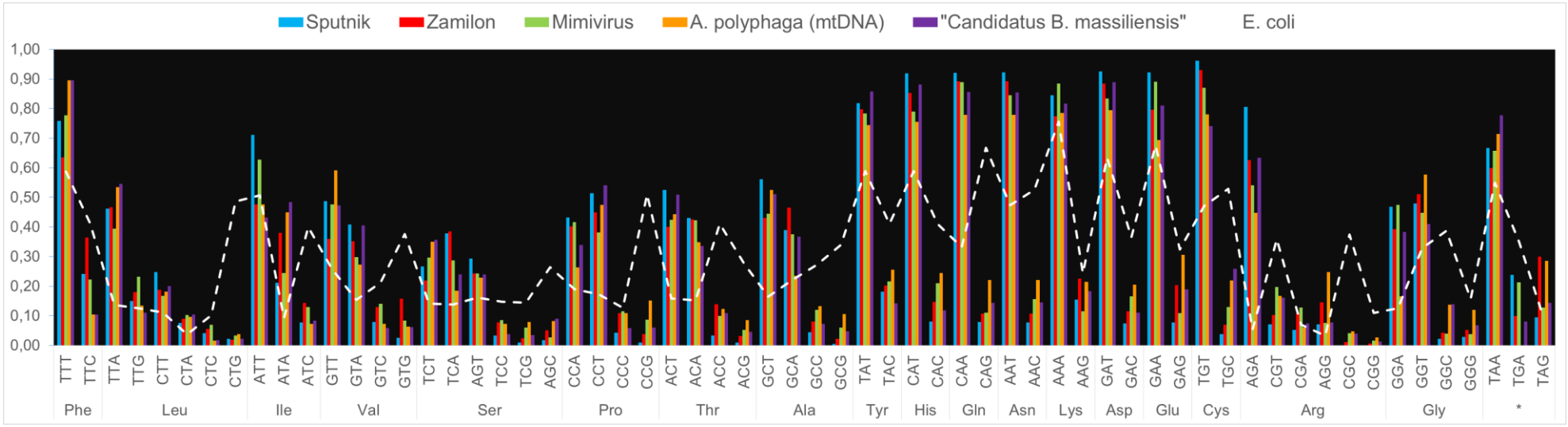


Figura 4. Frequência Relativa de Uso de Códons Sinônimos (FRUCS) com base nas frequências observadas em todas as CDS dos genomas. Os códons estão agrupados pelos seus aminoácidos cognatos. Os códons sinônimos estão ordenados em ordem descrescente com base na sua FR de uso, tendo Mimivirus como referencial. Os códons de Metionina e Triptofano não foram incluídos por não possuírem códons sinônimos. Diferente do RSCU, os códons de parada (*) foram incluídos nesta análise. O genoma de E. coli (grupo externo) está representado por uma linha tracejada de cor branca, para facilitar a comparação com os demais genomas (intra-amebais).

Fonte: Elaborada pelo autor.

Em relação à *Mimivirus*, *Sputnik* mostrou diferença apenas no códon preferencial para os aminoácidos Prolina (CCT). *Zamilon* mostrou diferença nos códons preferenciais dos aminoácidos Serina (TCA), Prolina (CCT), Alanina (GCA) e Glicina (GGT). A mitocôndria de *A. polyphaga* apresentou diferenças no códon preferido apenas no aminoácido Glicina (GGT). "*Candidatus* B. massiliensis" mostrou diferença nos códons preferenciais dos aminoácidos Isoleucina (ATA), Prolina (CCT) e Glicina (GGT). Nos casos observados (todos sendo terminados em A ou T), a diferença em relação aos códons sinônimos preferenciais de *Mimivirus* ocorreu no nucleotídeo da 3ª posição, de forma que o códon preferido por *Mimivirus* mostrou-se terminado em uma base complementar (ie.: XY**A** → XY**T** ou XY**T** → XY**A**, sendo X e Y nucleotídeos aleatórios).

Nota-se também que há divergências entre os vírus quanto ao códon de parada menos usado. Apesar de APMV, Sputnik e Zamilon terem TAA como códon de parada mais utilizado (com FR = ~60%), *Zamilon* evita o códon de parada TGA, o segundo mais utilizado por APMV (FR = 21,3%) e *Sputnik* (FR = 23,8%), utilizando-o em apenas duas ORFs (FR = 10%). Por sua vez, *Sputnik* e *Mimivirus* evitaram o códon de parada TAG (FR^{APMV} = 0,13; $FR^{Sputnik}$ = 0,1; $FR^{Zamilon}$ = 0,3). Já a mitocôndria de *A. polyphaga* evitou totalmente o códon de parada TGA.

A bactéria "*Candidatus* B. massiliensis" mostrou maior similaridade com a mitocôndria de *A. polyphaga* nos códons sinônimos preferidos TTT (Fenilalanina), TTA (Leucina), GCT (Alanina), TCT (Serina), TGT (Cisteína) e TAA (parada).

5.5) Análise de Correlação de Pearson entre valores de FRUCS dos genomas

Após as análises de FRUCS procurou-se ver qual era a correlação entre as FRUCS das CDS dos genomas. As análises de Correlação de Pearson foram calculadas no *software* PAST e apresentaram valores maiores que os observados nos trinucleotídeos ao longo dos genomas, de forma que os CEOs apresentaram as maiores correlações positivas observadas, como observado na Tabela 3.

	Sputnik	*Zamilon*	*Mimivirus*	*A. polyphaga* (mtDNA)	"*Candidatus* B. massiliensis"	*E. coli*
Sputnik		3,41E-42	1,05E-46	5,43E-27	8,53E-37	5,44E-05
Zamilon	0,97496		6,44E-41	4,08E-29	1,02E-35	7,82E-06
Mimivirus	0,98215	0,97244		1,91E-30	2,15E-37	7,73E-07
A. polyphaga (mtDNA)	0,92051	0,9325	0,93904		1,48E-38	9,06E-08
"*Candidatus* B. massiliensis"	0,96239	0,9592	0,96405	0,96707		2,62E-06
E. coli	0,48232	0,52663	0,57243	0,60946	0,54913	

Tabela 3. Análise de Correlação de Pearson entre as FRUCS observadas na análise relativa à Figura 4. Os valores da diagonal inferior representam a correlação (r) entre as frequências relativas de códons dos genomas. Em verde estão os valores mais próximos de 1 e em vermelho os mais distantes. Na diagonal superior estão os valores da significância estatística (p). O valor de corte de p foi escolhido com base no menor valor de significância observado nos genomas intra-amebais.

Fonte: Elaborada pelo autor.

De acordo com os resultados da Tabela 3, APMV compartilha uma maior similaridade no uso de códons com *Sputnik* (r = 0,982 p = 1,05e-46) e, em seguida, com *Zamilon* (r = 0,972 p = 6,44e-41) do que com os demais organismos comparados. *Sputnik* mostrou maior correlação com APMV enquanto *Zamilon* mostrou uma maior correlação com *Sputnik* (r = 0,974 p = 3,41e-42) do que com APMV, porém a diferença foi muito baixa (< 0,0025) para ser considerada significante.

"*Candidatus* B. massiliensis" mostrou maior correlação com a mitocôndria de *A. polyphaga* (r = 0,967 p = 1,48e-38). Apesar disso, a menor correlação significativa observada mostrou-se apenas um pouco menor e foi observada com *Zamilon* (r = 0,959 p = 1,02e-35).

O genoma de *E. coli* mostrou correlação com o genoma da mitocôndria de *A. polyphaga* (r = 0,604, p = 9,06e-08) e, em seguida, com APMV (r = 0,572 p = 7,73e-07). Apesar dos valores destas correlações normalmente serem considerados

moderados, suas significâncias estatísticas foram baixas quando comparadas com as encontradas entre os demais genomas, que por sua vez, atingiram um mínimo de 5,43e-27.

As correlações observadas aqui na análise de FRUCS mostraram-se maiores do que os valores encontrados na correlação de frequências trinucleotídicas (Tabela 2).

5.6) Frequência Relativa de Códons (FRC)

A análise de FRC, por fim, mostrou quais eram os códons mais prevalentes nas CDS dos genomas, como observado na Figura 5. Verificou-se que, em relação a APMV, os virófagos foram os que apresentaram padrões de uso mais similares e que "*Candidatus* B. massiliensis" apresentou padrões de uso similares aos virófagos e com a mitocôndria.

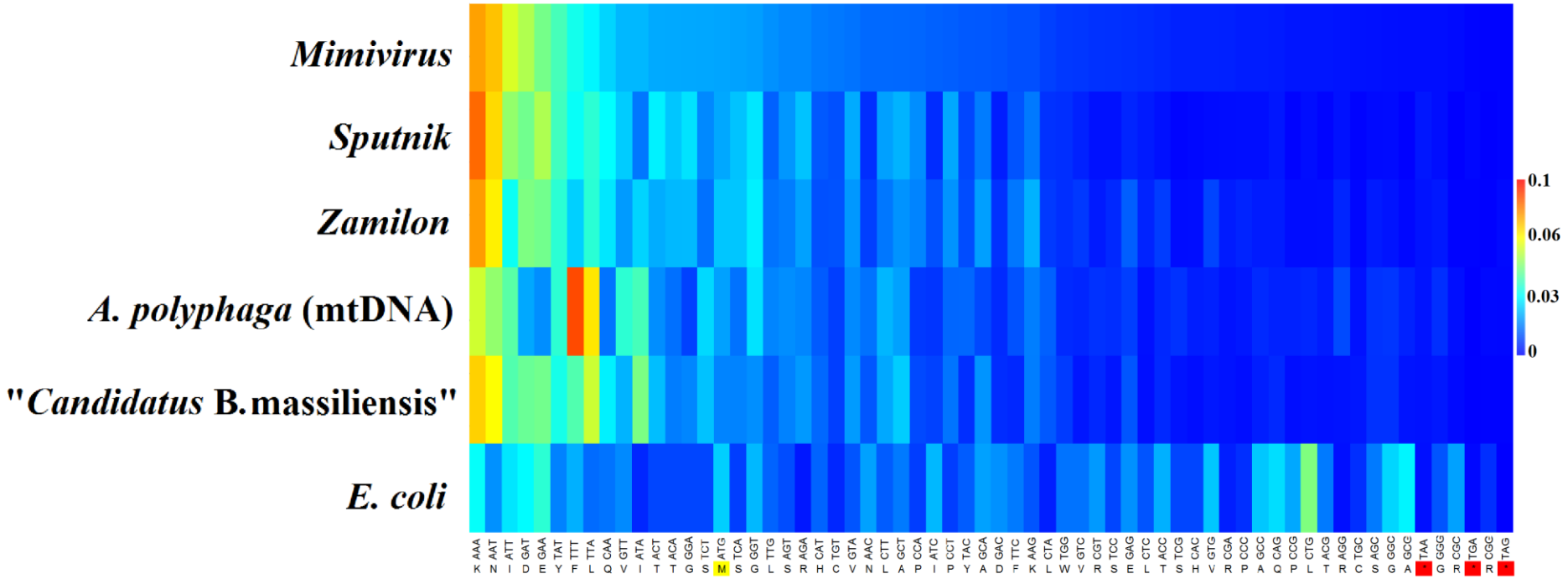


Figura 5. Frequência relativa dos códons utilizados pelas CDS dos genomas. Nesta análise, foram utilizadas todas as CDS dos genomas comparados e foram considerados os códons não analisados no RSCU e/ou FRUCS: Metionina (M), Triptofano (W) e códons de parada (*). Os códons de parada estão destacados em vermelho, enquanto que o códon de início/metionina está destacado em amarelo. Os códons estão ordenados em ordem decrescente de frequência com base nos valores observados em APMV.

Fonte: Elaborada pelo autor.

Observando as cores nas bandas relativas aos genomas de *Sputnik* e *Zamilon*, nota-se que eles mostram preferência similar a APMV do códon AAA até o códon TAT (FR = 0,03 ~ 1). Ainda neste intervalo, *Zamilon* apresentou reduzida preferência de uso pelo códon ATT (FR = ~0,03).

"*Candidatus* B. massiliensis" mostrou preferência pelos códons AAA (Lisina; FR ≥ 0,06) e AAT (Asparagina; FR ≥ 0,06), também preferidos por APMV, *Sputnik* e *Zamilon*. Por outro lado, "*Candidatus* B. massiliensis" mostrou relativa semelhança com a frequência de uso do códon TTA (Leucina; FR = 0,06 ~ 0,1), preferidos pela mitocôndria de *A. polyphaga*.

A mitocôndria de *A. polyphaga* mostrou um padrão de uso de códons distinto dos organismos intra-amebais, tendo TTT (Fenilalanina) e TTA (Leucina) como códons com maior frequência relativa (FR > 0,06). GAT (Ácido Aspártico) e GAA (Ácido Glutâmico) mostraram baixa frequência relativa (FR < 0,03), enquanto que os três vírus e a bactéria "Candidatus B. massiliensis" apresentaram frequências moderadas (FR = 0,03 ~ 0,06).

E. coli mostrou um padrão de UC relativamente bem distribuído, por ter um conteúdo GC equilibrado. O códon com maior utilização em *E. coli* foi CTG (Leucina; FR > 0,03), diferente de todos os outros genomas que apresentaram baixa utilização (FR < 0,03).

6) Discussão

Os resultados encontrados indicaram a existência de uma elevada similaridade nas AGs (observadas nas análises de FDN e UC) dos virófagos *Sputnik* e *Zamilon* com o vírus-auxilador APMV. Os genomas da mitocôndria de *A. polyphaga* e de "*Candidatus* Babela massiliensis" apresentaram maior similaridade entre si do que com os outros genomas (*APMV*, *Zamilon*, *Sputnik* e *E. coli)*, chegando a ter mais similaridades com os vírus (APMV, *Zamilon* e *Sputnik*) do que com *E. coli*. Já *E. coli* mostrou um padrão distinto dos demais genomas em ambas análises de AG, confirmando-se como um grupo externo.

A semelhança observada nos padrões observados de FDN (Figura 2) de APMV e seus virófagos pode indicar a ocorrência de uma "convergência" nas suas AG. A "convergência" nas AGs de APMV e seus virófagos pode ser explicada pelo uso do mesmo maquinário de replicação (LA SCOLA, 2008) e reparo (MOLINER; FOURNIER; RAOULT, 2010 apud RAOULT et al., 2004), supostamente responsáveis pela constância da AG (KARLIN, 1998), ou como uma adaptação para escapar de mecanismos de reconhecimento de padrão (GREENBAUM et al., 2008; LOBO et al., 2009) do APMV, como observado no sistema MIMIVIRE (LEVASSEUR, 2016). Desta forma, a convergência nas AGs seria então regida por um modelo coevolutivo de DCA, requerendo assim um período relativamente longo de coevolução com o hospedeiro. Assim, como *Sputnik* foi o virófago que apresentou maior semelhança com as AGs de APMV, como visto nas correlações de frequências di-, tri-, tetranucleotídicas (ver Tabela 2), e UC (ver Figuras 3 – 5), pode ser que ele seja mais adaptado ao APMV por ter um histórico coevolutivo mais antigo que o de *Zamilon*. Isto leva a crer que o sucesso de replicação viral (ie.: o sucessivo uso do maquinário de replicação e reparo por vírus/virófagos) acarreta na ocorrência de adaptação à sistemas de reconhecimento de padrão, por levar a uma eventual imitação (mímica) da AG do hospedeiro, confirmando a proposta de Karlin (1998).

Neste estudo, a mitocôndria e todos os CEO e REO intra-amebais mostraram conteúdo GC baixo. Isto também é observado em genomas de bactérias capazes de reproduzir no interior de amebas como *Legionella pneumophila, Legionella drancourtii, Rickettsia bellii, "Candidatus* Protochlamydia amoebophila" e *"Candidatus*

Amoebophilus asiaticus" que possuem conteúdo GC variando de 31 a 38% (MOLINER; FOURNIER; RAOULT, 2010). Assim, pode ser que a tendência em apresentar elevado conteúdo AT pode estar relacionado com a capacidade de sobrevivência intracelular na ameba. Ainda, pode ser que semelhanças no conteúdo AT e, de forma mais específica, nas AGs favoreçam a ocorrência de transferência lateral gênica, comum nestes organismos intra-amebais (MOLINER; FOURNIER; RAOULT, 2010; PAGNIER et al., 2015).

As correlações observadas entre os genomas com base nas frequências oligonucleotídicas (ver Tabela 2) mostraram que os virófagos tinham maior correlação (nas três análises) com APMV, depois com "*Candidatus* Babela massiliensis" e então entre os seus genomas. A alta correlação das frequências oligonucleotídicas dos dois virófagos com APMV, reforça a idéia de que o uso do mesmo maquinário de reprodução e reparo (KARLIN; BURGE, 1995; KARLIN, 1998) podem estar envolvidos na convergência das AGs com a de APMV. A maior correlação oligonucleotídica de "*Candidatus* Babela massiliensis" foi observada com *Sputnik*, porém apenas nas frequências de di- e trinucleotídeos. Já nas frequências tetranucleotídicas, APMV passa a ser o genoma com maior correlação com *"Candidatus* Babela massiliensis", por muito pouca diferença com *Sputnik*. O uso de oligonucleotídeos maiores poderiam ajudar a explicar melhor esta evidência, visto que eles apresentam maior espécie-especificidade (DESCHAVANNE et al., 2000).

As análises de RSCU mostaram que há uma tendência nos genomas da mitocôndria de *A. polyphaga*, "*Candidatus* Babela massiliensis" e dos CEOs quanto ao nucleotídeo da 3ª posição, sendo ele claramente influenciado pelo conteúdo GC. Assim, nota-se pouca diferença entre os resultados dos CEOs, da mitocôndria e de "*Candidatus* Babela massiliensis". Apesar disso, "*Candidatus* Babela massiliensis" também mostrou preferências por códons terminados em C (GGC e TGC), como também observado em *E. coli*. Campbell, Mrázek e Karlin (1999) propuseram que genomas mitocondriais retêm AGs próximas de seus ancestrais procarióticos que possuíam enzimas de reparo já que as mesmas não possuem. Assim, no caso de "*Candidatus* B. massiliensis", que possui um gene para enzima de reparo (situado na ORF "BABL1_RS04045") e um sistema de replicação reduzido (PAGNIER et al., 2015). Pode então ser que estas proteínas estejam mantendo a AG do seu genoma (KARLIN, 1998) similar à época que a mesma era independente da ameba.

A análise de FRUCS (Figura 4) reforçou os resultados encontrados no RSCU, visto que os códons sinônimos preferidos foram todos terminados em A e T, incluindo os códons de parada. Como no RSCU, *Sputnik* mostrou-se mais semelhante à APMV do que *Zamilon*, também evidenciada pelo uso de códons de parada. O fato da mitocôndria não apresentar CDS terminadas em TGA indica que a mesma não possui fatores de liberação capazes de atuar neste códon de parada ou que o códon TGA foi reassignado para um códon funcional (KORKMAZ et al., 2014 apud TATE et al., 1999).

A análise de correlação com os dados de FRUCS, mostrou que há uma correlação maior entre as CDS dos genomas do que a encontrada utilizando as frequências de trinucleotídeos ao longo dos genomas (ver Tabela 2). Porém, diferente da análise de trinucleotídeos, a correlação entre *Sputnik* e *Zamilon* mostrou-se maior do que a de *Sputnik* e "*Candidatus* B. massiliensis". Isto indica que as pressões que mantêm a AG das regiões codificantes são diferentes daquelas que ocorrem ao longo do genoma inteiro, provavelmente relacionadas com a estrutura das proteínas (GREENBAUM et al., 2008) ou dos mRNAs (DUAN; ANTEZANA, 2003). Ainda, isto pode explicar o considerável aumento evidenciado na correlação de FRUCS das CDS de *E. coli* com os demais genomas.

Assim como no RSCU, a análise de FRC nas CDS (ver Figura 5) mostrou-se diretamente relacionada com o conteúdo GC do organismo. Esta análise também mostrou que os virófagos tem UC mais similar à APMV. "*Candidatus* B. massiliensis" mostrou semelhança com os CEOs nos códons mais utilizados. Estas evidências indicam que provavelmente há algo que pode estar favorecendo um UC rico em A e T nos parasitas da ameba. Isto pode estar relacionado ao repertório de tRNAs cognatos (SHARP, TUOHY, MOSURSKI, 1986; KUMAR, KUMARI, SHARMA, 2016; DUAN, ANTEZANA, 2003), velocidade e eficiência de transcrição (PLOTKIN; KUDLA, 2011; LOBO et al, 2009) ou com estrutura secundária do mRNA (DUAN; ANTEZANA, 2003; MCDOWALL; LIN-CHAO; COHEN, 1994).

7) Considerações Finais:

Os resultados levam a crer que dentre os dois virófagos, *Sputnik* é o que apresenta maior grau de adaptação a APMV (dada a maior similaridade entre suas AGs) pois, provavelmente, compartilha um histórico coevolutivo maior que o observado entre APMV e *Zamilon*.

Como se sabe, *Acanthamoeba* sp. não suprime o seu conteúdo GC (KARLIN; BURGE, 1995). Os projetos de sequenciamento do genoma de *A. polyphaga*, que ainda estão em andamento (Número do projeto no NCBI: 35827), indicam que seu genoma possui um conteúdo GC elevado (próximo de 58%), ao contrário dos CEOs aqui estudados. Este fato, aliado às evidências aqui achadas, indicam que é bem provável que as AGs dos CEOs estejam evoluindo de forma indepedente da ameba, implicando que a similaridade na AG de *Sputnik* e *Zamilon* dependem diretamente, se é que não exclusivamente, do APMV e suas fábricas virais.

Por fim, os métodos utilizados neste trabalho permitiram a obtenção de resultados independente de homologias entre os genomas e suas CDS. Estudos futuros podem vir a aprofundar as relações aqui observadas entre APMV e seus virófagos.

Referências

1. ABE, Takashi et al. Informatics for unveiling hidden genome signatures. **Genome research**, v. 13, n. 4, p. 693-702, 2003.
2. ALBERTS, Bruce et al. **Biologia molecular da célula**. Artmed Editora, 2009.
3. BENSON, Dennis A. et al. GenBank. **Nucleic acids research**, v. 41, n. D1, p. D36-D42, 2013. Disponível em: https://www.ncbi.nlm.nih.gov/genbank. Acesso em: 07 nov. 2016
4. CAMPBELL, Allan; MRAZEK, Jan; KARLIN, Samuel. Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. **Proceedings of the National Academy of Sciences**, v. 96, n. 16, p. 9184-9189, 1999.
5. DESCHAVANNE, Patrick et al. Genomic signature is preserved in short DNA fragments. In: **Bio-Informatics and Biomedical Engineering, 2000. Proceedings. IEEE International Symposium on**. IEEE, 2000. p. 161-167.
6. DUAN, Jubao; ANTEZANA, Marcos A. Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. **Journal of Molecular Evolution**, v. 57, n. 6, p. 694-701, 2003.
7. EHRLICH, Paul R.; RAVEN, Peter H. Butterflies and plants: a study in coevolution. **Evolution**, p. 586-608, 1964.
8. FORTERRE, Patrick. Defining life: the virus viewpoint. **Origins of Life and Evolution of Biospheres**, v. 40, n. 2, p. 151-160, 2010.
9. GAIA, Morgan et al. Zamilon, a novel virophage with Mimiviridae host specificity. **PLoS One**, v. 9, n. 4, p. e94923, 2014.
10. GREENBAUM, Benjamin D. et al. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. **PLoS Pathog**, v. 4, n. 6, p. e1000079, 2008.
11. HAMMER, Ø.; HARPER, D. A. T.; RYAN, P. D. PAST: Paleontological Statistics Software Package for education and data analysis. Palaeontolia Electronica 4. 2001.
12. JANZEN, Daniel H. When is it coevolution. **Evolution**, v. 34, n. 3, p. 611-612, 1980.
13. JERNIGAN, Robert W.; BARAN, Robert H. Pervasive properties of the genomic signature. **BMC genomics**, v. 3, n. 1, p. 1, 2002.
14. KARLIN, Samuel. Global dinucleotide signatures and analysis of genomic heterogeneity. **Current opinion in microbiology**, v. 1, n. 5, p. 598-610, 1998.
15. KARLIN, Samuel; BURGE, Chris. Dinucleotide relative abundance extremes: a genomic signature. **Trends in genetics**, v. 11, n. 7, p. 283-290, 1995.
16. KORKMAZ, Gürkan et al. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. **Journal of Biological Chemistry**, v. 289, n. 44, p. 30334-30342, 2014.
17. KUMAR, Sushil; KUMARI, Renu; SHARMA, Vishakha. Coevolution mechanisms that adapt viruses to genetic code variations implemented in their hosts. **Journal of genetics**, v. 95, n. 1, p. 3-12, 2016.
18. LA SCOLA, Bernard et al. A giant virus in amoebae. **Science**, v. 299, n. 5615, p. 2033-2033, 2003.
19. LA SCOLA, Bernard et al. The virophage as a unique parasite of the giant mimivirus. **Nature**, v. 455, n. 7209, p. 100-104, 2008.

20. LEGENDRE, Matthieu et al. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. **Virology journal**, v. 8, n. 1, p. 1, 2011.
21. LEVASSEUR, Anthony et al. MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. **Nature**, v. 531, n. 7593, p. 249-252, 2016.
22. LEWIN, Benjamin et al. **genes IX**. Mc Graw-Hill Interamericana,, 2008.
23. LOBO, Francisco P. et al. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. **PloS one**, v. 4, n. 7, p. e6282, 2009.
24. MARTINY, Jennifer BH et al. Antagonistic coevolution of marine planktonic viruses and their hosts. **Marine Science**, v. 6, 2014.
25. MCDOWALL, Kenneth J.; LIN-CHAO, Sue; COHEN, Stanley N. A+ U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. **Journal of Biological Chemistry**, v. 269, n. 14, p. 10790-10796, 1994.
26. MITCHELL, David. GC content and genome length in Chargaff compliant genomes. **Biochemical and biophysical research communications**, v. 353, n. 1, p. 207-210, 2007.
27. MOLINER, Claire; FOURNIER, Pierre-Edouard; RAOULT, Didier. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. **FEMS microbiology reviews**, v. 34, n. 3, p. 281-294, 2010.
28. PAGNIER, Isabelle et al. Babela massiliensis, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. **Biology direct**, v. 10, n. 1, p. 1, 2015.
29. PEARSON, Helen. 'Virophage'suggests viruses are alive. **Nature News**, v. 454, n. 7205, p. 677-677, 2008.
30. PLOTKIN, Joshua B.; KUDLA, Grzegorz. Synonymous but not the same: the causes and consequences of codon bias. **Nature Reviews Genetics**, v. 12, n. 1, p. 32-42, 2011.
31. PUIGBÒ, Pere; BRAVO, Ignacio G.; GARCIA-VALLVE, Santiago. CAIcal: a combined set of tools to assess codon usage adaptation. **Biology direct**, v. 3, n. 1, p. 1, 2008. Disponível em: http://genomes.urv.es/CAIcal/. Acesso em: 7 nov. 2016.
32. RAOULT, Didier et al. The 1.2-megabase genome sequence of Mimivirus. **Science**, v. 306, n. 5700, p. 1344-1350, 2004.
33. RAOULT, Didier; FORTERRE, Patrick. Redefining viruses: lessons from Mimivirus. **Nature Reviews Microbiology**, v. 6, n. 4, p. 315-319, 2008.
34. RICKLEFS, Robert E. **A economia da natureza.** 6. ed. Guanabara Koogan, 2010.
35. RIDLEY, Mark. **Evolução**. 3. ed. Artmed, 2006.
36. SHARP, Paul M.; LI, Wen-Hsiung. Codon usage in regulatory genes in Escherichia coli does not reflect selection for ‘rare’codons. **Nucleic Acids Research**, v. 14, n. 19, p. 7737-7749, 1986.
37. SHARP, Paul M.; TUOHY, Therese MF; MOSURSKI, Krzysztof R. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. **Nucleic acids research**, v. 14, n. 13, p. 5125-5143, 1986.
38. TATE, W. P. et al. UGA: a Dual Signal for" Stop" and for Recoding in Protein Synthesis. **BIOCHEMISTRY C/C OF BIOKHIMIIA**, v. 64, n. 12, p. 1342-1353, 1999.

39. VAN ROSSUM, Guido; DRAKE JR, Fred L. **Python tutorial**. Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica, 1995.
40. WODARZ, Dominik. Evolutionary dynamics of giant viruses and their virophages. **Ecology and evolution**, v. 3, n. 7, p. 2103-2115, 2013.
41. YAU, Sheree et al. Virophage control of antarctic algal host–virus dynamics. **Proceedings of the National Academy of Sciences**, v. 108, n. 15, p. 6163-6168, 2011.

Apêndice

Apêndice A: Pseudocódigo do algoritmo de contagem de oligonucleotídeos.

INÍCIO

VARIÁVEIS

genoma: É a sequência FASTA convertida em uma única *string*.

nuc: vetor contendo os 4 nucleotídeos (A,T,C e G) e suas quantidades no genoma

dinuc: vetor contendo os 16 dinucleotídeos e suas quantidades no genoma

trinuc: vetor contendo os 64 trinucleotídeos e suas quantidades no genoma

tetranuc: vetor contendo os 256 tetranucleotídeos e suas quantidades no genoma

i: iterador

PARA_CADA nucleotideo **EM** genoma **FAÇA:**

nuc[nucleotideo] += 1 //Incrementa a contagem do dado nucleotídeo

dinuc[nucleotideo + (nucleotídeo+1)] += 1 //Incrementa a contagem do dinucleotídeo

trinuc[nucleotideo + (nucleotídeo+2)]] += 1 //Incrementa a contagem do trinucleotídeo

tetranuc[nucleotideo + (nucleotídeo+3)]] += 1 //Incrementa a contagem do tetranucleotídeo

FIM_PARA_CADA

i <- 0 //Zerando o iterador

PARA_CADA i **EM** nuc: //Mostra a Frequência Relativa de nucleotídeos

ESCREVER (“Frequência de nucleotídeos: ”, nuc / tamanho(genoma))

FIM_PARA_CADA

i <- 0 //Zerando o iterador

PARA_CADA i **EM** dinuc**:**

ESCREVER (“Frequência de dinucleotídeos: ”, dinuc / tamanho(genoma))

FIM_PARA_CADA

i <- 0 //Zerando o iterador

PARA_CADA i **EM** trinuc:

    ESCREVER (“Frequência de trinucleotídeos: ”, trinuc / tamanho(genoma))

FIM_PARA_CADA

i <- 0 //Zerando o iterador

PARA_CADA i **EM** tetranuc:

    ESCREVER (“Frequência de tetranucleotídeos: ”, tetranuc / tamanho(genoma))

FIM_PARA_CADA

FIM