



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS**

Análise de ilhas genômicas em grandes vírus nucleocitoplasmáticos de DNA da Família Mimiviridae

Victor Montenegro Marcelino

Orientador: Prof. Dr. Sávio Torres de Farias

Co-orientadora: Prof. Dra. Thaís Gaudêncio do Rego

João Pessoa- PB

Novembro - 2016

**Análise de ilhas genômicas em grandes vírus nucleocitoplasmáticos de DNA da Família
Mimiviridae**

Victor Montenegro Marcelino

Orientador: Prof. Dr. Sávio Torres de Farias

Co-orientadora: Prof. Dra. Thaís Gaudêncio do Rego

Trabalho - 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 - 2016

Agradecimentos

Agradeço a minha mãe Ana Lucia e ao meu pai Francisco Tomas e minha irmã Inngrid por me encorajarem e apoiarem nessa durante todo o curso de biologia.

Agradeço ao professor Sávio Torres de Farias e a professora Thaís Gaudêncio do Rego por me orientarem nesse trabalho de conclusão de curso.

Agradeço aos amigos do Laboratório de genética evolutiva Paulo Lemiski por me ajudarem com este presente trabalho, em especial : Camila , Paulo e Vanessa.

Agradeço aos colegas Daniel Brito e Thais Ratis do Laboratório Visio, por me ajudarem na produção deste trabalho , juntamente com Paulo e Lucas da biologia.

Agradeço ao professor Edson Luiz Folador e ao doutorando Víctor Manuel Serrano Solís por participar da banca do meu trabalho de conclusão de curso.

RESUMO

Durante os últimos anos vem se discutindo bastante a evolução dos vírus. A descoberta dos grandes vírus de DNA nucleocitoplasmáticos (do inglês, Nucleocytoplasmic large DNA viroses, NCLDV) foi de grande ajuda para entender como os vírus evoluíram visto que os vírus comuns sofrem muita mutação e apresentam um genoma muito pequeno dificultando o encontro de um sinal evolutivo, diferentemente dos NCLDV que apresentam maquinaria de reparo de DNA e um genoma grande. Existe, atualmente, duas hipóteses de como esses vírus poderiam ter adquirido o tamanho gigante do seu genoma em relação aos demais vírus. A hipótese de aquisição de genes por transferência horizontal formando ilhas genômicas e a hipótese de degradação do genoma, onde os grandes vírus estariam perdendo material genético para o meio. Neste trabalho procuro encontrar evidências que corroborem com a hipótese de aquisição de material genético para os grandes vírus verificando a existência de ilhas genômicas em cada um dos genomas já sequenciado da família de Mimivirus, utilizando um algoritmo de clusterização *Mean-Shift* e para confirmar se de fato o genoma apresenta uma ilha genômica, utilizo filogenias para identificar possíveis transferências laterais de genes.

Palavras-chave: Evolução, Ilhas Genômicas, Mimivirus

ABSTRACT

During the last few years the evolution of the virus has been much discussed. The discovery of the large nucleocytoplasmic DNA viruses (Nucleocytoplasmic Large DNA Viruses, NCLDV) was a great help to understand how viruses evolved since the common viruses undergo much mutation and present a very small genome making it difficult to find an evolutionary signal, NCLDV systems that present DNA repair machinery and a large genome. There are currently two hypotheses as to how these viruses could have acquired the giant size of their genome in relation to the other viruses. The hypothesis of acquisition of genes by horizontal transfer forming genomic islands and the hypothesis of degradation of the genome, where the large viruses are losing genetic material to the medium. In this work, I try to find evidence that supports the hypothesis of genetic material acquisition for large viruses by verifying the existence of genomic islands in each of the genomes already sequenced from the Mimivirus family using a *Mean-shift* clustering algorithm and to confirm if Fact the genome presents a genomic island, I use phylogenies to identify possible lateral gene transfers.

Keywords: Evolution, Genomic Islands, Mimivirus

LISTA DE FIGURAS

Figura 1 – Fator de libertação da cadeia peptídica eRF1.....	13
Figura 2 – Provável Endonuclease VIII.....	14
Figura 3 – D5-ATPase-Helicase.....	15
Figura 4 – Provável proteína CFxQ.....	16
Figura 5 – RNA Ligase 2.....	17
Figura 6 – Provável proteína p13.....	18
Figura 7 – Provável proteína colagénio 7.....	19
Figura 8 – Provável DnaK Hsp70.....	20
Figura 9 – Provável proteína chemotaxis CheD.....	21
Figura 10 – Provável Glicosiltransferase.....	22
Figura 11 – Fator de Troca de nucleotídeo Guanina.....	23
Figura 12 – Provável proteína Bifuncional Metalloproteaseubiquitina ligase.....	24
Figura 13 – Provável proteína Serinethreonine quinase.....	25
Figura 14 – Provável UvrDREP helicase.....	26
Figura 15 – Provável enzima ativadora da ubiquitina E1.....	27

LISTA DE TABELAS

Tabela 1 – Ilhas genômicas.....	11
Tabela 2 – Fatores de transcrição basal.....	28

Sumário

1. Introdução	1
2. Objetivos	2
2.1 - Objetivos gerais	2
2.2 - Objetivos específicos	2
3. Fundamentação teórica	2
3.1 – Mimivirus	2
3.2 - Transferência Lateral de genes e ilhas genômicas	4
3.3 - Transferência lateral de genes em Mimivirus.	5
3.4 - Detecção de ilhas genômicas usando mean-shift	7
4. Metodologia	8
4.1 - Escolha dos organismos	8
4.2 - Obtenção das proteínas das ilhas genômicas	8
4.3 - Obtenção das sequências de proteínas para filogenia	8
4.4 - Detecção de ilhas genômicas	9
4.5 - Alinhamento e modelo evolutivo	9
4.6-Árvores filogenéticas	10
4.7 - Comparação de vias metabólicas	10
5. Resultados	10
5.1 - Ilhas genômicas	10
5.2 - Filogenias das proteínas das ilhas	12
5.3 - Fator de libertação da cadeia peptídica eRF1	13
5.4 - Provável Endonuclease VIII	14
5.5 - D5-ATPase-Helicase	15
5.6 - Provável proteína CFxQ	16
5.7 - RNA Ligase 2	17
5.8 - Provável proteína p13	18
5.9 - Provável proteína colagénio 7	19
5.10 - Provável DnaK Hsp70	20
5.11 - Provável proteína chemotaxis CheD	21
5.12 - Provável Glicosiltransferase	22
5.13 - Fator de Troca de nucleotídeo Guanina	23
5.14 - Provável proteína Bifuncional Metalloproteaseubiquitina ligase	24
5.15 - Provável proteína Serinethreonine quinase	25
5.16 - Provável UvrDREP helicase	26
5.17 - Provável enzima ativadora da ubiquitina E1	27

5.18 - Fatores de transcrição basal.....	28
6. Discussão	30
7. Considerações finais	34
Referências	35
ANEXO A.....	40
Anexo B Tabela de transcrição	56
Anexo C	58
Anexo D	58

1. Introdução

Durante os últimos anos vem se discutindo bastante a evolução dos vírus para tentar inferir sua provável origem. Porém, a maior parte dos estudos utiliza uma abordagem de genéticas de populações onde procura-se mostrar novos polimorfismos genéticos entre diferentes linhagens. Outros estudos filogenéticos com vírus tem apresentado dificuldades na execução pela propensão dos vírus a adquirirem material genético aleatório dos seus hospedeiros e do ambiente, pela sua taxa de divergência alta em suas sequências (MONIER; CLAVERIE; OGATA, 2007). Com a descoberta dos grandes vírus nucleocitoplasmáticos de DNA (do inglês Nucleocytoplasmic large DNA viroses, NCLDV), essas dificuldades mudaram, pois estes vírus apresentam proteínas de reparo de DNA (RAOULT *et al.*, 2004), que impedem a alta taxa de divergência dentro dos vírus, facilitando encontrar um sinal filogenético para esses organismos e podendo traçar mais facilmente sua história evolutiva na árvore da vida.

Os modelos de evolução dos mimivirus é um campo de forte discussão na ciência. Atualmente, existem duas hipóteses de como estes teriam surgido: a primeira é a de que esses vírus são grupo irmão dos protoeucariotos, uma vez que seu genoma teria evoluído de um organismo celular, de maneira que uma célula teria perdido material genético e se tornado um vírus. Entretanto, existe uma outra hipótese que sugere que eles sofreram diversas transferências horizontais gênicas onde adquiriram mais de quinhentos genes, que se adaptam ao seu hospedeiro e transformaram o vírus pequeno em um grande vírus, explicando a origem dos genes celulares presentes nos Mimivirus (PEREIRA, 2014). Por causa destas características únicas dos Mimivirus, além do fato de serem parte de um grupo descoberto recentemente e que está em ascensão na comunidade científica, neste trabalho tenta-se esclarecer a provável origem deste grupo baseando-se na hipótese que estes vírus adquiriram muito material genético de transferência lateral gênica e desenvolveram um grande tamanho em comparação aos demais vírus.

2. Objetivos

2.1 - Objetivos gerais

Verificar a hipótese de que os Mimivirus teriam evoluído por aquisição de genes advindas de transferência horizontal gênica.

2.2 - Objetivos específicos

Fazer uma comparação da maquinaria basal de transcrição de Mimivirus com modelos de organismos eucariotos e procariotos.

Fazer uma análise filogenética das prováveis proteínas das ilhas genômicas dos Mimivirus para descobrir a provável origem de cada ilha.

3. Fundamentação teórica

3.1 – Mimivirus

Os Mimivirus foram descobertos em 2003 por isolamento em co-cultura com a ameba *Acanthamoeba polyphaga* e, desde então, quebraram paradigmas relacionados a vírus e a definição de vida. Mimivirus e outros vírus gigantes são definidos pelo seu vírion e genoma de tamanho gigante, por sua complexidade, pela presença de DNA e mRNA, a presença de dezenas ou centenas de proteínas em vírions e componentes associados com os processos de tradução (SHARMA et al., 2016), são visíveis em microscopia de luz e são retidos por filtros e seu material genético é muitas vezes ainda maior do que as de bactérias. Foram encontrados alguns de seus genes codificantes para proteínas únicas entre os vírus e envolvidos em

processos como a reparação do DNA, dobras de proteínas, de síntese de nucleotídeos, metabolismo de aminoácidos, proteína de modificação de lipídios ou metabolismos de polissacarídeos (RAOULT *et al.*, 2004), bem como, tradução e transcrição de proteínas, que é algo que não tem precedentes em vírus. Tais proteínas mostram que esses organismos têm sua própria maquinaria de replicação completa, com várias proteínas de tradução (AHERFI *et al.* 2016). Foi detectado RNA mensageiro no capsídeo de Mimivirus e no genoma foram encontrados genes para codificar RNAs de transferência e abriga os genes promotores iniciais e tardios (RAOULT *et al.*, 2004; RENESTO *et al.*, 2006; LEGENDRE *et al.*, 2010 apud AHERFI *et al.*, 2016).

Também foi descoberto que Mimivirus servem como hospedeiro para um outro tipo menor de vírus, que foi chamado de virófago, como uma analogia para bacteriófago (dentro desses vírus, sendo os exemplos mais notáveis, Sputnik e Zamilon). Estes virófagos não podem se replicar ao entrar em uma ameba que não está infectado com *Acanthamoeba polyphaga mimivirus* (APMV) pois os virófagos necessitam da maquinaria de replicação do APMV para a sua própria replicação, fazendo com que os Mimivirus sejam impactados profunda e negativamente, uma vez que isso causa uma redução da sua taxa de replicação e no aumento da criação de Mimivirus anormais que não conseguem sobreviver (LA SCOLA 2008). Além disso, também foi descoberto um sistema de defesa contra o virófago Zamilon em vírus da linhagem A de APMV, que é muito semelhante ao sistema CRISPR-cas encontrado em bactérias e é algo que era impensável em vírus até muito recentemente (LEVASSEUR *et al.*, 2016).

Considerando que os grandes vírus de DNA citoplasmático (do inglês, *Nucleocytoplasmic large DNA viruses*, NCLDV) podem integrar em genomas eucarióticos, é provável que a assimilação dos genomas ancestrais dos NCLDV também ocorreu nos genomas proto-eucarióticos. Apropriadamente, estabelecida por seu possível envolvimento na origem do núcleo (BELL, 2001), NCLDV pode ter desempenhado um papel na origem dos eucariotos modernos, introduzindo novos genes (e, portanto, possivelmente, novas funções) em proto-eucariotos.

No entanto, no momento da descrição do genoma Mimivírus, os mesmos foram mostrados como um novo ramo na árvore da vida, sendo colocados próximos da origem do domínio eucariótica em uma reconstrução filogenética estabelecida por sequências concatenadas de sete proteínas conservadas (AHERFI *et al.*, 2016).

3.2 - Transferência Lateral de genes e ilhas genômicas

Genomas podem evoluir por um processo chamado transferência horizontal de genes (transferência lateral de genes). Este processo se dá quando um gene de um genoma de uma espécie é incorporado no genoma de outra espécie (RIDLEY, 2006). A transferência lateral de genes ocorre entre os procariotos por três mecanismos principais, conjugação, transformação e transdução.

Conjugação ocorre pelo contato físico de um pili de conjugação de uma célula doadora com uma célula receptora. O DNA de cadeia simples é transportado da célula doadora para a célula receptora. Na transformação, a célula procariótica capta o DNA exógeno do meio. A transdução é a entrada de material genético em procariotos através da captura de fagos, devido a integração do material genético do hospedeiro no genoma do fago. Existem ainda outros mecanismos de transferência lateral de genes. Os agentes de transferência de genes fazem parte de um sistema de entrega de genes por fagos que não reconhecem mais o próprio DNA e passam fragmentos aleatórios para o cromossoma do hospedeiro e são encontrados tanto em bactérias, quanto em arqueias.

A fusão celular é um mecanismo só encontrado em arqueias, onde as células ficam fusionadas através da membrana celular que forma uma ponte onde ocorre uma troca bidirecional de material genético. Em eucariotos, a transferência lateral de genes pode ocorrer por dois mecanismos que são a transferência gênica intracelular ou endossimbionte, que acontece quando um organismo endossimbionte ou organela é incorporada no genoma do hospedeiro. Outro mecanismo que pode ocorrer é através da introgressão em um evento de hibridização entre duas espécies divergentes (SOUCY; HUANG; GOGARTEN, 2015).

Como exemplo de transferência lateral de genes entre vírus e eucariotos temos os retrovírus são singulares entre os vírus animais pois, como passo exigido do seu ciclo de vida, eles se integram nos genomas do hospedeiro como provírus. Caso um evento de integração ocorra em células germinativas, os provírus podem ser passados da geração parental para os descendente e verticalmente herdados dentro da espécie. Oito por cento de todo o genoma humano é derivado de mais de 100.000 fósseis retrovirais, muitos dos quais estão presentes

em todos os primatas símios que compartilharam o último ancestral comum há mais de 30 milhões de anos (PATEL; EMERMAN; MALIK, 2011).

Tais transferências geram ilhas genômicas que são um grupo de genes que se fixam em um genoma de uma espécie e que são advindos de outras espécies. As regiões de ilhas genômicas contêm genes que variam de 10-200 kb de comprimento e os genes são geralmente elementos genéticos móveis como profagos, integrons, transposons e elementos conjugativos integrativos (LANGILLE; HSIAO; BRINKMAN, 2010).

3.3 - Transferência lateral de genes em Mimivirus.

Atualmente, existem duas frentes de discussão sobre a evolução dos mimivirus: uma afirma que Mimivírus e outros NCLDV adquiriram seu genoma gigantesco devido a vários eventos de transferência horizontal de genes (do inglês, Horizontal Gene Transfer, HGT) com células eucarióticas (MOREIRA; BROCHIER-Armanet, 2008;. WILLIAMS et al, 2011). A outra hipótese estabelece que estes vírus gigantes são originados a partir de organismos celulares antigos que passaram por uma redução do genoma e adquiriram um estilo de vida viral (esta ideia baseia-se em genes que são semelhantes aos encontrados nos outros 3 domínios) (Claverie, 2006).

Em análises filogenéticas de replicação dos genes virais, estes geralmente ficam agrupados com mais frequência na base da árvore e longe de sequências celulares relacionadas, sugerindo que os genes de replicação estavam presentes no ancestral NCLDV e evoluíram de forma independente, raramente afetados por transferência lateral de genes. Estudos recentes de filogenômica dos genomas de Mimivirus levantaram novas questões sobre a aquisição maciça de genes de hospedeiros celulares de bactérias e eucariotos.

Filée (2009) e Filée, Pouget e Chandler (2008) sugeriram que vários genes dos Mimivírus foram adquiridos de outros protistas, como o Heterolobosea ou Kinetoplastida. Os genoma dos Mimivirus e dos Phycodnavirus (outro vírus gigante de DNA) apresentam sequências de inserção (do inglês, *insertion sequences*, ISs), tipicamente encontradas em bactérias e Archaea. Esses elementos pertencem às famílias IS200 / IS605 e IS4 e

provavelmente foram adquiridos através de transferência horizontal de genes a partir de fontes procariotas.

A complexidade e o tamanho do genoma dos Mimivírus e *Chlorella PhycoDNVirus* parece ser substancialmente explicado pelo acréscimo de um conjunto diverso de genes bacterianos e elementos genéticos móveis. Apesar de ter genomas maiores, a *Chlorella PhycoDNVirus* e o Mimivírus tem a menor proporção (número comprimento / genoma) de genes de origem potencialmente eucariótica. Por outro lado, se alguns genes foram adquiridos por transferência lateral com o hospedeiro, o gene correspondente aparece polifilético, e como um táxon irmão das sequências hospedeiras. Filogenia dos genes nucleares que estão envolvidas no metabolismo do DNA fornecem evidências de transferências laterais frequentes de diferentes fontes celulares para NCLDV (FILÉE, 2009; FILÉE, POUGET e CHANDLER, 2008).

Em procariotas, os genes adquiridos por transferência lateral são muitas vezes agrupados no cromossoma e fazem ilhas genômicas. Nos vírus foi feita a co-localização dos genes de composição anômala ao longo de cromossomos dos grandes vírus de DNA de fita dupla (do inglês, Large Double Stranded DNA viruses, LDVs) que é o grupo ao qual os NCLDV fazem parte (do inglês, Nucleocytoplasmic Large DNA Viruses, NCLDV) usando uma simulação de Monte-Carlo. Nenhum dos genomas de entomopoxvírus, baculovírus, iridovírus, Mimivírus, nimavírus ou phycoDNVirus exibiu uma tendência significativa para os genes de composição anômala para ser agrupado ao longo do genoma. Esses genes de composição anômala poderiam fornecer aos LDVs a capacidade de prolongar a integridade das células hospedeiras. Entretanto, não houve correlação significativa entre a proporção de genes de composição anômala e o tamanho dos genomas LDV. Esta observação propõe que tamanhos extremamente grandes dos genomas de alguns LDVs (tais como Mimivírus) não são devido à recente aquisição de genes estranhos. Com base em análises de composição de nucleotídeos, reconstruções de árvores filogenéticas e pesquisas de bancos de dados sugerem que a transferência horizontal de genes para estes vírus de seus hospedeiros atuais é pouco frequente e não levam em conta a grande variação no tamanho do genoma (MONIER; CLAVERIE; OGATA, 2007).

Nos Mimivírus, foi sugerido que de um total de mais de 900 genes, mais de 100 genes são adquiridos por transferência lateral vinda de bactérias e organismos próximos a eucariotas. No entanto, o Mimivírus codifica para mais de 700 genes órfãos (são genes sem homologia em outras linhagens) ou genes mal conservados, assim, entre os 200 genes com

homologia nas bases de dados confiáveis, metade parecem resultar de transferências lateral de genes. Os virófagos, que foram recentemente descritos, podem ser um mecanismo alternativo de transferência de genes entre NCLDV's. O virófago possui um genoma mosaico composto de genes de origens distintas (por exemplo, eucariotos e bactérias) (FILÉE; CHANDLER, 2010).

3.4 - Detecção de ilhas genômicas usando *mean-shift*

Os genes de uma determinada espécie são normalmente semelhantes na sua composição de base e os padrões de utilização de códons, de tal maneira que os fragmentos de sequências adquiridas horizontalmente a partir de outros organismos podem ser selecionadas computacionalmente. Os métodos atuais de previsão de ilhas genômicas com base na composição de sequência depende fortemente desse fato. No entanto, isso conduz à não identificação de ilhas genômicas em que os genomas de dois organismos, doadores e receptores têm conteúdos semelhantes ou idênticos de G + C (guanina mais citosina).

Para poder separar elementos de um conjunto de dados, em grupos de objetos de acordo com sua semelhança são utilizado metodos de clusterização (agrupamento) como o método *mean-shift*, ou seja, separar padrões de G + C de um genoma e tentar agrupar esses padrões em duas categorias: uma contendo os padrões do genoma e o outro que esses padrões sejam diferentes do genoma.

O *mean-shift* é um método não paramétrico de agrupamento onde não é preciso definir o número de agrupamentos para a operação, mas necessita de largura de banda como parâmetro a ser definido, que influencia no número de aglomerados formados e afeta a velocidade de convergência do algoritmo. A definição de parâmetro de largura de banda é crítica, influencia fortemente o resultado do algoritmo e utiliza uma heurística própria (BRITO et al., 2016).

4. Metodologia

4.1 - Escolha dos organismos

Para este trabalho foram selecionados 17 genomas de vírus da família *Mimiviridae*: *Acanthamoeba castellanii* mamavirus strain Hal-V (gb/JF801956.1), *Acanthamoeba polyphaga* mimivirus (NC_014649.1), *Cafeteria roenbergensis* virus BV-PW1 (NC_014637.1), *Hirudovirus* strain Sangsue (KF493731.1), *Mimivirus* Bombay (gb/KU761889.1), *Mimivirus* fauteuil , (emb/LN871163.1), *Samba virus* (gb/KF959826.2), *Mimivirus* terra2 (NC_023639.1), *Acanthamoeba polyphaga* moumouvirus (NC_020104.1), *Moumouvirus* goulette (gb/KC008572.1), *Megavirus* chiliensis (NC_016072.1), *Megavirus* courdo11 (gb/JX975216.1), *Megavirus* lba (gb/JX885207.1), *Megavirus* terra1 (NC_023640.1), *Megavirus* avenue9 (LN867403.1), *Megavirus* battle43 (LN868256.1) e *Powai lake megavirus* (gb/KU877344.1). Estes genomas virais foram selecionados pois são os únicos com genoma completo e montado até a data (10/08/2016). Seus dados estão disponíveis no banco de dados online NCBI (<https://www.ncbi.nlm.nih.gov>)

4.2 - Obtenção das proteínas das ilhas genômicas

Para obtenção das proteínas das ilhas genômicas encontradas foi utilizada a ferramenta online Blastx (KORF, 2003) disponível no NCBI (<https://blast.ncbi.nlm.nih.gov/>) .Onde foi colocada uma sequência de nucleotídeos encontrados como ilha genômica nos genomas de *Mimivirus* e foi traduzida em sequências de aminoácidos que geraram algumas proteínas.

4.3 - Obtenção das sequências de proteínas para filogenia

Foram utilizada quinze proteínas de ilhas genômicas de *Mimivirus* sendo elas: Provável proteína colagénio 7 (AGD92194.1), Provável Endonuclease VIII

(YP_004894323.1), Fator de troca de nucleotídeo guanina (YP_004894327.1), Provável proteína Bifuncional Metalloproteaseubiquitina ligase (AGD92181.1), Provável proteína CFxQ (AEX61379.1), Provável proteína chemotaxis CheD (AEX61366.1), Provável DnaK Hsp70 (YP_003970036.1), Provável Glicosiltransferas (YP_004894341.1), Provável proteína Serinethreonine quinase (AEX61407.1), Provável enzima ativadora da ubiquitina E1 (YP_003970068.1), Provável UvrDREP helicase (YP_003970059.1), D5-ATPase-helicase (AEY99293.1), Provável proteína p13 (AGD92165.1), Fator de libertação da cadeia peptídica eRF1 (YP_004894331.1) e RNA Ligase 2 (YP_004894359.1). Foram obtidas a maior quantidade destas sequências pela ferramenta BLAST (MORENO-HAGELSIEB; LATIMER, 2007) de cada um dos domínios da vida e todas as sequências de Mimivírus na plataforma online UniProt (<http://www.uniprot.org/>) e no no banco de dados online NCBI (<https://www.ncbi.nlm.nih.gov>).

4.4 - Detecção de ilhas genômicas

Para detecção das ilhas genômicas nos genomas de mimivirus selecionados, foi utilizado o algoritmo de clusterização *mean-shift* (BRITO et al., 2016) disponível no site <http://integrativebioinformatics.me/msgip/>.

4.5 - Alinhamento e modelo evolutivo

Para o alinhamento das sequências encontradas nas ilhas genômicas foi utilizado o software mafft v.7 (Kato; STANDLEY, 2016) utilizando os parâmetros mostrados no anexo C.. Depois disso, buscamos o melhor modelo evolutivo usando Prottest 3 (Darriba et al., 2011) onde foram encontradas as proteínas provável DnaK Hsp70, Fator de libertação da cadeia peptídica eRF1 e provável enzima ativadora da ubiquitina E1 de acordo com o índice AIC: LG+I+G+F, para a provável proteína colágeno 7 foi índice BIC: RtREV e para as demais proteínas foi encontrado o índice BIC: VT.

4.6-Árvores filogenéticas

As árvores filogenéticas de máxima verossimilhança de cada proteína das ilhas genômicas foram criadas utilizando o software T-REX (*Tree and Reticulogram Reconstructor*) (BOC; DIALLO; MAKARENKOV 2012, MAKARENKOV 2001), que utilizou o modelo de probabilidade máxima com os parâmetros a partir dos modelos gerados pelo Prottest 3 e os seguintes parâmetros presentes no anexo D. Para interpretar estas árvores filogenéticas foi utilizado o software TreeGraph 2 (Stover; MÜLLER, 2010).

4.7 - Comparação de vias metabólicas

Para comparar as vias metabólicas de transcrição foram utilizadas as proteínas dos genoma do *Cafeteria roenbergensis virus BV-PW1*, *Acanthamoeba polyphaga mimivirus*, *Megavirus chiliensis* e *Acanthamoeba polyphaga moumouvirus*. no site do NCBI (<https://www.ncbi.nlm.nih.gov>) e comparadas com as vias metabólicas de eucariotos e procariotos na plataforma online KEGG (<http://www.kegg.jp/>).

5. Resultados

5.1 - Ilhas genômicas

Foram encontradas cinco ilhas genômicas utilizando o algoritmo de *mean-shift*. Para o genoma do *Cafeteria roenbergensis virus BV-PW1* foi encontrada uma provável região de ilha da posição 0,450 - 0,500Mb do genoma. Para o genoma de *Megavirus chiliensis* foi encontrada a provável região de ilha na posição 0,250 - 0,300Mb. Já no genoma de *Megavirus courdo11* foi encontrado a provável região na posição 0,250 - 0,300Mb. Em *Megavirus avenue 9* foi encontrada uma provável ilha na região 0,950 - 1,000Mb e para o genoma do *Megavirus battle 43* foi encontrada uma região de provável ilha na posição 0,950 - 1,000Mb. Utilizando a ferramenta online BLASTX (disponível em: <https://blast.ncbi.nlm.nih.gov/>) que traduz os nucleotídeos em proteínas, foram obtidas as proteínas que segue na Tabela 1.

Proteína \ Organismo	<i>Cafeteria roenbergensis</i> virus BV-PWI	<i>Megavirus chiliensis</i>	<i>Megavirus courdo11</i>	<i>Megavirus avenue9</i>	<i>Megavirus battle43</i>
Provável proteína colagénio 7		X	X	X	X
D5-ATPase-helicase			X		
Provável Endonuclease VIII				X	X
Fator de troca de nucleotídeo guanina		X	X	X	X
Provável proteína p13					X
Fator de libertação da cadeia peptídica eRF1		X			X
Provável proteína CFxQ		X	X	X	X
Provável proteína chemotaxis CheD					X
Provável DnaK Hsp70	X				
Provável Glicosiltransferase				X	X
Provável proteína Serinethreonine quinase		X	X		
Provável enzima ativadora da ubiquitina E1	X				
Provável proteína Bifuncional Metalloproteaseubiquitin ligase		X		X	X
Provável UvrDREP helicase	X				
RNA Ligase 2			X		

Tabela 1. Correspondência entre as proteínas de ilhas genômicas e seus respectivos organismos

Também foram encontradas proteínas hipotéticas sendo seis para o, *Cafeteria roenbergensis* virus BV-PWI que são hypothetical protein crov406(YP_003970039.1); hypothetical protein crov409(YP_003970042.1), hypothetical protein crov415 (YP_003970048.1), hypothetical protein crov424 (YP_003970057.1), hypothetical protein crov427 (YP_003970060.1), hypothetical protein crov440 (YP_003970073.1); para o, *Megavirus chiliensis* foram cinco, sendo elas: hypothetical protein MegaChil _gp0271 (YP_004894322.1), hypothetical protein MegaChil _gp0264 (YP_004894315.1), hypothetical protein MegaChil _gp0279 (YP_004894330.1), hypothetical protein MegaChil _gp0282 (YP_004894333.1), hypothetical protein MegaChil _gp0301(YP_004894352.1); para *Megavirus courdo11* foram oito, sendo as seguintes : hypothetical protein (ANB50420.1) , hypothetical protein c7_R346 (AEX61412.1), hypothetical protein CE11_00289 (AFX92319.1), hypothetical protein CE11_00307

(AFX92337.1), hypothetical protein CE11_00314 (AFX92344.1), hypothetical protein CE11_00319 (AFX92349.1), hypothetical protein MegaChil _gp0301 (YP_004894352.1), hypothetical protein MegaChil _gp0302 (YP_004894353.1); para o *Megavirus avenue9* foram seis sendo elas : hypothetical protein CE11_00289 (AFX92319.1), hypothetical protein MegaChil _gp0254 (YP_004894305.1), hypothetical protein MegaChil _gp0256 (YP_004894307.1), hypothetical protein MegaChil _gp0262 (YP_004894313.1), hypothetical protein MegaChil _gp0264 (YP_004894315.1), hypothetical protein MegaChil _gp0283 (YP_004894334.1); para *Megavirus battle43* foram cinco sendo elas : hypothetical protein c7_L289 (AEX61355.1), hypothetical protein CE11_00289 (AFX92319.1), hypothetical protein LBA_00249 (AGD92169.1), hypothetical protein LBA_00251 (AGD92171.1), hypothetical protein MegaChil _gp0263 (YP_004894314.1). Os demais Mimivirus não tiveram ilhas genômicas.

5.2 - Filogenias das proteínas das ilhas

Foram construídas árvores filogenéticas utilizando as proteínas encontradas nas ilhas genômicas dos genomas dos Mimivirus e proteínas homólogas encontradas nos três domínios da vida e em outros Mimivirus para inferir a provável origem de cada uma das proteínas das ilhas, analisando onde cada proteína se agrupa na árvore filogenética. Se de fato forem ilhas genômicas essas proteínas devem se agrupar com as proteínas homólogas, em um dos três domínios da vida mostrando que houve transferência horizontal de genes entre um determinado domínio da vida e os Mimivirus.

5.3 - Fator de libertação da cadeia peptídica eRF1

O Fator de libertação da cadeia peptídica eRF1 é uma proteína que participa da maquinaria de tradução dos Mimivirus. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 1, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

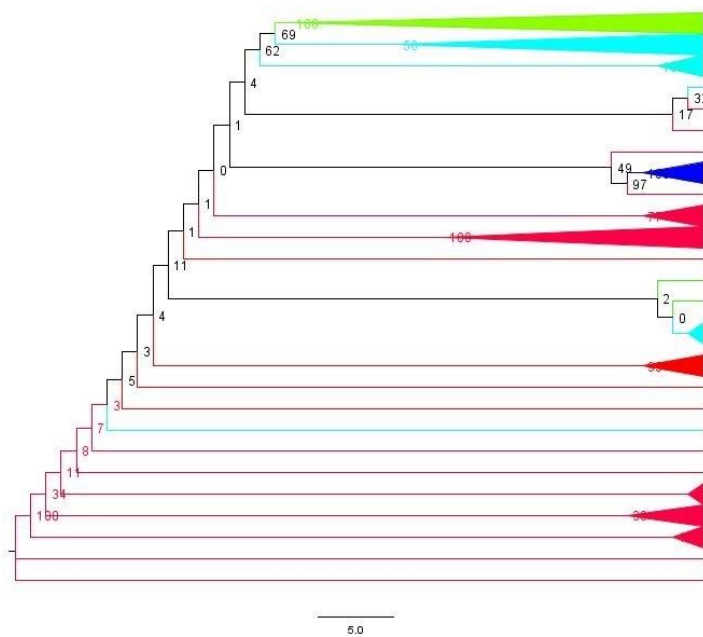


Figura 1. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.4 - Provável Endonuclease VIII

A provável Endonuclease VIII atua nos processos biológicos de reparo de base-excisão, hidrólise da ligação fosfodiéster do ácido nucleico e reparo de excisão de nucleotídeo. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 2, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

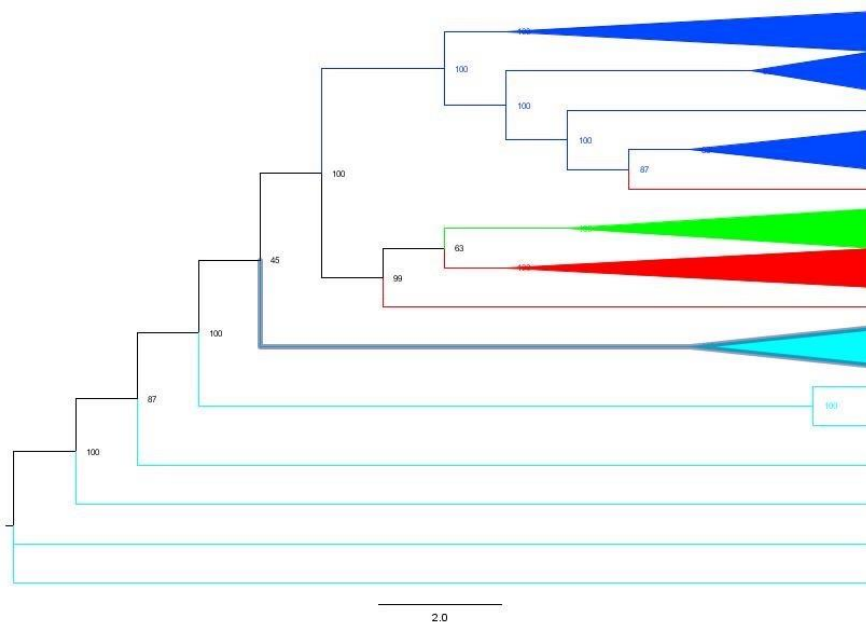


Figura 2. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.5 - D5-ATPase-Helicase

A D5-ATPase-Helicase tem função de ligação de nucleotídeos, atividade de helicase, hidrolase, atuando sobre ácido anidrido e ligação de ATP. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 3, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

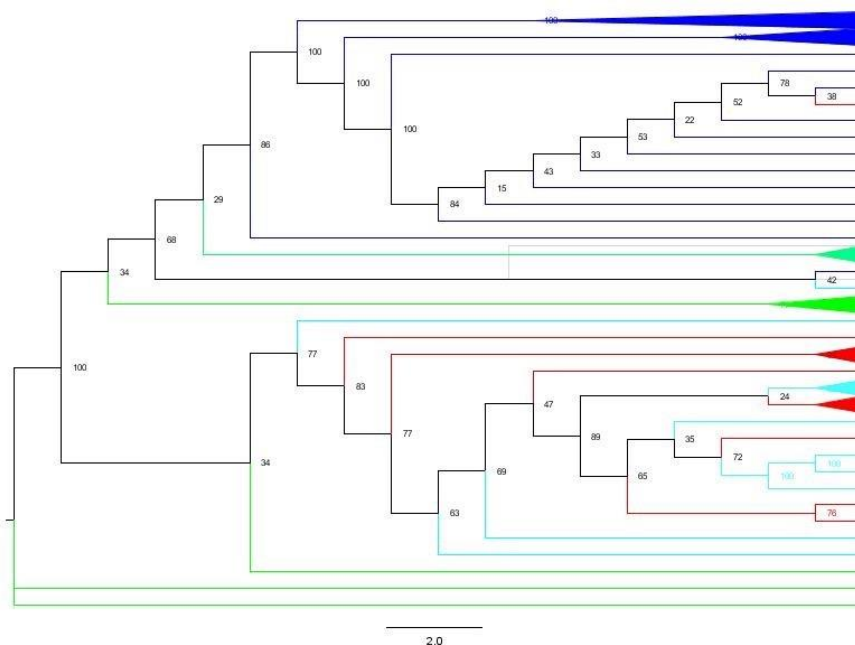


Figura 3. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.6 - Provável proteína CFxQ

A provável proteína CFxQ é uma ligante de ATP. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 4, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

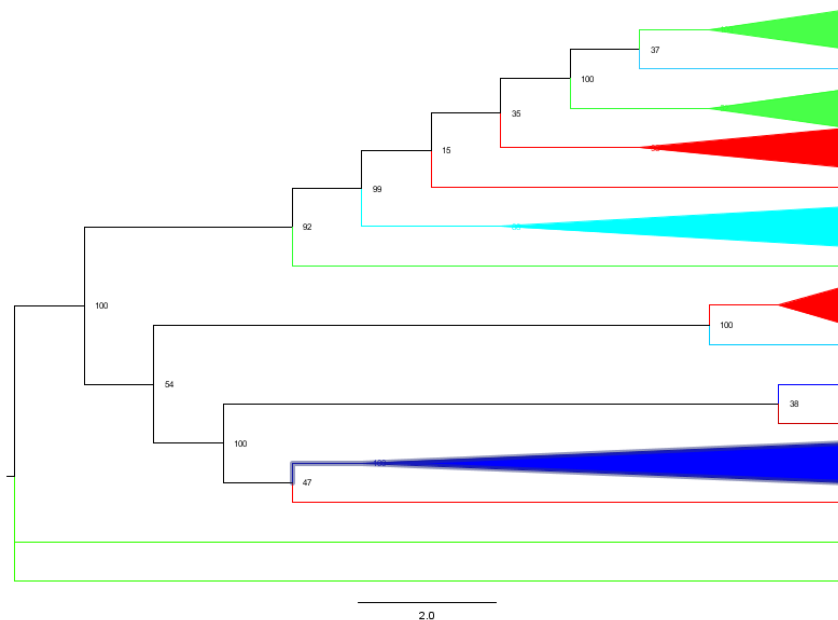


Figura 4. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.7 - RNA Ligase 2

A RNA Ligase 2 tem função de catálise da união de um ribonucleotídeo, com a hidrólise concomitante da ligação difosfato em ATP ou um trifosfato semelhante. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 5, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

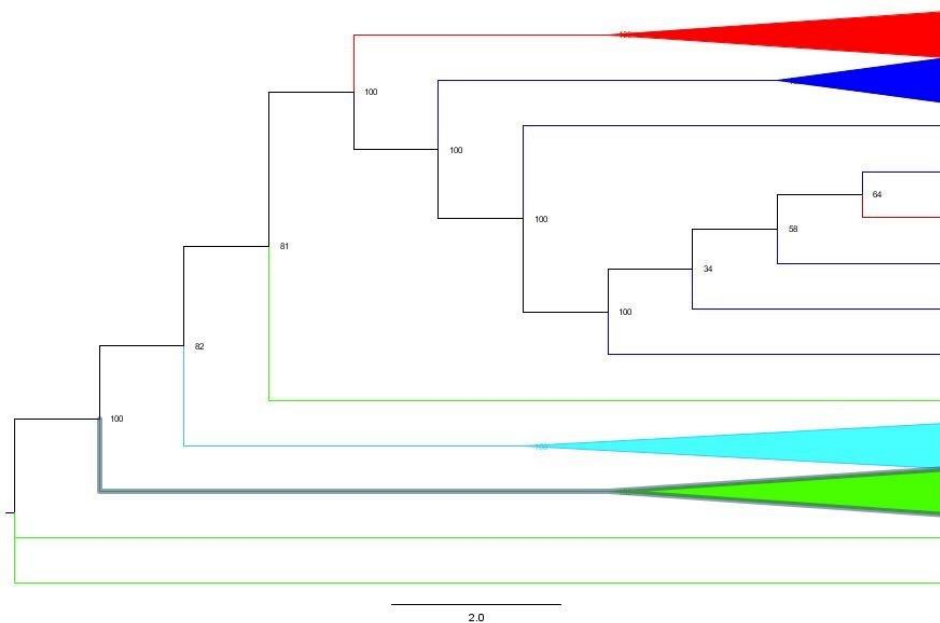


Figura 5. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.8 - Provável proteína p13

A provável proteína P13-like tem função de transferase, transferindo grupos glicosil. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 6, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

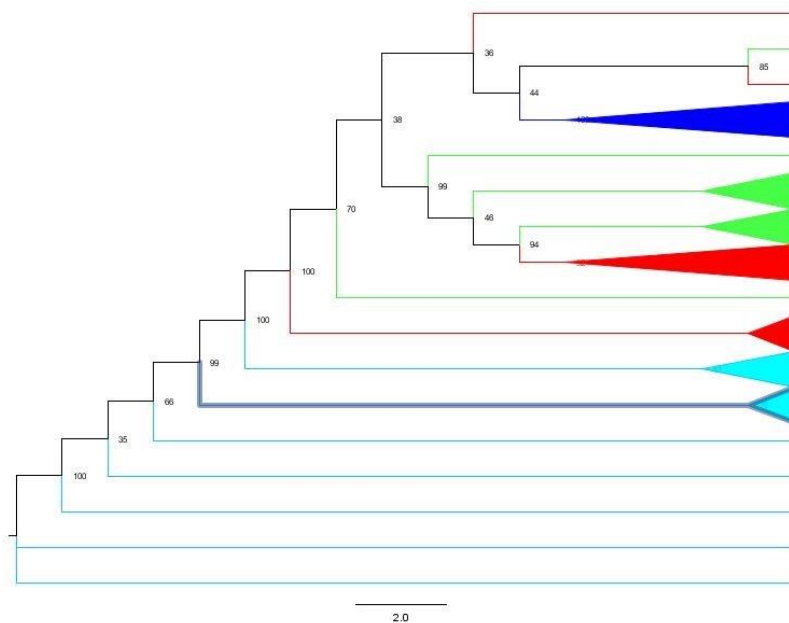


Figura 6. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea

5.9 - Provável proteína colagénio 7

A provável proteína colagénio 7 pode participar na formação de uma camada de fibrilas reticuladas glicosiladas na superfície viral, dando-lhe assim uma aparência “peluda”. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 7, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

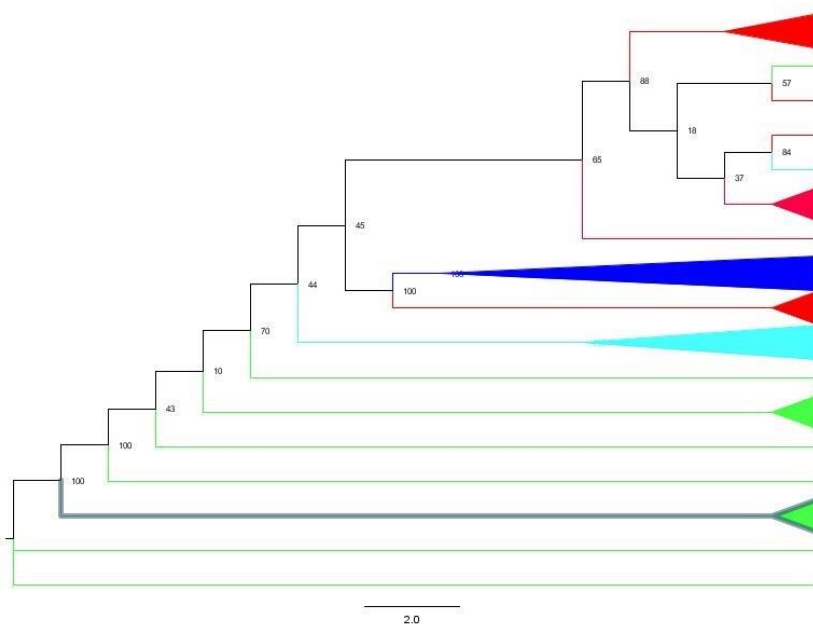


Figura 7. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.10 - Provável DnaK Hsp70

A provável DnaK Hsp70 é uma proteína de choque térmico que participa da dobra de proteínas. A maioria das proteína homólogas em Mimivirus foi agrupada em um clado com a exceção de uma que ficou agrupado como grupo irmão das proteínas homólogas encontradas em Archaea e Bacteria como pode ser visto na Figura 8, indicando que pode ter ocorrido uma transferência lateral de genes entre um desses dois domínios e os Mimivirus.

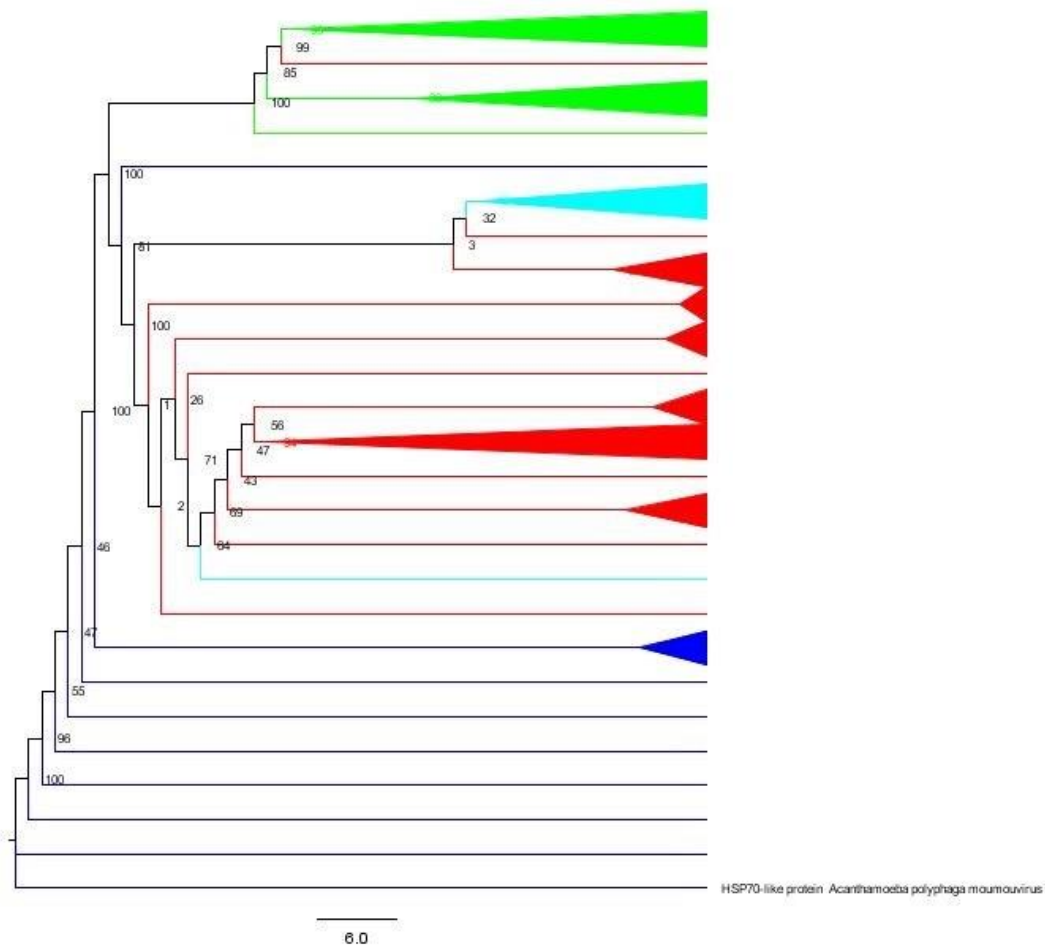


Figura 8 Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.11 - Provável proteína chemotaxis CheD

A provável proteína chemotaxis CheD é um transdutor de sinal de quimiotaxia. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 9, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

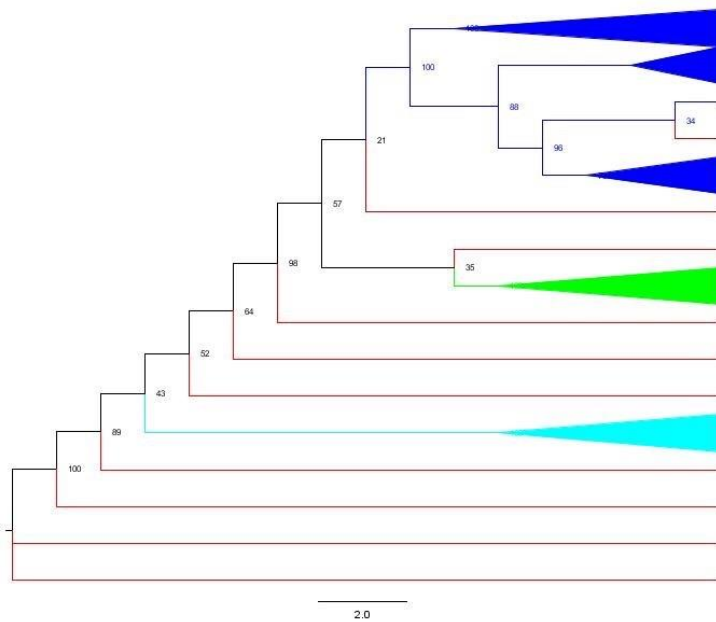


Figura 9. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.12 - Provável Glicosiltransferase

A provável Glicosiltransferase é uma proteína de transferência de grupo glicosil. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 10, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

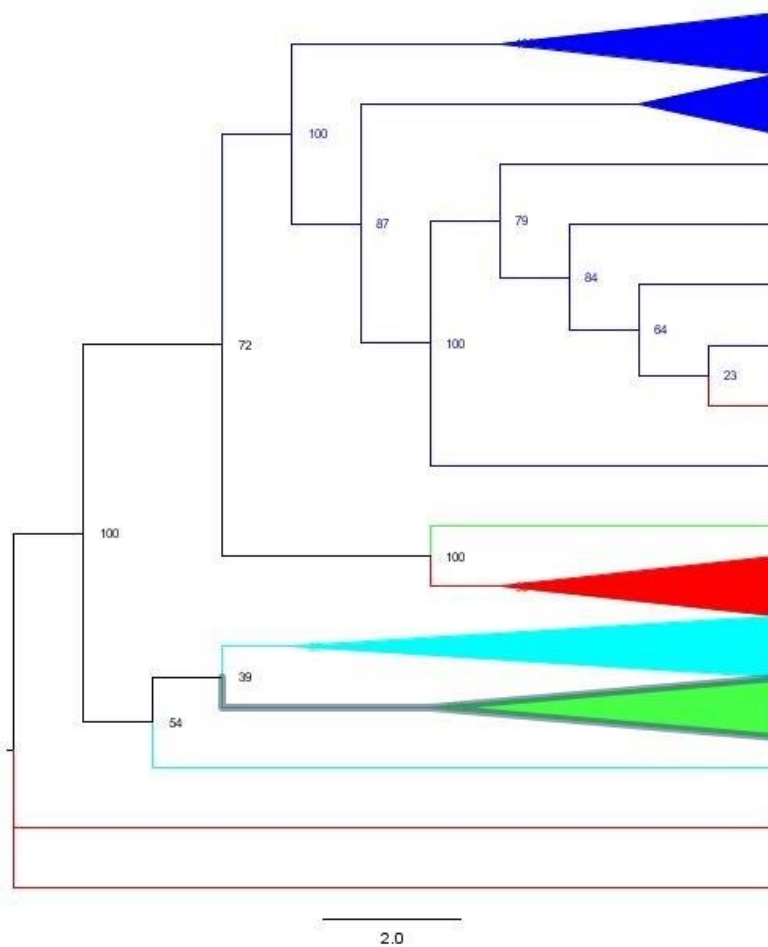


Figura 10. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.13 - Fator de Troca de nucleotídeo Guanina

O fator de troca de nucleotídeo guanina como o próprio nome sugere tem atividade de troca de nucleotídeos guanina. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 11, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

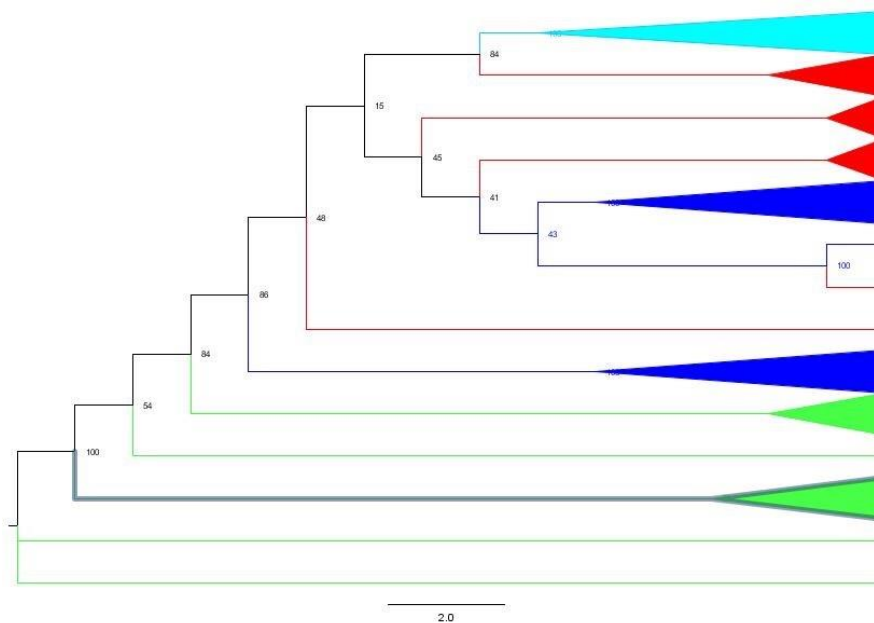


Figura 11. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.14 - Provável proteína Bifuncional Metalloproteaseubiquitina ligase

A provável proteína Bifuncional Metalloproteaseubiquitina ligase tem atividade de metalopeptidase, proteólise e de ligase. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 12, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

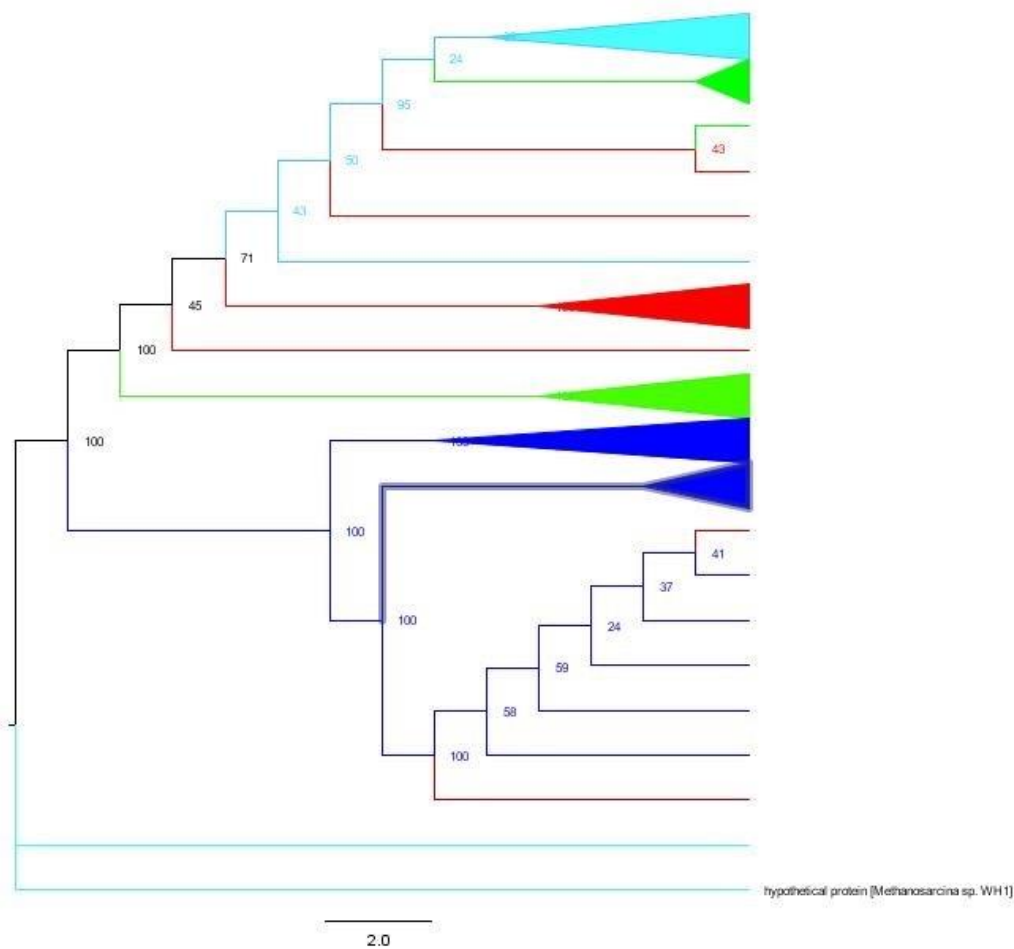


Figura 12. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.15 - Provável proteína Serinethreonine quinase

A provável proteína Serinethreonine quinase tem função de fosforilação proteica e de transdução de sinal intracelular. A maioria das proteína homólogas em Mimivirus foi agrupada em um clado com a exceção de uma que ficou agrupado como grupo próximo das proteínas homólogas encontradas em Eukarya como pode ser visto na Figura 13, indicando que pode ter ocorrido uma transferência lateral de genes entre um desses dois domínios e os Mimivirus.

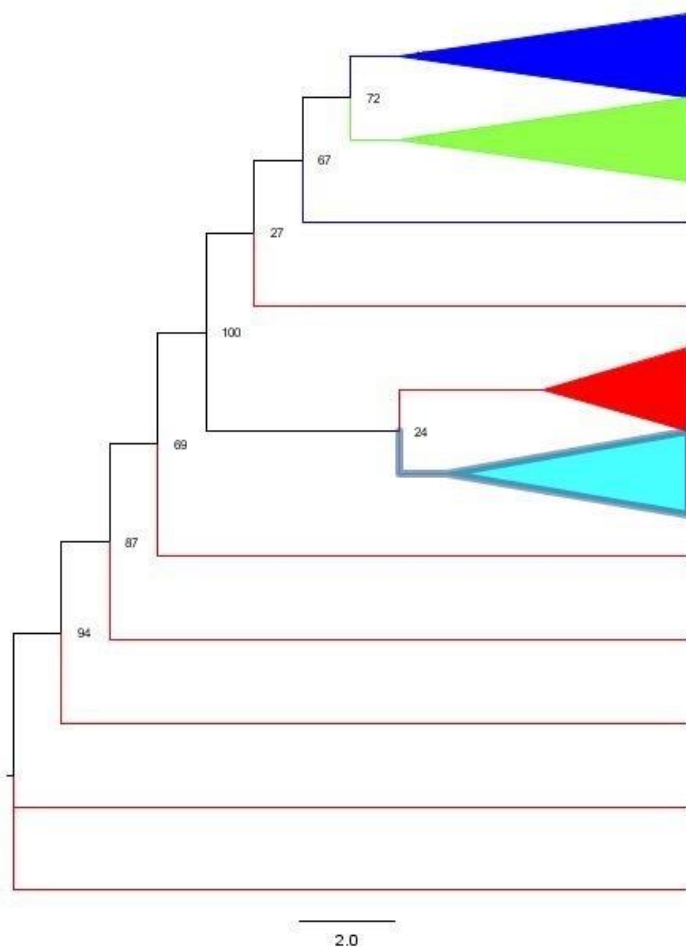


Figura 13. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.16 - Provável UvrDREP helicase

A Provável UvrDREP helicase interage seletivamente e não covalentemente com ATP, adenosina 5'-trifosfato, uma coenzima universalmente importante e regulador enzimático. A maioria das proteínas homólogas em Mimivirus foi agrupada em um clado com a exceção de uma que ficou agrupado como grupo próximo das proteínas homólogas encontradas em Bacteria como pode ser visto na Figura 14, indicando que pode ter ocorrido uma transferência lateral de genes entre um desses dois domínios e os Mimivirus.

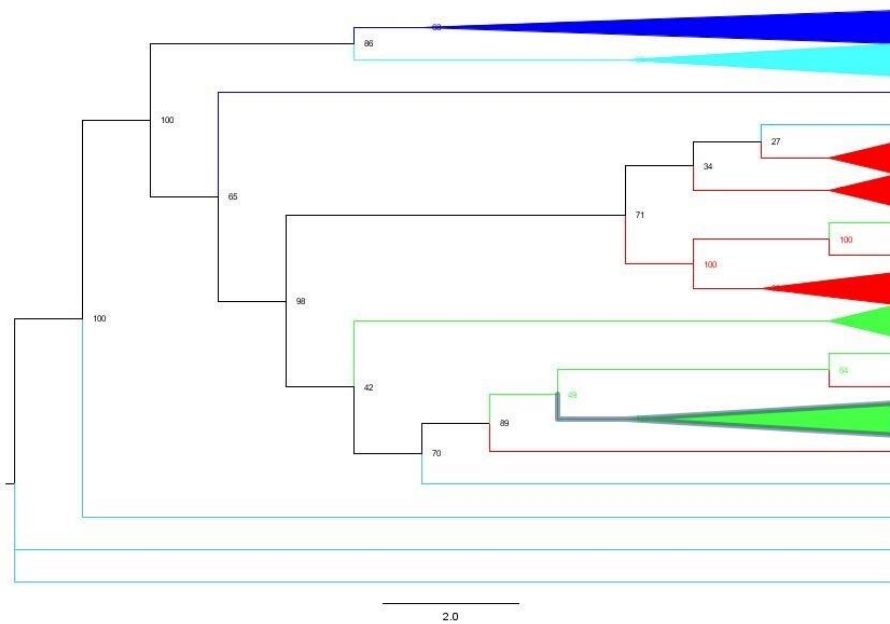


Figura 14. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

5.17 - Provável enzima ativadora da ubiquitina E1

A Provável enzima ativadora da ubiquitina E1 tem função de catálise da ativação de proteínas pequenas, como a ubiquitina ou proteínas semelhantes à ubiquitina. Esta proteína foi agrupada em sua filogenia com as proteínas homólogas em um clado formado pelos Mimivirus, como pode-se observar em azul na Figura 15, mostrando que provavelmente essa proteína não faz parte de uma ilha genômica visto que nenhuma proteína do grupo dos Mimivirus se agrupou com eucariotos ou bactérias ou arqueias.

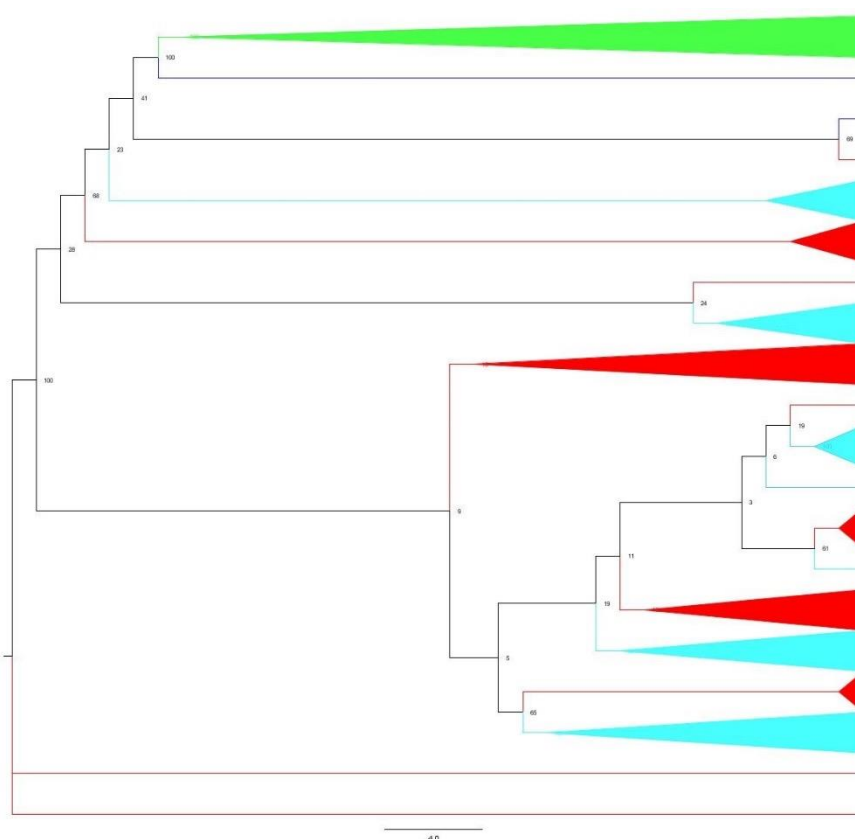


Figura 15. Em vermelho estão as proteínas agrupadas no domínio Bacteria, em azul escuro os Mimivirus, em verde os Eukarya e azul claro os Archaea.

Os resultados das filogenias mostram que quando ocorre um evento de ilha as proteínas não se agrupam todos no mesmo grupo pois essas proteínas teoricamente vem de outros organismos por meio de transferência lateral gênica, que foi o que aconteceu com as filogenias das proteínas Putativa Serinethreonine-proteína quinase, Putativa DnaKHsp70 e Putativa UvrDREP helicase. As demais filogenias das proteínas indicadas como sendo filogenias pertencente a uma ilha pelo *mean-shift* ficaram todas juntas mostrando que não

aconteceu um evento de transferência lateral de genes podendo indicar uma falha na metodologia do *mean-shift* que gerou falsos positivos ou que as ilhas genômicas das demais proteínas só ocorram entre os próprios Mimivírus.

5.18 - Fatores de transcrição basal

Utilizando as proteínas da maquinaria basal de transcrição foi feita a Tabela 2 que apresenta a comparação das proteínas de transcrição do *Acanthamoeba polyphaga mouloumovirus*, *Megavirus chilensis*, *Acanthamoeba polyphaga mimivirus* e *Cafeteria roenbergensis virus BV-PWI* disponíveis no NCBI (<https://www.ncbi.nlm.nih.gov/>) com as proteínas da maquinaria de transcrição dos três domínios da vida, disponível no KEGG (http://www.genome.jp/kegg-bin/show_pathway?map=ko03020&show_description=show e http://www.genome.jp/kegg-bin/show_pathway?map=ko03022&show_description=show).

E B A Proteína

Acanthamoeba polyphaga mouloumovirus

- x x x RNA- Polimerase dirigida por DNA subunidade B
- x x x RNA- Polimerase dirigida por DNA subunidade N
- x x x Fragmento RNA- Polimerase dirigida por DNA subunidade 1
- x x x Intein-contendo RNA- Polimerase dirigida por DNA subunidade 2
- x x x Putativa RNA- Polimerase dirigida por DNA subunidade
- x Fator de iniciação da transcrição TFIIB

Megavirus chilensis

- x x x RNA- Polimerase dirigida por DNA subunidade 1
- x x x RNA- Polimerase dirigida por DNA subunidade 5
- x x x RNA- Polimerase dirigida por DNA subunidade 6
- x x x Intein-contendo RNA- Polimerase dirigida por DNA subunidade 2
- x x x Putativa RNA- Polimerase II dirigida por DNA subunidade n
- x x x Putativa RNA- Polimerase dirigida por DNA subunidade
- x x x Putativa RNA- Polimerase dirigida por DNA subunidade 1 inactive-like protein

x		Putativo Fator de iniciação da transcrição iib-like protein
<i>Acanthamoeba polyphaga mimivirus</i>		
x	x	x RNA- Polimerase dirigida por DNA subunidade 1
x	x	x RNA- Polimerase dirigida por DNA subunidade 2
x	x	x RNA- Polimerase dirigida por DNA subunidade 5
x	x	x RNA- Polimerase dirigida por DNA subunidade 6
x	x	x Putativa DNA-directed RNA polymerase II subunidade N
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade 1 inactive-like protein
x		Putativo Fator de iniciação da transcrição Iib-like protein
<i>Cafeteria roenbergensis virus BV-PW1</i>		
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade Rpb1
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade Rpb10
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade Rpb2 precursor
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade Rpb5
x	x	x Putativa RNA- Polimerase dirigida por DNA subunidade Rpb7
x		Putativo Fator de iniciação da transcrição TFIIB

Tabela 2. E = Archaea ; B = Bacteria ; E = Eukarya

Nesta tabela podemos ver que os Mimivirus compartilham de uma maquinaria de transcrição basal comum com os três domínios da vida. Todos os Mimivirus possuem RNA-Polimerase dirigida por DNA assim com todos os eucariotos e procariotos, entretanto, vale ressaltar que eles possuem fator de iniciação da transcrição que é presente só nos eucariotos (Maiores detalhes ver tabela em anexo B).

6. Discussão

Diferentemente dos resultados encontrados por Filée (2009) ,Filée, Pouget e Chandler (2008) e Filée (2015) que encontraram transferência lateral de genes da família de sequência de inserção, genes órfãos, genes com funções desconhecidas e DNA metiltransferase, nesse trabalho foram encontradas ilhas genômicas que codificam algumas proteínas de envelhecimento, proteínas de choque térmico, proteínas de capsídeo e proteínas de tradução. Esta diferença de ilhas encontradas provavelmente se deve ao fato da utilização de metodologias diferentes, enquanto o presente trabalho usa a nova metodologia de *mean-shift* (BRITO et al., 2016), os outros trabalhos utilizaram BLAST, distância entre assinaturas genômicas, análise de sintenia e alinhamento múltiplo dos genomas para encontrar transferência horizontal, posteriormente, confirmar por meio de filogenias (FILÉE 2015; FILÉE, 2009 ; FILÉE; POUGET; CHANDLER, 2008; MONIER; CLAVERIE; OGATA, 2007).

As filogenias foram usadas para confirmar se existe realmente a ocorrência de ilhas genômicas nos genomas de Mimivirus, entretanto a maioria das proteínas que foram encontradas como sendo pertencentes as ilhas ficaram agrupadas juntas no táxon dos Mimivirus, com a exceção das ilhas de Putativa Serinethreonine-proteína quinase, Putativa DnaKHsp70 e Putativa UvrDREP helicase pertencentes ao vírus *Cafeteria roenbergensis virus BV-PWI*. O fato de ter se confirmado poucas ilhas pode indicar duas possibilidades, que existe uma transferência lateral de genes, porém só entre os Mimivirus, ou que o método encontrou falsos positivos e não existe transferências laterais nos Mimivirus, que pode ser corroborado pelo fato das ilhas encontradas que pertencem ao *Cafeteria roenbergensis virus BV-PWI* se agruparem em um clado diferente, pois apesar desse vírus se agrupar filogeneticamente com os demais NCLDV e ficar mais próximo dos Mímivirus, os genes do *Cafeteria roenbergensis virus BV-PWI* possui menos de um terço de homologia com os genes dos Mimmivirus (FISCHER et al., 2010). Sendo assim, existe a possibilidade do *Cafeteria roenbergensis virus BV-PWI* não pertencer a família Mimiviridae e portanto, nas filogenias de proteínas de prováveis ilhas genômicas, eles não se agrupam em um mesmo clado com os demais Mimivirus.

Levando em consideração que existam ilhas genômicas nos Mimivirus, estas ilhas são poucas e os eventos de transferência horizontal de genes são raros e recentes, sendo assim,

como afirma Monier, Claverie e Ogata (2007,) isso não contribuiu para o aumento no tamanho dos genomas desses vírus corroborando com a hipótese de degradação do genoma de Claverie (2006), onde os grande vírus derivam de um antepassado celular pois, diferente dos demais vírus, eles tem uma maquinaria complexa com fatores de transcrição, proteínas de tradução e foi perdendo seu genoma através de cada linhagem, diferentemente da hipótese de aquisição de genes onde os vírus seriam “pickpockets” que teriam herdado seus genes de seus hospedeiros celulares (FORTERRE, 2016).

Os dados desse trabalho também se contradizem com a hipótese do escape antigo, onde os primeiros vírus teriam surgido no período do mundo RNA onde as primeiras células e proteínas baseadas em RNA abrigavam diferentes tipos de replicons incluindo os parasitas. Como nesse período a maquinaria de divisão celular acoplados com a replicação do RNA não eram sofisticados, haveria uma forte pressão seletiva em favor dos replicadores que burlavam o processo de divisão celular usando proteínas de capsídeo primitivo como transporte para penetrar e replicar em novas células, que seriam os primeiros vírions (genoma viral com capsídeo) a se originar e que, ao passar do tempo, algumas linhagens de virions de DNA foram adquirindo maior complexidade estrutural e maior tamanho (FORTERRE, 2016).

,Para Forterre (2016), a hipótese do escape antigo é a mais congruente de como os vírus teriam se originado, e segundo ele, a hipótese de regressão parece improvável pois é difícil compreender de como uma célula, mesmo que primitiva, teria se transformado em um vírion principalmente os virions minimalistas. Os defensores da hipótese de regressão afirmam que esta hipótese se aplica somente aos NCLDVs, no entanto, Forterre (2016) afirma que existe um gradiente na virosfera (diversidade de todos os vírus em todos os possíveis habitats do planeta) do tamanho do genoma que vai do pequeno até o maior vírus e qualquer divisão na virosfera com base no tamanho do genoma é arbitrária.

De acordo com a hipótese levantada por Forterre (2016), vírus seria um grupo monofilético onde todos teriam surgido do escape de células primitivas do mundo RNA, porém Nasir e Caetano-anolles (2015) fizeram uma filogenia para os vírus com base nas dobras de superfamília de proteínas e seus dados suportam uma origem polifilética para os vírus e que uma origem monofilética é fraca, mas poderia ser explicada pela hipótese da evolução redutiva que sugere a perda da maquinaria do ribossomo primitivo em proto-virocélulas (ancestrais virais com natureza celular) e subsequente redução em vírus e pode existir um ou mais ancestrais para os vírus modernos. Contudo, Claverie e Abergel (2016)

dizem que vírus não sofrem evolução redutiva pois organismos que sofreram redução evolutiva, como bactérias parasitas intracelulares, tem ausência de vários componentes centrais na maquinaria de tradução, incluindo aminoacil-tRNA ligase e proteínas ribossomais, e os Mimivirus possuem em seu genoma aminoacil-tRNA ligase, logo, eles não poderiam estar sofrendo uma redução evolutiva, mas sim, uma ganho de genes como na hipótese “pick-pockets” porém, mais extrema, onde todos os vírus de DNA fita dupla são considerados derivados do mesmo elemento móvel ancestral da qual são derivados os Polintons que possuem capacidade de codificar proteínas de cápsideo de vírus, o que sugere que esses elementos móveis podem formar viriões, sob algumas condições.

Considerando a origem polifilética dos vírus, Filée (2015) levanta uma nova hipótese de evolução dos NCLDV's com base no seus dados usando análise de sintenia e alinhamento múltiplo dos genomas, chamada : “genomic accordion” onde os vírus gigantes evoluíram por um processo equilibrado de ganhos de genes (duplicação principalmente mas, também transferência horizontal de genes) compensados por perdas de genes e ainda afirma que a taxa de expansão do genoma tem desacelerado antes da diversificação dos Mimiviridae e sendo assim, os antepassados dessas famílias virais já eram vírus gigantes, indicando que o gigantismo do genoma emergiu precocemente durante a evolução dos Megavirales (FILÉE 2015). Entretanto, FILÉE (2015) afirma que a hipótese de degradação do genoma de Claverie (2006) só é suportada pois os vírus gigantes possuem genes da maquinaria de tradução mas que esses genes teriam vindo por transferência lateral de organismos celulares.

Os dados deste trabalho mostram que com a exceção do fator de liberação da cadeia peptídica *erf1* que foi encontrado como ilha genômica porém não foi confirmado por filogenia, não existem ilhas genômicas de maquinaria de tradução, ou seja, esses genes não vieram por meio de transferência lateral de genes. Além disso podemos fazer uma comparação com a Tabela 2 de maquinaria de transcrição, onde os Mimivirus compartilham o fator de iniciação da transcrição em conjunto com os eucariotos mostrando que eles estão filogeneticamente próximos corroborando com a hipótese de degradação do genoma, onde provavelmente a origem do genoma desses NCLDV é derivada de um antepassado celular que pode ser da mesma linhagem que deu origem aos eucariotos, entretanto, a linhagem dos Mimivirus sofreu uma pressão seletiva devido à adaptação ao modo de vida parasita, sendo assim, reduziu seu genoma.

Apesar de terem sido encontradas e confirmadas poucas ilhas genômicas, não podemos descartar a hipótese do “genomic accordion” pois o principal mecanismo de aumento de genoma pode ser por duplicação dos genes e compensado por perda de genes, sendo, uma hipótese não é completamente excludente a outra, logo, os vírus gigantes podem ter surgido como organismos celulares da linhagem dos protoeucariotos que tinham hábitos parasitas e por isso precisou degradar parte do seu genoma mas, com capacidade de duplicação dos genes do seu genoma ao longo do tempo, eles não ficaram diminutos como os demais vírus.

7. Considerações finais

Este trabalho contribui com mais dados para a hipótese de degradação do genoma, mostrando que a hipótese ainda não foi refutada, dando indícios que a maquinaria de tradução dos Mimivirus não vieram por transferência lateral de genes, sendo assim estes vírus tem uma origem celular e estão em um processo de perde e ganho de genes por duplicação do seus genes como na hipótese do “genomic accordion”. Entretanto, é necessário comparar os resultados deste trabalho onde foram encontradas ilhas genômicas com o método *mean-shift* com outros métodos de procura de ilhas genômicas.

Referências

AHERFI, Sarah et al. Giant Viruses of Amoebas: An Update. **Front. Microbiol.**, [s.l.], v. 7, n. 349, p.1-14, 22 mar. 2016. Frontiers Media SA. <http://dx.doi.org/10.3389/fmicb.2016.00349>.

BELL, Philip John Livingstone. Viral Eukaryogenesis: Was the Ancestor of the Nucleus a Complex DNA Virus?. **Journal Of Molecular Evolution**, [s.l.], v. 53, n. 3, p.251-256, 1 set. 2001. Springer Science + Business Media. <http://dx.doi.org/10.1007/s002390010215>.

BOC, A.; DIALLO, A. B.; MAKARENKOV, V.. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. **Nucleic Acids Research**, [s.l.], v. 40, n. 1, p.573-579, 6 jun. 2012. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gks485>.

BRITO, Daniel M. de et al. A Novel Method to Predict Genomic Islands Based on Mean Shift Clustering Algorithm. **Plos One**, [s.l.], v. 11, n. 1, p.1-12, 5 jan. 2016. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0146352>.

CLAVERIE, Jean-Michel. Viruses take center stage in cellular evolution. **Genome biology**, v. 7, n. 6, p. 1, 2006.

CLAVERIE, Jean-michel; ABERGEL, Chantal. Giant viruses: The difficult breaking of multiple epistemological barriers. **Studies In History And Philosophy Of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences**, [s.l.], v. 59, p.89-99, out. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.shpsc.2016.02.015>.

DARRIBA, D. et al. ProtTest 3: fast selection of best-fit models of protein evolution. **Bioinformatics**, [s.l.], v. 27, n. 8, p.1164-1165, 17 fev. 2011. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btr088>.

FILÉE, Jonathan; POUGET, Noëlle; CHANDLER, Mick. Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. **Bmc Evol Biol**, [s.l.], v. 8, n. 1, p.320-333, 2008. Springer Nature. <http://dx.doi.org/10.1186/1471-2148-8-320>.

FILÉE, Jonathan. Lateral gene transfer, lineage-specific gene expansion and the evolution of Nucleo Cytoplasmic Large DNA viruses. **Journal Of Invertebrate Pathology**, [s.l.], v. 101, n. 3, p.169-171, jul. 2009. Elsevier BV. <http://dx.doi.org/10.1016/j.jip.2009.03.010>.

FILÉE, Jonathan; CHANDLER, Michael. Gene Exchange and the Origin of Giant Viruses. **Intervirology**, [s.l.], v. 53, n. 5, p.354-361, 2010. S. Karger AG. <http://dx.doi.org/10.1159/000312920>.

FILÉE, Jonathan. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. **Front. Microbiol.**, [s.l.], v. 6, p.1-13, 16 jun. 2015. Frontiers Media SA. <http://dx.doi.org/10.3389/fmicb.2015.00593>.

FISCHER, M. G. et al. Giant virus with a remarkable complement of genes infects marine zooplankton. **Proceedings Of The National Academy Of Sciences**, [s.l.], v. 107, n. 45, p.19508-19513, 25 out. 2010. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.1007615107>.

FORTERRE, Patrick. To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. **Studies In History And Philosophy Of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences**, [s.l.], v. 59, p.100-108, out. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.shpsc.2016.02.013>.

KATOH, Kazutaka; STANDLEY, Daron M.. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. **Bioinformatics**, [s.l.], p.1-10, 26 fev. 2016. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btw108>.

KORF, I. Serial BLAST searching. **Bioinformatics**, [s.l.], v. 19, n. 12, p.1492-1496, 11 ago. 2003. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btg199>.

LA SCOLA, B. *et al.* (2008). The virophage as a unique parasite of the giant mimivirus. **Nature** 455, 100–104. doi: 10.1038/nature07218.

LANGILLE, Morgan G. I.; HSIAO, William W. L.; BRINKMAN, Fiona S. L.. Detecting genomic islands using bioinformatics approaches. **Nature Reviews Microbiology**, [s.l.], v. 8, n. 5, p.373-382, maio 2010. Springer Nature. <http://dx.doi.org/10.1038/nrmicro2350>.

LEGENDRE, M. et al. MRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. **Genome Research**, [s.l.], v. 20, n. 5, p.664-674, 1 abr. 2010. Cold Spring Harbor Laboratory Press. <http://dx.doi.org/10.1101/gr.102582.109>.

LEVASSEUR, Anthony et al. MIMIVIRE is a defense system in mimivirus that confers resistance to virophage. **Nature**, [s.l.], v. 531, n. 7593, p.249-252, 29 fev. 2016. **Nature Publishing Group**. <http://dx.doi.org/10.1038/nature17146>. Disponível em: <<http://www.nature.com/nature/journal/v531/n7593/full/nature17146.html>>. Acesso em: 18 maio 2016.

MAKARENKOV, Vladimir. T-REX: reconstructing and Visualizing Phylogenetic Trees and Reticulation Networks. **Bioinformatics**, [s.i.], v. 17, n. 72001, p.664-668, 03 abr. 2001.

MONIER, Adam; CLAVERIE, Jean-michel; OGATA, Hiroyuki. Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. **Bmc Genomics**, [s.l.], v. 8, n. 1, p.456-473, 2007. Springer Nature. <http://dx.doi.org/10.1186/1471-2164-8-456>.

MOREIRA, D.; BROCHIER-ARMANET, C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. **BMC evolutionary biology**, v. 8, p. 12, 2008.

MORENO-HAGELSIEB, G.; LATIMER, K.. Choosing BLAST options for better detection of orthologs as reciprocal best hits. **Bioinformatics**, [s.l.], v. 24, n. 3, p.319-324, 26 nov. 2007.

NASIR, A.; CAETANO-ANOLLES, G.. A phylogenomic data-driven exploration of viral origins and evolution. **Science Advances**, [s.l.], v. 1, n. 8, p.1-24, 25 set. 2015. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/sciadv.1500527>.

PATEL, Maulik R; EMERMAN, Michael; MALIK, Harmit S. Paleovirology—ghosts and gifts of viruses past. **Current Opinion In Virology**, [s.l.], v. 1, n. 4, p.304-309, out. 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.coviro.2011.06.007>.

PEREIRA, Soraia Fernanda Costa Inácio. **Estudo Evolutivo dos Grandes Vírus Nucleocitoplasmáticos de DNA da Família Mimiviridae**. 2014. 177 f. Dissertação

(Mestrado) - Curso de Ciências Biológicas, Universidade Federal da Paraíba, João Pessoa, 2014.

RAOULT, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabasegenome sequence of Mimivirus,. **Science**, 306, 1344–1350. doi: 10.1126/science.1101485.

RAOULT, D. (2014). How the virophage compels the need to readdress the classification of microbes. **Virology** 477, 119–124. doi: 10.1016/j.virol.2014.11.014.

RENESTO, P. et al. Mimivirus Giant Particles Incorporate a Large Fraction of Anonymous and Unique Gene Products. **Journal Of Virology**, [s.l.], v. 80, n. 23, p.11678-11685, 13 set. 2006. American Society for Microbiology. <http://dx.doi.org/10.1128/jvi.00940-06>.

RIDLEY, Mark. Genômica evolutiva: incorporações simbióticas e transferências gênicas horizontais entre espécies influem na evolução do genoma. In: RIDLEY, Mark. **Evolução**. 3. ed. Porto Alegre: Artmed Editora, 2006. Cap. 19. p. 577-591.

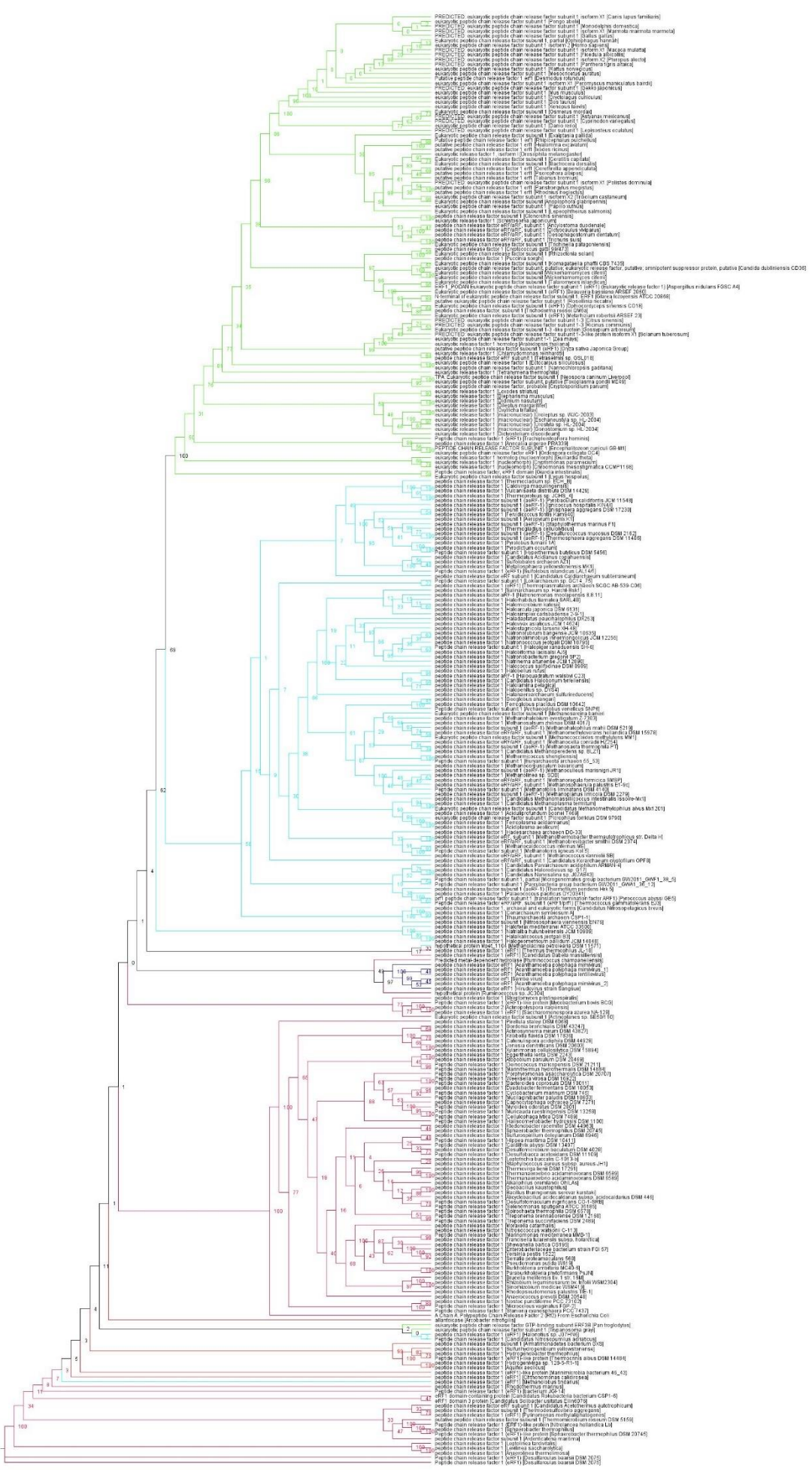
SHARMA, Vikas et al. Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. **Current Opinion In Microbiology**, [s.l.], v. 31, p.16-24, jun. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.mib.2015.12.010>.

SOUICY, Shannon M.; HUANG, Jinling; GOGARTEN, Johann Peter. Horizontal gene transfer: building the web of life. **Nat Rev Genet**, [s.l.], v. 16, n. 8, p.472-482, 17 jul. 2015. Nature Publishing Group. <http://dx.doi.org/10.1038/nrg3962>.

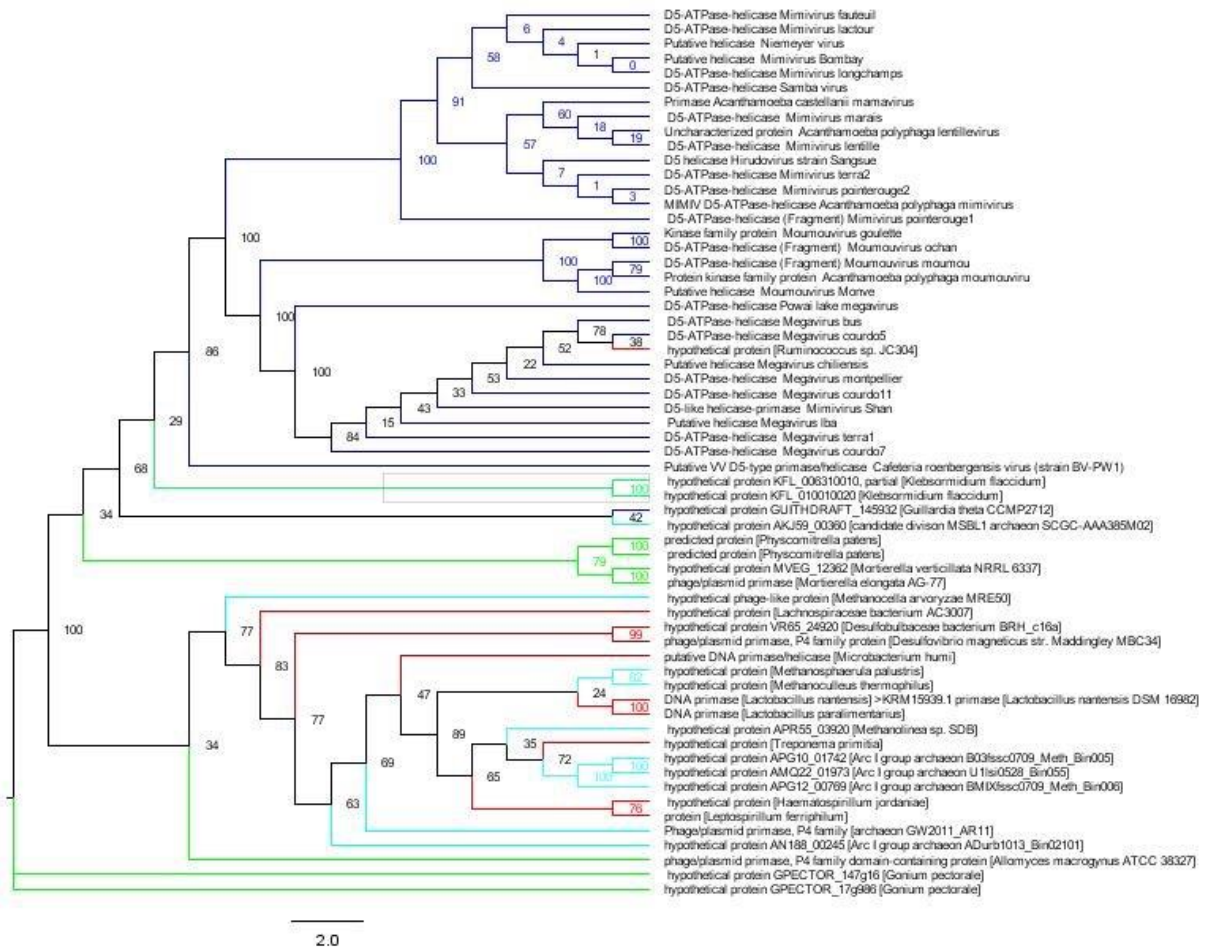
STÖVER, Ben C; MÜLLER, Kai F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. **Bmc Bioinformatics**, [s.l.], v. 11, n. 1, p.1-9, 2010. Springer Nature. <http://dx.doi.org/10.1186/1471-2105-11-7>.

WILLIAMS, T. A.; EMBLEY, T. M.; HEINZ, E. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. **PloS one**, v. 6, n. 6, p. e221080, 2011.

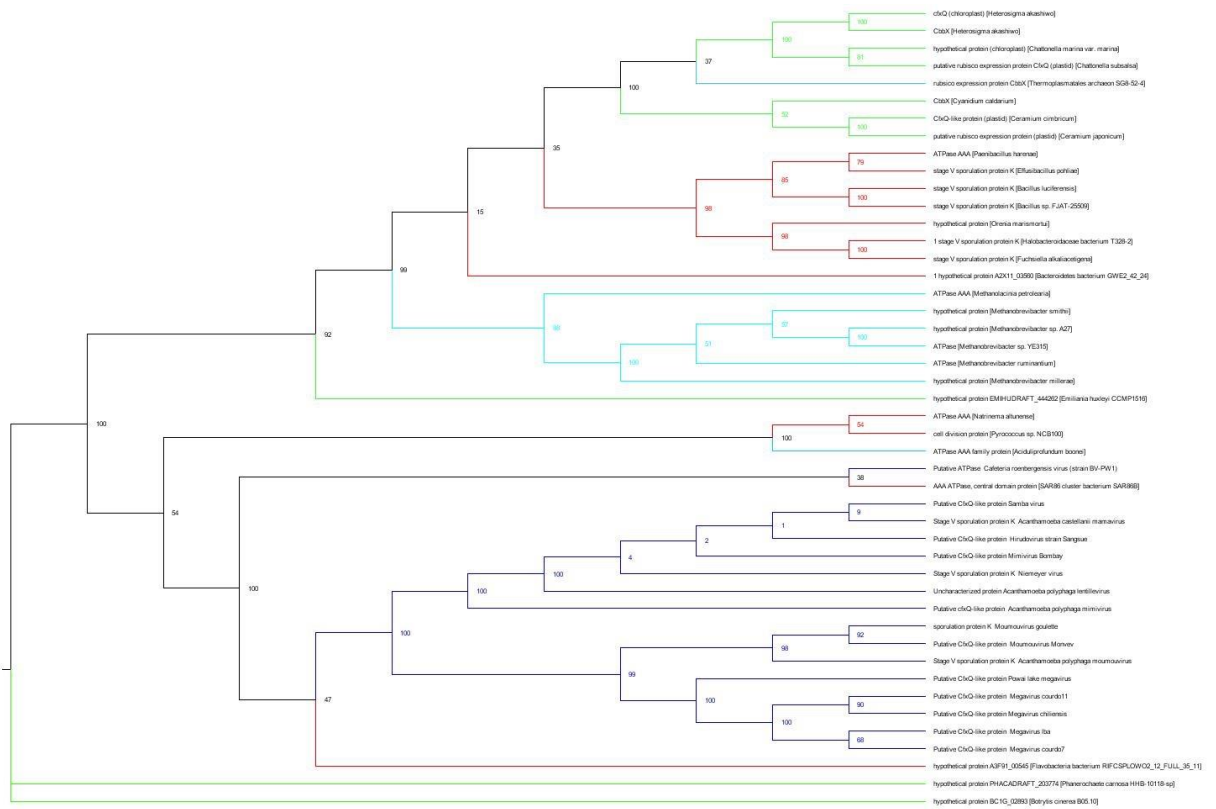
ANEXO A**Peptide chain release fator eRF1**



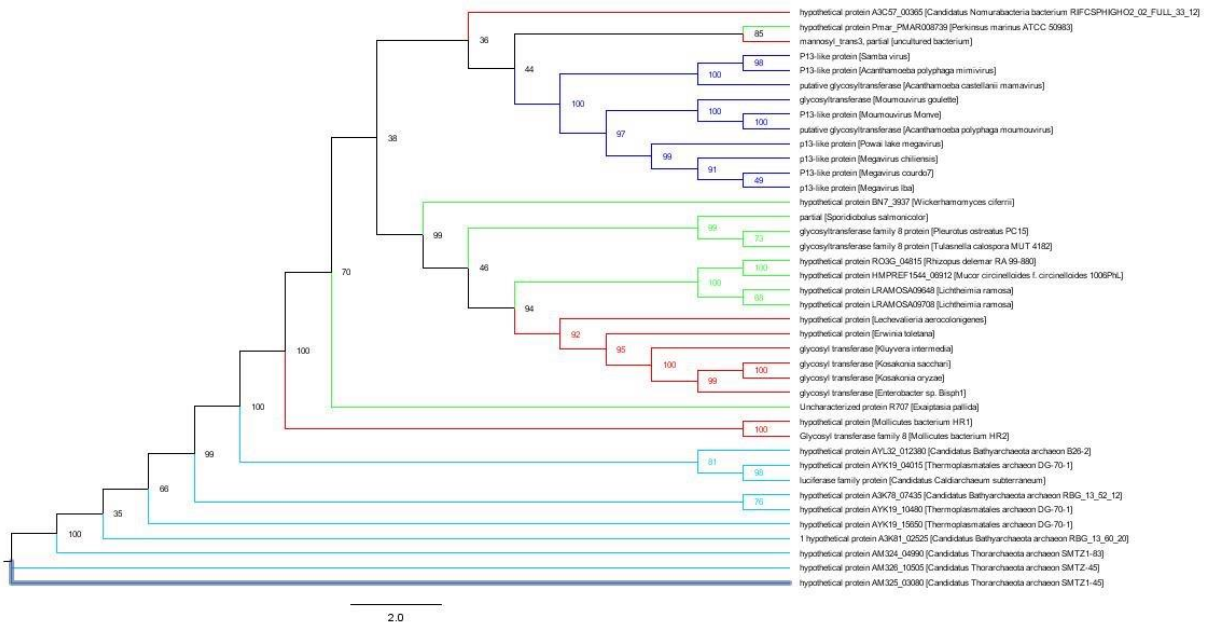
D5-ATPase-Helicase



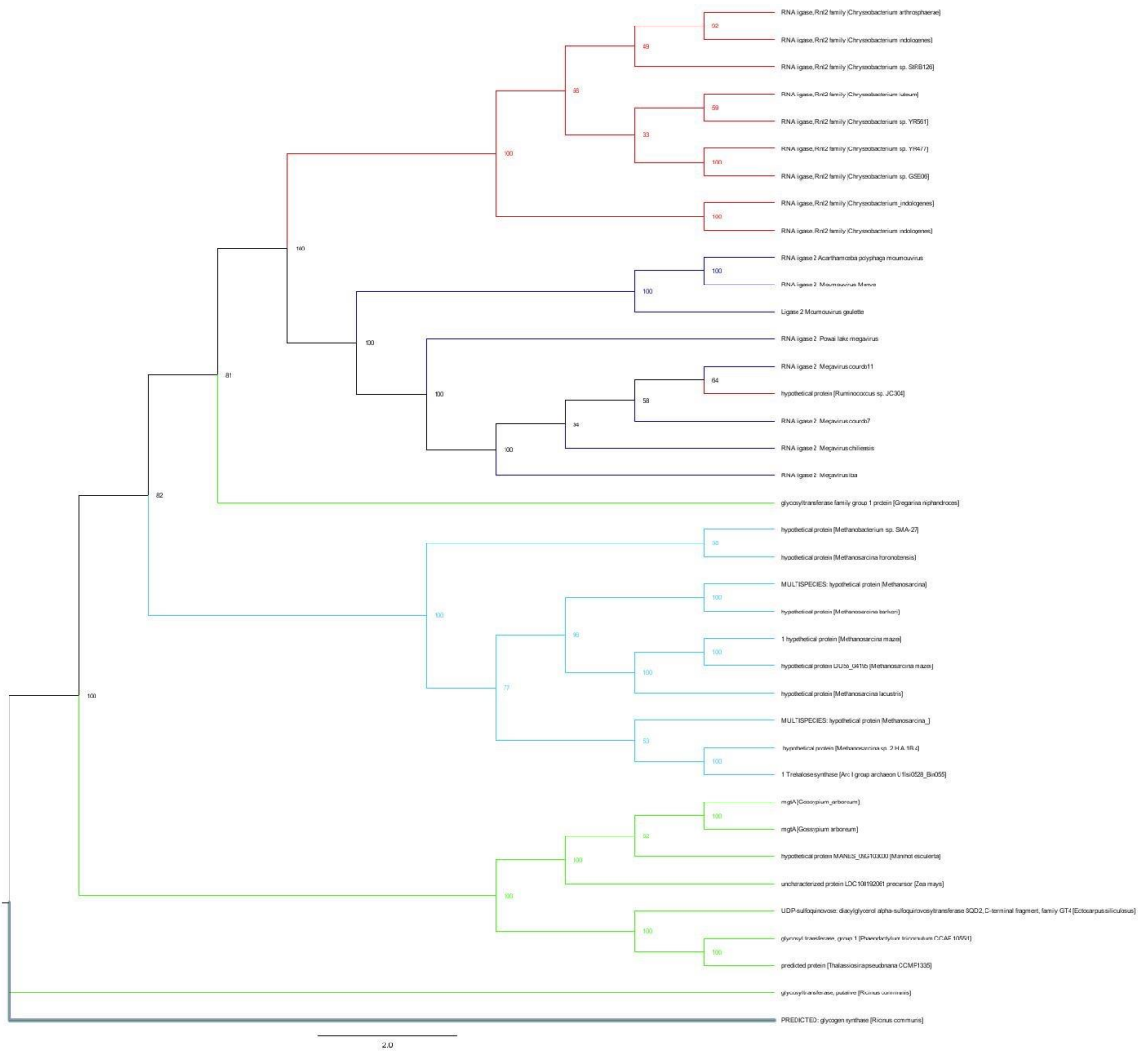
CFXQ



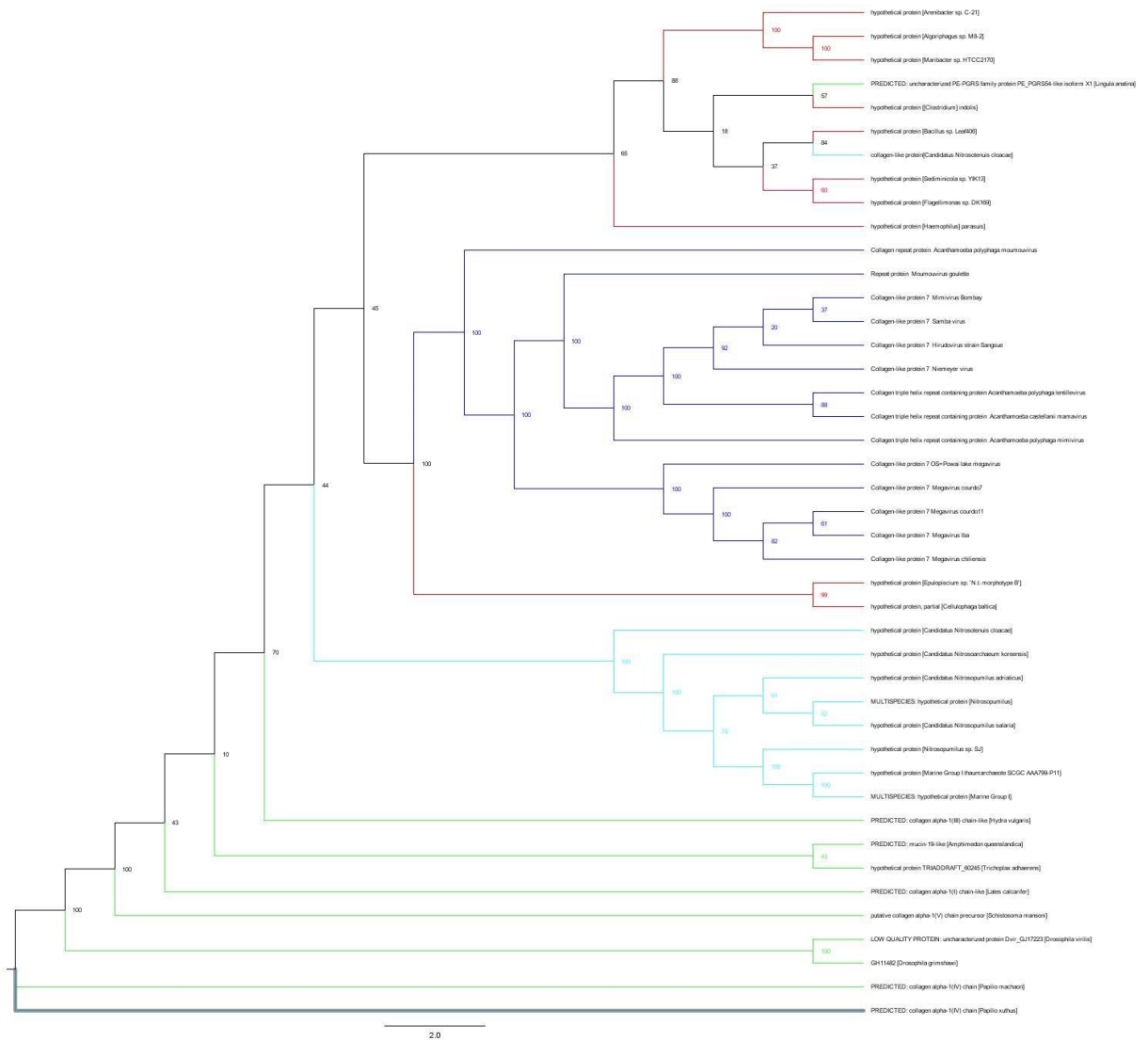
p13-like protein



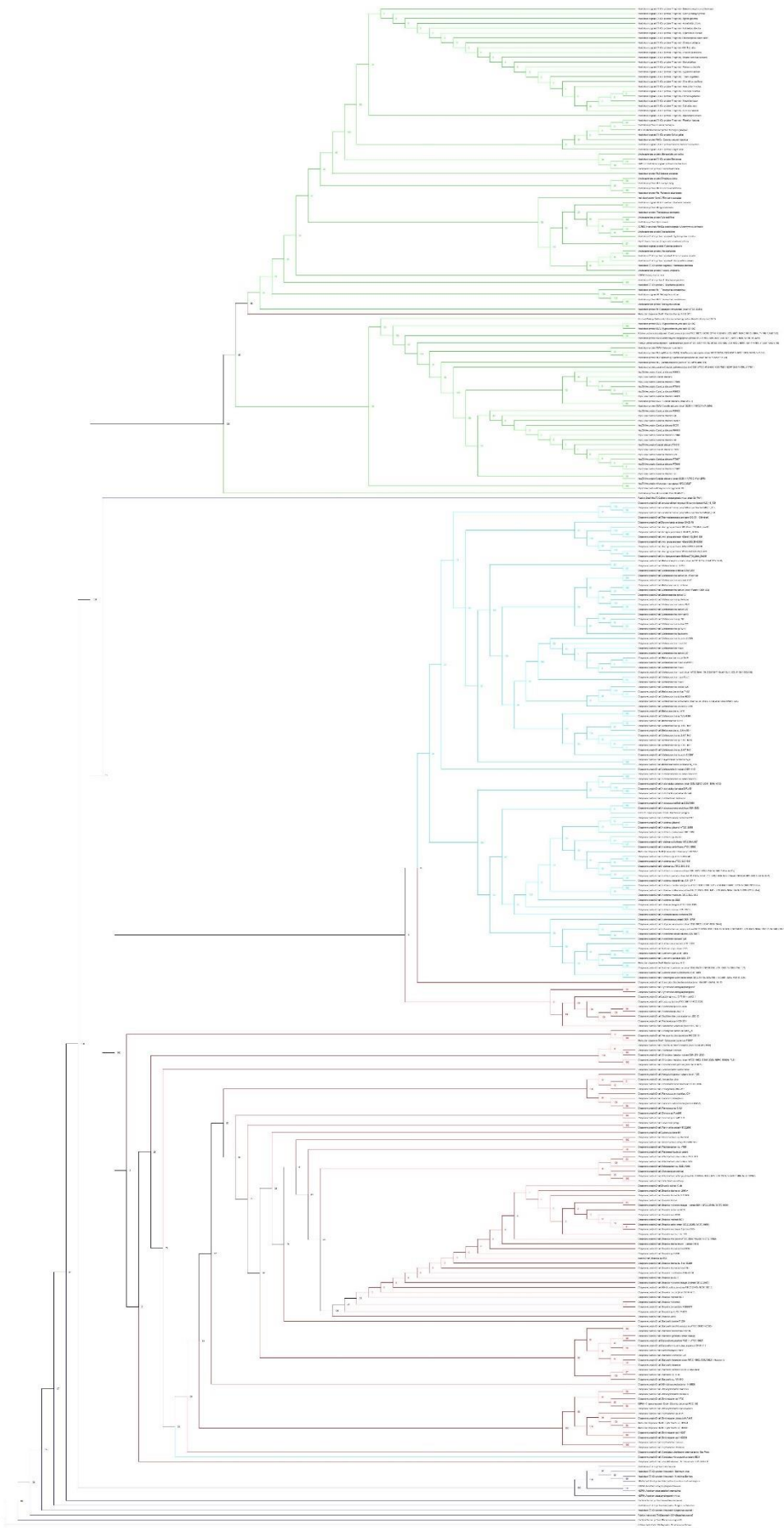
RNA Ligase 2



Collagen like protein 7



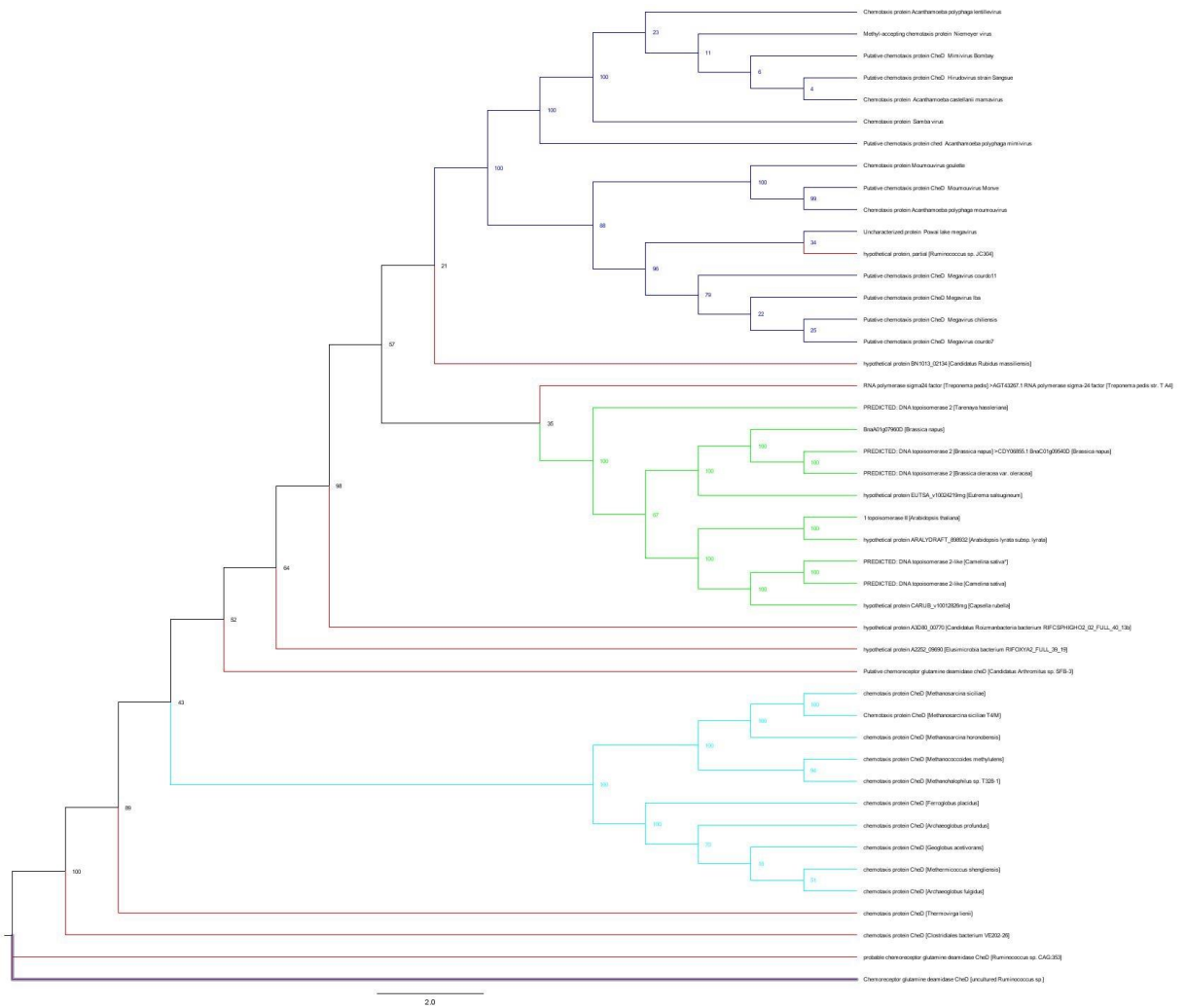
Putative DnaKHsp70



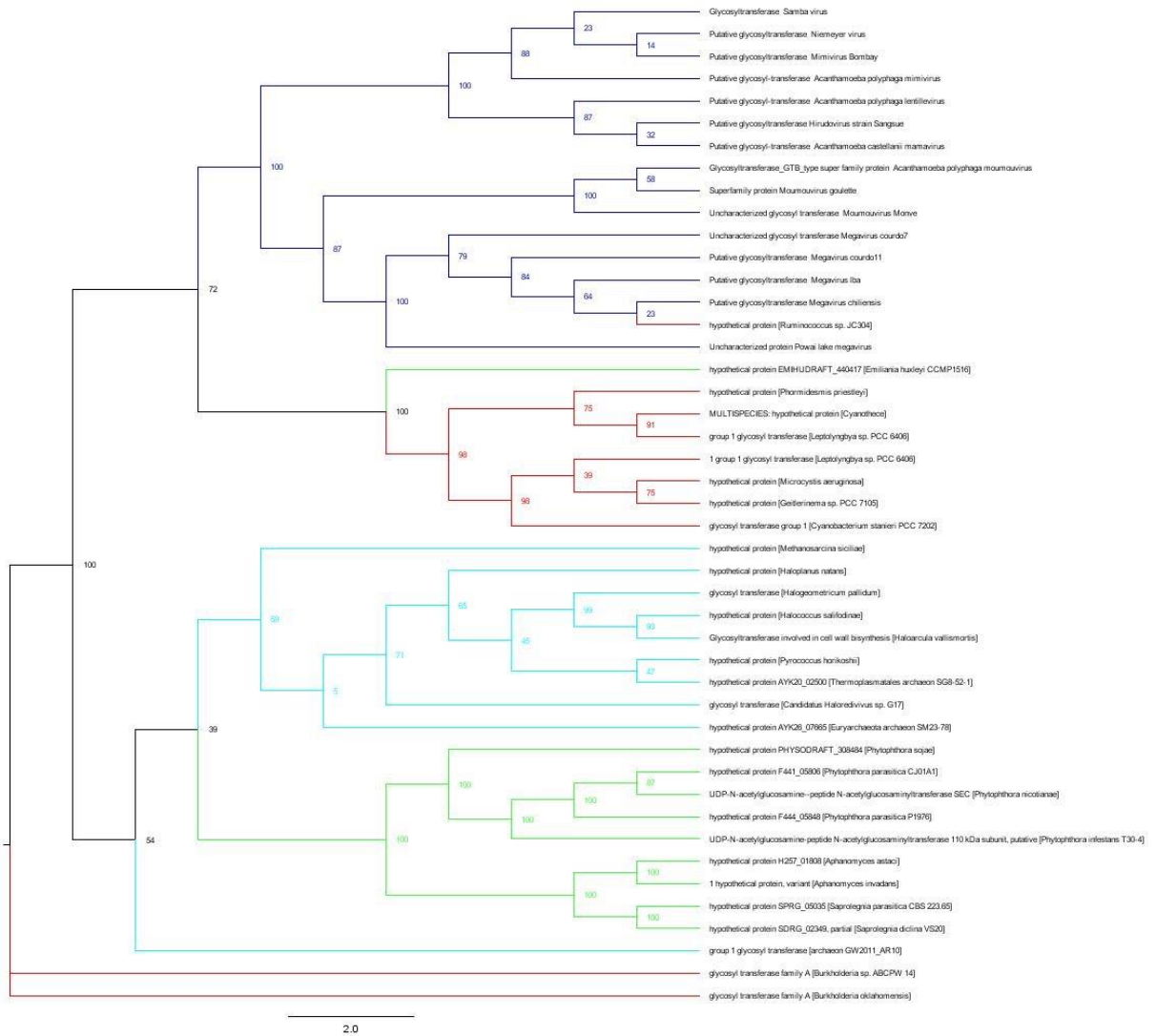
Endonuclease VIII-like



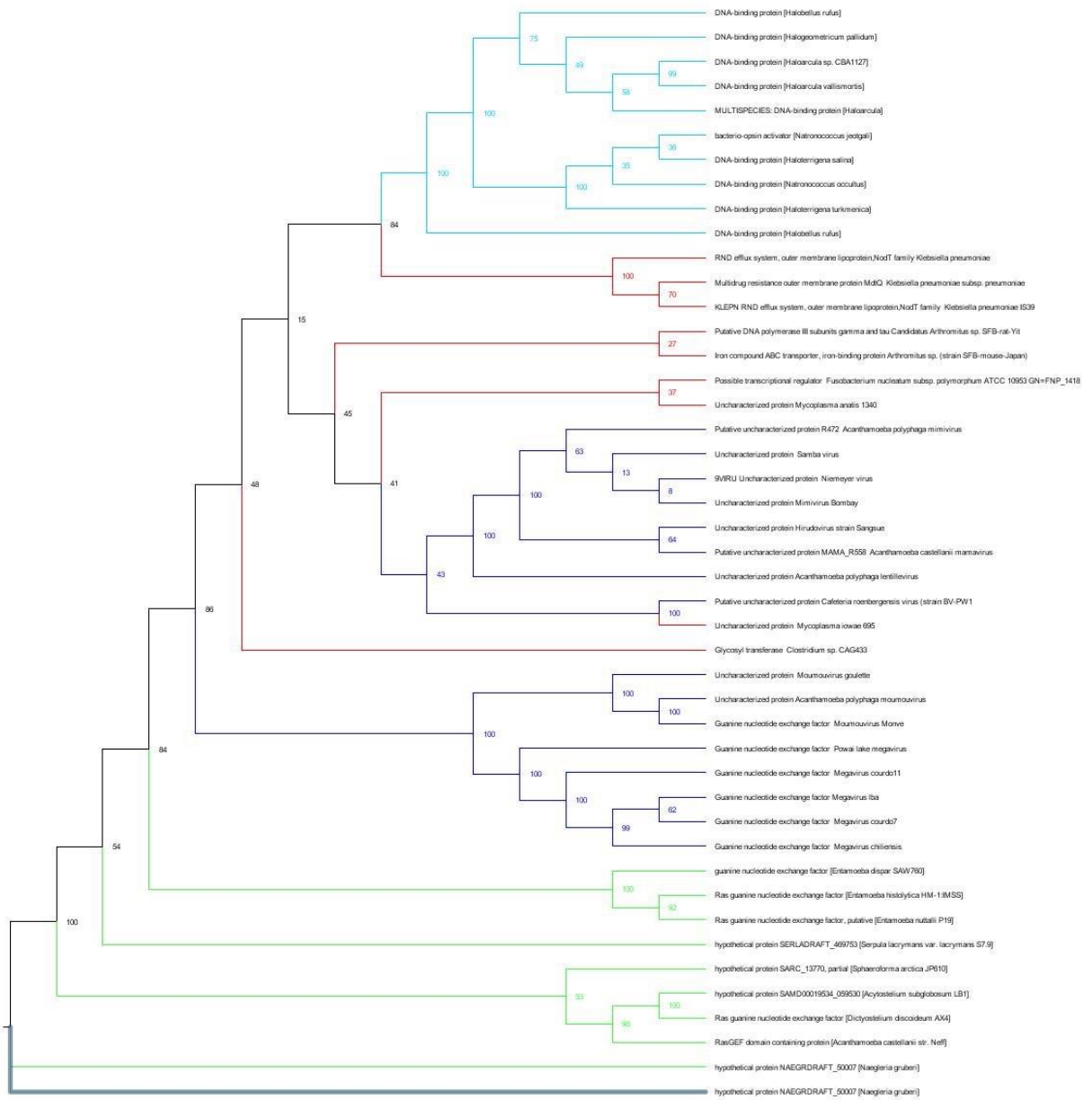
Putative chemotaxis protein CheD



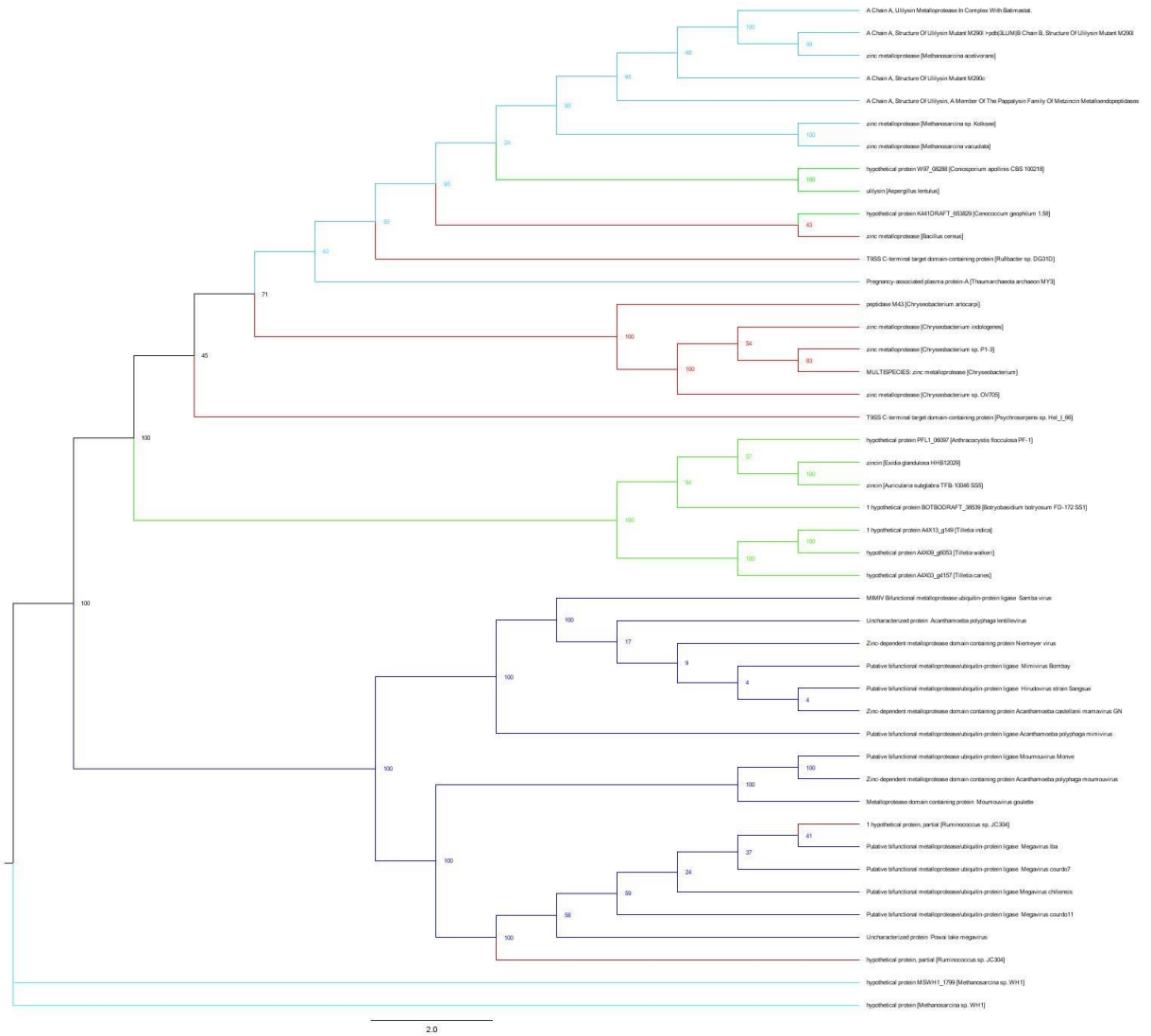
Putative Glycosyl transferase



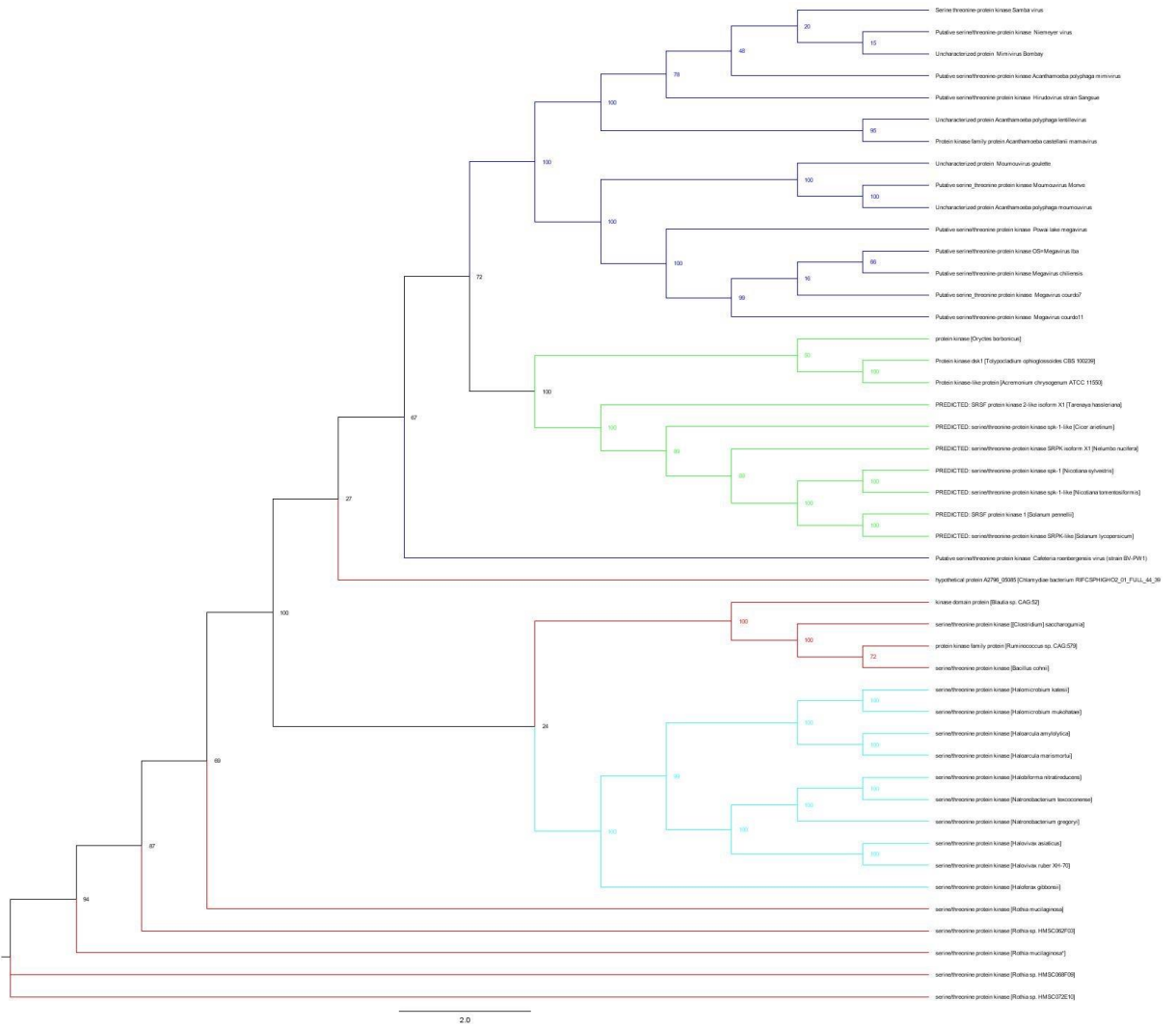
Guanine nucleotide Exchange fator



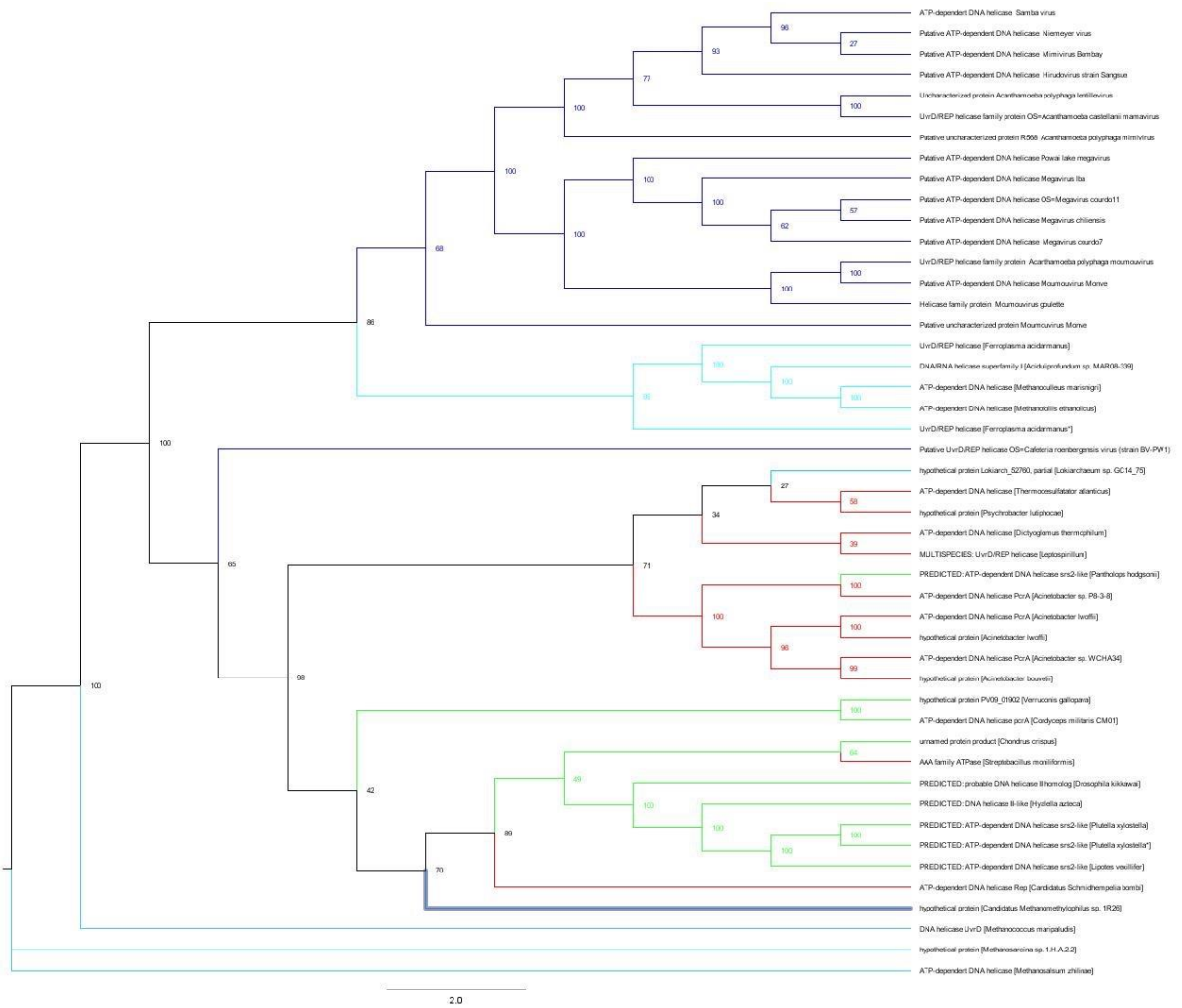
Putative Bifunctional Metalloproteaseubiquitin-protein ligase



Putative Serinethreonine-protein kinase



Putative UvrDREP helicase



Putative Ubiquitin-activating enzyme E1



Anexo B Tabela de transcrição

Proteínas Eukarya

Factor de montagem de quinase activadora de CDK MAT1

Ciclina H

Ciclina-dependente de quinase7

Proteína de reparo de excisão de DNA ERCC-2

Proteína de reparo de excisão de DNA ERCC-3 Proteína de reparo de excisão de DNA ERCC-2

RNA-polimerase dirigida por DNA I subunidade RPA12,RPA14,RPA34,RPA49 e RPA43

RNA-polimerase dirigida por DNA I subunidade RPA2 e RPA1

RNA-polimerase dirigida por DNA II subunidade RPB2, RPB3,RPB1 e RPB11

RNA-polimerase dirigida por DNA II subunidade RPB4, RPB7e RPB9

RNA-polimerase dirigida por DNA III subunidade RPC1 e RPC2

RNA-polimerase dirigida por DNA III subunidade RPC3,RPC4,RPC11,RPC8,RPC7,RPC6,RPC5

RNA-polimerase dirigida por DNAs I and III subunidade RPAC1 e RPAC2

RNA-polimerase dirigida por DNAs I and III subunidade RPAC1 e RPAC2

TFIIH Complexo basal de transcrição subunidade TTD-A

Fator de iniciação da transcrição TFIIA subunidade grande

Fator de iniciação da transcrição TFIIA subunidade pequena

Fator de iniciação da transcrição TFIIB

Fator de iniciação da transcrição TFIID subunidade 1

Fator de iniciação da transcrição TFIID subunidade 10

Fator de iniciação da transcrição TFIID subunidade 11

Fator de iniciação da transcrição TFIID subunidade 12

Fator de iniciação da transcrição TFIID subunidade 13

Fator de iniciação da transcrição TFIID subunidade 15

Fator de iniciação da transcrição TFIID subunidade 2

Fator de iniciação da transcrição TFIID subunidade 3

Fator de iniciação da transcrição TFIID subunidade 4

Fator de iniciação da transcrição TFIID subunidade 5

Fator de iniciação da transcrição TFIID subunidade 6
Fator de iniciação da transcrição TFIID subunidade 7
Fator de iniciação da transcrição TFIID subunidade 8
Fator de iniciação da transcrição TFIID subunidade 9
Fator de iniciação da transcrição TFIID subunidade 9B
Fator de iniciação da transcrição TFIID TATA-box-proteína de ligação
Fator de iniciação da transcrição TFIID/TFIIF subunidade
Fator de iniciação da transcrição TFIIE subunidade alpha
Fator de iniciação da transcrição TFIIE subunidade beta
Fator de iniciação da transcrição TFIIF subunidade alpha
Fator de iniciação da transcrição TFIIF subunidade beta
Fator de iniciação da transcrição TFIIH subunidade 1
Fator de iniciação da transcrição TFIIH subunidade 2
Fator de iniciação da transcrição TFIIH subunidade 3
Fator de iniciação da transcrição TFIIH subunidade 4
Fator de iniciação da transcrição TFII-I

Proteínas Bacteria

RNA- Polimerase dirigida por DNA subunidade alpha
RNA- Polimerase dirigida por DNA subunidade beta
RNA- Polimerase dirigida por DNA subunidade beta'
RNA- Polimerase dirigida por DNA subunidade delta
RNA- Polimerase dirigida por DNA subunidade omega

Proteínas Archaea

RNA- Polimerase dirigida por DNA subunidade A'
RNA- Polimerase dirigida por DNA subunidade B'

RNA- Polimerase dirigida por DNA subunidade D

RNA- Polimerase dirigida por DNA subunidade E'

RNA- Polimerase dirigida por DNA subunidade F

RNA- Polimerase dirigida por DNA subunidade G

RNA- Polimerase dirigida por DNA subunidade H

RNA- Polimerase dirigida por DNA subunidade K

RNA- Polimerase dirigida por DNA subunidade L

RNA- Polimerase dirigida por DNA subunidade N

RNA- Polimerase dirigida por DNA subunidade P

Anexo C

Parametros de alinhamento

Aminoácidos com letras maiúsculas; direção das sequências de nucleotídeos: o mesmo que a entrada; ordem de saída: Alinhado; estratégia: Auto (FFT-NS-1, FFT-NS-2, FFT-NS-i ou L-INS-i depende do tamanho dos dados); tentar alinhar as regiões gappy de qualquer maneira; matriz de pontuação para sequências de aminoácidos Blossum 62; a sequência superior contra as outras e Plot e alinhamento.

Anexo D

Parametros para filogenia

sequências de entrada intercaladas, frequência de aminoácidos modelo, proporção de sítios invariáveis quando presentes, número de categorias de taxa de substituição relativa quatro, parâmetro de distribuição gama quando presente, média de cada classe de taxa, otimização da topologia de árvore com comprimentos de ramificação, árvore de partida BioNJ, operações de pesquisa de topologia de árvore: movimentos de SPR (lentos, precisos), adicionar cinco árvores de início aleatória e análise não paramétrica de bootstrap com cem replicatas.