

Universidade Federal do Rio Grande do Sul

Centro de Biotecnologia

Programa de Pós-Graduação em Biologia Celular e Molecular

**Desenvolvimento de *pipeline* para filogenômica de procariotos e genômica evolutiva de *Mycoplasma* spp.**

Dissertação de Mestrado

**Meiski Mariá Vedovatto**

Porto Alegre, Outubro 2022

Universidade Federal do Rio Grande do Sul

Centro de Biotecnologia

Programa de Pós-Graduação em Biologia Celular e Molecular

**Desenvolvimento de *pipeline* para filogenômica de procariotos e genômica evolutiva de *Mycoplasma* spp.**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para obtenção do grau de Mestre.

**Meiski Mariá Vedovatto**

Profa. Dra. Claudia Elizabeth Thompson – Orientadora

Prof. Dr. Henrique Bunselmeyer Ferreira – Orientador

Porto Alegre, Outubro de 2022

## **AGRADECIMENTOS**

A meu orientador, Henrique, e à coordenação do Programa de Pós-Graduação em Biologia Celular e Molecular, agradeço por todo suporte, preocupação e instrução. A realização desse projeto foi num momento difícil, onde o incentivo constante vindo do Henrique foi essencial. Obrigada por ser uma referência no rigor científico e por todas as reuniões onde demonstrou grande compreensão.

A minha orientadora Claudia, um grande agradecimento por ter apresentado o mundo da bioinformática de uma forma tão apaixonante. Agradeço aos meus amigos Brisa, Lívia e John, assíduos no alívio cômico. Aos meus colegas de laboratório Andrea, Fábio, Gabriela, Renato, Rodrigo e Vinícius deixo o agradecimento carregado do sentimento saudade.

*There are many remaining assumptions and simplifications, which are often made to retain computational feasibility, and in certain cases the methods can be misled by recombination, small sample size, or biological processes not included in the model. We take great care to highlight such possible shortcomings, because it is our firm belief that the knowledge of a methods' limitations is as important as the knowledge of their power.*

Lemey *et al.*, The Phylogenetic Handbook

p. 423

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	<b>5</b>
1.1 Genômica, filogenômica e genômica evolutiva	5
1.2 Classe Mollicutes	7
1.2.1 Taxonomia e características gerais	7
1.2.2 Micoplasmas	9
1.2.3 Evolução de micoplasmas	11
<b>2. JUSTIFICATIVAS</b>	<b>13</b>
<b>3. OBJETIVOS</b>	<b>15</b>
3.1 Objetivo geral	15
2.2 Objetivos específicos	15
<b>4. MATERIAIS E MÉTODOS</b>	<b>16</b>
4.1 Configurações do computador utilizado	16
4.1 Sequências e bancos de dados	16
4.2 Sistema operacional, linguagem de programação e ferramentas utilizadas nas análises filogenômicas	17
<b>5. RESULTADOS</b>	<b>18</b>
5.1 Desenvolvimento do pipeline para inferência filogenética	18
5.2 Aplicação do Pipelution para o estudo filogenômico de micoplasmas	21
5.3 Desempenho	27
<b>6. DISCUSSÃO</b>	<b>29</b>
6.1 Paradigma experimental do pipeline	29
6.2 Validação do Pipelution com micoplasmas	31
<b>7. CONCLUSÕES</b>	<b>34</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>35</b>
<b>CURRICULUM VITÆ resumido</b>	<b>44</b>
<b>APÊNDICE 1</b>	<b>47</b>
<b>APÊNDICE 2</b>	<b>50</b>
<b>APÊNDICE 3</b>	<b>52</b>
<b>APÊNDICE 4</b>	<b>72</b>
<b>APÊNDICE 5</b>	<b>73</b>

## RESUMO

Como resultado das novas tecnologias de sequenciamento, um grande número de espécies teve seus genomas completamente sequenciados e os dados gerados foram disponibilizados em bancos de dados públicos. Esses dados impulsionaram estudos evolutivos baseados em genômica comparativa usando ferramentas filogenéticas. Tais estudos computacionais fornecem subsídios para o entendimento de processos que moldaram os genomas das espécies e ajudam a inferir suas relações evolutivas. Bactérias do gênero *Mycoplasma* constituem um caso interessante para testar um *pipeline* de análise filogenômica, uma vez que possuem alguns dos menores genomas conhecidos dentre organismos capazes de realizar autorreplicação e que há um crescente número de novas espécies que estão sendo identificadas e tendo seus genomas sequenciados. Portanto, com o objetivo de automatizar o processo de análise filogenômica para a inferência e a reconstrução evolutiva de espécies, foi desenvolvido um *pipeline* denominado Pipelution e micoplasmas foram utilizados para uma primeira análise e validação da ferramenta. O *pipeline* foi testado usando genomas de 105 espécies de micoplasmas, todos recuperados de bancos de dados públicos. Ele foi implementado em Python 3, fazendo uso de bibliotecas auxiliares. Na primeira etapa do processamento, os grupos de ortólogos são identificados com a ferramenta OrthoFinder e, em seguida, são alinhados com os programas PRANK ou Clustal Omega, podendo ser aparados pela ferramenta trimAl. A ferramenta ModelTest-NG é usada para calcular o modelo evolutivo mais adequado para a reconstrução filogenética, que pode ser realizada com as ferramentas RAxML, PhyML, ou com o programa MrBayes. Dezesesseis filogenias foram inferidas com a combinação dos métodos implementados no *pipeline*, cada uma usando sequências de 43 proteínas ortólogas identificadas em todas as espécies analisadas. Destas 16 filogenias, duas tiveram todos os ramos com suporte acima de 80%, ambas utilizando o alinhador PRANK e reconstrução por inferência bayesiana. Não houve diferença significativa no suporte ao comparar o resultado dos alinhadores e o nível de completude do alinhamento. Os métodos de inferência, entretanto, demonstraram ter impacto no suporte e na topologia, sendo a reconstrução por distância a melhor para solucionar a ordem dos grandes grupos. Todas as 16 filogenias reproduziram a topologia clássica de micoplasmas, validando a estratégia utilizada no Pipelution.

## ABSTRACT

As a result of new sequencing technologies, a large number of whole genome sequences has become available in public databases. These data boosted evolutionary studies based on comparative genomics using phylogenetic tools. Such theoretical and computational studies have provided clues on the processes that shaped the genomes of current species and have helped inferring their relationships. Bacteria of the *Mycoplasma* genus (mycoplasmas) constitute an interesting test case for a phylogenomic pipeline, as they have some of the smallest genomes of known self-replicating organisms and new mycoplasma species have been identified and had their genomes sequenced. Therefore, in order to automate the process of phylogenomic analysis for phylogenetic reconstruction and inference of evolutive models, a pipeline called Pipelution was developed using mycoplasmas for its first analysis and validation. The pipeline was tested using the genomes of 105 mycoplasma species, all retrieved from public databases. In the developed pipeline, first, orthogroups are identified by OrthoFinder. Next, the ortholog groups are aligned by PRANK or Clustal Omega, with trimming by trimAl. ModelTest-NG is used to infer the best-fit evolutionary model for phylogenetic reconstruction, which can be performed by distance method, with RAxML, PhyML or MrBayes tools. Sixteen trees were inferred by the combination of methods implemented in the pipeline, each one using sequences of 43 proteins, with orthologs shared by all assessed species. Of these 16 phylogenies, two had all branches with support above 80%, both with alignment by PRANK and reconstruction with bayesian inference. There was no significant difference when the alignment algorithm and the level of completeness of alignment were compared. Inference methods had a major impact in support and topology and distance reconstruction was the best for solving inner great groups. All 16 phylogenies reproduced the classical topology of mycoplasmas, validating the strategy used in the developed pipeline.

## LISTA DE ABREVIATURAS E UNIDADES

<b>A</b>	Adenina
<b>aa</b>	Aminoácidos
<b>AICc</b>	Critério de Informação Akaike com correção
<b>C</b>	Citosina
<b>CDS</b>	Sequência de DNA codificadora
<b>CPU</b>	Unidade Central de Processamento (do inglês, <i>Central Processing Unit</i> )
<b>DNA</b>	Ácido desoxirribonucleico
<b>G</b>	Guanina
<b>GO</b>	Grupos ortólogos
<b>h</b>	Horas(s)
<b>HMM</b>	Modelo Oculto de Markov (do inglês, <i>hidden Markov models</i> )
<b>IB</b>	Inferência bayesiana
<b>kb</b>	quilobase(s)
<b>Mb</b>	megabase(s)
<b>min</b>	Minuto(s)
<b>MV</b>	Máxima verossimilhança
<b>MP</b>	Máxima parcimônia
<b>NJ</b>	Neighbour Joining
<b>OTU</b>	Unidade taxonômica operacional (do inglês, <i>operational taxonomic unit</i> )
<b>RAM</b>	Memória de Acesso Aleatório (do inglês, <i>Random Access Memory</i> )
<b>RNA</b>	Ácido ribonucleico
<b>rRNA</b>	Ácido ribonucleico ribossômico
<b>SH-like</b>	Similar a Shimodaira-Hasegawa
<b>SNP</b>	Polimorfismos de nucleotídeo único (do inglês <i>single nucleotide polymorphisms</i> )
<b>TGH</b>	Transferência genética horizontal
<b>T</b>	Timina
<b>U</b>	Uracila
<b>NCBI</b>	Centro Nacional de Informação Biotecnológica (do inglês, <i>National Center for Biotechnology Information</i> )



## 1. INTRODUÇÃO

### 1.1 Genômica, filogenômica e genômica evolutiva

A partir da publicação da primeira versão da sequência completa do genoma humano, em 2001 (J *et al.*, 2001; LANDER *et al.*, 2001), e com a evolução constante das tecnologias de sequenciamento de DNA, houve o crescimento do número de organismos com genomas disponíveis completamente sequenciados. Informações biológicas de diferentes grandezas passam a ser adquiridas em diversas áreas, como a microbiologia, veterinária e ecologia (HASIN; SELDIN; LUSIS, 2017). Esses dados tornam-se importantes por representarem um avanço na intersecção desses domínios e viabilizar o estudo da biologia dos organismos de modo minucioso.

Na pecuária, por exemplo, infecções bacterianas diversas acometem as criações de bovinos, suínos, ovinos e caprinos (NICHOLAS; AYLING; MCAULIFFE, 2008), levando à redução da produção e gerando impacto econômico negativo (ABDELHAY KAOUD, 2019). Com o estudo das informações genética dos agente etiológicos é possível rastrear a origem desses surtos bacterianos (ABRAHAMIAN *et al.*, 2019; HOLT *et al.*, 2012), o que auxilia no combate aos focos de infecção e na prevenção de novos surtos. Adicionalmente, por meio da identificação de padrões evolutivos, por exemplo, adquirem-se dados que auxiliam no desenvolvimento de fármacos e de vacinas (CALCUTT *et al.*, 2018; DOMMARAJU, 2021), na compreensão de linhagens resistentes e que acabam por contribuir para o entendimento das forças evolutivas operantes em diferentes condições (TOPRAK *et al.*, 2011; DAVIES *et al.*, 2021; DIDELOT, 2021).

A filogenética, área que estuda a história evolutiva de segmentos específicos das sequências nucleotídicas, é dependente de algoritmos para a identificação de ortologia, alinhamento múltiplo de sequências, inferência de modelos evolutivos e reconstrução de árvores filogenéticas (CURRAT *et al.*, 2015; ZHANG; LIN, 2015). Esse tipo de análise considera diversas características, dentre elas a história evolutiva do grupo, as taxas de substituição de nucleotídeos

(BROWN, 2002; LAING; WHITESIDE; GANNON, 2017), identificação de regiões repetidas (LORETO; FERREIRA, 2014; LEI *et al.*, 2017) e sintenia (GUIMARAES *et al.*, 2014). Estão disponíveis na literatura estudos filogenéticos descrevendo inferências das relações evolutivas entre genes de diferentes organismos (HODGE; COPE, 2000), famílias gênicas e proteínas (GABALDÓN, 2007), relatando como suas sequências e funções sofreram alterações no decorrer do tempo.

De modo complementar, a filogenômica estuda genomas, considerando, assim, um número bem maior de sequências ortólogas a serem comparadas entre os diferentes organismos. A filogenômica possui, portanto, o potencial de inferir com maior confiabilidade a história evolutiva de organismos pertencentes a diferentes grupos taxonômicos (EISEN; FRASER, 2003; DELSUC; BRINKMANN; PHILIPPE, 2005; KLENK; GÖKER, 2010). Marcadores biológicos (LEMEY; SALEMI; VANDAMME, 2009), presença de SNPs (LAING; WHITESIDE; GANNON, 2017) e o número de genes resultante da identificação de genes ortólogos são variáveis importantes utilizadas para tal (THOMPSON *et al.*, 2015).

A evolução do *Candidatus Hepatoplasma crinochetorum*, por exemplo, foi inferida a partir de 127 genes da classe Mollicutes, identificando *H. chinochetorum* como um ramo-irmão do clado Hominis de *Mycoplasma* (LECLERCQ *et al.*, 2014). Evidências como essa são proveitosas para o diagnóstico de agentes patogênicos, classificação dos organismos (MAKIMURA *et al.*, 1999), conhecimento dos componentes mínimos necessários à sobrevivência sob longos períodos de seleção natural e no rastreamento de eventos de TGH (LIU *et al.*, 2012), todos eventos que exercem grande influência na evolução de organismos procarióticos.

A genômica evolutiva, contudo, se apresenta como um desafio teórico-computacional para a comunidade científica, devido a sua abordagem multivariada (ITAN; GERBAULT; PINES, 2015) e à magnitude do processamento de dados envolvido (BAICHO; OUZOUNIS, 2017). A inevitável interação com diversos domínios científicos aumenta o grau de complexidade das análises de genômica evolutiva e exige um nível de automação cada vez maior, atentando para fatores como o desempenho e a acurácia.

Por essa razão, é crescente o desenvolvimento de *pipelines*, ou seja, a construção de um fluxo de trabalho computadorizado, em bioinformática (LEIPZIG, 2017). O propósito principal é o de estabelecer parâmetros no desenvolvimento que propiciem o crescimento da área e a consequente disseminação do conhecimento a ela relacionado. Isso implica na necessidade de algoritmos robustos que compreendam as técnicas de manipulação de dados e possibilitem controle maximizado da análise.

## **1.2 Classe Mollicutes**

### **1.2.1 Taxonomia e características gerais**

A classe Mollicutes, composta por 6 famílias e mais de 10 gêneros (FEDERHEN, 2012), demonstra estar em expansão devido à recente identificação de organismos que posicionam-se em seu clado (DESIRÒ *et al.*, 2018; SPERGSEER *et al.*, 2022) e está passando por uma reorganização taxonômica já aceita por parte da comunidade científica (GUPTA; SON; OREN, 2018; BALISH *et al.*, 2019). Uma das premissas para a reorganização taxonômica é a presença de polifilia na classe, visto que espécies do gênero *Mycoplasma*, por exemplo, estão presentes em 3 dos 4 clados principais do grupo.

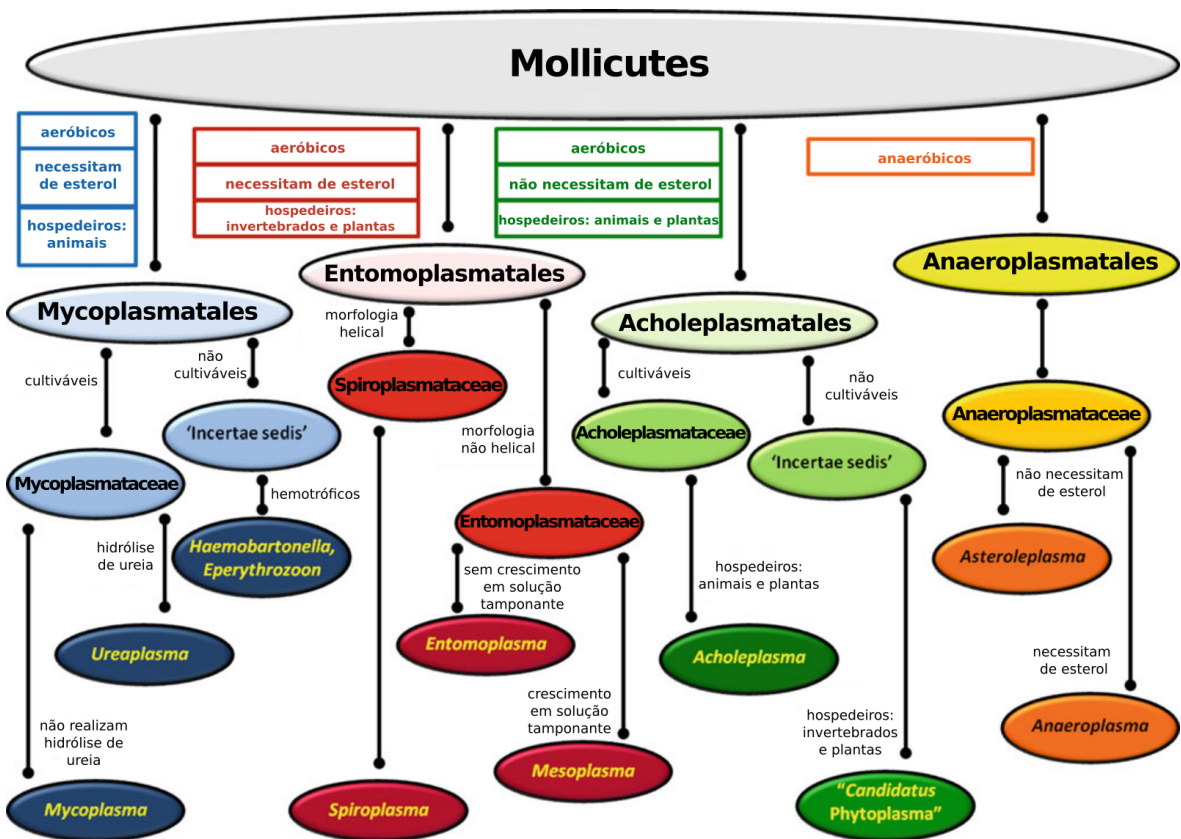
A presença de polifilia na classe se dá devido a uma classificação taxonômica baseada em marcador filogenético único e propriedades fenotípicas, como morfologia celular, precisar ou não de esterol para o crescimento e ser cultivável. Em busca de marcadores que sejam capazes de distinguir tanto de modo molecular como filogenético, foi identificado e proposta a utilização de sinapomorfias, como deleções, inserções e assinaturas únicas entre as espécies para a reorganização taxonômica (GUPTA *et al.*, 2018). A partir disso, a nomenclatura passou a ser coerente com o conceito de que os clados da classe são monofiléticos.

A alteração sugerida é de que a ordem Mycoplasmatales deve ser alterada para abranger apenas as espécies de *Mycoplasma* dentro do *cluster*

“Spiroplasma” e que uma nova ordem, Mycoplasmoidales ord. nov., deve ser criada para abranger as demais espécies de *Mycoplasma*. Os agrupamentos “Pneumoniae” e “Hominis” são propostos como duas novas famílias, Mycoplasmoidaceae fam. nov., que inclui os gêneros *Eperythrozoon*, *Ureaplasma* e os gêneros propostos *Malacoplasma* e *Mycoplasmoides*, e *Metamycoplasmataceae* fam. nov. para conter os gêneros também propostos *Metamycoplasma*, *Mycoplasmopsis* e *Mesomycoplasma*. Entretanto, ainda não há consenso sobre qual nomenclatura utilizar.

A variedade de hospedeiros atualmente conhecidos abrange plantas, insetos e outros animais (RAZIN, 2006; CITTI *et al.*, 2018). Dentre os gêneros, com exceção de *Mycoplasma* e *Ureaplasma*, são todos compostos por espécies incapazes de colonizar seres humanos. Os organismos da classe Mollicutes são caracterizados pela ausência de parede celular e genomas de tamanhos relativamente pequenos, entre 0,58 Mb e 2 Mb. Isso resulta em uma grande dependência do hospedeiro para o suprimento de nutrientes devido à incapacidade de sintetizar componentes da parede celular, ácidos nucleicos e precursores de algumas proteínas (IPOUTCHA *et al.*, 2019). A organização taxonômica da classe está esquematizada na Figura 1, considerando a taxonomia clássica.

Os gêneros polifiléticos *Spiroplasma*, *Mycoplasma*, *Ureaplasma* e *Mesoplasma* usam o códon UGA para codificar triptofano (INAMINE *et al.*, 1990), não o utilizando de códon de terminação, como ocorre no código genético universal. Esses organismos utilizam, portanto, UAA e UAG para terminação (RAZIN; YOGEV; NAOT, 1998). Estima-se que essa mudança se deu pela pressão sobre o processo de tradução com uso preferencial de códons (MUTO; OSAWA, 1987), onde o códon UAA tem uso preferencial em micoplasmas com genoma de conteúdo GC reduzido. A abundância de conteúdo AT pode chegar a 90% (RAZIN; YOGEV; NAOT, 1998), principalmente porque a composição das trincas pode exibir até 93% de AT na terceira posição, como em *Mycoplasma capricolum*, por exemplo (SHARP *et al.*, 1993).



**Figura 1. Organização da classe Mollicutes e diferentes características fenotípicas.** As cores separam as ordens, famílias e gêneros em gradientes, estes últimos representados pelos círculos inferiores. Modificado de MAY *et al.* (2014). As ordens indicadas são aquelas definidas pela taxonomia clássica (WEISBURG *et al.*, 1989).

### 1.2.2 Micoplasmas

O primeiro estudo com micoplasmas, termo doravante utilizado para referenciar bactérias pertencentes ao gênero *Mycoplasma*, foi feito em 1898 com o agente etiológico da pleuropneumonia bovina, *Mycoplasma mycoides* subsp. *mycoides* (RAZIN; HAYFLICK, 2010). Na época, esses organismos foram classificados como partículas virais e somente em 1960, a partir do método de hibridização de DNA, foi possível identificá-los como bactérias.

Já foram descritas mais de 200 espécies diferentes de micoplasmas (PARTE, 2018), as quais são capazes de causar doenças em seres humanos e animais (GUPTA; SON; OREN, 2018). Recentemente, micoplasmas foram isolados de animais até então não identificados como hospedeiros destas bactérias, como peixes da família Salmonidae e invertebrados (RAMÍREZ *et al.*,

2019; MORA-SÁNCHEZ; PÉREZ-SÁNCHEZ; BALCÁZAR, 2020), demonstrando a vasta gama taxonômica que micoplasmas são capazes de colonizar.

A primeira espécie de micoplasma identificada como agente patogênico de seres humanos foi *Mycoplasma pneumoniae*, causadora de doenças no trato respiratório inferior e superior. Alguns micoplasmas também são agentes etiológicos de doenças urogenitais em humanos, como é o caso de *Mycoplasma genitalium* (LJUBIN-STERNAK; MEŠTROVIĆ, 2014) e *Mycoplasma penetrans* (SASAKI *et al.*, 2002). O gênero também possui espécies de estilo de vida parasito oportunista, como *Mycoplasma fermentans*, *Mycoplasma hominis* (MARGARITA; FIORI; RAPPELLI, 2020) e outras, as quais foram encontradas associadas a infecções em quadros clínicos já estabelecidos (PITCHER; NICHOLAS, 2005; SCHREIBER; MÜLLER, 2020).

Apesar dessa associação ainda não ser clara, é possível afirmar que há uma correlação entre infecções por *Mycoplasma salivarum* (BORTOLOTTI *et al.*, 2020) e *Mycoplasma orale* (LIU *et al.*, 2021) e a imunossupressão de seus respectivos hospedeiros. O perfil oportunista do gênero é corroborado também devido a presença de espécies em pacientes acometidos por tumores ou em pessoas acometidas pelo vírus da imunodeficiência humana (WANG *et al.*, 1992; HANNAN, 1998; NAPIERALA MAVEDZENGE; WEISS, 2009; BARYKOVA *et al.*, 2011; XIE *et al.*, 2017).

Em relação aos nichos preferenciais para colonização por parte de micoplasmas, os tratos respiratório e urogenital, glândulas mamárias, mucosas e olhos estão entre os principais em seres humanos. Entretanto, já foram encontradas micoplasmas no sistema nervoso central de seres humanos (CHRISTO *et al.*, 2010), de focas e bovinos vinculadas a quadros de encefalites, meningite asséptica e coinfeções (TSIODRAS *et al.*, 2005; ROSALES *et al.*, 2017).

Micoplasmas também são patógenos de uma vasta gama de animais domésticos de uso na agropecuária. Bovinos, suínos, caprinos e aves são os grupos mais afetados, impactando negativamente a produção e a economia da indústria pecuária (HOLST; YESKE; PIETERS, 2015; MAES *et al.*, 2018; DUDEK

*et al.*, 2020; BERGERON *et al.*, 2021). Adicionalmente, existem agravantes no que tange infecções causadas por espécies potencialmente resistentes a múltiplos fármacos (BEYLEFELD *et al.*, 2018; PEREYRE; TARDY, 2021).

Outra particularidade de espécies do gênero *Mycoplasma* é um aparente tropismo por determinados nichos dos hospedeiros (DAWOOD *et al.*, 2022). Micoplasmas funcionalmente semelhantes e que aderem ao mesmo tipo de nicho, mas que têm hospedeiros diferentes, não são distinguíveis entre si, de acordo com Kamminga. Isso demonstra um repertório de genes similar, mas uma capacidade de resposta variada. Presume-se, portanto, que há o uso ótimo dos mecanismos de infecção e de manipulação do microambiente nos hospedeiros e seus respectivos nichos (sangue e demais tecidos), o que é coerente com o tamanho reduzido do genoma desses organismos (KAMMINGA *et al.*, 2017). Esse uso otimizado vai ao encontro de evidências de diferentes expressões entre espécies de conteúdo gênico muito similar (MACHADO *et al.*, 2020).

### **1.2.3 Evolução de micoplasmas**

A primeira hipótese evolutiva para micoplasmas, proposta por Morowitz e Wallace (MOROWITZ; WALLACE, 1973), é de que elas seriam os seres mais primitivos existentes, tendo se perpetuado antes do desenvolvimento da parede celular de peptidoglicano. Entretanto, desde 1960, Neimark defendeu a ideia de que, na verdade, o grupo passou por um processo de evolução degenerativa (NEIMARK, 1986), confirmado por filogenia de RNA ribossômico (WOESE; MANILOFF; ZABLEN, 1980).

A hipótese evolutiva melhor estabelecida, portanto, é a de que essas bactérias teriam passado por uma série de ciclos que resultaram na diminuição de seu genoma e perda de parede celular. Atualmente, micoplasmas são considerados um dos menores e mais simples organismos capazes de autorreplacação (RAZIN; HAYFLICK, 2010). Micoplasmas compartilham um ancestral comum com bactérias gram-positivas do subgrupo *Bacillus-Lactobacillus-Streptococcus* (WOESE; MANILOFF; ZABLEN, 1980), do qual estima-se que eles divergiram há 600 milhões de anos (RAZIN; YOGEV;

NAOT, 1998). Essa divergência resultou em dois ramos principais, os quais se diferenciaram há 400 milhões de anos (MANILOFF, 1996). Micoplasmas estão no ramo constituído também por *Spiroplasma* spp. e *Entomoplasma* spp. Na Figura 1, é possível observar suas características gerais: são organismos aeróbicos, requerem esterol, estão presentes em animais e não realizam hidrólise de ureia. O gênero *Mycoplasma* é dividido em três grupos internos: Hominis, Pneumoniae e Mycoides (BROWN, 2010), sendo o último mais proximamente relacionado a espécies dos gêneros *Mesoplasma* e *Entemoplasma*.

Micoplasmas do grupo Mycoides, por sua vez, possuem uma história evolutiva diferente dos outros dois clados, pois seriam originários de um ancestral associado a insetos. Tornaram-se fenotipicamente semelhantes a outras linhagens de micoplasmas por meio de eventos independentes de evolução convergente, provavelmente envolvendo TGH (LO; GASPARICH; KUO, 2018). Supõe-se que cenário semelhante tenha ocorrido com *Mycoplasma agalactiae* e aproximadamente 18% de seu genoma tenha como origem TGH com espécies do grupo Mycoides (SIRAND-PUGNET *et al.*, 2007).

Micoplasmas apresentam um mecanismo distinto de TGH (DORDET-FRISONI *et al.*, 2014). A troca de material genético entre as células bacterianas do grupo pode ser tão grande e ampla que recebe o nome de transferência cromossômica, contribuindo fortemente para a constituição de um genoma-mosaico (DORDET-FRISONI *et al.*, 2019). Esse tipo de dinâmica evolutiva se apresenta como central na diversificação e adaptação a ambientes em mudança, oferecendo vantagem significativa para micoplasmas (FAUCHER *et al.*, 2019).

Elementos genéticos móveis podem também ter parte na plasticidade fenotípica de micoplasmas. Estudos de genômica comparativa mostraram a presença de regiões repetidas em diversas espécies. Em *M. pneumoniae*, por exemplo, elas já foram identificadas cobrindo uma porção de até até 2,2 kb (MUSATOVOVA; KANNAN; BASEMAN, 2012). Tais repetições se mostram responsáveis por conferir variação antigênica às espécies. Isso foi observado também em *Mycoplasma hyopneumoniae* (BARATE *et al.*, 2014; MACHADO *et al.*, 2020), *M. genitalium* (MA *et al.*, 2010), *Mycoplasma hyorhinis* (YOGEV *et al.*,



1991) e em outras espécies (ZHANG; WISE, 1996; LIU *et al.*, 2000; LYSNYANSKY I; ROSENGARTEN R; YOGEV D, 1996; SIMMONS WARREN L.; DENISON AMY M.; DYBVIG KEVIN, 2004). A evidência dessas repetições é importante considerando o processo evolutivo do gênero que, mesmo sob pressão para redução do genoma, conservou esses "reservatórios evolutivos" (CATTANI, 2016), os quais provavelmente estariam sob pressão seletiva mais intensa (ROCHA; BLANCHARD, 2002).

Em 2012 foi identificada até então mais recente adaptação no gênero: a transição de hospedeiro de *Mycoplasma gallisepticum*, o qual antes era patogênico em aves domesticadas e passou a ser efetivo em colonizar o passarinho silvestre *Carpodacus mexicanus* (DELANEY *et al.*, 2012). Há evidências de concomitantes mutações em *M. gallisepticum* e em seu hospedeiro anterior, ou seja, um processo de coevolução com aves domesticadas (BONNEAUD *et al.*, 2018; STALEY *et al.*, 2018). Esse evento tem como uma das explicações a presença de plasmídeos, também encontrados em espécies do grupo Mycoïdes (BRETON *et al.*, 2012).

Todos esses elementos de alteração e comutação entre espécies de micoplasmas configura um novo horizonte sobre a forma como esses organismos evoluem. Delinea-se que eles são competentes em adquirir e ceder material genético, não tendo como força evolutiva principal a redução de seus genomas enquanto bactérias endossimbiontes dependentes do hospedeiro (SIRAND-PUGNET *et al.*, 2007; CITTI *et al.*, 2018).

## **2. JUSTIFICATIVAS**

O crescimento constante das bases de dados de sequências genômicas completas impõe um desafio para a análise dessas informações, pois, além do volume crescente de dados, demanda a utilização de técnicas computacionais que considerem o significado biológico dos fragmentos. Para tanto, é necessário o desenvolvimento de ferramentas de montagem de genomas, de anotação funcional de genes, de similaridades e de análises comparativas e evolutivas. A análise evolutiva, por sua vez, é de grande relevância para o entendimento da

evolução no contexto genômico, permitindo identificar e descrever processos como os de TGH, de duplicação e de geração de elementos móveis (BROWN, 2002) que moldam a evolução dos organismos (MARTINEZ-URTAZA *et al.*, 2017).

Como mencionado anteriormente, uma das formas de realizar a análise evolutiva é a partir de estudos de filogenômica, os quais normalmente conseguem reproduzir com maior suporte estatístico as filogenias (KLENK; GÖKER, 2010; ROKAS *et al.*, 2003). Para tanto, existem diversas metodologias e abordagens que partem de centenas de sequências de nucleotídeos ou aminoácidos. O intuito principal é encontrar a árvore filogenética que melhor explique os padrões observados nos dados, levando em consideração que a atribuição de significado biológico e, principalmente, a escolha do paradigma metodológico devem ser guiados pela pergunta biológica norteadora.

Estes fatores configuram um cenário de desenvolvimento tecnológico complexo, onde o rastreamento e a manutenção da integridade dos dados entre as diversas metodologias são importantes para não ocorrer a inserção de ruído na análise. O desenvolvimento é complexo também considerando o contexto de dados biológicos, que são gerados em grande quantidade e faz-se necessário construir ferramentas que façam uso de todos os recursos computacionais disponíveis com a implementação de processos paralelizáveis (LEIPZIG, 2017; WRATTEN; WILM; GÖKE, 2021).

Considerando esses fatores sobre o volume de dados e propondo o desenvolvimento de uma ferramenta que exige validação e testes constantes, faz-se necessário o uso de um *dataset* com características conhecidas e de tamanho otimizado para o processamento. Genomas de bactérias do gênero *Mycoplasma* se enquadram nestes critérios devido a serem organismos de genoma reduzido em comparação com bactérias de outros gêneros, o que permite a inclusão de uma grande quantidade de táxons sem um maior aumento no tempo de processamento. Taxonomicamente também são interessantes, pois apesar de novas espécies estarem sendo identificadas, há extenso material sobre a evolução do gênero. Por fim, questionamentos sobre a evolução, em especial sobre espécies patogênicas, têm surgido (CITTI *et al.*, 2020; CRUZ *et al.*, 2020;

TARDY *et al.*, 2020; HSIEH *et al.*, 2022) e, por consequência, aumentado a relevância de estudos filogenéticos do grupo.

### **3. OBJETIVOS**

#### **3.1 Objetivo geral**

Desenvolvimento de um *pipeline* para estudos filogenômicos de procariotos utilizando métodos teórico-computacionais, com a realização de análise de aspectos evolutivos de genomas de micoplasmas como prova de conceito.

#### **2.2 Objetivos específicos**

- i. Desenvolvimento de um *pipeline* para análise evolutiva de genomas bacterianos.
- ii. Automatização do processo de remoção de ruído filogenético com a aplicação de *cut-offs* descritos na literatura.
- iii. Aplicação do *pipeline* desenvolvido na análise filogenômica de bactérias do gênero *Mycoplasma*.
- iv. Validação da ferramenta tendo como hipótese a monofilia do gênero *Mycoplasma*.
- v. Posicionamento filogenético de espécies recentemente identificadas ainda não incluídas em análises filogenômicas.

## 4. MATERIAIS E MÉTODOS

### 4.1 Configurações do computador utilizado

O computador utilizado na análise possui o processador Intel Core™ i7-3770 CPU @ 3.40GHz x 4 e 32 GB de memória RAM. Sistema operacional Linux Mint 19.3 64 bit. O tempo total para alinhamento e cálculo de cada árvore variou de acordo com a complexidade dos testes filogenéticos realizados.

### 4.2 Sequências e bancos de dados

As sequências de proteínas deduzidas dos genomas de micoplasmas foram obtidas através do banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/>). A seleção dos genomas representativos das espécies incluídas nas análises seguiu os seguintes critérios, por ordem de prioridade: (i) linhagem-modelo indicada como genoma referência para a espécie; (ii) genoma com indicação de referência; e (iii) maior nível de montagem do genoma (completo, em cromossomo, em *scaffold* ou em *contig*). Para espécies sem montagens disponíveis no RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>), foram selecionadas as montagens do GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

A partir dos critérios estabelecidos, foram selecionados manualmente 105 genomas de micoplasmas, sendo que 54,7% desses genomas estão completos, 22,6% estão em *scaffolds*, 21,7% estão em *contigs* e 0,9% estão em nível de cromossomo. A esse conjunto, foi acrescido o genoma da unidade taxonômica externa escolhida, que foi *Bacillus subtilis* subsp. *subtilis* 168, totalizando 106 genomas selecionados. A relação dos genomas utilizados está no Apêndice 1.

### 4.3 Sistema operacional, linguagem de programação e ferramentas utilizadas nas análises filogenômicas

O *pipeline* foi construído para ambientes Linux 64-86 bit e a linguagem de programação utilizada foi o Python 3.8 (VAN ROSSUM, 2007). As ferramentas para análise filogenética utilizadas estão listadas na Tabela 1.

**Tabela 1: Ferramentas externas incorporadas no *pipeline* desenvolvido e seus usos, parâmetros e critérios de utilização.**

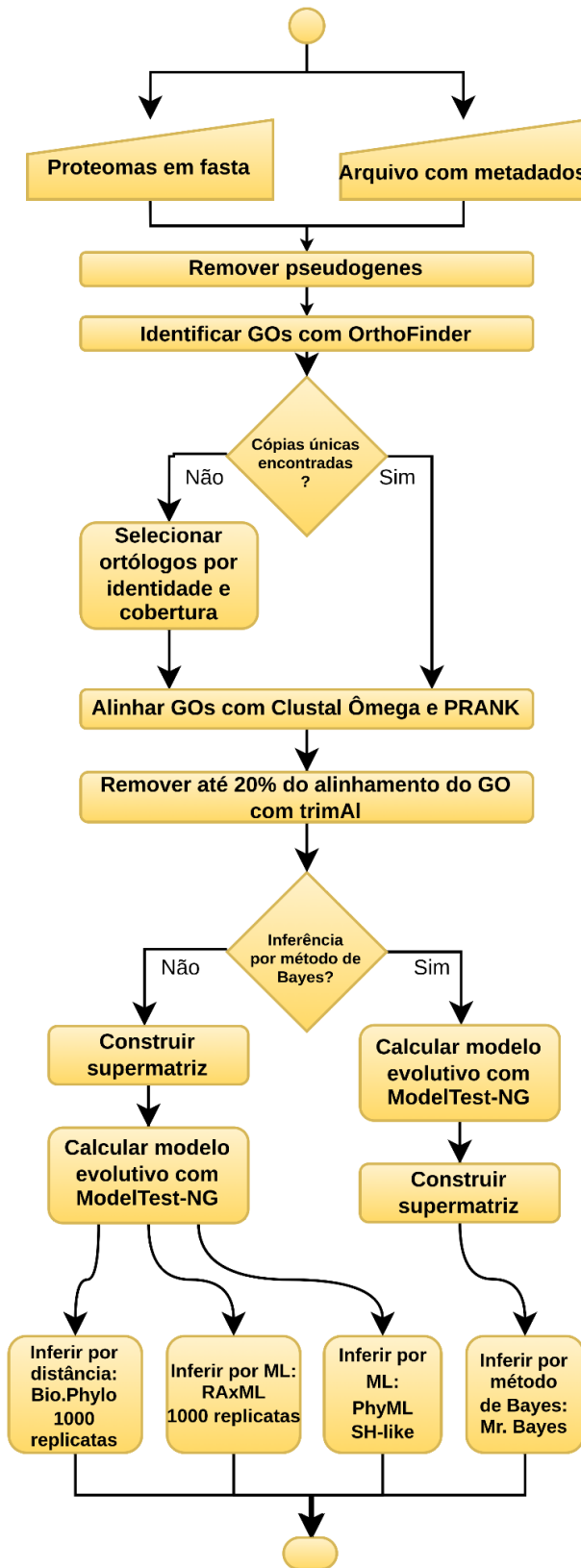
Ferramenta	Uso	Parâmetros e critérios
OrthoFinder 2.2.7 (EMMS; KELLY, 2015)	Identificação de ortólogos	Utilizado com parâmetros-padrão
PRANK (LÖYTYNOJA; GOLDMAN, 2008)	Alinhamento	Utilizados com parâmetros-padrão
Clustal Omega (SIEVERS <i>et al.</i> , 2011)		
trimAl (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009)	Aparar alinhamento	Remoção de no máximo 20% do alinhamento
ModelTest-NG (DARRIBA <i>et al.</i> , 2020)	Cálculo dos modelos evolutivos e identificação do melhor modelo para cada grupo de ortólogo	As opções de teste para modelos que não sejam de procariotos foram removidas
SeqKit (SHEN <i>et al.</i> , 2016)	Remoção de pseudogenes	Procedimento realizado somente em arquivos de CDS
Bio.Phylo (Talevich <i>et al.</i> 2012)		
RAxML 8.2.12 (STAMATAKIS, 2014)		
MrBayes (HUELSENBECK; RONQUIST, 2001)	Inferência filogenética	Suporte estatístico de 1000 replicatas ou equivalentes, quando aplicável. A análise Bayesiana foi feita com 1 milhão de gerações, amostradas a cada 100 passos. Foram utilizadas quatro correntes e valores padrão para os <i>priors</i> . 25% das árvores filogenéticas foram descartadas como <i>burn-in</i> .
PhyML 3.3 (GUINDON; GASCUEL, 2003)		

## 5. RESULTADOS

### 5.1 Desenvolvimento do *pipeline* para inferência filogenética

Inicialmente, foi desenvolvido o *pipeline* denominado Pipelution, como ferramenta para a análise evolutiva de genomas de organismos procarióticos. A seguir, serão apresentados os resultados de automação de cada etapa do processo, que se dá pela integração, com programação em Python, das ferramentas apresentadas na Tabela 1. O Pipelution utiliza as bibliotecas da linguagem Python para a manipulação dos dados e, com ele, é possível realizar toda a cadeia de processos que vão desde a inserção de dados inicial até a saída de arquivos como árvores filogenéticas.

O Pipelution foi arquitetado com paradigma *multi-thread* e, devido à abordagem metodológica implementada, pode gerar até 16 árvores filogenéticas a partir do mesmo conjunto de dados. Todas as combinações metodológicas são executadas para os genomas selecionados, gerando 16 árvores filogenéticas para validação e exploração do paradigma experimental. O fluxo de trabalho geral do Pipelution está ilustrado na Figura 2.



**Figura 2. Arquitetura geral do Pipelution.** Representação esquemática da integração das metodologias empregadas para construção do *pipeline* e do fluxo de dados. Trapézios representam entrada de dados manual; retângulos representam etapas de processamento; losangos representam pontos de decisão.

Os dados de entrada no *Pipelution* são os conjuntos completos de proteínas de cada organismo que se pretende estudar em formato fasta, daqui em diante referenciado como proteoma. Esses arquivos podem ser sequências de aminoácidos deduzidas das CDS ou proteínas em formato fasta; ambos os tipos de arquivos estão disponíveis no NCBI. Um arquivo contendo informações adicionais sobre as montagens dos genomas, o qual contém identificadores das montagens dos genomas e a quais espécies e linhagens elas pertencem também é necessário para o pré-processamento.

Com o objetivo de diminuir o tamanho do banco de dados e considerando que sequências marcadas como pseudogenes não são aceitas pelas ferramentas de análise evolutiva utilizadas, tais sequências foram removidas dos arquivos fastas. Caso essa informação esteja presente no arquivo de entrada do usuário, o *Pipelution* a remove automaticamente antes de iniciar a análise com a ferramenta SeqKit. Também foram automatizadas, com *scripts*, as alterações de nome de arquivo e de cabeçalhos.

A identificação dos ortólogos, primeiro passo para inferência filogenética, foi implementada no *Pipelution* através da inclusão do algoritmo OrthoFinder. Para as proteínas que não foram encontradas em cópia única nos GOs definidos pelo OrthoFinder, foi desenvolvido um módulo para identificação de ortólogos verdadeiros. Utilizando o programa Clustal Omega, o módulo desenvolvido constrói uma matriz de identidade do GO que contém cópias identificadas como potenciais ortólogos. Essas identidades foram comparadas com a identidade média das proteínas em cópia única do GO e a sequência de maior valor foi selecionada como a representativa do genoma. Todo esse processo de cálculo e seleção é feito automaticamente pelo *pipeline*.

Para o alinhamento dos GOs foram utilizados no *Pipelution* os programas PRANK e Clustal Omega, em suas configurações-padrão. Em seguida, os alinhamentos foram aparados pela ferramenta trimAl com remoção de até 20% da sequência para redução do ruído filogenético. Essas etapas de alinhamento por mais de um algoritmo e diminuição do ruído, entretanto, são opcionais para o usuário, podendo ser selecionado somente um alinhador dentre as opções e utilizar o alinhamento completo.



Para os métodos de inferência por distância e MV, as sequências alinhadas de todos os GOs são unidas em um único arquivo, denominado supermatriz. O melhor modelo evolutivo pode ser então calculado para todos os GOs através da ferramenta ModelTest-NG e selecionado de acordo com o AICc. Para o método de IB, o cálculo do modelo evolutivo é feito individualmente para cada GO, visto que essa ferramenta permite o particionamento dos dados de acordo com cada GO. Para entrada e saída dos dados entre essas ferramentas são necessárias várias adaptações nos arquivos e transformação dos dados, bem como seleção de informações. Através do Pipelution, por meio de *scripts*, todo esse processo foi automatizado, tornando-se imperceptível para o usuário.

Os grupos ortólogos processados na etapa anterior seguem para a etapa de inferência filogenética, que tem também como parâmetro de entrada os modelos evolutivos anteriormente calculados. Para a inferência filogenética, as ferramentas incluídas no *pipeline* e utilizadas foram a biblioteca Bio.Phylo, a ferramenta RAxML e o programa MrBayes, que aplicam os métodos de distância, MV e IB, respectivamente. Além disso, a inferência também é implementada com PhyML e suporte estatístico *SH-like*, por MV. A saída final do *Pipelution*, são as árvores filogenéticas geradas nos respectivos formatos por cada ferramenta de inferência.

## **5.2 Aplicação do Pipelution para o estudo filogenômico de micoplasmas**

Para os 106 genomas, 105 de micoplasmas e 1 do grupo externo, foram realizadas todas as combinações metodológicas possíveis no Pipelution. O resultado foram 16 árvores filogenéticas construídas a partir de 43 GOs de proteínas de cópia única. Informações sobre o tamanho de cada grupo ortólogo estão no Apêndice 2. Todas as árvores filogenéticas geradas pelos diferentes métodos foram comparadas entre si, levando em consideração o posicionamento das espécies em estudo (GUPTA *et al.*, 2018) e o suporte estatístico dos nós, visto que 80% foi considerado o mínimo para definição de um nó suportado. Todos os dados apresentados daqui em diante estão detalhados na Tabela 2.

**Tabela 2: Relação de nós com suporte inferior a 80% em todas as 16 árvores filogenéticas geradas no Pipelution para genomas de micoplasmas, considerando todas as combinações metodológicas.** Os resultados com o programa MEGA estão presentes somente para fins comparativos entre o método de distância e os demais métodos.

Método <sup>1</sup>	Alinhador	Alinhamento <sup>2</sup>	Modeltest-NG <sup>3</sup>	Suporte < 80% <sup>4</sup>	Suporte < 90% <sup>5</sup>	Árvore <sup>6</sup>
IB (Mr. Bayes)	Clustal Ω	Completo	OG	3	3	1
	Clustal Ω	Aparado	OG	4	4	2
	PRANK	Completo	OG	0	2	3*
	PRANK	Aparado	OG	0	2	4
MV (PhyML - SH-like)	Clustal Ω	Completo	Supermatriz	7	13	5
	Clustal Ω	Aparado	Supermatriz	7	10	6
	PRANK	Completo	Supermatriz	3	6	7
MV (RAxML)	PRANK	Aparado	Supermatriz	3	6	8
	Clustal Ω	Completo	Supermatriz	12	17	9
	Clustal Ω	Aparado	Supermatriz	11	18	10
	PRANK	Completo	Supermatriz	12	16	11
Distância (MEGA)	PRANK	Aparado	Supermatriz	10	16	12
	Clustal Ω	Completo	Supermatriz	10	11	13
	Clustal Ω	Aparado	Supermatriz	9	12	14
	PRANK	Completo	Supermatriz	12	13	15
Distância (Bio.Phylo)	PRANK	Aparado	Supermatriz	10	13	16
	Clustal Ω	Completo	Supermatriz	106	106	17
	Clustal Ω	Aparado	Supermatriz	106	106	18
	PRANK	Completo	Supermatriz	106	106	19
	PRANK	Aparado	Supermatriz	106	106	20

\*Árvore com melhor resultado em relação ao suporte e escolhida para as seguintes comparações.

<sup>1</sup> A Reconstrução filogenética pelo método de distância com a ferramenta MEGA foi realizada por NJ, distância p com 1000 replicatas considerando a Árvore Original.

<sup>2</sup> Alinhamento utilizado completo ou processado pelo algoritmo da ferramenta trimAl. O número de aminoácidos resultantes para o alinhamento com a ferramenta Clustal Ômega é de 15.808 e 13.049 aa completo e aparado, respectivamente. O número de aminoácidos resultantes para o alinhamento com a ferramenta PRANK é de 23.042 e 18.616 aa completo e aparado, respectivamente

<sup>3</sup> Modelo evolutivo calculado para supermatriz ou para cada grupo de ortólogos.

<sup>4</sup> Número de nós com suporte inferior a 80% para cada combinação metodológica.

<sup>5</sup> Número de nós com suporte inferior a 90% para cada combinação metodológica.

<sup>6</sup> Identificação de cada árvore gerada.

As árvores filogenéticas construídas com o programa Mr. Bayes e de alinhamento com a ferramenta PRANK obtiveram suporte para todos os ramos, com o alinhamento completo e aparado (árvores 3 e 4). As árvores filogenéticas construídas com o programa Mr. Bayes e alinhamento realizado com a ferramenta Clustal Ômega apresentaram 3 e 4 bifurcações com suporte inferior a 80% para alinhamento completo e aparado, respectivamente (árvores 1 e 2). As árvores

filogenéticas de inferência bayesiana obtiveram um total de 7 ramos sem suporte. Ao considerar 90% como suporte mínimo, 11 bifurcações não foram suportadas.

A inferência por MV com a ferramenta PhyML apresentou um total de 20 ramos sem suporte. Com o alinhador PRANK, cada árvore filogenética teve 3 ramos sem suporte com o alinhamento aparado e completo (árvores 7 e 8). Com o alinhador Clustal Ômega, as árvores filogenéticas inferidas com a ferramenta PhyML tiveram 7 ramos sem suporte cada, considerando alinhamento completo e aparado (árvores 5 e 6). Ao considerar 90% como suporte mínimo, 35 bifurcações não foram suportadas.

Outra forma de inferência por MV, com a ferramenta RAxML, obteve um total de 45 ramos sem suporte. Destes ramos, 23 são de árvores inferidas utilizando o alinhador Clustal Ômega (árvores 9 e 10) e 22 utilizando o alinhador PRANK (árvores 11 e 12). Não houve diferença maior que 2 ramos sem suporte quando comparadas as árvores que utilizam alinhamento completo e aparado. Os ramos com suporte inferior a 90% com a ferramenta RAxML totalizam 67.

O método de distância, com inferência pelo programa MEGA (TAMURA *et al.*, 2013), apresentou um total de 41 clados com suporte inferior a 80%. Quando foi utilizado o pacote Clustal Ômega para realizar o alinhamento, as árvores tiveram 10 e 9 ramos sem suporte (árvores 13 e 14) com alinhamento completo e aparado, respectivamente. As árvores filogenéticas inferidas com a ferramenta MEGA e utilizando o alinhador PRANK obtiveram 12 e 10 ramos sem suporte para alinhamento completo e aparado, respectivamente (árvores 15 e 16). Considerando 90% como suporte mínimo, o método de distância apresentou 50 ramos sem suporte.

O método de distância, implementado no *pipeline* com a biblioteca Bio.Phylo obteve suporte inferior a 80% para todos os clados de todas as árvores com ele inferidas (árvores 17, 18, 19 e 20). As árvores 17, 18, 19 e 20 foram, portanto, desconsideradas. Nenhuma árvore filogenética inferida pelo *pipeline* obteve o mesmo resultado de suporte e topologia. O suporte dos ramos das árvores filogenéticas inferidas a partir do alinhamento com a ferramenta PRANK e com o pacote Clustal Ômega tiveram uma diferença de 13 ramos sem suporte (50

para PRANK e 63 para Clustal Ômega). Quando 90% foi considerado como suporte mínimo, a diferença foi de 15 (74 para PRANK e 89 para Clustal Ômega).

De todas as 16 árvores filogenéticas consideradas (árvores filogenéticas de número 1 a 16), as inferidas a partir dos alinhamentos completos, com ambos os alinhadores, apresentaram 59 ramos sem suporte. As árvores filogenéticas inferidas a partir do alinhamento aparado com a ferramenta trimAl apresentaram 54 ramos sem suporte. O número total de ramos identificados sem suporte nas 16 árvores filogenéticas geradas é de 113. Todas as 20 árvores geradas, 16 através do *pipeline* e 4 com o programa MEGA, estão no Apêndice 3.

Os 113 ramos sem suporte identificados nas 16 árvores filogenéticas estão distribuídos conforme indicado na Figura 3. Destes 113 ramos, 63 podem estar associados a diferenças topológicas entre as árvores, tópico doravante explorado. Em relação aos clados que apresentaram suporte em todas as árvores estão os agrupamentos formados por *Candidatus Mycoplasma haemohominis*, *Mycoplasma amphoriforme*, *Mycoplasma marinum* e *Mycoplasma todarodis*, *Mycoplasma neurolyticum*, *Mycoplasma subdolum*, *Mycoplasma phocidea* e *Mycoplasma equirhinis*, *Mycoplasma struthionis*, *Mycoplasma falconis*, *Mycoplasma neophronis*, *Mycoplasma testudineum*, *Mycoplasma caviae*, *Mycoplasma maculosum*, *Mycoplasma columbinasale*, *Mycoplasma pullorum*, *Mycoplasma anserisalpingitis*, *Mycoplasma citelli*, *Mycoplasma gallopavonis*, *Mycoplasma nasistruthionis*, *Mycoplasma mustelae*, *Mycoplasma bovirhinis* e *Mycoplasma edwardii*. As espécies supracitadas agruparam com suporte com as mesmas OTUS em todas as 16 árvores.

O clado II do grupo Hominis também foi representado com suporte em todas as árvores, com a mesma topologia em seus nós mais terminais. *Mycoplasma mobile* foi encontrado tanto mais basal em relação aos clados I e II como mais basal dentro do clado II. Entretanto, seu posicionamento dentro do clado II foi encontrado sem suporte em 3 árvores filogenéticas (árvores 13, 14 e 16). Informações adicionais sobre os resultados de topologia e suporte podem ser encontradas no Apêndice 4.

Algumas vezes também foram observados ramos sem suporte mas com topologia semelhante ao que está descrito no estudo referência (GUPTA *et al.*, 2018). O clado A7, por exemplo, foi encontrado sem suporte em 7 árvores filogenéticas, mas a topologia de suas OTUs é sempre a mesma e está de acordo com o esperado.



Candidatus *Mycoplasma haemolamae* e Candidatus *Mycoplasma haemominutum*. Árvores com e sem suporte nesses ramos apresentam ambas as espécies agrupadas ou com posicionamentos mais basais. O mesmo ocorre nos clados U4, W8 e X4, onde *Mycoplasma verecundum* e *Mycoplasma nasistruthionis* seguidamente são agrupados com *Mycoplasma felis* e *Mycoplasma mustalae* (árvores 1, 2, 3, 7, 8, 11 e 12) e com *Mycoplasma buteonis* e *Mycoplasma glycyphilum* (árvores 4, 13, 14, 15 e 16).

As 16 árvores filogenéticas apresentaram ramos com e sem suporte para bifurcações que envolvem *Mycoplasma imitans* e *Mycoplasma gallisepticum* (clados D1 e E5). Das 16 árvores em comparação, somente a árvore filogenética 9 apresentou *M. imitans* e *M. gallisepticum* como grupo mais basal em relação a *Mycoplasma genitalium* e *Mycoplasma pneumoniae*, em uma bifurcação sem suporte. O posicionamento de *Mycoplasma mucosicanis* também foi encontrado diversas vezes em bifurcações sem suporte. Os clados S9 e T2 indicam que 12 vezes o posicionamento de *M. mucosicanis* não obteve suporte.

Os nós mais terminais do clado III não possuem suporte em diversas árvores filogenéticas (bifurcações I1, H4, J11, K4, M8, L3 e N1). O posicionamento de *Mycoplasma arthridis*, por exemplo, está associado a ramos que não foram suportados em no mínimo 11 árvores. Contraditoriamente, *M. arthridis* está posicionado em ramos suportados nas árvores inferidas pelo método de distância. Estes dois posicionamentos são, entretanto, diferentes do estudo referência e possuem muitas alterações topológicas internas no clado III. Também no clado III, *Mycoplasma arginini* não possui suporte em 8 árvores (5, 6, 7, 8, 9, 10, 11, 12), quando agrupado mais próximo de *Mycoplasma phocicerebrale*.

O posicionamento da espécie *Mycoplasma synoviae* também foi diferente com o método de distância. As árvores de número 1 a 12 apresentaram *M. synoviae* como grupo irmão de *M. gallinaceum* e *M. verecundum*. As árvores inferidas com a ferramenta MEGA (13, 14, 15 e 16) apresentaram *M. synoviae* mais basal em relação a *Mycoplasma crocodyli* e *Mycoplasma alligatoris*.

Dos resultados obtidos, 24 ramos não possuem suporte em no mínimo uma árvore filogenética (representados nos círculos pretos da Figura 3). A reconstrução filogenética por inferência bayesiana foi a metodologia que obteve os melhores suportes estatísticos. A inferência com MV através da ferramenta PhyML com suporte *SH-like* obteve no máximo 7 clados sem suporte. A inferência por MV com a ferramenta RAxML apresentou resultados inferiores, com até 12 ramos sem suporte.

O resultado com o método de distância utilizando o programa MEGA foi similar à inferência com a ferramenta RAxML, em relação ao suporte. O método de distância se diferencia, entretanto, ao conseguir representar de forma mais coerente as divisões internas do grupo Hominis, como a ordem dos clados III, II e I. Ao analisar suporte e topologia das árvores geradas, é possível afirmar que apesar de ter apresentado suportes mais baixos, o método de distância foi o que trouxe topologias mais semelhantes ao observado na literatura.

Adicionalmente, ao avaliar o suporte estatístico de modo mais estrito, ou seja, considerando 90% como suporte mínimo, a IB e o método de distância, árvores de número 1 a 4 e árvores de número 13 a 16, respectivamente, tiveram um aumento de 1 a até 4 ramos sem suporte. As árvores filogenéticas inferidas por MV, árvores de número 5 a 12, apresentaram o maior aumento de ramos não suportados, variando de 3 a 7.

### **5.3 Desempenho**

A entrada para os alinhadores Clustal Ômega e PRANK foram 1.261.591 aminoácidos distribuídos em 43 arquivos (Apêndice 2). O programa Clustal Ômega, utilizando uma unidade de processamento, precisou de 3,6 min para completar o alinhamento. Quando configurado para utilizar toda a capacidade do computador, precisou de 1,5 min. Utilizando toda a capacidade de processamento do computador, o alinhador PRANK precisou de 24 min. O tempo de processamento da ferramenta trimAl foi de 1 min, em média.

As árvores filogenéticas inferidas pelo método de distância com o programa MEGA demoraram aproximadamente 3 min cada. As árvores inferidas



pelo método de distância adaptadas para utilizar todas as unidades de processamento com a biblioteca Bio.Phylo levaram, em média, 20 h. As árvores filogenéticas inferidas por MV, utilizando a capacidade máxima do computador, levaram em média 3 h com o algoritmo do programa PhyML e em média 6 dias com o algoritmo do programa RAxML. Por fim, o programa Mr. Bayes, em sua versão que permite multiprocessamento, precisou em média 3 dias para alcançar o valor de parada configurado ( $stopval = 0.01$ ).

O tempo necessário em relação à manipulação de dados interna do *pipeline*, como mudança de nomes, configuração de arquivos e entrada e saída de dados entre as ferramentas é em média de 3 min.

## 6. DISCUSSÃO

### 6.1 Paradigma experimental do *pipeline*

O *pipeline* recebe o nome Pipelution por fazer alusão aos conceitos *pipeline*, *evolution* e *solution*. Todos os termos supracitados estão relacionados ao objetivo de auxiliar a entender questões evolutivas de procariotos englobando diversas soluções de análises filogenéticas. Ele foi construído com ferramentas não comerciais, de uso livre e é projetado para ambientes Linux, o qual já é sistema nativo de ambientes e ferramentas de análise filogenética.

*Scripts* modulares de análise filogenética, desenvolvidos ou disponibilizados individualmente, normalmente carecem de automatização. A automatização é essencial para agilizar o processo e diminuir atividades de gerenciamento em cada etapa do fluxo dos dados (LEIPZIG, 2017; WRATTEN; WILM; GÖKE, 2021). O desenvolvimento do Pipelution é feito considerando o objetivo de produzir diversas árvores filogenéticas de forma automatizada, incorporando ferramentas que tenham como saída dados que são a entrada para a ferramenta seguinte.

O trimAl é um exemplo de algoritmo que, se incorporado em um *pipeline*, pode ser automatizado através de parâmetros de entrada pré-estabelecidos. É importante adicionar ferramentas assim pois elas se propõem a remover ruído filogenético do alinhamento, aumentando as chances de obter árvores filogenéticas com suporte estatístico mais alto. Utilizando parâmetros pré-estabelecidos de remoção de ruído, torna-se facultativa a escolha desse parâmetro por parte do usuário e surge a possibilidade de aplicar tais padrões, ou identificar padrões melhores, para outros níveis taxonômicos.

De forma complementar, a automatização na transformação dos arquivos de entrada e saída entre as ferramentas é de valor imensurável, pois conforme aumenta o volume de dados, aumenta também a dificuldade de realizar tais tarefas manualmente. O tempo informado para processamento interno dos dados, 3 min, é irrisório se comparado com o tempo que seria necessário para realizar tais atividades manualmente em uma abordagem filogenômica.

Evitando etapas manuais e considerando as características das ferramentas utilizadas, o desenvolvimento do *pipeline* foi feito objetivando otimizar os processos e usar por padrão todo recurso disponível para trabalho. A elaboração de paralelização no próprio *pipeline* para a execução dos alinhadores permitiu uma redução de 2 min no tempo total de alinhamento. Apesar de não ser um grande ganho de tempo, a elaboração desse método mostra-se relevante considerando análises com maior número de dados, como genomas maiores de procariotos.

A paralelização também foi realizada com o método de reconstrução filogenética por distância, utilizando a biblioteca Bio.Phylo. Esse método, como explicitado nos resultados, não foi capaz de reproduzir de forma coerente os grupos mais clássicos do gênero *Mycoplasma*. Isso sugere ser necessário explorar outros métodos de construção de árvores por distância ou aprimorar o método atual.

Além dos módulos de paralelização de processos, também foi desenvolvido um módulo para escolha de proteínas que não sejam encontradas em cópia única em GOs. Esse módulo para identificação de ortólogos verdadeiros é importante por permitir inserir mais sequências na análise, as quais seriam descartadas. Talvez essa seja uma simplificação em demasia do problema de origem dessas proteínas, mas é também, entretanto, uma forma de aumentar o número de sequências utilizadas para inferência filogenética.

Em relação à escolha dos métodos de inferência implementados, duas ferramentas diferentes para MV foram selecionadas. A razão é devido a PhyML e RAxML conterem tipos diferentes de suporte estatístico, sendo o suporte SH-like (PhyML) calculado mais rapidamente e tão confiável quanto o uso de replicatas. A ferramenta Mr. Bayes oferece ainda um terceiro tipo de cálculo de suporte, chamado de probabilidade posterior. A escolha de dois alinhadores também é importante dada a proposta de oferecer diferentes combinações metodológicas. O alinhador Clustal Ômega é reconhecido por sua rapidez e o PRANK conhecido por seu viés evolutivo, sendo os dois ótimas opções para um *pipeline* de filogenômica.

Este último ponto é importante em relação aos alinhadores. Uma característica intrínseca da ferramenta PRANK é que ela é projetada para ser utilizada em análises filogenéticas. Na prática o que ocorre é a inserção de um número maior de *gaps* no alinhamento, visando manter bem alinhadas regiões conservadas e que representem homologia. Isso reflete no tamanho dos alinhamentos gerados: o alinhamento realizado com a ferramenta PRANK possui 7.234 posições a mais que o alinhamento realizado com o Clustal Ômega. Isso refletiu também nas 16 árvores filogenéticas inferidas, que tiveram diferença no número de clados com e sem suporte utilizando os dois alinhadores. O alinhador PRANK pode ter apresentado mais ramos com suporte justamente devido a seu viés evolutivo, melhor alinhando regiões de sinal filogenético forte.

## 6.2 Validação do Pipelution com micoplasmas

Durante o desenvolvimento deste *pipeline* houve a proposta de revisão da taxonomia de Tenericutes, grupo que envolve o gênero *Mycoplasma*, proposto como *Mesomycoplasma* (GUPTA *et al.*, 2018). A nova organização é aceita por uma parte da comunidade científica, mas o debate sobre a nomenclatura ainda está se desenvolvendo. Algumas formas intermediárias, por exemplo, são encontradas na literatura, como *Mycoplasma (Mesomycoplasma) hyorhinitis*, (KÄBISCH *et al.*, 2021). Apesar de a nova proposta ser mais coerente com características fenotípicas e genômicas das espécies, não há consenso sobre sua utilização e por isso optou-se por seguir a nomenclatura tradicional e mais referida (WEISBURG *et al.*, 1989).

No presente estudo, para análise de bactérias do gênero *Mycoplasma*, foi priorizada a seleção de montagens de genomas do banco RefSeq, que é um banco de dados não redundante, mais completo e curado, resultando em um conjunto de espécies representativo e, portanto, adequado para a validação do Pipelution (PHILIPPE *et al.*, 2011). A avaliação das árvores geradas com o *pipeline* foi feita com o estudo que utilizou maior número de genomas (80) de espécies de *Mycoplasma* (GUPTA *et al.*, 2018).

As árvores filogenéticas aqui utilizadas como referência foram 2, uma inferida com sequências concatenadas de 63 proteínas conservadas, identificadas com HMM a partir do perfil de marcadores do filo Tenericutes (WANG; WU, 2013). A segunda foi inferida a partir de sequências de 45 proteínas ribossomais conservadas. Ambas as árvores foram inferidas por MV com o programa RAXML, utilizando o suporte SH-*like*, com alinhamento aparado pelo trimAl e realizado com pacote Clustal Ômega (GUPTA *et al.*, 2018).

O critério de escolha da melhor filogenia gerada pelo *pipeline*, a ser comparada com as árvores referência, foi o suporte dos ramos, pois o suporte estatístico em si é um processo neutro que reflete apenas o sinal filogenético (ou ruído) nos dados, conforme detectado pelo método de construção de árvore usado (LEMEY; SALEMI; VANDAMME, 2009). Das 16 árvores geradas com o *pipeline*, a árvore por IB e com alinhamento completo pela ferramenta PRANK teve 80% ou mais de suporte para todos os ramos, sendo ela a escolhida (árvore 3). Todas as outras filogenias foram sistematicamente comparadas com a de número 3 e conseqüentemente avaliadas em relação à literatura.

Como explicado anteriormente, as árvores filogenéticas geradas pelo método de distância implementado no Pipelution com a biblioteca Bio.Phylo apresentaram os piores suportes e foram desconsideradas para análise (árvores 17, 18, 19 e 20). Mesmo ao intercalar os alinhadores e o nível de alteração no alinhamento (alinhamento completo e aparado), não houve melhora no suporte dessas árvores. Portanto, porque o método de distância do *pipeline* foi incapaz de solucionar a relação entre as OTUs, foi executada adicionalmente a inferência por distância com o programa MEGA, inferida à parte do *pipeline* e utilizada para fins comparativos. As árvores filogenéticas geradas com o MEGA (árvores 13, 14, 15 e 16) apresentaram as topologias dos grandes grupos e suportes acima de 80% e foram utilizadas nas comparações com as outras abordagens.

A escolha dos alinhadores não apresentou ser impactante na inferência filogenética de micoplasmas. As divergências identificadas em relação às árvores filogenéticas referência (GUPTA *et al.*, 2018) estão presentes nas árvores filogenéticas inferidas tanto a partir do alinhamento com o pacote Clustal Ômega, como a partir do alinhamento com o algoritmo do PRANK. Apesar de o alinhador

PRANK ter aumentado o número de ramos suportados, é importante observar que os ramos sem suporte não representam, necessariamente, alterações topológicas.

Uma grande melhoria entre as árvores filogenéticas também não é percebida quando a ferramenta trimAl é utilizada com a finalidade de aumentar o sinal filogenético. A escolha do método de inferência, entretanto, mostrou-se importante por trazer topologias diferentes para os grandes clados nas árvores filogenéticas geradas. O método de distância, que tende a não ser o melhor em solucionar relações entre espécies evolutivamente distantes, foi o mais eficaz em resolver a ordem dos clados internos do grupo Hominis.

Entre os métodos de inferência por MV, observa-se diferença significativa entre o número de ramos sem suporte ao utilizar a ferramenta PhyML e a ferramenta RAxML. Essa diferença, entretanto, não implicou em ramos não suportados afetando a topologia, ou seja, os ramos não suportados apresentam quase a mesma topologia de ramos suportados entre as árvores filogenéticas de números 5 a 12.

A análise do suporte dos ramos considerando a topologia levanta uma questão sobre os grandes clados do grupo Hominis. Os clados I, II e III dentro do grupo Hominis, possuem uma ordem de especiação onde o grupo III se apresenta como mais basal em relação ao II e I, sendo o clado I o mais distante. Das 16 árvores filogenéticas aqui inferidas e analisadas, somente 4, inferidas pelo método de distância, apresentam essa topologia vista na literatura. As bifurcações desses clados foram encontradas, entretanto, em ramos sem suporte estatístico, indicando que o valor mínimo de confiança estatística poderia ser diferente para o método de distância quando consideradas OTUs evolutivamente próximas, por exemplo.

O método de inferência por distância foi também o único capaz de agrupar *M. edwardii* com *Mycoplasma canis* (CHALKER; BROWNLIE, 2004) em todas as suas filogenias. Dentre as 4 filogenias construídas pelo método de distância estão 3 árvores filogenéticas que apresentam uma organização, com suporte estatístico, de acordo com o esperado dentro do clado I de Hominis, onde o clado de *M.*

*opalescens* é mais basal. Outro posicionamento que o método de distância apresenta de acordo com o esperado é o de *M. arthritidis*, que é mais basal em relação a *Mycoplasma cloacale*. No entanto, a resolução da topologia interna no clado I foi comprometida, onde bifurcações mais terminais são diferentes do estudo utilizado para comparação.

Apesar de o suporte representar a probabilidade de uma filogenia ser a verdadeira, ou ainda a probabilidade de o mesmo resultado ser encontrado utilizando outro conjunto de marcadores moleculares (HILLIS; BULL, 1993), é necessário reconhecer limitações técnicas que podem afetar a interpretação dos valores de suporte. Os tipos de erros que podem ser introduzidos pelo processo de inferência filogenética podem ser de dois tipos: associados à amostragem, denominados erros estocásticos que tendem a ser minimizados de acordo com o número de marcadores moleculares utilizados. E erros associados à natureza simples dos modelos evolutivos em comparação com a complexidade composicional das sequências, que são denominados erros sistemáticos e tendem a ser maximizados em grandes conjuntos de dados (PHILLIPS; DELSUC; PENNY, 2004).

A interpretação dos resultados considerando possíveis erros sistemáticos traz questionamentos sobre as inferências com MV e IB. Apesar de as árvores filogenéticas por IB terem todos os ramos com probabilidade posterior igual ou superior a 80% (0.8), a topologia nos ramos internos dos grupos III, II e I foi comprometida, bem como o posicionamento das OTUs supracitadas. Mesmo ao utilizar parâmetros específicos de modelo evolutivo para cada GO e código genético de *Mycoplasmas*, configurações permitidas somente pelo programa Mr. Bayes, não foi observada nas árvores filogenéticas inferidas por IB uma topologia mais semelhante às filogenias referidas .

Há ainda outra problemática em relação à IB, visto que é recomendado utilizar valores de no mínimo 90% de probabilidade posterior para esse método (ANISIMOVA *et al.*, 2011) e ele possui a tendência de superestimar a confiança do nó (SUZUKI; GLAZKO; NEI, 2002). Sendo assim, considerando 95% (0.95) para análise de suporte estatístico, as árvores filogenéticas 1, 2, 3 e 4 tiveram 6, 4, 2 e 3 ramos com confiança inferior a 95%, respectivamente. Desses ramos, há

somente uma sobreposição entre as 4 filogenias, que envolve as bifurcações R8, S9 e T2, também não suportadas em diversas outras árvores filogenéticas (Figura 3). Entretanto, nenhum desses clados sem suporte estatístico está associado às bifurcações dos grupos III, II e I, que são topologicamente diferentes das árvores utilizadas como referência. Isso pode indicar que os grupos III, II e I possuem um padrão distinto de especiação que viola as premissas dos modelos evolutivos utilizados, resultando em um sinal filogenético não adequado para ser resolvido por métodos mais complexos, como IB e MV.

Como decorrência, as maiores diferenças topológicas e de suporte entre as ferramentas utilizadas foram observadas comparando o método de distância e os demais, IB e MV. E, apesar das divergências entre as filogenias aqui inferidas e as filogenias utilizadas como referência, todos os métodos considerados conseguiram reproduzir os grupos Mycoides, Pneumoniae e Hominis, característicos do gênero *Mycoplasma*. A seguir serão descritas abordagens que poderiam contribuir na compreensão dos resultados gerados através do *pipeline*.

A ferramenta OrthoFinder, baseada em busca local, de viés evolutivo e utilizada no Pipelution, identificou 43 grupos ortólogos. As duas árvores filogenéticas referenciadas utilizaram 63 e 45 proteínas cada, identificadas com heurísticas diferentes do OrthoFinder. Isso significa que com os métodos de inferência aqui explorados foi possível reproduzir o mesmo esqueleto da filogenia de micoplasmas inferida com marcadores específicos do gênero. Mostra-se pertinente, portanto, realizar uma comparação futura com as proteínas integrantes dos 43 GOs e as proteínas utilizadas no estudo de referência para identificar possíveis sobreposições e explorar o conjunto de sequências utilizadas.

Adicionalmente, seria interessante realizar uma nova inferência com as mesmas configurações da árvore de número 3, alterando o conjunto de dados. Essa alteração consistiria em aumentar o número de GOs incluídos selecionando grupos de proteínas que não são encontradas em cópia única. Para tal, fazer uso do módulo de escolha de ortólogos verdadeiros, desenvolvido no Pipelution. Considerando as filogenias já obtidas e a árvore filogenética gerada a partir dessa proposta, seria possível analisar se há aumento ou diminuição do suporte estatístico e mudança na topologia entre as filogenias. Essa inferência seria



também uma primeira forma de teste do método de escolha de ortólogo verdadeiro aqui desenvolvido.

Também pode ser realizada a inclusão de sequências genômicas de espécies que não estavam disponíveis durante a execução deste trabalho, bem como a inclusão de grupos irmãos ao gênero, que auxiliariam na validação do *pipeline*, devido a recente revisão da taxonomia do gênero *Mycoplasma* e de seus grupos irmãos. De forma complementar, mais comparações, como cálculos de distância, podem ser executadas entre as 16 filogenias inferidas. Robinson Foulds e outras métricas de análise de comprimento de ramos seriam relevantes para melhor entender as diferenças entre as metodologias aplicadas.

Por fim, mostra-se pertinente executar uma abordagem que contemple a complexidade dos eventos evolutivos e que pondere a especificidade de hospedeiros de micoplasmas. Assim sendo, é importante considerar a investigação de TGH e co-evolução entre espécies do gênero *Mycoplasma* e seus respectivos hospedeiros. Já existem estudos similares com algumas espécies ou *clusters* de espécies (FISCHER *et al.*, 2012; TOOMER *et al.*, 2015; DESIRÒ *et al.*, 2018; ELMNASRI *et al.*, 2018; RASMUSSEN *et al.*, 2022), mas uma análise que utilize a sequência genômica de diversas espécies de micoplasma e seus hospedeiros seria de grande valia para entender a ecologia e evolução do gênero.

## 7. CONCLUSÕES

- I. Bactérias do gênero *Mycoplasma* possuem três grandes grupos bem definidos, os quais foram identificados nas filogenias inferidas por diferentes metodologias no *pipeline*.
- II. Dentre as muitas ferramentas que fazem parte da reconstrução filogenética, o método de inferência demonstrou ser o mais responsável por resultados divergentes em relação a topologia e suporte estatístico.
- III. O método de distância foi o mais eficiente em solucionar a ordem das relações entre grupos mais internos do gênero *Mycoplasma*.
- IV. A automatização de processos com uso de *pipelines* permite a inclusão de muitos dados, bem como o processamento facilitado deles, o que é muito vantajoso para a filogenômica.
- V. O desenvolvimento do Pipelution foi validado com sequências de genomas de *Mycoplasma*, demonstrando ser competente no objetivo proposto.
- VI. A escolha de métodos a serem utilizados para reconstrução filogenética com o pipeline pode ser guiada por um conhecimento prévio do grupo taxonômico ou pode ser com viés exploratório dos dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABDELHAY KAOUD, Hussien. Introductory Chapter: Bacterial Cattle Diseases - Economic Impact and Their Control. *Em*: ABDEL HAY EL-SAYED KAOUD, Hussein (org.). Bacterial Cattle Diseases. IntechOpen, 2019.

ABRAHAMIAN, Peter *et al.* Molecular Epidemiology of *Xanthomonas perforans* Outbreaks in Tomato Plants from Transplant to Field as Determined by Single-Nucleotide Polymorphism Analysis. *Applied and environmental microbiology*, v. 85, n. 18, 2019. Disponível em: <http://dx.doi.org/10.1128/AEM.01220-19>.

ANISIMOVA, Maria *et al.* Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology*, v. 60, n. 5, p. 685–699, 2011.

BAICHO, Shakuntala; OUZOUNIS, Christos A. Computational complexity of algorithms for sequence comparison, short-read assembly and genome alignment. *Bio Systems*, v. 156-157, p. 72–85, 2017.

BALISH, Mitchell *et al.* Recommended rejection of the names *Malacoplasma* gen. nov., *Mesomycoplasma* gen. nov., *Metamycoplasma* gen. nov., *Metamycoplasmataceae* fam. nov., *Mycoplasmoidaceae* fam. nov., *Mycoplasmoidales* ord. nov., *Mycoplasmoides* gen. nov., *Mycoplasmopsis* gen. nov. [Gupta, Sawnani, Adeolu, Alnajar and Oren 2018] and all proposed species comb. nov. placed therein. *International journal of systematic and evolutionary microbiology*, v. 69, n. 11, p. 3650–3653, 2019.

BARATE, Abhijit K. *et al.* Immunogenicity of IMS 1113 plus soluble subunit and chimeric proteins containing *Mycoplasma hyopneumoniae* P97 C-terminal repeat regions. *FEMS microbiology letters*, v. 352, n. 2, p. 213–220, 2014.

BARYKOVA, Yulia A. *et al.* Association of *Mycoplasma hominis* infection with prostate cancer. *Oncotarget*, v. 2, n. 4, p. 289–297, 2011.

BERGERON, Nadia *et al.* Prevalence of *Mycoplasma synoviae* and Its Impact on Productivity in Commercial Poultry Farms in Quebec, Canada. *Avian diseases*, v. 65, n. 4, p. 546–552, 2021.

BEYLEFELD, Amanda *et al.* Evidence for Multidrug Resistance in Nonpathogenic *Mycoplasma* Species Isolated from South African Poultry. *Applied and environmental microbiology*, v. 84, n. 21, 2018. Disponível em: <http://dx.doi.org/10.1128/AEM.01660-18>.

BONNEAUD, Camille *et al.* Rapid Antagonistic Coevolution in an Emerging Pathogen and Its Vertebrate Host. *Current biology: CB*, v. 28, n. 18, p. 2978–2983.e5, 2018.

BORTOLOTTI, Fernanda *et al.* Oral mucosa infection by *Mycoplasma salivarium* in a patient with chronic graft-versus-host disease: A diagnostic challenge. *Hematology, transfusion and cell therapy*, 2020. Disponível em:

<http://dx.doi.org/10.1016/j.htct.2020.08.002>.

BRETON, Marc *et al.* Distribution and diversity of mycoplasma plasmids: lessons from cryptic genetic elements. *BMC microbiology*, v. 12, p. 257, 2012.

BROWN, Terence A. *How Genomes Evolve*. Wiley-Liss, 2002. *E-book*.

BROWN, Daniel R. Phylum XVI. Tenericutes Murray 1984a, 356VP (Effective publication: Murray 1984b, 33.). *Em*: KRIEG, Noel R. *et al.* (org.). *Bergey's Manual® of Systematic Bacteriology: Volume Four The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*. New York, NY: Springer New York, 2010. p. 567–723.

CALCUTT, M. J. *et al.* Gap analysis of *Mycoplasma bovis* disease, diagnosis and control: An aid to identify future development requirements. *Transboundary and emerging diseases*, v. 65 Suppl 1, p. 91–109, 2018.

CAPELLA-GUTIÉRREZ, Salvador; SILLA-MARTÍNEZ, José M.; GABALDÓN, Toni. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, v. 25, n. 15, p. 1972–1973, 2009.

CATTANI, A. M. Elementos repetitivos na regulação da transcrição de *Mycoplasma hyopneumoniae*. [s. l.], 2016. Disponível em: <https://www.lume.ufrgs.br/handle/10183/143890>.

CHALKER, Victoria J.; BROWNLIE, Joe. Taxonomy of the canine Mollicutes by 16S rRNA gene and 16S/23S rRNA intergenic spacer region sequence comparison. *International journal of systematic and evolutionary microbiology*, v. 54, n. Pt 2, p. 537–542, 2004.

CHRISTO, Paulo Pereira *et al.* Rhombencephalitis possibly caused by *Mycoplasma pneumoniae*. *Arquivos de neuro-psiquiatria*, v. 68, n. 4, p. 656–658, 2010.

CITTI, Christine *et al.* Genomic Islands in Mycoplasmas. *Genes*, v. 11, n. 8, 2020. Disponível em: <http://dx.doi.org/10.3390/genes11080836>.

CITTI, C. *et al.* Horizontal Gene Transfers in Mycoplasmas (Mollicutes). *Current issues in molecular biology*, v. 29, p. 3–22, 2018.

CRUZ, Laura De la *et al.* Unraveling the Global Phylodynamic and Phylogeographic Expansion of *Mycoplasma gallisepticum*: Understanding the Origin and Expansion of This Pathogen in Ecuador. [S. l.: s. n.], 2020. Disponível em: <http://dx.doi.org/10.3390/pathogens9090674>.

CURRAT, Mathias *et al.* Forward-in-Time, Spatially Explicit Modeling Software to Simulate Genetic Lineages Under Selection. *Evolutionary bioinformatics online*, v. 11, n. Suppl 2, p. 27–39, 2015.

DARRIBA, Diego *et al.* ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular biology and evolution*, v. 37, n. 1, p. 291–294, 2020.

DAVIES, Nicholas G. *et al.* Modeling the effect of vaccination on selection for antibiotic resistance in *Streptococcus pneumoniae*. *Science translational medicine*, v. 13, n. 606, p. eaaz8690, 2021.

DAWOOD, Ali *et al.* Mycoplasmas as Host Pantropic and Specific Pathogens: Clinical Implications, Gene Transfer, Virulence Factors, and Future Perspectives. *Frontiers in cellular and infection microbiology*, v. 12, p. 855731, 2022.

DELANEY, Nigel F. *et al.* Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS genetics*, v. 8, n. 2, p. e1002511, 2012.

DELSUC, Frédéric; BRINKMANN, Henner; PHILIPPE, Hervé. Phylogenomics and the reconstruction of the tree of life. *Nature reviews. Genetics*, v. 6, n. 5, p. 361–375, 2005.

DESIRÒ, Alessandro *et al.* Mycoplasma-related endobacteria within *Mortierellomycotina* fungi: diversity, distribution and functional insights into their lifestyle. *The ISME journal*, v. 12, n. 7, p. 1743–1757, 2018.

DIDELOT, Xavier. Phylogenetic Methods for Genome-Wide Association Studies in Bacteria. *Methods in molecular biology*, v. 2242, p. 205–220, 2021.

DOMMARAJU, Kalpana. Computational Identification of Viral/Bacterial Epitopes for Type and Strain Characterization, Leading to the Development of Vaccines Against Human Adenovirus, HIV, and *Staphylococcus Aureus*. 2021. - George Mason University, 2021. Disponível em: <https://search.proquest.com/openview/e0ff3c67dd2736fd2a14f6f45151f2e5/1?pq-origsite=gscholar&cbl=18750&diss=y>.

DORDET-FRISONI, Emilie *et al.* Chromosomal transfers in mycoplasmas: when minimal genomes go mobile. *mBio*, v. 5, n. 6, p. e01958, 2014.

DORDET-FRISONI, Emilie *et al.* Mycoplasma Chromosomal Transfer: A Distributive, Conjugative Process Creating an Infinite Variety of Mosaic Genomes. *Frontiers in microbiology*, v. 10, p. 2441, 2019.

DUDEK, Katarzyna *et al.* Mycoplasma bovis Infections-Occurrence, Diagnosis and Control. *Pathogens*, v. 9, n. 8, 2020. Disponível em: <http://dx.doi.org/10.3390/pathogens9080640>.

EISEN, Jonathan A.; FRASER, Claire M. Phylogenomics: intersection of evolution and genomics. *Science*, v. 300, n. 5626, p. 1706–1707, 2003.

ELMNASRI, Khaled *et al.* Highly divergent Mollicutes symbionts coexist in the scorpion *Androctonus australis*. *Journal of basic microbiology*, v. 58, n. 10, p. 827–835, 2018.

EMMS, David M.; KELLY, Steven. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*, v. 16, p. 157, 2015.

FAUCHER, Marion *et al.* Mycoplasmas under experimental antimicrobial selection: The unpredicted contribution of horizontal chromosomal transfer. *PLoS genetics*, v. 15, n. 1, p. e1007910, 2019.

FEDERHEN, Scott. The NCBI Taxonomy database. *Nucleic acids research*, v. 40, n. Database issue, p. D136–D143, 2012.

FISCHER, Anne *et al.* The Origin of the “Mycoplasma mycoides Cluster” Coincides with Domestication of Ruminants. [*S. l.: s. n.*], 2012. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0036150>.

GABALDÓN, Toni. Evolution of proteins and proteomes: a phylogenetics approach. *Evolutionary bioinformatics online*, v. 1, p. 51–61, 2007.

GUIMARAES, Ana M. S. *et al.* Comparative genomics and phylogenomics of hemotrophic mycoplasmas. *PloS one*, v. 9, n. 3, p. e91445, 2014.

GUINDON, Stéphane; GASCUEL, Olivier. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, v. 52, n. 5, p. 696–704, 2003.

GUPTA, Radhey S. *et al.* Phylogenetic framework for the phylum Tenericutes based on genome sequence data: proposal for the creation of a new order Mycoplasmoidales ord. nov., containing two new families Mycoplasmoidaceae fam. nov. and Metamycoplasmataceae fam. nov. harbouring Eperythrozoon, Ureaplasma and five novel genera. *Antonie van Leeuwenhoek*, v. 111, n. 9, p. 1583–1630, 2018.

GUPTA, Radhey S.; SON, Jeen; OREN, Aharon. A phylogenomic and molecular markers based taxonomic framework for members of the order Entomoplasmatales: proposal for an emended order Mycoplasmatales containing the family Spiroplasmataceae and emended family Mycoplasmataceae comprised of six genera. *Antonie van Leeuwenhoek*, 2018. Disponível em: <http://dx.doi.org/10.1007/s10482-018-1188-4>.

HANNAN, P. C. Comparative susceptibilities of various AIDS-associated and human urogenital tract mycoplasmas and strains of *Mycoplasma pneumoniae* to 10 classes of antimicrobial agent in vitro. *Journal of medical microbiology*, v. 47, n. 12, p. 1115–1122, 1998.

HASIN, Yehudit; SELDIN, Marcus; LUSIS, Aldons. Multi-omics approaches to disease. *Genome biology*, v. 18, n. 1, p. 83, 2017.

HILLIS, David M.; BULL, James J. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic biology*, v. 42, n. 2, p. 182–192, 1993.

HODGE, T.; COPE, M. J. A myosin family tree. *Journal of cell science*, v. 113 Pt 19, p. 3353–3354, 2000.

HOLST, Sam; YESKE, Paul; PIETERS, Maria. Elimination of *Mycoplasma hyopneumoniae* from breed-to-wean farms: A review of current protocols with emphasis on herd closure and medication. *Journal of Swine Health and Production*, v. 23, n. 6, p. 321–330, 2015.

HOLT, Kathryn E. *et al.* *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nature genetics*, v. 44, n. 9, p. 1056–1059, 2012.

HSIEH, Yu-Chia *et al.* Global Genome Diversity and Recombination in *Mycoplasma pneumoniae*. *Emerging infectious diseases*, v. 28, n. 1, p. 111–117, 2022.

HUELSENBECK, J. P.; RONQUIST, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, v. 17, n. 8, p. 754–755, 2001.

INAMINE, J. M. *et al.* Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. *Journal of bacteriology*, v. 172, n. 1, p. 504–506, 1990.

IPOUTCHA, Thomas *et al.* Multiple Origins and Specific Evolution of CRISPR/Cas9 Systems in Minimal Bacteria (Mollicutes). *Frontiers in microbiology*, v. 10, p. 2701, 2019.

ITAN, Yuval; GERBAULT, Pascale; PINES, Gur. Evolutionary Genomics: Supplement Aims and Scope. *Evolutionary bioinformatics online*, v. 11s2, p. EBO.S39729, 2015.

J, Venter Craig *et al.* The Sequence of the Human Genome. *Science*, v. 291, n. 5507, p. 1304–1351, 2001.

KÄBISCH, Lisa *et al.* Whole-Genome Sequence of the *Mycoplasma* (*Mesomycoplasma*) *hyorhinis* DSM 25591 Type Strain. *Microbiology resource announcements*, v. 10, n. 16, 2021. Disponível em: <http://dx.doi.org/10.1128/MRA.00164-21>.

KAMMINGA, Tjerko *et al.* Persistence of Functional Protein Domains in *Mycoplasma* Species and their Role in Host Specificity and Synthetic Minimal Life. *Frontiers in cellular and infection microbiology*, v. 7, p. 31, 2017.

KLENK, H-P; GÖKER, M. En route to a genome-based classification of Archaea and Bacteria?. *Systematic and applied microbiology*, v. 33, n. 4, p. 175–182, 2010.

LAING, Chad R.; WHITESIDE, Matthew D.; GANNON, Victor P. J. Pan-genome Analyses of the Species *Salmonella enterica*, and Identification of Genomic Markers Predictive for Species, Subspecies, and Serovar. *Frontiers in microbiology*, v. 8, p. 1345, 2017.

LANDER, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature*, v. 409, n. 6822, p. 860–921, 2001.

LECLERCQ, Sébastien *et al.* Phylogenomics of “Candidatus Hepatoplasma crinochetorum,” a lineage of mollicutes associated with noninsect arthropods. *Genome biology and evolution*, v. 6, n. 2, p. 407–415, 2014.

LEI, Ning *et al.* Phylogeny and expression pattern analysis of TCP transcription factors in cassava seedlings exposed to cold and/or drought stress. *Scientific reports*, v. 7, n. 1, p. 10016, 2017.

LEIPZIG, Jeremy. A review of bioinformatic pipeline frameworks. *Briefings in bioinformatics*, v. 18, n. 3, p. 530–536, 2017.

LEMEY, Philippe; SALEMI, Marco; VANDAMME, Anne-Mieke. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, 2009.

LIU, Wei *et al.* Comparative genomics of *Mycoplasma*: analysis of conserved essential genes and diversity of the pan-genome. *PloS one*, v. 7, n. 4, p. e35698, 2012.

LIU, L. *et al.* GAA trinucleotide repeat region regulates M9/pMGA gene expression in *Mycoplasma gallisepticum*. *Infection and immunity*, v. 68, n. 2, p. 871–876, 2000.

LIU, Aaron C. *et al.* Septic arthritis due to *Mycoplasma orale* in a young patient with hypogammaglobulinemia. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada*, v. 6, n. 4, p. 333–336, 2021.

LJUBIN-STERNAK, Sunčanica; MEŠTROVIĆ, Tomislav. Chlamydia trachomatis and Genital Mycoplasmas: Pathogens with an Impact on Human Reproductive Health. *Journal of pathogens*, v. 2014, p. 183167, 2014.

LO, Wen-Sui; GASPARICH, Gail E.; KUO, Chih-Horng. Convergent Evolution among Ruminant-Pathogenic *Mycoplasma* Involved Extensive Gene Content Changes. *Genome biology and evolution*, v. 10, n. 8, p. 2130–2139, 2018.

LORETO, E.; FERREIRA, H. B. Genes e genomas procarióticos. *Em: ZAHA, A.; FERREIRA, H. B.; PASSAGLIA, Lmp (org.). Biologia Molecular Básica*. Artmed, 2014. p. 58–82.

LÖYTYNOJA, Ari; GOLDMAN, Nick. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, v. 320, n. 5883, p. 1632–1635, 2008.

LYSNYANSKY I; ROSENGARTEN R; YOGEV D. Phenotypic switching of variable surface lipoproteins in *Mycoplasma bovis* involves high-frequency chromosomal rearrangements. *Journal of bacteriology*, v. 178, n. 18, p. 5395–5401, 1996.

MA, Liang *et al.* Genetic variation in the complete MgPa operon and its repetitive



chromosomal elements in clinical strains of *Mycoplasma genitalium*. PloS one, v. 5, n. 12, p. e15660, 2010.

MACHADO, Lais Del Prá Netto *et al.* Evidences of differential endoproteolytic processing on the surfaces of *Mycoplasma hyopneumoniae* and *Mycoplasma flocculare*. Microbial pathogenesis, v. 140, p. 103958, 2020.

MAES, D. *et al.* Update on *Mycoplasma hyopneumoniae* infections in pigs: Knowledge gaps for improved disease control. Transboundary and emerging diseases, v. 65 Suppl 1, p. 110–124, 2018.

MAKIMURA, K. *et al.* Phylogenetic classification and species identification of dermatophyte strains based on DNA sequences of nuclear ribosomal internal transcribed spacer 1 regions. Journal of clinical microbiology, v. 37, n. 4, p. 920–924, 1999.

MANILOFF, J. The minimal cell genome: “on being the right size”. Proceedings of the National Academy of Sciences of the United States of America, v. 93, n. 19, p. 10004–10006, 1996.

MARGARITA, Valentina; FIORI, Pier Luigi; RAPPELLI, Paola. Impact of Symbiosis Between *Trichomonas vaginalis* and *Mycoplasma hominis* on Vaginal Dysbiosis: A Mini Review. Frontiers in cellular and infection microbiology, v. 10, p. 179, 2020.

MARTINEZ-URTAZA, Jaime *et al.* Genomic Variation and Evolution of *Vibrio parahaemolyticus* ST36 over the Course of a Transcontinental Epidemic Expansion. mBio, v. 8, n. 6, 2017. Disponível em: <http://dx.doi.org/10.1128/mBio.01425-17>.

MORA-SÁNCHEZ, Brenda; PÉREZ-SÁNCHEZ, Tania; BALCÁZAR, José Luis. Phylogenetic analysis of intestinal microbiota reveals novel *Mycoplasma* phylotypes in salmonid species. Microbial pathogenesis, v. 145, p. 104210, 2020.

MOROWITZ, Harold J.; WALLACE, Douglas C. Genome size and life cycle of the *Mycoplasma*. Annals of the New York Academy of Sciences, v. 225, n. 1 *Mycoplasma an*, p. 62–73, 1973.

MUSATOVOVA, Oxana; KANNAN, T. R.; BASEMAN, Joel B. *Mycoplasma pneumoniae* large DNA repetitive elements RepMP1 show type specific organization among strains. PloS one, v. 7, n. 10, p. e47625, 2012.

MUTO, A.; OSAWA, S. The guanine and cytosine content of genomic DNA and bacterial evolution. Proceedings of the National Academy of Sciences of the United States of America, v. 84, n. 1, p. 166–169, 1987.

NAPIERALA MAVEDZENGE, Sue; WEISS, Helen Anne. Association of *Mycoplasma genitalium* and HIV infection: a systematic review and meta-analysis. AIDS, v. 23, n. 5, p. 611–620, 2009.

NICHOLAS, Robin; AYLING, Roger; MCAULIFFE, Laura. *Mycoplasma* Diseases of Ruminants. CABI, 2008.

PARTE, Aidan C. LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International journal of systematic and evolutionary microbiology*, v. 68, n. 6, p. 1825–1829, 2018.

PEREYRE, Sabine; TARDY, Florence. Integrating the Human and Animal Sides of Mycoplasmas Resistance to Antimicrobials. *Antibiotics* (Basel, Switzerland), v. 10, n. 10, 2021. Disponível em: <http://dx.doi.org/10.3390/antibiotics10101216>.

PHILIPPE, Hervé *et al.* Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, v. 9, n. 3, p. e1000602, 2011.

PHILLIPS, Matthew J.; DELSUC, Frédéric; PENNY, David. Genome-scale phylogeny and the detection of systematic biases. *Molecular biology and evolution*, v. 21, n. 7, p. 1455–1458, 2004.

PITCHER, D. G.; NICHOLAS, R. A. J. Mycoplasma host specificity: fact or fiction?. *Veterinary journal*, v. 170, n. 3, p. 300–306, 2005.

RAMÍREZ, Ana S. *et al.* First description of two moderately halophilic and psychrotolerant Mycoplasma species isolated from cephalopods and proposal of Mycoplasma marinum sp. nov. and Mycoplasma todarodis sp. nov. *Systematic and applied microbiology*, v. 42, n. 4, p. 457–467, 2019.

RASMUSSEN, Jacob *et al.* Hologenomic insights demonstrate co-evolution between an intestinal Mycoplasma and its salmonid host. 1 jul. 2022. Disponível em: <https://www.researchsquare.com/article/rs-1814912/latest.pdf>.

RAZIN, Shmuel. The genus Mycoplasma and related genera (class Mollicutes). *The prokaryotes*, v. 4, p. 836–904, 2006.

RAZIN, Shmuel; HAYFLICK, Leonard. Highlights of mycoplasma research—An historical perspective. *Biologicals: journal of the International Association of Biological Standardization*, v. 38, n. 2, p. 183–190, 2010.

RAZIN, S.; YOGEV, D.; NAOT, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiology and molecular biology reviews: MMBR*, v. 62, n. 4, p. 1094–1156, 1998.

ROCHA, Eduardo P. C.; BLANCHARD, Alain. Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. *Nucleic acids research*, v. 30, n. 9, p. 2031–2042, 2002.

ROKAS, Antonis *et al.* Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, v. 425, n. 6960, p. 798–804, 2003.

ROSALES, Rubén S. *et al.* Mycoplasmas: Brain invaders?. *Research in veterinary science*, v. 113, p. 56–61, 2017.

SASAKI, Yuko *et al.* The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans. *Nucleic acids research*, v. 30, n. 23, p. 5293–5300, 2002.

SCHREIBER, Peter W.; MÜLLER, Nicolas. Mycoplasma and Ureaplasma Infections in Transplantation: A Big Impact Despite the Lack of a Cell Wall. *Em: MORRIS, Michele I.; KOTTON, Camille Nelson; WOLFE, Camille (org.). Emerging Transplant Infections: Clinical Challenges and Implications. Cham: Springer International Publishing, 2020. p. 1–14.*

SHARP, P. M. *et al.* Codon usage: mutational bias, translational selection, or both?. *Biochemical Society transactions*, v. 21, n. 4, p. 835–841, 1993.

SHEN, Wei *et al.* SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS one*, v. 11, n. 10, p. e0163962, 2016.

SIEVERS, Fabian *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, v. 7, p. 539, 2011.

SIMMONS WARREN L.; DENISON AMY M.; DYBVIG KEVIN. Resistance of Mycoplasma pulmonis to Complement Lysis Is Dependent on the Number of Vsa Tandem Repeats: Shield Hypothesis. *Infection and immunity*, v. 72, n. 12, p. 6846–6851, 2004.

SIRAND-PUGNET, Pascal *et al.* Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS genetics*, v. 3, n. 5, p. e75, 2007.

SPERGSER, Joachim *et al.* Mycoplasma mucosicanis sp. nov., isolated from the mucosa of dogs. *International journal of systematic and evolutionary microbiology*, v. 61, n. Pt 4, p. 716–721, 2011.

SPERGSER, Joachim *et al.* Mycoplasma nasistruthionis sp. nov. and Mycoplasma struthionis sp. nov. isolated from ostriches with respiratory disease. *Systematic and applied microbiology*, v. 43, n. 1, p. 126047, 2020.

SPERGSER, Joachim *et al.* Mycoplasma tauri sp. nov. isolated from the bovine genital tract. *Systematic and applied microbiology*, v. 45, n. 1, p. 126292, 2022.

STALEY, Molly *et al.* Bacterial Pathogen Emergence Requires More than Direct Contact with a Novel Passerine Host. *Infection and immunity*, v. 86, n. 3, 2018. Disponível em: <http://dx.doi.org/10.1128/IAI.00863-17>.

STAMATAKIS, Alexandros. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, v. 30, n. 9, p. 1312–1313, 2014.

SUZUKI, Yoshiyuki; GLAZKO, Galina V.; NEI, Masatoshi. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America*, v. 99, n. 25, p. 16138–16143, 2002.

TAMURA, Koichiro *et al.* MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, v. 30, n. 12, p. 2725–2729, 2013.

TARDY, Florence *et al.* Mycoplasma bovis in Nordic European Countries: Emergence and Dominance of a New Clone. *Pathogens*, v. 9, n. 11, 2020. Disponível em: <http://dx.doi.org/10.3390/pathogens9110875>.

THOMPSON, Cristiane C. *et al.* Microbial taxonomy in the post-genomic era: rebuilding from scratch?. *Archives of microbiology*, v. 197, n. 3, p. 359–370, 2015.

TOOMER, Kevin H. *et al.* Molecular evolution patterns reveal life history features of mycoplasma-related endobacteria associated with arbuscular mycorrhizal fungi. *Molecular ecology*, v. 24, n. 13, p. 3485–3500, 2015.

TOPRAK, Erdal *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature genetics*, v. 44, n. 1, p. 101–105, 2011.

TSIODRAS, S. *et al.* Central nervous system manifestations of Mycoplasma pneumoniae infections. *The Journal of infection*, v. 51, n. 5, p. 343–354, 2005.

VAN ROSSUM, G. Python Programming language. *Proceedings of the USENIX ... annual Technical Conference*. USENIX Technical Conference, 2007. Disponível em: [https://www.wikizero.com/en/Python\\_\(language\)](https://www.wikizero.com/en/Python_(language)).

WANG, R. Y. *et al.* High frequency of antibodies to Mycoplasma penetrans in HIV-infected patients. *The Lancet*, v. 340, n. 8831, p. 1312–1316, 1992.

WANG, Zhang; WU, Martin. A phylum-level bacterial phylogenetic marker database. *Molecular biology and evolution*, v. 30, n. 6, p. 1258–1262, 2013.

WEISBURG, W. G. *et al.* A phylogenetic analysis of the mycoplasmas: basis for their classification. *Journal of bacteriology*, v. 171, n. 12, p. 6455–6467, 1989.

WOESE, C. R.; MANILOFF, J.; ZABLEN, L. B. Phylogenetic analysis of the mycoplasmas. *Proceedings of the National Academy of Sciences of the United States of America*, v. 77, n. 1, p. 494–498, 1980.

WRATTEN, Laura; WILM, Andreas; GÖKE, Jonathan. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nature methods*, v. 18, n. 10, p. 1161–1168, 2021.

XIE, Xiaohui *et al.* Microbial infection, inflammation and epithelial ovarian cancer. *Oncology letters*, v. 14, n. 2, p. 1911–1919, 2017.

YOGEV, D. *et al.* Sequence and TnpA analysis of a Mycoplasma hyorhinis protein with membrane export function. *Journal of bacteriology*, v. 173, n. 6, p. 2035–2044, 1991.

ZHANG, Yan-Cong; LIN, Kui. Phylogeny Inference of Closely Related Bacterial Genomes: Combining the Features of Both Overlapping Genes and Collinear Genomic Regions. *Evolutionary bioinformatics online*, v. 11, n. Suppl 2, p. 1–9, 2015.

ZHANG, Q.; WISE, K. S. Molecular basis of size and antigenic variation of a

*Mycoplasma hominis* adhesin encoded by divergent *vaa* genes. *Infection and immunity*, v. 64, n. 7, p. 2737–2744, 1996.

## ***CURRICULUM VITÆ* resumido**

VEDOVATTO, M.

(VEDOVATTO, M. M.)

### **1. DADOS PESSOAIS**

**Nome:** Meiski Mariá Vedovatto

**Local e data de nascimento:** São Leopoldo, Rio Grande do Sul, Brasil -  
23/10/1996

**Endereço profissional:** N/A

**Telefone profissional:** N/A      **E-mail:** meiskiv@gmail.com

### **2. FORMAÇÃO:**

- 2015 - 2018: Bacharel em Informática Biomédica, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA).
- 2019: Introdução a Análises Transcritômicas de Dados de RNA-seq. (Carga horária: 15h). UFRGS.
- 2019: V Curso de Verão em Biologia Evolutiva. (Carga horária: 30h). UFRGS.
- 2017 Extensão universitária em Modelagem Molecular Aplicada ao Desenvolvimento Racional de Fármacos. (Carga horária: 20h). UFRGS.
- 2017: Extensão universitária em Montagem e Anotação de Genomas. (Carga horária: 15h). UFRGS.
- 2017: Metagenômica - Capacitação em Bioinformática Aplicada à Saúde 2º Edição. (Carga horária: 10h). UFRGS.
- 2015: Extensão universitária em Introdução à Biologia de Sistemas. (Carga horária: 15h). UFRGS.

### 3. ESTÁGIOS

#### a. UFCSPA

- Estágio Curricular (2018): Desenvolvimento e implementação de *pipeline* para análise de dados genômicos de fungos patogênicos. UFRGS, Unidade de Biologia Teórico e Computacional. Responsável: Claudia Elizabeth Thompson
- Bolsista de Iniciação Científica (2018): Desenvolvimento de *scripts* e construção de um *pipeline* para análise evolutiva de procariotos. Manipulação e transformação dos dados, aquisição de dados de forma remota. Responsabilidade sobre a visualização e interpretação dos dados. Responsável: Claudia Elizabeth Thompson. Agência de fomento: PIC/UFCSPA
- Bolsista de Iniciação Científica (2017 - 2018): Desenvolvimento de *scripts* e execução de ferramentas de genômica comparativa. Auxílio no desenvolvimento de uma ferramenta de genômica comparativa em linguagem Python. Responsável: Adriana Seixas. Agência de fomento: PIC/UFCSPA
- Bolsista de Iniciação Científica (2015 - 2016): Prática experimental ao executar protocolos laboratoriais como PCR e atividades de rotina de laboratório. Responsável: Ana Beatriz Gorini da Veiga. Agência de fomento: PIC/UFCSPA

### 4. EXPERIÊNCIA PROFISSIONAL OU DIDÁTICA ANTERIOR

- 2021 - Atual: Carga horária: 44h semanais. Desenvolvimento e aprimoramento de *pipelines* para análises de microbioma. Agrega Pesquisa e Desenvolvimento em Biotecnologia Ltda.
- 2018 - Monitoria Voluntária, Carga horária: 10. Disciplina de Biologia Molecular para o curso de Informática Biomédica.
- 2018 - Monitoria Voluntária, Carga horária: 10. Disciplina de Aspectos de Genômica e Filogenética para o curso de Informática Biomédica.

## 5. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

- CASTRO, I. M. S. ; PRICHULA, J ; HERBERT, J. S. ; SEIXAS, A. ; VEDOVATTO, M. M. ; LUZ, R. C. Z. ; DAZEVEDO, P. ; FRAZZON, A. P. G. . Desenvolvimento De Uma Plataforma Web Para Caracterização Genômica De Cepas De Enterococcus Spp.. 2018. (Apresentação de Trabalho).
- VEDOVATTO, M. M.; PRICHULA, J ; CASTRO, I. M. S. ; LUZ, R. C. Z. ; FRAZZON, A. P. G. ; DAZEVEDO, P. ; SEIXAS, A. . Genômica Comparativa De Enterococcus Spp. Isolados De Fezes De Pinguins-de-magalhães Selvagens. 2018. (Apresentação de Trabalho).
- LUZ, R. C. Z. ; PRICHULA, J ; VEDOVATTO, M. M. ; CASTRO, I. M. S. ; FRAZZON, A. P. G. ; DAZEVEDO, P. ; SEIXAS, A. . Uma Visão Genômica De Isolados De Enterococcus Spp. Associados A Animais Marinhos Selvagens. 2018. (Apresentação de Trabalho).

VEIGA, A. B. G. ; VEDOVATTO, M. M. ; BELLO, A. G. D. ; HOCHHEGGER, B. ; PASQUALOTTO, A. C. ; AMANTEA, S. ; DAZEVEDO, P. . Estudo epidemiológico para a detecção molecular e análise filogenética de Bocavirus humano associado à gastroenterite aguda. 2016. (Apresentação de Trabalho - Congresso).



## APÊNDICE 1

**Tabela A3 - Informações sobre os genomas das espécies utilizadas no pipeline.** Cento e cinco genomas pertencem a espécies de micoplasma e 1 genoma pertence ao grupo externo. Tabela ordenada alfabeticamente.

N°	Espécie	ID linhagem	Montagem <sup>1</sup>	Nível
1	<i>Bacillus subtilis subsp. subtilis</i>	168	GCF_002009135.1	Completo
2	Candidatus <i>Mycoplasma girerdii</i>	VCU_M1	GCA_000770195.1	Completo
3	Candidatus <i>Mycoplasma haemobos</i>	INIFAP01	GCA_001645765.1	Contig
4	Candidatus <i>Mycoplasma haemolamae</i>	Purdue	GCA_000281235.1	Completo
5	Candidatus <i>Mycoplasma haemominutum</i>	Birmingham 1	GCA_000319365.1	Scaffold
6	Candidatus <i>Mycoplasma haemohominis</i>	SWG34-3	GCA_008326325.1	Contig
7	<i>Mycoplasma agalactiae</i>	PG2	GCA_000063605.1	Completo
8	<i>Mycoplasma agassizii</i>	PS6	GCA_002272945.1	Scaffold
9	<i>Mycoplasma alkalescens</i>	ATCC 29103	GCA_003208575.1	Scaffold
10	<i>Mycoplasma alligatoris</i>	A21JP2	GCA_000178375.1	Contig
11	<i>Mycoplasma alvi</i>	ATCC 29626	GCA_000701785.1	Scaffold
12	<i>Mycoplasma amphoriforme</i>	A39	GCA_000723365.1	Completo
13	<i>Mycoplasma anatis</i>	NCTC 10156	GCA_003285065.1	Completo
14	<i>Mycoplasma anseris</i>	ATCC 49234	GCA_003285045.1	Completo
15	<i>Mycoplasma arginini</i>	HAZ145_1	GCA_001547975.1	Completo
16	<i>Mycoplasma arthritidis</i>	158L3-1	GCA_000020065.1	Completo
17	<i>Mycoplasma auris</i>	ATCC 51348	GCA_003253435.1	Scaffold
18	<i>Mycoplasma bovigenitalium</i>	HAZ596	GCA_002356075.1	Completo
19	<i>Mycoplasma bovirhinis</i>	GS01	GCA_002688685.1	Completo
20	<i>Mycoplasma bovis</i>	PG45	GCA_000183385.1	Completo
21	<i>Mycoplasma bovoculi</i>	M165/69	GCA_000524555.1	Completo
22	<i>Mycoplasma buteonis</i>	ATCC 51371	GCA_000733865.1	Scaffold
23	<i>Mycoplasma californicum</i>	ST-6	GCA_000695835.1	Completo
24	<i>Mycoplasma canadense</i>	HAZ360_1	GCA_000828855.1	Completo
25	<i>Mycoplasma canis</i>	PG 14	GCA_001553195.1	Completo
26	<i>Mycoplasma capricolum subsp capricolum</i>	ATCC 27343	GCA_000012765.1	Completo
27	<i>Mycoplasma caviae</i>	NCTC10126	GCF_900631685.1	Contig
28	<i>Mycoplasma citelli</i>	NCTC10181	GCF_900660645.1	Completo
29	<i>Mycoplasma cloacale</i>	NCTC 10199	GCA_003269445.1	Completo
30	<i>Mycoplasma collis</i>	ATCC 35278	GCA_000701825.1	Scaffold
31	<i>Mycoplasma columbinasale</i>	NCTC10184	GCF_900660705.1	Completo

<sup>1</sup>De acordo com NCBI.

N°	Espécie	ID linhagem	Montagem <sup>1</sup>	Nível
32	<i>Mycoplasma columbinum</i>	ATCC 29257	GCA_000712175.1	Scaffold
33	<i>Mycoplasma columborale</i>	ATCC 29258	GCA_000701845.1	Scaffold
34	<i>Mycoplasma conjunctivae</i>	HRC/581T	GCA_000026765.1	Cromossomo
35	<i>Mycoplasma cricetuli</i>	ATCC 35279	GCA_000526955.1	Contig
36	<i>Mycoplasma crocodyli</i>	MP145	GCA_000025845.1	Completo
37	<i>Mycoplasma cynos</i>	C142	GCA_000328725.1	Completo
38	<i>Mycoplasma dispar</i>	ATCC 27140	GCA_000941075.1	Completo
39	<i>Mycoplasma edwardii</i>	NCTC10132	GCF_900476105.1	Completo
40	<i>Mycoplasma elephantis</i>	ATCC 51980	GCA_000687815.1	Contig
41	<i>Mycoplasma equirhinis</i>	ATCC 43428	GCF_006385185.1	Contig
42	<i>Mycoplasma felis</i>	ATCC 23391	GCA_000701865.1	Scaffold
43	<i>Mycoplasma feriruminatoris</i>	G5847	GCA_000327395.1	Scaffold
44	<i>Mycoplasma fermentans</i>	JER	GCA_000148625.1	Completo
45	<i>Mycoplasma fermentans</i>	M64	GCA_000186005.1	Completo
46	<i>Mycoplasma gallinaceum</i>	NCTC10183	GCF_900660495.1	Completo
47	<i>Mycoplasma gallinarum</i>	DSM 19816	GCA_000621085.1	Scaffold
48	<i>Mycoplasma gallisepticum</i>	R	GCA_000092585.1	Completo
49	<i>Mycoplasma gallopavonis</i>	NCTC10186	GCF_900660635.1	Completo
50	<i>Mycoplasma genitalium</i>	G-37	GCA_000027325.1	Completo
51	<i>Mycoplasma glycyphilum</i>	ATCC 35277	GCA_000687855.1	Contig
52	<i>Mycoplasma haemocanis</i>	Illinois	GCA_000238995.1	Completo
53	<i>Mycoplasma haemofelis</i>	Langford 1	GCA_000200735.1	Completo
54	<i>Mycoplasma hominis</i>	ATCC 23114	GCA_000085865.1	Completo
55	<i>Mycoplasma hyopneumoniae</i>	J	GCA_000008205.1	Completo
56	<i>Mycoplasma hyorhinis</i>	HUB-1	GCA_000145705.1	Completo
57	<i>Mycoplasma hyosynoviae</i>	NPL4	GCA_000691345.1	Contig
58	<i>Mycoplasma imitans</i>	ATCC 51306	GCA_000518305.1	Scaffold
59	<i>Mycoplasma iners</i>	ATCC 19705	GCA_000701805.1	Scaffold
60	<i>Mycoplasma iowae</i>	695	GCA_000227355.2	Contig
61	<i>Mycoplasma leachii</i>	PG50	GCA_000183365.1	Completo
62	<i>Mycoplasma leonicaptivi</i>	ATCC 49890	GCA_000622205.1	Scaffold
63	<i>Mycoplasma lipofaciens</i>	ATCC 35015	GCA_000686585.1	Scaffold
64	<i>Mycoplasma maculosum</i>	NCTC10168	GCF_900660665.1	Completo
65	<i>Mycoplasma marinum</i>	PE	GCF_004335975.1	Contig
66	<i>Mycoplasma meleagridis</i>	IZSVE/2944/9/2011	GCA_001484825.1	Scaffold
67	<i>Mycoplasma mobile</i>	163K	GCA_000008365.1	Completo
68	<i>Mycoplasma molare</i>	ATCC 27746	GCA_000622165.1	Scaffold
69	<i>Mycoplasma mustelae</i>	ATCC 35214	GCF_004365095.1	Scaffold
70	<i>Mycoplasma mycoides subsp capri</i>	95010	GCA_000253075.1	Scaffold

<sup>1</sup>De acordo com NCBI.

N°	Espécie	ID linhagem	Montagem <sup>1</sup>	Nível
71	<i>Mycoplasma neurolyticum</i>	NCTC10166	GCF_900660485.1	Completo
72	<i>Mycoplasma opalescens</i>	ATCC 27921	GCA_000712185.1	Contig
73	<i>Mycoplasma orale</i>	ATCC 23714	GCA_000420105.1	Scaffold
74	<i>Mycoplasma ovipneumoniae</i>	NCTC 10156	GCA_000753815.1	Completo
75	<i>Mycoplasma ovis</i>	Michigan	GCA_000508245.1	Completo
76	<i>Mycoplasma parvum</i>	Indiana	GCA_000477415.1	Completo
77	<i>Mycoplasma penetrans</i>	HF-2	GCA_000011225.1	Completo
78	<i>Mycoplasma phocicerebrale</i>	1049	GCF_003383595.3	Completo
79	<i>Mycoplasma phocidae</i>	105	GCF_003332325.1	Completo
80	<i>Mycoplasma phocirhinis</i>	852	GCF_004216495.1	Completo
81	<i>Mycoplasma pirum</i>	ATCC 25960	GCA_000685905.1	Completo
82	<i>Mycoplasma pneumoniae</i>	M129; ATCC 29342	GCA_000027345.1	Contig
83	<i>Mycoplasma primatum</i>	ATCC 25948	GCA_000702785.1	Completo
84	<i>Mycoplasma pullorum</i>	B359_6	GCA_001900245.1	Completo
85	<i>Mycoplasma pulmonis</i>	UAB CTIP	GCA_000195875.1	Completo
86	<i>Mycoplasma putrefaciens</i>	KS1	GCA_000224105.1	Contig
87	<i>Mycoplasma salivarium</i>	ATCC 23064	GCA_000485555.1	Contig
88	<i>Mycoplasma simbae</i>	ATCC 49888	GCA_000702705.1	Scaffold
89	<i>Mycoplasma spumans</i>	ATCC 19526	GCA_000620005.1	Scaffold
90	<i>Mycoplasma sturni</i>	DSM 22021	GCA_000701485.1	Contig
91	<i>Mycoplasma subdolum</i>	ATCC 29870	GCF_003688445.1	Completo
92	<i>Mycoplasma suis</i>	Illinois	GCA_000179035.2	Completo
93	<i>Mycoplasma synoviae</i>	53	GCA_000008245.1	Scaffold
94	<i>Mycoplasma testudineum</i>	BH29	GCA_002245785.1	Contig
95	<i>Mycoplasma testudinis</i>	ATCC 43263	GCA_000687795.1	Contig
96	<i>Mycoplasma todarodis</i>	5H	GCF_004335995.1	Contig
97	<i>Mycoplasma verecundum</i>	ATCC 27862	GCA_900167035.1	Completo
98	<i>Mycoplasma wenyonii</i>	Massachusetts	GCA_000277795.1	Completo
99	<i>Mycoplasma yeatsii</i>	GM274B	GCA_000875755.1	Completo
100	<i>Mycoplasma anserisalpingitis</i>	ATCC:BAA-2147	GCA_007859615.1	Completo
101	<i>Mycoplasma falconis</i>	ATCC 51372	GCF_006385795.1	Contig
102	<i>Mycoplasma felifaucium</i>	ATCC 43428	GCF_000687775.1	Contig
103	<i>Mycoplasma mucosicanis</i>	1642	GCF_006546935.1	Contig
104	<i>Mycoplasma nasistruthionis</i>	2F1A	GCF_006228185.1	Completo
105	<i>Mycoplasma neophronis</i>	24097	GCF_006491995.1	Contig
106	<i>Mycoplasma struthionis</i>	2371A	GCF_003855455.1	Completo

<sup>1</sup>De acordo com NCBI.

## APÊNDICE 2

**Tabela 4 - Grupos ortólogos identificados com a ferramenta OrthoFinder.** Cada GO possui 106 seqüências de aminoácidos que constituem um arquivo no formato FASTA.

GO	Tamanho (aa)			Total (aa)
	Mínimo	Médio	Máximo	
OG0000114	69	72,7	86	7710
OG0000115	151	220,4	256	23364
OG0000116	66	116,6	123	12364
OG0000117	129	132,4	142	14032
OG0000118	120	121,8	124	12913
OG0000119	80	89,6	100	9497
OG0000120	136	141,3	160	14980
OG0000121	216	235,4	417	24948
OG0000122	87	91,1	100	9655
OG0000123	277	281,3	287	29822
OG0000124	98	102,8	121	10900
OG0000126	321	332,9	355	35289
OG0000127	302	313,6	346	33237
OG0000128	300	339,3	361	35971
OG0000129	142	145,1	153	15377
OG0000130	129	133,4	150	14143
OG0000131	393	411,1	771	43576
OG0000135	65	86,8	90	9197
OG0000136	106	111,9	127	11863
OG0000137	250	260,6	292	27624
OG0000140	33	113,1	147	11988
OG0000141	110	149	281	15793
OG0000142	436	447,5	460	47431
OG0000144	137	165,7	204	17562
OG0000145	199	245,2	282	25987
OG0000146	682	889,7	921	94312
OG0000147	179	186,6	204	19784
OG0000148	188	205,5	299	21780
OG0000150	371	535,9	636	56807
OG0000151	567	602,8	716	63900
OG0000152	369	423,8	657	44926
OG0000155	360	377,1	657	39970
OG0000156	201	270	287	28618
OG0000158	502	524,4	598	55584
OG0000159	167	182,8	195	19372
OG0000160	226	263,1	314	27892
OG0000162	295	302,8	341	32098

GO	Tamanho (aa)			Total (aa)
	Mínimo	Médio	Máximo	
OG0000164	777	836,1	880	88628
OG0000165	680	694,3	702	73595
OG0000166	148	155,8	161	16518
OG0000169	314	336,5	361	35665
OG0000170	108	128	141	13565
OG0000171	44	126	162	13354

### APÊNDICE 3

Nesta seção são apresentadas todas as 16 árvores geradas com o Pipelution e as 4 árvores geradas com a ferramenta MEGA. As figuras estão numeradas de acordo com a Tabela 2.



**Figura A1. Árvore filogenética inferida por inferência bayesiana com o programa Mr. Bayes. O alinhamento foi realizado com o Clustal Ômega e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.**

Tree scale: 1

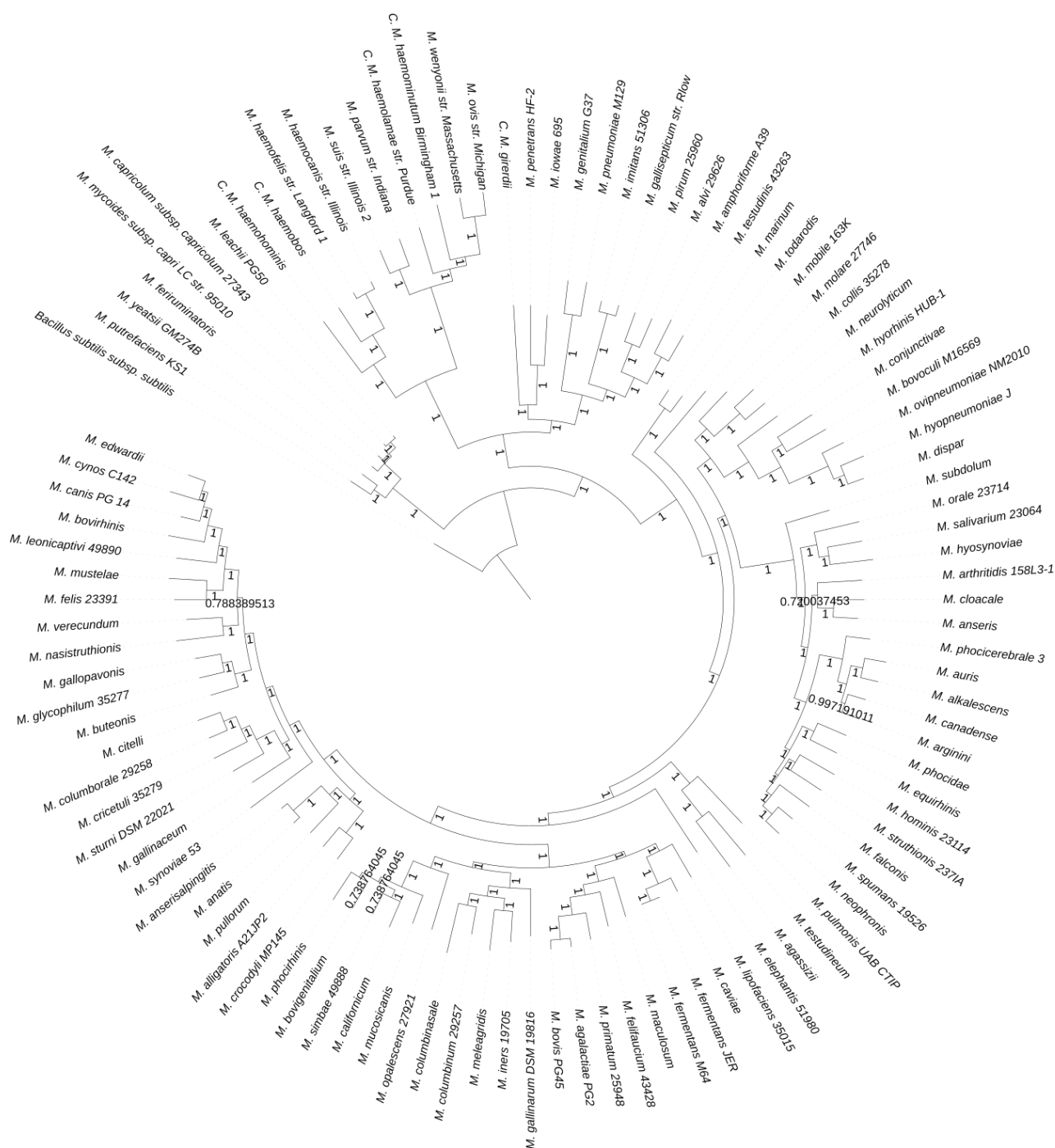
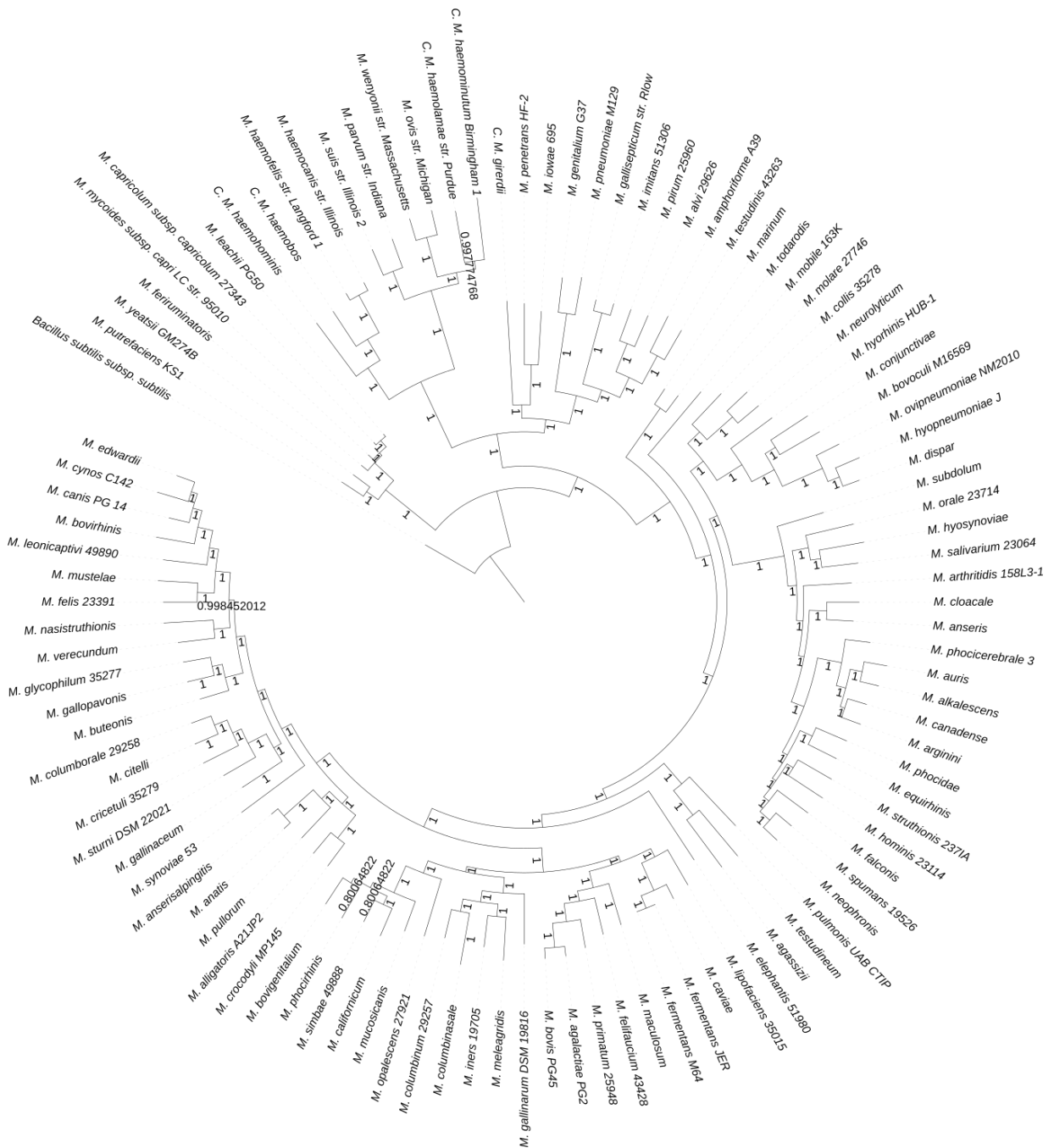


Figura A2. Árvore filogenética inferida por inferência bayesiana com o programa Mr. Bayes e alinhador Clustal Ômega. O alinhamento foi aparado com a ferramenta trimAl. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.

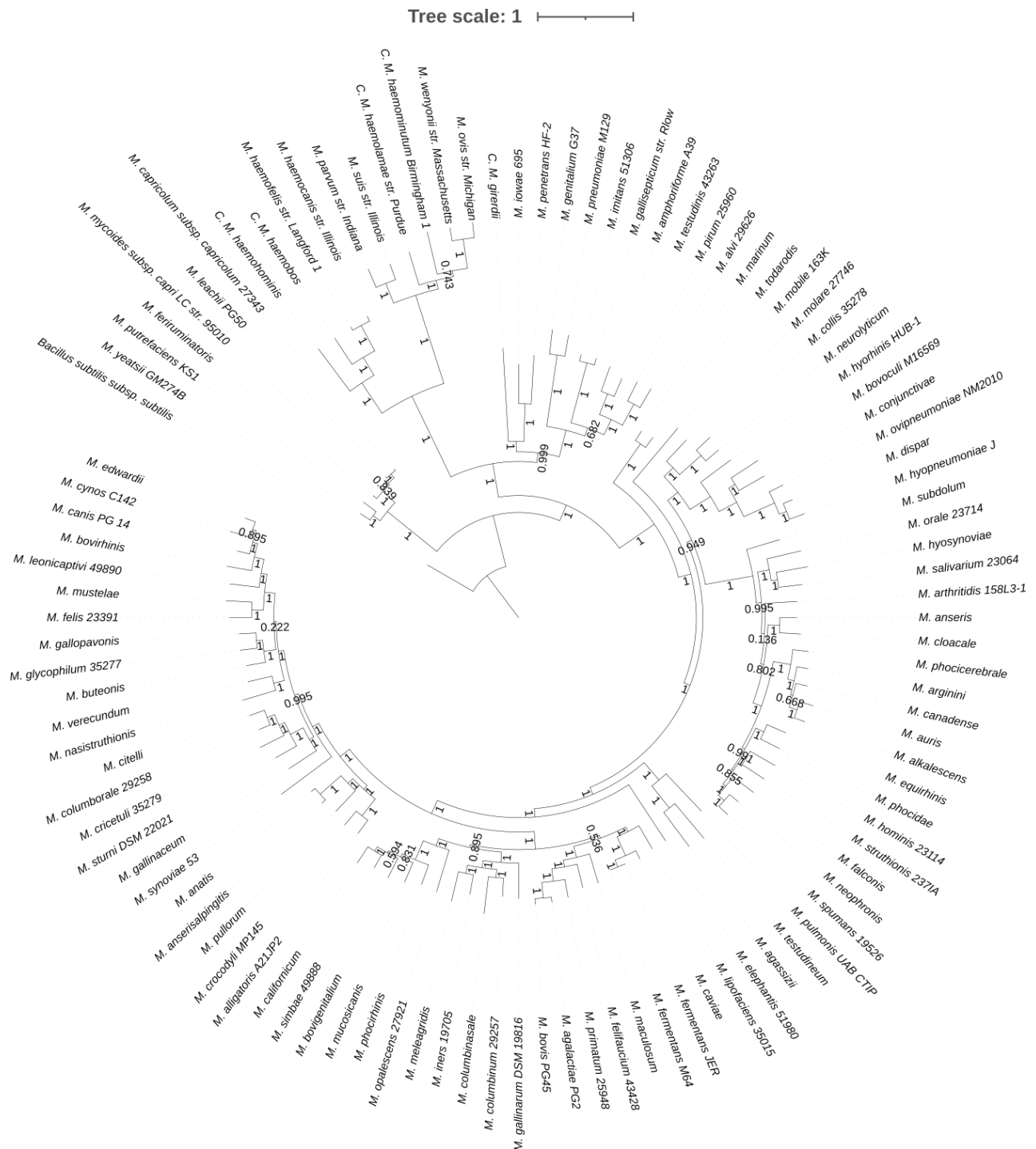
Tree scale: 1



**Figura A3. Árvore filogenética inferida por inferência bayesiana com o programa Mr. Bayes. O alinhamento foi realizado com o PRANK e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.**







**Figura A5.** Árvore filogenética inferida pelo método de máxima verossimilhança com a ferramenta PhyML. O alinhamento foi realizado com o Clustal Ômega e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.

Tree scale: 1

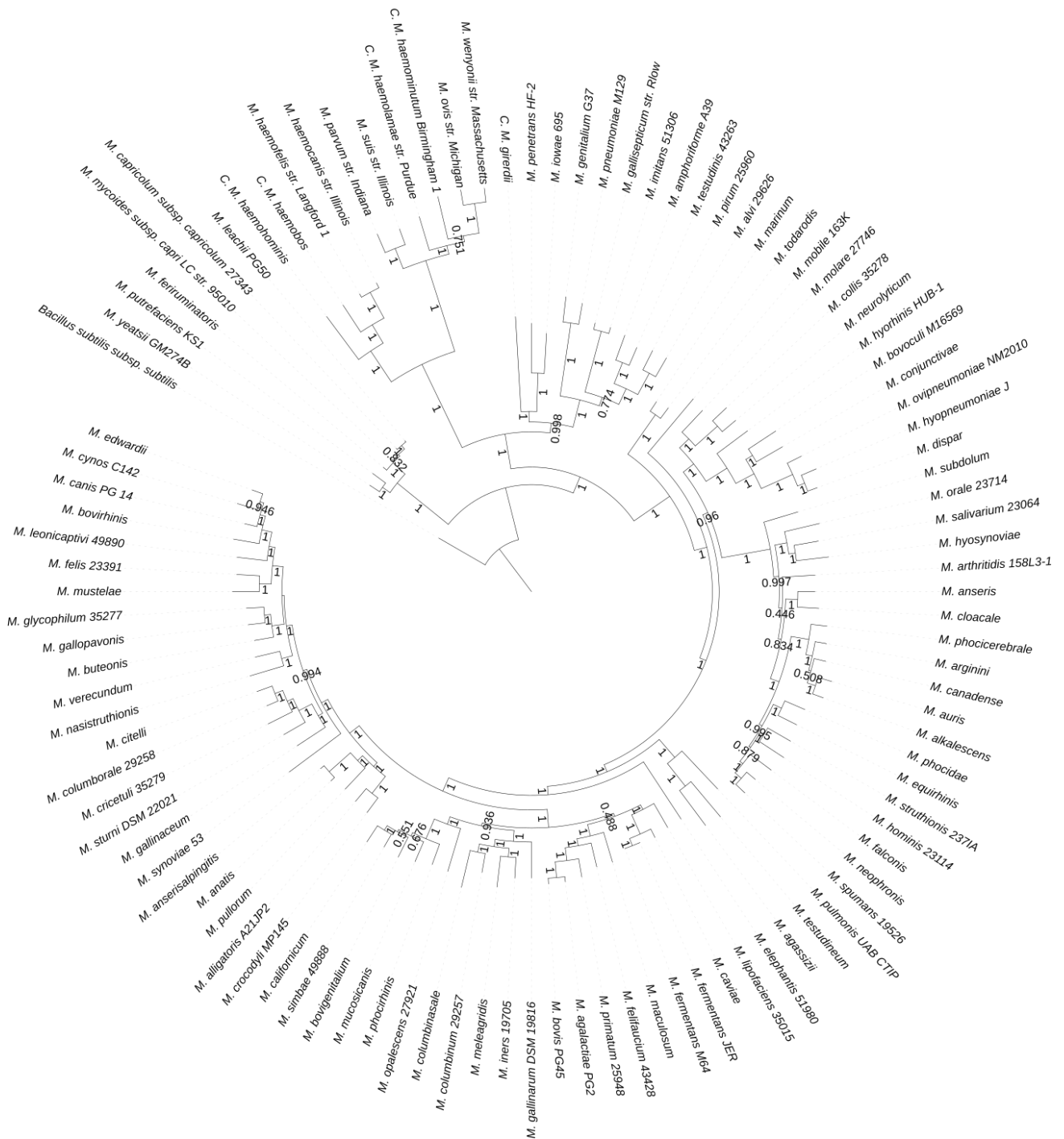
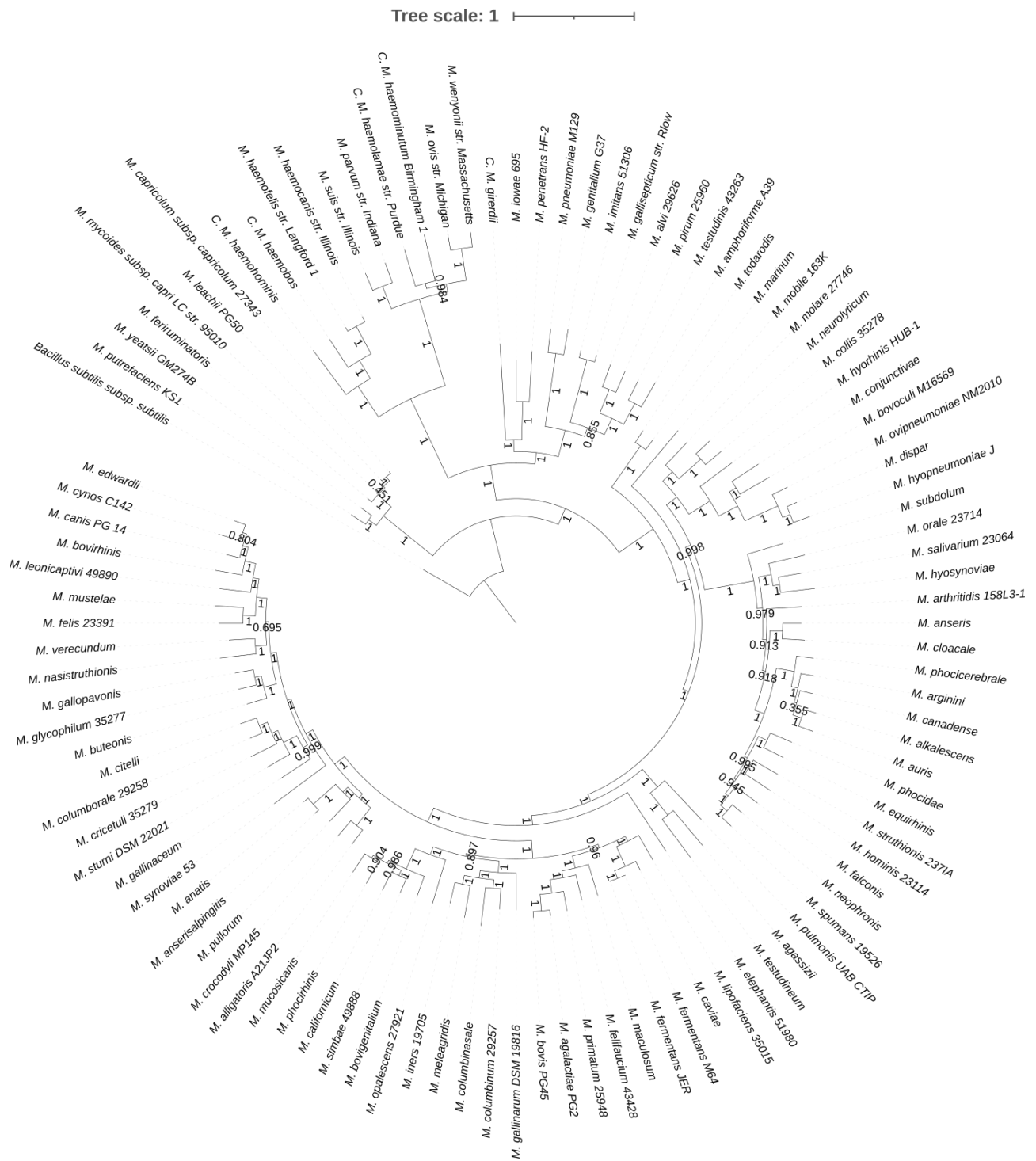


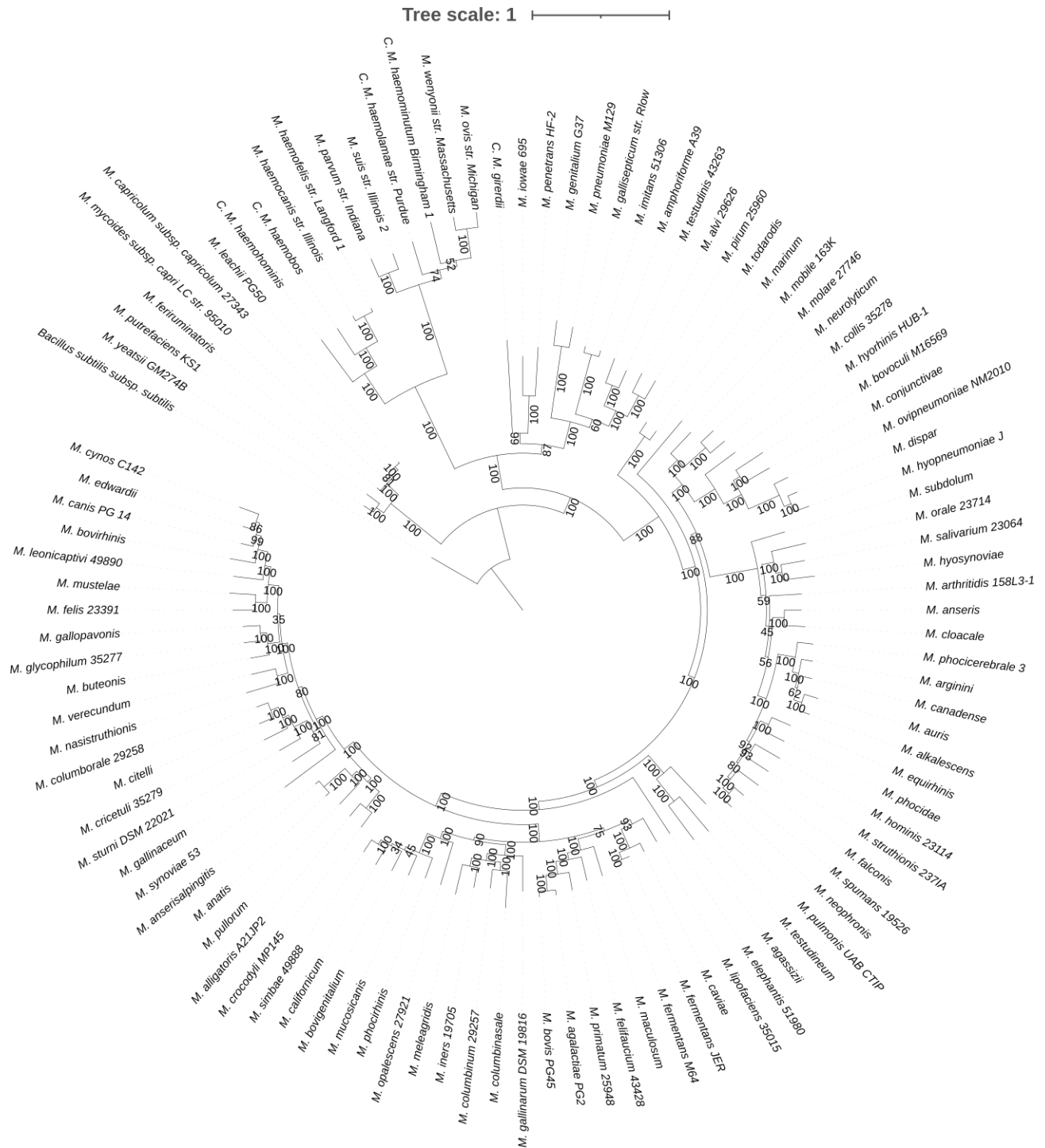
Figura A6. Árvore filogenética inferida pelo método de máxima verossimilhança com a ferramenta PhyML e alinhador Clustal Ômega. O alinhamento foi aparado com a ferramenta trimAl. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.





**Figura A8. Árvore filogenética inferida pelo método de máxima verossimilhança com a ferramenta PhyML e alinhador PRANK. O alinhamento foi aparado com a ferramenta trimAl. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.**





**Figura A10.** Árvore filogenética inferida pelo método de máxima verossimilhança com a ferramenta RAxML e alinhador Clustal Ômega. O alinhamento foi aparado com a ferramenta trimAl. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.

Tree scale: 1

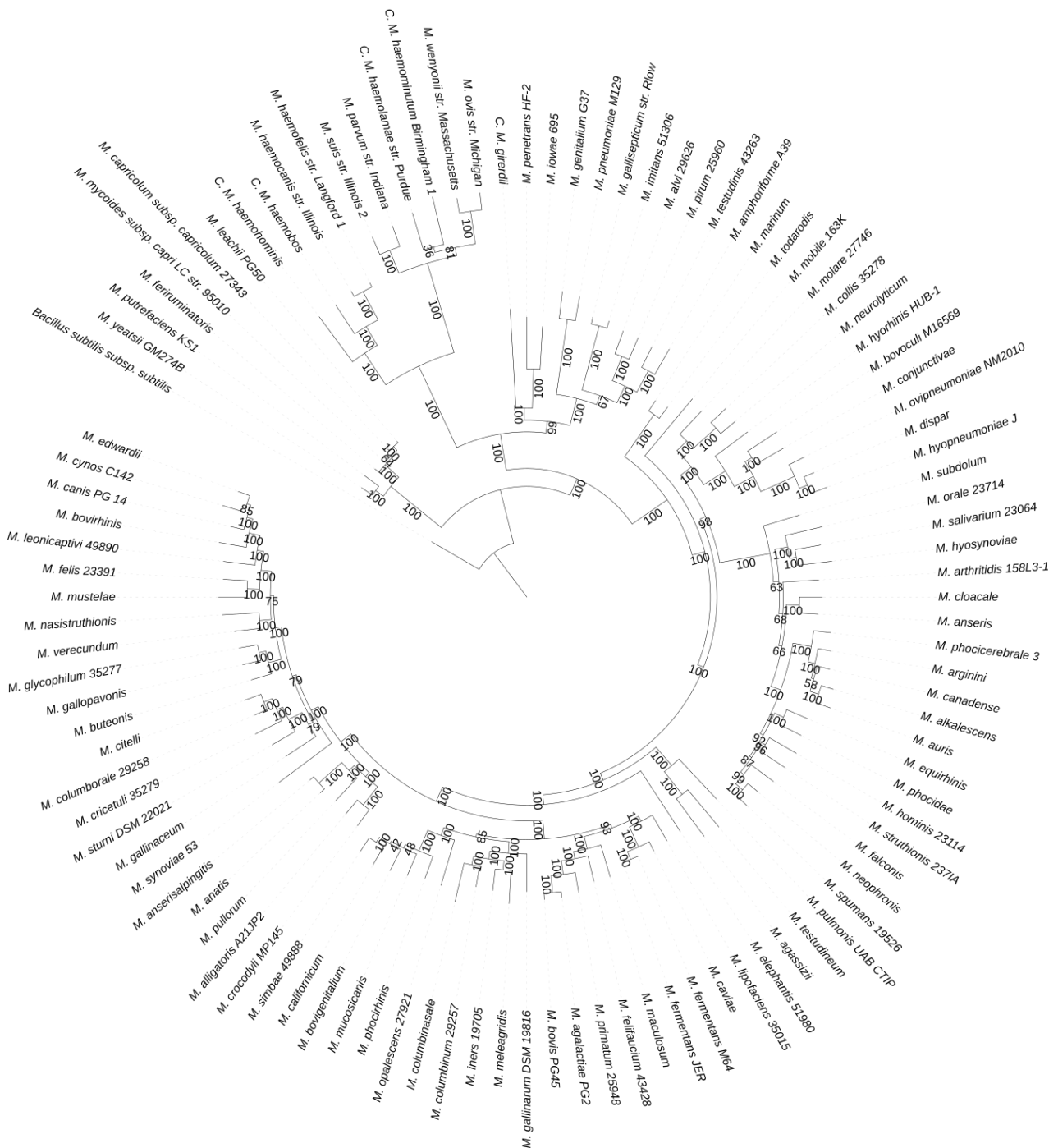
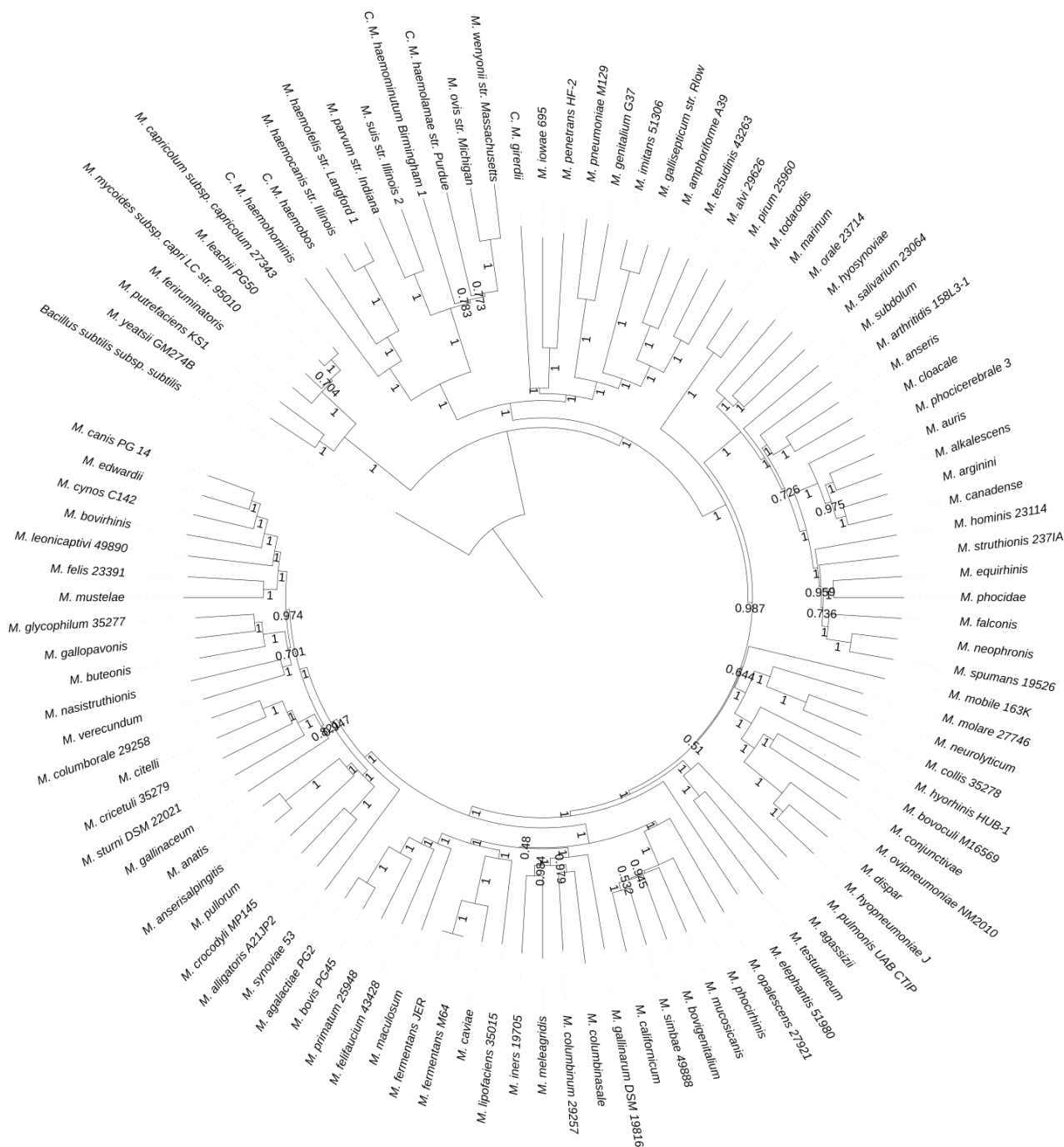


Figura A11. Árvore filogenética inferida pelo método de máxima verossimilhança com a ferramenta RAxML. O alinhamento foi realizado com o PRANK e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.





Tree scale: 0.1



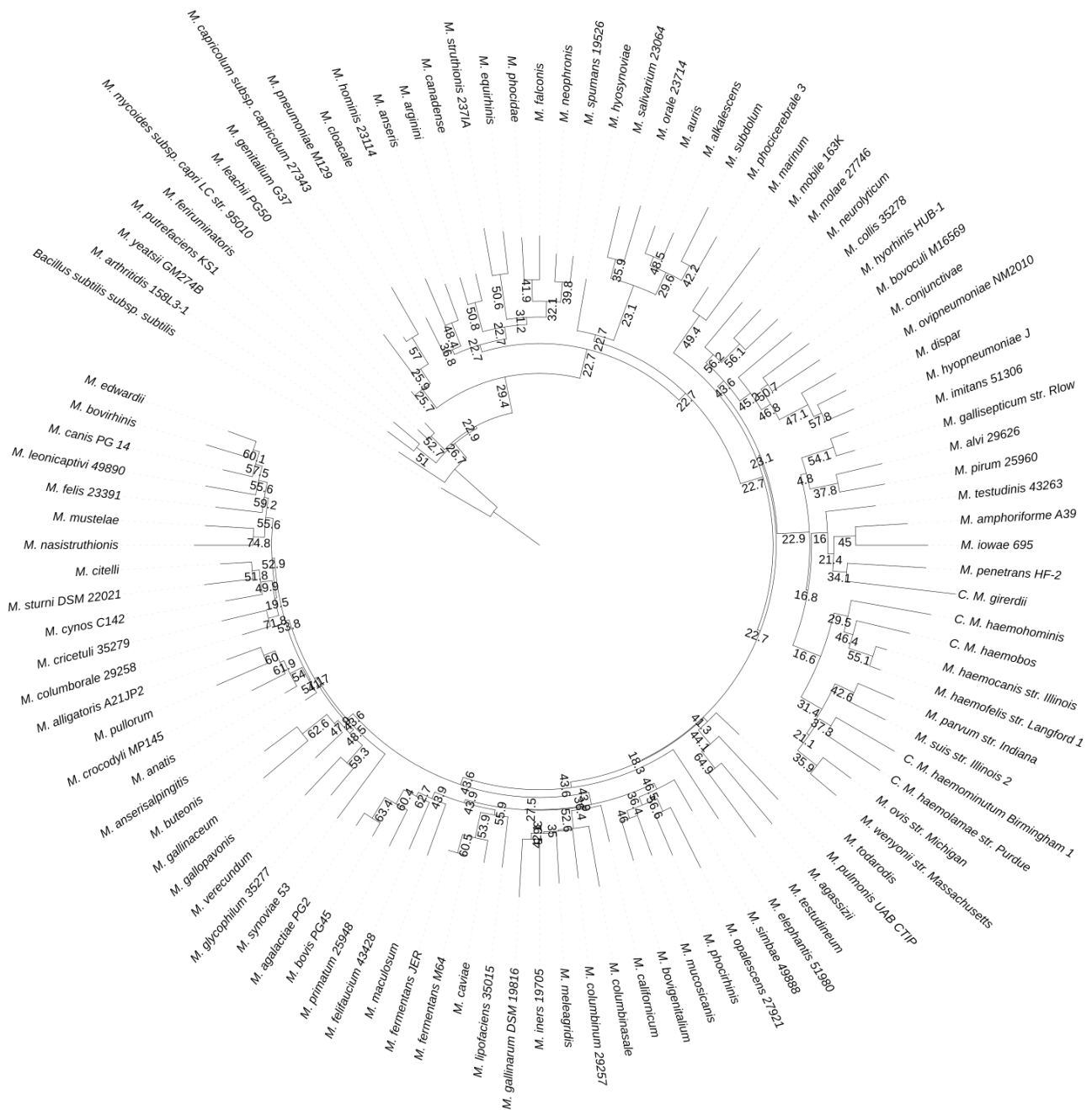
**Figura A13. Árvore filogenética inferida pelo método de distância com a ferramenta MEGA. O alinhamento foi realizado com o Clustal Ômega e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.**







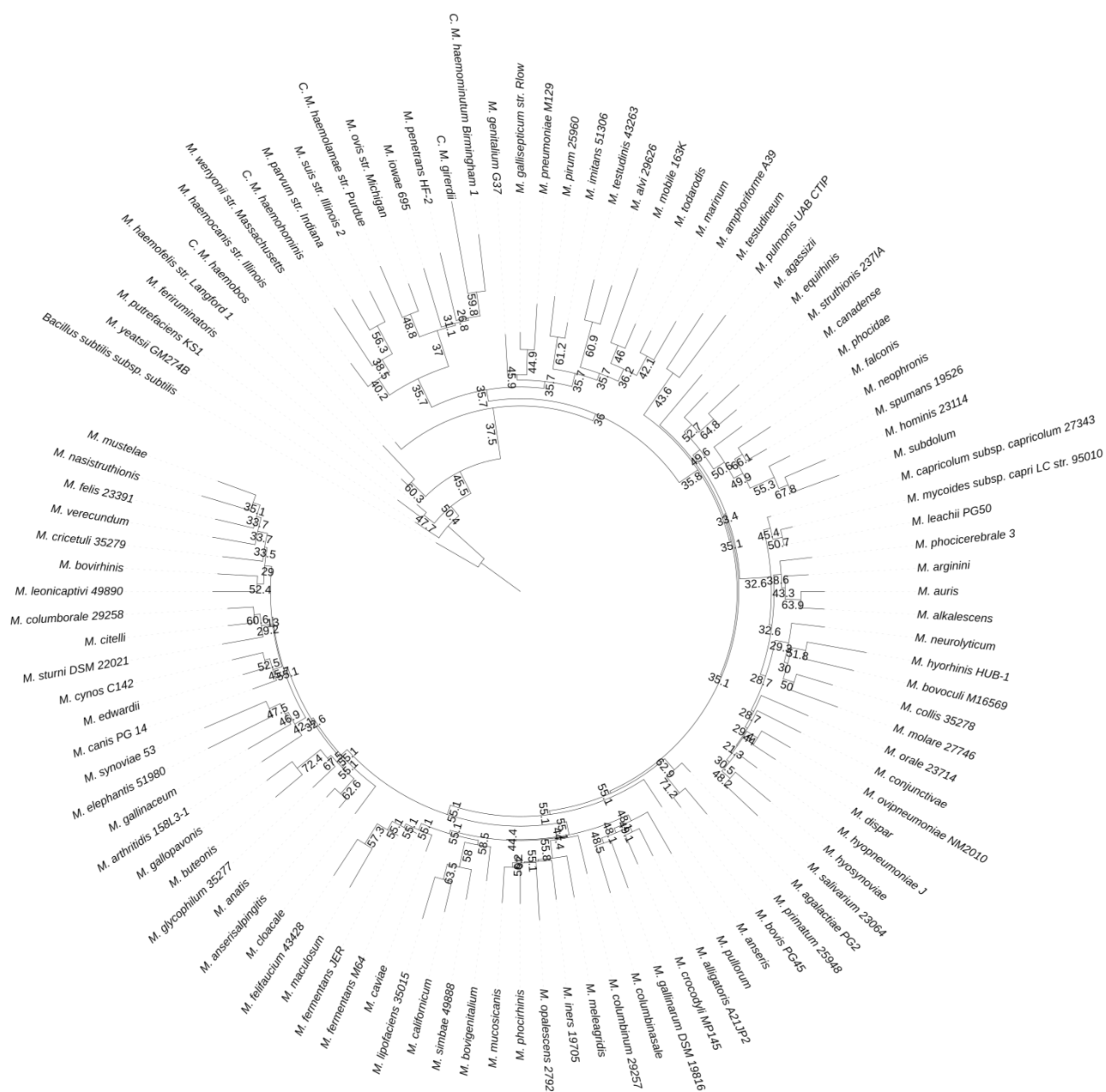
Tree scale: 0.1



**Figura A17. Árvore filogenética inferida pelo método de distância com a biblioteca Bio.Phylo. O alinhamento foi realizado com o Clustal Ômega e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.**



Tree scale: 0.1



**Figura A19.** Árvore filogenética inferida pelo método de distância com a biblioteca Bio.Phylo. O alinhamento foi realizado com o PRANK e utilizado completo. O suporte dos ramos está indicado em cada bifurcação. Foram utilizadas 1000 replicatas.





## APÊNDICE 4

As Tabelas de A5 a A8 apresentam detalhes das comparações topológicas realizadas com as 16 árvores filogenéticas geradas com o Pipelution e as árvores referenciadas (GUPTA *et al.*, 2018). Quando necessário é utilizado como referência o estudo de publicação da espécie. Alguns dados levantados na comparação estão descritos abaixo. Eles se aplicam a todas as tabelas (A5, A6, A7 e A8).

Bifurcações / OTUs envolvidas:

- B7 E C6 / Candidatus *M. haemolamae* e Candidatus *M. haemominutum*: Entre as árvores referência não há consistência sobre qual espécie é mais basal, Candidatus *M. haemolamae* ou Candidatus *M. haemominutum*.
- D1 E E5 / *M. imitans* e *M. gallisepticum*: Na filogenia referenciada, *M. imitans* e *M. gallisepticum* são mais basais (formam um grupo irmão) em relação a *M. genitallium* e *M. pneumoniae*.
- G4 / Ordem clado III>II>I: Clados I e II deveriam ser mais agrupados entre si em relação ao clado III. Em outras palavras, o clado III deveria estar mais basal em relação ao clado II em uma bifurcação. A ordem é: III > II > I.
- F4 / *M. mobile*: *M. mobile* deveria estar ou dentro do clado II (nas primeiras bifurcações) ou mais basal em relação aos clados I e II.
- H4 / *M. arthritidis*: Nas árvores referenciadas, *M. arthritidis* é mais basal que *M. cloacale*.
- M8 / *M. arginini*: *M. arginini* está posicionado no clado I nas árvores referenciadas. Mas *M. arginini* também tem proteínas homólogas com espécies do clado II, sem posicionamento fixo.
- R8, S9 E T2 / *M. mucosicanis*: no artigo de publicação da espécie, *M. mucosicanis* é mais basal em relação a *M. phocirhinis* (SPERGSEER *et al.*, 2011)
- O3 / Topologia no primeiro clado dentro do grupo Hominis: Na filogenia de referência o clado de *M. opalescens* é mais basal em relação a todas as outras espécies desta quarta bifurcação (clado I) dentro do grupo Hominis.
- U4, W8 E X4 / *M. verecundum* e *M. nasistruthionis*: Na filogenia referência não há consenso sobre o posicionamento de *M. verecundum*. A espécie está tanto basal em relação a *M. buteonis* e *M. glycyphilum*, como agrupada com *M. felis*. *M. nasistruthionis* está agrupada com *M. verecundum* na filogenia publicada da espécie (SPERGSEER *et al.*, 2020).
- N/A / *M. synoviae*: Nas árvores referenciadas, *M. synoviae* está uma bifurcação antes do clado de *M. gallinaceum*.
- N/A / *M. canis*: *M. canis* está agrupado com *M. cynos* nas árvores filogenéticas de referência. Na literatura, *M. canis* agrupa com *M. edwardii* (CHALKER; BROWNLIE, 2004).

**Tabela A5: Descrição das diferenças encontradas nas árvores de número 1 a 4 em relação às árvores filogenéticas inferidas por Gupta.** Estão descritas as diferenças topológicas em relação ao posicionamento das espécies, as quais estão mapeadas em nós identificados conforme descrito no corpo do texto, quando possível. Posicionamentos em negrito representam nós com suporte inferior a 80%.

Método	IB (Mr. Bayes)			
	Alinhamento	Clustal completo	Clustal aparado	PRANK completo
Árvore	1	2	3	4
Suporte < 80%	3	4	0	0
Suporte < 90%	3	4	2	2
<b>Bifurcações / OTU envolvida</b>				
<b>B7 E C6 / Candidatus <i>M. haemolamae</i> e Candidatus <i>M. haemominutum</i></b>	Candidatus <i>M. haemolamae</i> mais basal	Candidatus <i>M. haemolamae</i> mais basal	Agrupados	Candidatus <i>M. haemominutum</i> mais basal
<b>D1 E E5 / <i>M. imitans</i> e <i>M. gallisepticum</i></b>	Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i>	<b>Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i></b>	Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i>	Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i>
<b>G4 / Ordem clado III&gt;II&gt;I</b>	II > III	II > III	II > III	II > III
<b>F4 / <i>M. mobile</i></b>	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III
<b>H4 / <i>M. arthritidis</i></b>	<b>Em um nó mais basal que <i>M. orale</i></b>	<b>Em um nó mais basal que <i>M. cloacale</i></b>	Em um nó mais basal que <i>M. cloacale</i>	Em um nó mais basal que <i>M. cloacale</i>
<b>M8 / <i>M. arginini</i></b>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>
<b>R8, S9 E T2 / <i>M. mucosicanis</i></b>	<b>Agrupado com <i>M. phocirhinis</i></b>	Basal em relação ao clado (4 OTUs)	Basal em relação ao clado (4 OTUs)	Agrupado com <i>M. phocirhinis</i>
<b>O3 / Topologia no primeiro clado dentro do grupo Hominis</b>	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal
<b>U4, W8 E X4 / <i>M. verecundum</i> e <i>M. nasistruthionis</i></b>	Agrupado com <i>M. felis</i> e <i>M. mustalae</i>	<b>Agrupado com <i>M. felis</i> e <i>M. mustalae</i></b>	Agrupado com <i>M. felis</i> e <i>M. mustalae</i>	Agrupado com <i>M. buteonis</i> e <i>M. glycyphilum</i>
<b>N/A / <i>M. synoviae</i></b>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>
<b>N/A / <i>M. canis</i></b>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>
<b>Nº de ramos sem suporte que podem impactar a topologia</b>	2	3	0	0

**Tabela A6: Descrição das diferenças encontradas nas árvores de número 5 a 8 em relação às árvores filogenéticas inferidas por Gupta.** Estão descritas as diferenças topológicas em relação ao posicionamento das espécies, as quais estão mapeadas em nós identificados conforme descrito no corpo do texto, quando possível. Posicionamentos em negrito representam nós com suporte inferior a 80%.

Método	MV (PhyML - SH-like)			
	Alinhamento	Clustal completo	Clustal aparado	PRANK completo
Árvore	5	6	7	8
Suporte < 80%	7	7	3	3
Suporte < 90%	13	10	6	6
<b>Bifurcações / OTU envolvida</b>				
<b>B7 E C6 / Candidatus <i>M. haemolamae</i> e Candidatus <i>M. haemominutum</i></b>	<b>Candidatus <i>M. haemolamae</i> mais basal</b>	<b>Candidatus <i>M. haemolamae</i> mais basal</b>	Agrupados	Agrupados
<b>D1 E E5 / <i>M. imitans</i> e <i>M. gallisepticum</i></b>	<b>Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i></b>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>	<b>Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i></b>	<b>Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i></b>
<b>G4 / Ordem clado III&gt;II&gt;I</b>	II > III	II > III	II > III	II > III
<b>F4 / <i>M. mobile</i></b>	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III
<b>H4 / <i>M. arthritidis</i></b>	Em um nó mais basal que <i>M. cloacale</i>	<b>Em um nó mais basal que <i>M. cloacale</i></b>	Em um nó mais basal que <i>M. cloacale</i>	Em um nó mais basal que <i>M. cloacale</i>
<b>M8 / <i>M. arginini</i></b>	<b>Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i></b>	<b>Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i></b>	<b>Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i></b>	<b>Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i></b>
<b>R8, S9 E T2 / <i>M. mucosicanis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	Agrupado com <i>M. phocirhinis</i>	Agrupado com <i>M. phocirhinis</i>
<b>O3 / Topologia no primeiro clado dentro do grupo Hominis</b>	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal
<b>U4, W8 E X4 / <i>M. verecundum</i> e <i>M. nasistruthionis</i></b>	<b>um grupo mais basal que <i>M. buteonis</i> e <i>M. glycyphilum</i></b>	um grupo mais basal que <i>M. buteonis</i> e <i>M. glycyphilum</i>	<b>Agrupado com <i>M. felis</i> e <i>M. mustalae</i></b>	<b>Agrupado com <i>M. felis</i> e <i>M. mustalae</i></b>
<b>N/A / <i>M. synoviae</i></b>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>
<b>N/A / <i>M. canis</i></b>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>
<b>Nº de ramos sem suporte que podem impactar a topologia</b>	5	4	3	3

**Tabela A7: Descrição das diferenças encontradas nas árvores de número 9 a 12 em relação às árvores filogenéticas inferidas por Gupta.** Estão descritas as diferenças topológicas em relação ao posicionamento das espécies, as quais estão mapeadas em nós identificados conforme descrito no corpo do texto, quando possível. Posicionamentos em negrito representam nós com suporte inferior a 80%.

Método	MV (RAxML)			
	Clustal completo	Clustal aparado	PRANK completo	PRANK aparado
Árvore	9	10	11	12
Suporte < 80%	12	11	12	10
Suporte < 90%	17	18	16	16
<b>Bifurcações / OTU envolvida</b>				
<b>B7 E C6 / Candidatus <i>M. haemolamae</i> e Candidatus <i>M. haemominutum</i></b>	<b>Candidatus <i>M. haemolamae</i> mais basal</b>	<b>Candidatus <i>M. haemolamae</i> mais basal</b>	<b>Agrupados</b>	<b>Agrupados</b>
<b>D1 E E5 / <i>M. imitans</i> e <i>M. gallisepticum</i></b>	<b>Grupo mais basal que <i>M. genitalium</i> e <i>M. pneumoniae</i></b>	<b>Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i></b>	<b>Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i></b>	Em ramo mais terminal que <i>M. genitalium</i> e <i>M. pneumoniae</i>
<b>G4 / Ordem clado III&gt;II&gt;I</b>	II > III	II > III	II > III	II > III
<b>F4 / <i>M. mobile</i></b>	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III	Basal em relação aos Clados II e III
<b>H4 / <i>M. arthritidis</i></b>	<b>Em um nó mais basal que <i>M. cloacale</i></b>	<b>Em um nó mais basal que <i>M. cloacale</i></b>	<b>Em um nó mais basal que <i>M. cloacale</i></b>	<b>Em um nó mais basal que <i>M. cloacale</i></b>
<b>M8 / <i>M. arginini</i></b>	Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i>	Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i>	Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i>	Posicionado no Clado III mas mais próximo de <i>M. phocicerebrale</i>
<b>R8, S9 E T2 / <i>M. mucosicanis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>
<b>O3 / Topologia no primeiro clado dentro do grupo Hominis</b>	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal	Errado, sem clado de <i>M. opalescens</i> ser mais basal
<b>U4, W8 E X4 / <i>M. verecundum</i> e <i>M. nasistruthionis</i></b>	<b>E um grupo mais basal que <i>M. buteonis</i> e <i>M. glycyphilum</i></b>	E um grupo mais basal que <i>M. buteonis</i> e <i>M. glycyphilum</i>	<b>Agrupado com <i>M. felis</i> e <i>M. mustalae</i></b>	<b>Agrupado com <i>M. felis</i> e <i>M. mustalae</i></b>
<b>N/A / <i>M. synoviae</i></b>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>	Grupo irmão mais basal do clado de <i>M. gallinaceum</i> e <i>M. verecundum</i>
<b>N/A / <i>M. canis</i></b>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>	Basal em relação a <i>M. edwardii</i>
<b>Nº de ramos sem suporte que podem impactar a topologia</b>	6	5	6	5

**Tabela A8: Descrição das diferenças encontradas nas árvores de número 13 a 16 em relação às árvores filogenéticas inferidas por Gupta.** Estão descritas as diferenças topológicas em relação ao posicionamento das espécies, as quais estão mapeadas em nós identificados conforme descrito no corpo do texto, quando possível. Posicionamentos em **negrito** representam nós com suporte inferior a 80%.

Método	Distância (MEGA)				
	Alinhamento	Clustal completo	Clustal aparado	PRANK completo	PRANK aparado
Árvore		13	14	15	16
Suporte < 80%		10	9	12	10
Suporte < 90%		11	13	13	13
<b>Bifurcações / OTU envolvida</b>					
<b>B7 E C6 / Candidatus <i>M. haemolamae</i> e Candidatus <i>M. haemominutum</i></b>	<b>Candidatus <i>M. haemominutum</i> mais basal</b>	<b>Candidatus <i>M. haemominutum</i> mais basal</b>	<b>Candidatus <i>M. haemominutum</i> mais basal</b>	<b>Candidatus <i>M. haemominutum</i> mais basal</b>	<b>Candidatus <i>M. haemominutum</i> mais basal</b>
<b>D1 E E5 / <i>M. imitans</i> e <i>M. gallisepticum</i></b>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>	Em ramo mais terminal que <i>M. genitallium</i> e <i>M. pneumoniae</i>
<b>G4 / Ordem clado III&gt;II&gt;I</b>	<b>III&gt;II&gt;I, mas com muitas alterações topológicas internas no clado III</b>	<b>III&gt;II&gt;I, mas com muitas alterações topológicas internas no clado III</b>	<b>III&gt;II&gt;I, mas com muitas alterações topológicas internas no clado III</b>	<b>III&gt;II&gt;I, mas com muitas alterações topológicas internas no clado III</b>	<b>III&gt;II&gt;I, mas com muitas alterações topológicas internas no clado III</b>
<b>F4 / <i>M. mobile</i></b>	<b>Basal em relação ao Clado II</b>	<b>Basal em relação ao Clado II</b>	Basal em relação aos Clados II e III	<b>Basal em relação ao Clado II</b>	<b>Basal em relação ao Clado II</b>
<b>H4 / <i>M. arthritidis</i></b>	Em um nó mais basal que <i>M. cloacale</i> . Com muitas alterações topológicas internas no clado III	Em um nó mais basal que <i>M. cloacale</i> . Com muitas alterações topológicas internas no clado III	Em um nó mais basal que <i>M. cloacale</i> . Com muitas alterações topológicas internas no clado III	Em um nó mais basal que <i>M. cloacale</i> . Com muitas alterações topológicas internas no clado III	Em um nó mais basal que <i>M. cloacale</i> . Com muitas alterações topológicas internas no clado III
<b>M8 / <i>M. arginini</i></b>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>	Posicionado no Clado III agrupado com <i>M. canadense</i>
<b>R8, S9 E T2 / <i>M. mucosicanis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i>.</b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Mais terminal que <i>M. phocirhinis</i></b>	<b>Basal em relação ao clado (4 OTUs)</b>
<b>O3 / Topologia no primeiro clado dentro do grupo Hominis</b>	Correto, com clado de <i>M. opalescens</i> posicionado mais basalmente	Errado, com clado de <i>M. gallinarum</i> mais basal	Correto, com clado de <i>M. opalescens</i> posicionado mais basalmente	Correto, com clado de <i>M. opalescens</i> posicionado mais basalmente	Correto, com clado de <i>M. opalescens</i> posicionado mais basalmente
<b>U4, W8 E X4 / <i>M. verecundum</i> e <i>M. nasistruthionis</i></b>	<b>Agrupado com <i>M. buteonis</i> e <i>M. glycyophilum</i></b>	<b>Agrupado com <i>M. buteonis</i> e <i>M. glycyophilum</i></b>	<b>Agrupado com <i>M. buteonis</i> e <i>M. glycyophilum</i></b>	<b>Agrupado com <i>M. buteonis</i> e <i>M. glycyophilum</i></b>	<b>Agrupado com <i>M. buteonis</i> e <i>M. glycyophilum</i></b>
<b>N/A / <i>M. synoviae</i></b>	Grupo irmão mais basal de <i>M. crocodyli</i> e <i>M. alligatoris</i>	Grupo irmão mais basal de <i>M. crocodyli</i> e <i>M. alligatoris</i>	Grupo irmão mais basal de <i>M. crocodyli</i> e <i>M. alligatoris</i>	Grupo irmão mais basal de <i>M. crocodyli</i> e <i>M. alligatoris</i>	<b>Grupo irmão mais basal de <i>M. crocodyli</i> e <i>M. alligatoris</i></b>
<b>N/A / <i>M. canis</i></b>	Agrupado com <i>M. edwardii</i>	Agrupado com <i>M. edwardii</i>	Agrupado com <i>M. edwardii</i>	Agrupado com <i>M. edwardii</i>	<b>Agrupado com <i>M. edwardii</i></b>
<b>Nº de ramos sem suporte que podem impactar a topologia</b>	5	5	4	7	7

## APÊNDICE 5

Tree scale: 1



### bootstrap

- 34
- 45.5
- 57
- 68.49
- 79.99

Alinhador	Alinhamento	Distância - MEGA	IB - Mr. Bayes	ML - PhyML (SH)	ML - RAxML
Clustal Ômega	Completo	●	●	●	●
	Aparado	●	●	●	●
PRANK	Completo	●	N/A	●	●
	Aparado	●	N/A	●	●

**Figura A21. Detalhes adicionais de suporte das 16 árvores filogenéticas analisadas.** A topologia apresentada é referente a árvore filogenética de número 3. Cada círculo colorido representa uma bifurcação que obteve suporte inferior a 80% em no mínimo uma árvore filogenética. No total estão posicionados 113 círculos coloridos de acordo com metodologia utilizada (esquematizadas na tabela inferior). As bifurcações sem círculo apresentaram suporte acima de 80% em todas as árvores filogenéticas. Os círculos estão em escala de *bootstrap*.